



UNIVERSIDADE DE  
COIMBRA



UNIVERSIDADE DE  
COIMBRA



**Proceedings of the  
18th European Conference on  
Cyber Warfare and Security**  
**University of Coimbra**  
**Portugal**  
**4-5 July 2019**



**Edited by**  
**Tiago Cruz and Paulo Simoes**

**acpi**

A conference managed by ACPI, UK

**Proceedings of the**

**18th European Conference on**

**Cyber Warfare and Security**

**ECCWS 2019**

**Hosted By**

**University of Coimbra**

**Portugal**

**4-5 July 2019**

**Edited by**

**Tiago Cruz and Paulo Simoes**

Copyright The Authors, 2019. All Rights Reserved.

No reproduction, copy or transmission may be made without written permission from the individual authors.

#### **Review Process**

Papers submitted to this conference have been double-blind peer reviewed before final acceptance to the conference. Initially, abstracts were reviewed for relevance and accessibility and successful authors were invited to submit full papers. Many thanks to the reviewers who helped ensure the quality of all the submissions.

#### **Ethics and Publication Malpractice Policy**

ACPIL adheres to a strict ethics and publication malpractice policy for all publications – details of which can be found here:

<http://www.academic-conferences.org/policies/ethics-policy-for-publishing-in-the-conference-proceedings-of-academic-conferences-and-publishing-international-limited/>

#### **Conference Proceedings**

The Conference Proceedings is a book published with an ISBN and ISSN. The proceedings have been submitted to a number of accreditation, citation and indexing bodies including Thomson ISI Web of Science and Elsevier Scopus.

Author affiliation details in these proceedings have been reproduced as supplied by the authors themselves.

The Electronic version of the Conference Proceedings is available to download from DROPBOX

<https://tinyurl.com/ECCWS19>. Select Download and then Direct Download to access the Pdf file. Free download is available for conference participants for a period of 2 weeks after the conference.

The Conference Proceedings for this year and previous years can be purchased from

[www.academic-bookshop.com/](http://www.academic-bookshop.com/)

E-Book ISBN: 978-1-912764-29-7

E-Book ISSN: 2048-8610

Book version ISBN: 978-1-912764-28-0

Book Version ISSN: 2048-8602

Published by Academic Conferences and Publishing International Limited

Reading

UK

44-118-972-4148

[www.academic-conferences.org](http://www.academic-conferences.org)

## Contents

Paper Title	Author(s)	Page No
<b>Preface</b>		vii
<b>Committee</b>		viii
<b>Biographies</b>		x
<b>Research papers</b>		
Using Hypervisors to Overcome Structured Exception Handler Attacks	Asaf Algawi, Michael Kiperberg, Roee Leon and Nezer Zaidenberg	1
Creating Modern Blue Pills and Red Pills	Asaf Algawi, Michael Kiperberg, Roee Leon, Amit Resh and Nezer Zaidenberg	6
Information Technology Governance: The Role of Board of Directors in Cybersecurity Oversight	Abdalmuttaleb Al-Sartawi	15
Revisiting Cyber Definition	Riza Azmi and Kautsarina	22
Secure Application for Health Monitoring	Vijay Bhuse and Harsh Sinha	31
Crossed Swords: A Cyber Red Team Oriented Technical Exercise	Bernhards Blumbergs, Rain Ottis and Risto Vaarandi	37
The far Right of the UK and Ukraine's Views About European Integration: An Analysis of Online Political Discourse	Radomir Bolgov and Olga Filatova	45
A Comparison of Chat Applications in Terms of Security and Privacy	Johnny Botha, Carien Van't Wout and Louise Leenen	55
Detection of Premeditated Security Vulnerabilities in Mobile Applications	Agné Brilingaitė, Linas Bukauskas and Eduardas Kutka	63
Online Glossary of Cyber Security	Ladislav Burita	72
A Framework for Managing Cybersecurity Effectiveness in the Digital Context	Marian Carcary, Eileen Doherty and Gerry Conway	78
Personal Data Protection (PDP): A Conceptual Framework for Organisational Management of Personal Data in the Digital Context	Marian Carcary, Eileen Doherty and Gerry Conway	87
A Capability Approach to Managing Organisational Information Security	Marian Carcary, Eileen Doherty and Gerry Conway	97
Leveraging the OODA Loop with Digital Analytics to Counter Disinformation	Jami Carroll	106
A Framework for Improved Home Network Security	António Craveiro, Ana Oliveira, Jorge Proença, Tiago Cruz and Paulo Simões	114
Hackers and the Military: How to Recruit and Manage Hidden Talents?	Didier Danet	124
Artificial Intelligence Cybersecurity Framework: Preparing for the Here and now With AI	Emily Darraj, Char Sample and Connie Justice	132
A Performance Study of the Ethical Hacking Capabilities of Single Board Computers	Adolfo Jose Arias de la Vega and Christina Thorpe	142

Paper Title	Author(s)	Page No
Investigation and Surveillance on the Darknet: An Architecture to Reconcile Legal Aspects With Technology	Maxence Delong, Eric Filiol and Baptiste David	151
Cybersecurity Assessment of the Public Sector in Greece	George Drivas, Leandros Maglaras, Helge Janicke and Sotiris Ioannidis	162
Improving Phishing Awareness in the United States Department of Defense	Christopher Dukarm, Richard Dill and Mark Reith	172
Putting Together the Pieces: A Concept for Holistic Industrial Intrusion Detection	Simon Duque Antón and Hans Dieter Schotten	178
A Cyber Counterintelligence Matrix for Outsmarting Your Adversaries	Petrus Duvenage, Victor Jaquire and Sebastian von Solms	187
The Offensive Cyber Operations Playbook	Dennis Granåsen and Margarita Jaitner	194
Building a (French) National Priority: The Grammar of the French Cyber Defence Strategy	Saïd Haddad	202
Where Cyber Meets the Electromagnetic Spectrum	Gerhard Henselmann and Martti Lehto	209
Protection of Critical Infrastructure in National Cyber Security Strategies	Eduardo Izycki and Rodrigo Colli	219
Towards a Framework for the Selection and Prioritisation of National Cybersecurity Functions	Pierre Jacobs, Sebastiaan von Solms, Marthie Grobler and Brett van Niekerk	229
Information Security Management Scaffold for Mobile Money Systems in Uganda	Fredrick Kanobe, Margaret Patricia Alexander and Kelvin Joseph Bwalya	239
Optimal Sensor Placement in Network Topology From the Defence Point of View	Vesa Kuikka and Juha-Pekka Nikkarila	248
Operator Impressions of 3D Visualizations for Cybersecurity Analysts	Kaur Kullman, Noam Ben Asher and Char Sample	257
The Balanced Digitalization and Digital Security: Case of Regional Authorities	Tuija Kuusisto and Rauno Kuusisto	267
Operational Tempo in Cyber Operations	Antoine Lemay and Sylvain Leblanc	275
In Search of a Social Contract for Cybersecurity	Andrew Liaropoulos	282
The Importance of Strategic Leadership in Cyber Security: Case of Finland	Jarno Limnälä and Martti Lehto	288
PhySec in Cellular Networks: Enhancing Security in the IIoT	Christoph Lipps, Mathias Strufe, Sachinkumar Bavikatti Mallikarjun and Hans Dieter Schotten	297
Technical Guidelines for Evaluating and Selecting Data Sources for Cybersecurity Threat Intelligence	Jabu Mtsweni and Muyowa Mutemwa	305
Cyberpsychological Threat Intelligence	Julie Murphy and Anthony Keane	314
Strategic Foresight and Resilience Through Cyber-Wargaming	David Ormrod and Keith Scott	319

Paper Title	Author(s)	Page No
The Persuasion Game: Developing a Serious Game Based Model for Information Warfare and Influence Studies	David Ormrod, Keith Scott, Lynn Scheinman, Thorsten Kodalle, Char Sample and Benjamin Turnbull	328
Cyber Attribution 2.0: Capture the False Flag	Timea Pahi and Florian Skopik	338
Operational Risk Assessment on Internet of Things: Mitigating Inherent Vulnerabilities	Youngjun Park, Mark Reith and Barry Mullins	346
Cyber Security of Vehicle CAN bus	Jouni Pöyhönen, Pyry Kotilainen, Janne Poikolainen, Janne Kalmari and Pekka Neittaanmäki	354
How to Apply Privacy by Design in OSINT and big Data Analytics?	Jyri Rajamäki and Jussi Simola	364
Research Challenges for Cybersecurity and Cyberwarfare: A South African Perspective	Trishana Ramluckan, Brett van Niekerk and Louise Leenen	372
Cyber-Influence Operations: A Legal Perspective	Trishana Ramluckan, Alicia Wanless and Brett van Niekerk	379
Online Expression and Spending on Personal Cybersecurity	Juhani Rauhala, Pasi Tyrväinen and Nezer Zaidenberg	387
Does Time Spent on Device Security and Privacy Inhibit Online Expression?	Juhani Rauhala, Pasi Tyrväinen and Nezer Zaidenberg	394
The Need for More Sophisticated Cyber-Physical Systems war Gaming Exercises	Aunshul Rege and Joe Adams	403
Western World Order in the Crosshairs? A Theoretical Review and Application of the Russian ‘Information Weapon’	Mari Ristolainen and Juha Kukkola	412
A Cross-Discipline Approach to Countering 4th Generation Espionage	Char Sample, Keith Scott and Emily Darraj	420
A Novel Intrusion Detection System Architecture for Internet of Things Networks	Leonel Santos, Ramiro Gonçalves and Carlos Rabadão	428
An Analysis of Security Features on Web Browsers	Siddhartha Sengupta and Joon Park	436
Alert Correlation Using Diamond Model for Cyber Threat Intelligence	Youngsup Shin, Changwan Lim, Mookyu Park, Sungyoung Cho, Insung Han, Haengrok Oh and Kyungho Lee	444
Digital Dematerialisation in the Portuguese tax System	Ana Paula Silva and Susana Aldeia	451
Information Security is More Than Just Policy; It is in Your Personality	Petteri Simola, Toni Virtanen and Miika Sartonen	459
Cyber Security Risk Modelling and Assessment: A Quantitative Approach	Abderrahmane Sokri	466
Taxonomies of Cybercrime: An Overview and Proposal to be Used in Mapping Cyber Criminal Journeys	Tiia Somer	475
On Neutrality and Cyber Defence	Marcel Stoltz	484

Paper Title	Author(s)	Page No
Integrative Approach to Understand Vulnerabilities and Enhance the Security of Cyber-Bio-Cognitive-Physical Systems	Todor Tagarev, Nikolai Stoianov and George Sharkov	492
Cyber Resilience Strategy and Attribution in the Context of International law	Pardis Moslemzadeh Tehrani	501
A Biological Framework for Characterizing Mimicry in Cyber-Deception	Steven Templeton, Matt Bishop, Karl Levitt and Mark Heckman	508
Agile Technology Development to Improve Scenario-Based Learning Exercises	Benjamin Turnbull, David Ormrod, Nour Moustafa and Nicholas Micallef	518
The Possibilities of Cyber Methods as Part of Maritime Warfare: Baltic Sea	Maija Turunen	527
Grasping Cybersecurity: A set of Essential Mental Models	Jan van den Berg	534
A Legal Perspective of the Cyber Security Dilemma	Brett van Niekerk and Trishana Ramluckan	544
An Analysis of Selected Cyber Intelligence Texts	Brett van Niekerk, Trishana Ramluckan and Petrus Duvenage	554
A Legal Understanding of State-Linked Cyberattacks and Malicious Cyber Activities	Murdoch Watney	560
A Methodology for the Comparative Analysis of Strategic Culture and Cyber Warfare	Andrew Williams	568
Cyber Warfare, Terrorist Narratives and Counter Terrorist Narratives: An Anticipatory Ethical Analysis	Richard Wilson	577
Cambridge Analytica, Facebook, and Influence Operations: A Case Study and Anticipatory Ethical Analysis	Richard Wilson	587
Information Warfare: Fabrication, Distortion and Disinformation: A Case Study and Anticipatory Ethical Analysis	Richard Wilson	596
ARM Security Alternatives	Raz Ben Yehuda, Roee Leon and Nezer Zaidenberg	604
IoT Architectures for Critical Infrastructures Protection	Alin Zamfiroiu, Bogdan Iancu, Catalin Boja, Tiberiu Georgescu and Cosmin Cartas	613
<b>Phd Research Papers</b>		621
A Conceptual Model for the Development of Cybersecurity Capacity in Mozambique	Martina Jennifer Zucule de Barros and Horst Lazarek	623
The Peculiarities of Securitising Cyberspace: A Multi-Actor Analysis of the Construction of Cyber Threats in the US (2003-2016)	Noran Shafik Fouad	633
E-Health Systems in Digital Environments	Aarne Hummelholm	641
Undersea Optical Cable Network and Cyber Threats	Aarne Hummelholm	650

Paper Title	Author(s)	Page No
The Current State of Research in Offensive Cyberspace Operations	Gazmend Huskaj	660
Cyber Sanctions: The Embargo of Flagged Data in a Geo-Cultural Internet	Ion Iftimie	668
The Transformation of Islamic Terrorism Through Cyberspace: The Case of ISIS	Eleni Kapsokoli	677
Protecting the Besieged Cyber Fortress: Russia's Response to Cyber Threats	Martti Kari	685
Government Efforts Toward Promoting IoT Security Awareness for end Users: A Study of Existing Initiatives	Kautsarina and Bayu Anggoro jati	692
Cyber Entanglement: A Framework for the Study of U.S.–China Relations	Karine Pontbriand	702
Effects of Cyber Domain in Crisis Management	Jussi Simola and Martti Lehto	710
A Data Security Framework for Cloud Computing Adoption: Mozambican Government Cloud Computing	Ambrósio Patrício Vumo, Josef Spillner and Stefan Köpsell	720
<b>Masters Research Papers</b>		731
A Technical Overview on the Usage of Cloud Encryption Services	Daniel Carvalho, João Morais, João Almeida, Pedro Martins, Carlos Quental and Filipe Caldeira	733
Description Logics and Axiom Formation for a Digital Forensics Ontology	Dagney Ellison, Adeyemi Richard Ikuesan and Hein Venter	742
Self-Directed Learning Tools in USAF Multi-Domain Operations Education	Nathaniel Flack and Mark Reith	752
Detecting Advanced Persistent Threat Malware Using Machine Learning-Based Threat Hunting	Tien-Chih Lin, Cheng-Chung Guo and Chu-Sing Yang	760
A Standardization Guide for Homomorphic Encryption Applications for Digital Forensic Investigations	Nnana Mogano	769
Fake News Detection Using Ensemble Machine Learning	Potsane Mohale and Wai Sze Leung	777
Synthetic Data Generation With Machine Learning for Network Intrusion Detection Systems	Marvin Newlin, Mark Reith and Mark DeYoung	785
Human Factors of Cyber Operations: Decision Making Behind Advanced Persistence Threat Operations	Veikko Siukonen	790
<b>Non Academic Papers</b>		799
Thoughts About a General Theory of Influence in a DIME/PMESII/ASCOP/IRC2 Model	Thorsten Kodalle, Char Sample, David Ormrod and Keith Scott	801
Success Factors and Pitfalls in Security Certifications	Helvi Salminen	811

<b>Paper Title</b>	<b>Author(s)</b>	<b>Page No</b>
Instilling Digital Citizenship Skills Through Education: A Malaysian Perspective	Ramona Susanty, Zahri Yunos, Mustaffa Ahmad and Norrizan Razali	819
Smart Citizens Wanted! How to act Responsibly With Data Security and Privacy?	Christine Ziske and Ulf Ziske	828
<b>Work In Progress Papers</b>		837
Towards Automated Threat-Based Risk Assessment for Cyber Security in Smarthomes	Pankaj Pandey, Anastasija Collen, Niels Nijdam, Marios Anagnostopoulos, Sokratis Katsikas and Dimitri Konstantas	839
Applicability of Resilience Metrics in the Context of Telecommunications Services	Tarja Rusi	845
Mitigating Return Oriented Programming	Lee Speakman, Thaddeus Eze, David Baker and Samuel Wairimu	852
US-Russian Relations in Cybersecurity: The Constructivist Dimension	Ilona Stadnik	858

## Preface

These proceedings represent the work of contributors to the 18th European Conference on Cyber Warfare and Security (ECCWS 2019), hosted by University of Coimbra, Portugal on 4 - 5 July 2019. The Conference Chair is Tiago Cruz and the Programme Chair is Paulo Simoes from University of Coimbra.

ECCWS is a well-established event on the academic research calendar and now in its 18th year, the key aim remains the opportunity for participants to share ideas and meet the people who hold them. The scope of papers will ensure an interesting two days. The subjects covered illustrate the wide range of topics that fall into this important and ever-growing area of research.

The opening keynote presentation is given by Dr. Leandros A. Maglaras, from the National Cyber Security Authority of Greece on the topic of *Protection of Critical National Infrastructures*. The second day of the conference will open with an address by Nuno Medeiros, EDP Distribuição, Lisbon, Portugal, who will talk about *Cyber Security Challenges for a Critical Infrastructure Operator*.

With an initial submission of 157 abstracts, after the double blind, peer review process there are 74 Academic research papers, 12 PhD research papers, 8 Masters Research papers, 4 non-academic papers and 4 work-in-progress papers published in these Conference Proceedings. These papers represent research from Australia, Austria, Bahrain, Brazil, Bulgaria, Canada, Czech Republic, Estonia, Finland, France, Germany, Greece, Indonesia, Ireland, Israel, Latvia, Lithuania, Malaysia, Mozambique, Netherlands, Norway, Portugal, Romania, Russia, South Africa, South Korea, Sweden, Taiwan, United Kingdom and the United States of America.

We hope you enjoy the conference.

Tiago Cruz and Paulo Simoes

University of Coimbra, Portugal

July 2019

## ECCWS Committee

Dr. Mohd Faizal Abdollah, University Technical Malaysia Melaka, Malaysia; Dr. Nasser Abouzakhar, University of Hertfordshire, UK; Dr William ("Joe") Adams, Univ of Michigan/Merit Network, USA; Dr. Tariq Ahamad, Prince Sattam Bin Abdulaziz University, Saudi Arabia; Prof Hamid Alasadi, Basra University, Iraq; Dr. Kari Alenius, University of Oulu, Finland; Chaminda Alocious, University of Hertfordshire, UK; Prof. Antonios Andreatos, Hellenic Air Force Academy, Greece; Dr. Olga Angelopoulou, University of Hertfordshire, UK; Dr. Leigh Armistead, Edith Cowan University, Australia; Colin Armstrong, Curtin University, Australia, Australia; Johnnes Arreymbi, University of East London, UK; Dr. Hayretdin Bahsi, Tallinn University of Technology, Estonia; Dr. Darya Bazarkina, Sholokhov Moscow State Humanitarian University, Russia; Ass Prof. Maumita Bhattacharya, Charles Sturt University, Australia; Mr Robert Bird, Coventry University, UK; Prof. Matt Bishop, University of California at Davis, USA; Andrew Blyth, University of Glamorgan, UK; Dr Radomir Bolgov, Saint Petersburg State University, Russia; Colonel (ret) Colin Brand, Graduate School of Business Leadership, South Africa; Dr. Svet Braynov, University of Illinois at Springfield, USA; Prof. Larisa Breton, FullCircle Communications, LLC, USA; Bill Buchanan, Napier University, UK; Dr Jim Chen, DoD National Defense University, USA; Bruce Christianson, University of Hertfordshire, UK; Dr. Maura Conway, Dublin City University, Ireland; Dr. Paul Crocker, Universidade de Beira Interior, Portugal; Prof. Tiago Cruz, University of Coimbra, Portugal; Dr. Christian Czosseck, CERT Bundeswehr (German Armed Forces CERT), Germany; Geoffrey Darnton, Bournemouth University, UK; Josef Demergis, University of Macedonia, Greece; Associate Dean Paul Dowland, Edith Cowan University, Australia; Marios Efthymiopoulos, Political Science Department University of Cyprus, Cyprus; Dr. Colin Egan, University of Hertfordshire, Hatfield, UK; Dr Ruben Elamiryan, Public Administration Academy of the Republic of Armenia, Armenia; Daniel Eng, C-PISA/HTCIA, China; Prof. Dr. Alptekin Erkollar, ETCOP, Austria; John Fawcett, University of Cambridge, UK; Prof. Eric Filiol, Ecole Supérieure en Informatique, Electronique et Automatique, France; Dr. Chris Flaherty, University of New South Wales, Australia; Prof. Steve Furnell, University of Plymouth, UK; Ass. Prof. Javier Garci'a Villalba, Universidad Complutense de Madrid, Spain; Prof Bela Genge, Petru Maior University of Tigră Mures, Romania; Mr. Tushar Gokhale, Hewlett Packard Enterprise, USA; Dr. Michael Grimalia, Air Force Institute of Technology, USA; Prof. Stefanos Gritzalis, University of the Aegean, Greece; Dr. Mils Hills, Northampton Business School, UK; Ulrike Hugl, University of Innsbruck, Austria; Aki Huhtinen, National Defence College, Finland; Mrs Rose Hunt, University of Wolverhampton, UK; Bill Hutchinson, Edith Cowan University, Australia; Dr. Berg Hyacinthe, State University of Haiti, Haiti; Dr. Abhaya Induruwa, Canterbury Christ Church University, UK; Dr. Md Ruhul Islam, Sikkim Manipal Institute of Technology, India; Hamid Jahankhani, University of East London, UK; Dr. Helge Janicke, De Montfort University, UK; Joey Jansen van Vuuren, CSIR, South Africa; Saara Jantunen, University of Helsinki, Finland; Dr. Audun Josang, University of Oslo, Norway; James Joshi, University of Pittsburgh, USA; Nor Badrul Anuar Jumaat, University of Malaya, Malaysia; Maria Karyda, University of the Aegean, Greece; Ass. Prof. Vasilis Katos , Democritus University of Thrace, Greece; Dr. Anthony Keane, Technological University Dublin, Ireland; Jyri Kivimaa, Cooperative Cyber Defence and Centre of Excellence, Tallinn, Estonia; Prof. Ahmet Koltukuz, Yasar University, Dept. of Comp. Eng, Turkey; Theodoros Kostis, Hellenic Army Academy, Greece; Prashant Krishnamurthy, University of Pittsburgh, USA; Mr. Peter Kunz, HiSolutions AG, Germany; Dr. Erikk Kurkinen, University of Jyväskylä, Finland; Takakazu Kurokawa, National Defence Acadamy, Japan; Rauno Kuusisto, Finnish Defence Force, Finland; Dr. Laouamer Lamri, Al Qassim University and European University of Brittany, Saudi Arabia; Martti Lehto, National Defence University, Finland; Mr Trupil Limbasiya, NIIT University, Neemrana, Rajasthan, India; Dr Efstratios Livanis, University of Macedonia, Greece; Peeter Lorents, CCD COE, Tallinn, Estonia; James Malcolm, University of Hertfordshire, UK; Dr Mary Manjikian, Regent University, USA; Mario Marques Freire, University of Beira Interior, Covilhā, Portugal; Ioannis Mavridis , University of Macedonia, Greece; Rob McCusker, Teeside University, Middlesborough, UK; Dr Imran Memon, zhejiang university, china; Dr Shahzad Memon, University of Sindh, Pakistan; Jean-Pierre Molton Michel, Ministry of Agriculture, Haiti; Durgesh Mishra, Acropolis Institute of Technology and Research, India; Dr. Yonathan Mizrachi, University of Haifa, Israel; Edmundo Monteiro, University of Coimbra, Portugal; Dr Pardis Moslemzadeh Tehrani, University of Malaya, Malaysia; Evangelos Moustakas, Middlesex University, London, UK; Antonio Muñoz, University of Málaga , Spain; Dr. Kara Nance, University of Alaska Fairbanks, USA; Dr. Funminiyi Olajide, Nottingham Trent University, UK; Rain Ottis, Tallinn University of Technology, Estonia; Dr Mahmut Ozcan, Webster University, USA; Prof Teresa Pereira, Instituto Politécnico de Viana do Castelo, Portugal; Michael Pilgermann, University of Glamorgan, UK; Engur Pisirici, Govermental - Independent, Turkey; Dr Bernardi Pranggono, Sheffield Hallam University, UK; Prof Carlos Rabadão, Politechnic of Leiria, Portugal; Dr. Muttukrishnan Rajarajan, City University London, UK; Prof Saripalli Ramanamurthy, Pragati Engineering College, India; Dr Trishana

*Ramlukan, University of KwaZulu-Natal, South Africa; Dr Aunshul Rege, Temple University, United States; Dr. Neil Rowe, US Naval Postgraduate School, Monterey, USA; Prof Vitor Sa, Catholic University of Portugal, Portugal; Filipe Sa Soares, University of Minho, Portugal; Dr. Char Sample, Carnegie Mellon University/CERT, USA; Prof. Henrique Santos, University of Minho, Portugal; Dr Keith Scott, De Montfort University, UK; Prof. Dr. Richard Sethmann, University of Applied Sciences Bremen, Germany; Dr. Yilun Shang, Northumbria University, UK; Mr Armin Simma, Vorarlberg University of Applied Sciences, Austria; Prof. Paulo Simoes, University of Coimbra, Portugal; Dr Umesh Kumar Singh, Vikram University, Ujjain, India; Prof. Jill Slay, University of South Australia, Australia; Dr Joseph Spring, University of Hertfordshire, UK; Anna Squicciarini, University of Milano, Italy; Iain Sutherland, Noroff University College, Kristiansand, Norway; Dr Hamed Taherdoost, Hamta Group / Hamta Business Solution Sdn Bhd, Malaysia; Unal Tatar, Old Dominion University, USA; Dr. Selma Tekir, Izmir Institute of Technology, Turkey; Prof. Sérgio Tenreiro de Magalhães, Champlain College, USA; Prof. Dr. Peter Trommler, Georg Simon Ohm University Nuremberg, Germany; Prof Tuna USLU, Istanbul Gedik University, Occupational Health and Safety Program, Türkiye; Craig Valli, Edith Cowan University, Australia; Dr Brett van Niekerk, University of KwaZulu-Natal & Transnet, South Africa; Rudi Vansnick, Internet Society, Belgium; Richard Vaughan, General Dynamics UK Ltd, UK; Dr Sangapu Venkata Appaji, KKR & KSR Institute of Technology and Sciences, India; Stilianos Vidalis, School of Computer Science, University of Hertfordshire, UK; Dr. Natarajan Vijayarangan, Tata Consultancy Services Ltd, India; Marja Vuorinen, University of Helsinki, Finland; Dr Khan Ferdous Wahid, Airbus Group, Germany; Prof Mat Warren, Deakin University, Australia, Australia; Dr. Santoso Wibowo, Central Queensland University, Australia; Prof. Trish Williams, Flinders University, Australia; Prof Richard Wilson, Towson University, USA; Simos Xenitellis, Royal Holloway University, London, UK; Dr. Hannan Xiao, University of Hertfordshire, UK; Dr. Omar Zakaria, National Defence University of Malaysia, Malaysia.*

## Biographies

### Conference and Programme Chairs



**Tiago Cruz** is Assistant Professor at the University of Coimbra. His research interests cover areas such as management systems for communications infrastructures and services, critical infrastructure security, broadband access network device and service management, internet of things, software defined networking and network function virtualization (among others), being the author of more than 60 publications, including chapters in books, journal articles and conference papers.



**Paulo Simões** is Assistant Professor at the University of Coimbra. His main research interests are Security, Network Management and Critical Infrastructure Protection. He has over 150 journal and conference publications in these areas. He has been involved in several European research projects. He has been involved in several European- and industry-funded research projects, with both technical and management activities. He was also co-founder of two technological spin-off companies in these areas.

### Keynote Speakers



**Dr. Leandros A. Maglaras** received the B.Sc. degree from Aristotle University of Thessaloniki, Greece in 1998, M.Sc. in Industrial Production and Management from University of Thessaly in 2004, and M.Sc. and PhD degrees in Electrical & Computer Engineering from University of Volos, in 2008 and 2014 respectively. In 2018 he was awarded a PhD in Intrusion Detection in SCADA systems from University of Huddersfield. He is the head of the National Cyber Security Authority of Greece and a part time Senior-Lecturer in the School of Computer Science and Informatics at De Montfort University, U.K. He serves on the Editorial Board of several International peer-reviewed journals such as IEEE Access and Wiley Journal on Security & Communication Networks. He is an author of more than 90 papers in scientific magazines and conferences and is a senior member of IEEE.



**Nuno Medeiros** gained a bachelor's degree in electrical and computer engineering from the University of Oporto, Portugal, in 2008. At the end of 2011, he obtained two MSc degrees in information security from Carnegie Mellon University (USA), as well as the University of Lisbon. In 2009 Nuno began working at EDP Distribuição, becoming part of the Directorate for Automation and Telecontrol. In 2012 he started working in the cybersecurity area, which is responsible for the end-to-end security architecture and the definition, management and coordination of the security and privacy requirements, based on risk analysis methodologies for the critical systems of the organization. From 2017, Nuno has been a Chief Information Security Officer (CCISO) Certified by the EC-Council and he became the top responsible for Cybersecurity at EDP Distribuição. Nuno is an industry representative in several projects and working groups at European level, he is regularly invited as a lecturer at several European conferences, which has allowed him to share the vision and the paths adopted by EDP Distribuição regarding Privacy and Cybersecurity.

### Mini Track chairs



**Dr Nasser S. Abouzakhar** is a senior lecturer at the University of Hertfordshire, UK. Currently, his research area is mainly focused on critical infrastructure security, cloud security and applying machine learning solutions to various Internet and Web security and forensics related problems. He received PhD in Computer Sci Engg in 2004 from the University of Sheffield, UK. Nasser worked as a lecturer at the University of Hull, UK in 2004-06 and a research associate at the University of Sheffield in 2006-08. He is a technical studio guest to various BBC World Service Programmes such as Arabic 4Tech show, News-hour programme and Breakfast radio programme. Nasser is a BCS chartered IT professional (CITP), CEng and CSci and is a BCS assessor for the accreditation of Higher Education Institutions (HEIs) in the UK. His research papers were published in various international journals and conferences.



**Filipe Caldeira** is an Adjunct Professor at the Polytechnic Institute of Viseu, Portugal. He is a researcher at the CI&DETS research centre of the Polytechnic Institute of Viseu and at the Centre for Informatics and Systems of the University of Coimbra. His main research interests include ICT security, namely, trust and reputation systems, Smart Cities and Critical Infrastructure Protection. His research papers were published in various international conferences, journals and book chapters. He has been recently involved in some international and national research projects.



**Prof. Walter Dorn** is Professor of Defence Studies at the Royal Military College of Canada (RMC) and the Canadian Forces College (CFC). He teaches officers of rank major to brigadier-general from Canada and about 20 other countries.



**Prof. Helge Janicke** is the Technical Director of De Montfort University's Cyber Technology Institute. He is a general chair of the International Symposium on SCADA and Industrial Control Systems Cyber Security Research (ICS-CSR). He serves on the editorial board and as reviewer for a number of international journals.



**Dr. Michael Robinson** is a cyber security research engineer at Airbus. As part of the architecture, innovation and scouting team he provides cyber expertise to the business and supports state of the art research into new and novel cyber security solutions.



**Dr. Char Sample** is a researcher focused on Threat Intelligence at MITRE and a visiting fellow at the University of Warwick. Dr. Sample has most recently focused her studies on the role of culture in cyber-attack and defence behaviours. Additionally, she has interest in metrics, traffic analysis, risk management and measurement, and predictive models. Dr. Sample's background encompasses commercial, government and most recently academic environments. She continues to try to merge the best features of all three environments.

## Biographies of Contributing Authors

**Ramona Susanty Ab Hamid** has been with CyberSecurity Malaysia for the past 17 years, an agency under the Ministry of Communications and Multimedia, Malaysia. Ramona holds a Degree in Applied Statistics and Operational Research from the University of Science Malaysia (USM) and holds a Postgraduate Diploma in Protective Security Management from International Islamic University Malaysia. She has contributed various publications and presentations related to cyber security and cyber safety besides managing content for CyberSAFETM Program since 2010

**Sokri Abderrahmane** has a Ph.D. in administration from HEC-Montreal. He serves as a Data Scientist for the Canadian Department of National Defence. He taught economics and statistics at different universities. He has published in high-level international journals. His current research interest includes game theory applied to military operations.

**Susana Aldeia** is a full-time Assistant Professor at the Portucalense University. She holds a Phd with mention in Tax Law. She is a researcher at REMIT and IJP. She develops research activities on taxation and accounting. She has been a chartered accountant since 2003 in exercise.

**Vijay Bhuse** is an Assistant Professor of Computer Science at the Grand Valley State University. He is received his Ph.D. from Western Michigan University in 2007 and completed his postdoctoral fellowship from the Dartmouth College. He worked in industry before returning to academia. His research interests are Network Security, Wireless Sensor Networks and Secure Coding.

**Lt Col Mustaffa Bin Ahmad** (RETIRED) C|CISO psc is the Senior Vice President, Outreach and Capacity Building Division of CyberSecurity Malaysia (CSM) – an agency under the purview of Ministry of Communications and Multimedia Malaysia. Prior to joining CyberSecurity Malaysia in 2007, he has served more than 18 years in the Malaysian Armed Forces in various capacities. He holds a Bachelor and Master's Degree in Mass Communications and Political Science from the University of Wisconsin, USA, a Post Graduate Diploma in Strategic Studies from University Malaya and a graduate of the prestigious Malaysian Armed Forces Staff College. Mustaffa has been recognized as a leader in internet safety and digital citizenship.

**B. Blumbergs** is a CERT.LV team member and the ambassador of the NATO Cooperative Cyber Defence Centre of Excellence. He is a certified exploit researcher, advanced penetration tester, industrial cyber security professional, and a cyber security PhD candidate at TalTech focusing on specialized cyber red teaming and responsive computer network operation execution.

**Cătălin Boja** is professor at the Economic Informatics and Cybernetics Department at the Academy of Economic Studies in Bucharest, Romania. He is a team member in various undergoing university research projects where he applied most of his project management knowledge. His work currently focuses on the analysis of mobile computing, information security and cryptography.

**Radomir Bolgov** Associate professor at the School of International Relations, St. Petersburg State University. He achieved a PhD in Political Science in 2011. He teaches the courses "Internet and World Politics", "Information Security", and "Information Society and International Relations". His current studies focus on the Information Society policies and Information/Cyber-Security in post-Soviet countries.

**Johnny Botha** Project Manager, software developer & researcher at the Council for Scientific and Industrial Research (CSIR). Masters (MTech) degree in Information Technology, at University of South Africa (UNISA). Topic: "Personal Identifiable Information Disclosure Since the Protection of Personal Information Act Adoption in South Africa" Obtained NDip and BTech degree in Computer Systems Engineering at the Tswane University of Technology (TUT).

**Dr Linas Bukauskas** holds PhD in computer science from Aalborg University, Denmark. He is a head of Cybersecurity Laboratory in the Institute of Computer Science at Vilnius University. He was one of the organisers of National Cybersecurity Training "Cyber Shield" and "Amber Mist" (2016-2018). His research interests include Cybersecurity, Data Mining, and Natural Language Processing.

**Prof. Ladislav Burita**, University of Defence, Brno and Tomas Bata University, Zlín, Czech Republic. PhD. from 1985 and professor from 2003. He worked in NATO/MIP team; was head of the faculty research program and MENTAL defense project. His area of interest are information and knowledge systems. He has published hundred papers at conferences and in journals.

**Filipe Caldeira** is an Adjunct Professor at the Polytechnic Institute of Viseu (IPV), Portugal. He is a researcher at the CI&DETS research centre of the IPV and at the CISUC of the University of Coimbra. His main research interests include ICT security, namely, trust and reputation systems, Smart Cities and Critical Infrastructure Protection.

**Dr Marian Carcary** is a senior lead researcher with the Innovation Value Institute at Maynooth University, Ireland. Her research interests include research methodology, design science, organizational transformation, risk and information security management, among others. Marian researches the development and deployment of the IT Capability Maturity Framework (IT-CMF), and has managed and worked on national and European-funded projects in the areas of e-skills, and IT management for small and medium-sized enterprises. Contact her at [marian.carcary@mu.ie](mailto:marian.carcary@mu.ie).

**Dr. Jami Carroll** retired from the U.S. Navy after two decades with C4ISR electronic systems. As the founder of Prisidian Security Solutions, he provides security-related services to the U.S. government. His research areas are cyber operations, threat intelligence, and denial and deception. His terminal degree is a D.Sc. in Cybersecurity from Capitol Technology University.

**Dave Chatterjee** is Associate Professor, in the MIS department, at the The University of Georgia. His expertise lies in the various facets of technology management – from sensemaking to information security, implementation, and change management. His work has been published in outlets such as MIS Quarterly, The Wall Street Journal, and MIT Sloan Management Review.

**Gerry Conway** is a Senior Research fellow with the Innovation Value Institute, Maynooth University, Ireland. Gerry has lead the research on the development of IVI's approach for Cloud and Data Centre Efficiency. His main focus areas are Business Process Management, Technical Infrastructure Management and Sustainability.

**Didier Danet** is holding a PhD in management sciences (Rennes Univ.) His research topics are dealing with cyberdefense. He is Head of a research department in Saint-Cyr Military Academy which is focusing on the changing character of war and director of the Saint-Cyr Master Program in Cyberdefense.

**Dr. Emily Darraj**, C|CISO, has 20+ years cybersecurity, computer forensics, and cyber warfare experience consulting to public, federal and state clients, specializing in AI. Dr. Darraj is a doctorate professor at Capitol Technical University. Dr. Darraj earned degrees from Notre Dame of Maryland University, Johns Hopkins University, and a cybersecurity doctorate from Capitol Technology University.

**Martina Jennifer Zucule de Barros** has Master's degree in Distributed Systems Engineering (2014) from Technical University of Dresden. She is currently a PhD student in cybersecurity at Technical University of Dresden, Germany and Eduardo Mondlane University, Mozambique. Her research focuses is on cybersecurity education.

**Maxence Delong** is a French last graduated engineerstudent at ESIEA, CVO Lab whose areas of expertise and research are network and information security, OSINT techniques and warfare techniques.

**Dr. Eileen Doherty** is a Research Fellow with the Innovation Value Institute, Maynooth University, Ireland. Her responsibilities include development of key aspects of IT-CMF. She has published 3 books and has published academic peer review journal papers in the area of SME, Innovation adoption and diffusion, and the utilisation of such technologies.

**Second Lieutenant Christopher Dukarm** is a graduate student at the Air Force Institute of Technology. He received his undergraduate degree in Computer and Network security at the United States Air Force Academy in 2018. His main areas of research are cloud computing, and cyber education.

**Simon Duque Anton** is a Ph.D. researcher at the German Research Center for Artificial Intelligence's Intelligent Networks research group. He received his Diploma of Engineering in 2015 from the University of Kaiserslautern. He is working in research projects about industrial cyber security, with a focus on machine learning methods for intrusion detection.

**Dr. Petrus Duvenage** is a counterintelligence specialist with extensive practical experience in various aspects of this field. In the course of his career, he served as an officer in the South African Defense Force, the National Intelligence Service and the State Security Agency. Duvenage holds a PhD from the University of Pretoria and is currently a Senior Research Fellow at the Academy for Computer Science and Software Engineering (University of Johannesburg).

**Dagney Ellison** has studied both at the University of Aberdeen, UK and the University of Pretoria, South Africa. Dagney completed her undergraduate degree in Computer Science at the University of Aberdeen in 2013 and is currently working on her Masters in digital forensics at the University of Pretoria.

**Dagney Ellison Dagney** has studied both at the University of Aberdeen, UK and the University of Pretoria, South Africa. Dagney completed her undergraduate degree in Computer Science at the University of Aberdeen in 2013 and is currently working on her Masters in digital forensics at the University of Pretoria.

**Thaddeus Eze Thaddeus** gained his PhD in Trustworthy Autonomic Computing and Communications from the University of Greenwich, UK, in 2014. Thaddeus also has experience in self-organising (ad hoc) networks in

disaster area scenarios and for rescue efforts. Thaddeus joined the University of Chester in 2015 to develop and deliver the University's new Cybersecurity programme and research.

**Eric Filoli** is the head of (C+V)O research lab at ESIEA, France and senior consultant in offensive cybersecurity and intelligence. He spent 22 years in the French Army (Infantry/Marine Corps). He holds an Engineer diploma in Cryptology, a PhD in applied mathematics and computer science and a Habilitation Thesis in Computer Science. He is graduated from NATO in InfoOps. He is the Editor-in-chief of the Journal in Computer Virology. He has been a speaker at international security events including Black Hat, CCC, CanSecWest, PacSec, Hack.lu, Brucon, H2HC...

**Nathaniel Flack** is pursuing a MS in Cyberspace Operations at the Air Force Institute of Technology in Ohio. He received his BS in Computer Engineering from Cedarville University in Cedarville, Ohio in 2012. His main research areas are cyber education, multi-domain operations, and serious games.

**Noran Shafik Fouad** is a doctoral researcher in international relations at the University of Sussex, and a recipient of the university's Chancellor International Research Scholarship. Her research examines the peculiarities of digital information and its implications on cybersecurity policy and theory, with a particular focus on the US as a case study.

**Saïd Haddad**, Ph.D in Political science (René Descartes University, Paris), is Senior lecturer in Sociology and member of the research team, Conflits in Mutation of the Saint-Cyr Research Center at the Saint-Cyr Military Academy, France. His current research focuses on the construction of cyber as a French national priority and the sociology of "cyber warriors".

**Gerhard Henselmann**, Dipl.-Ing. MBA, graduated Flighttest-Engineer was educated in Aerospace Engineering at Technical University of Munich/Germany and is working over 35 years in aerospace with expert experience in testing, flighttesting of airborne military platforms and has a wide experience in avionics, electronic warfare and self-defence of military platforms. He started his PhD studies in summer 2016 at the University of Jyväskylä on Cyber Security.

**Aarne Hummelholm** graduated from Helsinki University of Technology in 2000. Since then he has been involved in the design, development of architectures' of authorities' telecommunications networks and information systems. Key themes in his work have been critical service availability, cyber security and preparedness issues. In 2017 he started his doctoral dissertations at the University of Jyväskylä.

**Gazmend Huskaj** is a PhD candidate in Cyber Operations at the Swedish Defence University. He received his MSc in Information Security from Stockholm University in 2015 as a distinguished graduate. Previously, he was Director Intelligence in the Swedish Armed Forces focusing on cyber-related issues. He is also a ISACA Certified Information Security Manager (CISM).

**Ion A. Iftimie** is a Doctoral Candidate in Vienna, Austria. Previously, he served as the Deputy Chief for Information Operations at the United States Cyber Command. He graduated from top defense colleges in the United States, Germany, and Sweden, and is an alumnus of the Harvard Kennedy School Executive Program in Cybersecurity Policies.

**Eduardo Arthur Izzycki** is a Student of Master in International Relations by the University of Brasília (UnB) and public servant. Eduardo Izzycki worked on developing solutions for risk assessments in the cycle of major events in Brazil (2012-2016). He currently works in the Critical Infrastructure Protection Coordination of the Brazilian Institutional Security Office (GSI).

**Margarita Jaitner** is an analyst at the Swedish Defense Research Agency. She received her MSSc in Societal Risk Management from Karlstad University Sweden. She has authored several academic publications within the area of information warfare in cyberspace, hybrid warfare and other policy-related research within cyber security.

**Dr. Victor Jaquire** has been within the field of cyber and information security for over 20 years within Government and the Private sector focusing on strategy, performance management and operations. He holds

an Honours Degree in Management from Henley University and a Master's and PhD in Informatics from the University of Johannesburg - specialising in strategies for cyber counterintelligence maturity and the security of cyberspace. He has published various academic papers on cyber strategies and cyber counterintelligence maturity. His professional certifications include CISSP, CISM and CCISO.

**Dr. Connie Justice** has over 30 years' experience in cybersecurity, computer, and systems engineering. She designed courses in cybersecurity curriculum to NSA/DHS Center of Academic Excellence and NIST National Initiative for Cybersecurity Education standards. Research areas include: fake news, industrial controls risk, experiential learning, information and security risk management, digital forensics.

**Fredrick Kanobe** was PhD candidate at Tshwane University of Technology, South Africa. His research domain is ICT4D.

**Mrs. Eleni Kapsokoli** is PhD Candidate in University of Piraeus, Department of International and European Studies, Greece. She also holds a bachelor degree from the [National and Kapodistrian University of Athens](#) at the faculty of Political Science and Public Administration. She earned her Master's Degree on International Relations and Strategic Studies at the Panteion University of Social and Political Sciences. Her main research interests include international security, terrorism, cybersecurity and cyberterrorism. She is also a researcher in the Institute of International Relations (I.I.R). She is also a PhD Fellow at the European Security and Defence College (ESDC).

**Martti J Kari** is university teacher and PhD student of cyber security in Jyväskylä University, Finland. He retired as colonel from Finnish Defense Intelligence in the end of year 2017. His last post was Assistant Chief of Defense Intelligence. He has MA in Russian language (1993) and literature and MA in cyber security (2017) in Jyväskylä University. Kari has worked as a university teacher from the beginning of year 2018 In Jyväskylä University. He is specialized in Russian cyber and hybrid warfare.

**Kaur Kullman** is researching at the US ARL whether stereoscopically perceivable 3D data visualizations would be helpful for cybersecurity analysts, incident responders and other operational roles. He's been in IT since '90s, focusing on cybersecurity since late '00s. His interests are hands-on technical (OS-hardening, malware analysis, pentests), while his duties at EISA were more various.

**Kautsarina** is a government researcher at Ministry of Communication and Information Technology (MCIT), Republics of Indonesia since 2009. She also works as an ISO 27001 Lead Auditor for public institution since 2011. She is involved in developing policy research and ICT master plan. Now she is full-time PhD student at Computer Science Faculty, University of Indonesia. Her interest is about information security awareness improvement for end-user.

**Anthony Keane**, MSc, PhD has a background in astrophysics research and computer science and is currently the Head of the School of Informatics & Engineering in the Technological University Dublin, Ireland. He is also a Principal Investigator in the Cyber Security Education & Research Centre with interests in Cyber Bullying, Cyber Warfare and Cloud Forensics.

**Thorsten Kodalle** is lecture on security policy at the Command and Staff College of the German Armed Forces with a special focus on NATO, Critical Infrastructure and Cyber. He has a diploma in Social Science, assignments as a youth information officer, in the MoD, lecture on management and leadership and supported for several years computer assisted exercises at the Command and Staff College with constructive simulation. He is a member of the NATO research task group "Gamification of Cyber Defense/Resilience", an experienced facilitator of manual wargaming on the operational level for courses of action analysis, for operational analysis, operations research, serious gaming and especially for matrix wargaming.

**Tuija Kuusisto** is a Senior Ministerial Advisor at Ministry of Finance and an Adjunct Professor at National Defence University and University of Jyväskylä in Finland. Her expertise covers information analysis and management for decision-making, as well as information and cyber security strategies and policies. She have contributed to several international research and experiment projects and working groups organized by EU, UN and OECD. She has about 70 scientific publications in international and national journals, conference proceedings and books.

**Sylvain (Sly) Leblanc** is an Associate Professor and Interim Chair for Cyber Security at the Royal Military College of Canada (RMC). Sly was a Canadian Army Signals Officer for over 20 years, where he developed his interest in computer network operations. His research interests are in computer security, cyber operations development and cyber education.

**Wai Sze Leung** is an associate professor at the Academy of Computer Science and Software Engineering at the University of Johannesburg. Her current research interests include digital forensics and the application of Artificial Intelligence in enhancing cyber security.

**Dr. Andrew N. Liaropoulos** is Assistant Professor in University of Piraeus, Department of International and European Studies, Greece. His research interests include international security, intelligence reform, strategy, foreign policy analysis, European security policy and cyber security. Dr. Liaropoulos is also a member of the editorial board of the Journal of Information Warfare (JIW).

**Jarno Limnéll** is the Professor of cybersecurity in Aalto University, Finland. Martti Lehto is the Professor of cybersecurity in University of Jyväskylä, Finland.

**Christoph Lipps** graduated in Electrical and Computer Engineering at the University of Kaiserslautern. Born in Pirmasens, Germany in 1986, he started working as a Researcher and Ph.D. candidate at the German Research Center for Artificial Intelligence (DFKI) in Kaiserslautern. His research focuses on Physical Layer Security (PhySec), Physically Unclonable Functions (PUFs) and entity authentication.

**Dr. Leandros A. Maglaras** received the B.Sc. degree from Aristotle University of Thessaloniki, Greece in 1998, M.Sc. in Industrial Production and Management from University of Thessaly in 2004, and M.Sc. and PhD degrees in Electrical & Computer Engineering from University of Volos, in 2008 and 2014 respectively. In 2018 he was awarded a PhD in Intrusion Detection in SCADA systems from University of Huddersfield. He is the head of the National Cyber Security Authority of Greece and a part time Senior-Lecturer in the School of Computer Science and Informatics at De Montfort University, U.K. He serves on the Editorial Board of several International peer-reviewed journals such as IEEE Access and Wiley Journal on Security & Communication Networks. He is an author of more than 100 papers in scientific magazines and conferences and is a senior member of IEEE.

**Nnana Mogano** is a professor of computing at University of Pretoria, South Africa. She received her Bachelor of Science in Statistics and Computer Science and Honours in Statistics from the University of Limpopo, South Africa in 2014. Her main interests are data analytics to enhance business models and operations.

**Potsane Mohale** is a master's student of the Academy of Computer Science and Software Engineering at the University of Johannesburg. He works as a software engineer based in Johannesburg, South Africa.

**Pardis Moslemzadeh Tehrani** is a Senior Lecturer in the Faculty of Law, University of Malaya where she has been a faculty member since 2015. Her research interests lie in the area of Cyber Terroism, Human Right, International Humanitarian Law, Cloud Computing in Law. Pardis has widely published papers in a number of national and international Conferences.

**Dr Jabu Mtsweni** is a Research Group Leader for Cyber Defence at the Council for Scientific and Industrial Research (CSIR), Research Fellow at University of South Africa, and Advisory Board Member at ITWeb Security Summit. His research interests and technical expertise are in cyber warfare, cybersecurity, and cybercrimes. He has over 15 years academic and industry experience with over 60 peer-reviewed conference and journal articles.

**Julie Murphy** has over 10 years of experience in telecommunications working primarily with Fortune 500 companies, and currently works as a Security Expert with IBM X-Force Red. Julie lectures part-time in the Technological University Dublin and is actively involved in promoting cybersecurity awareness and training.

**Abdalmuttaleb M.A. Musleh Al-Sartawi** is the Editor-in-Chief of the International Journal of Electronic Banking (IJEBank). He received his PhD in Accounting, from UBFS. He has chaired as well as served as a member in various editorial boards and technical committees in international refereed journals and conferences.

**Marvin Newlin** is a graduate student at the Air Force Institute of Technology studying to obtain an M.S. degree in Cyber Operations. Marvin graduated in 2014 from North Carolina State University with B.S. degrees in Mathematics and Computer Science.

**Captain (Eng.), Dr. Juha-Pekka Nikkarila** has a PhD in Physics (2008) and serves as a researcher and a special officer at the Finnish Defence Research Agency (FDRA). He obtained his MSc in Physics (2006) and MSc(Tech.) in Electrical Engineering (2016). He has served at FDRA since 2012 with research interests in operation analysis, electronic warfare and Cyber studies. His current research interests include modelling Cyber influencing, resilience and warfare. Earlier he served as a researcher in Marioff Corporation / United Technologies Corporation (2009-2012), Inspecta (2007-2009) and at the University of Jyväskylä (2006-2007), as well as a physics lecturer in Metropolia University of Applied Sciences (2009-2014).

**Dr Dave Ormrod** is a cyber security professional. Over his 22 years in the military, Dave has served on operations in Iraq and worked with multi-national forces in Europe, the United Kingdom and the United States. Dave has extensive practical leadership, management, cyber security, wargaming and simulation experience. He has a PhD in computer science from the University of New South Wales (UNSW) in 2017. Dave is certified with the Australian Information Security Registered Assessors Program (IRAP), Certified Information Systems Security Professional (CISSP) and Certified Information Security Auditor (CISA). Dave is also a Project Management Professional (PMP), a member of the SAP Defence Interest Group Security Working Group, the UNSW Australian Centre for Cyber Security Cyber War and Peace Research Group and the United States Military Cyber Professionals Association.

**Timea Pahi** is an engineer in IT Security. She works currently as a Junior Scientist at the Austrian Institute of Technology on several research projects focusing on threat intelligence, cyber attribution and on establishing cyber range trainings. Her area of expertise includes national cyber security, the protection of critical infrastructures, and cyber situational awareness.

**Dr. Pankaj Pandey** is a Research Scientist at the Norwegian University of Science and Technology, Gjøvik. Dr. Pandey's research is focussed on the economics of cyber security, risk management, etc. and he has contributed to risk assessment and application of blockchain technology for IoT under the ambit of GHOST project.

**Youngjun Park** received the B.S. degree in Computational and Systems Biology from University of California, Los Angeles in 2018. He is currently pursuing a masters degree on Cyber Operations at the Air Force Institute of Technology. His research focuses on investigating information leakage in Internet of Things networks.

**Dr. Joon S. Park** is a professor at Syracuse University, Syracuse, New York, USA. Over the past decades Dr. Park has been involved with theoretical/practical research and education in Cybersecurity. He is Syracuse University's point of contact (PoC) for the Center of Academic Excellence (CAE) in Cyber Defense (Education) and CAE-R (Research) programs, which are designated by the National Security Agency (NSA) and the Department of Homeland Security (DHS).

**Pierre Jacobs** is an Cybersecurity Operations Manager at Cisco. He received his Masters degree n 2016 at Rhodes University with specialisation in information security, and is currently a PhD candidate at the University of Johannesburg. He specialises in Security Operation Centers (SOCs), and have implemented SOCs at both national and organisational level. His main areas of research are cybersecurity frameworks and models.

**Karine Pontbriand** is a PhD Candidate in Cyber Security and a member of the Research Group on Cyber War and Peace at UNSW Canberra Cyber. She is a former policy analyst at Global Affairs Canada. She holds a B.A. in International Relations and International Law and a M.A. in International and Intercultural Communication (with Distinction).

**Jouni Pöyhönen**, MSc. (Industrial Development and Management), Col (ret.) is PhD-student in Cybersecurity at the Faculty of Information Technology in the University of Jyväskylä. He has over 30 years' experience as developer and leader of C4ISR Systems in Finnish Air Forces. Now he is also a project researcher of the Cyber Security programs. He has in all a few research reports and articles on areas of cyber security.

**Jyri Rajamäki** is Principal Lecturer in Information Technology at Laurea University of Applied Sciences and Adjunct Professor of Critical Infrastructure Protection and Cyber Security at University of Jyväskylä, Finland. He holds D.Sc. degrees in electrical and communications engineering from Helsinki University of Technology, and PhD degree in mathematical information technology from University of Jyväskylä.

**Dr Trishana Ramlukan** a Post- Doctoral Researcher in International Cyber Law, College of Law and Management Studies, UKZN. Her areas of research include IT and Governance. She is a member of the IFIP working group on ICT Uses in Peace and War and is an Academic Advocate for ISACA.

**Juhani Rauhala** is a Research Affiliate and PhD student of cybersecurity at the University of Jyväskylä. He has over ten years' experience in the telecommunications industry and has been awarded two patents related to cloud storage. Juhani is a designated Eur Ing by the Federation of European Engineers and holds BScEE (1992) and an MScE (1996) degrees from San Francisco State University. His research interests include the weaponization of ubiquitous technologies and technology abuse.

**Dr. Aunshul Rege** is an Associate Professor with the Department of Criminal Justice at Temple University. Her cybercrime/security research on adversarial decision-making and adaptation, organizational and operational dynamics, and proactive cybersecurity is funded by several National Science Foundation grants.

**Dr. Mari Ristolainen** is a Researcher at the Finnish Defence Research Agency. She has studied psychology at the Moscow State University and she earned a doctorate in Russian Language and Cultural Studies from the University of Joensuu in 2008. She has been conducting postdoctoral research in the field of Russian and Border Studies in several Academy of Finland- and EU-funded projects at the University of Eastern Finland and at the University of Tromso, Norway. Her current research interests include cyber warfare as a phenomenon, Russian digital sovereignty, and the governance of cyber/information space.

**M.Sc. (Cognitive Science) Tarja Rusi** is a Cyber Security Master's student (University of Jyväskylä, Finland). She has 20+ years career in the telecommunications industry and has been lecturing on cyber threats and cyber terrorism. She has participated in governmental cyber threat evaluation work. Research interests: critical infrastructure protection, cyber threats, cyber terrorism, state sponsored cyber-attacks.

**Helvi Salminen** has worked in information security since June 1990. Before her security career 12 years of experience in systems development. Helvi is founder member of Finnish Information Security Association and president of ISACA Finland Chapter Helvi is qualified CISA, CISSP & SABSA & was awarded as CISO of the year in Finland 2014.

**Dr. Char Sample** is research fellow for ICF Inc. at the US Army Research Laboratory in Maryland, and is a visiting academic at the University of Warwick, UK. She has over 20 years experience in the information security industry and focuses her research on Fake News, cultural values in cyber security events, and data resilience.

**Leonel Santos** is an Equiv. Assistant Professor at Polytechnic Institute of Leiria. He is a researcher on the Computer Science and Communication Research Centre and is a PhD student in University of Trás-os-Montes e Alto Douro. His major research interests include Cybersecurity, Information and Networks Security, Internet of Things, Intrusion Detection Systems and Computer Forensics.

**Mr. Lynn Scheinman** is a technology consultant specialising in defence, security, and applied data science at SAP in Walldorf, Germany. He received his Masters in Science of Project Management and Operations Research at Florida Institute of Technology. His defence experience comes from 10 years in the US Army Special Forces as a Green Beret.

**Dr Keith Scott** is the Subject Leader for Languages at De Montfort University, where he is also a member of the Cyber Security Centre. He is also a member of the Cyber Policy Centre, a UK-based independent public policy centre devoted exclusively to the consideration of cyber as a socio-technical phenomenon. His research interests include the social and cultural implications of 'cyber' as a concept, influence, online communication, and the use of gaming as a teaching and research tool.

**Youngsup Shin** is a researcher in Agency for Defense Development, South Korea. He is in an integrated PhD program in Korea University. His main research areas are cyber situational awareness and cyber warfare.

**Ph.D. Petteri Simola** is a senior psychologist at the Finnish Defence Research Agency, Human Performance Division. His work involves human aspects of information security, Human Factors (especially sleep) and aptitude testing in recruitment.

**Jussi Simola** is a PhD student of cyber security in University of Jyväskylä. His area of expertise includes decision support technologies, SA systems, information security and continuity management. His current research is focused on effects of cyber domain as a part of Hybrid Emergency Response Model.

**Mr Veikko Siukonen** is a research officer at Finnish Defence Research Agency (FDRA). He received his Master's Degree in Military Sciences from The National Defence University in 2007. He is a master of science student (cyber security program) in University of Jyväskylä. His main research areas are cyber warfare and cyber threat intelligence.

**Tiia Sõmer** is early stage researcher at Tallinn University of Technology. She is conducting PhD level studies, focusing on modelling cyber criminal journey mapping. In addition to research she does teaching and has co-authored educational materials for general education. Before starting academic career, she served for more than twenty years in the Estonian defence forces.

**Lee Speakman**, Lee gained his PhD in the area of Mobile Ad hoc Networks from Niigata, Japan, in 2009. Since then Lee worked in the area of networks, network security, and software exploitation and protection measures in Defence. Lee joined the University of Chester in 2015 to develop and deliver the University's new Cybersecurity programmes and research.

**Ilona Stadnik** is a PhD student at the School of International Relations, Saint-Petersburg State University, Russia. During 2018-2019 academic year she was a Fulbright visiting researcher at Georgia Institute of Technology, USA, working with Internet Governance Project. She has been a regular participant and speaker at major cybersecurity events such as the United Nations Internet Governance Forum (IGF), CyFy conference, European Dialogue on Internet Governance (EuroDIG). Her research covers international cyber norm-making, Russia-US relations in cybersecurity, and global Internet governance.

**Dr. Nikolai Stoianov** is Colonel in the Bulgarian armed forces, Deputy Director of the Bulgarian Defence Institute "Prof. Tsvetan Lazarov" and principal member of NATO's Science and Technology Board. He is also associate professor and leads several international research projects on cybersecurity and related issues.

**Mr Marcel Stolz** is a doctoral student in cyber security at the University of Oxford, UK. He has a background in Computer Science and has served as a First Lieutenant in the Swiss Armed Forces. His research interests lie in global cyber security and regulation of data companies, such as Facebook.

**Dr. Steven Templeton** is a researcher at the University of California, Davis, USA. Since 1999 he has operated a consulting firm specializing in ICS security and compliance. Originally a wildlife biologist, in 2018 he received his PhD in computer science. His research spans multiple area of computer security, in particular intrusion detection, monitoring, and attack modelling.

**Dr Ben Turnbull** is a Senior Lecturer for the University of New South Wales, Australia. He is an expert in cyber security and digital forensics with 16 years in the industry. He is also a Certified Information Systems Security Professional (CISSP). He has previously worked as a research scientist for the Defence Science and Technology Organisation in the field of cyber network defence and analysis. In his spare time, Ben plays too many card and board games.

**Maija Turunen** is a PhD Student at the Finnish National Defense University. Her main research areas consist of cyber warfare, Russia and strategic communication. Maija Turunen works as a legal counsel at the Finnish Transport Infrastructure Agency.

**Prof.dr.ir. Jan van den Berg** is a (n emeritus) Full Professor of Cyber Security at Delft University of Technology and at Leiden University. He received his PhD in Computer Sciences on a topic of Computational Intelligence from Erasmus University Rotterdam in 1996. He published on a variety of topics, mostly related to data analytics (both theory and applications) and cyber security. Till 1/1/2017 he headed the Section of Cyber Security at the EEMCS Faculty of TUDelft and till 1/1/2019 he acted as Scientific Director of the Cyber Security Academy The Hague.

**Brett van Niekerk** is a computer science academic at the University of KwaZulu-Natal. He serves as secretary for the IFIP Working Group on ICT in Peace and War and the co-Editor-in-Chief of the International Journal of Cyber Warfare and Terrorism. In 2012 he graduated with his PhD focusing on information operations and critical infrastructure protection.

**Prof. SH (Basie) von Solms** is a Research Professor in the Academy for Computer Science and Software Engineering at the University of Johannesburg. He is also the Director of the Centre for Cyber Security at the University of Johannesburg, an Associate Director of the Global Cybersecurity Capacity Centre of the University of Oxford (UK), serves on the Word Economic Forum's Global Council on Cybersecurity and is a Past President of the International Federation for Information Processing (IFIP). Prof von Solms specialises in research and consultancy in the area of Information and Cyber Security, Critical Information Infrastructure Protection, Cyber Crime and other related cyber aspects.

**Ambrósio Patrício Vumo** is a Doctoral student between TU-Dresden in Germany and UEM University in Mozambique. He received his Master in Network Engineering and Communication Service from Minho University Portugal. He is an author of paper entitled Analysis of Mozambican Websites: How do they protect their users. His main research areas are cloud computing security, network security and advanced IP network.

**Murdoch Watney** is a professor at the University of Johannesburg, South Africa. She is Head of the Department: Private Law. Murdoch is an NRF rated established researcher. She contributed to four textbooks and has published on the law of criminal law, and cyber law and has delivered peer-reviewed papers at national and international conferences.

**Mr Andrew Williams** is a PhD student at the University of New South Wales, Canberra and is a member of the Research Group on Cyber War and Peace. He is an Army Officer and holds four Masters degrees across a range of fields. His main research areas are strategy, cyber security policy and cyber law.

**Richard L. Wilson** is a Professor of Philosophy at Towson University in Towson, MD. Teaching Ethics in the Philosophy and Computer and Information Sciences departments and Senior Research Fellow in the Hoffberger Center for Professional Ethics at the University of Baltimore.

**Chu-Sing Yang** is a Professor of Electrical Engineering in the Institute of Computer and Communication Engineering at National Cheng Kung University (NCKU). His research interests include software-defined networking, network management, cloud computing, and cyber-security.

**Dr. Zahri Yunos** is the Chief Operating Officer of CyberSecurity Malaysia, an agency under the Ministry of Communications and Multimedia, Malaysia. Zahri holds a PhD in Information Security from the Universiti Teknikal Malaysia Melaka (UTeM), Malaysia. He has contributed various publications and presentations related to cyber security. Zahri also has been appointed as Adjunct Professor at UTeM.

**Christine Ziske** is a Scientific Analyst (MSc) with major in Information Science and holds a master's Degree in General Business Management. She is CEO of the Kikusema AB in Sweden and co-founder of the KikuSema GmbH in Berlin since 2001. The KikuSema GmbH/AB identifies future challenges of cyber security and transmits them into IT security applications.

**Ulf Ziske** as an independent digital professional is the founder and CEO of the KikuSema GmbH. He is the inventor and the developer of the multi-decorated security app "FabulaRosa and the Five New Protocols". His educational background as a Scientific Analyst enables him to identify future challenges of Cyber security and to transfer them into apps.

# Using Hypervisors to Overcome Structured Exception Handler Attacks

Asaf Algawi<sup>1</sup>, Michael Kiperberg<sup>3</sup>, Roee Leon<sup>1</sup> and Nezer Zaidenberg<sup>2</sup>

<sup>1</sup>University of Jyväskylä, Finland

<sup>2</sup>College of Management Academic Studies, Rishon LeZion, Israel

[asaf@trulyprotect.com](mailto:asaf@trulyprotect.com)

[michael@trulyprotect.com](mailto:michael@trulyprotect.com)

[roee@trulyprotect.com](mailto:roee@trulyprotect.com)

[nezer@trulyprotect.com](mailto:nezer@trulyprotect.com)

**Abstract:** Microsoft windows is a family of client and server operating systems that needs no introduction. Microsoft windows operating system family has a feature to handle exceptions by storing in the stack the address of an exception handler. This feature of Microsoft Windows operating system family is called SEH (Structured exception handlers). When using SEH the exception handler address is specifically located on the stack like the function return address. When an exception occurs the address acts as a trampoline and the EIP jumps to the SEH address. By overwriting the stack one can create a unique type of return oriented programming (ROP) exploit that force the instruction pointer to jump to a random memory address. This memory address may contain random malicious code. Multiple Microsoft Windows applications are particularly vulnerable to this type of exploit. Attacks on Microsoft Window application that exploit these mechanisms are found in many common windows applications (including Microsoft Office, Adobe Acrobat, Flash and other popular software). These attacks are well documented in CVE database in numerous exploits. We previously described how hypervisors can be used to white list an end point and provide application control for a workstation and servers and protect against malware and viruses that may run on the end point computer. In this work we extend the protection mechanism for end points and servers that uses the hypervisor to white list the machine. The hypervisor detects permission elevation from user space to kernel space (system calls invocation) and detects anomalies in the software execution. The hypervisor based mechanism allows for detection and prevention of SEH return oriented exploits execution. Our hypervisor based SEH-exploit prevention mechanism was tested on multiple well documented CVE vulnerabilities. Our hypervisor was found to prevent a large collection of different types of SEH exploits in multiple applications and multiple flavours and versions of Windows OS in both 32 and 64 bit environments

---

**Keywords:** SEH, rootkit, application control, hypervisor

## 1. Introduction

Structured exception handlers (SEH) are a family of Return oriented Programming (ROP) (Roemer et al 2012) attacks that infest windows hosts.

Hypervisors (Zaidenberg 2018) can be used for various security purposes. One common usage for hypervisors for system security is to protect End points and allow only predefined software will receive execute permissions and allowed to run as was shown by Seshadri et al (2007) and Resh et al (2017) and others.

However even by controlling the entire guest system memory the hypervisor is still not capable of protecting against return oriented programming. In this work we would like to briefly describe prior attempts at using hypervisors to control and attest the guest system and explain why they are not sufficient for preventing SEH and other ROP attacks.

We will describe our method that allows a thin hypervisor to provide the guest operating system protection against SEH attacks.

## 2. Background

Hypervisors are modern software components that are designed to run multiple operating systems on a single hardware device. Thus the hypervisor (called the host) is catching all memory access and all hardware interrupts and delivers them to the operating systems that it runs (called guests). Thus the hypervisor has a relationship with the guest that is similar to the relationship that the normal operating system has with a process.

However a special type of hypervisor has been proposed the thin hypervisor. The thin hypervisor supports only one operating system and leaves all (or most) of the handling of hardware interrupts, memory allocation and other system events to the guest the operating system. The thin hypervisor is only a monitor that supervises a

specific set of predefined event. TrulyProtect thin hypervisor for execution protection or anti reverse engineering (Averbuch et al 2013, Kiperberg et al 2016) as well as SecVisor for end point security and the blue pill are examples for such thin hypervisor.

However when implementing thin hypervisor for system protection we discovered a class of attacks against code that does not run as an application but as complex class of ROP attacks called SEH.

## **2.1 Using hypervisor for forensics**

Hypervisors have long been used as development tools. (Khen et al 2011, Khen et al 2013) They can even be used in order to detect security weaknesses on a target systems (Zaidenberg et al 2015). Later Kiperberg (Kiperberg 2019) has shown that the hypervisor can inspect the entire guest memory. It follows such a tool can be used in ensuring the end point security. Thus the goal is, given an end-point which is running off the shelf operating system. The hypervisor can be verified and found to be Trustworthy (Zaidenberg et al 2015). Thus a chain of trust be formed. I.e. can a remote 3<sup>rd</sup> party ensure that the end point is running code from pre-defined white list.

## **2.2 Using hypervisors for end point security**

A remote host can attest the hypervisor and verify the remote system correctness using side effects of computation (Zaidenberg and Resh 2015). This was shown in principle by Kennell (Kennell et al 2002) and in practice on modern x86 processors. (Kiperberg et al 2013, Kiperberg et al 2015) once the end point is running a trusted hypervisor it was shown by Seshadri et al 2007 and Resh et al 2017 that the end point can be counted on to only intentionally run only code from predefined white lists. However, code can come from many sources not only intentionally. Self-modifying code and Return oriented programming code can modify in memory and thus execute code (as an attack) that was not originally present in the white list. W^X or Data execution prevention (DEP) may prevent some attacks of self-modifying code but on a modern window host some pages are required to have write (W) and execute (X) permissions thus W^X is impossible for those pages. Some of these memory pages are SEH pages which are thus vulnerable to exploit even if a protecting hypervisor is running. Furthermore, These SEH pages are well known as part of a complete class of attacks against various software. Thus, it is required to provide windows hosts protection against SEH attacks.

## **2.3 Windows SEH handlers and SEH attacks**

In Intel's x86 instruction set architecture an exception is an event that is triggered during the execution of a program. The exception requires the immediate execution of code outside the normal flow of control. In x86 architecture there are two kinds of exceptions: hardware exceptions and software exceptions. Hardware exceptions are initiated by the CPU. They can result from the execution of certain instruction sequences, such as division by zero or an attempt to access an invalid memory address. Software exceptions are initiated explicitly by applications or the operating system. For example, the system can detect when an invalid parameter value is specified. The exception mechanism is employed by many programming languages ranging from native languages such as C++ to interpreted or JIT languages such as Java and Javascript. In windows Microsoft had developed a mechanism called SEH, which stands for Structured Exception Handling, this mechanism allows native windows code to handle both software and hardware exceptions. The SEH mechanism allows developers for the native platform to use \_try, \_except & \_finally keywords to handle exceptions. It is worth noting that unlike other C++ exception mechanisms, SEH allows for \_finally which is not standard exception handling in RAII languages like C++.

SEH mechanism works in the following manner, each TIB (Win32 Thread Information Block) hold a pointer to a list of structures called *\_EXCEPTION\_REGISTRATION\_RECORD* as shown in Figure 1

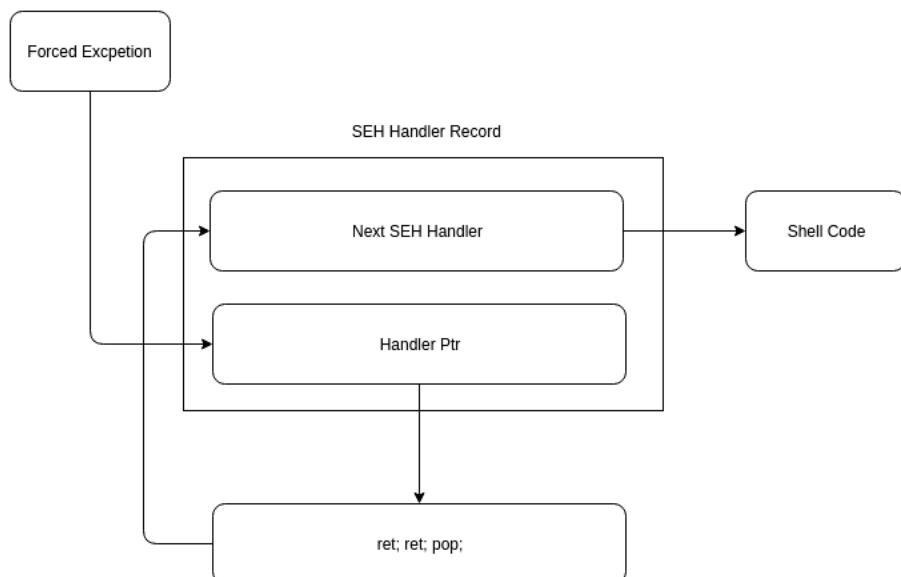
```
typedef struct _EXCEPTION_REGISTRATION_RECORD
{
    PEXCEPTION_REGISTRATION_RECORD Next;
    PEXCEPTION_DISPOSITION Handler;
} EXCEPTION_REGISTRATION_RECORD, *PEXCEPTION_REGISTRATION_RECORD;
```

**Figure 1** Windows \_EXCEPTION\_REGISTRATION\_RECORD

Each link on the list has two pointers, the first one is a pointer to the next link on the chain, the second is a pointer to the function `_except_handler3` which in turn calls the exception `_except` and `_finally` blocks for the developer.

SEH is entirely implemented in the Microsoft windows's visual studio compiler. When Microsoft (visual studio) compiler "sees" a `_try` block it will generate a call to a function called `EH_PROLOG`. The `EH_PROLOG` function will in turn create the `_EXCEPTION_REGISTRATION_RECORD` structure on the user stack and put it at the head of the chain. Once the Microsoft compiler "sees" an exit from the protected block the compiler will generate a call to `EG_EPILOG` which will remove the aforementioned structure from the stack. Every SEH chain in Windows ends with `kernel32!UnhandledExceptionFilter`, a predefined function which shows the "A Program had stopped working" dialog and kills the program's process.

The SEH mechanism is notorious for its vulnerability for security exploits. Since SEH handlers and pointers are written to the stack, it is possible for a user to control these pointers after an invalid memory operation or assignment such as `memcpy` or `strcpy` without (or with incorrect) bound checking. The basic idea is to control the pointer of the second link in the chain, making it jump into some random shell code inside the user stack. This is shown in Figure 2 below.



**Figure 2:** SEH attack operation logic.

We utilize the fact that the prolog function `EG_PROLOG` is called with the stack setup as follows, the top of the stack contains a ptr to the second link in the chain + 8 bytes, therefore issuing the sequence `pop; pop; ret;` will setup the stack in such manner the 8 bytes are gone, the top of the stack contains a ptr to the second link in the chain to which we return.

The attacker has to prepare a payload which contains the following:

**[PAD][Next SEH Record][SEH Record][Shell code]**

While we remember that each record is exactly 8 bytes, the pad is the distance between the buffer and the top of the chain. The exact payload obviously differs between each program but the structure is mostly the same, the pad can be any junk, the 'Next SEH Record' contains a JMP instruction with an address of the appended shell code (which should fit in 8 bytes), the 'SEH Record' part of the payload is a pointer to an address inside the program containing the special `pop; pop; ret;` code. Figure 2 demonstrates SHE operation logic

Some well-known attacks that utilize the above system are CVE-2018-9059, Exploit DB 38486, Exploit DB 38516 and Exploit DB 44218 and many others. All the above exploits use the same principal highlighted above for similar attacks on different applications.

### **3. Solution**

In order to circumvent attacks based on manipulating the SEH chain we use a special hardware features available as part of Intel's (and AMD's) hardware virtualization instruction set. The hypervisor can specify certain 'events' which cause an exit (VMEXIT) from the guest code to the hypervisor exit handler. One such event that causes a VMEXIT is the CPUID instruction. The thin hypervisor uses this feature to force an exit from guest to host whenever the function *KiUserExceptionDispatcher* starts. We catch this function's execution because it is responsible for unwinding the SEH chain during an exception. Once we catch that function we can inspect the entire SEH chain and make sure no function contains the malicious *pop;pop;ret* sequence or any sort of code located on the stack, especially such code which does a near jump to the stack. Once such template of malicious code is found, the Hypervisor can easily close the program safely without the malicious payload executing.

#### **3.1 Performance impact**

Systems such as SecVisor or TrulyProtect bears some performance cost from the sole reason of running a hypervisor.

Since the systems require look up two layers of memory indirection (OS and Hypervisor level) as opposed to only single level some memory performance penalty is expected. This penalty was measured by VMWare and other sources and was found to be between 0 and 5% depending on application.

The proposed system still suffer from this performance penalty. (Since the system is using a thin hypervisor it suffers from slightly reduced penalty when compared with full hypervisor such as VMWare ESXi but some penalty must exist)

Furthermore, slight additional penalty is added due to the loading of programs to memory and calculating the page hashes when the software is first loaded to memory but not on standard usage. This penalty is also associated with the general hypervisor-based protection and not with the costs associated with the protection system proposed here-in.

The only cost that is associated with the system described herein is the cost involved from handling exceptions during standard operations. Since exceptions are not called regularly in normal applications (we measured less than 1 exception per second) the costs associated with the system are truly minimal. (less than 0.1% in our system testing and within random operation variance)

### **4. Conclusion**

We have shown a critical weakness in using a well-known weakness family in hypervisor based end point protection. The aforementioned weakness allows execution of random code through a call for ROP. The ROP code execution is performed by exploiting the Windows SEH weakness. When using the weakness an attacker can inject random code to a hypervisor protected windows host. The code will execute despite the hypervisor based white listing as the memory pages are modified in memory.

We have also shown means to mitigate this weakness. By specifying the needed values on VMCS (Intel) or VMCB (AMD) structure hypervisors can mitigate this SEH attack and enforce the white list. This method is not preventing all ROP attacks not associated with Exceptions. DEP should still be used to defeat these attacks. However since there are several attacks against DEP as well (e.g. Stojanovski 2007) additional research is required to provide full defence against all forms of Return oriented programming (ROP).

### **References**

- Averbuch, A., Kiperberg, M., & Zaidenberg, N. J. (2013). Truly-protect: An efficient VM-based software protection. *IEEE Systems Journal*, 7(3), 455-466.
- CVE-2018-9059 "Stack-based buffer overflow in Easy File Sharing (EFS)" NIST <https://nvd.nist.gov/vuln/detail/CVE-2018-9059>
- Exploit DB 38486 Tomabo MP4 Player 3.11.6 - Local Stack Overflow (SEH) <https://www.exploit-db.com/exploits/38486>
- Exploit DB 38526 Easy File Sharing Web Server 7.2 - Remote Overflow (SEH) <https://www.exploit-db.com/exploits/38526>
- Exploit DB 44218 IrfanView 4.50 Email Plugin - Buffer Overflow (SEH Unicode) <https://www.exploit-db.com/exploits/44218>

- Khen, E., Zaidenberg, N. J., & Averbuch, A. (2011, June). Using virtualization for online kernel profiling, code coverage and instrumentation. In *2011 International Symposium on Performance Evaluation of Computer & Telecommunication Systems* (pp. 104-110). IEEE.
- Khen, E., Zaidenberg, N. J., Averbuch, A., & Fraimovitch, E. (2013, July). Lgdb 2.0: Using lguest for kernel profiling, code coverage and simulation. In *2013 International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS)* (pp. 78-85). IEEE.
- Kiperberg, M., & Zaidenberg, N. (2013) Efficient Remote Authentication. In *Proceedings of the 12th European Conference on Information Warfare and Security: ECIW 2013*(p. 144). Academic Conferences Limited.
- Kiperberg, M., Resh, A., & Zaidenberg, N. J. (2015, November). Remote Attestation of Software and Execution-Environment in Modern Machines. In *2015 IEEE 2nd International Conference on Cyber Security and Cloud Computing* (pp. 335-341). IEEE.
- Kiperberg, M., Resh, A., & Zaidenberg, N. (2016). U.S. Patent No. 9,471,511. Washington, DC: U.S. Patent and Trademark Office.
- Kiperberg M, Algawi A, Leon R, Resh A. & Zaidenberg N. J. Hypervisor-assisted Atomic Memory Acquisition in Modern Systems (2019) in Proceedings of 5<sup>th</sup> international conference on information system security and privacy ICISSP 2019
- Resh, A., Kiperberg, M., Leon, R., & Zaidenberg, N. J. (2017). Preventing Execution of Unauthorized Native-Code Software. *International Journal of Digital Content Technology and its Applications*, 11.
- Roemer, R., Buchanan, E., Shacham, H., & Savage, S. (2012). Return-oriented programming: Systems, languages, and applications. *ACM Transactions on Information and System Security (TISSEC)*, 15(1), 2.
- Seshadri, A., Luk, M., Qu, N., & Perrig, A. (2007, October). SecVisor: A tiny hypervisor to provide lifetime kernel code integrity for commodity OSes. In *ACM SIGOPS Operating Systems Review* (Vol. 41, No. 6, pp. 335-350). ACM.
- Stojanovski, N., Gusev, M., Gligoroski, D., & Knapskog, S. J. (2007, April). Bypassing data execution prevention on microsoftwindows xp sp2. In *The Second International Conference on Availability, Reliability and Security (ARES'07)*(pp. 1222-1226). IEEE.
- Zaidenberg, N. J. (2018). Hardware Rooted Security in Industry 4.0 Systems. *Cyber Defence in Industry 4.0 Systems and Related Logistics and IT Infrastructures*, 51, (pp 135-151).
- Zaidenberg, N. J., & Khen, E. (2015, November). Detecting Kernel Vulnerabilities During the Development Phase. In *2015 IEEE 2nd International Conference on Cyber Security and Cloud Computing* (pp. 224-230). IEEE
- Zaidenberg, N., & Resh, A. (2015). Timing and side channel attacks. In *Cyber Security: Analytics, Technology and Automation* (pp. 183-194). Springer, Cham.
- Zaidenberg, N., Neittaanmäki, P., Kiperberg, M., & Resh, A. (2015). Trusted Computing and DRM. In *Cyber Security: Analytics, Technology and Automation* (pp. 205-212). Springer, Cham.

# Creating Modern Blue Pills and Red Pills

**Asaf Algawi<sup>1</sup>, Michael Kiperberg<sup>4</sup>, Roee Leon<sup>1</sup>, Amit Resh<sup>3</sup> and Nezer Zaidenberg<sup>2</sup>**

**<sup>1</sup>University of Jyväskylä, Finland**

**<sup>2</sup>College of Management Academic Studies, Rishon LeZion, Israel**

**<sup>3</sup>Shenkar College, Ramat Gan, Israel**

**<sup>4</sup>Holon Institute of Technology, Holon, Israel**

[asaf@trulyprotect.com](mailto:asaf@trulyprotect.com)

[michael@trulyprotect.com](mailto:michael@trulyprotect.com)

[roee@trulyprotect.com](mailto:roee@trulyprotect.com)

[amit@trulyprotect.com](mailto:amit@trulyprotect.com)

[nezer@trulyprotect.com](mailto:nezer@trulyprotect.com)

**Abstract:** The blue pill is a malicious stealthy hypervisor-based rootkit. The red pill is a software package that is designed to detect such blue pills. Since the blue pill was originally proposed there has been an ongoing arms race between developers that try to develop stealthy hypervisors and developers that try to detect such stealthy hypervisors. Furthermore, hardware advances have made several stealth attempts impossible while other advances enable even more stealthy operation. In this paper we describe the current status of detecting stealth hypervisors and methods to counter them.

**Keywords:** virtualization, forensics, information security

---

## 1. Introduction

The blue pill was introduced by Johanna Rutkowska in Blackhat 2006 (Rutkowska 2006b). There were others who used hypervisor for security it was with the introduction of the concept of the blue pill and red pill that the virtualization concept became so closely related to cyber security.

The blue pill is a rootkit that takes control of a victim host computer. Unlike other rootkits the blue pill is actually a malicious hypervisor. The original blue pill starts after the OS has already booted and through a series of hardware instruction, the blue pill gains control of the victim host. In fact, after the blue pill is deployed it has gained higher privileges than the OS that started it. Of course, like all rootkits the blue pill must camouflage its existence, or it will be removed by the user. Using a hypervisor assists in camouflage because the blue pill can now start tasks outside the OS scope. In order to counter the blue pill the red pill was invented. The red pill is a hardware or software tool that is designed to detect such malicious blue pill rootkit.

The red pill is a special case of the related “trusted computing” and the attestation concept (Zaidenberg et al. 2015d), In Trusted computing attestation a remote 3<sup>rd</sup> party or even local software tries to ensure the integrity of the local machine in terms of software (mainly) and hardware (sometimes).

When hypervisors are concerned the attestation of lack of hypervisor or blue pill was first researched by Kennell et al. (Kennell et al 2003) in order to establish the “genuinity of the host” (i.e. ensure that the host is a physical machine running the correct software as opposed to an emulator or a virtual machine or a physical machine running non-genuine software) Kennell proposes running a series of tests that will pass only if the inspected system is genuine. The test will fail if a hypervisor is running due to side effects involved with the running of hypervisors, mainly more expensive memory traversal.

Today the original Kennell hypothesis are questionable due to the availability of specific hardware instructions and features designed to remove memory access side effects (Second Level Address Translation instructions such as EPT<sup>TM</sup> in Intel’s case and RVI<sup>TM</sup> in AMD’s case) However the core idea of using side effects is still in use in many modern red pills.

Since Rutkowska introduced the blue pills malware concept, multiple attempts to create red pills that detect such blue pills have also been proposed. However, more advanced blue pills have been designed to avoid detection. Which led to more advanced red pills and so on and so forth. So, the goals of the blue and red pills are conflicting and naturally technology advances in one front requires advancement by the other front in order

to keep up. This paper describes multiple modern methods in which one can construct a blue pill or red pill defeating most blue pills.

## **2. Modern hardware capabilities**

Nowadays, modern CPUs by Intel, AMD and ARM feature hardware-assisted virtualization. Hardware-assisted virtualization provides new capabilities for implementing virtual machines and emulator software. Thus, hardware-assisted virtualization makes several “red pill” attempts futile, for example by eliminating several virtualization side effects(Zaidenberg et al 2015c).

However, hardware-assisted virtualization provides new forensics methods. Therefore, many new openings to create new red pills now exist and many side effects that were previously used are eliminated or are much subtler.

This paper describes the situation of red pills and blue pills primarily on Intel virtualizations architecture circa 2019 and the ninth generation of intel’s core CPUs.

## **3. Background**

Blue pill technology relies on hardware assisted virtualization (“hypervisor instructions”) and hardware assisted virtualization technology. Recent advances in x86 hardware-assisted virtualization allow for better blue pills and red pills. These new instruction families are called VT-x, VT-d, EPT on the intel architecture. AMD-v, IOMMU and RVI are the corresponding names on AMD architectures. The new instructions enable the blue pill and red pill capabilities.

### **3.1 Hypervisors and thin hypervisors**

A hypervisor is a computer software concept that is designed to run multiple operating systems on the same hardware.

As its name implies, a hypervisor has higher privileges (hyper= “above”) than the operating system (i.e. The operating system = the supervisor).

The operating system supervises memory and hardware resources for the processes that runs on top of it. Likewise, the hypervisor supervises the hardware resources for each operating system that runs on top of it.

Hypervisors research started with Popek et al. (1974) who classify hypervisors into two main categories:

- 1. Type I hypervisors, or boot hypervisors, are hypervisors that the machine starts from on boot. The machine starts the guest operating system after booting the hypervisor. VMWare ESXi is an example of a modern Type I hypervisor.
- 2. Type II hypervisors or hosted hypervisors are hypervisors that start only after the operating system has started. A modern example for a Type II hypervisor is VMWare Desktop or Oracle Virtual Box hypervisors.

The original blue pill that was described by Rutkowsa was a Type II hypervisor, however type I “blue pill” hypervisors (boot kits) are also possible.

Regular hypervisors reside logically between the hardware and the supervisor (OS) layers. The hypervisors catch interrupts and deliver them to the correct operating system and control memory addresses. The hypervisor uses its own translation tables for deciding which operating system owns each memory address and which operating system should handle each hardware interrupt. This is analogues to the MMU in OS environment where each memory address is assigned to a different process.

However, there also exists a special case of hypervisors that do not attempt to run multiple operating systems. Instead, these hypervisors, called “thin hypervisors”, supports running only one operating system on the target hardware. All interrupts and memory accesses are either blocked or transferred to the operating systems for handling. Indeed, Thin hypervisors act as a microkernel that provides certain services to the underlying operating system. The thin hypervisor includes very little memory management and relies on the single guest OS system for both memory management and interrupt handling.

### 3.2 Hypervisors, forensics and cyber security

Zaidenberg (Zaidenberg 2018) summarized hypervisor usage in cyber security. Hypervisors can be used to inspect the target system. Such forensics effort can be used to assist developers (Khen 2011), profile the code (Khen 2013) or directly to obtain the memory of the inspected system (Kiperberg et al 2019a) or detect malware (Zaidenberg et al 2015b). Malicious software may desire to inspect the hypervisor presence before deployment.

Furthermore, the hypervisor may provide security services to the guest system. Microsoft's Deviceguard, TrulyProtect hypervisor for protection against reverse engineering (Averbuch et al. 2013, Kiperberg et al 2019b) and Execution Whitelisting (Kiperberg et al. 2017) are examples of thin hypervisors. Other thin hypervisors monitor Video DRM (David et al 2014), provide forensics data or provide end point security (Leon et al 2019). Virtually all blue pills are thin hypervisors.

### 3.3 x86 virtualization

On intel's x86 CPU architecture, hardware-assisted virtualization is provided by unique instruction families. Intel architecture and AMD architecture each provide three such instructions families for handling hypervisors.

These instructions are further optimized with each new processor generations. Newer generations include new hardware assisted virtualization capabilities such as shadow VMCS. Furthermore, virtualization instructions take less CPU cycles to complete in newer CPU generations.

The x86 instruction families are presented in Table 1.

**Table 1:** x86 Virtualization Instructions

	Intel Architecture Name	AMD Architecture Name	Usages
<b>Virtualization instructions</b>	VT-x	AMD-v	This family is required for starting a hypervisor
<b>SLAT (second-level address translation)</b>	EPT (Extended page tables)	RVI (Rapid virtualization indexing)	Allows the HV to run Multiple MMUs for multiple guest operating system
<b>IO MMU</b>	VT-d	IOMMU	Allow the HV to assign IO memory to specific guest operating systems
<b>VM data structure</b>	VMCS	VMCB	Holds all the VM information (one per VM)

### 3.4 Rootkits and the blue pill

The recommended and common system administrator's cyber incident-response protocol once an attack is detected on any networked server, is to format and reinstall the effected server's operating system. If the operating system is reinstalled (and fully patched with security patches), then the hacker may find herself locked outside of the compromised system. It follows that the hackers have a desire to hide their tracks. Thus, if the hack was not detected, the hacker can maintain a persistent access to the hacked servers. Therefore, hackers frequently install software packages known as a "rootkit." A rootkit is a software package that allows the hacker access to the victim system resources. Furthermore, the rootkit hides itself and all the processes that the hacker runs on the infected system in an effort to mask its existence. Therefore, the rootkit has two goals. First the rootkit provides the hacker ease of access to victim computer resources. Second the rootkit provides measures to hide the hack and its own existence on the victim system.

There are many ways to build rootkits from hijacking system calls and library functions and installing *setuid* programs to replacing innocent-looking binaries.

The blue pill is a special type of rootkit. Unlike normal rootkits that modify the operating system in order to hide files and gain access to the system, the blue pill starts a hypervisor. Thus, the blue pill gains more permissions than the operating system. The blue pill can run processes in the hypervisor address space. Thus, the processes or their address spaces are not visible by the operating system.

### **3.5 Bootkits**

One of the ways that the rootkit can operate is as a “bootkit”. A bootkit is a special type of rootkit that boots (from the hard drive master boot record, UEFI, PXE, or other means) before the operating system starts. The bootkit runs its own software (i.e. the hypervisor in the blue pill case) before the OS starts and later boots the OS (by calling the OS boot). The bootkit may also patch the OS system calls in order to hide its processes and files.

### **3.6 The original blue pill and subverting attacks**

Rutkowska (2006a) has introduced the original blue pill concept in 2006. The blue pill approach was an innovative rootkit approach that was relatively unresearched at the time. Since Rutkowska introduced the first blue pill, there have been several suggestions on how a blue pill can be detected (for example, the memory location of the IDT vector). However advances in x86 virtualization technology make such detection more difficult or even impossible. For example, it was initially proposed to check the address of the IDT in order to detect hypervisors. However, using EPT (new virtualization technology) it is completely possible to maintain addresses that appear to be genuine IDT address by the guest but point to different address.

### **3.7 Blue chicken**

Rutkowska et al. (2007) introduced the blue chicken blue pill one year after the original blue pill. The blue chicken blue pill detects the red pill inspection attempts and unloads itself, and then reinstalls itself after the detection attempts are over. The blue chicken method is not recommended for two reasons.

- A hypervisor can be loaded after the blue pill is unloaded. Then, if the blue pill tries to reload itself, the new hypervisor can detect it and prevent the blue pill from loading. The new hypervisor can even act as a blue pill, allowing the original blue pill to load in order to perform forensics inspect its operational logic).
- Most red pills do not consume too much time and may be able to detect hypervisor present before the blue chicken manages to unload.

We summarize that the blue chicken method is not a threat today.

## **4. Local red pills**

Local red pills are tests performed by the tested machine and contained within the tested machine. These tests cannot be considered reliable as computation is performed on an untrustworthy machine (the very machine that is being inspected). However, in many real-life scenarios there is no TPM chip available or similar third-party chip or server that one can use for attestation.

If no third party, trustworthy, root of trust is available, then one is left with running local code on an untrustworthy machine to provide attestation in a best effort attempt to detect blue pills on the inspected machine.

### **4.1 Paranoid fish and other modern red pills**

Paranoid fish (Pafish) (Ortega 2016) is currently the *de facto* standard in “red pill for blue pill detection.” software Pafish includes multiple tests capable of detecting most known blue pills when running under Linux or Windows. Many of these tests to detect a hypervisor assume the hypervisor is not really attempting to hide by looking up specific values in memory. (For example, VMWare routinely reports 440BX (1990’s era hardware) on all machines. These chipsets are over 20 years old components that are virtually non-existent on real physical modern systems)

However, some of Pafish tests are specifically geared toward hypervisor that do try to hide. These methods rely on timing and side effects (Zaidenberg 2015c) of running hypervisors. Local timing methods are employed to try to flush out these blue pills. The local timing tests perform the following steps:

- Take local time (for example, RDTSC instruction).
- Execute an operation that must be intercepted by a hypervisor (for example, CPUID instruction).
- Take local time again to obtain elapsed time.

The underlying assumption is that step (2) takes significantly longer to execute when a hypervisor is active.

## 4.2 Paranoid fish tests

Paranoid Fish (Pafish) tests include the following:

- Specific and paravirtualization tests – These tests are designed to detect specific commercial hypervisors (that do not try to hide)
- Timing tests (Pafish includes two timing tests)
- *RDTSC, CPUID, RDTSC < 1000 cycles. This test runs in user space. It runs the CPUID instruction, which must be intercepted by a hypervisor. As a result of the required context switch, the time required for the three instructions acts as a tell-tale sign for hypervisor presence (the red pill).*
- *RDTSC, RDTSC < 750 cycles - This test involves calling two RDTSC instructions and measuring the response times. It is designed to ensure that the hypervisor is not intercepting the RDTSC instruction.*
- *In both tests, an irregular high result will be obtained due to an interrupt. To avoid false positives due to irregular time required for the context switch, Pafish performs an average over 10 runs.*
- Sandbox tests – These tests detect if the system is some sort of sandbox. The tests measure if there is no user interaction or mouse movement.

## 4.3 Defeating the paranoid fish timing tests

Most Pafish tests can be defeated easily using the following methods:

- Instruction replacement – The CPUID operating can be detected and replaced by the hypervisor. In that method, the first time that CPUID is called a context switch will occur and, indeed, take a long time; however, further instructions can be replaced with another instruction that generates a similar response (such as generating an interrupt and returning with iret).
- Tampering with RDTSC - The hypervisor is allowed to intercept calls to RDTSC and similar calls and provide its own responses. The hypervisor may run its own code and return values that are smaller than the values obtained from the CPU. These values can take into consideration time spent in the hypervisor. However, RDTSC time measurements are needed for the operating system to function properly. Tampering with RDTSC return values may cause system instability.
- Intel allows the time measurements (in CPU cycles) on the VM to be slower than on the actual machine; for example, every two CPU cycles on the physical PC can be considered as only one CPU cycle on the guest. This method can offset several of the red pill tests as operations no longer take “too much time.” However this test works with a constant multiplication factor. This is not the case in real life. Some CPU opcodes require context switches and are much slower when a hypervisor is present while other operations take exactly the same amount of time regardless of hypervisor presence. Thus, these tests can be offset by performing operations that do not take longer on a VM (such as ADDing, ORing, and NOPing). It is possible to create a sequence of very fast operations that do not run slower on a VM. If the VM is running some sort of clock adjustment were VM cycles are slower than real world cycles, then these operations may seem to take too little time on the VM. (creating yet another type of red pill)

## 4.4 Augmenting the paranoid fish tests

The Pafish timing test runs in user mode and examines only a limited set of instructions (for example, only RDTSC and not RDTSCP). By augmenting Pafish timing tests to run in kernel mode and trying multiple instructions (RDTSC, RDTSCP etc.), the Pafish tests can become more potent and harder to avoid.

Other well-known tests not found in Pafish include verifying the IDT pointer. (that is easily avoided in modern systems with EPT)

## 4.5 Nested virtualization tests

One of the tests not performed by Pafish is performing virtualization instructions (calling VMXON, VMREAD etc.) and performing complex virtualization operations such as EPT or IOMMU operations.

These instructions require a significant effort to support nested virtualization. Running these instructions efficiently poses extra complexity on the blue pill hypervisor, which is not even implemented by some commercial hypervisors such as Oracle's Virtual Box.

Calling these instructions introduces additional complexity, which the normal blue pill hypervisor may not meet or may not meet within time and complexity limits.

Like the CPUID instructions, all of Intel's VMX instructions induce an exit from guest to host mode, making it mandatory to emulate and, thus, extending timing attacks to new instruction types and even mixing instructions from the family. Additionally, VMX instructions also modify the flags register in order to notify VMM software of success or failure of the instructions.

#### **4.6 Undocumented opcode-based attacks**

The x86 opcodes vary in length from 1 to 19 bytes. This massive address space leaves huge room for opcodes.

Domas (2017) mapped the opcodes and compared them with the x86 specs. Domas found many ( $\sim 10^5$ ) undocumented opcodes in the x86 architecture. It is speculated that these instructions are used by Intel's engineers for internal testing. While the behaviour of these opcodes is undefined, by definition it is not unlikely that by calling one of these opcodes one could generate a different behaviour if a red pill hypervisor is running. Further research on this subject is required.

### **5. Remote red pills**

"Remote red pills" refers to situations in which there is a trusted third party that is available to test the inspected system.

These tests make more sense than local test cases as they do not rely on a local system that is currently considered untrustworthy.

If such a remote system is available, then these tests can provide responses with greater confidence than local tests. Unfortunately, there are many real-world cases and scenarios where such reliable third party systems do not exist.

#### **5.1 Kennell's timing method and derived attacks**

Kennell et al. (Kennell 2003) proposed a method to perform a remote hypervisor red pill based on computation side effects on the inspected system. Such side effects include TLB and cache hits and misses as well as real world time that performing the computation consumes.

According to Kennell's method, the attested computer receives a "challenge" (computation request) from the trustworthy remote server.

The computation of the challenge causes several computational side effects such as TLB hits, TLB misses, cache hits, cache evictions, etc. The challenge contains several stages in each of these stages the side effects of the prior stage are added to the computation result. Thereby the prior stages result, and side effects affect the result of the next phase computation. To pass the test, the tested computer must not only produce the same results for the computation itself but also compute accurate results for the side effects as well. Furthermore, the entire computation (of all stages) must be completed in a short time (the time it would take a non-virtual machine to calculate). The kennell test relies on the fact that if a blue pill or an emulator is running then the computation side effects are bound to be different. Thus calculation of side effects must be done separately and consume more time. Therefore, it is impossible for the response to the challenge on an emulated machine or virtual machine to be right and arrive on time. Furthermore, since the challenge is constructed from many stages that must be computed in the correct order the Kennell test cannot be emulated on parallel machines. The Kennell test will declare the machine is not genuine if the answer is wrong or arrives too late.

Kennell's method came under direct attack the following year. Shankar et al. (2004) claimed that performance side effects are not sufficient as a method for software detection.

Kennell et al (Kennell 2004) has answered these claims and the matter rested until virtualization became commonly available in modern PCs.

Kennell's method cannot be emulated directly on modern system as modern systems are more complex than the model assumed by kennell with multiple caches. Also EPT provides much faster memory traversal even if an hyper visor is present.

Kiperberg et al (Kiperberg et al 2013 and Kiperberg et al 2015) claimed that the Kennell method can be replicated on modern PCs with hardware virtualization. This result was short-lived as Intel changed their caching algorithm the following year (between 2<sup>nd</sup> and 3<sup>rd</sup> generation of core processor). Furthermore, Intel has not shared their caching algorithms.

However, Kennell tests rely on the availability of certain algorithms, such as CPU caching algorithms, which are not commonly available. These algorithms are considered trade secrets. Furthermore, Intel has changed the caching algorithms of their core platform between the second and third generations and changed them again to combat the "meltdown" (Lipp et al. 2018) and "specter" (Kocher et al. 2018) weaknesses. Thus supporting the Kennell tests on modern hardware require reversing the architecture of the caching algorithm. It is difficult and extremely time consuming. Supporting the Kennell tests on all recent intel/AMD architectures can be a menial task that will require further research.

## **5.2 Network-based attacks**

Because Kennell has shown that several operations take a significantly longer time when a blue pill is running, it is possible to construct a network-based rootkit using other network servers instead of a Kennell-based attestation server.

For example, if a network-based NTP (network time) server is available and can be considered trustworthy, then it is possible to construct a "network-based blue pill."

The network-based red pill scheme is as follows:

- Query a network time server.
- Perform instructions that may take significantly longer if a blue pill is present (for example CPUID) \* N times.
- Query the network time server and obtain the difference.
- Compare the time required to perform CPUID against a threshold.

The time measurements can be done more accurately by pinging other computers in the network as well as the default gateway.

## **6. TPM-based red pills**

The Trusted Platform Module (TPM) is a device that is found on most modern computer systems. The TPM includes several performance counters that take measurements during the system boot.

The TPMs are developed using hardware obfuscation methods that cannot be reversed easily.

One of the main goals of the TPM is to attest the hardware and software currently running.

This is performed using the following steps:

- Establish a trust nexus.
- Maintain a chain of trust.
- Perform TPM attestation (remote and local).

### **6.1 Establishing a trust nexus and chain of trust**

The TPM chip itself is the trust nexus. The TPM is built using hardware obfuscations and its internals cannot be easily reversed. Despite attacks on older TPM models, there are no known attacks on recent TPM modules. If

the risk that a modern TPM chip will be hacked is accepted (or considered negligible), then the TPM can serve as a root of trust (or trust nexus) for the application. Once the TPM itself has been attested the new modern technologies such as secure boot can be used to create a chain of trust.

## **6.2 Maintain the chain of trust**

Once the TPM has attested the BIOS or initial hypervisor it is up to the next phase to maintain trust. For example, the hypervisor instructions may be disabled in BIOS and the OS can ensure that only trusted software is loaded. Alternatively a hypervisor can be started from UEFI or MBA and ensure that only trusted OS is loaded. The trusted OS will load only trusted applications and so on and so forth.

## **6.3 Perform TPM attestation**

The TPM can be attested directly or through trusted third party thus assuming 6.1 and 6.2 holds the system can be considered attested.

## **7. Conclusion and recommendation**

As demonstrated above, there are numerous ways in which a hypervisor can hide and avoid detection and numerous ways in which a hypervisor can be detected.

It is recommended taking the following precautions as a malicious blue pill must be avoided at all costs:

- Install a TPM on the system and attest the UEFI BIOS.
- Start an attested trusted hypervisor that will prevent any subversion attacks.
- Update your code if new subverting attacks are discovered and published (0-days).

If a red pill is required and a primordial hypervisor cannot be installed, then the recommended red pill method is to use remote attestation methods whenever possible.

Local attestation by the local system should be avoided, especially if the malicious hypervisor's designers can review the local attestation code and design the hypervisor so that it passes the local test with high probability.

However, if local attestation cannot be avoided, then it is recommended to use the multiple cores available in all modern CPUs for time-based attestation. Such methods should be used as a last resort and remote attestation or network timing-based methods are preferred. A malicious blue pill hypervisor can block these methods; however, tampering with NMIs between cores will usually introduce noticeable latency to the system.

It is also possible that unknown opcodes can be researched to provide new red pills.

## **References**

- Averbuch, A., Kiperberg, M., & Zaidenberg, N. J. (2013). Truly-Protect: An efficient VM-based software protection. *IEEE Systems Journal*, 7(3), 455-466.
- David, A. and Zaidenberg, N., (David et al 2014), October. Maintaining streaming video DRM. In *Proc. Int. Conf. Cloud Security Manage (ICCSM)* (p. 36).
- Domas, C. (2017). Breaking the x86 ISA Blackhat, USA
- Garfinkel, T., Pfaff, B., Chow, J., Rosenblum, M., & Boneh, D. (2003, October). Terra: A virtual machine-based platform for trusted computing. In *ACM SIGOPS Operating Systems Review* (Vol. 37, No. 5, pp. 193-206). ACM
- Kennell, R., & Jamieson, L. H. (2003). Establishing the Genuinity of Remote Computer Systems. In *USENIX Security Symposium* (pp. 295-308).
- Kennell, R., & Jamieson, L. H. (2004). An analysis of proposed attacks against genuinity tests. *CERIAS, Purdue Univ., West Lafayette, IN, USA, Tech. Rep*, 27, 2004.
- Khen, E., Zaidenberg, N. J., & Averbuch, A. (2011, June). Using virtualization for online kernel profiling, code coverage and instrumentation. In *2011 International Symposium on Performance Evaluation of Computer & Telecommunication Systems* (pp. 104-110). IEEE.
- Khen, E., Zaidenberg, N. J., Averbuch, A., & Fraimovitch, E. (khen et al 2013). Lgdb 2.0: Using Iguest for kernel profiling, code coverage and simulation. In *2013 International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS)* (pp. 78-85). IEEE.
- Kiperberg, M., & Zaidenberg, N. (Kiperberg et al 2013) Efficient Remote Authentication. In *Proceedings of the 12th European Conference on Information Warfare and Security: ECIW 2013*(p. 144). Academic Conferences Limited.

- Kiperberg, M., Resh, A., & Zaidenberg, N. J. (Kiperberg et al 2015). Remote Attestation of Software and Execution-Environment in Modern Machines. In *2015 IEEE 2nd International Conference on Cyber Security and Cloud Computing* (pp. 335-341). IEEE.
- Kiperberg M, Algawi A, Leon R, Resh A. & Zaidenberg N. J. Hypervisor-assisted Atomic Memory Acquisition in Modern Systems (Kiperberg et al2019a) in Proceedings of 5<sup>th</sup> international conference on information system security and privacy ICISSP 2019
- Kiperberg, M., Leon, R., Resh, A., Algawi, A. and Zaidenberg, N.J., (Kiperberg et a; 2019b). Hypervisor-based Protection of Code. In *IEEE Transactions on Information Forensics and Security*.
- Kocher, P., Genkin, D., Gruss, D., Haas, W., Hamburg, M., Lipp, M., ... & Yarom, Y. (2018). Spectre attacks: Exploiting speculative execution. *arXiv preprint arXiv:1801.01203*.
- Lipp, M., Schwarz, M., Gruss, D., Prescher, T., Haas, W., Mangard, S., ... & Hamburg, M. (2018). Meltdown. *arXiv preprint arXiv:1801.01207*
- Ortega, A. (2016) Paranoid Fish <https://github.com/a0rtega/pafish>
- Popek, G. J., & Goldberg, R. P. (1974). Formal requirements for virtualizable third generation architectures. *Communications of the ACM*, 17(7), 412-421.
- Resh, A., Kiperberg, M., Leon, R., & Zaidenberg, N. J. (2017). Preventing Execution of Unauthorized Native-Code Software. *International Journal of Digital Content Technology and its Applications*, 11.
- Rutkowska, J. (Ruthkowska 2006a). Introducing blue pill. *The official blog of the invisiblethings. org*, 22, 23
- Rutkowska, J. (Ruthkowska 2006b). Subverting VistaTM kernel for fun and profit. *Black Hat Briefings*.
- Rutkowska, J., & Tereshkin, A. (2007). IsGameOver () anyone. *Black Hat, USA*.
- Shankar, U., Chew, M., & Tygar, J. D. (2004). Side effects are not sufficient to authenticate software. In *USENIX Security Symposium* (Vol. 8, No. 3).
- Zaidenberg, N. J. (Zaidenberg 2018). Hardware Rooted Security in Industry 4.0 Systems. *Cyber Defence in Industry 4.0 Systems and Related Logistics and IT Infrastructures*, 51, (pp 135-151).
- Zaidenberg, N. J., & Khen, E. (Zaidenberg 2015b). Detecting Kernel Vulnerabilities During the Development Phase. In *2015 IEEE 2nd International Conference on Cyber Security and Cloud Computing* (pp. 224-230). IEEE.
- Zaidenberg, N. J., & Resh, A. (Zaidenberg et al 2015c). Timing and side channel attacks. In *Cyber Security: Analytics, Technology and Automation* (pp. 183-194). Springer, Cham.
- Zaidenberg, N., Neittaanmäki, P., Kiperberg, M., & Resh, A. (Zaidenberg et al 2015d). Trusted Computing and DRM. In *Cyber Security: Analytics, Technology and Automation* (pp. 205-212). Springer, Cham.

# Information Technology Governance: The Role of Board of Directors in Cybersecurity Oversight

**Abdalmuttaleb Al-Sartawi**

**Ahlia University, Manama, Bahrain**

[amasartawi@hotmail.com](mailto:amasartawi@hotmail.com)

**Abstract:** The MENA region is a target for many cyber-attacks, as they are consumers of technology instead of being innovators. MENA companies need to protect their data, and the BOD need to embed a culture of information security in the company starting from the top managers to the lower ranked employees. This makes cybersecurity a subset of a BOD's responsibilities. The aim of this paper is to measure the level of cybersecurity disclosure and governance in the MENA region and examine the relationship between cybersecurity governance and the level of companies' cybersecurity. The study used a checklist to collect data from a sample of 57 firms listed in the financial stock markets of the MENA countries for the year 2018. The study concludes that there is a significant and direct relationship between IT governance and the level of a firm's cybersecurity. This indicates the importance of appointing board members with IT knowledge and experience which leads to better decision-making when faced with cyber-threats and challenges.

**Keywords:** cybersecurity, information technology governance, MENA region, board of directors

---

## 1. Introduction

Due to the advancement of the new digital age and the rise of the internet, information has become a very valuable asset, making it vulnerable to alteration, theft, and denial of access (Gordon et al., 2010). More than 50 countries around the world have officially published strategy documents defining their stance on information security breaches, cybercrimes, cybersecurity readiness (Solms and Niekerk, 2013). Based on the Information Telecommunications Union (ITU) (2008), cybersecurity is the compilation of policies, procedures, security concepts, guidelines, risk management approaches, training, best practices, assurance, and technologies that can be used to safeguard the cyber environment and individual or organizational assets.

Cybercrimes and threats continue to evolve and become more sophisticated. Poor cybersecurity policies lead to a disruption of a business by endangering its clients' information and risking the exposure of its secrets to competitors and hackers. Ishiguro et al., (2006) argues that cybersecurity breaches might have adverse effects on firm value. The question is: Is cybersecurity a board issue? From a corporate governance standpoint, the role of board of directors (BOD) in overseeing cybersecurity has become threefold since the large data breaches in the recent years, which made cybersecurity a central issue for governments, regulators, shareholders as well as potential investors. Data breaches damage a company's image in the eye of shareholders and investors which lead to a fall in stock prices. Osterman Research (2016) conducted a survey to gauge the perceptions of board members on security and cyber-risk analytics. The study found that the chief reason for cybersecurity being a top priority is the complexity of compliance regulations. About 33% of the board members agreed that cybersecurity is mainly a technical issue, i.e., it requires technical experience.

Another significant corporate governance issue currently faced by the modern firms is related to Information Technology Governance (ITG). Previous research shows that information technology has a direct effect on corporate governance (Farhanghi et al., 2013). According to the website of the IT Governance Institute (ITGI), ITG is a subset of corporate governance which focuses on information technology systems, and how their performances are measured, and risks managed. McCollum (2006) claim that the Sarbanes-Oxley Act of 2002 has alerted board of directors to their firm's need to prioritize Information technology governance. Therefore, due to the critical need of ITG, Huff et al. (2005) demands that an IT expert should be appointed to the board to provide various views based on practical experience or background they have in the field of information technology.

As with developed countries such as the USA, the MENA region face more or less the same issues. The primary objective of this paper is to link two fundamental topics of corporate governance: cybersecurity and IT Governance. The study was undertaken to address the following main research questions. First, what is the extent of Information Technology Governance (ITG) by MENA listed firms? What is the level of cybersecurity by MENA listed firms? Finally, what is the relationship between ITG and the level of cybersecurity in the MENA region? To the researcher' knowledge, there are negligible studies which investigate such an issue, especially

from a MENA region perspective. Thus, this paper aims to contribute to the literature on ITG and cybersecurity in the MENA.

## **2. Literature review and hypothesis development**

Solms and Niekerk (2013) differentiated between the ever-interchangeable cybersecurity and information security. Although the two synonymous words share several characteristics such as availability, integrity (authenticity and non-repudiation), and confidentiality. The authors concluded that while information security is the protection of information from possible harm resulting from various threats and vulnerabilities, cyber security is more than that. Cybersecurity not only includes the protection of cyberspace itself, but also the protection of those that function in cyberspace and any of their assets that can be reached via cyberspace.

Newbound (2016) claims that there are several high-profile companies whose reputation has suffered due to their failure in safeguarding their customers' data from hackers. These scandals mean that cybersecurity should become a standing item on boardroom agendas. The study further argues that board of directors should ensure they understand all security measures, procedures and policies that their firm has in place, and failure to do so could result in personal liability. Similarly, based on the Cheng and Groysberg (2017), board members fail to make the connection between the continuous cybersecurity lapses and their own firms' vulnerabilities to cyber-threats. They suggest that board of directors should take corrective actions such as holding regular discussions of cybersecurity at board meetings. In addition, managers need to bring in outside experts to conduct briefings. This study proposes a third solution: IT governance.

Rau (2004) defines IT governance as the way senior management interacts and communicates with IT leaders to ensure that technology investments enable the achievement of business strategy in an effective and efficient manner. Likewise, Li et al, (2007) describe IT governance as leadership, organizational structures, and control processes which ensure that the IT of the firms is able support and expand the company and achieve its objectives. Nolan and McFarlan (2005) regarded ITG as "a vital asset that requires intense board-security and assistance". Therefore, as ITG increases firms' responsiveness to stakeholders, it can be applied to electronic financial disclosure.

Moreover, along with Nolan and McFarlan (2005), Andriole (2009) recommends that an ITG group of expert independent directors need to be appointed to the board, as in the case with the audit committee, due to economic and regulatory matters related to the Sarbanes-Oxley compliance and corporate governance. Thus, these directors are required to be experts in IT, and understand the dynamic potential of information technology in changing the business environment. This notion is supported by the resource dependence theory that state the importance of board of directors as resources for the firm in order to reach external resources (Ribeiro and Colauto, 2016). Despite this need for an IT expert on the board, Valentine and Stewart (2013) found that board level willingness to reduce the gap between awareness and action is very low or non-existent. On the other hand, in the case of the MENA, more specifically the UAE, Nicho and Khan (2017) found that the objectivity of ITG is well-defined and emphasized.

Alan (2014) states that the board needs to have a full understanding of IT governance issues as ITG is an essential part of corporate governance. The study further found that board members believe that fear of retribution might discourage the IT department from fully disclosing details of cyber breaches to top management. Therefore, the involvement of the board in ITG is of utmost importance. Moreover, Alan (2014) found that many boards still lack the necessary knowledge and qualifications to oversee cybersecurity effectively, and that their firms had no plans to prioritize training due to budgetary constraints. The study concluded that a lack of boardroom expertise may partly explain the reluctance of some companies to give up outdated cyber security goals and evolve with the ever-evolving cybercrimes and threats. Similarly, Dramis (2015) claim that today every firm has become a technology firm with data, marketing, compliance, logistics and finance all dependent on digital support structures. Hence, ITG in the form of a cyber/technological committee is crucial.

This study therefore hypothesizes that:

**H1: There is a relationship between the level of Information Technology Governance and the level of Cybersecurity.**

### 3. Research methodology

In order to address the research questions, mainly the relationship between ITG and the level of cybersecurity in the MENA region, the current study collected data from a sample of 57 firms listed in the financial stock markets of the, MENA for the year 2018. To measure the level of cybersecurity, a checklist was developed using a number of different sources such as: (1) the National Institute of Standards and Technology's (NIST) cybersecurity framework and, (2) the Federal Communications Commission's cybersecurity planning guide. When a firm check marked any item on the checklist the item received a score of 1, and 0 when otherwise.

Data collection went through two phases. Phase one included collecting data from the official websites of the firms, strategy, governance or policy documents, as well as the annual reports for the year ended December 2017 and March 2018. However, as not all the information was available from the secondary sources, the researcher sent emails soliciting the data required to complete the checklist. IT specialists, IT security consultants, and IT security engineers were targeted, and their emails were retrieved from the websites of the listed firms in the financial sectors of the selected countries' stock exchanges. The checklist was then sent to the targeted potential respondents at the end of September 2018. Follow-up emails were sent, and by the end of November 2018, 57 responses were received. Therefore, only completed checklists were included in the study, while the rest of the firms were discarded from the sample. Table 1 shows the distribution of the sample among the selected MENA countries.

Additionally, the study measures the level of ITG in the MENA firms by determining the percentage of the members of the board of directors who have an IT background or experience. This data was extracted from the annual reports and the websites of the sample firms.

**Table 1:** Distribution of sample among countries

Bahrain				UAE (Dubai)			
Sectors	Total	Sample	%	Sectors	Total	Sample	%
Commercial banks	7	3	13.64	Banks	13	8	23.5
Investment banks	10	4	18.18	Insurance and Financial	21	9	26.5
Insurance	5	0	0				
Total Sample	<b>22</b>	<b>7</b>	31.82%	Total Sample	<b>34</b>	<b>17</b>	50%
Jordan				Palestine			
Sectors	Total	Sample	%	Sectors	Total	Sample	%
Banks	15	6	9.09	Financial Services	7	3	12.5
Insurance	20	7	10.61	Investment	10	3	12.5
Financial Services	31	12	18.18	Insurance	7	2	8.33
Total Sample	<b>66</b>	<b>25</b>	37.88%	Total Sample	<b>24</b>	<b>8</b>	33.33 %
<b>Total firms in the sample: 57 out of a total of 146 firms</b>							

To test the hypothesis, the following regression model was developed using the level of cybersecurity as a dependent variable, and ITG as an independent variable. Additionally, the study used firm size, board size, and firm age as control variables.

**Model:**

$$CySec_i = \beta_0 + \beta_1 ITG_i + \beta_2 L\_FSZ_i + \beta_3 BD\_size_i + \beta_4 AGE_i + \varepsilon_i$$

**Table 2:** Study variables

Code	Variable Name	Operationalization
<b>Dependent variable</b>		
CySec	Cybersecurity	Total scored items by the company/Total maximum scores
<b>Independent Variables</b>		
ITG	Directors with IT background and experience %	The percentage of members who have IT background and experience to the total board size
<b>Control Variables:</b>		
L_FSZ	Firm size	Natural logarithm of Total Assets

<b>Code</b>	<b>Variable Name</b>	<b>Operationalization</b>
BD_size	Board size	Number of members on the board
AGE	Firm Age	The difference between the establishing date of the firm and the report date
$\epsilon_i$		Error

#### 4. Data analysis

Table 3 reports the overall level of cybersecurity among the selected sample of MENA countries, while table 4 reports the descriptive analysis of the independent, dependent and control variables. The UAE had the highest level of cybersecurity (83.4 %) among the sample countries, while Palestine had the lowest level (69.4%). One reason for the high level of cybersecurity in the UAE is that cybersecurity is the number one priority for firms in the UAE (Ryan, 2018). Based on the ‘Dubai Cyber Security Strategy’, accessed through the Dubai’s government website, the cybersecurity strategy was developed to unify the efforts of government institutions and firms to make Dubai the safest electronic city in the world.

**Table 3:** Level of cybersecurity among the selected MENA countries

<b>Country</b>	<b>No.</b>	<b>Mean</b>	<b>S.D.</b>
Bahrain	7	0.798	0.152
UAE	17	0.834	0.078
Jordan	25	0.726	0.174
Palestine	8	0.694	0.190
Total	57	0.763	0.149

Table 4 reports that the maximum level of information technology governance (ITG) was 55% by the sample firms while the minimum was 0 % with a mean of 7.6% indicating a low level of ITG. According to Valentine and Stewart (2013), there is a gap between the need for ITG need and reality. With regards to the level of cybersecurity, the MENA sample countries achieved a maximum level of 92%, with an overall mean of 76.3%, which is considered as a moderate level.

**Table 4:** Descriptive statistics

Descriptive Statistics Continues Variables					
	N	Minimum	Maximum	Mean	Std. Deviation
ITG	57	0.00	0.54	0.076	0.10719
CySec	57	0.58	0.92	0.763	1.022
L_FSZ	57	84526	173965317	9.3E49	4.679E9
B_size	57	5	13	8.72	1.27718
AGE	57	1	174	36.22	14.665

On the other hand, the descriptive statistics for control variables show that the mean of firm size, i.e. Total Assets, was 9.3 million, with a minimum of 84526 and a maximum 173 million, indicating large firms. The normality distributions of the total Assets were skewed, so natural logarithm was used in the regression analysis to reduce skewness and bring the distribution of the variables nearer to normality. The average board size is 9 members, while, firm age ranges from 1 to 174 with a mean of 36.22.

Two main tests were used to test the validity of data. Shapiro-Wilk was used to test the normal distribution. In accordance to (Gujarati, 2003) the data is normally distributed if the significance is greater than 0.05. The researcher concluded in table (5) that the variables of this research generally are normally distributed, except for firm size which had a significance level of 0.000. To solve this issue, the natural logarithm was used in the regression analysis to bring the distribution of the variables nearer to normality.

**Table 5:** Normality test

	Shapiro-Wilk Test		
	Statistic	df	Sig.
ITG	2.717	57	0.458
CySec	1.925	57	0.481
L_FSZ	2.892	57	0.000
B_size	2.727	57	0.302
AGE	1.843	57	0.210

Furthermore, the Variance Inflation Factor (VIF) test was used to check the data for multicollinearity. The results as shown in table 6 indicate that since no VIF score exceeded 10 for any variable in the model, and as no Tolerance score was below 0.2, it was concluded that there is no threat of multicollinearity.

**Table 6:** Collinearity test

Model	Collinearity Statistics	
	Tolerance	VIF
ITG	0.857	1.004
L_FSZ	0.885	1.035
B_size	0.997	1.214
AGE	0.976	1.162

Table 7 reports the findings of the regression analysis. The findings indicate that the model demonstrates the relationship between the dependent and independent variables in a statistically appropriate way. According to the table, the model has an adjusted R<sup>2</sup> of 0.189 which shows that the model explains approximately 18.9% of the variation in the level of cybersecurity amongst the MENA listed firms. Moreover, the probability of the F-statistic with a significance 0.000 means that the ITG was significant in interpreting the level of cybersecurity.

**Table 7:** Regression analysis

	Beta	T-test	Sig.	R <sup>2</sup>	F	Sig. (F)
Model (CySec)	ITG	0.238	2.207	0.036		
	L_FSZ	0.008	0.063	0.028		
	B_size	-0.412	-6.473	0.641		
	AGE	0.181	1.720	0.338	8.220	0.000

The main hypothesis of the study states that there is a relationship between ITG and the level of cybersecurity by firms listed in the MENA stock markets. Table 7 shows that there is a significant and direct relationship between study variables. We can therefore assume that the higher the level of ITG on the board, the higher the level of cybersecurity. This is in line with many of the previous studies which state the importance of IT literacy at the board level. As cyber security issues are directly related to the cyber security architecture, which is part of the IT architecture. This indicates that board members with IT qualifications, knowledge and experience have a better understanding of the challenges which their firms face and can hence make well-informed decisions.

With regards to the control variables, the study found a significant and positive relationship between the level of cybersecurity and firm size. As for the rest of the control variables, the results show no relationship between board size, age and the level of cybersecurity.

## 5. Conclusion

The primary objective of this study was to investigate the relationship between the level of information technology governance (ITG) and the level of cybersecurity by MENA listed firms. To answer the research questions, the researcher collected data from a sample of 57 firms listed in the financial stock markets of the MENA region countries for the year 2018. The countries included in the sample were Bahrain, UAE, Jordan and Palestine. The findings indicated that the UAE had the highest level of cybersecurity (83.4 %) among the sample countries, while Palestine had the lowest level (69.4%). Moreover, the results show that there is a significant and direct relationship between study variables. Therefore, the higher the level of ITG on the board, the higher the level of cybersecurity. This indicates the importance of appointing board members with IT knowledge and experience which leads to better decision-making when faced with cyber-threats and challenges.

This paper contributes a new topic to the MENA region literature by combining two significant areas, namely information technology governance and cybersecurity. The study would also be of interest to the international investment community, regulators, policy makers and governments in the MENA region and other developing countries due to the similarities in their technological infrastructure. However, the checklist developed for this study could be used by both developing and developed countries. In addition, this paper offers a practical contribution that could be useful to shareholders when appointing board members or forming technological/cyber committees.

While cyber-threats and the awareness of cyber-security are constantly evolving in the MENA region, this paper recommends firms to develop effective cybersecurity programs that address current regulatory compliance requirements and prepare for emergency cyber responses. Furthermore, to address a gap in the MENA region literature, this paper suggests conducting a study that further investigates the relationship between the level of cybersecurity and firm performance in the form of market value, earnings per share or profitability.

Cyber Security Level (CSL) Index			
1	Cybersecurity is a precondition to work and is considered in all business decisions	18	Background credentialing check on employees are conducted before hiring them.
2	Using a Firewall software.	19	Implementing network and cloud monitoring.
3	Cybersecurity policies are documented.	20	Fault tolerant architecture is in place.
4	Employees are trained on the company's network policy securities.	21	Intrusion detection and prevention systems are in place (IDS/IPS).
5	Enforcing password protection policies.	22	Implementing and managing access agreements with employees (nondisclosure agreements, acceptable use agreements, access agreements)
6	Checking backup regularly to ensure that it is functioning correctly.	23	Third-party personnel security processes are in place.
7	Anti-malware software installed on all devices and the network.	24	Assigning risk designation to organizational positions.
8	Have developed an Incident Response Plan (IRP).	25	Performing regular internal vulnerability audits or assessments.
9	Protocol to revoke access to terminated employees	26	Using a spam email filter.
10	Using encryption software to protect sensitive data.	27	Employees are able to recognize and avoid phishing.
11	Using password-security software.	28	Creating safe-use flash drive policies.
12	Having cybersecurity insurance.	29	Email retention and usage policies are in place.
13	There is a specific guidance on when to disclose company activities using social media, and what kinds of details can be discussed in a public forum	30	Ensuring all smartphones, computers, laptops are wiped clean before disposal.
14	Holding regular workshops and meetings on best cybersecurity practices.	31	There is an employee internet usage policy that is personal breaks to surf the web are be limited to a reasonable amount of time and to certain types of activities.
15	Notice triggering information system is in place.	32	Executives and managers are involved in cybersecurity.
16	There is a clear strategic plan in place for the protection of critical data and essential services	33	Wi-Fi network is secured by setting up wireless access point or router.
17	Have created a mobile device action plan.	34	Prioritizing services based on analysis of the potential impact if the services are disrupted.

## References

- Alan, C. (2014). "The Boardroom View on Cyber Security", *Corporate Board*, 35 (208), p11-15.
- Andriole, S. J. (2009). "Boards of directors and technology governance: The surprising state of the practice", *Communications of the Association for Information Systems*, 24(1), 22.
- Cheng, J.Y.J., and Groysberg, B. (2017). "Why Boards Aren't Dealing with Cyberthreats", *Harvard Business Review*, [Available at: <https://hbr.org/2017/02/why-boards-arent-dealing-with-cyberthreats>]
- Dramis, F.A. (2015). "Time for A Board Cyber/Tech Committee?", *Corporate Board*, 36 (213), p1-5.
- Farhanghi, A.A., Abbaspour, A. and Ghassemi, R.A., (2013). "The effect of information technology on organizational structure and firm performance: An analysis of Consultant Engineers Firms (CEF) in Iran", *Procedia-Social and Behavioral Sciences*, 81, pp.644-649.
- Gordon, L. A., Loeb, M. P., and Sohail, T. (2010). "Market value of voluntary disclosures concerning information security", *MIS quarterly*, 567-594.
- Harvard Business Review. 2017. Vol. 95 Issue 3, p36-36. 1/3p
- Huff, S.L., Maher, P.M. & Munro, M.C. (2005). "Adding value: The case for adding IT savvy directors to the board", *Ivey Business Journal*, 2, 1-5.

***Abdalmuttaleb Al-Sartawi***

- International Telecommunications Union (ITU). (2008). "ITU-TX.1205: series X: data networks, open system communications and security: telecommunication security: overview of cybersecurity".
- Ishiguro, M., Tanaka, H., Matsuura, K., and Murase, I. (2006). "The effect of information security incidents on corporate values in the Japanese stock market", In International Workshop on the Economics of Securing the Information Infrastructure (WESII).
- Li, C., Lim, J., and Wang, Q. (2007). "Internal and external influences on IT control governance", *International Journal of Accounting Information Systems*. 8, 225-239.
- Li, C., Peters, G. F., Richardson, V. J., & Watson, M. W. (2012). "The consequences of information technology control weaknesses on management information systems: The case of Sarbanes-Oxley internal control reports", *MIS Quarterly*, 36(1), 179-203
- McCollum, T. (2006). "Bridging the Great Divide", *Internal Auditor*, 1, 49-53
- Newbound, D. (2016). "Why Cyber Security Matters?", *Credit Control*, Vol. 37 Issue 3/4, p19-21. 3p
- NICHO, M. and KHAN, S. (2017). "IT governance measurement tools and its application in IT-business alignment", *Journal of international technology and information management*, 26(1), Article 5.
- Nolan, R. and McFarlan, F.W. (2005). "Information Technology and the board of directors", *Harvard Business Review* [Available at: <https://pdfs.semanticscholar.org/9149/ab6cb4c7fa9a3d39709f9ae75f804b0db5a4.pdf>].
- Osterman Research. (2016). "What's Driving Boards of Directors to Make Cyber Security a Top Priority?", *Bay Dynamics*, [Available at: <https://baydynamics.com/content/uploads/2016/09/BoardSecurityOstermanReport.pdf>]
- Rau, K.G. (2004). "Effective Governance of IT: Design Objectives, Roles and Relationships", *Information Systems Management*. 21, p. 35-42.
- Ribeiro, F., and Colauto, R. (2016). "The Relationship Between Board Interlocking and Income Smoothing Practices", *R. Cont. Fin. – USP, São Paulo*, 27(70), 55-66.
- Solms, R., & Van Niekerk, J. (2013). "From information security to cyber security", *Computers & security*, (38) 97-102.
- Valentine, E. & Stewart, G. (2013). "The emerging role of the Board of Directors in enterprise business technology governance", *International Journal of Disclosure and Governance*, 10 (4), 346-362.

# Revisiting Cyber Definition

Riza Azmi<sup>1</sup> and Kautsarina<sup>2</sup>

<sup>1</sup>University of Wollongong, Australia

<sup>2</sup>Universitas Indonesia, Depok, Indonesia

[ra873@uow.mail.edu.au](mailto:ra873@uow.mail.edu.au)

[kautsarina61@ui.ac.id](mailto:kautsarina61@ui.ac.id)

**Abstract:** We often use the term cyber in many recent conversations and statements, as well as various official documents, but understanding the meaning still leaves a question mark. The term cyber has remains in various perspectives spanning from definition and domain. While the meaning can be easily grasp which associates to computers or computer networks (such as the Internet), there is currently no consensus on this definition. Up to recent, we found that there are several diverge understanding on this jargon, such as on the semantic discussion (adjective vs noun), and domain (such as the confusion that tangling information security and cyber security). The Cooperative Cyber Defence Center of Excellence (CCDCOE) collects various definitions that show this jargon, although prevalence in many national and international statements is interpreted differently. From those definitions, the term cyber can be cascaded to five areas: 1)Physical infrastructure; 2) Communication; 3) System, 4) Devices and 5) Virtual environment. In this article, we will revisit the term cyber taken from nations/states, different organisation, and scholars interpret and picture the term cyber. This article aims to abridge definitions of cyber from several perspectives. In going revisiting the cyber definition, we walk through from several contexts to abridge the definition of cyber. We start by walking from discussing the diverse perspective of the word cyber, not only the adjective vs the noun that is widely used, but also the confusion to other domains, such as information security vs cyber security. This will be covered in the Section 1. In the Section 2, we will revisit the term by seeing this jargon from some perspectives: linguistic genealogy, context of cyber world, and we use of this word in today context, which we will find vary. In the Section 3, we will coincide different pictures and interpretations to cyber into a single frame of definition. To conclude, we propose the abridge definition of cyber from overall context.

**Keywords:** definition, cyber, cybernetics, cyberspace, cyber security

---

## 1. Introduction

The term cyber is commonly used as a jargon to describe computer, network, and related things to broadly described as internet and its virtual environment (Merriam-Webster 2017; Oxford Dictionary 2017). Despite its prevalence used in the media, national, and international organisation conversations and statements, this term is largely understood in many different ways (CCDCOE 2017; Lehto 2015). In this article, we will revisit the term cyber taken from nations/states, different organisation, and scholars interpret and picture the term cyber. This article aims to abridge definitions of cyber from several perspectives.

This paper start in discussing the diverse perspective of the word cyber, not only *the adjective vs the noun* that is widely used, but also the confusion to other domains, such as information security vs cyber security. This will be covered in the Section 1. In the Section 2, we will revisit the term by seeing this jargon from some perspectives: linguistic genealogy, context of cyber world, and we use of this word in today context. In the Section 3, we will coincide different pictures and interpretations to cyber into a single frame of definition. To conclude, we propose the abridge definition of cyber from overall context.

## 2. “Cyber”: Jargon in the diverge understanding

Although the term *cyber* is a common jargon used in modern life, understanding its meaning is elusive. Until the recent, the term cyber has remains in various perspectives spanning from definition and domain. While the meaning can be easily grasp which associates to computers or computer networks (such as the Internet) (Merriam-Webster 2017; Oxford Dictionary 2017), there is currently no consensus on this definition. We found that there are several diverge understanding on this jargon, such as on the semantic discussion (adjective vs noun), and domain (such as the confusion that tangling information security and cyber security).

### 2.1 Semantic debate: Noun vs. adjective

In a semantic debate, the term cyber is often used as an *adjective* that is emphasising to its corresponding domain, such as the use of term *cyber space* and *cyber security* to modify term “space” which refer to virtual room, and to add a meaning “security” in the cyber space (see semantic discussion in Bayuk et al. 2012; and Ramirez and Choucri 2016). On the other side, the term *cyber* is also used as a noun combined with its

corresponding domain, such as in the use of cyberspace, and cybersecurity. One example of different, the Oxford University (GCS 2014) uses the term “Cyber Security” (with space), while the International Telecommunication Union (ITU 2012), and International Organization for Standardization (ISO/IEC 2012) use the noun word “Cybersecurity” in their definition of framework.

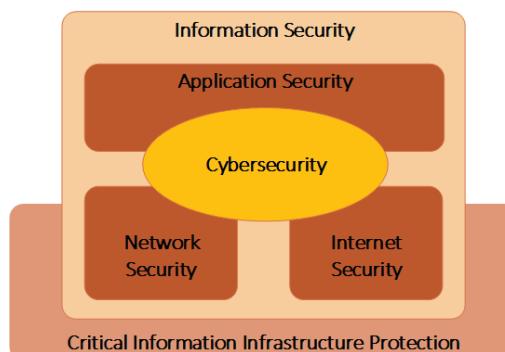
There are arguments scattered which show the different style of *noun vs adjectives*. Yet, those two ways of writing are still used largely on conversations and statements. While the semantic debate seems superfluous for the purpose that the jargon “cyber” is commonly understood as a virtual space (Bayuk et al. 2012), some argue that a term should follow linguistic basis such as clear etymology, enjoy the popular and historical usage by the global community based on its trends, searchable, and well-defined (Ramirez and Choucri 2016). However, those two styles remain used widely.

## **2.2 Domain**

Another confusion is that the use of this jargon is commonly used in occurrence interchangeably with another security domain, for such, cyber security and information security (Luijif et al. 2013). While in contrast to the CIA Triad (Confidentiality, Integrity and Availability) in McCumber’s cube [1991]) which can directly describe what underpins information security, the domain of cyber security is still vague. For example, ITU (2012) and ISO (2012) use the term cyber security and information security as an interchangeable domain.

ITU (2012) defines cyber security as in conjunction to information security which the main aim is to attain and maintain the security properties of confidentiality, integrity, and availability (ITU 2012). The distinctions are information security started in its stand alone and not traverse into other jurisdictions. It implies that information security is only securing its organisation system’s perimeter, not as a global challenge.

In another perspective, ISO defines cyber security seems to be interrelated to information security (see ISO/IEC 2012). To distinguish these two domains as two separated aspects, the ISO defines this into two separated standards. The Information Security (see ISO/IEC 2013) focuses on organisation, while Cyber Security (see ISO/IEC 2012) focuses on collaboration to address issues on security domains in cyberspace. In the ISO/IEC 27032 (2012), security is classified by five domains which is cyber security as the central (Figure 1). It consists of Information Security, Application Security, Network Security, Internet Security, and Critical Information Infrastructure Protection.



**Figure 1:** Relationship between cyber security and others security domain based on ISO/IEC 27032 (ISO/IEC 2012)

In the policy statement, these two domains also seem tied to each other that we can see in some National Cyber Security Strategy (NCSS). In this context, it implies that cyber and information security are tangled between each other. Since information security and cyber security seem to be interrelated, the domain remains to be used interchangeably. The term cyber, although it is widely used and in our modern conversations and documents, seems to be agreed at no consensus of use and to be used in a diverse perspective. Through those diverge perspective of cyber definitions, in the next chapter, we try to detangle the definition of cyber. We start by elucidating this term from the word origin to the context of word use. We also see the common theme on the use of the definition by the organisation. This will be discussed in the Section 3.1 to Section 3.3.

### 3. Revisiting the term “cyber”

#### 3.1 Context 1: History of the word cyber

The term “cyber” has been strayed from its original context. Gibson (1984) introduced the term *cyberspace* in his science fiction novel which is something that represents a virtual environment/an alternative environment. The term *cyber* was derived from the word *cybernetics* (Johnson 2015; Ottis and Lorents 2010; Solomon 2007). Though, this term is widely misunderstood (such as in Ottis and Lorents 2010) of which the integration between biological and non-biological thing which further brings a new thought of possibility to integrate human and machine. It is since the prefix “cyb” in *cybernetics* is widely understood referring to *cyborg* or robot/android (Pangaro 2013).

In linguistic genealogy, the word “cyber” is rooted and can be traced from the ancient Greek word “*kybereo*” (*κυβερνεω*) which means “to assist”, “to steer”, “to guide”, “to control”, “to govern” (Lehto 2013; Maathuis et al. 2017), which strays from the nowadays understanding. The earliest used of this word can be found in the dialog between Plato and Alcibiades (see *Alcibiades I*) which is *kybernetikes* (*κυβερνητικης*) or a steersman/pilot/governor (Liddell and Scott 1940). Plato used this word to highlight the importance of skill in the navigation (Johnson 2015).

In 1843, André-Marie Ampère coined the word “*La cybernétique*” to the France language. In his essay, he wrote that “*the future science of government should be called 'la cybernétique'*” (Ampère 1843, pp. 140–141). Etymologically, this word was adopted from Greek word to the France language which then aspired Norbert Wiener to use this word to introduce the new study field in communication and control (as he previously used the term “*angelos*” ('messenger') but finally settled on “*cybernetics*” which express the same meaning (Johnson 2015)).

It was in 1948, the term *Cybernetics*, which is a derivation of word *kybernetikes* ('steersman') or *cybernétique* ('to govern'), became widely known since Norbert Wiener defines *Cybernetics* in his book as the science of control and communication, in the animal and the machine. It is the science of automated control system in both machines and the living things that require ‘communication’ and ‘feedback’ (Ashby 1957; Johnson 2015; Wiener 1948). Back to that era, this idea was peculiar since non-living (machines) things can have ‘a purpose’ (Pangaro 2013). Since the input is defined and the output is pre-defined, cybernetics study is a teleological mechanism which about the art of studying the black box (system) that includes the interaction between the object and the surrounding elements (Lucas 2004). The zeitgeist of cybernetics is spanning to some study fields, from technical to sociology (Johnson 2015; Pangaro 2017).

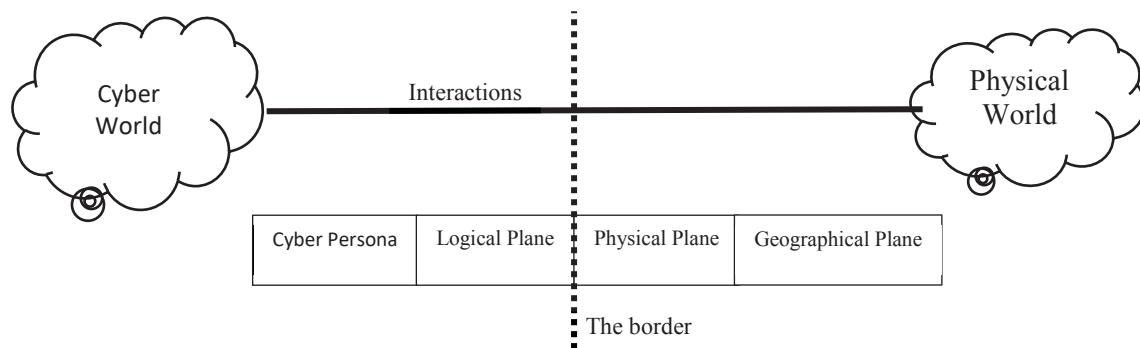
Three decades after, Gibson (1984) in early 1980’s firstly used this word, and brought new term that we use now widely: the cyberspace which rooted from cybernetics (Solomon 2007). This term, “*cyberspace*”, in his science-fiction book describes as “a graphic representation of data abstracted from banks of every computer in the human system.” (Gibson 1984). Since then, the term has entered common usage, including the study of information system's study (Ottis and Lorents 2010). This term began to raise its use in the academic since 1990 (Ramirez and Choucri 2016) and thereafter up until now in modern conversations and statements.

#### 3.2 Context 2: The social structure of cyber world

As discussed in the previous section, the word *cyber* in the nowadays context is widely understood as the abstract and alternative environment enabled by the internet and computer. Therefore, it raises a question of “what is inside this world”. It implies a question on understanding the structure of this abstract word. There are some views to capture this abstract world that we can characterise the cyber world into two characteristics. *First*, cyber world is interdependent to the physical world (Kuusisto and Kuusisto 2015, p. 33). *Second*, cyber world is built by interaction between each node connected to the internet (Kuusisto and Kuusisto 2015; WEF 2014).

The transgression in the cyber world may affect the physical world. Taking as an example the recent event in the end of 2017, the Bitcoin hacking, although asset is virtual and happened in virtual space, it caused people lost their investment in real life, which 64 million dollar worth stolen (Gibbs 2017; Khatwani 2017). *Vice versa*, the

transgression in the physical world may also trigger the cyber world. For example, the crisis between Georgia and Russia, made Georgia creates their National Cyber Security Strategy(Azmi et al. 2016).



**Figure 2:** “The border” of cyber world and physical world

Expanding the thought of Kuusisto (2015) in relation with the tangled system between these two world, “the border” between cyber world and physical world can be distinguish into four layers (borrowed idea from Fanelli and Conti 2012; Lehto 2015; Raymond et al. 2013, 2014), which are Persona, Logical Plane, Physical Plane, and Geographical Plane (see Figure 2).

Cyber persona means plane of identity in the cyber world. This can be referred to one physical individual, an ego, a group of people, a machine, an autonomous system, a software, or an animal, which may hinder real identity anonymously. The logical plane consists of software, operating system, application, and any logical system. The physical layer is the information infrastructure in that is one of pillar of cyber world. This can be hardware, a router, switches, or electricity. The geographical plane is where the location in the real world that impact to the operation in cyberspace. In both direction of interaction, the result of the interaction may affect both space. When the interaction traversing from physical to cyber world, it ‘converts’ information resides in the physical world to the persona (or node), while in vice versa, it converts an idea into a construction/destruction.

Other characteristic of cyber world is that it is consists of collection of node connected to the internet (which is described as persona). Other additional perspective that can outline the social structure of cyber world context is the Social Network Theory, has been adapted by the World Economic Forum (WEF) in viewing cyber world. The WEF perceives the cyber world as the *hyper-connected* environment. It also views that the structural relationship has been changed, from the hierarchical to the “flat” structure. Every single object joined the cyberspace act as an entity with the same social-level, including human, animal, machine, and any other objects.

### 3.3 Context 3: Cyber in today context

Thus far from the previous section we saw the context of cyber from the word history, and what is inside the cyber world. In this Section, we will discuss the term *cyber* used in modern conversations and statements. Table 1 consists of some collection of definition *cyber* and *cyberspace* (see detail list on CCDCOE 2017).

**Table 1:** Cyber definitions

ID	Term	Definition	References
	Cyber	“Cyber” refers to the interdependent network of information technology infrastructures, and includes technology “tools” such as the Internet, telecommunications networks, computer systems, and embedded processors and controllers in critical industries.	WEF (2012b)
	Cyber	The word ‘cyber’ is almost invariably the prefix for a term or the modifier of a compound word, rather than a stand-alone word.	Finland (2013)
	Cyber	“Cyber” is defined as: “anything relating to, or involving, computers or computer networks (such as Internet)”.	Montenegro (2013)
	Cyberspace	The global environment that is created through the interconnection of communication and information systems.	Belgium (2012)
	Cyberspace	Cyberspace is the electronic world created by interconnected networks of information technology and the information on those networks.	Canada (2010)

ID	Term	Definition	References
	Cyberspace	Cyber space means digital environment, enabling to create, process and exchange information, created by information systems and services and electronic communication networks.	Czech Republic (2015)
	Cyberspace	Cyberspace - space in which communication among information systems takes place.	Croatia (2015)
	Cyberspace	Cyberspace is the virtual space of all IT systems linked at data level on a global scale.	Germany (2011)
	Cyberspace	Cyber space is a complex environment consisting of interactions between people, software, and services, supported by worldwide distribution of information and communication technology (ICT) devices and networks.	India (2013)
	Cyberspace	"Cyberspace" – the physical and non-physical domain that is created or composed of part or all of the following components: mechanized and computerized systems, computer and communications networks, programs, computerized information, content conveyed by computer, traffic and supervisory data and those who use such data.	Israel (2011)

From those definitions, we can see that the term *cyber/cyberspace* remains in many ways. However, in the broad sense we can capture that the term *cyber* can be cascaded or related to five areas which are (Table 2):

- Physical / IT Infrastructure, such as physical network, critical information infrastructure, and fibre optic.
- Communication / Network, such as telecommunication, and internet.
- System, which is information system.
- Devices, such as computers, servers, routers, and any IT devices
- Virtual Environment, which is complex environment that related to national space

**Table 2:** The term cyber related

Related to	ID
Physical / IT Infrastructure	[1], [4], [5], [7], [9], [10]
Communication/ Network	[1], [2], [3], [5], [6], [8], [9], [10]
System	[1], [4], [6], [7], [9], [10]
Devices	[1], [3], [4], [8], [9], [10]
Virtual Environment	[2], [4], [5], [6], [8], [9], [10]

#### 4. Discussions

In the previous section, we have seen that the definition of "cyber" has been strayed from its original meaning, from "steermanship" (Plato), to "to govern" Ampère, to "the alternative environment", (which we use this definition widely after Gibson introduce it). Although the definition of cyber scattered in different perspectives, and seems to be agreed at no consensus of use, we can grasp the definition of cyber that related to the four parts, which are: IT Infrastructure, Communication / Network, System, Devices, and Virtual Environment. We also see that cyberspace is a "virtual space" which tangled to the "real life". In this section, we try to abridge the different perspective by consolidating the point of view of the various definitions.

##### 4.1 Semantic used of cyber

Abridging the semantic debate, the four linguistic bases discussed in Section 1 can be used to gain a consensus on in writing the term "cyber" (i.e. clear etymology, enjoy the popular and historical usage by the global community based on its trends, searchable, and well-defined (Ramirez and Choucri 2016)). Taking as an example "cyber security" vs "cybersecurity", these two words are largely used interchangeably and both styles are acceptable. However, the term "cyber security" is more acceptable rather than "cybersecurity" (Ramirez and Choucri 2016). In other hand, some words containing "cyber" can be used in a portmanteaus style (noun word), such as cyberspace, while other new jargons are recommended to use the adjectives style (cyber<space>[domain]), with the reason that we can see clearly the domain being discussed. Etymologically, separating the word cyber also favours in most forms (Ramirez and Choucri 2016). The semantic suggestion is shown in Table 3.

**Table 3:** Semantic used of term cyber

Style	Favourable	Disfavour
Portmanteaus style (noun word)	Cybernetics	
	Cyberspace	Cyber Space
	Cyber Security	Cybersecurity
Adjective style	Cyber<space>[domain] Example: cyber world, cyber culture, cyber-attack, cyber warfare, cyber terrorism	

#### 4.2 Context of cyber

As we can see in previous discussion, understanding the domain of *cyber* can be found as vary, especially in the security domain. Some organisations and Governments use *cyber security*, *information security*, *network security*, *internet security* and *critical information infrastructure security* as an interrelated domain (ISO/IEC 2012, p. 4; see ITU-T X.1205 ITU 2009; The Republic of Croatia 2015), which sometimes cyber security is described to comprise Confidentiality, Integrity, and Availability of all aspect inside cyber space (just as CIA model of McCumber 1991).

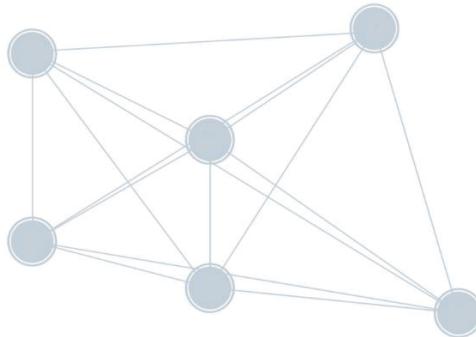
Academics also attempt to elaborate on the definition of cybersecurity in their studies. Harel et al. (2018) propose the concept of cyber related to cyber security and cyber phenomenon. They stated that *Cyber security pertains to all actions that can take place on a computerised platform, with or without the knowledge of the owner of the platform, as well as all of technologies, products and efforts that can be used to defend against such actions*. Von Solms on his recent paper also mentions that *Cybersecurity as that part of information security which specifically focuses on protecting the CIA of digital information assets against any threats, which may arise from such assets being compromised using internet*(von Solms & von Solms, 2018). While Mueller(2017) make the distinction between national cybersecurity and social cybersecurity. National cybersecurity is to strengthen or protect the particular nation-state; Societal cybersecurity are intend to strengthen or protect the users and uses of cyberspace regardless of their nationality. Other escalates the domain of cyber security as not only on securing their own security perimeter, just like in securing other security domains but beyond that, cyber security is intended to secure the virtual hyper-connected environment including the interaction between entities, just as in the ITU and ISO definition(ITU 2012) and ISO/IEC Cybersecurity Guidelines (ISO/IEC 2012).

To abridge the various perspectives, *first*, we try to detangle security domains between cyber security and any other security domains (i.e. information security, critical infrastructure, and network security) using perspective of von Solms and van Niekerk (2013), which *cyber security is securing all assets in the online form including information and non-information asset*.

*Second*, taking into account the domain of cyber security, we can say that cyber security is more global, and provides hyper-connected environment which there are interactions between each entity in the virtual environment. Therefore, consolidated from the various perspectives, we can abridge the term cyber security as “*securing a virtual environment, including its assets* (i.e. *information and non-information based*), *entities* (such as: *end-users, organisations, governments, societies, machines, and software*), and *its interactions* (*enabled by: IT Infrastructure, Communication/Network, System, and Devices*)”.

We therefore can imply that *cyberspace* is orchestrated by three parts – entities (node), interactions (line), and assets (illustrated as in Figure 3).

*The entity* refers to every object connected to cyberspace. In the perspective of ISO/IEC (2012), this entity refers to “*providers*” and “*consumer*”, while others refer this to stakeholders, i.e. *the government, the private, and the societies* (ITU 2012; Klumburg 2012; OAS 2004; WEF 2012a). In this context favour to use term “*entity*” which is represented as a node in the Figure 3. In this context, while we recite to term used in the CCDCOE “the 3 Dimensions” which refers to the Government (the central), National (all actors related to cyber space), and International, however the term “*entity*” is preferred to use to illustrate that the entity is more global and is emphasising the flattening concept of “*end user*” / “*nodes*” in cyberspace. This is also to show that the nature of relationship in cyber space is changed from the hierarchical system to the flattened and hyperconnected relationship (see the WEF concept).



**Legend:** Nodes: entity, such as: end-users, organisations, governments, societies, machines, and software  
**Line:** interactions, which enabled by technology, system, and network  
**Asset:** something need to be protected which is valuable whether inside the entity or information crossing the interaction

**Figure 3:** Illustration of cyberspace

Since those entities live in the hyperconnected environment (Kuusisto and Kuusisto 2015; WEF 2012a, 2012b, 2014, 2015), they naturally have interactions. *This interaction* between the entity in the cyber space can be illustrated as connected lines between one to others on Figure 3, which we can view this from the perspective of the Social Network Theory. These interactions are enabled by IT Infrastructure, Communication/Network, System, and Devices (see Table 2 for the discussion).

The last aspect of cyber is the assets. *The assets* are something valuable which need to be protected whether inside the entity or information crossing the interaction. Since the assets differ from one to other organisation, defining how valuable they are, depends on how the entity view it. Unlike the information asset which if it is stolen, the information still resides in the owner, the transgression of cyber asset (or non-information asset as described by von Solms et all (2013)) means to remove the ownership of the asset virtually.

## 5. Conclusion

This paper began with the aim to revisit the term cyber by walking through its used in contexts. On revisiting the term ‘cyber’, this paper started with the history of the word cyber, explained what is inside the cyber world, and the nowadays context of this term from organisations and governments. Abridging the various context, we can grasp that the context of term *cyber* was shifted from not only described as internet and its virtual environment, but also there are entities, its interactions, and assets, which those interdependent to the physical world.

This work may contribute in several ways. *First*, for the policy-makers in their modern conversations and statements, revisiting the context “*cyber*” may realign the focus. For example, discussing information security and cyber security is, although overlapped, but they are on different domain. *Second*, researchers may benefit from refocusing the subject of their research. Also, “*the Interaction*” discussed in this paper, may illuminate a new study of social structure of cyber world.

## References

- Ampère, A.-M. 1843. *Essai sur la philosophie des sciences, ou, Exposition analytique d'une classification de toutes les connaissances humaines*, Paris.
- Ashby, W. R. 1957. *An Introduction to Cybernetics* (Second Imp.), London: Chapman & Hall Ltd.
- Azmi, R., Tibben, W., and Win, K. T. 2016. “Motives behind Cyber Security Strategy Development: A Literature Review of National Cyber Security Strategy,” in Australasian Conference on Information Systems, Wollongong: University of Wollongong.
- Bayuk, J. L., Healey, J., Rohmeyer, P., Sachs, M. H., Schmidt, J., and Weiss, J. 2012. *Cyber Security Policy Guidebook*, Canada: Wiley (doi: 10.1002/9781118241530).
- BMI. 2011. *Cyber Security Strategy for Germany*, Berlin: Federal Ministry of the Interior (available at [www.bmi.bund.de](http://www.bmi.bund.de)).
- CCDCOE. 2017. “Cyber Definitions,” Resources (available at <https://ccdcoe.org/cyber-definitions.html>; retrieved February 1, 2018).
- Chan, S. 2001. “Complex Adaptive Systems,” ESD.83 Research Seminar in Engineering Systems (Vol. 31) (doi: 10.1002/cplx.20316).
- Cabinet Office of UK. 2011. *The UK Cyber Security Strategy: Protecting and Promoting the UK in a digital world*, London: Cabinet Office of UK (doi: 10.1109/MC.2013.72).

**Riza Azmi and Kautsarina**

- Department of Defense. 2015. The DoD Cyberstrategy, Washington: Department of Defence - United States of America (doi: 10.1017/CBO9781107415324.004).
- Fanelli, R., and Conti, G. 2012. "A methodology for cyber operations targeting and control of collateral damage in the context of lawful armed conflict," 2012 4th International Conference on Cyber Conflict (CYCON), pp. 1–13 (available at <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?reload=true&arnumber=6243983>).
- Finland. 2013. Finland's Cyber security Strategy.
- GCSCC. 2014. Cyber Security Capability Maturity Model (CMM) (Version 1.), Oxford: Global Cyber Security Capacity Centre (GCSCC), University of Oxford (available at [http://www.sbs.ox.ac.uk/cybersecurity-capacity/system/files/CMM\\_Version\\_1\\_2\\_0.pdf](http://www.sbs.ox.ac.uk/cybersecurity-capacity/system/files/CMM_Version_1_2_0.pdf)).
- Gibbs, S. 2017. "Bitcoin: \$64m in cryptocurrency stolen in 'sophisticated' hack, exchange says," The Guardian Online (available at <https://www.theguardian.com/technology/2017/dec/07/bitcoin-64m-cryptocurrency-stolen-hack-attack-marketplace-nicehash-passwords>).
- Gibson, W. 1984. Neuromancer (T. Carr, ed.), New York: Ace Books.
- Government of Belgium. 2012. Cyber Security Strategy: Securing Cyberspace, Belgium: Government of Belgium.
- Government of Japan. 2015. Cybersecurity Strategy (Provisional Translation), Tokyo, japan: The Government of Japan (GoJ) (available at <http://www.nisc.go.jp/active/kihon/pdf/cybersecuritystrategy-en.pdf>).
- GoM. 2013. National Cyber Security Strategy, Podgorica, Montenegro: The Government of Montenegro (GoM) (available at [http://www.enisa.europa.eu/activities/Resilience-and-CIIP/national-cyber-security-strategies-ncsss/NCSS\\_ESen.pdf](http://www.enisa.europa.eu/activities/Resilience-and-CIIP/national-cyber-security-strategies-ncsss/NCSS_ESen.pdf)).
- Harel, Y. Gal, I.B., and Elovici, Y. 2017. Cyber security and the role of intelligent systems in addressing its challenges. ACM Trans. Intell.Syst.Technol. 8, 4, Article 49 (May 2017), doi: 10.1145/3057729.
- IMCCS. 2012. Government of the Republic of Trinidad & Tobago National Cyber Security Strategy, Trinidad & Tobago: Inter-Ministerial Committee for Cyber Security - the Republic of Trinidad & Tobago National.
- ISO/IEC. 2012. "ISO/IEC 27032:2012 Information technology — Security techniques — Guidelines for cybersecurity," Geneva (available at [http://www.iso.org/iso/catalogue\\_detail?csnumber=44375](http://www.iso.org/iso/catalogue_detail?csnumber=44375)).
- ISO/IEC. 2013. "ISO/IEC 27001:2013 - Information Security Management," Geneva.
- ITU. 2009. "Definition of cybersecurity," ITU-D Cybersecurity: Study Group 17 (available at <http://www.itu.int/en/ITU-T/studygroups/com17/Pages/cybersecurity.aspx>; retrieved August 17, 2017).
- ITU. 2012. ITU National Cybersecurity Strategy Guide (F. Wamala, ed.), Geneva: International Telecommunication Union (available at <http://www.itu.int/ITU-D/cyb/cybersecurity/docs//ITUNationalCybersecurityStrategyGuide.pdf>).
- Johnson, C. 2015. "'French' cybernetics," French Studies (69:1), pp. 60–78 (doi: 10.1093/fs/knu229).
- Khatwani, S. 2017. "Top 5 Biggest Bitcoin Hacks Ever," Coin Sutra (available at <https://coinsutra.com/biggest-bitcoin-hacks/>).
- Klimburg, A. (Ed.). 2012. National Cyber Security Framework Manual The NATO Science for Peace and Security Programme, Tallin: NATO Cooperative Cyber Defence Centre of Excellence (doi: 9789949921119).
- Kuusisto, T., and Kuusisto, R. 2015. "Cyber World as a Social System," in Cyber Security: Analytics, Technology and AutomationM. Lehto and P. Neittaanmäki (eds.) (Vol. 78), Springer, pp. 31–44 (doi: 10.1007/978-3-319-18302-2).
- Lehto, M. 2013. "The Cyberspace Threats and Cyber Security Objectives in the Cyber Security Strategies," International Journal of Cyber Warfare and Terrorism (3:3), pp. 1–18 (doi: 10.4018/ijcwt.2013070101).
- Lehto, M. 2015. "Phenomena in the Cyber World," in Cyber Security: Analytics, Technology and AutomationM. Lehto and P. Neittaanmäki (eds.), Cham: Springer International Publishing, pp. 3–29 (doi: 10.1007/978-3-319-18302-2\_1).
- Liddell, H. G., and Scott, R. 1940. A Greek-English Lexicon, Oxford: Clarendon Press (available at <http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.04.0057:entry=kubernh/ths>).
- Lucas, C. 2004. "Cybernetics and Stochastic Systems," The Complexity & Artificial Life Research Concept for Self-Organizing Systems (available at <http://www.calresco.org/lucas/systems.htm>; retrieved February 26, 2018).
- Luijif, E., Besseling, K., and de Graaf, P. 2013. "Nineteen national cyber security strategies," International journal of critical ... (July 2015), pp. 2–31 (doi: 10.1504/IJCIS.2013.051608).
- Maathuis, C., Pieters, W., and Van Den Berg, J. 2017. "Cyber weapons: A profiling framework," in 2016 IEEE International Conference on Cyber Conflict, CyCon U.S. 2016 (doi: 10.1109/CYCONUS.2016.7836621).
- MADISA. 2013. Cyberspace Protection Policy of the Republic of Poland, Warsaw, Poland: Ministry of Administration and Digitisation, Internal Security Agency - Republic of Poland.
- McCumber, J. R. 1991. "Information System Security: A Comprehensive Model," in 14th National Computer Security Conference, Washington: National Institute of Standards and Technology, National Computer Security Center.
- MCIT. 2013. National Cyber Security Policy 2013 (NCSP-2013), India.
- MED. 2011. New Zealand'S Cyber Security Strategy, Wellington: Ministry of Economic Development - New Zealand.
- Merriam-Webster. 2017. "Definition of Cyber," (available at <https://www.merriam-webster.com/dictionary/cyber>; retrieved February 1, 2018).
- Microsoft. 2013. Developing a National Strategy for Cybersecurity: Foundations for Security, Growth, and Innovation (C. F. Goodwin and J. P. Nicholas, eds.), Microsoft.
- MICT. 2013. National Cyber Security Strategy (Ministry of Information and Communications Technology - Kingdom of Qatar, ed.), Qatar: Ministry of Information and Communications Technology - Kingdom of Qatar.
- MICT. 2014. Cybersecurity Strategy, Nairobi, Kenya: Ministry of Information Communications and Technology - Republic of Kenya.
- MOD. 2014. Cyber Security Strategy of Latvia 2014 - 2018, Riga, Latvia: Ministry of Defence - Republic of Latvia.

- MPS. 2010. Canada ' s Cyber Security Strategy, Ottawa: Ministry of Public Safety.
- Mueller, M. 2017. "Is Cybersecurity eating internet governance? Causes and consequences of alternative framings", Digital Policy, Regulation and Governance, Vol. 19 Issue: 6, pp 415-428 (doi : 10.1108/dprg-05-2017-0025).
- NCKB. 2015. National Cyber Security Centre of The Czech Republic for the Period from 2015 to 2020, Czech: National Cyber Security Centre (NCKB) - The Czech Republic.
- NIST. 2014. Framework for Improving Critical Infrastructure Cybersecurity National Institute of S (Version 1.), New York: National Institute of Standards and Technology (available at <http://www.nist.gov/cyberframework/upload/cybersecurity-framework-021214.pdf>).
- OAS. 2004. "A Comprehensive Inter-American Cybersecurity Strategy: a Multidimensional and Multidisciplinary Approach to Creating a Culture of Cybersecurity," in AG/RES. 204 (XXXIV-O/04) on Adoption of Comprehensive Inter-American Strategy to Combat Threats to Cybersecurity: A Multidimensional and Multidisciplinary Approach to Creating a Culture of CybersecurityInter-American Committee Against Terrorism (CICTE) (ed.), Montevideo, Uruguay: Organization of American States (OAS) (available at [https://www.oas.org/juridico/english/cyb\\_pry\\_strategy.pdf](https://www.oas.org/juridico/english/cyb_pry_strategy.pdf)).
- Ottis, R., and Lorents, P. 2010. "Cyberspace: Definition and Implication," in Proceeding of the 5th International Conference Information Warfare and Security, Ohio, USA: The Air Force Institute of Technology, pp. 267–269.
- Oxford Dictionary. 2017. "Definition of Cyber," (available at <https://en.oxforddictionaries.com/definition/cyber>; retrieved February 1, 2018).
- Pangaro, P. 2013. "Cybernetics - A Definition," (available at <http://www.pangaro.com/definition-cybernetics.html>; retrieved February 26, 2018).
- Pangaro, P. 2017. "Cybernetics as Phoenix: Why Ashes, What New Life," in Conversations. Cybernetics: State of the ArtL. C. Werner (ed.) (First Edit., Vol. 1), Berlin: Technische Universität Berlin (doi: 10.1007/978-3-642-01310-2\_2).
- PCM. 2013. The National Plan for Cyberspace Protection and ICT security, Rome, Italy: Presidency of the Council of Ministers - Italian Republic.
- Plato. 2014. "The First Alcibiades (Translated with an Introduction by Benjamin Jowett)," The University of Adelaide Library (available at <https://ebooks.adelaide.edu.au/p/plato/p71al/complete.html>; retrieved February 23, 2018).
- PMO. 2011. Advancing National Cyberspace Capabilities: Resolution No. 3611 of the Government of August 7, 2011 (Vol. 2002), Jerusalem, Israel: Prime Minister's Office (PMO) of Israel.
- Ramirez, R., and Choucri, N. 2016. "Improving Interdisciplinary Communication With Standardized Cyber Security Terminology: A Literature Review," IEEE Access (4), pp. 2216–2243 (doi: 10.1109/ACCESS.2016.2544381).
- Raymond, D., Conti, G., Cross, T., and Fanelli, R. 2013. "A Control Measure Framework to Limit Collateral Damage and Propagation of Cyber Weapons," 5th International Conference on Cyber Conflict (CyCon), pp. 1–16.
- Raymond, D., Cross, T., Conti, G., and Nowakowski, M. 2014. "Key terrain in cyberspace: Seeking the high ground," International Conference on Cyber Conflict, CYCON, pp. 287–300 (doi: 10.1109/CYCON.2014.6916409).
- Republic of Turkey. 2013. National Cyber Security Strategy and 2013-2014 Action Plan.
- von Solms, R., and van Niekerk, J. 2013. "From information security to cyber security," Computers & Security (38), Elsevier Ltd, pp. 97–102 (doi: 10.1016/j.cose.2013.04.004).
- von Solms, B. and von Solms, R. 2018. "Cybersecurity and information security - what goes where?" Information and Computer Security. Vol. 26 No. 1, pp 2-9 (doi: 10.1108/ics-04-2017-0025).
- Solomon, D. 2007. "Back From the Future: Questions for William Gibson," The New York Times Magazine, New York (available at <http://www.nytimes.com/2007/08/19/magazine/19wwln-q4-t.html>).
- The Republic of Croatia. 2015. The National Cyber Security Strategy of the Republic of Croatia (Vol. 2015), Zagreb: Republic of Croatia.
- Wafa, Z. 2014. National Cyber Security Strategy of Afghanistan (NCSA) (2<sup>nd</sup> ed.), Kabul, Afganistan: Ministry of Communications and IT - Islamic Republic of Afghanistan.
- WEF. 2012b. Risk and Responsibility in a Hyperconnected World: Pathways to Global Cyber Resilience, Geneva, Switzerland: World Economic Forum (WEF) (doi: 270912).
- WEF. 2012a. Partnering for Cyber Resilience: Risk and Responsibility in a Hyperconnected World - Principles and Guidelines, Geneva, Switzerland: World Economic Forum (WEF) (available at [http://www3.weforum.org/docs/WEF\\_IT\\_PartneringCyberResilience\\_Guidelines\\_2012.pdf](http://www3.weforum.org/docs/WEF_IT_PartneringCyberResilience_Guidelines_2012.pdf)).
- WEF. 2014. Risk and responsibility in a hyperconnected world: Implications for enterprises World Economic Forum In collaboration with McKinsey & Company (J. Kaplan, A. Weinberg, and D. Chinn, eds.), Geneva, Switzerland.
- WEF. 2015. Partnering for Cyber Resilience: Towards the Quantification of Cyber Threats, Geneva: World Economic Forum (WEF).
- Wiener, N. 1948. Cybernetics: Control and Communication in the Animal and the Machine (second edi.), Cambridge, Massachusetts: The M.I.T Press.

# Secure Application for Health Monitoring

Vijay Bhuse<sup>1</sup> and Harsh Sinha<sup>2</sup>

<sup>1</sup>Grand Valley State University, Allendale, USA

<sup>2</sup>Western Michigan University, Kalamazoo, USA

[bhusevij@gvsu.edu](mailto:bhusevij@gvsu.edu)

[harsh.sinha@wmich.edu](mailto:harsh.sinha@wmich.edu)

**Abstract:** Long term collection of health-related data can be very useful for healthcare providers in diagnosing and treating illnesses. This data can also help athletes improve their performance. People are able to continuously monitor their health and fitness because of innovations in wearable sensors that interface with smartphones. However, there are numerous security and privacy concerns that stem from the transmission and sharing of data of personal nature. Breach of this data violates patient privacy and puts liability on healthcare providers. This paper outlines our research project aimed at building secure and private Android “app” to monitor health using Bluetooth based sensors to track heart rate and blood pressure. We use Public Key Infrastructure (PKI) for a secure and private transfer of data (1) from sensors to the smartphone and (2) from the smartphone to healthcare providers. We consider an entire life cycle of data while designing and developing our solutions to provide an end to end encryption, authentication and data integrity.

**Keywords:** smartphone, android, sensor, healthcare, public key infrastructure (PKI), certificate authority (CA).

---

## 1. Project description

Advancing healthcare is always going to be an important effort. In such effort, we are creating a platform that is meant to simplify the connection between doctors and patients. The application strives to provide a simpler mean of communicating information between doctors and patients, allowing for more comprehensive, intuitive care. With all the revolutionary advances in technology, more and more people are wearing smart devices that are capable of taking measurements of the body's vitals. The app would be designed to communicate with all such devices a person is wearing using a secure Bluetooth connection. The data would be stored locally on the device until the user decides to do otherwise. After taking readings from the smart devices, the user would be able to select between a list of doctors to communicate with. The user would request a secure connection between a doctor, and the doctor would accept the request. Once the secure connection is established, the user can start chatting with the doctor, and share data from the smart device with the doctor, if he or she chooses to do so. The data chosen to be shared would be uploaded to the cloud by the patient, and then pulled down by the doctor. Both the upload and download would happen using a secure communication between the device and the cloud. The benefit of uploading to the cloud to transfer data, rather than send data between the devices, is privacy. If the server in the cloud does get hacked, then the only data that is stolen is random readings from random devices from random people. The cloud does not store sensitive information such as social security numbers, email addresses, home addresses, telephone numbers, names, age and gender. Whereas, if data was transferred directly between devices, and the link was hacked, then the hacker could gain all of the sensitive information listed above. The cloud-based data transfer is the best way to ensure both the patient's and doctor's privacy. Another major benefit is that with the cloud-based data transfers all the user would require is a simple internet connection in order to connect to a doctor or a patient. If data was transferred directly, then the connection would require both the doctor and the patient to be on the same network whether that be through LAN or the same mobile network.

This paper is an extension and implementation of our ideas proposed in our paper *Bhuse et al. (2014)*. We focus on building wearable sensors to monitor vital signs (body temperature, blood pressure, heart rate and pulse rate) and record readings periodically using Smartphone (Android based devices). Next step will be sharing this data with healthcare providers. Our primary goal would be the privacy of the patient data and secure communication. The specific goals are as follows.

- Sensors with an ability to communicate using a wireless technology (Bluetooth and RFID) with Smartphones. The sensing of vital signs and reporting the reading could be continuous or the device might sense, store-and-sync with Smartphone later on.
- Smartphone will synthesize the data and send it securely to the patient, a doctor or whoever the patient wants to share with (may be a health insurance company). We plan to use state of the art encryption, data integrity and authentication techniques. We will also consider access control mechanisms so that a patient

might be able to share data with a physical trainer but not an insurance company. Data that has no significance over long term will have ability to self-destruct if that is what patient desires.

Specific aims of the project are listed as follows: (a) Building wearable sensors, (b) establishing communication between sensors and Smartphone and (c) communicating data securely with healthcare providers with high priority to patients' privacy are the specific aims of this paper.

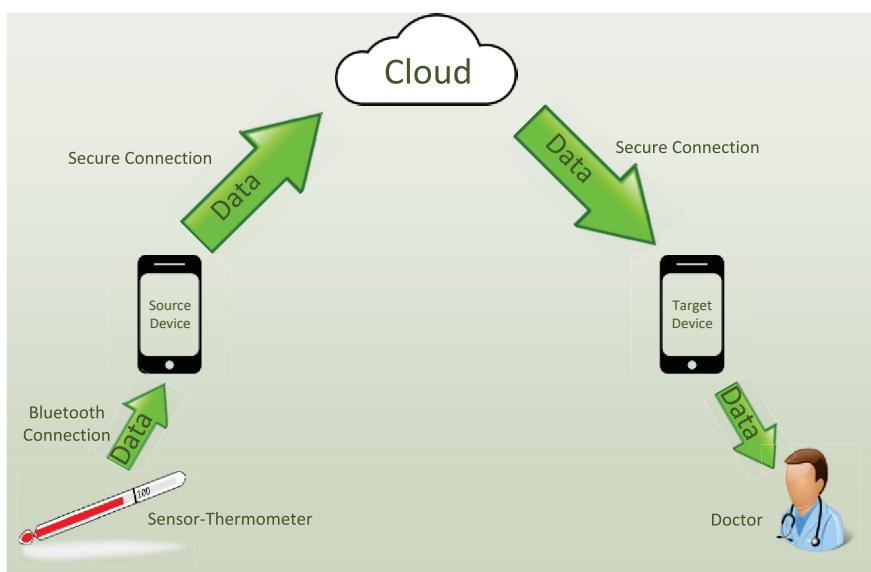
This work would have a significant impact on people's lives because it integrates the Smartphone with sensors to monitor health. It would improve the quality of life by providing people with fitness and healthcare related data. Since we are going to focus on secure communication and storage and privacy of patients, most of the people would have little to no objection using the sensors.

## 2. Solution for secure and private sensing in healthcare

Below we describe the current state of the project. Data breach is the most critical threat in healthcare related cloud computing solution. The major concern for patients is the loss of health-related data which is very private in nature. Another issue is loss of personally identifiable information such as a social security number, address, and date of birth etc. which can be used for identity theft. We propose a multi-factor authentication scheme using Public Key Infrastructure (PKI) framework, smartcards and biometrics to address data breach issue in healthcare related cloud solution.

PKI is a service framework needed to manage a large-scale public key, based technologies. Certificate is a document, which binds together the name of the entity and its public key and has been signed by the Certificate Authority (CA). CA is the trusted third party that signs the public keys of entities in a PKI-based system. CA is trusted by the patient, healthcare providers and the hospital. We use the following notations.  $P_{public}$  is the Public key of the patient.  $P_{private}$  is the Private key of the patient.  $HP_{public}$  is the Public key of the healthcare provider.  $HP_{private}$  is the Private key of the healthcare provider.  $E(Data)_{HP_{public}}$  is the Patient data encrypted with the public key of the healthcare provider.  $H(Data)$  is the Hash value of the data.  $E(H(Data))_{P_{private}}$  is the Digital signature of the patient data.  $D(Data)_{HP_{private}}$  is the Data decrypted with the healthcare provider's private key. Patient's smartphones sends  $E(Data)_{HP_{public}} | E(H(Data))_{P_{private}}$  to the healthcare provider through untrusted network or cloud. The healthcare provider gets data by using the decryption function  $D(E(Data)_{HP_{public}})_{HP_{private}}$ . The hash value of this data should be same as  $D(E(H(Data))_{P_{private}})_{P_{public}}$ . This guarantees privacy of the patient data, integrity of the patient data, authentication of the patient and non-repudiation.

In a healthcare related cloud computing solution, every user may not need the full strong authentication. A user role may be assigned to each user in an organization and the appropriate security provided to each user role. The level of security provided will depend on the role of the user. For example, an administrator role needs a highly secured mechanism, the healthcare provider role needs to be secured, and the patient needs normal security level. This way, the overhead caused by smartcards and biometric hardware can be optimized.



**Figure 1:** Connectivity using android based smartphones

### **3. Android/Java APIs used**

We used the following Android/Java APIs.

#### *KeyPairGenerator*

This is used to generate a random pair of private and public keys that will be used for encryption and authentication. The generation of the pair of private keys is random based on the different initialization method.

#### *Android Keystore System*

This allows the pair of private and public keys generated, or given to the device, to be stored in a more secure location. This makes it so before any key can be used, the Keystore System needs to be provided with authentication: which is provided when generating the key pair itself. This system protects the pairs from unauthorized usage from the android system or processes not related to the app. The first measure of security is that the key material never enters the application process. Therefore, if a hacker compromises the app, they will not be able to extract the key material for usage outside of the android device. The second security measure is that the key material may be bound to secure hardware. This way the key material is never exposed outside of secure hardware on the android device. This also prevents hacker from extracting key material from the android device.

#### *Sensors*

Sensors are both physical components on board and software based. This API allows us access to the raw data that any sensor on board is providing.

#### *Bluetooth*

The android platform supports the Bluetooth Network Stack. This allows an android application to scan for Bluetooth devices, query local Bluetooth adapters for paired Bluetooth devices, establish RFCOMM channels, connect to other devices through service discovery, transfer data to and from connected Bluetooth device, and manage multiple connections. We plan to use both Classic Bluetooth and Bluetooth Low Energy to measure the efficiency of one versus the other. The android system needs to provide permission to the application to use Bluetooth.

#### *Session*

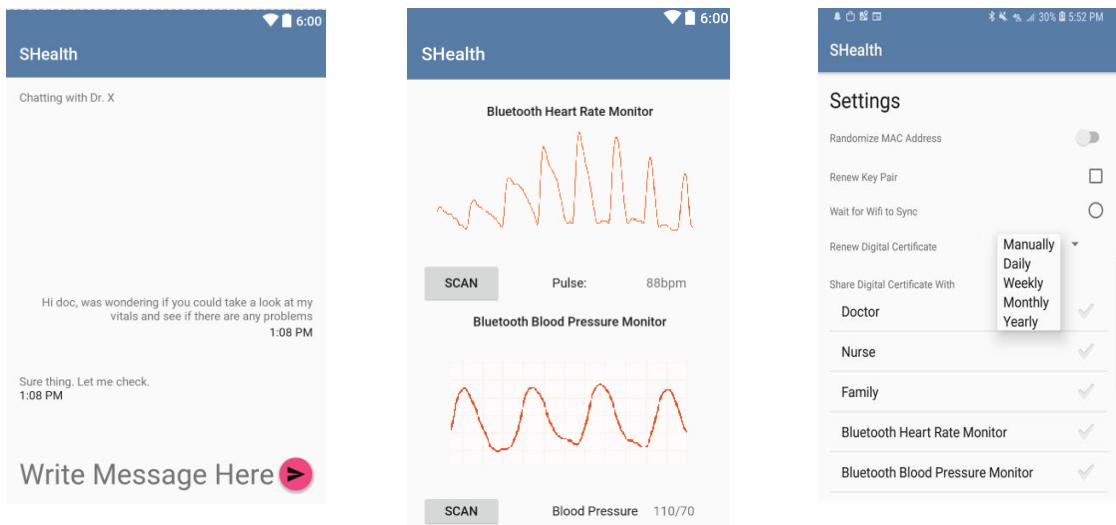
A session is a time interval specified by an application that stores associated metadata. Sessions allow for easy querying of data in a detailed or aggregated fashion.

#### *CloudRail Universal Cloud*

This is a free third-party library meant for securely uploading and downloading data from a host of cloud services. It has easy authentication independent of the cloud service being used. Moreover, there is no need to change the android app when the provider API changes. It uses asymmetric encryption in order to ensure privacy and security.

#### *Activities*

Activities serve as an entry point for a user's interaction with an app, and also are crucial to how a user navigates within an app. This class allows the app to draw its UI on, and handle interactions with the UI. One activity is one screen in an app.



**Figure 2:** Screenshots of the app with sensing and security features

#### 4. Need for wearable sensors and data collection

The Centers for Disease Control (CDC) reports that 75% of health care dollars in the United States goes to treatment of chronic diseases (2014). These persistent conditions, which are the nation's leading causes of death and disability, lead to deaths that could have been prevented, lifelong disability, compromised quality of life, and ever-increasing health care costs. Chronic diseases include but are not limited to diabetes, cardiovascular disease, chronic obstructive pulmonary disease and cancer. Health care providers have the task of managing these illnesses from data collected intermittently from either the patient or specific tests that are performed every few months. These snapshots of the patient's condition can be helpful but are limited.

The challenge for health care providers and patients is the sharing of accurate information between visits. Very important data can be missed during this time period. This data can range from information that would benefit from immediate attention to educational opportunities that would aid in decreasing exacerbations of the chronic illness.

Monitoring patients at home with the use of holter monitors (2014) for cardiac assessment has been a success. The holter monitor requires the patient to physically go to the health care provider's office, be connected to the machine by a professional, and then wear the device with wires that will record around the clock for several days. The monitor is cumbersome which makes it difficult to perform every day activities especially bathing and changing clothing. It also requires a trip back to the health care provider's office for removal and downloading of the information collected. The health care professional then has the task of reviewing and interpreting several days' worth of data.

New technology has made it possible to remotely monitor many different types of conditions and illnesses and capture the information in real-time. According to Rosenthal *et al.* (2006), the benefits of real-time monitoring of symptoms include quicker data transmission, increased patient safety, and enhanced communication.

Quicker data transmission to the health care provider will increase the speed in which symptoms are identified and treated. The monitoring of patients with real-time results allow patients to be monitored at any time in any location. Varshney *et al.* (2007) discussed the actions that can be taken when the real-time information is received. They identified that an alert message can be sent to the nearest ambulance and the patient's provider can contact the patient. These actions could reduce the time between the occurrence of an emergency and the arrival of needed help.

By providing health care providers real-time access to a patient's health status, they can provide patients with appropriate preventive interventions, helping to avoid hospitalization and to improve the patient's quality of care and quality of life. As per Blount *et al.* (2007) the use of real-time monitoring may reduce health-care costs by focusing on preventive measures and monitoring instead of emergency care and hospital admissions. The overall care of the patient with chronic illness will improve as a result of early intervention and increased

educational opportunities. The Smartphone and the wearable sensors provide us with a great opportunity to record and share data securely in real time.

## **5. Related work**

Halperin *et al.* (2008) cover the security issues of implanted medical devices such as a pacemaker or a cardiac defibrillator that a physician could remotely connect to and monitor vitals or activate the machines. The issues with this type of connection arise when considering malicious attacks via unwanted connections that could prove fatal. The future of these machines relies on the need to balance security and accessibility in a way that would prevent attacks while allowing the necessary access.

Olanrewaju *et al.* (2011) broach the subject of accessing medical images and records via electronic delivery. The main factors to consider in telemedicine are availability, confidentiality, and integrity. Proper balances of these factors are important as medical record privacy is of the utmost importance. Image encryption and watermarking are the two prevention measures discussed at length in the article, and both need to be sufficient in order to prevent attacks, theft, or record modification.

Wilkowska *et al.* (2012) used surveys and focus groups to collect data on the importance of security and privacy in medical assistive devices found in the home. The results of these polls found that accepted adoption of these devices by users would require strict data protection and self-enforced data storage in transfer. With the need for remote devices rising alongside the population of elderly or citizens in poor health, it is important for these security measures to be achieved in order for the adoption of new medical technology to be successful.

Bigfoot Biomedical (2019) is a company that makes an automated insulin delivery system. For those with diabetes, it is a constant chore to keep their insulin levels up. Bigfoot have developed a device and app that when synced up can automate the delivery of insulin. This takes a huge hassle out of the lives of the patients and helps simplify their medical experience. This company has proven the effectiveness of using smart devices to aid the simplification, and ease of access to healthcare. Our app would be much like Bigfoot's, however, it would act as a hub for multiple smart wearable technologies instead of only being compatible with one.

Panjwani *et al.* (2017) state that Startup Health Care Originals utilizes the power of artificial intelligence to predict and prevent asthmatic attacks. A common misconception of asthmatic attacks is that by using an inhaler while aggressively coughing or wheezing you can prevent it. The fact is that if someone is at that stage then they are close to an asthmatic attack. Startup Health Care Originals device keeps track of a person's vitals and then using artificial intelligence to interpolate when they are in the earliest stages of an asthmatic attack - before the coughing and wheezing.

Haghi *et al.* (2019) discuss the emerging importance of smart wearable technologies and their place in health care systems. The paper discusses that these emerging technologies can serve to help reduce the cost of monitoring patients and trying to prevent other lethal or harmful situations. They introduce the idea of the Medical Internet of Things (MIoT) in which they would design an architecture for how the hardware and software would interact, and what safety features the whole process would have. It summarizes the top healthcare sensors in the market at the time of publication and lists possible uses for the sensors. The biggest concern given was the usability in the real world and the power consumption on the smart phones.

## **6. Conclusions and future work**

Wearable sensors are used by large number of people regardless of the security and privacy concerns. We demonstrate that it is possible to build an Android based app that is secure and private by using PKI. Our security solutions achieve desired security and privacy expectations for patients to use these sensors and store their health-related data permanently or temporarily in the cloud.

## **References**

- Bhuse (2014), "Security and privacy challenges for healthcare records and wearable sensors in cloud.", Transaction on IoT and Cloud Computing, Vol. 2(3), 2014, pg. 11-17.  
Bigfoot Biomedical (2019), <https://www.bigfootbiomedical.com/>, last accessed January 2019.  
Blount, M., Batra, V., Capella, A., Ebling, M., Jerome, W., Martin, S., et al. (2007). "Remote health-care monitoring using Personal Care Connect". IBM Systems Journal, 46(1), 2007, pg. 95-113.

**Vijay Bhuse and Harsh Sinha**

- CDC (2014) Chronic disease prevention and health promotion, Centers for Disease Control and Prevention. Retrieved February 5, 2014, from <http://www.cdc.gov/chronicdisease/index.htm>
- Haghi, M., Thurow, K., Habil, I., Stoll, R., Habil, M. (2019). "Wearable Devices in Medical Internet of Things: Scientific Research and Commercially Available Devices". National Center for Biotechnology Information. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5334130/>, last accessed January 2019.
- Halperin D., Heydt-Benjamin T., Fu K., Kohno T., and Maisel W. (2008). "Security and Privacy for Implantable Medical Devices", IEEE Pervasive Computing Mobile and Ubiquitous Systems, Vol. 7, No. 1, January 2008
- Holter Monitor (2014). [http://www.hopkinsmedicine.org/healthlibrary/test\\_procedures/cardiovascular/holter\\_monitor\\_92,P07976/](http://www.hopkinsmedicine.org/healthlibrary/test_procedures/cardiovascular/holter_monitor_92,P07976/), last accessed August 2014.
- Olanrewaju, R. F., et al. (2013). "ICT in Telemedicine: Conquering Privacy and Security Issues In Health Care Services." electronic Journal of Computer Science and Information Technology 4.1 (2013).
- Panjwani, L. (2017). "Wearable IoT Device Predicts an Asthma Attack Before it Happens". <https://www.rdmag.com/article/2017/09/wearable-iot-device-predicts-asthma-attack-it-happens>, last accessed January 2019.
- Roenthal, K. (2006). Enjoy "smarter" patient monitoring. Nursing Management, 37(5), 52.
- Snowden E. (2014). <http://www.biography.com/people/edward-snowden-21262897>, last accessed August 2014.
- Varshney, U. (2007). Pervasive healthcare and wireless health monitoring. Mobile Network Applications, 12, 2007, pg 113-127.
- Wilkowska, W., and Ziefle M. (2012) "Privacy and data security in E-health: Requirements from the user's perspective." Health informatics journal 18.3 (2012): 191-201.

# Crossed Swords: A Cyber Red Team Oriented Technical Exercise

Bernhards Blumbergs<sup>1, 2</sup>, Rain Ottis<sup>2</sup> and Risto Vaarandi<sup>2</sup>

<sup>1</sup>CERT.LV, IMCS University of Latvia, Riga, Latvia

<sup>2</sup>Centre for Digital Forensics and Cyber Security, TalTech, Tallinn, Estonia

[bernhards.blumbergs@cert.lv](mailto:bernhards.blumbergs@cert.lv)

[rain.ottis@taltech.ee](mailto:rain.ottis@taltech.ee)

[risto.vaarandi@taltech.ee](mailto:risto.vaarandi@taltech.ee)

**Abstract:** This paper describes the use-case of international technical cyber exercise “Crossed Swords” aimed at training the NATO nation cyber red teams within a responsive cyber defence scenario. This exercise plays a full-spectrum cyber operation, incorporates novel red teaming techniques, tools, tactics and procedures (TTPPs), assesses team design and management, trains the skills for target information system covert infiltration, precision take-down, cyberattack attribution, and considers legal implications. Exercise developers and participants have confirmed the learning benefits, significant improvements in understanding the employed TTPPs, cyber-kinetic interaction, stealthy computer network infiltration and full-spectrum cyber operation execution.

**Keywords:** technical cyber exercise, cyber red teaming, responsive cyber defence, computer network operations

---

## 1. Introduction

The majority of exercises are cyber defence oriented with the defending blue team (BT) being the primary training audience and the attacking cyber red team (CRT) role-playing the adversary to provide the learning experience for the defenders (Leblanc, et al., 2011; Ogee, et al., 2015; Dewar, 2018; Fox, et al., 2018). However, technical exercises, oriented at advancing the readiness level and experience of a CRT, are lacking, limited in scope, not mentioned or described publicly (Lewis, 2015). To enable the development of defensive approaches, both sides – blue and red, must be exercised, especially if they are dependent on each other in real operations. To integrate and explore the concepts of responsive cyber defence (RCD), computer network operations (CNO), CRT techniques, tools, tactics and procedures (TTPPs), detection mechanisms, cyber deceptions, and cyber red team operational infrastructure, a unified environment is required. A technical cyber exercise oriented not only at training single facet of the CRT but implementing a full cyber operation environment would improve the CRT training experience. Additionally, CRT interactions and inter-dependencies with other operational entities, such as conventional kinetic or special operations forces (SOF), should also be explored to see how cyber operations fit within the larger picture. This paper defines and assesses the CRT oriented technical cyber exercise “Crossed Swords” since year 2014.

The main contribution of this paper is a detailed description of the CRT oriented technical cyber exercise “Crossed Swords” that has been conducted since 2014. To the best of our knowledge, no research papers on CRT exercise design have been published before, and this paper fills this gap. Also, “Crossed Swords” exercise has several unique design features, such as multi-disciplinary nature of the exercise and training audience, complex scenario which adds geo-political, strategic and cyber-kinetic dimensions to advanced technical challenges, and near real-time feedback to exercise participants via situational awareness system.

The remainder of this paper is organized as follows – section 2 provides an overview of related work, section 3 focuses on various design aspects of “Crossed Swords” exercise, and section 4 concludes the paper.

## 2. Related work

Leblanc et.al. (Leblanc, et al., 2011) explore and analyse multiple war-gaming exercises and implemented exercise support tool-sets, as described in this paragraph. “CyberStorm” is the US Department of Homeland Security developed exercise with the aim of examining readiness and response mechanisms to a simulated cyber event. Participation is strictly limited to Five Eyes alliance members. “Piranet” is the French developed response plan and a simulation exercise of a major cyber-attack against France’s Critical Information Infrastructure (CII). “Divine Matrix” is the India’s war-gaming exercise to simulate a nuclear attack accompanied by a massive cyber-attack with kinetic effects against India. “Standoff”, organized by PHDays conference, involves a competition without fixed scenario between attackers, defenders, and security monitoring teams. Airbus commercial Cyber-Range platform hosts the playground for “European Cyber Week Challenge Final”. Similarly, NATO Cyber-Range is used for running “Crossed Swords” exercise.

Mauer, Stackpole and Johnson (Mauer, et al., 2012) look at developing small team-based cyber security exercises for use at the university as a practical hands-on part within the courses. The research explains the management and roles of the engaged parties in the exercise creation and execution. The developed game network is comprised of a small set of virtualized machines used by a group of participants to attack, defend and monitor the event. DeLooze, McKeen, Mostow and Graig (DeLooze, et al., 2004) examine the US Strategic Command developed simulation environment to train and exercise CNO and determine if these complex concepts can be more effectively taught in the classroom. The simulation environment consists of “Virtual Network Simulator”, comprised of two or more networked computers designed to represent attack effects in an interactive graphical environment, and the “Internet Attack Simulator”, presenting a set of simple attacks, ranging from reconnaissance to DoS, available for launching against the network simulator’s virtual network. Mostow and Graig confirm the benefit of CNO simulation exercises by measuring the increase of knowledge of the participants.

The issues tackled, in the related work for the cyber defence exercises are either narrow in scope, specific to a nation or small set of nations, restricted only to exercising just the decision-making process or a small subset of a full-scale cyber operation, or limited to just simulation of common cyber-attacks. Additionally, the majority of the exercises are delivered for the defensive capability building, and the CRT is either simulated or role-playing the adversary. However, a dedicated technical exercise for training CRT capabilities is required.

### **3. Cyber red team exercise design**

Cyber exercise “Crossed Swords” (XS) (NATO CCD CoE, 2019), organized jointly by NATO CCD CoE and CERT.LV, is an annual international technical exercise oriented at training CRT with the latest technologies and striving to deliver highly realistic training. Since its inception, the exercise has grown in complexity and size. The exercise spans across three consecutive days representing a 24-hour fast-paced and intensive operation. More importantly, this exercise has served as a platform for implementing, testing, confirming, and conducting academic studies in the areas of CRT TTPPs, learning effectiveness, and near real-time situational awareness (Kont, et al., 2017).

The exercise is designed to implement the following cyber red team training objectives (TO):

- 1. Perform defended system compromise assessment, practice evidence gathering and information analysis for technical attribution, identify the origins of malicious activities and take actions stop them;
- 2. Execute a responsive cyber defence scenario for adversarial information system infiltration;
- 3. Employ stealthy attack approaches, and evaluate applicable TTPPs for fast-paced covert operations;
- 4. Exercise working as a united team in achieving the laid-out mission objectives;
- 5. Develop specialized cyber red teaming soft and technical skills needed for operation management, information flow, and target information system takeover; and
- 6. Explore and evaluate the full-spectrum operation’s cyber-kinetic interdependencies.

The following subsections will provide a detailed description of the exercise.

#### **3.1 Cyber red team structure and chain-of-command**

The exercise developers, execution managers, and the participants are allocated to various teams and sub-teams based on the specifics and activity focus area. The exercise has the following teams based on the area of operations: cyber-kinetic operations team (Red Team – RT), adversary and user simulation team (Blue Team – BT), exercise control and scenario management (White Team – WT), near real-time attack and situational awareness team (Yellow Team – YT), and game network infrastructure development and support team (Green Team – GT). As stated, the structure and chain-of-command for such cyber-kinetic operations has not been publicly discussed or disclosed by any nation; therefore, this exercise strives to experiment and uncover the organizational model providing simple chain-of-command and separation of duties.

The designed chain-of-command model for “Crossed Swords 2019” exercise is depicted in Figure 1, where the grey boxes represent the cyber red team at strategic, operational and tactical levels (with respective grey colour shading for every level), and the white boxes indicate the white team presence and assistance to the CRT. In the “Crossed Swords” exercise the CRT is divided into the sub-teams based on the expertise in technologies to be

targeted (e.g., web applications, network protocols). This exercise favours the speciality based sub-team creation to allow participant engagement throughout the exercise game-play and not only for explicit phases of the cyber operation.



**Figure 1:** “Crossed Swords 2019” chain-of-command

These teams are subdivided into sub-teams according to the specialization and operational management level:

- **1. Red Team:** being the largest team of around fifty experts consists of exercise training audience. The chain-of-command and various sub-teams are the following:
    - *a. Cyber commander.* The top-level officer in charge of commanding the cyber operation at a political level. Cyber commander, as part of the training audience, manages and coordinates the cyber operation to reach the set mission goals and coordinates the high-level activities, based on the desired effects, of the sub-teams;
    - *b. Strategic adviser.* Is a member of the exercise development and management team (WT) with the role to provide the advice to the cyber commander, and, if needed, give minor hints to keep the red team activities on the course of designed scenario. This option allows exercise developers to explore the nuances and alternative paths for the developed scenario, allowing deviations if the result objectives are met;
    - *c. Red team leader group.* This small group of experts, operating at an operational level and under direct command of the cyber commander, are responsible for fulfilling assigned operational effects by working with the sub-team leaders and ensuring that objectives, force protection, and intelligence activities are correctly executed and reached. This group consists of three experts: the overall red team leader, OPSEC officer, and an intelligence officer;
    - *d. Sub-team leaders.* The main red team consists of five expert-focused sub-teams, at a tactical level, which are represented by their leaders. The purpose of every sub-team is to deliver the intended effects in their responsibility area. Over the exercise iterations it has been identified that sub-team size of 6-10 experts offers the best trade-off between required capability and management efficiency. The role and purpose of every sub-team is as follows:
      - *i. Client-side attack sub-team.* This team focuses on executing attacks targeted at exploiting end-user (i.e., human vulnerabilities) to get the initial foothold, such as creating a *spear-phishing* campaign, setting up *watering-hole* or *drive-by* attacks. Once initial breach has succeeded this team ensures persistence in the target computer network;
      - *ii. Network attack and exploit development sub-team.* The goal for this team is to target exposed computer network services, gain control over them by abusing the misconfiguration, poor implementation, abuse, or developing and exploiting software vulnerabilities. Additionally, this team is conducting IP network (IPv4 and IPv6) and service mapping, and executing attacks against specialized systems, such as tactical radio networks, mobile operator base stations, and ICS elements;
      - *iii. Web-application attack sub-team.* For this team all web-based systems and technologies, such as web-applications, services, and back-end relational databases, are the target. This team extracts valuable information from the web-applications, such as user credentials, e-mails, or application source code, as well

- as breaches the security of an exposed web-application to gain access to the internal network services, and establish persistence;
- *iv. Digital battle-field forensics sub-team.* The main effort of this team is to perform data carving and artefact extraction from various sources, such as hardware devices (e.g., smart phones, portable computers and other electronics), computer memory or hard disk images. This team serves as the bridge between the cyber and kinetic operational components as it is tasked to perform analysis of forensic evidence extracted either by the cyber sub-teams or brought in from the field by the kinetic team; and
  - *v. Kinetic forces sub-team.* This team formed from trained military and law-enforcement experts performing various kinetic operations, including high interaction with the CRT capability, such as forced entry, covert access, hardware extraction, target capture or take-down, intelligence collection, surveillance, or kinetic activities on enemy territory. The interaction with the rest of the red team provides one of the key aspects for cyber-kinetic game-play. This team is managed and trained by industry experts (e.g., HTCI – High-Tech Crime Institute) and SOF instructors (e.g., NATO SOF School). The created scenario is designed to have the interdependencies within the CRT and anticipates the cyber-kinetic cooperation.
  - *e. Sub-team liaisons.* For every mentioned sub-team there is an attached WT liaison responsible for observing and, if required, providing minor hints to the sub-team leader to ensure that the team is not wasting too much time on some targets, such as cyber decoys and honeypots, and does not deviate from the intended scenario significantly; and
  - *f. Legal advisers.* Legal advisers are embedded to assist every level of the chain-of-command. The role of these experts is to provide their assessment of the activities from the international and domestic law perspectives.
  - *2. Blue Team.* A small team of up to four experts experienced in conducting cyber red team activities. This team is under direct control and supervision of WT and is used to manage the CRT progression within the adversary's computer networks. The main tasks for this team are user and adversary simulation. As a user simulation role-player, they are directly engaged in client-side conducted activities, such as examining and deciding to open received malicious attachments, visiting the web links, or browsing the in-game web services. Their task is to observe, assess and deny, or permit, the CRT initial foothold, based on the quality and delivery method sophistication level;
  - *3. White Team.* A small group of experts, typically no more than two, responsible for controlling and steering the exercise according to the developed scenario. As mentioned before, deviations from scenario are accepted and sometimes encouraged if the overall focus is not lost, and mission objectives can be reached. The exercise does not have the goal of succeeding in accomplishing the intended scenario and fulfilling entirely the laid mission goals by any means necessary. Depending on the activities pursued by the CRT, their course-of-action and time limitations, the mission might be a failure, as it can happen in real life. Within the past iterations of the exercise the red team only once successfully accomplished all the mission objectives, and in other cases completed them partially.
  - *4. Yellow Team.* This team focuses various areas of threat and anomaly detection, such as monitoring, big data analytics, intrusion detection, and situational awareness. The most crucial task for this team is to provide the near-real time situational awareness picture to the CRT from the perspective of neutral and hostile actors. This feedback allows the CRT to immediately spot mistakes and adjust their operations and tool usage to avoid detection, therefore not only increasing the level of stealth, but also having a better understanding on the used tools and performed actions; and
  - *5. Green Team.* Is responsible for tasks, such as maintaining the cyber range platform, supporting the game network technical requirements, developing the game network hosts and targets, and integrating new technologies.

### 3.2 Technical environment and technical exercise scenario

The “Crossed Swords” (XS) exercise game network is hosted on a cyber range running VMware ESXi hypervisor and it consists of around 200 virtual machines for in-game core networking, simulated Internet, CRT segment, and a set of target networks. Not all intended technical game-play elements can be virtualized, therefore the game network is expanded by connecting physical hosts and systems through the cyber range infrastructure. Before creating the overarching geo-political scenario, the technical scenario is established based on the core development team ideas and intended technical game-play intentions. Due to XS being relatively small, with

respect to the game-network scale and training audience size, experimentation and introduction of new, recently prototyped, and unorthodox technologies can be afforded making the technical game-play more attractive and as close to the real-life as possible. The network also uses the traditional IT systems to provide the networking and common workstation operating systems, such as MS Windows and GNU/Linux, to provide replicate the structure of a regular office and business networks. The following list briefly summarizes some of the technologies introduced in the XS game series, to highlight the technical level:

- Bunker door – control system employing a set of interconnected Siemens developed S7-1200 based PROFINET IO-devices. The CRT must reverse-engineer the communication protocol to inject remotely the commands controlling the bunker door;
- Alarm system – protected premises by the Paradox alarm system, which must be targeted remotely by analysing the used bus-protocol, and capturing and decoding the PIN code;
- CCTV IP camera – CRT has to find and exploit the flaws in the IP-based surveillance camera's web interface to gain full control remotely;
- Distributed power-grid – based on IEC-60870-5-104 industrial Ethernet protocol series and a Martem produced remote terminal unit (RTU) is used to manage and supervise the power-grid. The CRT has to reverse-engineer the protocol and perform remote command injection to control the power supply;
- Unmanned aerial vehicle (UAV) – Threod manufactured UAVs flying over the protected territory must be targeted to gain control over the steering and video stream;
- Unmanned ground vehicle (UGV) – Milrem developed UGVs serve as an adversary-controlled tank force and the cyber red team is tasked to take full control over them by targeting either the used network protocols or the controlling workstation;
- Maritime navigation – a vessel's steering and tracking system based on the AIS (Automatic Identification System) maritime protocol is targeted by the CRT to gain control over the ship and inject fake naval tracks;
- Radio communication network – Harris-based military-grade data network must be infiltrated by the CRT by extracting the encryption keys;
- Mobile network base stations – the cyber red team must infiltrate the LMT (Latvian Mobile Telephone) operator provided base stations connected to the actual mobile network, analyse and parse the intercepted communications to decode the adversary agent's message exchange (SMS) and pinpoint their physical location; and
- Railroad control station – a system based on Siemens created S7-1200 PLC running s7comm+ protocol, controls the in-game railroad network. The CRT is tasked to gain control over the railroad control stations to stop or derail the train.

The various technical challenges implemented across nearly all game-net systems, are designed in a way, that no single CRT sub-team can solve them on its own. Instead, cooperation, information exchange, objective tracking, and operation management is emphasized to provide the collaborative training experience and attempting to push the participants out of their comfort zones. The technical scenario, being time limited and fast-paced, cannot be fully solved, therefore the CRT has to consider ways and approaches on how to prioritize the technical objectives and manage the focus of force to accomplish the overall mission objectives within the exercise time.

The integration of real-life vulnerabilities and systems (Blumbergs & Vaarandi, 2017) (Blumbergs, 2019) deliver the learning perspective to the exercise participants. Examining, developing exploits, and attacking the systems which are widely used for automation and industrial process control are challenging and allow the training audience to comprehend the actual state of security for such industrial components. Furthermore, some participants might have such systems in their organizations, but are not allowed to execute attacks or tests due to them being in a production state. CRT members, with some guidance by the instructors, follow the full weakness identification, vulnerability determination, and exploit development life-cycle. This allows the participants to successfully exploit the industrial control protocols and devices.

### **3.3 Geo-political exercise scenario**

The technical scenario, describing the interdependencies, attack vectors, and alternative paths, only covers the part for the actual work to be conducted by the exercise participants. To deliver the context, reasoning, and

clear objectives, the overarching scenario is required. This scenario provides the elements, such as the state-of-the-world background, geo-political situation, intelligence information on what has happened, why the response is being triggered, what are the objectives and rules of engagement. The main geo-political story revolves around a fictitious group of Cyberian islands, where every island is a sovereign country with its technological advancements, political stance, alliances, and intentions. The three island-countries are Berylia, Crimsonia, and Revalia. Berylia being the smallest with a modest military force, part of NATO alliance, and its main economic income originating from the electronics manufacturing. Crimsonia is the largest island with a strong military, rich in natural resources, not part of any alliance, and is expressing some signs of aggression against its neighbouring island-countries. Revalia is a small, self-sustained, and politically neutral country. Within the scenario, the exercise participants assume the role of Berylian team, which is assembled to address the looming crisis. Every year, with a new exercise edition, the scenario evolves and the tensions between Berylia and Crimsonia have been escalating, ranging from Crimsonia conducting a series of debilitating cyber-attacks against Berylian CII, abuse of a neutral nation infrastructure, placing insiders and double-agents, forming military blockades, up to launching a military invasion of Berylia. The various levels of conflict are designed to explore the technical, cyber-kinetic, and legal game-plays as every scenario opens new opportunities and provides flexibility in conducting the responsive computer network operations. The operational environment for the kinetic force's unit is extremely important, as this restricts, or enables, some types of activities to be exercised.

### 3.4 Legal considerations

The legal aspects are incorporated in the form of legal scenario injects aimed to trigger the discussion and legal implication consideration by the command element. Legal advisers are assigned to the chain-of-command to assess and consult the exercise participants. The legal aspects of the conducted cyber-kinetic operations and applied TTPPs, within the context of the scenario, typically tackle at least the following legal considerations as covered in Tallinn Manual 2.0 (Schmitt, et al., 2017):

- *Applicable law.* Depending on the scenario, the lawyers are tasked to ascertain which regimes of public international law apply to the cyber operations occurring during the exercise;
- *States entitled to take countermeasures.* Only state affiliated institutions and organizations, such as military or intelligence, can conduct responsive activities on the state's behalf as long as the activities they engage in do not constitute an internationally wrongful act;
- *Effect of RCD on third parties.* Since RCD has extraterritorial nature and implicates pursuing the adversary, as well as, performing malicious service take-down within the cyberspace, the legal advisers are required to assess the legality of the RCD effects on the third parties. For the CRT to complete their mission objectives, the RCD activities have to be deemed lawful;
- *Limitations on RCD.* Depending on the legal qualification of the RCD operations, various limitations, such as concerning necessity, proportionality, imminence and immediacy, are attached to this operation. The legal advisers are tasked to identify any applicable limitations, such as requirements for the RCD to be necessary and proportional, and provide these legal implications to the commander or sub-team leaders;
- *Self-defence against an armed attack.* The scenario is designed in such a manner that the severity of the offensive action against the victim state amounts to an armed attack, thus permitting to respond in self-defence with immediate asymmetric responsive cyber operations against a stronger and advanced adversary;
- *Geographical limitations of cyber operations.* The effects of cyber operations have to be limited to the intended target information systems and geographical locations. This, although not always being possible to limit geographically, is taken into consideration by the CRT when executing the cyber operation which may include the activities, such as placement of drive-by exploit-kits on third-party services;
- *Means and methods of cyber warfare.* The exercise scenario plays on the various levels of aggression and conflicts;
- *Precautions.* For the executed cyber operations, the CRT is asked to exercise constant care, perform verification of targets, choice of means or methods, choice of targets, evaluate proportionality, and estimate the effects of cyber-attack whenever it is reasonably possible and applicable;
- *Cyber operations in neutral territory.* The adversary may proxy their cyber-attacks or route the kinetic attack, such as drone flying through neutral state's air space before heading to the intended target. In such cases,

the red team's response might have uncertainty and limitations on taken actions in the neutral state's cyberspace.; and

- *False-flag and no-flag operations.* For the CRT to protect their identity, assets and intended objectives, a false-flag or no-flag operation could be considered to be executed to imply uncertainty and make attribution harder. From the technical perspective, the cyber red team might adapt the known TTPs of a chosen threat actor to deceive the adversary. From the legal point of view, it is not clear if such operations are permitted when, for example, impersonating and adversarial profile of a threat actor with high certainty attributable to a third state.

### 3.5 Training assessment and real-time feedback

One of the key aspects of the “Crossed Swords” exercise is to provide the environment, where the CRT can experiment, practice applicable TTPs and observe their effects in near real-time. Such opportunity provides the necessary feedback to the exercise participants for their tool and procedure stealth and efficiency, as well as, to the exercise management to evaluate the progress of the CRT and the fulfilment of training objectives. To accomplish this, a dedicated framework, called the *Frankenstack* (Kont, et al., 2017), is developed to deliver the required visibility through meaningful visual means and notifications. The *Frankenstack* development is facilitated and coordinated by NATO CCD CoE since 2016. The development team is assembled from technical experts in the field of monitoring, data visualization, threat detection and assessment, and big data analytics. The contributions include NATO CCD CoE partners, such as Arc4dia, Stamus Networks (Suricata IDS), Greycorex, Cymmetria, Tallinn University of Technology, CERT.LV, and CERT-EE. The *Frankencoding* events (<https://github.com/ccdcoe/Frankencoding>) have resulted in an ongoing *Frankenstack* development with its source code released publicly on GitHub under the MIT license (<https://github.com/ccdcoe/frankenstack>).

The solution is easily deployable in the game network and can accept any possible sources of information to be further processed, which can be from at least the following origins:

- *ERSPAN (Encapsulated Remote Switched Port ANalyser)* traffic mirror collecting all the network data recording, parsing, and deep packet inspection;
- *NetFlow* from game network routers for traffic statistical analysis and evaluation;
- *data from the systems*, such as system performance metrics (e.g., CPU load, HDD utilization, network interface card statistics), and logs (e.g., Syslog, and application textual log-files);
- *honeypots and cyber decoys* placed in the network to attract and deceive the cyber red team into revealing its TTPs; and
- *aggregates the information from all sources* in textual format allowing this to be reduced to a log correlation and analysis problem.

During the “Crossed Swords 2017” execution the members of WT performed the assessment of the deployed *Frankenstack* solution for its usefulness and training benefits (Kont, et al., 2017). The identified findings were addressed and incorporated into the following exercise editions. The conducted expert qualitative interviews and online survey results reflected the following:

- the deployed tools themselves do not increase the learning perspective, but is up to how red team members perceive and use the tools;
- the addition of situational awareness solutions to the exercise is welcome and seen as a necessary component;
- the four large screens in the execution room, showing the yellow team provided information, was preferred and checked approximately every 45 minutes by most of the training audience;
- exercise participants also used the opportunity to access the *Frankenstack* dashboards locally on their computers and dig deeper when attempting new attack vectors;
- *Alerta* tool, showing the identified attacks as priority categorized alerts, was found most useful by most of the trainees;
- it was acknowledged, that ease of use should be further improved especially when considering the merger of high intensity technical exercise with monitoring tools not known to all participants;

- majority of the training audience strongly agreed that the provided situational awareness was beneficial to the learning process, was accurate and delivered in acceptable speed;
- the larger part of the training audience agreed that they learned more regarding how their actions can be detected and tried to be stealthier; and
- integration of various tools into the *Frankenstack* has to be evaluated carefully to avoid visual distractions and making the output more self-explanatory.

#### **4. Conclusions and future work**

In this paper, we have presented the “Crossed Swords” exercise which involves intense game-play scenario and near real-time feedback, and explores novel concepts of CRT structure, cyber operation management and execution, and TTTPs applicability and stealth. For the future work, we plan to study the impact of the exercise on participant learning efficiency, on participant knowledge retention and change of perception about red team cyber operations, and on best practices for red team cyber operations.

#### **Acknowledgements**

The authors thank Liis Vihul and Joonsoo Kim for their valuable contribution.

#### **References**

- Blumbergs, B., 2019. Remote Exploit Development for Cyber Red Team Computer Network Operations Targeting Industrial Control Systems. Prague, Scitepress, 5th International Conference on Information Systems Security and Privacy.
- Blumbergs, B. & Vaarandi, R., 2017. Bbuzz: A Bit-aware Fuzzing Framework for Network Protocol Systematic Reverse Engineering and Analysis. Baltimore, IEEE Milcom.
- DeLooze, L., McKean, P., Mostow, J. & Graig, C., 2004. Simulation for training computer network operations. WestPoint, IEEE Fifth SMC Annual Information Assurance Workshop.
- Dewar, R. S., 2018. Cyber Defense Report: Cyber Security and Cyber Defense Exercises, Zürich: Center for Security Studies (CSS), ETH Zürich.
- Fox, D. B., McCollum, C. D., Arnoth, E. I. & Mak, D. J., 2018. Cyber Wargaming: Framework for Enhancing Cyber Wargaming with Realistic Business Context, Massachusetts : The MITRE Corporation.
- Kont, M. et al., 2017. Frankenstack: Toward Real-time Red Team Feedback. Baltimore, IEEE Milcom.
- Leblanc, S., Partington, A., Chapman, I. & Bernier, M., 2011. An Overview of Cyber Attack and Computer Network Operations Simulatio. SanDiego, Proceedings of the 2011 Military Modeling & Simulation Symposium.
- Lewis, J., 2015. The Role of Offensive Cyber Operations in NATO’s Collective Defence. Tallinn, NATO CCD CoE.
- Mauer, B., Stackpole, W. & Johnson, D., 2012. Developing Small Team-based Cyber Security Exercises. LasVegas, International Conference on Security and Management.
- NATO CCD CoE, 2019. Crossed Swords. [Online] Available at: <https://ccdcoc.org/exercises/crossed-swords/>
- Ogee, A., Gavrilis, R. & Trimintzios, P., 2015. The 2015 Report on National and International Cyber Security Exercises: Survey, Analysis and Recommendations, Athens: ENISA.
- Schmitt, M. et al., 2017. Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations. Tallinn: Cambridge University Press.

# The far Right of the UK and Ukraine's Views About European Integration: An Analysis of Online Political Discourse

Radomir Bolgov and Olga Filatova

Saint Petersburg State University, Saint Petersburg, Russia

[rbolgov@yandex.ru](mailto:rbolgov@yandex.ru)

[filatovo@gmail.com](mailto:filatovo@gmail.com)

**Abstract:** This study is an example of interdisciplinary research, which combines the linguistic aspects with the study of public political discourse in social media. This research provides an analysis of far right political discourse in social media (Facebook and Twitter) on European integration. We collect data from the official accounts of British National Party (BNP) and Ukrainian far right parties and movements: Freedom Party (Svoboda), Radical Party and Right Sector (Pravyi Sektor). The authors hypothesize the existence of two divergent discourses which characterize the European integration in social media in different ways. The authors propose an analysis technique that is most appropriate for discourse analysis in political studies. The technique involves two levels of discourse analysis: identifying key conceptual metaphors in both alleged discourses (positively or negatively characterizing the aspects of European integration) and the identification of semantic opposition "us-them", implemented within the framework of metaphors. On the basis of this approach to textual analysis we can identify the metaphorical images of European integration, the concepts used, and, subsequently, the conceptual metaphors they create. We conclude the research hypothesis was confirmed. The views of the BNP on European integration are definitely negative, while the views of the Ukrainian far right parties and movements can be called, to some extent, positive for the most aspects of European integration.

**Keywords:** online political discourse, discourse analysis, Twitter, information war, Brexit

---

## 1. Introduction

Every large-scale political event causes social response, becoming the topic of public discourse. At the same time, social media are becoming a main platform for discussion of social problems that blurs the traditional boundaries of the public space (Bolgov, Filatova & Tarnavsky, 2016). For instance, Twitter has registered more than 330 million users around the world, and therefore it is not surprising that its information capabilities are used for political aims.

The aim of our research is to discover the trends in social media discourse generated by the British and Ukrainian far right in the light of two opposite processes: Brexit and Ukraine's aspiration to join the EU. Despite the different political culture and living standards, the Ukrainian and British far right share the values which are traditionally common to the far right: family, religion, nation etc. Therefore they address the same statements. Initially, we make a hypothesis that two forms of discourse are generated by social media on topics related to the European integration. That is discourses which positively or negatively characterize European integration. Then we present the results of the research, with a focus on the method, as we believe that it can be used in similar studies in the future.

## 2. Discourse and social media

We base this study on Van Dijk's model of discourse, which means social interaction based on linguistic communication. The key aspect of the discourse is not the fact of "live" communication and not the specific linguistic parameters of the produced text. The most important component of texts construction and perception is the judgments of social situations behind them and their cognitive representation (van Dijk, 1985, p. 122).

The notion of the final understanding of discourse as verbally mediated social interaction was developed by Habermas, who proposed considering communication and discourse not just as the interaction of at least two subjects (able to speak and act) entering (through verbal and non-verbal means) in interpersonal relationships (Habermas, 2008), but the interaction that takes place on important public and political issues.

Discourse in social media is a collection of open (accessible for change and expansion), verbally mediated discussions on certain topics, and conducted by peer-to-peer actors. However, a full part of the public discourse is only those discussions devoted to problems that matter to the whole of society that have already fallen into the public sphere. Discourse in social media is a part of the general public discourse, with the difference that any subject can become its actor in social media (Bolgov, Filatova & Tarnavsky, 2016).

The computer semantic analysis of the discourse of social networks and media today is one of the most actively developing areas of computer linguistics and, in particular, computer semantics. A significant contribution to the development of this direction is made by numerous modern studies of methods of using automatic linguistic processing of texts to effectively solve such problems as automatic classification of messages, recognition of named entities, data mining, sentiment analysis, automatic referencing and other tasks, the main difficulty of solving them is due to the multi-valued and multivariate nature of language. As a separate direction, it is worth noting the analysis of online discourse (not necessarily political) within the framework of applied linguistics (Potapova, 2014), for example, the sentiment analysis in speech communication. Baker and McEnery present a methodology, employing automated Corpus Linguistics analysis (Baker & McEnery, 2005).

The analysis of political discourse has long included content analysis, operational coding, cognitive mapping etc. A number of works are devoted to the discourse analysis (Bodrunova et al., 2015; Bolgov, Filatova & Tarnavsky, 2016). It is worth noting in-depth content analysis of political realism and political idealism (Beer & Balleck, 1994). These authors even announced the beginning of a rhetorical turn in international relations. Also there are works on the discourse of Donald Trump in particular (Yakoba, 2017) and populism / far right in general (Stavrakakis, 2017). In addition, some researchers analyze the experience of political leaders in a comparative perspective (Aharony, 2012; Bolgov et al., 2019).

### **3. Conceptual metaphors in political discourse**

Political discourse is a special genre of discourse that includes social and political actors at levels ranging from individual to local community, and entire state (Schmidt, Radaelli, 2004). Following this line of reasoning, discourse is regarded as arising from the interaction of many complex dynamic contexts and operates over several time levels (Cameron, Deignan, 2006).

The social-cognitive trend in the linguistic paradigm considers political discourse as a social space mediated by knowledge (Wodak, 2002). This mediation is often associated with the mechanism of conceptual metaphor (Musolff, 2004).

Unlike traditional rhetorical paradigms, conceptual metaphors are considered as a fundamental aspect of human cognition (Gibbs, 2014). Conceptual metaphors are considered to be present in many non-linguistic areas of everyday knowledge (Semino, 2008). To clarify, conceptual metaphors constitute a categorization mechanism that gives a more concrete, vivid image of an abstract concept (Musolff, 2004). Metaphors provide a cognitive framework for social and political issues (Charteris-Black, 2006). Conceptual metaphors are an important means of studying political discourse (Musolff, 2004), because they interpret and limit the development of political narration, especially in the political discourse of international relations.

### **4. Political discourse in social media**

Twitter is designed to publish 140-character texts on the Internet (Marwick & Boyd, 2010). Twitter has a default function that displays the number of followers on each user's page. This feature allows the evaluation of the audience and popularity of Twitter accounts. While Twitter was created as a platform for the exchange of non-political information (Munson & Resnik, 2011; Page, 2012), today it is widely used as a platform for online communication on a variety of politically (Woolley et al., 2010) and socially significant topics (Yardi & Boyd, 2010). Twitter has become a frequently used channel of political discourse (Tumasjan et al., 2010). Politicians and public persons have their own Twitter accounts, where they publish their messages (Fischer & Reuber, 2011; Marwick & Boyd, 2010). Twitter is used in political discourse to spread political views and opinions, as well as to maintain online presence (Spina & Cancila, 2013).

Twitter, as one of the social networks, can be conceptualized as a public online sphere. Baumer and colleagues (Baumer et al., 2010) indicate that the political microblogging on Twitter has increasingly become an influential and democratizing source of news and information. Government officials who have their own Twitter accounts have the same publishing rights with non-public Twitter users. It is noteworthy that politicians in Twitter adhere to the same ethical principles as all other Twitter users. Twitter triggers an egalitarian type of online discourse that is democratic, user-friendly and multimodal (Gillen & Merchant, 2013; Zappavigna, 2013). The Twitter discourse has unique characteristics, such as linking messages with users, hyperlinks to external Internet sources and hashtagging (Ausserhofer & Maireder, 2013). The style of the discourse in Twitter includes fluidity of meaning, innovation and creativity (Gillen & Merchant, 2013), along with consciously controlled and fixed users'

views of the outside world (Marwick & Boyd, 2010). In addition, the discourse on Twitter is extremely dynamic due to the speed with which the texts are published on Twitter (Yardi & Boyd, 2010). Presumably, such a discursive space includes Twitter-specific forms of political discourse, in particular, foreign policy discourse.

The choice of Twitter by political parties as a social-network platform of discourse is due to its openness to all participants, the conciseness of messages, as well as the ease of sharing and tagging information. In addition, participants often repeat information in all social media to increase the audience reach. In previous studies, the authors also preferred Twitter as a platform, in particular, for researching social movements, protests and election campaigns (Doroshenko et al., 2018; Woolley et al., 2010). For example, Hong and Nadler studied the use of social media by U.S. presidential candidates (Hong & Nadler, 2012). In particular, the authors studied the number of mentions of a candidate in Twitter. The results of the conducted research showed that with the advent of social media, the number of channels for broadcasting information to the audience increases. It turned out that the high level of activity of candidates in social media, as a result, has a minimal impact on the level of public attention in the online environment.

## **5. Methods of research**

Initially, we decided to focus our research on subjects directly related to political activities and exclude the media, civil society organizations, journalists, ordinary users and other "non-political" actors in the public discourse.

Among the investigated technology platforms we examined were well-known and popular social media platforms like Facebook and Twitter. Also, the study does not include blogging platforms like LiveJournal because of the presence of large texts which require another discourse analysis approach.

The first task was the sampling which contained the list of public discourse actors and their pages of textual discourse that formed the empirical basis of the study. After the final sampling we identified and selected empirical content for the discourse analysis. It is a set of texts (statements, messages, etc) created by actors in social media. We collected data from the official accounts of British (British National Party, BNP) and Ukrainian far right parties and movements (Freedom Party, Radical Party and Right Sector). 79 posts (from February 1, 2014 to April 1, 2018) of the Ukrainian far right and 938 posts of British far right (836 on Twitter and 102 on Facebook) were analyzed.

As for method of analysis, so we decided to limit the research to two levels of discourse analysis: identifying key conceptual metaphors of both alleged discourses (positively and negatively characterizing the European integration), and the identification of semantic oppositions "us-them" implemented within the framework of metaphors. Thus, we had formed a technique which, in our opinion, is most appropriate for discourse analysis in political studies that was not directly related to linguistics.

The first component of our discourse analysis is an identification of key conceptual metaphors generated in both discourses. The use of conceptual metaphors in discourse allows to transfer onto an object negative or positive qualities that the subject of naming has, even without naming them (and in some cases avoiding direct naming). All this makes the metaphor a tool of making of recipient' desired attitude to problems, events, personalities.

The second stage of our discourse analysis was to identify the semantic opposition of "us-them", which was embedded in the framework of the identified metaphors. The concept of semantic opposition in discourse study was used by van Dijk (van Dijk, 1985). His work was devoted to the study of images of the immigrants that are emerging in the discourse of the native Dutch. A specific feature of such discourse is a special structural parameter - the opposition of "us-them"("friend or foe"). This opposition reflects the conflict between the ethnic groups, majority and minority, and simultaneously defines different assessments of the situation ("view"). The style of the stories, the presence of specific pronouns, various rhetorical techniques and structure of stories, and other linguistic means express the opposition and their corresponding points of view (van Dijk, 2000, p.183).

The presence of opposition in which we, or "friends", are always right and "good", and they, or "enemies", are a priori wrong and "bad", is a mandatory component of any discourse. This opposition is implemented on different levels: not only at the structural level of discourse (i.e. macro level) (which is more typical for the more

or less big connected texts) but also at its micro level by using different speech means. The specific feature of this opposition is the rising of "friends" and the humiliation of "enemies" at all textual levels.

Thus, the second part of our discourse analysis is the identification and analysis of the semantic opposition "us-them", which is implemented in the previously identified conceptual metaphors. This analysis of oppositions is concentrated on detecting speech features (lexical, stylistic etc.) because of the empirical basis of our research, which is a set of more or less short texts in social networks written by different actors.

## **6. Results of research**

### **6.1 Analysis of British National Party's public discourse about European integration in social media**

First of all, it is worth noting that we did not find even one metaphor, which describes European integration in a positive way. All the metaphors have a harsh negative connotation and describe especially Muslim immigrants.

All in all our research of the social media public discourse highlighted the fact that British Nation Party's discourse is concentrated on defamation of the European integration, the immigration process in Britain, immigrants and European Union (Table 1). During our research, we found that all the metaphors were focused on describing immigration in Britain and Europe in a negative way. However, when it comes to semantic oppositions we observed a weakening of the linguistic features and a clearly delineation of "friends" and "enemies" of the BNP and immigration policy (Table 2, Table3).

As we can see in our first table the majority of the negative conceptual metaphors were used to define immigrants, immigration, Muslims, Islam and multiculturalism. According to BNP opinion the greatest danger to Britain at this moment are Muslim immigrants. It is very important in this context to mention that the immigration in BNP discourse has been represented only as an immigration of Muslims. The public discourse is highly negative and detrimental to European Union values and vision on immigrants. Brexit is considered as the only efficient solution for immigration crises that Europe is facing.

Semantic oppositions are giving us a better understanding on those who are for or against immigration. As we can see in the tables above the EU policy, Emmanuel Macron, Angela Merkel and Theresa May are those who have pro-immigration views in comparison with BNP members, Marine Le Pen, Putin and Trump who represent the opposition to immigration.

It's also important to emphasize the fact that there is a clear difference in BNP public discourse between Europe and the EU. Europe is defined as originally a perfect place to live in since the EU was created, that's why leaving the EU represents the only way to return Europe safety, economic stability and progress.

The British National Party is using an aggressive and negative rhetoric regarding the migration flood in the UK and the Brexit process. BNP describes itself as „the only choice”, „the only party that can save Great Britain” from migration and high criminality. Also, BNP describes itself as „the first brexiteers” and „the savers of the British nation.”

In opposition we have the immigrants, who are mostly described as „terrorists”, „rapists”, „criminals” and the British government. The British Government and the Prime-Minister Theresa May are described as „traitors”, „cowards”, „pathetic politicians”, who are unable to stop the crimes against the „white girls” and to fulfill the Brexit resolutions. BNP is upset that the British Government is unable to finish the Brexit process, which is currently stalled.

Also, BNP is criticizing the policy of the British Government, which has not the right priorities. For instance, BNP is criticizing the fact that Great Britain is expelling 23 Russian diplomats „whitout any evidence regarding the suspect poisoning”, but doesn't spelling immigrants, „who are raping for more than 40 years British white girls.”

### **6.2 Analysis of Ukraine's far right public discourse about European integration in social media**

Ukrainian nationalists have different views on the possible integration of Ukraine into the European Union. For example, such parties as "Praviy sector" and the Radical party of Oleg Lyashko are very negative about European integration (Table 4). They believe that Ukraine has its own path of development. Ukraine's accession to the

European Union may lead to the loss of cultural identity. Another position of the party "Svoboda" is that its representatives believe that Ukraine's accession to the European Union will lead to the economic prosperity of the country and make it independent from the influence of the Russian Federation.

There are also disagreements among users. Some believe that Ukraine should remain a sovereign and independent state from the policy of the European Union, while others believe that accession will lead to changes in the quality of life (Table 5, Table 6). Among the users there are those who believe that Ukraine should maintain cooperation with Russia, because these countries are fraternal.

## **7. Conclusion and future steps**

The views of the BNP on European integration are definitely negative, while the views of the Freedom Party can be called positive for most aspects of European integration. The difference in views on European integration of these Ukrainian and British right-wing radical parties shows that the Ukrainian party cannot be called nationalistic to the full extent. Rather, it is a populist party. This difference is explained by the fact that Ukraine, unlike the UK, does not have experience of membership in European integration institutions. Ukrainian citizens have positive expectations of European integration, in contrast to the BNP, which focuses its messages on the negative sides of integration. Another explanation for this difference is that the Ukrainian Freedom Party sits in the parliament, unlike the BNP, and therefore has to rely on the broad masses of voters.

The British National Party uses aggressive and negative rhetoric on Twitter regarding the flow of migrants in the UK and Brexit. BNP describes itself as "the only choice", "the only side that can save the UK" from migration and crime.

Within the framework of semantic oppositions, immigrants are mainly called "terrorists", "rapists", and "criminals". The British government and Prime Minister Theresa May are described as "traitors", "cowards", "pathetic politicians" who cannot stop crimes against "white girls" and implement Brexit's resolutions.

In addition, BNP criticizes the policy of British government, which does not have the right priorities. For example, BNP criticizes the fact that the UK sends 23 Russian diplomats "without any evidence of suspicious poisoning," but does not extradite immigrants "who have raped more than 40 British white girls."

The BNP discourse in Facebook focuses on criticism of European integration, the immigration process in the UK, immigrants and the European Union. All metaphors were designed to emphasize negative attitudes towards immigration in the UK and Europe. However, when it comes to semantic oppositions, we observe a weakening of language features and a clear nomination of "friends" and "foes" of the BNP and immigration policy.

Most negative conceptual metaphors have been used to define immigrants, immigration, Muslims, Islam, and multiculturalism. Semantic oppositions give us a better understanding of those who are for or against immigration. It is also important to emphasize the fact that there is a clear difference in the public discourse of BNP between Europe and the EU. Europe is defined as the initially ideal place to live, so leaving the EU is the only way to return Europe.

Ukrainian nationalists revealed different views on the possible integration of Ukraine into the European Union. For example, such parties as the Right Sector and Radical Party led by Oleg Lyashko have a negative attitude towards European integration. They believe that Ukraine has a unique path of development. Ukraine's accession to the European Union may lead to a loss of cultural identity. The opposite position is that of the Svoboda party. Its representatives believe that Ukraine's accession to the European Union will lead to economic prosperity of the country and make it independent from Russia.

Our work is an example of interdisciplinary research, which couples the linguistic aspects with the study of public political discourse in social media. Many discourse studies in political science deal often, in our opinion, with very specific linguistic parameters. Those limits primarily consist of counting the quantitative indicators, transforming discourse analysis into content analysis. The study does not claim to be comprehensive and is intended in some way to fill this gap.

Metaphors per se appeal to the emotions of the audience. The texts with metaphors are characterized by extreme exaggeration and wide use of expressive words, which facilitates the influence on public opinion. The same features are characteristic of fake news and the extreme right/populist discourse. Any metaphor in political or news discourse has a focus and purpose (to influence public opinion). Therefore, the analysis of metaphors in general and this methodology in particular can be adapted for fake news (deception) detection and analysis. The term "fake news" is not currently used to clarify a social phenomenon. On the contrary, different subjects use it primarily for a broad description of those elements of the media space that they dislike and that harm their political, social or media goals. Thus, Donald Trump calls "fake" the media publications which criticize him. In the future, the authors will test various techniques for fake news detection.

## Appendix A

**Table 1:** BNP's conceptual metaphors, which describe European integration in a negative way (Extract from Final research table)

Metaphor	Source of signification (concept)	Content of the concept "+" / "-"	Purpose of signification
Immigrants are Grooming Gangs	Grooming Gangs		Immigrants
Immigrants are parasites	Parasites		Immigrants
Immigrants are an uncontrolled mass of people	Uncontrolled mass of people		Immigrants
Immigrants are Islamists	Islamists		Immigrants
Immigrants are an alien race	Alien race		Immigrants
Immigrants are illegal	Illegality		Immigrants
Immigrants are terrorists	Terrorism		Immigrants
Immigrants are expensive black slaves	Black slavery		Immigrants
Immigrants are abusers	Abuse		Immigrants
Immigrants are criminals	Criminality		Immigrants

**Table 2:** BNP's semantic oppositions, which describe European integration in a negative way

Metaphor	Articulation "US"	Articulation "THEM"
Immigration is a great danger to Europe	Absent	Angela Merkel
Immigration policy is a failure	Marine Le Pen	Emmanuel Macron
Immigration is a threat	Europe	EU
Immigration is not safe for Europe	Absent	Theresa May
Immigration is war	Absent	Islamists
Immigration is criminal	Absent	Muslims
Immigration is an invasion	Europe	Muslims
Immigration is an invasion	Europe	Syrian refugees
Immigration is terrorism	Absent	Islamic State
Immigration is war	Absent	Islamists
Immigration is a danger to Britain	Britain	Islamic State
Immigration is White Genocide on the peoples of the West	West	East
Immigrants should leave Europe	Brexit	EU
Immigrants are illegal	BNP	EU
Immigrants are Third World people	West	Est
Immigrants are Third World people	Absent	Pakistan
Immigrants are pedophiles	Absent	Muslims
Immigrants are aliens	Absent	Orientals and Asians

**Table 3:** BNP's positive semantic oppositions

Metaphor	Articulation "US"	Articulation "THEM"
British people are brave	British people	EU
BNP members are dedicated activists	BNP members	Absent

Metaphor	Articulation "US"	Articulation "THEM"
BNP members are proud British people	BNP members	Absent
BNP is a courageous party	BNP	Islamists
BNP is the protector of British Identity	BNP	EU
BNP is the supporter of traditions	BNP	Absent
BNP is putting local people first	BNP	Absent
BNP is the proud Nationalist movement in the UK	BNP	Absent
BNP is protecting children	BNP	Grooming Gangs
BNP is protecting Christian Faith	BNP	Islam
BNP is correct	BNP	Absent
Brexit is a victory	Brexit	EU
Britain have the ability for independent thought	Britain	EU
Marine Le Pen is the only one who can free France from the EU and protect France from Islam	Marine Le Pen	Emmanuel Macron
Trump is a support of anti-immigration policy	Trump	EU
Putin is a strong leader who keeps his country safe	Putin	Theresa May

**Table 4:** Ukrainian far right's conceptual metaphors, which describe European integration in positive or negative way

Metaphor	Source of signification (concept)	Content of the concept "+" / "-"	Purpose of signification
The European Union is anti-Christian	Anti-Christian structure	«-»	The European Union
The European Union is an anti-European space	Anti-European space	«-»	The European Union
The European Union is religious discrimination	Religious discrimination	«-»	The European Union
Europeans are enemies of God	Enemies of God	«-»	Europeans
European integration is a threat to the national economy	The threat to the national economy	«-»	European integration
The European Union will lead to the extinction of the European race	The Death of a European Race	«-»	European Race
European integration is the destruction of the environment	Destruction of the environment	«-»	European integration
The European Union is imperialism	Imperialism	«-»	The European Union
The European Union is globalism	Globalism	«-»	The European Union
Europe is a friend of the USA	Friend of the USA	-/+	Europe, USA
European integration is a threat to Ukrainian traditions	The threat to Ukrainian traditions	«-»	European integration
European integration is dangerous by the influx of emigrants	Dangerous by the influx of emigrants	«-»	Angela Merkel, The European Union.
European integration is a threat to independence	The threat to independence	«-»	European integration
European integration is the loss of identity	Loss of identity	«-»	European integration

Metaphor	Source of signification (concept)	Content of the concept "+" / "-"	Purpose of signification
European integration is a threat to family values	The threat to family values	«-»	European integration
Europe is a high culture	High culture	«+»	Europe
Europe is a great history	Great history	«+»	Europe
The European Union is a high quality of life	High quality of life	«+»	The European Union
The European Union is an important economic partner	Important economic partner	«+»	The European Union
The European Union is a defense against Russian aggression	Defense against Russian aggression	«+»	The European Union
The European Union is a defense against Russian aggression	Defense against Russian aggression	«+»	Russia

**Table 5:** Ukrainian far right's semantic oppositions, which describe European integration in a positive way

Metaphor	Articulation "US"	Articulation "THEM"
The European Union is an important economic partner	«Svoboda»	National Front (Ukraine)
The European Union is a high quality of life	Ukrainian citizens Batkivshina «Svoboda»	National Front (Ukraine)
The European Union is a defense against Russian aggression	«Svoboda» Ukrainian citizens «Batkivshina»	Vladimir Putin Russian politicians Marine Le Pen

**Table 6:** Ukrainian far right's semantic oppositions describing European integration in a negative way

Metaphor	Articulation "US"	Articulation "THEM"
European integration is the loss of identity	«Praviy sector»	Absent
The European Union is imperialism	«Praviy sector» Dmitriy Yarosh Vasiliy labaychuk	Peter Poroshenko «Batkavshina»
European integration is a threat to Ukrainian traditions	«Praviy sector»	«Batkavshina» Peter Poroshenko «Svoboda»
European integration is dangerous by the influx of emigrants	Ukrainian citizens	Peter Poroshenko «Batkavshina» «Svoboda»
The European Union is globalism and imperialism	«Praviy sector»	«Svoboda» «Batkavshina»
Europeans are enemies of God	«Praviy sector»	Absent
The European Union is religious discrimination	«Praviy sector»	Absent
The European Union is an anti-European space	Oleg Lyashko «Praviy sector»	Absent

## Acknowledgements

This paper was prepared with the financial support of Russian Science Foundation, RSF, Project 18-18-00360.

## References

- Aharony, N. (2012) Twitter use by three political leaders: An exploratory analysis. *Online Information Review*, Vol. 36, No. 4, pp. 587-603. DOI: 10.1108/14684521211254086

- Ausserhofer, J., Maireder, A. (2013) National Politics on Twitter. *Information, Communication and Society*, Vol. 16, No. 3, pp. 291–314. DOI: 10.1080/1369118X.2012.756050
- Baker, P., and McEnery, T. (2005). A corpus-based approach to discourses of refugees and asylum seekers in UN and newspaper texts. *Journal of Language and Politics*, Vol.4, Issue 2, pp. 197-226. DOI: 10.1075/jlp.4.2.04bak
- Baumer, E., Sinclair, J., Irvine, B. (2010) 'America Is Like Metamucil': Fostering Critical and Creative Thinking about Metaphor in Political Blogs. *CHI 2010: Expressing and Understanding Opinions in Social Media*, pp. 1437–1446. DOI: 10.1145/1753326.1753541
- Beer, F., Balleck, B. (1994) Realist/Idealist Texts: Psychometry and Semantics. *Peace Psychology Review*, Vol. 1, No. 1, pp. 38-44.
- Bodrunova, S.S., Litvinenko, A.A., Gavra, D.P., Yakunin, A.V. (2015) Twitter-Based Discourse on Migrants in Russia: The Case of 2013 Bashings in Biryulyovo. *International Review of Management and Marketing*, No. 5, pp. 97-104.
- Bolgov, R., Chernov, I., Katsy, D., Ivannikov, I. (2019) Battle in Twitter: Comparative Analysis of Online Political Discourse (cases of Macron, Trump, Putin, and Medvedev). *International Conference on Electronic Governance and Open Society: Challenges in Eurasia, EGOSE 2018. Communications in Computer and Information Science (CCIS)*, Vol.947. Pp. 374-383. DOI: 10.1007/978-3-030-13283-5\_28
- Bolgov, R., Filatova, O., Tarnavsky, A. (2016) Analysis of public discourse about Donbas conflict in Russian social media. *Proceedings of the 11th International Conference on Cyber Warfare and Security, ICCWS 2016*, pp. 37-46.
- Cameron L., Degnan A. (2006). The Emergence of Metaphor in Discourse. *Applied Linguistics*, Vol. 27, No. 4, pp. 671–690.
- Charteris-Blac, J. (2006). Britain as a Container: Immigration Metaphors in the 2005 Election Campaign. *Discourse and Society*, Vol. 17, No. 5, pp. 563–581.
- Dijk, van, T.A. (1985) Cognitive Situation Models in Discourse Production: The Expression of Ethnic Situations in Prejudiced Discourse, Language and Social Situations. *Springer Series in Social Psychology*.
- Dijk, van, T.A. (2000) *Yazyk, poznanie, kommunikaciya (Language, cognition, communication)* [in Russian]. Blagoveshchensk.
- Doroshenko, L., Schneider, T., Kofanov, D. et al. (2018) Ukrainian nationalist parties and connective action: an analysis of electoral campaigning and social media sentiments. *Information, Communications & Society*, January, pp. 1-20. DOI: 10.1080/1369118X.2018.1426777
- Fischer, E., Reuber, R. A. (2011) Social Interaction via New Social Media: (How) Can Interactions on Twitter Affect Effectual Thinking and Behavior? *Journal of Business Venturing*, No. 26. DOI: 10.1016/j.jbusvent.2010.09.002
- Gibbs, R. W. (2014). Conceptual Metaphor in Thought and Social Action. *The Power of Metaphor: Examining Its Influence on Social Life*. Eds. M. J. Landau, M. D. Robinson, B. P. Meier. American Psychological Association, Washington, pp. 17–40.
- Gillen, J., Merchant, G. (2013) Contact Calls: Twitter as a Dialogic Social and Linguistic Practice. *Language Science*, No. 35, pp. 47–58. DOI: 10.1016/j.langsci.2012.04.015
- Habermas, J. (2008) Relationship to the world and rational aspects of action in four sociological concepts of action [in Russian]. *Sociologicheskoe obozrenie (Sociological Review)*, Vol. 7, No. 1.
- Hong, S., Nadler, D. (2012) Which candidates do the public discuss online in an election campaign?: The use of social media by 2012 presidential candidates and its impact on candidate salience. *Government Information Quarterly*, Vol. 29, No. 4, pp. 455-461.
- Marwick, A., Boyd, D. (2010) I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience. *New Media and Society*, Vol. 13, No. 1, pp. 114–133. DOI: 10.1177/1461444810365313
- Munson, S., Resnik, P. (2011) The Prevalence of Political Discourse in Non-Political Blogs. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*.
- Musolff, A. (2004). *Metaphor and Political Discourse: Analogical Reasoning in Debates across Europe*. Palgrave Macmillan, London.
- Page, R. (2012) The Linguistics of Self-Branding and Micro-Celebrity in Twitter: The Role of Hashtags. *Discourse and Communication*, Vol. 6, No. 2, pp. 181–210. DOI: 10.1177/1750481312437441
- Potapova, R.K. (2014) Social network discourse as an object of interdisciplinary research [in Russian]. *Proceedings of the 2nd international conference "Discourse as social activity: priorities and prospects"*, pp. 20-22.
- Spina, S., Cancila, J. (2013) Gender Issues in the Interactions of Italian Politicians on Twitter: Identity, Representation and Flows of Conversation. *International Journal of Cross-Cultural Studies and Environmental Communication*, Vol. 2, No. 2, pp. 147–157.
- Schmidt V., Radaelli C. (2004). Policy Change and Discourse in Europe: Conceptual and Methodological Issues. *West European Politics*, Vol. 27, No. 2, pp. 183–210.
- Semino E. (2008). *Metaphor in Discourse*. Cambridge University Press, Cambridge.
- Stavrakakis, Y. (2017) Discourse theory in populism research. *Journal of Language and Politics*, Vol. 16, Issue 4, pp. 523-534. DOI: 10.1075/jlp.17025.sta
- Tumasjan, A., Sprenger, T., Sandner, P., Welpe, I. (2010) Election Forecast With Twitter: How 140 Characters Reflect the Political Landscape. *Social Science Computer Review*, pp. 1–17. DOI: 10.1177/0894439310386557
- Wodak R. (2002). What CDA is About. A Summary of its History, Important Concepts and Developments. *Methods of Critical Discourse Analysis: Introducing Qualitative Methods*. Eds. R. Wodak, M. Meyer. Sage Publications, Thousand Oaks/New Delhi, 1–14.

***Radomir Bolgov and Olga Filatova***

- Woolley, J., Limperos, A., Oliver, M. (2010) The 2008 presidential election, 2.0: A content analysis of user-generated political facebook groups. *Mass Communication and Society*, Vol. 13, No. 5, pp. 631-652. DOI: 10.1080/15205436.2010.516864
- Xifra, J., Grau, F. (2010) Nanoblogging PR: The Discourse on Public Relations in Twitter. *Public Relations Review*, No. 36, pp. 171–174. DOI: 10.1016/j.pubrev.2010.02.005
- Yakoba, I.A. (2017) Deconstruction of Donald Trump's discourse (cases of his 2016 elections speeches) [in Russian]. *Diskurs Pi*, Vol. 26, No. 1, pp. 164-169.
- Yardi, S., Boyd, D. (2010) Dynamic Debates: An Analysis of Group Polarization Over Time on Twitter. *Bulletin of Science. Technology and Society*, Vol. 30, No. 5, pp. 316–327.
- Zappavigna, M. (2013) Enacting Identity in Microblogging through Ambient Affiliation. *Discourse and Communication*, pp. 1–20. DOI: 10.1177/1750481313510816

# A Comparison of Chat Applications in Terms of Security and Privacy

Johnny Botha<sup>1</sup>, Carien Van 't Wout<sup>1</sup> and Louise Leenen<sup>2</sup>

<sup>1</sup>Council for Scientific and Industrial Research (CSIR), Pretoria, South Africa

<sup>2</sup>University of the Western Cape, South Africa

[jbotha1@csir.co.za](mailto:jbotha1@csir.co.za)

[cvtwout@csir.co.za](mailto:cvtwout@csir.co.za)

[lleenan@uwc.ac.za](mailto:lleenan@uwc.ac.za)

**Abstract:** Mobile messaging or chat Applications (Apps) have gained increasing popularity over the past decade. Large amounts of data are being transmitted over the internet when people make use of these Apps. Metadata and personal information are being collected and stored every day while consumers are seeking protection against surveillance as well as against attacks from hackers. There are countless Apps available but some are leading the way in popularity, platform availability and features. WhatsApp, one of the leading Apps, revealed in 2016 that it had more than one billion users. In March 2016, WikiLeaks released information that the CIA was able to bypass all security systems of both WhatsApp and Signal, another popular App, to read user messages. WikiLeaks also revealed that the CIA makes use of malware and hacking tools that allow them to remotely hack into smartphones. In 2017, a Guardian report indicated that Facebook, WhatsApp's parent company, could read encrypted messages due to a certain vulnerability found in the App. In terms of security, it is important to distinguish pure secure messaging Apps from the ones who are less secure and trustworthy. This paper compares the best and the supposedly most secure messaging Apps based on the built-in security and privacy features of the Apps, as well as the location and subsequent accessibility of stored data. Recommendations and best practice advisements for users are made on which Apps seem to be the most secure and private.

**Keywords:** chat app feature comparison, chat app data storage security, cybersecurity, message encryption, privacy

---

## 1. Introduction

Every day millions of people are exchanging messages via messaging or rather chat Applications (Apps). However, users do not know what happens to the messages once they have been sent. Initially, encryption was considered to be used only by paranoid users or people with a heightened need for secrecy. After the revelations of Edward Snowden, consumers have become more aware of online privacy and the dangers of digital scooping of data and identity theft. Surveillance activities are increasing globally and concerns amongst people all over the world has been raised considerably whilst data retention laws are being implemented (Ali, 2017). We live in a digital age where surveillance and data logging occur on almost all our communication. Companies want to collect as much as possible personal information about consumers. Some governments are hacking mobile devices to gain unauthorised access for surveillance and other unknown reasons (Curran, 2018). Recently, the Russian government requested the chat Application Telegram, on several occasions, to give them the encryption keys of citizens registered on the chat Application. Although Telegram did not comply, they are now at risk of being banned in Russia (Caffo, 2018).

Although messaging Apps have been around for a number of years, the development of secure mobile Apps are increasing, focusing on securing the privacy of users and meeting their demands (Corpuz, 2017; Das, 2017). Recent studies show that users are becoming concerned about protecting privacy on their smartphones and opposed apps that collected their contacts (Balebako et al., 2013). One survey of 2, 245 US adults showed that 57% of all smartphone app users have either deleted an app or refused to install an app for security and privacy reasons (Boyles et al. 2012).

This paper covers an overview of the most secure Apps of 2017 and 2018 and also highlight some known security flaws within messaging Apps in Section 2. The main focus is a comparison of these Apps in terms of security and privacy. Section 3 presents a comparison of some of the main security features on a number of the most used messaging Apps. Section 4 provides a comparison of where messaging data and media of Apps are stored and can subsequently be accessed or recovered from. In addition, recommendations are given, in section 5, on the most secure Apps to use as well as best practices for back-ups. The paper concludes in section 7.

## 2. An overview of the best and most secure messaging apps

This section gives an overview of the Apps that are regarded to be the best and the most secure; Facebook, WhatsApp, Telegram, Signal, WeChat, Line, Skype and Viber (Caffo, 2018). Two of the main reasons why chat

Applications have become so popular at a rapid pace, are firstly, the rapid growth in access to cell-phones and to the Internet. Secondly the death of SMS messages because chat Applications allow for “richer” methods of remote communication.

Facebook’s chat Application is called Messenger. This App is used by over two billion users registered on Facebook. The App can be accessed via Facebook and allows for normal chat messages, voice and video calls. End-to-end encryption is not enabled by default and has to be enabled with each and every chat by selecting the Secret Conversation option when messaging a contact. WhatsApp (WhatsApp, 2019) has a simple installation and setup by synchronising contacts on your phone automatically. It allows for text and multimedia messages with end-to-end encryption by default. It also periodically asks for a password to access the App. WhatsApp is owned by Facebook and there are rumours that Facebook intends to populate members’ Facebook profiles with their WhatsApp data. This idea has been blocked by the European Union, but it seems it is only a matter of time before this feature might be built in, posing more security and privacy risks to users (Caffo, 2018). WhatsApp is the most popular messaging App (Sutikno et al., 2016).

Telegram (Telegram, 2019) was launched after the Snowden revelations for user that are aware of the need for secure digital communication. It offers a client-server encryption for chat messages and secret chats where privacy cannot be violated. These chats self-destruct after a certain time on the devices at both ends (for both individual or group chats) as well as on the server.

Signal is developed by a company called Open Whisper Systems. Edward Snowden stated that this company can be trusted: *“Use anything by Open Whisper Systems”*. The App uses military-level end-to-end encryption. Signal is strengthened by an open-source platform, which is closely monitored and reviewed to improve security. It is the preferred App for hacktivists and leading security experts (Signal, 2019).

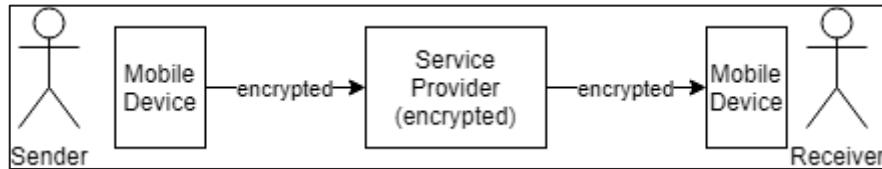
WeChat has more than 700 million users and dominates the Chinese web. The App offers text messages, voice and video calls, group chats and a rich multimedia experience. It also offers features such as “Friend Radar”, “People Nearby” and “Shake” to find new people online. It was one of the first Apps that was available on Android Wear and Apple Watch. It does not offer end-to-end encryption, but it does provide client-to-server and server-to-client protection. The App has an accreditation from the Privacy company TRUSTe, who provides solutions to manage privacy compliance for the General Data Protection Regulation (GDPR) and other global privacy regulations (TRUSTe, 2019). WeChat also complies with **ISO 27001-2013**, a very strict international standard, therefore it would be very difficult for hackers to breach the Application (Caffo, 2018).

Line is a Japanese App with more than 600 million users globally. Line offers additional features such as group chats and calls of up to 200 participants and allows calling of mobile and landline numbers via purchased credits. It also follows certain channels, news feeds and events. Skype was recently revamped in an attempt to make it more attractive to users. It is still known for its great audio and video capabilities and is used more by corporate users. Digital communication is secured by Transport Layer Security (TLS) and Advanced Encryption Standard (AES) encryption; however, there is no encryption when calling landline or mobile numbers (Websecurity.symantic.com, 2019). Viber has similar features to WhatsApp by using a mobile number to login and syncing contacts on the phone book of the device (Caffo, 2018; Viber, 2019).

Table 1 provides a summarised comparison of these Apps in terms of security and privacy features. According to the Amnesty International report of 2016, Snapchat ranked among the least secure Apps because it fails with respect to privacy by not implementing end-to-end encryption (Williams, 2016). In 2019, Snapchat announced that end-to-end encryption has been added to protect user’s messages (Titcomb, 2019).

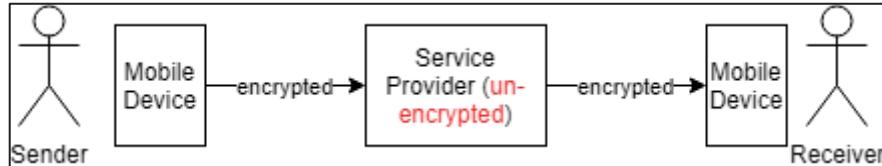
### **3. Comparing security and privacy features**

Since consumers demand better security and privacy in messaging Apps, software development companies have been attempting to address these issues. One of the features was to launch end-to end-encryption (see Figure 1). End-to-end encryption refers to when messages are encrypted during transmission and no copy is stored unencrypted on the servers of the service providers. Nobody apart from the people communicating can view these messages party; no third party, not even the government or the developers of these Apps. Communication is transmitted using a secret code rather than plain text (Rijnetu, 2018).



**Figure 1:** End-to-end encryption

Another type of encryption that is used is encryption in transit (see Figure 2). This means the message is encrypted between the user and the service provider, but stored as clear text on the server. This poses a risk as stored messages can be read by the service provider or other third parties that gain access to the server.



**Figure 2:** Encryption in transit

Table 1 presents a comparison on messaging Apps with regards to security and privacy features.

**Table 1:** Security and privacy features of messaging apps

Messaging App		End-to-end encryption	Encryption in	Private key not accessible by	Deleted from	Self-Destruct Messages	Open-Source	Password lock	Verification SMS/Email	Screenshot detection	Two-step Verification	Remote logout	Remotely Wipe Messages	Account self-destruct	Free
Confide		✓							✓						
CoverMe		✓				✓							✓		
Dust		✓			✓	✓				✓					✓
Hangouts			✓												✓
iMessage		✓		✓	✓	✓									✓
Line		✓							✓						✓
Messenger		✓(optional)	✓			✓									✓
Signal		✓				✓	✓	✓							✓
Skype		✓(optional)	✓												✓
Slack			✓												✓
Snapchat		✓													✓
Telegram		✓(optional)	✓	✓		✓	✓	✓			✓	✓	✓		✓
Threema		✓													
Viber		✓		✓	✓	✓		✓							✓
WeChat															✓
WhatsApp		✓		✓				✓	✓			✓			✓
Wickr Me		✓								✓					

Apps are sorted alphabetically

One can argue that an adequate level of **end-to-end encryption** should be the golden standard and default security feature to be included in messaging Apps. Most Applications listed in Table 1 provide end-to-end encryption. Signal, WhatsApp and Facebook Messenger make use of the Signal end-to-end encryption protocol.

Although Telegram, Skype and Messenger offers end-to-end encryption, it is not enabled by default. Telegram offers this feature as a “secret chat” option. If this is not enabled, encryption in transit is used. Skype recently added end-to-end encryption, but it is not on by default. One has to start a “private conversation” to enable end-to-end encryption (Deahl, 2018). Secret conversations on Messenger are currently only available in the Mobile App; it will thus not appear on Facebook chat or messenger.com and is furthermore only visible on the device where you create the conversation and the device the recipient uses to open the conversation (Woollaston, 2016). Therefore, Messenger, Skype and Telegram has checks for both end-to-end and in-transit encryption as seen in Table 1. (End-to-end encryption by default encrypts in transit as well).

Hangouts and Slack does not provide end-to-end encryption, but instead make use of **encryption in transit**. This immediately makes them less secure and trustworthy. According to (Corrigan, 2018), Google Hangouts is an App that should be avoided. The App allegedly has numerous security and privacy concerns. It uses encryption in transit, messages are stored on the server in clear text. Google can access anyone’s private messages at any time and relay the information to government agencies and other third parties (Corrigan, 2018). According to the 2016/17 Amnesty International Report (Amnesty.org, 2017, Griggs, 2018), WeChat has major privacy issues. They ranked last with a score of zero out of 100 when it comes to privacy. Facebook Messenger and WhatsApp scored 73 and Apple’s iMessage 67 out of 100 (Griggs, 2018). WeChat provides no end-to-end encryption and did not publish transparency reports on China’s government request for information. Based on this, WeChat was subjected to both censorship and surveillance. Due to the lack of privacy and security in WeChat, it is safer to delete the App from your device.

Since most of the Apps make use of end-to-end encryption, the next concern that arises is to know if the **private key is accessible by the service provider**. Apple, Telegram and WhatsApp claim that they cannot obtain the private key. No information was found on the remaining Apps in Table 1. Although Apple claims they cannot access the private key and are unable to read the user’s messages, a study done by Hackers indicated otherwise. They proved that technically Apple can read your iMessage messages whenever they want to (Blue, 2018). One concern is that Confide does not notify users if the primary when a new key is generated, whereas Apps does as iMessage and WhatsApp do alert users on this (TechCrunch, 2017).

iMessage, Viber and Dust are the only chat Apps that claim they **delete your messages from the server**. iMessage delete messages automatically after seven days from the server (Corrigan, 2018). Dust has a feature of never permanently storing your message on the server (Rijnetu, 2018). The message is stored in the random-access memory (RAM) of the server. Once the receiver has received the message, the message gets removed from the RAM. Dust also allows to erase messages from the receiver’s device sent by the particular sender. Telegram, iMessage, Viber, Messenger, CoverMe, Dust and Signal have a feature that allows for the messages to **self-destruct or disappear** after a certain amount of time for both the sender and recipients’ devices (Corrigan, 2018). Facebook is rolling out a self-destruct timer for messages that allows users to set a timer that will have messages disappear automatically (Woollaston, 2016).

Both Signal and Telegram have an **open-source** policy. Anyone can check the source code, protocol and API (Corrigan, 2018). With Threema, only the encryption part of the source is open, the rest is not open source (Decentralize.today, 2016). Signal, Telegram, Viber and WhatsApp has a **pass-code lock** on the chat App that one needs to enter before the App can be used (Corrigan, 2018). On registration of a new user, both WhatsApp and Line send a **verification code via SMS** that is required before the installation can be completed (Corrigan, 2018). Dust and Wickr Me introduced a new feature, **screenshot detection**, that notifies the user when a screenshot has been made of a chat sent by that particular user (Corrigan, 2018). Telegram offers a **two-step verification** feature where the App requires to use both a SMS code and password to log into the App. The App also allows for setting up a recovery email address for in case a user forgets the password (Corrigan, 2018). **Remote logout** is a feature offered by Telegram only. Most Apps allow to be logged into the App from multiple devices. With this feature one can logout from all devices from the current device in use (Corrigan, 2018). Another feature is the account **self-destruct**. Only Telegram offers this functionality. If the account has been inactive for a certain period, where six months is the default, the account will automatically self-destruct and all messages and media linked to the account will be erased (Corrigan, 2018).

Most of the Apps compared in Table 1 are free of charge. Apps such as Slack and Threema may be more suitable for private business chats, whereas the other Apps are aimed at personal use. It seems as if a new market has opened for the development of more secure Apps and to charge users for the secure chatting service. Such Apps,

which are not free of charge are Threema, Wickr Me, CoverMe and Confide. Wickr Me has a free version with limited functionality but the professional version is not free of charge. CoverMe has additional features to the ones compared to in Table 1, such as a private vault to lock your messages, passwords, documents and multimedia; it allows users to obtain a second private number to hide the callers personal number; military-graded encrypted phone calls, password protected call pickups; and it allows to disguise and hide the App with a news reader App for example (CoverMe, 2019).

The results in this section indicate that Signal, Telegram, WhatsApp and Viber are the most secure free Apps. All the paid for Apps in Table 1 include all of the basic security and privacy features as well as additional features, indicating that they might be more secure than the secure free Apps. CoverMe has taken the security and privacy to the next level with a number of additional features to hide and disguise the user's information. The least secure Apps are WeChat, Google Hangouts and Slack primarily due to not using end-to-end encryption. The next section compares the Apps in terms of accessibility of stored data.

#### **4. Comparing accessibility of stored data**

The next question in the comparison of messaging Apps in terms of security and privacy is where the data is stored and how easily it can be recovered – this would include chat history, messages, photos and videos sent via these Apps. This section compares messaging Apps based on this factor. The Apps (free versions) that were found to be the most secure from section 3, Table 1, are being compared in Table 2. LINE and WeChat were also added to this comparison, due to their popularity in the east.

**Table 2:** Location of stored chat app data

Apps	Device Back-up: Chat History	Device Back-up: Images & Videos	Cloud Back-up	Back-up to PC	Transfer Chat history between mobile devices	Copies sent to email
<b>LINE</b>	✓	✓	✓(optional)	✓(optional)	✓	✓
<b>Signal</b>	✓	✓			✓	✓
<b>Telegram</b>	✓	✓	✓(optional)	✓(optional)	✓	✓
<b>Viber</b>	✓	✓	✓(optional)	✓(optional)	✓	✓
<b>WeChat</b>	✓	✓	✓(optional)	✓(optional, default on web-version)	✓	✓
<b>WhatsApp</b>	✓	✓	✓(optional)	✓(optional, default on web-version)	✓	✓

Apps are sorted alphabetically

**LINE** allows users to choose to back-up the chat history in various places: LINE Keep or Memo (on device), share to email, Google Drive, OneDrive and to PC (Bruce, 2017). The app has optional features to back up on the cloud and on a PC. **Signal** only stores the metadata on the device that is required for the App to work. Signal does allow for a backup on the device as optional, but no PC, server or cloud backup is provided (Support.signal.org, 2019). **Telegram** stores all images on the image folder on the device internal storage or SD card. Deleted chats, images and videos are also stored in the cache folder of the SD card (Coline, 2016). Telegram stores all chats and media on their cloud service. Secret chats are not stored on the server, but secret media do get stored (encrypted).

**Viber** allows users to save a copy of their chats by sending it to their email via the user settings. Chats will be put into an archived .zip folder as .csv files with the names of contacts that you chatted to. A chat backup copy can be created and read as text, but it cannot be restored in the Application itself; copies of sent files (photos, videos etc) are not saved to such text files. Viber also keeps data in a separate folder located in the internal

system memory of the user's device. Such backup data can be accessed only with Root rights or by using a kind of Root explorer software (Cherniga, 2017).

**WeChat** data can be backed up on a personal computer (PC) using WeChat Client software downloaded from the internet. WeChat, as well as other Apps (e.g. WhatsApp, Line, Kik, Viber, etc), data can also be backed-up to PC using USB connection and dr.phone software without internet connection. A back-up may also be made by data transfer to another smartphone (Dr.phone, 2019). The Cloud backup feature is no longer available from WeChat 6.2.5. Chat history is kept permanently within the App on the device for as long as the App is not removed, and the phone has sufficient storage space. In the interest of privacy, WeChat does not store the chat history on its server unless a user explicitly chooses the backup feature. Chat history cannot be recovered once deleted (WeChat Help Center, 2019).

**WhatsApp** data is stored on the device of the sender and receiver. Backups are made on the user's device by default on a daily basis within the internal storage and backups may also be stored on the user's Google Drive. Back-up media and messages are not protected by WhatsApp end-to-end encryption while in Google Drive (WhatsApp FAQ, 2019). Messenger data is stored in the user's Facebook account. Message history of all messages created, whether on the Facebook website, Messenger for PC or for Android, is always transferred through the Facebook account and saved there. Facebook provides users with the opportunity to save a copy of all user information, including uploaded images and videos, contact information, friends and most importantly, the complete message history (Hetman Software, 2019). Messenger data is also stored to Android devices in the same manner as similar Apps (e.g. WhatsApp) in the internal storage or SD card Android's data folder (Anydata Recovery, 2019).

All Apps in Table 2 backs up the data on the device. Signal is the only App that does not allow backup to a PC or Cloud, all other Apps has this as an optional feature. All Apps allow to transfer messages from one mobile device to another. All of the Apps allow to send chats to email, unencrypted. This poses a risk for if a user's email gets compromised, all of the chat backups will be available to the attacker. Based on the finding in this section, Signal would be the most secure option due to the App not allowing backups on the cloud or on a PC. All the Apps allow this feature as optional, therefore, if this feature is not used, they would be on par with Signal in this comparison.

## 5. Recommendations

The data generated on and transmitted via messaging Apps may be vulnerable in terms of privacy and confidentiality. This paper compared messaging Apps based on the built-in security features of the Apps, as well as the location and subsequent accessibility of stored data. Recommendations are made for users to maintain privacy and confidentiality based on these two aspects.

Based on the findings in sections 3 and 4, the best and most secure free chat Apps are Signal, Telegram, WhatsApp, Viber and Line. WeChat is ruled out due to privacy concerns highlighted in section 3.

Users often have a need to recover their chat history or data and can hence make use of the back-up functions of the various Apps to enable later restoring or recovery of data. Such back-ups are usually stored on the device itself and/or somewhere in the cloud. Users have further options on the different Apps to make back-ups to email or on their PC. If devices contain classified information (personal or otherwise) the required security controls should be applied to ensure that the data is "safe" in case the device or PC gets compromised.

The choice of which messaging App is best depends on criteria that have varying importance to different users and also influences the required level of security. Such criteria include:

- Being able to connect with relevant others.
- Various countries use different Apps.
- Availability (if there is internet, there is messaging App). No available cellular phone minutes and exorbitant fees can also be relevant factors.
- Cellular phone service provider does not own messaging App data – therefore messaging App data is already more secure than SMSs.

- It would not make sense to the average person to switch to some very secure paid App if the people they need to engage with are not using the same App. It may be relevant for a business organisation to use such an App for communication between colleagues in order to protect information assets.

Users are therefore recommended to apply the correct settings in the use of their preferred messaging App, depending on whether they have a need to recover chat history and other media generated on these Apps. The safest option is not to backup chats and media to a PC or cloud service. If a device containing backups is lost, all backups are lost with it.

The forensic community uses tools enabled to access well-known Session Initiation Protocol (SIP) and Voice over Internet Protocol (VoIP) which applies to messaging Apps. Forensic recovery of important data is thus possible for some messaging Apps. This also implies that if users need to be more secure, they should make use of a less common App. This has its own challenges on the other hand, because it is not exactly clear where the servers are that store the data from these less known Apps.

## **6. Conclusion**

People choose messaging Apps based on different criteria and would hence have different requirements in terms of levels of security (confidentiality and privacy of their chat data). Users are recommended to apply the correct settings in the use of their preferred messaging App, depending on whether they have a need to recover chat history and other media generated on these Apps.

A comparison of the security features of various Apps suggests that Signal, Telegram, WhatsApp and Viber are the most secure free Apps. All of the paid for Apps in Table 1 seem to be equally as secure or even more secure than the most secure free Apps. CoverMe has taken security and privacy to the next level with a number of additional features to hide and disguise the user's information. The least secure Apps are WeChat, Google Hangouts and Slack, primarily due to not using end-to-end encryption. It has been highlighted that WeChat has major privacy issues and the safest option is to remove the App from your phone. Google Hangouts was also flagged for numerous security and privacy concerns.

A comparison in terms of the accessibility to stored App data suggest that Signal is the most secure. On the remainder of the Apps in Table 2, if the correct user settings are applied for back-up of data, they would be on par with Signal with regards to the storage and backup of data. The least secure Apps in Table 2 are WhatsApp and WeChat, primarily due to the fact that if the web-version is used, chats and media gets backed up on the PC by default.

## **References**

- Ali, Z. (2017). Best Secure Messaging Apps for Android and iOS - PrivacyEnd. Available at: <https://www.privacyend.com/best-encrypted-messaging-Apps/> [Accessed 15 Jan. 2018].<sup>9</sup>
- Amnesty.org. (2017). Amnesty International Report 2016/17. Available at: <https://www.amnesty.org/download/Documents/POL1048002017ENGLISH.PDF> [Accessed 10 Apr. 2019].
- Anydata Recovery (2019). How to Recover Deleted Facebook Messenger Messages on Android Device. Available at <https://www.any-data-recovery.com/android-data/recover-deleted-facebook-messenger-message-from-android-devices.html> [Accessed 11 Apr 2019].
- Balebako, R., Jun, J., Lu, W., Cranor, L.F., and Nguyen, C. (2013). "Little Brothers Watching you": Raising Awareness of Data Leaks on Smartphones. Proceedings of the Ninth Symposium on Usable Privacy and Security.
- Blue, V. (2018). Hackers: Here's how Apple's iMessage surveillance flaw works (video). Available at: <https://www.zdnet.com/article/hackers-heres-how-Apples-imessage-surveillance-flaw-works-video/> [Accessed 15 Mar 2019].
- Boyles, J.L., Smith, A., and Madden, M. (2012). Privacy and Data Management on Mobile Devices. Pew Internet and American Life Projects.
- Bruce, I. (2017). Ways to Back Up and Restore LINE Chat on Android. Available at: <https://www.recovery-android.com/backup-restore-line-android.html> [Accessed 15 Mar 2019].
- Caffo, A. (2018). The best (and most secure) chat Apps. Available at <https://blog.avira.com/best-chat-Apps-smartphone> [Accessed 7 Mar 2019].
- Cherniga, M. (2017). How to Recover Message History, Contacts and Viber Files on Android or Windows. Available at: <https://hetmanrecovery-com.cdn.ampproject.org> [Accessed 15 Mar 2019].
- Coline, N. (2016). How Can I Recover Deleted Telegram Chats? Available at: <https://www.quora.com/How-can-I-recover-deleted-Telegram-chats>. [Accessed 15 Mar 2019].

- Corrigan, C. (2018). The very best private messaging Apps. Available at <https://www.avg.com/en/signal/secure-message-Apps>. [Accessed 9 April 2019].
- Corpuz, J. (2017). Best Encrypted Messaging Apps. Available at: <https://www.tomsguide.com/us/pictures-story/761-best-encrypted-messaging-Apps.html>. [Accessed 15 Jan. 2019].
- CoverMe (2019). Corporate Websites. Available at: <http://www.coverme.ws/en/index.html>. [Accessed 10 Apr. 2019].
- Curran, D. (2018). Are your phone camera and microphone spying on you? The Guardian. Available at: <https://www.theguardian.com/commentisfree/2018/apr/06/phone-camera-microphone-spying> [Accessed 14 Apr 2019].
- Das, A. (2017). 8 Best Secure and Encrypted Messaging Apps for Android & iOS. Fossbytes. Available at: <https://fossbytes.com/best-secure-encrypted-messaging-Apps>. [Accessed 15 Jan 2019].
- Deahl, D. (2018). Skype now offers end-to-end encryption conversations. Available at: <https://www.theverge.com/2018/8/20/17725226/skype-private-conversation-end-to-end-encrypted-opt-in>. [Accessed 10 Apr. 2019].
- Decentralize.today (2016). Threema – Secure Messengers..or not so secure? Part 3 Available at: <https://decentralize.today/threema-secure-messengers-or-not-so-secure-part-3-6df427896caa>. [Accessed 10 Apr. 2019].
- Dr.fone (2019). How to Backup WeChat: 5 Ways You May Not Know. Available at: <https://drfone.wondershare.com/wechat/wechat-backup.html>. [Accessed 15 Mar 2019].
- Grigg, A (2018). WeChat's privacy issues mean you should delete China's No. 1 messaging App. Available at: <https://www.afr.com/news/world/asia/wechats-privacy-issues-mean-you-should-delete-chinas-no1-messaging-App-20180221-h0wgct>. [Accessed 20 Mar 2019].
- Hetman Software. (2019). How to Save or Restore Facebook Messenger Access and Data on Android or PC. Available at: [https://hetmanrecovery.com/recovery\\_news/how-to-save-or-restore-facebook-messenger-access-and-data-on-android-or-pc.htm](https://hetmanrecovery.com/recovery_news/how-to-save-or-restore-facebook-messenger-access-and-data-on-android-or-pc.htm) [Accessed 10 Apr 2019]
- Kim, L. (2018). The Top 7 Messenger Apps in the World. Available at: <https://www.inc.com/larry-kim/the-top-7-messenger-Apps-in-world.html>. [Accessed 15 Mar 2019].
- Messenger (2019). Corporate Website. Available at: <https://www.messenger.com>. [Accessed 7 Mar 2019].
- Pryvate (2019). Corporate Website. Available at: <https://www.pryvatenow.com>. [Accessed 7 Mar 2019].
- Rijnetu, I. (2018). The Best Encrypted Messaging Apps You Should Use Today. <https://heimdalsecurity.com/blog/the-best-encrypted-messaging-Apps>. [Accessed 25 Mar 2019].
- Signal (2019). Corporate Website. Available at: <https://signal.org>. [Accessed 7 Mar 2019].
- Support.signal.org (2019). Signal Support Website. Backup and Restore Messages. Available at: <https://support.signal.org/hc/en-us/articles/360007059752-Backup-and-Restore-Messages>. [Accessed 12 Apr 2019].
- Sutikno, T., Handayani, L., Stiawan, D., Riyadi, M.A. and Subroto, I.M.I., 2016. WhatsApp, viber and telegram: Which is the best for instant messaging? International Journal of Electrical & Computer Engineering (2088-8708), 6(3).
- TechCrunch.com (2017). Researchers critique security in messaging App Confide. Available at: <https://techcrunch.com/2017/03/08/researchers-critique-security-in-messaging-app-confide/> [Accessed 12 Apr 2019].
- Telegram (2019). Corporate Website. Available at: <https://telegram.org> [Accessed 7 Mar 2019].
- Telegram.org/privacy (2019). Telegram Privacy Policy. Available at: <https://telegram.org/privacy> [Accessed 12 Apr 2019].
- Titcomb, J (2019). Snapchat adds end-to-end encryption to protect users' messages Available at: <https://www.telegraph.co.uk/technology/2019/01/09/snapchat-adds-end-to-end-encryption-protect-users-messages> [Accessed 9 April 2019].
- TRUSTe (2019). Corporate Website. Available at <https://www.trustarc.com/> [Accessed 7 Mar 2019].
- Viber (2019). Corporate Website. Available at: <https://support.viber.com>. [Accessed 7 Mar 2019].
- WhatsApp (2018). WhatsApp Encryption Overview – Technical White paper. Available at: <https://www.whatsapp.com/security/WhatsApp-Security-Whitepaper.pdf>. [Accessed 15 Jan 2019].
- WhatsApp (2019). Corporate Website. Available at: <https://www.whatsapp.com>. [Accessed 7 Mar 2019].
- WhatsApp FAQ (2019). Restoring Your Chat History. Available at: <https://faq.whatsapp.com/en/android/20887921> [Accessed 15 Mar 2019].
- Websecurity.symantic.com (2019). The Ultimate Guid: What is SSL, TLS and HTTPS. Available at: <https://www.websecurity.symantec.com/security-topics/what-is-ssl-tls-https>. [Accessed 7 Mar 2019].
- WeChat (2019). Corporate Website. Available at: <https://www.wechat.com>. [Accessed 7 Mar 2019].
- WeChat Help Center (2019). Chat History. Available at: [https://help.wechat.com/cgi-bin/newreadtemplate?t=help\\_center/topic\\_list&plat=2&lang=en&Channel=helpcenter&detail=1001146](https://help.wechat.com/cgi-bin/newreadtemplate?t=help_center/topic_list&plat=2&lang=en&Channel=helpcenter&detail=1001146). [Accessed 15 Mar 2019].
- Williams, R. (2016). Snapchat among least secure Apps for data protection, report finds. Available at: <https://inews.co.uk/news/technology/snapchat-among-least-secure-Apps-data-protection-report-finds>. [Accessed 7 Mar 2019].
- Woollaston, V. (2016). How to find and use Facebook's Secret messages. Available at: <https://www.wired.co.uk/article/messenger-secret-messages-end-to-end-encryption> [Accessed 15 Mar 2019].

# Detection of Premeditated Security Vulnerabilities in Mobile Applications

Agnė Brilingaitė, Linas Bukauskas and Eduardas Kutka

Institute of Computer Science, Vilnius University, Lithuania

[agne.brilingaite@mif.vu.lt](mailto:agne.brilingaite@mif.vu.lt)

[linas.bukauskas@mif.vu.lt](mailto:linas.bukauskas@mif.vu.lt)

[eduardas.kutka@mif.vu.lt](mailto:eduardas.kutka@mif.vu.lt)

**Abstract:** Ubiquitous usage of mobile applications in the personal and business areas raises challenges when coping with data and device privacy and protection. Many organisations support the application of personal mobile devices for work-related matters. Mobile applications enable control of Internet of Things (IoT) remote devices and infrastructures. Insecure applications within the device pose a risk of business data breach. In many cases, a device becomes a part of the infrastructure, and it can be an exploitation point. As individuals, users maintain a vast amount of personal and sensitive data on smart gadgets. They support daily activities like calendaring, banking, health, location, social life, and communication. Granted permissions to access security layers such as contact list, files, photos, camera, and microphone enable application vendors to potentially abuse the trust. In the case of premeditated security vulnerabilities, gathered data become a threat to society or a dedicated group of people. We propose a methodology and framework for detection of security vulnerabilities in mobile applications at various abstraction levels. The proposed methods enable tracking down premeditated vulnerabilities or unexpected design flaws that infringe the trust boundaries. Our methodology covers security analysis of mobile applications at communication, data handling, and source code levels. The examination process enables the identification of mobile communication behavioural patterns. When considering software execution semantics, some patterns show possible breaches of trust and identify possible security vulnerabilities. The proposed framework assumes sandboxing as a compartmentalisation method to encapsulate and isolate communication activities. The isolation of communication traffic enables detection of end-points, communication patterns, data volumes. Reverse engineering is included within the methodology to analyse software flow, cross-referencing, and possibly insecure or superfluous data handling within the code. Methodology supports incremental application analysis and observation over time to detect modifications after software updates. To indicate the level of risk, we use traffic light methodology. The use case of the proposed methodology application is presented and discussed.

**Keywords:** mobile application security, security vulnerabilities, violation of trust, risk assessment, communication analysis, data breach

## 1. Introduction

Cybersecurity in mobile devices should become the biggest toothache for individuals and organisations who are concerned about their privacy and information security. In 2016, Statista forecasted that the number of mobile phone users in the world would pass the five billion mark by 2019, and almost three billion of them would be smartphones (Statista, 2016). In 2017, 178.1 billion mobile applications were downloaded, and it was forecasted that there would be 228.2 billion of downloads by 2022 (Statista, 2018). Most users of these devices do not care about cybersecurity. NowSecure Mobile Security 2016 Report (NowSecure, 2016) states that more than 75% of Android OS device users do not update OS for two or more years. Typically, daily each device connects to 160 unique IPs, and 35% of data traffic is unencrypted. A lot of popular applications have security flaws, and 43% of users do not use any access control on their devices.

Bring your own device (BYOD) policy erases limits by mixing personal and work environments. Non-responsible attitude towards security in mobile application can be an exploitation point or a gateway to confidential data in any device connected to the network. These devices might be essential components of the critical business infrastructure or smart home gadgets containing vital private information. The number of IoT devices is increasing. In many cases, they are controlled using vendor-specific applications that provide unlimited access to possibly sensitive user information. There exist specialised search engines with APIs that provide an easy access to unprotected IoT devices. Such a diverse variety of poorly controlled devices and their insecurity enable easy injection of malicious code into applications. Some applications should be identified as undesirable due to the features of malicious software.

Increasing the flexibility of applications in terms of a set of functionalities makes users grant access to more resources than mobile applications would actually need. Often, applications miss the possibility to grant permissions gradually when the particular service is required and to revoke permissions when the service is not

needed anymore. Users lose control of their device security during the default installation or after auto-updates. At the beginning of 2019, it was announced that one social network paid money to teenage users to collect and analyse their private data and to track their daily activities.

A lot of research is done to solve particular problems in malware detection, e.g. malware identification based on a required permission list or based on signatures of applications. However, the overall high-level methodology that defines the detection of security vulnerabilities without low-level details of algorithmic solutions is missing. The contributions of this paper are: 1) developed high-level framework to detect security vulnerabilities in mobile applications at physical communication level and logical programming code level, 2) sandbox construction design to enable simulation of the real environment and conditions, 3) developed methodology to detect and evaluate security vulnerabilities in mobile applications using the designed framework, 4) presentation of the methodology execution.

The paper is structured as follows. Section 2 covers related work. Section 3 presents the architecture of the proposed framework for the detection of security vulnerabilities in mobile applications with an emphasis on the required sandbox. Our methodology of vulnerability detection is covered in Section 4. Discussion of the methodology application is presented in Section 5. The paper ends with conclusions and future work.

## **2. Related work**

Users of mobile smart devices are sensitive to a number of attacks, e.g. phishing, spyware, surveillance attacks. Mobility, personalisation, and connectivity (He, Chan and Guizani, 2015) distinguish mobile security from traditional computer security. European Union Agency For Network And Information Security (ENISA, 2017) emphasised that various features of the mobile application development environment make an impact on privacy and security. Variety of data, multiple sensors, different types of identifiers, online social networks make a list of risks for users of mobile devices. Thus, privacy must be implemented by design. Data must be protected from excessive collection and sharing, disclosure, unlawful processing, loss, breach, destruction.

As security challenges become apparent due to an improper definition of access to resources, usage of external functionalities, decentralised software distribution, API-dependent design, and webification, a taxonomy for attacker capabilities (Acar et al., 2016) on Android OS ranging from dangerous permissions, piggybacking apps to dynamic code loading and network attacks.

APK Auditor (Talha, Alper and Aydin, 2015) classifies Android applications into malicious and non-threatening based on permissions requested by the application. The tool calculates malware scores for applications and permissions. Application's malware score defines the application's maliciousness mark.

He, Chan, and Guizani (2015) discussed threats of malware in mobile applications and malware detection techniques. They suggested deploying a two-level strategy to secure devices from the malware. First of all, the device must be secured from malware getting into the device. Secondly, the tools must be active in detecting existing malware. Malware detection techniques are signature based and anomaly based. They are also distinguished into host-based or cloud-based. Detection can be based on static or dynamic analysis. In the case of the dynamic analysis, the application is run and monitored in an isolated environment, while static analysis is made without the execution of the application.

Transmission of sensitive data does not always mean that there is a data breach. Thus, AppIntent framework (Yang et al., 2013) was developed to identify Android OS applications that leak the user's private data. The framework derives the user input/interaction information as context information. A sequence of user interface manipulations is presented to the human analyst for the final judgment of the application.

Alharbi and Yeh (2015) researched user interface design patterns of mobile applications. They decompiled applications to get layout files and byte code to extract various features, e.g. components, behaviour, usage of third-party libraries. A lot of mobile applications get access to private data from local sensors, e.g. GPS, camera, to use cloud-based services from the third-party. The problem is what data and where it is sent. TaintDroid (Enck et al., 2014) was created as a whole-system taint tracking framework to identify misbehaving applications. Taint tracking is based on semantic labelling of data items or variables. TaintDroid integrates four types of taint propagation: variable-level, method level, message-level, and file-level. FlowDroid (Arzt et al., 2014) analyses

the application's bytecode and configuration files to find potential privacy leaks. Thus, the framework considers full Android application lifecycle, correct callback handling, and UI widgets. FlowDroid generates a call graph and an inter-procedural control-flow graph to implement taint tracking.

Wang et al. (2016) introduced a behaviour chain based method to detect four types of Android malware. Privacy leakage is one of the types. A term behaviour chain refers to a sequence of behaviours. One process is behaviour that is made of actions or action groups.

Bottazzi et al. (2015) presented framework MP-Shield for phishing detection in Android mobile devices. MP-Shield natively intercepts IP packets. The framework uses a public phishing blacklist service provided by Google. Also, the tool runs machine learning algorithms based on WEKA framework to identify new phishing attacks based on URL attributes and properties. Modules of MP-Shield can be enabled / disabled on demand. Also, checks can be performed by the machine-learning engine and setup based on the needs. Mobile malware with self-updating capabilities can be detected using specific network traffic patterns learned locally, as semi-supervised machine learning methods can be applied to detect deviations from the application's expected normal behaviour (Shabtai et al., 2014).

Malware detection framework MalDozer (Karbab et al., 2018) is based on an artificial neural network that uses sequences of API method calls in the assembly code as an input. The tool can recognise malicious patterns automatically.

A lightweight mobile application certification (Enck, Ongtang and McDaniel, 2009) was implemented as a service that is based on security rules. At install time, the tool evaluates the application configuration. Specific combinations of security properties allow malfeasance. Geneiatakis et al. (2015) presented the permission certification technique that combined runtime information and static analysis to identify if applications are overprivileged and follow the least privilege principle. They also presented the risk assessment framework concerning granted permissions to sensitive sources. During the static analysis, a maximum set of permission is calculated, and their usage is validated during the dynamic analysis.

### **3. Framework architecture**

We develop the framework to detect security vulnerabilities in mobile applications to support cybersecurity experts in software assessment processes. Detection of vulnerabilities in mobile devices requires a sandbox as a compartmentalisation method to encapsulate and isolate communication activities. In this section, we present the design of the sandbox construction and the general architecture of the framework.

#### **3.1 Sandbox construction**

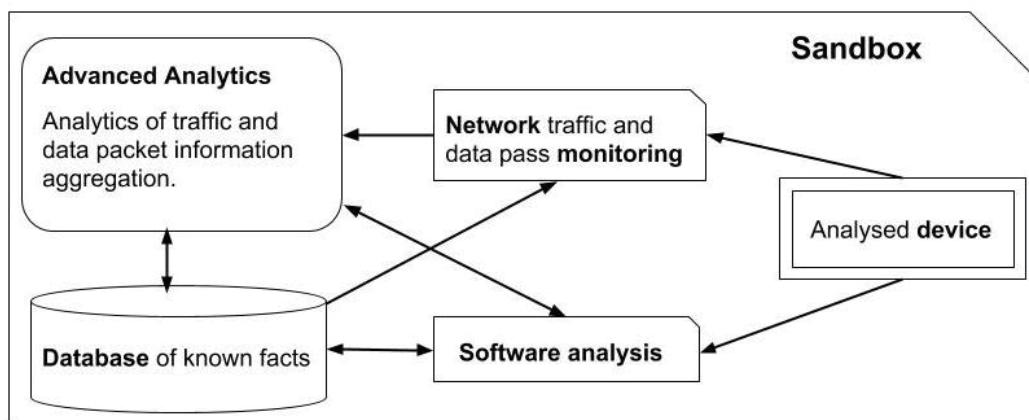
Application security assessment requires the application and the device placed in the prepared, isolated environment that mirrors the real-life situation. The sandbox fulfils this requirement. The sandbox construction includes initial setup, a configuration of system components, and preparation of data/information set to reflect a typical user of a mobile device. We distinguish six steps of sandbox construction.

- *1. setup.* Preparation of Network traffic and data pass monitoring component (NTDM) and its modules (if applicable). Set required GPS coordinates in GPS relocator. Connect to the Internet through Network Filter and Proxy + SSL Bridge. Prepare all Network Terminal Access Points (TAP) for gathering traffic. Setup a wireless access point and connect several neighbour devices to the local network. Simulate regular network communication in the local neighbourhood. Mark local network devices as trusted in NTDM, so their communications would be easily filtered out if needed.
- *2. Preparation of database of known facts.* Installation and configuration of the database management system(s), getting ready file system components to store various artefacts, e.g. certificates, data on specialised software, configuration templates and schemas, lists of sites, IP addresses.
- *3. Simulation of cloud service identities.* Creation of identities (user accounts) for different OS and cloud services, e.g. Google and Facebook accounts, required in typical mobile applications for identification and synchronisation.
- *4. Preparation for usage of payment service(s).* Setup of accounts/identities for payment services.

- 5. Preparation of data/information set. Making/generating a set of various personal and business data elements, e.g. contacts, photos, videos, calendar events, notes, office documents, of different timestamps for simulation of the real user, also, the creation of accounts on various services, e.g. e-mail, e-shop.
- 6. Integration and setup of components for advanced analytics. Setup of the virtual machine for components of advanced analytics, installation and integration of various tools, e.g. deep learning, neural networks, data mining algorithms for pattern mining (see Section 2 for available solutions)

### 3.2 General architecture

We present a general architecture of the framework for the detection of security vulnerabilities in mobile applications in Fig. 1. The framework is responsible for physical communication analysis and logical analysis of the application software at the code level. *Network traffic and data pass monitoring (NTDM)* component keeps track of network activities generated by the *Analysed device*. The configuration of the monitoring component is stored on the *Database of known facts*. Observed activities are sent to the component of *Advanced Analytics*. The *Software analysis* component is responsible for locating dangerous parts in the programming code. It uses *Database of known facts* for detection of suspicious sites, comparison of the device's application hash code to the hash code of the application version retrieved from the official store. Component of *Advanced Analytics* contains various artificial intelligence tools and algorithm implementations to detect malicious behaviour of applications based on malware pattern recognition, aggregated communication data. The component uses the *Database of known facts* to record historical data of physical communication and features of the application at the logical level. Also, it retrieves data to utilise *Advanced Analytics* tools. Thus, physical and logical analysis components use the central *Database of known facts* for continuous integration.

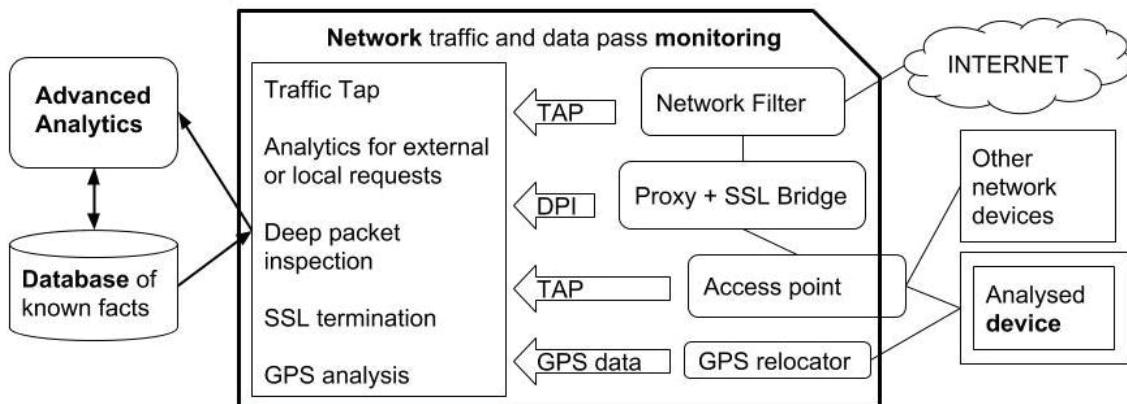


**Figure 1:** General architecture of the framework for detection of security vulnerabilities

NTDM component is illustrated in Fig. 2. Network level is analysed by listening to all the network traffic going from the *Analysed device* to the Internet and back. The device of interest is connected to the closed and entirely controlled network via the wireless access point, and the Internet connection goes through the *Network Filter*. Several additional network devices are connected to the same network to simulate regular neighbourhood interaction. Network Terminal Access Point (TAP) passes all traffic from access point and network filter to NTDM analytics module and later to *Advanced Analytics* component. *Database of known facts* is used to identify already known suspicious communication patterns and targets in the detected traffic.

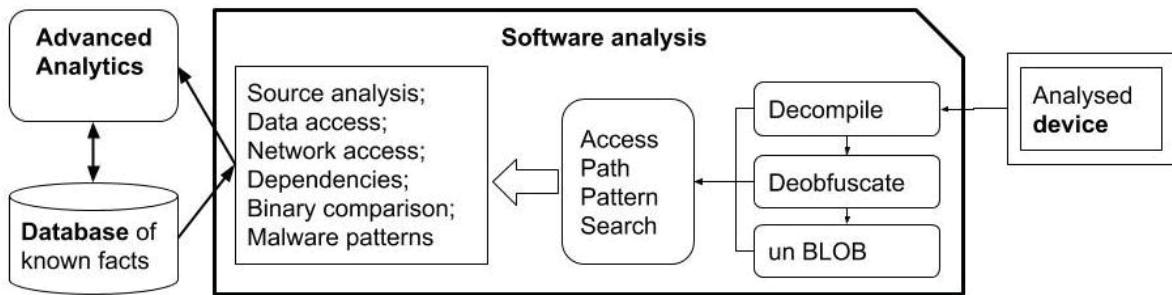
NTDM is a transparent network element where all the traffic is filtered and analysed. NTDM sub-component *Proxy + SSL Bridge* is based on a Next Generation Firewall. This component terminates all encrypted SSL traffic, and new SSL sessions are established to both parties of secure communication. Thus, decrypted data is accessible for further detailed analysis. Deep packet inspection (DPI) is done on all network data and to detect anomalies or undesired traffic.

*GPS relocator* emulates GPS activity on the fly or plays a predefined GPS route to the device. *GPS relocator* allows identifying how application behaviour changes in different locations. For example, offloading of gathered information might be done only in public places, or only after the device has visited some predefined locations. Also, *GPS relocator* provides the possibility to bypass Geo-fencing technologies.



**Figure 2:** Network traffic and data pass monitoring component

NTDM analysis module makes primary analysis and DPI of the collected data from *GPS relocator*, traffic pass-through and other network components. As all the data is already decrypted, the *Database of known facts* is used to detect and flag suspicious communication. Local requests to neighbouring devices are monitored, logged, analysed and in most cases flagged as suspicious. Requests to the Internet are distinguished by different parameters, e.g. manufacturer, country.



**Figure 3:** Software analysis component

*Software analysis* component is presented in Fig. 3. It analyses the behaviour of the application at the code level to detect unsafe parts. *Decompile* step creates a high-level source file of the mobile application taken from the device. Then, the analysis of the programming code is possible. As obfuscation might be used to hide malicious content in the software, the *Deobfuscate* step makes the decompiled code understandable by a human. Binary large objects (BLOBS) are identified, tagged, and tracked in the *un BLOB* step. BLOB objects are binary objects executed at run-time. Some BLOB objects are third-party binaries (libraries, drivers) used in the application or external calls to dynamic libraries written in other than Java code. Also, BLOB services are used to store images, documents, logs. Some BLOB objects might contain malicious software. *Software analysis* component locates dangerous parts in the programming code by comparing/analysing software code and matching patterns of data access, network access, etc. Application access requests from OS are analysed to define if they are required.

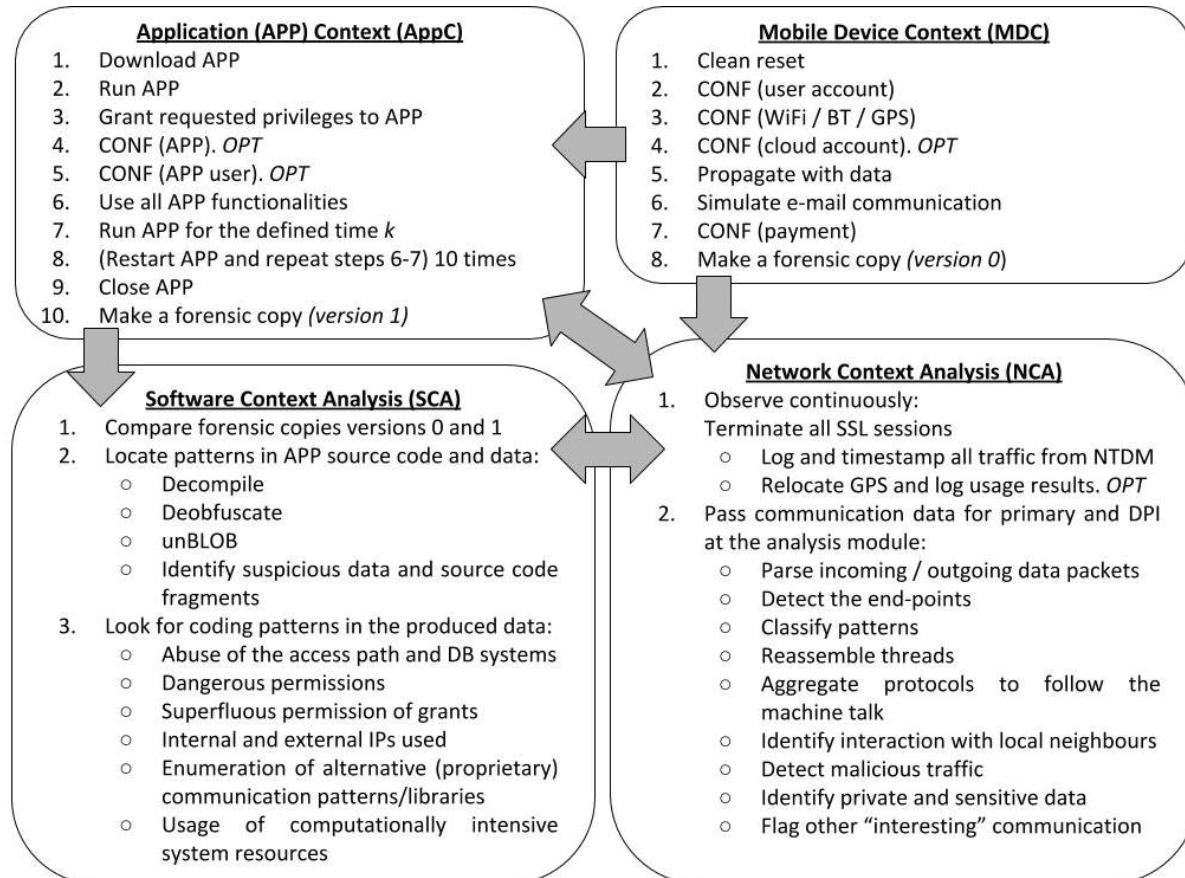
#### 4. Methodology for detection of security vulnerabilities

We developed a methodology to detect security vulnerabilities in mobile applications. The methodology assumes the sandbox and architecture presented in Section 3. The schematic view of the methodology is presented in Fig. 4. The methodology contains four phases: mobile device context (MDC), application context (AppC), software context analytics (SCA), and network context analytics (NCA). The figure uses notations CONF, APP, OPT to represent configuration, application interaction, and optional steps, respectively.

MDC makes the basis for the other phases with a configuration of the mobile device and propagating initial data prepared in sandbox construction steps. Thus, during MDC, the sandbox gets an initial network view with regular traffic and behavioural device patterns. Afterwards, AppC phase covers download, installation, configuration, execution, and repetitive testing/simulation of all functions of the tested application. During the AppC, the sandbox enables triggering and logging of the suspicious application behaviour in the isolated environment. MDC and AppC phases end by making a forensic copy of the device.

NCA phase is run in parallel with the MDC and AppC phases to monitor, log, and analyse all communication activities. Also, external and internal end-points are identified. NCA enables data packet parsing, classification of communication patterns. Therefore, traffic that goes beyond normal and safe application behaviour is detected and flagged, e.g. malicious traffic, data breaches.

SCA phase analyses forensic copies of the device (version 0 and 1) and the application itself to detect changes in the file system and suspicious code fragments, respectively. Thus, SCA includes decompilation, code deobfuscation, and unBLOBing. Therefore, the phase enables identification of superfluous or dangerous permissions, abuse of data access or encoded communication end-points. SCA can be run separately or in parallel with NCA to improve vulnerability detection rate and reduce false/positives.



**Figure 4:** Schematic view of the methodology to detect security vulnerabilities

After the methodology is applied to the tested application, the results are summarised to define the risk level. We use the Common Vulnerability Scoring System CVSS v3 score (FIRST.org, Inc., 2017) to assess the risk associated with possible software abuse. The scoring system associates base and *environmental* scores to particular value measures to compute a possible risk. The base metric measures the exploitability and intrinsic characteristics of the tested application. Attack vector, attack complexity, privileges required, and user interaction are related to the internal evaluation of technical means of detected exploits.

Extrinsic view on *Confidentiality*, *Integrity*, and *Availability* makes an impact on scoring, associating and calculating the risk from the environment. For the tested application, the string is generated to encode parameters with values for the vulnerability assessment.

For example,

AV:P:AC:L:PR:L:UI:N:S:C/C:H/I:H/A:L.../IR:M:AR:M/MAV:N/MAC:L/MPR:L/MUI:N/MS:U/MC:L/MI:L/MA:L

is evaluated as:

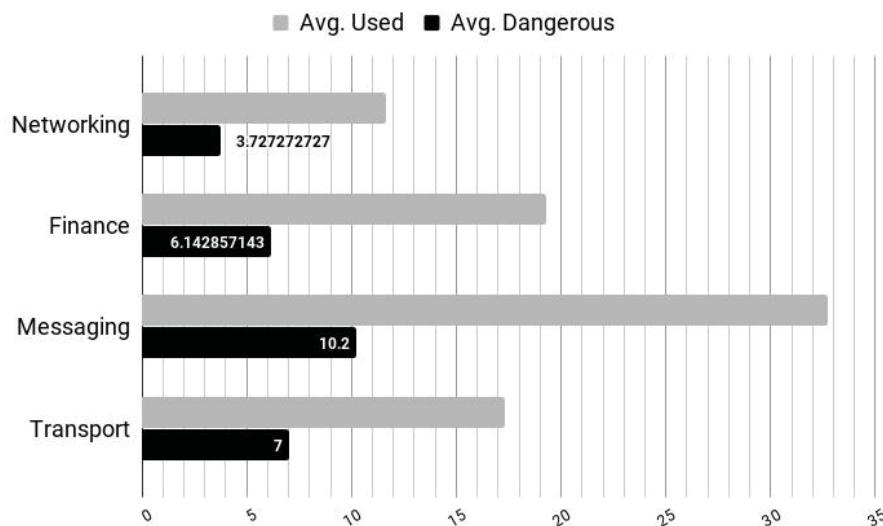
- CVSS 3.0 Base Score: 7.3, Rating : High;

- CVSS 3.0 Environmental Score: 6.8, Rating : Medium.

We visualise such results using Traffic Light Protocol (TLP) representation where High, Medium, and Low are mapped to Red, Amber, and Green, respectively. The CVSS v3 score computing is not fully automated, and it is a subject to the human judgement.

## 5. Discussion

We made an experiment on a number of applications in four categories: networking, finance, messaging, and transportation. In each category, at least 10 most popular and locally recognisable applications from different vendors were chosen. Analysed applications were downloaded from the official AppShop and tested in the sandbox of the framework. All tested applications were related to financial services, or a financial service was one of the features of the application and text/audio/video messaging. All applications initially have a known functionality for local users. The purpose of the experiment was to assess the security risk of the applications using our proposed methodology. Permissions are defined as dangerous if they enable access to user data or internal/external environment. Fig. 5 presents the average number of detected permissions in comparison with an average of dangerous permissions in the category. Results indicate that *messaging* applications request the most significant number of total and dangerous permissions on average. The second largest number of dangerous permissions is in the category of *transport* even though *finance* applications have the second largest number of total permissions on average. Such results are influenced by the specific use of hardware within the application.



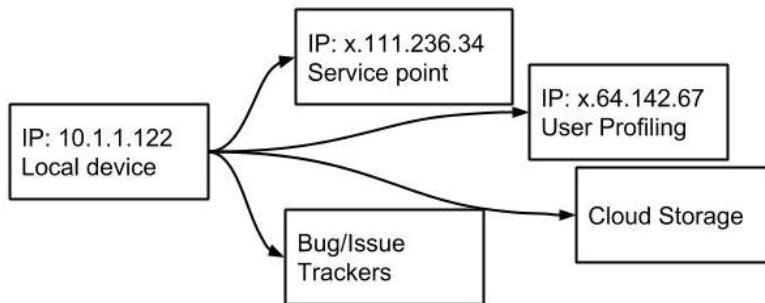
**Figure 5:** Average number of permissions tracked

SCA phase requires the expert to make a list of prioritised dangerous permissions. The expert considers the context when permissions are used as a part of known libraries, e.g., add, behaviour trackers. We prioritised breach of trust and private data. Thus, the priority for security permissions was set for *Confidentiality*. The most frequently requested permissions in the *finance* category were ACCESS\_FINE\_LOCATION, ACCESS\_COARSE\_LOCATION, READ\_CONTACTS, SEND\_SMS, READ\_CONTACTS, WRITE\_EXTERNAL\_STORAGE. In the *messaging* category, the most frequent dangerous permissions were READ\_CONTACTS, READ\_EXTERNAL\_STORAGE, CAMERA, READ\_SMS. Analysis of the source code produced external URLs that are not directly related to the service point of the vendor organisation (see Fig. 6). As shown in the figure, typically, most applications had led to external sites for user profiling, cloud services, developer bug reporting.

During the analysis, we discovered that on average a number of embedded remote site HTTP(S) URLs within the analysed applications is 2.63. Interesting to notice that some vendors use the same service providers. Thus, certain libraries lead to recurring websites throughout applications.

NCA phase enabled identification of the third-party websites requested during the initial phases of the tested application (HTTP URL requests). NCA phase also confirmed communication with external IP addresses to share information about customer usage behaviour. Non-secure HTTP protocol was used to access remote sites on

average 23 times per application session. In most detected cases, user behaviour trackers or client dataset acquisitions were sent. Secure communication to remote sites was observed on average 60 times per application session within the period of the test time.



**Figure 6:** Typical network traffic behaviour with request to remote sites

During the SCA phase, the local database schema was evaluated. The methodology enabled to identify schema creation and private data manipulation in plain text using local SQL storage (no encryptions whatsoever). We noticed that all analysed applications did use plain text information that disclosed private user behaviour. Also, most of the applications did not restrict the usage of the internal database. NCA indicated that applications mostly used the HTTPS protocol to access vendor services.

TLP visualisation of the risk level flagged most of the applications as Amber to draw the attention of the expert.

## 6. Conclusions and future work

We created the framework that reflects the real environment of the mobile application. In the sandbox, architectural solutions enable simulation of network communication and analysis of the application behaviour at physical and logical levels. The framework combines the available technological advances to detect anomalies and data breaches. Our methodology of vulnerability detection distinguishes four phases based on the device and application context. The methodology defines style and steps for application execution, analysis, and assessment. The application of the methodology shows that developers and providers sometimes abuse user's trust, request superfluous permissions, and possibly breach confidential data. Results showed that even financial sector applications did not encrypt sensitive data on a device, tracked user behaviour, and did not follow the recommendations for good coding practices applicable in mobile application development.

The work can be extended in several directions. Sometimes, the results of one analytical subcomponent require the execution of the particular algorithm, and at the same time, it makes some other tool ineffective due to the specifics of the found case. Thus, the modules of the Advanced Analytics component could be systematised by rationally integrating available technologies and algorithms. Furthermore, we could try to reverse engineer the application level protocols and decode their traffic. Also, the methodology could include fast detection of application changes due to in-app updates and the discovery of temporary unsafe code or code injected in remote resources. Finally, the detection of security vulnerabilities in hybrid applications should make a part of our methodology.

## References

- Acar, Y., Backes, M., Bugiel, S., Fahl, S., McDaniel, P. and Smith, M. (2016) "SoK: Lessons Learned from Android Security Research for Appified Software Platforms," in 2016 IEEE Symposium on Security and Privacy (SP). IEEE, pp. 433–451. doi: 10.1109/SP.2016.33.
- Alharbi, K. and Yeh, T. (2015) "Collect, Decompile, Extract, Stats, and Diff," in Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI). New York, New York, USA: ACM Press, pp. 515–524. doi: 10.1145/2785830.2785892.
- Arzt, S., Rasthofer, S., Fritz, C., Bodden, E., Bartel, A., Klein, J., Le Traon, Y., Octeau, D. and McDaniel, P. (2014) "FlowDroid: precise context, flow, field, object-sensitive and lifecycle-aware taint analysis for Android apps," in Proceedings of the 35th ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI). New York, USA: ACM Press, pp. 259–269. doi: 10.1145/2594291.2594299.
- Bottazzi, G., Casalicchio, E., Cingolani, D., Marturana, F. and Piu, M. (2015) "MP-Shield: A Framework for Phishing Detection in Mobile Devices," in 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous

- Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing. IEEE, pp. 1977–1983. doi: 10.1109/CIT/IUCC/DASC/PICOM.2015.293.
- Enck, W., Gilbert, P., Chun, B.-G., Cox, L. P., Jung, J., McDaniel, P. and Sheth, A. N. (2014) “TaintDroid: An Information Flow Tracking System For Real-Time Privacy Monitoring on Smartphones,” *Communications of the ACM*, 57(3), pp. 99–106. doi: 10.1145/2494522.
- Enck, W., Ongtang, M. and McDaniel, P. (2009) “On lightweight mobile phone application certification,” in Proceedings of the 16th ACM conference on Computer and communications security (CCS). New York, USA: ACM Press, pp. 235–245. doi: 10.1145/1653662.1653691.
- European Union Agency For Network And Information Security (ENISA) (2017) Privacy and data protection in mobile applications. A study on the app development ecosystem and the technical implementation of GDPR.
- FIRST.org, Inc. (2017) Common Vulnerability Scoring System SIG. Available at: <https://first.org/cvss/> (Accessed: January 31, 2019).
- Geneiatakis, D., Fovino, I. N., Kounelis, I. and Stirparo, P. (2015) “A Permission verification approach for android mobile applications,” *Computers & Security*. Elsevier Advanced Technology, 49, pp. 192–205. doi: 10.1016/J.COSE.2014.10.005.
- He, D., Chan, S. and Guizani, M. (2015) “Mobile application security: malware threats and defenses,” *IEEE Wireless Communications*, 22(1), pp. 138–144. doi: 10.1109/MWC.2015.7054729.
- Karbab, E. B., Debbabi, M., Derhab, A. and Mouheb, D. (2018) “MalDozer: Automatic framework for android malware detection using deep learning,” *Digital Investigation*. Elsevier, 24, pp. S48–S59. doi: 10.1016/J.DIIN.2018.01.007.
- Martini, B., Do, Q. and Choo, K.-K. R. (2015) “Conceptual evidence collection and analysis methodology for Android devices,” in Ko, R. and Choo, K.-K. R. (eds.) *The Cloud Security Ecosystem*. Syngress, pp. 285–307. doi: 10.1016/B978-0-12-801595-7.00014-8.
- NowSecure (2016) Mobile Security Report. Available at: <https://www.nowsecure.com/ebooks/2016-nowsecure-mobile-security-report/> (Accessed: January 31, 2019).
- Shabtai, A., Tenenboim-Chekina, L., Mimran, D., Rokach, L., Shapira, B. and Elovici, Y. (2014) “Mobile malware detection through analysis of deviations in application network behavior,” *Computers & Security*. Elsevier Advanced Technology, 43, pp. 1–18. doi: 10.1016/J.COSE.2014.02.009.
- Statista (2016) Number of smartphone users worldwide 2014-2020. Available at: <https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/> (Accessed: January 31, 2019).
- Statista (2018) Number of mobile app downloads worldwide in 2017, 2018 and 2022. Available at: <https://www.statista.com/statistics/271644/worldwide-free-and-paid-mobile-app-store-downloads/> (Accessed: January 31, 2019).
- Talha, K. A., Alper, D. I. and Aydin, C. (2015) “APK Auditor: Permission-based Android malware detection system,” *Digital Investigation*. Elsevier, 13, pp. 1–14. doi: 10.1016/J.DIIN.2015.01.001.
- Wang, Z., Li, C., Yuan, Z., Guan, Y. and Xue, Y. (2016) “DroidChain: A novel Android malware detection method based on behavior chains,” *Pervasive and Mobile Computing*. Elsevier, 32, pp. 3–14. doi: 10.1016/J.PMCJ.2016.06.018.
- Yang, Z., Yang, M., Zhang, Y., Gu, G., Ning, P. and Wang, X. S. (2013) “Applntent: analyzing sensitive data transmission in android for privacy leakage detection,” in Proceedings of the ACM SIGSAC conference on Computer & communications security (CCS). New York, USA: ACM Press, pp. 1043–1054. doi: 10.1145/2508859.2516676.
- Zonouz, S., Houmansadr, A., Berthier, R., Borisov, N. and Sanders, W. (2013) “Secloud: A cloud-based comprehensive and lightweight security solution for smartphones,” *Computers & Security*. Elsevier Advanced Technology, 37, pp. 215–227. doi: 10.1016/J.COSE.2013.02.002.

# Online Glossary of Cyber Security

Ladislav Burita

The Department of Informatics, Cyber Defence, and Robotics; Faculty of Military Technology, University of Defence, Brno

The Department of Industrial Engineering and IS, Faculty of Management and Economics, Tomas Bata University in Zlín; Czech Republic

[ladislav.burita@unob.cz](mailto:ladislav.burita@unob.cz)

**Abstract:** The research results published in the article aim to contribute to the solution for the absence of an internationally accepted and harmonized definition of cyber security and definitions of the related concepts. The goal of the case study is to create an Online Glossary of Cyber security as a Knowledge Management System (KMS) based on the software ATOM, a product of the Company AION-CS Zlín, Czech Republic. Characteristics and functions of the SW ATOM and its environments are mentioned in the paper. The literature review has confirmed that cyber security is a very frequently published topic, but the online dictionary or glossary is out of scientists' interest not only about cyber security. The working methodology of the research includes the steps: assembling of glossary documents and their analysis, determination of integrative themes and areas, selection of concepts for the glossary, creation of an ontology of the KMS and implementation of the glossary. The content of the online glossary in the theme of cyber security is divided into the three relevant areas: 1) Assets protected against cyber threats and attacks; 2) Threats and attacks from cyberspace; and 3) Set of means for protection against cyber threats and attacks. The volume of concepts in the glossary is about 30 and they were chosen for the case study to highlight the advantage of the online glossary layout in a KMS format. The glossary as a KMS offers complexity and flexibility more than a book; it is valuable for analysis, integration, education and study purposes. For example, the ontology includes recursive associations to the class CONCEPT that makes it possible to link terms according to their hierarchy or affinity, thus increasing understanding of the theme, when there are relationships between glossary terms. However, the successful finishing of the idea depends on fulfilling some requirements and the solution takes time for a team of scientists.

**Keywords:** cyber security, online glossary, knowledge management system, ontology, software ATOM

---

## 1. Introduction

Cyber security is a theme affecting the European Union (EU), individual states, companies or organizations, and individuals who work in cyberspace or just use its capabilities. The EU pays great attention to cyber security, EU has issued a series of recommendations and has a cyber security strategy (CSS) with the requirements to maintain cyber security (EU CSS, 2013). In addition to the EU directive, individual countries, such as the Czech Republic, are moving forward and have issued their own strategic documents (CR CSS, 2015).

For the individual strategies to be harmonized and the maximum effect achieved, individual actors who understand cyber security measures should have the same understanding of the basic concepts. In addition, here is the problem because interpretations of cyber security concepts are large and are often very different. It is documented by the research published in the article by Luijif, Besseling, and de Graaf (2013). The following results were published in the paper: "Observation 1: An internationally accepted and harmonized definition of 'Cyber security' is lacking. Observation 2: A global harmonized definition and understanding of 'cyber security' (and related terminology framework) would be beneficial to all nations."

## 2. The literature review

The source of the literature review was the portal Web of Knowledge (<http://apps.webofknowledge.com>). Cyber security is a very frequently published topic (5566 publications in the last 5 years, retrieved on 26 November 2018). The theme "Online dictionary or glossary of cyber security" was mentioned only in one paper (Murton, Johnston, and Waymire, 2014) but not in relation to analyzing the terms of the theme, but it was connected with being used in the simulation technology. "There are currently many Mod-Sim software tools available for use in characterizing various combat, physical, and cyber security scenarios. Even with the existence of a glossary distributed by the Modeling and Simulation Information Analysis Center, U.S. Department of Defense, there is a lack of requirement specifications and definitions for many of the terms and processes used in the modelling and simulation field."

The most cited publications (more than one hundred citations) of the cyber security theme were in 7 papers, oriented to Cyber security of the Industry 4.0: (1), Cyber security in the Internet of things (2), Protection from

attack against denial of service (1), Security of machine-to-machine network (1), Protection of vehicular systems (1), Security on demand response programs (1).

In the paper (Hwang, 2011) is presented a web-based object oriented model of language resources which are distributed in different places with variable forms in Internet. Language resources hosted in the web can be very easily utilized as components for application systems. So, web based linguistic components are particularly attractive for the maintenance of the web applications of the component, for changes immediately become effective to those applications. YDK (web based dictionary system for Korean language) can integrate variable sources of information in dictionaries on Internet, but also construct high quality dictionaries by wiring closely natural language systems residing in the network.

In the article (Medykovskiy and Chaplin, 2007) a method of semantic dictionary building is described. This method should be used for developing of semantic dictionary, which will be used to build semantic description of the graphic object. Later this semantic description can be used for search process in Automatic control systems organization.

### **3. Analysis of the glossary sources and ontology definition**

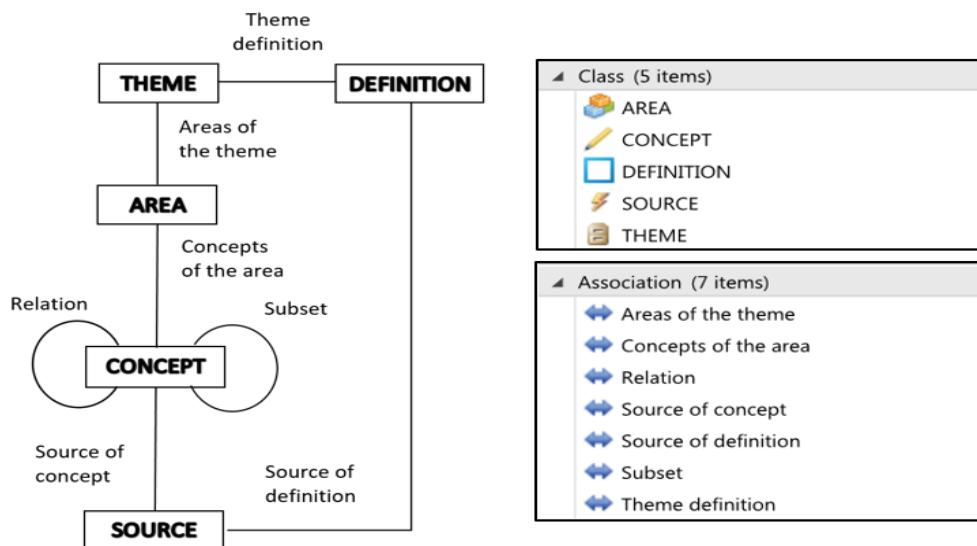
There are many glossaries of Cyber security available. Some of them are part of the documents of the cyber security strategy (CSS), for example Austria (AU CSS, 2013) with 22 terms, the Czech Republic (CZ CSS, 2015) with 8 terms, Poland (PL CSS, 2017) with 16 terms, and the United Kingdom (UK CSS, 2017) with 58 terms. National cyber security authorities issued some glossaries, for example US Glossary (2013) or CZ Glossary (2015).

Glossaries are published as original documents or a compilation from more sources (Maurer and Morgus, 2014).

As a theoretical basis for creating the online glossary was chosen the publication (Remenyi and Wilson, 2018) with respect to Copyright © 2018 APCIL "Except for quotation of short passages of critical review or use in research". The glossary and its complement with links to explanatory sources is an ideal and compact source for the online glossary development.

The ontology structure of the online glossary starts with class THEME, theme of a glossary. The online glossary can include more themes. Theme, as a core term in glossary, has included definitions in class DEFINITION that supported a privileged position theme in the KMS. Theme is explained in selected integration areas, the class AREA. The case study develops the theme of cyber security in three areas:

- Assets protected against cyber threats and attacks.
- Threats and attacks from cyberspace.
- Set of means of protection against cyber threats and attacks.



**Figure 1:** The ontology definition and implementation in SW ATOM, source own

Each area contains relevant concepts in class CONCEPT. The volume of concepts in the glossary is about 30 and they were chosen to highlight the advantage of the online glossary layout in a KMS format. Definitions and concepts are in relation to document sources (class SOURCE). The ontology definition and implementation in SW ATOM is depicted in Figure 1. Ontology contains five classes: AREA, CONCEPT, DEFINITION, SOURCE, and THEME; and seven associations between classes: Areas of the theme, Concepts of the area, Relation, Source of concept, Source of definition, Subset, and Theme definition.

The role of the association in the ontology is clear and do not need any explanation, except for the associations, Relation and Subset. Both associations are recursive to the class CONCEPT and make it possible to link terms according to their hierarchy, affinity, or any other relation.

#### 4. Characteristics of software ATOM

One of the goals and benefits of the SW ATOM (2018) is to support the implementation of projects of knowledge management systems (KMS), especially an effective development of web applications. The theoretical bases of the SW ATOM are Topic maps (Pepper, 2012), standardized in ISO/IEC 13250:2003. The Topic maps model consists of the three elements: topic (in SW ATOM class), association, and occurrences (in SW ATOM instances).

ATOM is a non-programming web database SW that does not require special knowledge. Anyone can easily construct a knowledge system on the web. The ATOM web database can be used as a construction kit for building web applications with powerful information retrieval used in encyclopaedias, dictionaries, collection of laws (Collection of Laws CR, 2018). The web application in the SW ATOM consists of the three components (environments): ATOM Studio, Data Editor, and User Portal.

The ATOM Studio creates and maintains ontologies and is used for batch exports and imports, advanced search, visualizations and complete system management. The Data Editor is used for inputting, editing, and deleting instances of classes, such as ORGANIZATION, PERSON or EVENT. The User Portal covers the knowledge base and it is designed for a user-friendly approach to information, stored in KMS. Company AION-CS should prepare a User Portal after the KMS evaluation is finished in the operational conditions.

#### 5. The online glossary as a KMS

KMS as an application in the ATOM environment as an ontology-driven solution is user-friendly and simple to use. The functionality of the online glossary includes data input and edit, development of the complex information structure, powerful information retrieval in class instances, built-in forms or full-text search.

The built-in form for data input of the class CONCEPT is shown at Figure 2; list of class AREA instances is shown at Figure 3; the complex data structure of the instance class CONCEPT, using associations Relation and Subset, is shown at Figure 4.

The screenshot shows a software interface for inputting data into a CONCEPT class. The title bar says "CONCEPT Biometrics". The main area has a left sidebar with checkboxes for "Other data", "Definition", "Definition-national", "Other Definition", "Other Definition-national", "Web-link to explanatory source", "Concepts in area", "Concept source", "Concept in relation to", "Relation of concept", "Contains", and "Is a subset of". The right side shows a detailed view of the "Definition" section. It contains a text area with the following content:

In cyberspace biometrics is the use of unique and identifiable bodily characteristics in order to identify individuals. Thus biometrics consists of fingerprints, iris impressions, hand geometry, voice recognition etc. It has been suggested that it is the most reliable way of identifying and individual.

Biometrics is increasingly considered to be a useful way to ensure the secure access of an individual to his or her computer and network.

Below this, there are two "Target" sections. The first Target is linked to "http://searchsecurity.techtarget.com/definition/biometrics" and describes biometrics as a means for protection against threats and attacks in cyberspace. The second Target is linked to "Glossary of Cyber Warfare, Cyber Crime and Cyber Security by Remenyi & Wilson, 2018" and describes it as a glossary of terms related to cyber security.

Figure 2: Form for the data input of the class CONCEPT, source own

The screenshot shows a web-based application interface for the 'AREA Threats and attacks from cyberspace' class. At the top, there is a navigation bar with a 'Return' button. Below it, a sidebar on the left lists various data types: 'Other data', 'Annotation', 'Area of the theme', and 'Concepts of the area'. The main content area displays a detailed description of the class, mentioning threats like Cyber threat, Cyber Attack, Cybercrime, Malware, and social engineering. A 'Target' section lists numerous specific cyber threats and activities, each preceded by a small icon.

Target
Cyber security
Computer intrusion
Cyber attack
Cyber crime
Cyber infiltration
Cyber target
Cyber threat
Denial of service (DoS)
Distributed DoS
Hacker
Malware
Phishing
Ransomware
Spyware
Target
Trojan horse
Virus
Worm

Figure 3: Class AREA, instance threats and attacks from cyberspace, source own

This screenshot shows a 'CONCEPT Cyber attack' page. It includes a 'Return' button at the top right. On the left, a sidebar lists 'Other data', 'Definition', 'Web-link to explanatory source', and several relationship types: 'Concepts in area', 'Concept source', 'Concept in relation to', 'Relation of concept', and 'Contains'. The main content area contains a detailed definition of a cyber attack as a hostile act through a computer or network. It also includes a 'source' link to a web page and a 'Target' section listing various cyber threats and activities, similar to Figure 3.

Target
Threats and attacks from cyberspace
Glossary of Cyber Warfare, Cyber Crime and Cyber Security by Remenyi & Wils
Incident response
Target
Denial of service (DoS)
Distributed DoS
Hacker
Malware
Phishing
Ransomware
Spyware

Figure 4: Class CONCEPT, instance cyber attack, source own

The form of class CONCEPT (Figure 2) contains all characteristics of the class defined in the ontology. There are Definition and other definitions in English, Definition and other definitions in the national language, Web links to explanatory sources, associations to class AREA (part of the glossary), SOURCE of the main definition, and associations to other instances of the class CONCEPT that are in the subset, hierarchy, or any relation.

The instance of the class AREA Threats and attacks from cyberspace, at Figure 3, is specified in the characteristic Annotation, contains the theme Cyber security, and includes a set of instances of the class CONCEPT.

The instance Cyber attack of the class CONCEPT (see Figure 4) contains the following characteristics: Definition in English, Web link to explanatory source, associations to class AREA, SOURCE, and associations that include other instances of the class CONCEPT. The online glossary of the theme cyber security is available at the <http://unobtest.atom3.cz/>, Ontology: Cyber security (ID: guest / PW: guest). Remarque: the TESTUNOB environment contains more KMS.

## **6. Conclusions**

The online glossary of the theme cyber security is a proof of concept. The glossary includes only a small set of concepts to highlight possibilities of the KMS to contribute to the goal “An internationally accepted and harmonized definition of cyber security and related terminology framework”.

The structure of the online glossary and the functionality of the SW ATOM allows work with more themes of the dictionary. The exception of the theme Cyber security should, for example, include the themes Cyber defense or Cyber warfare. An online glossary makes it possible to collect definitions of the theme and the concept at the one place and to create conditions that can help integrate the set of definition of the theme or concept.

The acceptance of that idea depends on its acceptance in the cyber security community and in the audience at the ECCWS-2019 conference. Any remarks, observations, and proposals for online glossary solution are invited.

The possible implementation of an online glossary in practice depends on:

- Obtaining permission from a dictionary owner to include terms into the online glossary.
- Support of Company AION-CS, Zlín; license to use the ATOM SW to create a dictionary.
- Getting a project for glossary development.
- Creating an international team that meets the project goals.

## **Acknowledgements**

First, thanks to the authors of the dictionary (Remenyi and Wilson, 2018) for the use of several concepts for the glossary as a case study. Thanks to the Company AION-CS Zlín for support with using the SW ATOM in the test environment.

The article presents the results of the research in cyber security as a part of the project (DZRO-209, 2018).

## **References**

- AU CSS. (2013) “Austrian cyber security strategy”, [online], [https://www.bmi.gv.at/504/files/130415\\_strategie\\_cybersicherheit\\_en\\_web.pdf](https://www.bmi.gv.at/504/files/130415_strategie_cybersicherheit_en_web.pdf)
- CZ CSS. (2015) “National cyber security strategy of the Czech Republic for the period from 2015 to 2020”, [online], NCKB, Prague, <https://www.enisa.europa.eu/about-enisa/structure-organization/national-liaison-office/news-from-the-member-states/czech-republic-national-cyber-security-strategy-2015-2020>
- CZ Glossary. (2015) “Cyber Security Glossary”, [online], NCKB, Prague, <https://www.govcert.cz/cs/informacni-servis/akce-udalosti/2193-vykladovy-slovnik-kyberneticke-bezpecnosti-druhe-vydani/>
- DZRO-209. (2018) “Development of systems C4I and cyber security,” Brno, University of Defence, 2016-2020.
- EU CSS. (2013) “Cybersecurity Strategy of the European Union: An Open, Safe and Secure Cyberspace”, [online], EU, Brussels, [https://eeas.europa.eu/archives/docs/policies/eu-cyber-security/cybsec\\_comm\\_en.pdf](https://eeas.europa.eu/archives/docs/policies/eu-cyber-security/cybsec_comm_en.pdf)
- Hwang, D. A. (2011). Dictionary Development System based on Web. INFORMATION-AN INTERNATIONAL INTERDISCIPLINARY JOURNAL, Volume: 14, Issue: 11, Pages: 3575-3582.
- Luijff, H.A.M., Besseling, K., Spoelstra, M. and de Graaf, P. (2013) “Ten National Cyber Security Strategies: a Comparison”, Conference proceedings Critical Information Infrastructure Security, [online], TNO, The Hague, <https://www.researchgate.net/publication/261987241>

**Ladislav Burita**

- Maurer, T., and Morgus, R. (2014) "Compilation of Existing Cybersecurity and Information Security Related Definitions", [online], Open Technology Institute New America, <https://www.newamerica.org/cybersecurity-initiative/policy-papers/compilation-of-existing-cybersecurity-and-information-security-related-definitions/>
- Medykovskiy, M., and Chaplahin, M. (2007). Semantic dictionary development method for building graphic object's semantic description. PROCEEDINGS OF THE 9TH INTERNATIONAL CONFERENCE ON THE EXPERIENCE OF DESIGNING AND APPLICATION OF CAD SYSTEMS IN MICROELECTRONICS, Pages: 541-541.
- Murton, M., Johnston, P., Waymire, R., and et al. (2014) "A Fidelity Framework for Small Arms Combat, Proceedings of the conference: 48th Annual IEEE International Conference Carnahan on Security Technology (ICCST), [online], Rome, Italy, [http://apps.webofknowledge.com/full\\_record.do?product=WOS&search\\_mode=GeneralSearch&qid=16&SID=D3eGGx6qdyMQWwur5FF&page=1&doc=1](http://apps.webofknowledge.com/full_record.do?product=WOS&search_mode=GeneralSearch&qid=16&SID=D3eGGx6qdyMQWwur5FF&page=1&doc=1)
- Pepper S. (2012) "The TAO of Topic Maps", [online], <http://www.ontopia.net/topicmaps/materials/tao.html>
- PL CSS. (2017) "National framework of cybersecurity policy of the Republic Poland for 2017-2022", [online], [https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/Cybersecuritystrategy\\_PL.pdf](https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/Cybersecuritystrategy_PL.pdf)
- Collection of Laws CR. (2018). "Zákony pro lidi (Laws for people)", [online], AION-CS, Zlín, <https://www.zakonyprolidi.cz/>
- Remenyi, Dan and Wilson, Richard L. (2018) Glossary of Cyber Warfare, Cyber Crime and Cyber Security, Academic Conferences and Publishing International Limited, Lighting Source.
- SW ATOM. (2018) "The web database software ATOM, developed by AION-CS", [online], AION-CS, Zlín, <https://www.aion.cz>
- UK CSS. (2017) "National cyber security strategy 2016-2021, [online], [https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/national\\_cyber\\_security\\_strategy\\_2016.pdf](https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/national_cyber_security_strategy_2016.pdf)
- US Glossary. (2013) "Glossary of Key Information Security Terms", [online], NIST, <https://www.hSDL.org/?view&did=738581>

# A Framework for Managing Cybersecurity Effectiveness in the Digital Context

**Marian Carcary, Eileen Doherty and Gerry Conway**  
**Innovation Value Institute, Maynooth University, Ireland**  
[Marian.carcary@mu.ie](mailto:Marian.carcary@mu.ie)

**Abstract:** The pace of digital transformation and new technology development and the growing sophistication of cyber criminals result in organisations facing greater scope and severity of cybersecurity attacks on a daily basis - estimated to cost between \$375 and \$575 billion per annum. It is anticipated that as more devices, systems, and infrastructure become interconnected and interdependent, and as more interfaces between customers, suppliers, and partners are leveraged, the IT 'attack surface' will continue to expand. Organisations vary in their approaches to attempting to prevent cybersecurity breaches: some are overly restrictive, making even routine business activities difficult, while others are too relaxed with poor oversight and inadequate protocols and procedures, creating unnecessary exposures. However, applying appropriate cybersecurity controls is now a particular necessity where digital leaders often have a higher tolerance and appetite for risk-taking and experimentation to identify key opportunities for the future. Organisations now need to rethink their cybersecurity management approaches, and recognise that traditional access control and perimeter defences alone are no longer sufficient. Rather holistic and proactive approaches that continually evolve and adapt to counter emerging threats and minimise the potential negative consequences of exposure are required. Understanding how effective the organisation is in its cybersecurity efforts is a prerequisite for ensuring controls remain abreast with, and appropriate for, the changing IT threat landscape. This paper presents a cybersecurity conceptual framework that can be used by organisations to provide a holistic analysis of their cybersecurity approaches. It details the key factors or management themes underpinning cybersecurity effectiveness and how the insights gained through assessing performance against these factors or management themes can be practically used to improve cybersecurity effectiveness.

**Keywords:** cybersecurity assessment, cybersecurity management, cybersecurity drivers, cybersecurity barriers, threats

---

## 1. Introduction

In recent years, the evolution of a relatively inexpensive and accessible worldwide digital infrastructure has served as a platform for the proliferation of a growing range of digital technologies, which are transforming the strategic context of the organization and its competitive environment (Bughin et al, 2017; Dahlström et al, 2017; Fichman et al, 2014; Ross et al, 2016; Sambamurthy and Zmud, 2012; Stanske and Kautz, 2018). While this era of digital transformation offers organisations considerable opportunities, it is also reflective of significant risks, threats, and uncertainties. The range of threats now faced by organisations is unprecedented, and the actors perpetrating such attacks span hacktivists with political agendas, lone wolf hackers, organised criminal syndicates, state sponsored attackers, external contractors, or corporate insiders, among others. In many instances, the purpose of such cyberattacks is the unauthorised access to and theft of corporate or personal data. This poses a particular challenge for organisations, given the potential legal, financial, and reputational implications that arise from a data breach (Arend et al, 2016).

Organisations certainly vary in their approaches to attempting to prevent cybersecurity breaches: some are overly restrictive, making even routine business activities difficult, while others are too relaxed with poor oversight and inadequate protocols and procedures, creating unnecessary exposures. In a recent industry survey, 87% of respondents believed that their cybersecurity budget or approaches did not meet their organisations' needs – protection levels were patchy and cybersecurity remained siloed or isolated; more than half of organisations did not make protection against cyber-attacks an integral part of their strategy (Ernst & Young, 2018). Applying appropriate levels of cybersecurity controls is a particular necessity in the current landscape where digital leaders often now have a higher tolerance and appetite for risk-taking and experimentation (Bradley et al, 2015; Hussain et al, 2007a, 2007b; Lee and Baby, 2013). CEOs/CIOs with a more 'risk-on' attitude to technology engage in risk-taking actions, experimentation, and digital business innovation using exploratory, fail fast approaches to identify the key opportunities for the future (Bradley et al, 2015). An organisational culture that embraces and supports more entrepreneurial and active risk-taking in digital programmes is characteristic of higher organisational performance (Rickards et al, 2015).

Implementing overly restrictive or excessively weak cybersecurity controls can result in regulatory, legislative, financial, and reputational implications that can impact business continuity (ISACA, 2012, 2013). Protection of the organisation's computing environment/infrastructure from cyberattacks that can impact business continuity

and the organisation's protection of key information assets must now be central to its core operations. The organisation needs to find the right balance in order to secure its IT resources without impeding effective business operations. In the digital era, it needs to rethink its cybersecurity management approaches, and recognise that traditional access control and perimeter defences alone are no longer sufficient. Rather holistic and proactive approaches that continually evolve and adapt to counter emerging threats and minimise the potential negative consequences of exposure are required (Accenture, 2018; Fraser et al, 2014). Understanding how effective the organisation is in its cybersecurity efforts is a prerequisite for ensuring controls remain abreast of the changing IT threat landscape.

This paper addresses the following objectives:

- Outline key factors or management themes underpinning cybersecurity effectiveness in the digital context.
- Present a conceptual framework that can be used by organisations to understand and improve their cybersecurity effectiveness.
- Elucidate how implementation of the key cybersecurity factors or management themes can result in enhanced cybersecurity management effectiveness

## **2. Research methodology**

A multi-method research approach was adopted.

Phase 1: An in-depth literature review was initially undertaken, focused on the key factors or management themes underpinning cybersecurity effectiveness in the digital context. The literature review covered both academic journals, IT industry publications, and industry and legislative standards (e.g. European Parliament, 2016a, 2016b; ISACA, 2012, 2013; ISO, 2009, 2013a, 2013b; National Institute of Standards and Technology, 2013; The Open Group, 2011). A content analysis of the material extracted from the literature was undertaken to establish the most common concepts. The authors followed the concept matrix method (Webster and Watson, 2002).

Phase 2: Building on the concepts identified through the in-depth literature review process, subject matter experts from both industry and academia worked collaboratively based on the principles of Open Innovation (Chesbrough, 2003) in a shared interest workgroup to create a conceptual framework for improving cybersecurity effectiveness (Chesbrough, 2003). These principles supported the leveraging of resources, knowledge, and expertise from multiple stakeholders whereby engaged participants communicate, exchange views, and develop a research output within a collaborative research environment. This resulted in the development of a conceptual framework and 'Cybersecurity Effectiveness Assessment (CEA)' which could be used by the organisation to holistically analyse its performance with respect to cybersecurity management. The value of adopting an open innovation approach centred on both the clear articulation of challenges by practitioners who worked in the cybersecurity context, and the grounding of the framework's development in these insights. A rigorous review of the developed CEA was also undertaken within the organisational context by a number of industry practitioners. Feedback from this review process was incorporated, resulting in the development of a CEA that had increased practical utility in the industry context.

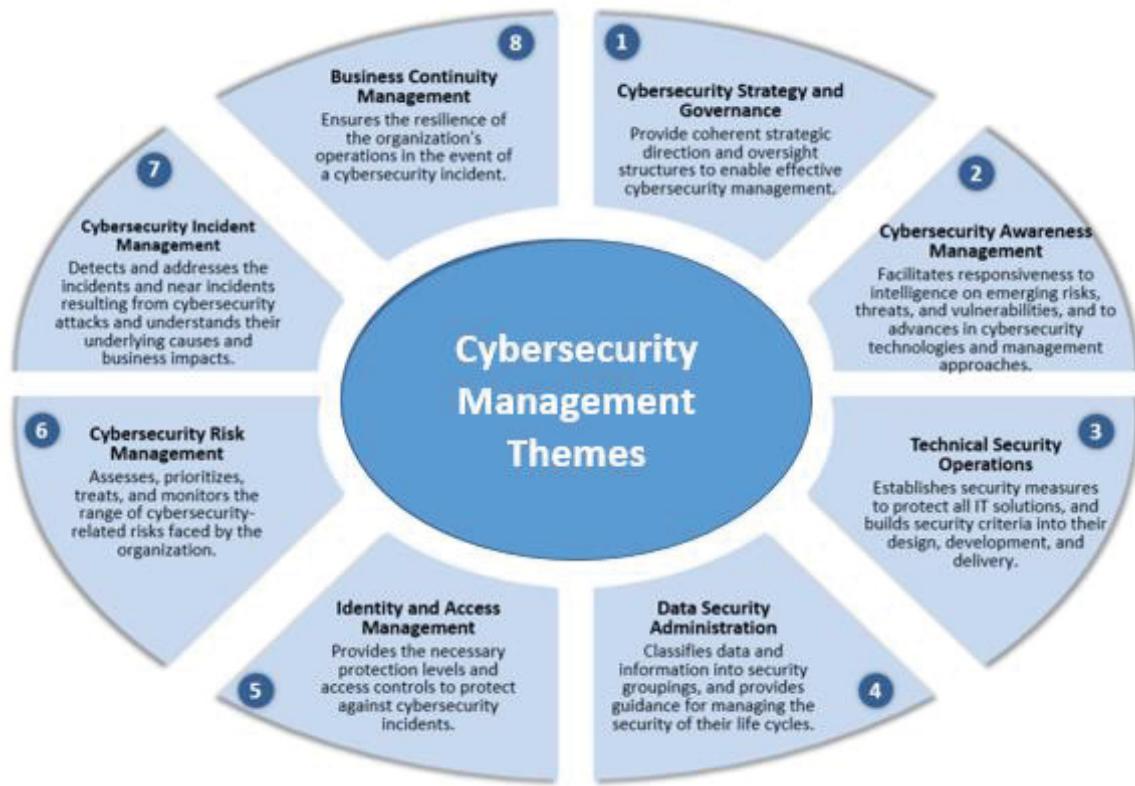
## **3. Cybersecurity conceptual framework**

Informed by an in-depth literature review and insights gained from engaging in collaborative research with subject matter experts, the cybersecurity conceptual framework is presented and discussed in this section. This framework categorises effective cybersecurity as being underpinned by eight key factors or management themes (Figure 1). Each factor or management theme and the impact on the organisation of improving in each area are briefly described below.

The *Cybersecurity Strategy and Governance* theme provides coherent strategic direction and oversight structures to enable effective cybersecurity management.

**Description:** Literature and industry expert insights suggest that organisations need to establish enterprise-wide governance structures for effective cybersecurity. Mitigating cybersecurity threats needs to be an organisation-wide strategic priority, and senior management's direction on and sponsorship of the cybersecurity programme needs to be evident, with supporting comprehensive cybersecurity strategies and policies. Collaborative,

partnership type relationships are needed between IT and business leaders and business ecosystem partners so that they can jointly agree on the required cybersecurity levels. Responsibilities and accountabilities also need to be allocated for cybersecurity activities to those who have the requisite experience, knowledge, skills, and authority, and cybersecurity performance measures regularly compared against targets and benchmarks.



**Figure 1:** Conceptual framework - cybersecurity factors or management themes

**Impact:** Through improving the organisation's *Cybersecurity Strategy and Governance*, cybersecurity becomes embedded in day-to-day work practices and protecting the organisation's information assets is regarded as part of everyone's job. Cybersecurity threats are more effectively managed across the entire value chain. The organisation is better able to demonstrate through regular audits that it complies with all relevant laws, regulations, standards, policies, and contractual requirements pertaining to cybersecurity.

The *Cybersecurity Awareness Management* theme facilitates responsiveness to intelligence on emerging risks, threats, and vulnerabilities, and to advances in cybersecurity technologies and management approaches.

**Description:** Literature and industry expert insights suggest that the organisation needs to be cognisant of the greater complexities associated with protecting its data and information assets, arising from the pervasiveness of digital technologies and unprecedented data volumes. Consequently, IT leaders need to keep abreast of new advances in security technologies and cybersecurity management approaches through, for example participating in industry sharing communities. Given the criticality of employees in protecting against cyber-attacks, the organisation needs to provide access to comprehensive employee-directed cybersecurity education, training and certifications, and other employee developmental mechanisms.

**Impact:** Through improving the organisation's *Cybersecurity Awareness Management*, the organisation has greater insight into potential attack profiles; can rapidly identify new and emerging cybersecurity risks, threats, and vulnerabilities; and act with a speed and scale of response that is proportionate to the business consequences of a cybersecurity breach/incident, or near incident. Employee decision-making on cybersecurity-related risks, threats, and incidents is better facilitated and the organisation is enabled to be more creative and agile in detecting and addressing potential future attacks.

The *Technical Security Operations* theme establishes security measures to protect all IT solutions, and builds security criteria into their design, development, and delivery.

**Description:** Literature and industry expert insights suggest that the security architecture needs to evolve with changes in the enterprise architecture and reflect all information domains including infrastructure, application, data, integration, and operations at conceptual, logical, technical, and physical levels for all IT solutions and services. Cybersecurity guidelines should be mandated to ensure security criteria is built into the design and development of all IT solutions and service components. Security measures should also be systematically implemented for all IT systems, components, and devices, and test approaches employed to validate that the required levels of security are provided.

**Impact:** Through improving the organisation's *Technical Security Operations*, security is considered throughout the entire IT solutions' life cycle and all IT systems, components, and devices provide the required levels of security to protect against cybersecurity incidents.

The *Data Security Administration* theme classifies data and information into security groupings, and provides guidance for managing the security of their life cycles.

**Description:** Literature and industry expert insights suggest that the organisation needs to adopt holistic and proactive approaches to data and information security management. Comprehensive security classifications need to be put in place for virtually all datasets; and comprehensive guidelines are required for managing the security of life cycle states and life cycle state transitions.

**Impact:** Through improving the organisation's *Data Security Administration*, the organisation's data and information security management approaches are continually adapted to keep pace with changing business requirements and to counter evolving threats. Data throughout its life cycle is appropriately secured against unauthorised or unlawful processing, accidental loss, or damage, and the integrity, confidentiality, accountability, usability, and availability of the organisation's data and information assets are effectively safeguarded.

The *Identity and Access Management* theme provides the necessary protection levels and access controls to protect against cybersecurity incidents.

**Description:** Literature and industry expert insights suggest that comprehensive protection level and access control guidelines and criteria are needed for all architecture layers, networks, IT solutions and services, and data stores. Authentication processes should be supported by sophisticated systems, tools, and techniques, and access rights should be dynamic and flexible. Data, applications, systems, and network access activity should be continually and rigorously audited, and scrutinised to identify opportunities for improvement.

**Impact:** Through improving the organisation's *Identity and Access Management*, the organisation's access controls reflect all required security and configuration settings and are responsive to IT, business, or personnel changes, such as changes in the scope of an employee's responsibilities or job role.

The *Cybersecurity Risk Management* theme assesses, prioritises, treats, and monitors the range of cybersecurity-related risks faced by the organisation.

**Description:** Literature and industry expert insights suggest that it is necessary for the organisation to adopt a strategic, enterprise-wide focus on cybersecurity risk management, with risk management philosophies embedded in all relevant processes and a holistic focus applied to both business exposures and technical risks. Cybersecurity risk profiles should be defined in collaboration with business ecosystem partners, be regularly benchmarked against external sources for validity and quality assurance, and be systematically used in the organisation's risk assessment, prioritisation, and treatment activities.

**Impact:** Through improving the organisation's *Cybersecurity Risk Management*, the organisation becomes more adept at identifying and prioritising the cybersecurity-related risks to which the organisation and the wider business ecosystem might be vulnerable, and applying appropriate cybersecurity risk treatment strategies in alignment with the organisation's risk tolerance and risk appetite.

The *Cybersecurity Incident Management* theme detects and addresses the incidents and near incidents resulting from cybersecurity attacks and identifies their underlying causes and business impacts.

**Description:** Literature and industry expert insights suggest that comprehensive approaches to cybersecurity incident management should be put in place and incidents prioritised based on asset criticality and the urgency to restore services as defined by Service Level Agreements (SLAs). Diagnostic metrics should focus on minimising repeat cybersecurity incidents and the advanced analytics-driven detection/prediction of potential incidents before they occur.

**Impact:** Through improving the organisation's *Cybersecurity Incident Management*, the classification, underlying cause, and impact of all incidents and near incidents, including recurring incidents, are effectively diagnosed and corrective measures effectively identified.

The *Business Continuity Management* theme ensures the resilience of the organisation's operations in the event of a cybersecurity incident.

**Description:** Literature and industry expert insights suggest that business continuity planning should cover all areas of the IT infrastructure and be part of all solutions development, procurement, testing, and deployment. Organisations require the ability to complete IT system restorations locally or remotely from other designated sites, with backup systems available to maintain critical systems online even if a site and its IT infrastructure are taken offline. Business continuity plans should be regularly reviewed and leverage the latest business recovery configuration options.

**Impact:** Through improving the organisation's *Business Continuity Management*, business continuity plans reflect business recovery priorities, business cycles, and operations criticality, and business continuity is effectively safeguarded in the event of cybersecurity incidents.

#### **4. Operationalising the cybersecurity conceptual framework – the CEA**

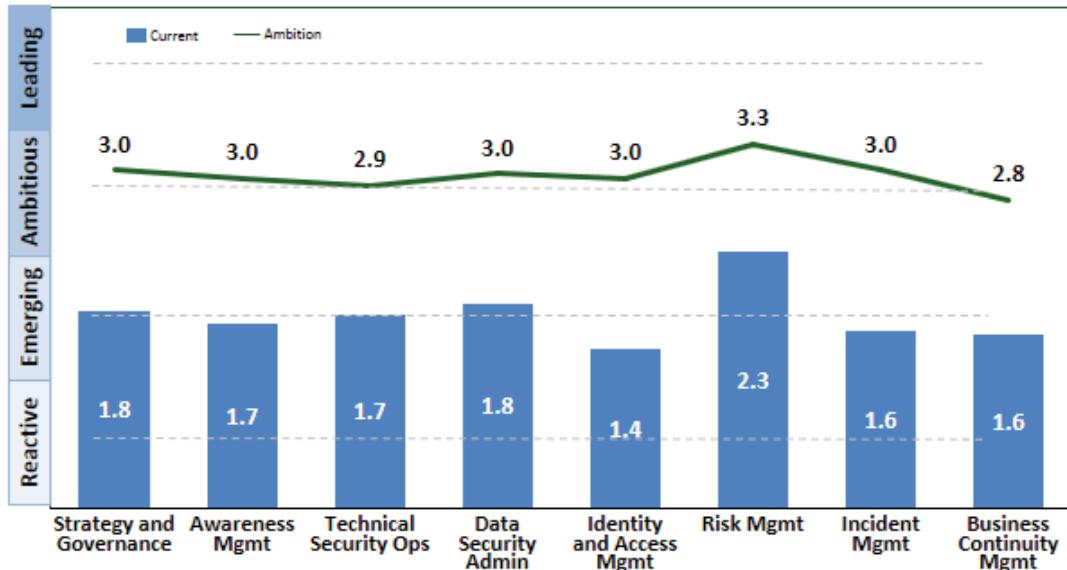
The conceptual framework discussed in section 3 served as the basis for developing a cybersecurity assessment instrument that can be used by organisations to provide a holistic analysis of their cybersecurity approaches and, through the insights gained from the analysis, practically improve their cybersecurity effectiveness. The CEA is based on a simple online maturity-based quantitative survey instrument that captures the organisation's current and future ambition level of achievement across the eight identified factors or management themes, which are decomposed into 45 individual 'behaviour' statements. The CEA should have the buy-in and commitment of the organisation's C-suite so that the importance of cybersecurity is reflected in the organisation's strategic direction and the insights gained can be executed with appropriate sponsorship and resourcing. The assessment typically captures multiple perspectives from participants who have a sufficient level of knowledge and are engaged in the organisation's cybersecurity efforts, such as chief information officers, chief technology officers, chief security officers, chief information security officers, information security directors, security architects, security auditors, data protection/privacy officers, and network security officers, among others. These participants are asked to score the organisation's current level of achievement and future target level of achievement across the 45 behaviour statements. The survey findings may also be validated through qualitative interviews with key stakeholders.

Insights gained from the CEA include a detailed understanding of the organisation's current and target maturity across all eight cybersecurity factors or management themes at a high level, and across all 45 behaviour statements at a detailed level. Figure 2 and Figure 3 provide sample insights that can be derived through undertaking the CEA. Figure 2 outlines the organisation's current versus target level of maturity across the eight management themes, while Figure 3 outlines the organisation's importance rating vis-à-vis its management theme maturity gap and provides key insights into target areas for improvement.

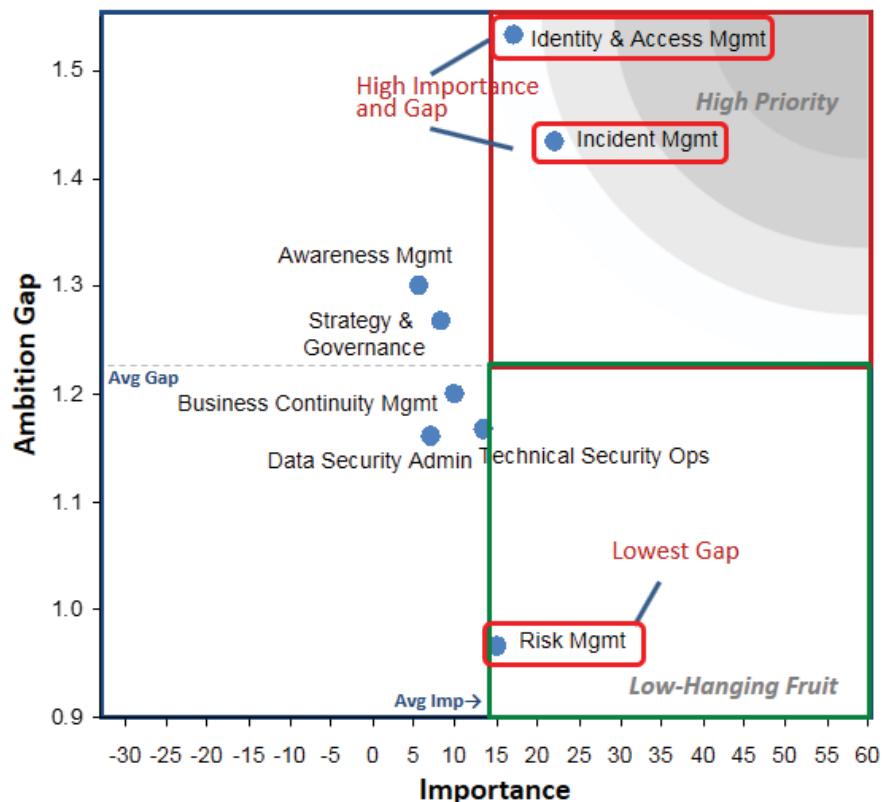
Through identifying the key target areas for improvement, this enables the organisation to concentrate on developing the capabilities of greatest importance to meet its specific security requirements (Figure 4). The CEA identifies a subset of organisational capabilities associated with the organisation's targeted areas of improvement and key recommended improvements in relation to the capability shortlist.

Many of the shortlisted capabilities identified (Figure 4) in the CEA are drawn from the IT Capability Maturity Framework (IT-CMF™) (Curley et al, 2017, 2016). IT-CMF™ reflects a compilation of 37 Critical Capabilities (CCs) (Figure 5) which are key for the successful management of IT and is designed to systematically lead the

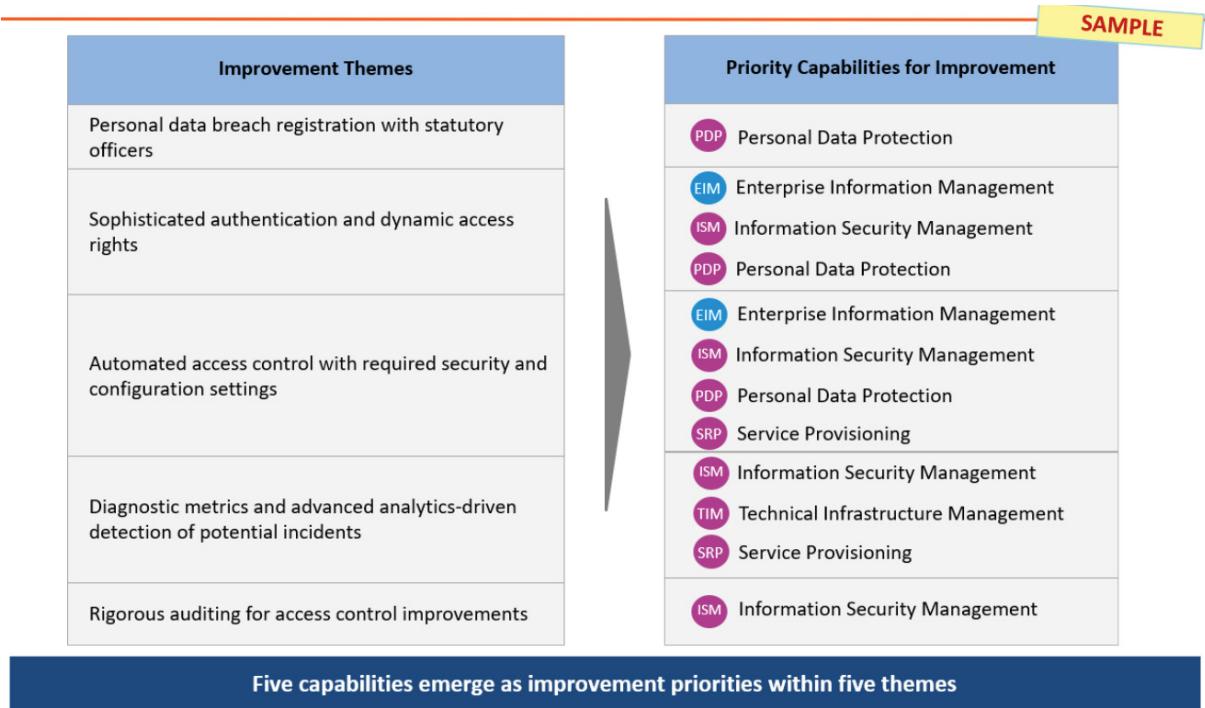
organisation towards developing the CCs necessary to support effective IT management in the digital context. It is important to note that, as with all digital transformation activities, there is a requirement to address not only IT-specific areas of the organisation for improvement activity but to ensure alignment and integration with business units on any approaches adopted. As is evident in the list of CCs in IT-CMF™, and in the core management themes within the CEA there is an increased focus on all aspects of management, inclusive of IT when adopting an approach to managing cybersecurity effectiveness.



**Figure 2:** Current versus target level of cybersecurity management maturity



**Figure 3:** Importance of cybersecurity management themes versus maturity gaps



**Figure 4:** CEA improvement priorities



**Figure 5:** IT-CMFTM

To enable improvements in the prioritised areas or capabilities, IT-CMF includes a wide range of material and tools to support improvement, as follows:

- **Capability Building Blocks (CBBs):** Each capability in IT-CMF™ is broken down into a series of building blocks which delineate the key components to focus on in order to holistically improve. Five-level maturity profiles support high-level CC assessments and in-depth CBB assessments of actual current maturity and desired target maturity.
- **Practices, Outcomes, and Metrics:** For each capability of IT-CMF™ there are representative *practices* that define the organisation's current maturity and how to progress to the next level of maturity. Each *practice* is accompanied by an *outcome* that states what benefits might result by following the *practice*. To help the organisation to measure how successful they are in achieving the *outcome*, one or more *metrics* are provided.

- *Capability Performance Indicators (CPIs)*: CPIs are directly related to each capability and are designed to make a connection between goals, improvement targets, and business outcomes. They are used to understand the organisation's progress towards expected outcomes. The CPIs are grouped into balance scorecard segments (i.e. financial, process, customer, learning and growth) to provide an overview of the target capability improvement.
- *Skills mapping*: To complement capability improvement, organisations need to identify the skills required to deliver the plan and also need to harness and coordinate those skills. Mapping of IT-CMF to industry standard skills frameworks provides structured guidance on key skills and competences required for the full range of IT practices at different levels.

## **5. Conclusions**

Organisations require cybersecurity effectiveness in order to protect against data theft, destruction, and unauthorised access, comply with legal and regulatory requirements, maintain visibility of the evolving threat landscape, and ensure effective management of actual cybersecurity incidents. However, they are often impeded due to the absence of strategy, standards, policies, and controls, limitations in the security architecture and cybersecurity management technologies and tools, cultural issues, and resource constraints, among other factors. These barriers result in the ongoing proliferation of high-profile security breaches, with many organisations still unable to detect when or where their systems have been breached.

This paper has elucidated a number of factors critical for cybersecurity effectiveness based on an in-depth review of relevant literature and empirical insights gained from adopting an open innovation approach with a workgroup of academic and industry subject matter experts. As such, it offers a number of theoretical and practical contributions. From a theoretical perspective, it presents a conceptual framework depicting eight factors or management themes that underpin cybersecurity effectiveness within the organisation. From a practical perspective, operationalisation of the conceptual framework through the CEA instrument enables the organisation to measure its cybersecurity effectiveness – holistically understand its key strengths and weaknesses in securing its IT assets, identify capability gaps in delivering effective cybersecurity, and identify priority areas to improve. The insights gained from the CEA serve as the basis for the organisation to understand what change it must effect in order to establish effective cybersecurity controls that evolve with the changing threat landscape.

This serves as the foundation for initiating the organisation's cybersecurity improvement roadmap, thereby enabling the organisation to support the culture of change and develop the structures required to analyse and address continually changing security considerations, and take the necessary steps to protect their resources proportionate to their organisational value. This research acts as a basis for a more in-depth development and expansion of the cybersecurity framework. As such, practitioners and academics interested in the area are asked to validate this framework to help determine its relevancy in a real-life industry setting.

### **5.1 Limitations and future research**

This paper was limited due to word-count. Further detailed discussion on the conceptual model is planned in future publications. This research was also restricted in that only secondary sources of data (academic journal articles, practitioner reports etc.) and insights from a small sample of subject matter experts were used in the development of the model due to time-constraints. However, further empirical testing is planned for the future with firms who are interested in cybersecurity in the digital context. As such, refinements to the model are expected.

## **References**

- Accenture. (2018). Gaining ground on the cyber attacker. 2018 state of cyber resilience. *Accenture*.
- Arend, C., Sundby, N., and Venkatraman, A. (2016). Reinventing data protection fit for digital transformation. *IDC*.
- Bradley, J., Loucks, J., McCaulay, J., Noronha, A., and Wade, M. (2015). Digital vortex - how digital disruption is redefining industries. *Global Centre for Digital Business Transformation*.
- Bughin, J., LaBerge, L., and Mellbye, A. (2017). The case for digital reinvention. *McKinsey Quarterly*.
- Chesbrough, H. (2003). *Open innovation: the new imperative for creating and profiting from technology*. Harvard Business Review Press.
- Curley, M., Kenneally, J., and Carcary, M. (eds) (2016). *IT Capability Maturity Framework (IT-CMF) – the body of knowledge guide*. 2<sup>nd</sup> edition. Van Haren.

- Curley, M., Kenneally, J., Carcary, M., and Kavanagh, D. (eds) (2017). *IT-CMF – a management guide*. Van Haren.
- Dahlström, P., Ericson, L., Khanna, S., and Meffert, J. (2017). From disrupted to disrupter: reinventing your business by transforming the core. *McKinsey*.
- Ernst & Young. (2018). Is cybersecurity about more than protection? EY's global information security survey 2018-19. *Ernst & Young*.
- European Parliament. (2016a). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC. [Online] Available: <<http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1489407324510&uri=CELEX:32016R0679>>.
- European Parliament. (2016b). Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA. [Online] Available: <<http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1489407324510&uri=CELEX:32016L0680>>.
- Fichman, R., Santos, B., and Zheng, E. (2014). Digital innovation as a fundamental and powerful concept in the Information Systems curriculum. *MIS Quarterly*, vol 38, no 2. 329-353.
- Fraser, J., Simkins, B., and Narvaez, K. (2014). *Implementing enterprise risk management: case studies and best practices*. Hoboken, NJ: Wiley.
- Hussain, Q., Chang, E., Hussain, F., and Dillon, T. (2007a). Ascertaining risk in financial terms in digital business ecosystem environments. *Proceedings of the 2007 Inaugural IEEE International Conference on Digital Ecosystems and Technologies*.
- Hussain, Q., Chang, E., Hussain, F., and Dillon, T. (2007b). Quantifying failure for risk-based decision-making in digital business ecosystem interactions. *Proceedings of the 2nd International Conference on Internet and Web Applications and Services*.
- ISACA. (2012). COBIT 5 for information security. [Online] Available: <<http://www.isaca.org/cobit/pages/info-sec.aspx>>.
- ISACA. (2013). COBIT 5 for risk. [Online] Available: <<http://www.isaca.org/cobit/pages/risk-product-page.aspx>>.
- International Organisation for Standardisation (ISO). (2009). ISO 31000 – Risk management. [Online] Available: <<http://www.iso.org/iso/home/standards/iso31000.htm>>.
- International Organisation for Standardisation (ISO). (2013a). ISO/IEC 27001: 2013. Information technology security techniques – information security management systems requirements. [Online] Available: <[http://www.iso.org/iso/catalogue\\_detail?csnumber=54534](http://www.iso.org/iso/catalogue_detail?csnumber=54534)>.
- International Organisation for Standardisation (ISO). (2013b). ISO/IEC 27002: 2013. Information technology – security techniques – code of practices for information security controls. [Online] Available: <[http://www.iso.org/iso/catalogue\\_detail?csnumber=54533](http://www.iso.org/iso/catalogue_detail?csnumber=54533)>.
- Lee, O. and Baby, D. (2013). Managing dynamic risk in global IT projects: agile risk management using the principles of service-oriented architecture. *International Journal of Information Technology and Decision Making*, vol. 12, no.6, 1121-1150.
- National Institute of Standards and Technology. (2014). Framework for improving critical infrastructure cybersecurity, Version 1.0. [Online] Available: <<https://www.nist.gov/sites/default/files/documents/cyberframework/cybersecurity-framework-021214.pdf>>.
- Rickards, T., Smaje, K., and Sohori, V. (2015). Transformer in chief: The new chief digital officer. *McKinsey*.
- Ross, J.W., Sebastian, I.M., Beath, C., Mocker, M., Moloney, K., and Fonstad, N. (2016). Designing and executing digital strategies. *Proceedings of the 37th International Conference on Information Systems*. December 11-14, Dublin.
- Sambamurthy, V. and Zmud, R. (2012). *Guiding the digital transformation of organisations*. Legerity Digital Press.
- Sing, A.N., Gupta, M.P., and Ojha, A. (2014). Identifying factors of organisational information security management, *Journal of Enterprise Information Management Decision*, vol. 27, no.5, 644-667.
- Stanske, S. and Kautz, K. (2018). Legitimising digital transformation: from system integration to platformisation, *Proceedings of the 39th International Conference on Information Systems*, December 13-16, San Francisco.
- The Open Group. (2011). Open information security management maturity model (O-ISM3). [Online] Available: <<https://www2.opengroup.org/ogsyst/jsp/publications/PublicationDetails.jsp?publicationid=12238>>.
- Webster, J. and Watson, R.T. (2002). Analysing the past to prepare for the future: writing a literature review, *MIS Quarterly*, vol. 26, no. 2, 13-23.

# **Personal Data Protection (PDP): A Conceptual Framework for Organisational Management of Personal Data in the Digital Context**

**Marian Carcary, Eileen Doherty and Gerry Conway**  
**Innovation Value Institute, Maynooth University, Ireland**  
[Marian.carcary@mu.ie](mailto:Marian.carcary@mu.ie)

**Abstract:** Leveraging insights from the personal data of customers can improve the decision-making capability of the organisation resulting in optimised operations, products, and services. However, in the digitally connected world the organisation is increasingly challenged to protect personal data and there is greater potential for inappropriate data use or disclosure, which can result in legal, financial, and reputational consequences for the organisation. Developments brought about by digital transformation increase the requirement for more stringent approaches to personal data protection to be implemented by the organisation. In most countries, data protection is taken very seriously and enforced through strict regulations (e.g. GDPR) which every organisation who holds personal data must adhere to. However, high-profile data breaches continue to occur. Research suggests that data protection controls have not kept pace with the degree to which organisations are experimenting with digital technologies and the unprecedented data volumes. Many organisations are unable to detect when or where their data systems have been breached. In order to comply with regulatory requirements to protect personal data and avoid potentially significant legal, financial, and reputational implications of a data breach, many organisations must improve their personal data protection approaches. This paper presents the key components or 'capability building blocks' of a conceptual framework in the area of Personal Data Protection (PDP) within the organisational context. This framework was developed based on the findings or themes that emerged from a systematic literature review (SLR) in this area followed by an open innovation approach. The resultant conceptual framework can be used by organisations to undertake a holistic analysis of their personal data protection capability. The framework also includes a set of POMs (Practices, Outcomes and Metrics) which acts as a roadmap for organisations to improve upon their current level of maturity in this area to effectively protect personal data and to demonstrate that the organisation is a trustworthy data custodian.

**Keywords:** personal data protection, data protection capability, data protection assessment, GDPR

---

## **1. Introduction**

Data is seen as central to the organisation's digital transformation journey (Carcary, 2018; Arend et al, 2016) and through leveraging insights from the personal data of customers, the organisation is in a position to make decisions to optimise its operations, products, and services, and thereby more effectively compete, and grow its market share and revenue streams (Carcary, 2018; Arend et al, 2016). However, in the digitally connected world, the organisation faces increasing challenges to ensure this personal data is protected due to the unprecedented scale of data collection, 'store everything' practices, and the seamless flow and processing of data across various platforms and applications. This increases the potential for inappropriate or illegal data use or disclosure (Carcary, 2018; Arend et al, 2016; Accenture, 2016; EU, 2016a)

Over the last number of years, people have become increasingly aware of their rights under data protection legislation - there is an expectation that organisations with which they share their data have adequate data protection policies and practices in place. People who seek legal reparation for inappropriate disclosure of their personal data are successful in approximately half of all legal actions (Black, 2013; Howard and Gulyas, 2014). Therefore, effectively protecting personal data and proving that the organisation is a trustworthy data curator is critical to the organisation's brand and competitiveness (Accenture, 2014; Brown, 2014; LeHong et al, 2013). However, research suggests that data protection controls have not kept abreast of the level at which organisations are experimenting with digital technologies and the unprecedented level of data (Carcary, 2018; Arend et al, 2016; Brown, 2014). In a recent Ernst & Young survey, 37% of all respondent organisations indicated that they did not have a data protection programme or only had ad hoc policies or processes in place (Ernst and Young, 2015). In a further study, data loss or damage was the primary concern for 41% of enterprise security professionals (Accenture, 2016). In addition, many organisations indicated that they were unable to identify when or where their data systems had been breached. This finding is echoed by figures which show that the FBI, banks, and credit card companies notify thousands of businesses each year that their data systems have been compromised (Carcary, 2018; Romanosky et al, 2014).

In most jurisdictions/countries, data protection is treated very seriously; evidenced through new regulations enforcing stricter obligations and responsibilities on the organisation to effectively protect the personal data it

handles, and imposing penalties for infringement of these regulations. Examples of this include the General Data Protection Regulation (GDPR) of the European Union (EU) which is applicable across the EU since 2018 (EU, 2016a; EU, 2016b), and global industry standards such as the Payment Card Industry Data Security Standard (PCI, 2016) which is now compulsory when dealing with payment card data. Additional regulations continue to emerge. For example, in 2017 a proposal for a new European regulation on privacy and electronic communications was announced to address the rapid evolution of IT products and services (EU, 2017).

However, the developments brought about by digital transformation continue to increase the need for more stringent approaches to personal data protection within the organisational context. In order to comply with requirements and avoid potentially severe legal, financial, and reputational implications of a personal data breach, many organisations need to improve their personal data protection approaches (Arend et al, 2016). By developing the organisation's personal data protection capability, the organisation can reduce its exposure to the risks associated with handling personal data, while setting the foundations for its compliance with relevant legislation and ultimately enhancing its reputation (Carcary, 2018).

This paper presents the key components or 'capability building blocks' of a conceptual framework in the area of Personal Data Protection (PDP) within the organisational context. This framework includes an assessment instrument and set of POMs (Practices, Outcomes and Metrics) which enable the organisation to improve on its current level of maturity to effectively protect personal data and to demonstrate that the organisation is a trustworthy data custodian.

## **2. Research methodology**

The Personal Data Protection (PDP) Critical Capability was developed through adopting a multi-method research approach based on (in phase 1) a synthesis of the personal data protection capability themes, derived through a systematic literature review approach, followed by (in phase 2) an open innovation approach, garnering expert input from subject matter experts in the area. This multi-method approach is outlined in greater detail below:

Phase 1: A systematic literature review was initially undertaken, focused on the key requirements for protecting personal data in the digital context. The literature review covered both academic journals, IT industry publications, and legislative standards. Due to the evolving nature of the digital landscape, every effort was made to ensure that these publications were up-to-date. A content analysis of the material extracted from the literature was undertaken to establish the most common concepts. The authors followed the concept matrix method (Webster and Watson, 2002) - the matrix rows provide the paper references from which the concepts were extracted, while frequency of occurrence of a particular theme is indicated by the number of 'Xs' in the table columns. This paper will not present the systematic literature review in detail due to word constraints. However, the key themes that emerged from this SLR are presented in section 3.1.

Phase 2: Building on the concepts identified through the systematic literature review process, subject matter experts from both industry and academia worked collaboratively in a shared interest workgroup to create a conceptual framework for developing a personal data protection capability, based on the principles of open innovation (Chesbrough, 2003). These principles supported the leveraging of resources, knowledge, and expertise from multiple stakeholders whereby engaged participants could communicate, exchange views, and develop a research output within a collaborative research environment. The value of adopting an open innovation approach, centred on both the clear articulation of challenges by practitioners who worked in the personal data protection context, and the grounding of the framework's development in these insights. A rigorous review of the developed PDP framework was also undertaken within the organisational context by four personal data protection industry practitioners. Feedback from this review process was incorporated, resulting in the development of a PDP Critical Capability that had increased practical utility in the industry context.

The conceptual framework for personal data protection discussed in this paper is positioned as part of an industry-standard IT management framework called the IT Capability Maturity Framework™ (IT-CMF™). IT-CMF™ reflects a compilation of 37 Critical Capabilities (see figure 1) which are key for the successful management of IT and is designed to systematically lead an organisation towards developing the Critical Capabilities necessary to support effective IT management in the digital context. It does this through systematically and continually improving the performance of the IT function in an organisation, and through measuring progress and value delivered. Each Critical Capability is made up of high-level Categories and

Capability Building Blocks. For each Critical Capability, the framework outlines 5 levels of maturity, where level 1 indicates a low level of maturity and level 5 indicates an example of best practice (Curley et al, 2015).



**Figure 1:** IT-CMF™

### 3. Findings and discussion

Section 3.1 below discusses the findings/themes that have emerged from the systematic literature review. Section 3.2 presents the conceptual model for PDP.

#### 3.1 Themes identified from systematic literature review and concept matrix

In order to better understand the personal data protection phenomenon, this study adopted a concept-centric examination of the extant literature, focused on the key requirements for protecting personal data in the digital context. The key literature insights are now discussed.

Data protection regulations are in place in order to protect the individual's fundamental rights in the protection of their personal data (Carcary, 2018; EU, 2016a). In order for organisations to be able to demonstrate compliance with such data protection regulations, it is critical that personal data protection is acknowledged as a Board level responsibility and priority (ISO, 2015). Driven by the most senior executive levels, the organisation may need to rethink its personal data protection strategy (Carcary, 2018; Arend et al, 2016), and should reflect privacy, trust, and security as the underpinning principles of the organisation's digital strategy and the organisation's brand (Accenture, 2014; LeHong et al, 2013). The organisation needs to develop and implement internal data protection policies and approved codes of conduct to guide its personal data protection efforts in line with its established strategies (EU, 2016a). Responsibilities and accountabilities must be assigned (Carcary, 2018; Accenture, 2016; Ernst and Young 2015), and organisational resources must be deployed to ensure data protection compliance.

In some cases, allocation of responsibility for data protection in the form of a designated officer is required to lead the organisation's data protection efforts (EU, 2016a; Druva, 2016). A good example of this is where data processing is performed by a public body or where data processing operations that require frequent, large-scale monitoring of data subjects or large scale processing of certain categories of personal data reflect the core activities of the organisation (Carcary, 2018; EU, 2016a). This role necessitates effective collaboration across the organisation, and success depends on the support of the senior management team (Druva, 2016). Responsibilities include informing stakeholders of their obligations in relation to data protection regulations, monitoring compliance with the regulations and data protection policies, and liaising/cooperating with relevant data protection supervisory authorities/public bodies (Carcary, 2018; EU, 2016a). Other responsibilities include development of a security aware culture, where heightened levels of awareness are necessary to keep abreast of the various threats and to develop new ways to respond to insights gained from known security breaches

(Accenture, 2013). He/she will drive awareness-raising activities and training to support the interpretation of the principles of data protection regulations for those involved in personal data processing<sup>1</sup> operations and audits (Carcary, 2018).

The basis upon which any data protection regulation is developed is the rights of the data subject. Therefore, the organisation has a responsibility to be mindful of protecting these rights in its day-to-day operations. Firstly, the organisation<sup>2</sup> is obliged to explicitly communicate to the data subject at the time of data collection the purposes for which his/her personal data is collected, the proposed processing of this data, and the associated potential risks and consequences. In addition, the data controller's contact details, the recipients of the personal data, the period for which it will be stored, and any information regarding the transfer of the personal data across multiple jurisdictions must also be clearly communicated at the outset (Carcary, 2018; EU, 2016a). Informed and unambiguous consent to the proposed data processing must be acquired from the data subject either through written or oral communication or by electronic means, and the organisation must respect that the data subject has the right to withdraw his/her consent at any time (EU, 2016a; EU, 2014). If the organisation wishes to process this personal data for additional purposes, these further purposes must also be communicated and the data subject's consent is necessary prior to such processing (Carcary, 2018; EU, 2016a).

The organisation has a responsibility to ensure that any processing of personal data is lawful, fair, and transparent. It should only be processed based on the specified and legitimate purposes for which it was obtained, and the data stored should be adequate and limited to only that which is required to fulfil the specified purposes (i.e. data minimisation). The organisation has a responsibility to maintain an accurate and up-to-date record of this personal data, and any inaccurate personal data should be deleted or remedied immediately. The organisation should also respect a data subject's right to request access to his/her personal data in the organisation' custody and the data subject's 'right to be forgotten'. This requires the removal of personal data when, for example, the data is no longer required for the purposes for which it was obtained, when data subject consent has been withdrawn, or when personal data has been unlawfully processed (Carcary, 2018; EU, 2016a). Where transfer of the personal data to other countries and international organisations needs to be facilitated, any lack of data protection in the foreign country should be compensated for by adequate safeguards. These may include contractual stipulations with the foreign recipient, binding corporate rules, and standard data protection clauses by a specific jurisdiction or supervisory authority including enforceable data subject rights and effective legal remedies (Carcary, 2018; EU, 2016a; EU, 2014).

The organisation should evaluate the likelihood and severity of risks to the rights of data subjects arising from any data processing activity. Where data processing is likely to result in a high risk to data subject rights (e.g. in instances of large scale processing of certain categories of personal data), a privacy impact assessment should be undertaken (Carcary, 2018). This is a systematic process to assess the nature and level of severity of the risks to data privacy from collection to destruction, and to inform the types of measures and safeguards to be implemented to mitigate against those risks. The organisation may also consider the potential damage to its own reputation if data is handled inappropriately. If the severity of the risks cannot be mitigated against using appropriate measures, the relevant data protection supervisory authority/public body should be consulted prior to any data processing (Carcary, 2018; EU, 2016a; ICO, 2011; ICO, 2014).

It is imperative that the personal data held by the organisation safeguards against unauthorised access or disclosure, unauthorised or unlawful processing, and accidental loss, destruction, or damage (Carcary, 2018) Therefore, the organisation must implement appropriate technical and organisational measures to ensure personal data is effectively protected in alignment with regulatory and legislative requirements. In so doing, the organisation should embrace the concept of privacy/data protection by design and by default, which involves considering and integrating appropriate safeguards both when determining the means by which personal data will be processed and during actual data processing itself. Developers of products, services, and applications

---

<sup>1</sup> Personal data processing is '*any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organization, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction'* [3].

<sup>2</sup> The reader should be cognizant of the terms 'data controller' and 'data processor' that are used in data protection regulations. A data controller is any legal person that either '*alone or jointly with others, determines the purposes and means of the processing of personal data*'. A data processor is any legal person that '*processes personal data on behalf of the controller*' [3]. The data processor processes data in line with conditions specified in a written contract agreed with the data controller.

that process personal data should consider the right to data protection during their design and development (Carcary, 2018; EU, 2016a; ICO, 2014).

In order to ensure an appropriate level of security is afforded to personal data, the organisation needs to establish a solid foundation of security measures to provide basic defence (Ernst and Young, 2014). Security protection needs to be rebalanced from ‘network centric’ to ‘data centric’ (CapGemini, 2016). Specific measures to secure and keep data confidential include, for example, strategies to anonymise/de-identify the personal data held in the organisation’s custody (e.g. ‘pseudonymization’<sup>3</sup> or other data anonymisation strategies) (EU, 2016a; ICO, 2012). Other measures include encryption of data held centrally and on mobile devices, and the capacity to delete the data held on a mobile device via a remotely issued command in the event of its loss or theft (Druva, 2016). Further measures are required to protect the availability, integrity, and resilience of processing systems and services, and to quickly restore availability and access to personal data in the event of an incident (EU, 2016a). Data protection solutions should reflect an analytics-led, adaptive approach to recovery and backup that evolves with changing business needs (Carcary, 2018; Arend et al, 2016). The effectiveness of both the technical and organisational measures should be regularly tested and evaluated vis-à-vis the risks associated with data processing, and any personal data breaches should be communicated to the appropriate data protection supervisory authority/public body without undue delay (EU, 2016a). In addition, the nature of any breach and recommendations to mitigate potential adverse consequences should be communicated to the data subject (Carcary, 2018; EU, 2016a; BakerHostetler, 2015) in a timely manner.

Three high-level concepts/categories emerged from the systematic literature review. Within these high-level categories, further sub-concepts or capability building blocks (CBBs) also emerged. These are presented in the form of the PDP conceptual model in Table 1.

### **3.2 Conceptual framework**

The Personal Data Protection (PDP) Critical Capability developed is defined as “*the ability to develop and deploy policies, systems, and controls for processing personal and sensitive personal data relating to living persons in all digital, automated, and manual forms. It ensures that the organization safeguards the right to privacy of individuals whose information it holds, and that the organization uses personal data strictly for specified purposes agreed with the data subjects*”.

Its key aims are to enable an organisation to:

- Comply with relevant data protection regulations.
- Manage the growing complexities of protecting personal data in the digital business context.
- Develop and deploy data protection policies, systems, and controls for appropriate acquisition, use, retention, and deletion/destruction of personal data.
- Verify the effectiveness of data protection policies, systems, and controls.
- Proactively identify and address any data protection issues.
- Manage timely communication and registration with statutory officers regarding data protection breaches and near incidents.
- Develop, test, and deploy incident management processes and procedures.
- Leverage valuable insights from personal data to enhance the organization’s operations without compromising data protection regulatory compliance.
- Increase stakeholder confidence that the organization can be regarded as a trustworthy custodian of personal data.

The conceptual framework for the PDP Critical Capability is decomposed into three manageable areas of focus or overarching categories namely: 1) Governance, Management, and Oversight; 2) People; and 3) Processing; and 12 sub-categories or Capability Building Blocks. This conceptual framework is presented in Table 1 below.

---

<sup>3</sup> Pseudonymization is ‘*the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person*

**Table 1:** PDP capability building blocks

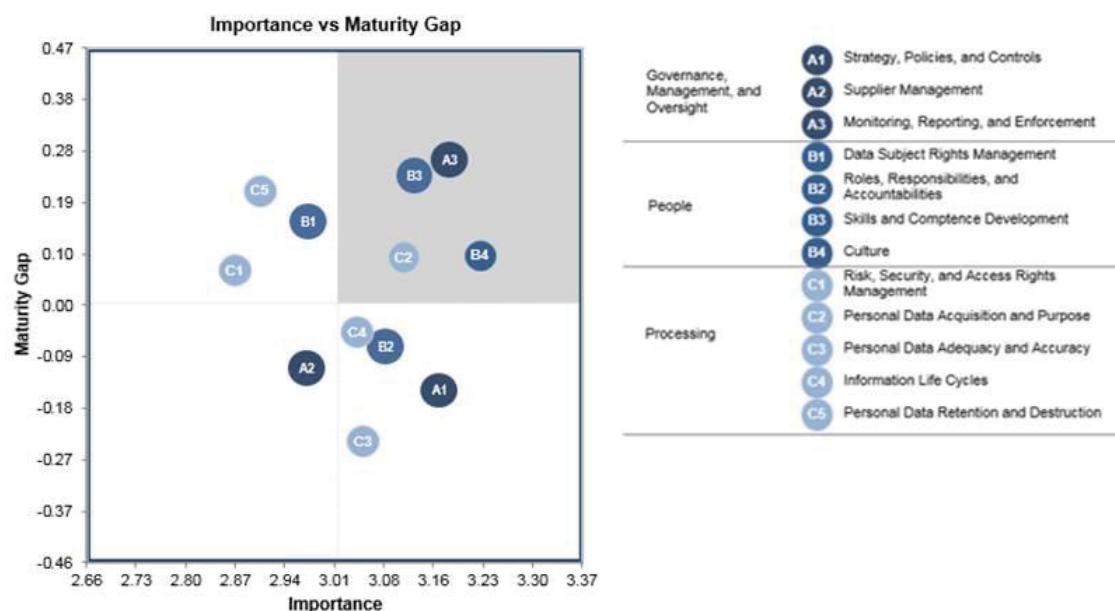
Category	Capability building block	Description
Governance, Management, and Oversight (A)	Strategy, Policies, and Controls (A1)	Establish a strategy for protecting personal data. Design, develop, and maintain personal data protection policies and controls that comply with relevant data protection standards, regulations, and laws, and that align with the organization's business model and objectives. Promote and drive personal data protection compliance.
	Supplier Management (A2)	Define personal data protection qualification criteria for identifying and validating suppliers, and select suppliers who are committed to observing the organization's personal data protection obligations. Draft and agree the data processor contract, and manage contract compliance with the suppliers.
	Monitoring, Reporting, and Enforcement (A3)	Establish appropriate measures for enforcing compliance and monitoring and reporting non-compliance with personal data protection policies, and for taking remedial action where necessary. Drive improvements based on lessons learned from incidents (e.g. data breaches and inappropriate or unauthorized data access) and near-incidents.
People (B)	Data Subject Rights Management (B1)	Manage requests by data subjects to access their personal data held by the organization (including the purposes for which it is held and to whom it may be disclosed), and to rectify or erase inaccurate data. Check that the communication channels and agents are authorized by the data subject.
	Roles, Responsibilities, and Accountabilities (B2)	Complete job and business process designs to identify the required roles for personal data protection tasks, and assign employees with the requisite knowledge and experience to the identified roles. Define and allocate the associated personal data protection responsibilities and accountabilities.
	Skills and Competence Development (B3)	Establish and make available a personal data protection training curriculum and other employee developmental mechanisms to ensure employees have the required skills and competences.
	Culture (B4)	Establish a personal data protection-aware culture. Inform stakeholders of key developments to build a shared understanding of how they can contribute to the realisation of personal data protection objectives.
Processing (C)	Risk, Security, and Access Rights Management (C1)	Establish and communicate personal data risk criteria, security criteria, and access rights controls (based on the life cycle state).
	Personal Data Acquisition and Purpose (C2)	Establish approaches to obtain data subject consent, provide fair notice, and manage the acquisition and lawful, fair, and transparent processing of personal data for explicit and legitimate purposes.
	Personal Data Adequacy and Accuracy (C3)	Ensure that personal data is only used and disclosed in line with the purposes for which it was acquired, and that the data held is adequate, relevant, and limited to what is necessary to meet those purposes. Monitor the quality of personal data held and remedy any data quality issues.
	Information Life Cycles (C4)	Provide input to information life cycle planning to identify, acquire, process, store, and/or destroy personal data in line with business, regulatory, and legal requirements and risks. Conduct privacy impact assessments at the planning stage of new or large change projects, and consider the potential damage or harm to both the data subject and the organization in whose custody the information has been placed.
	Personal Data Retention and Destruction (C5)	Develop and implement controls to verify that personal data is not retained beyond the time specified in data retention policies. Destroy data media (all forms – paper, digital, DNA encoded etc.) at the end of the data's life cycle and ensure that obsolete (or deleted) personal data is not inappropriately restored.

In order to effectively operationalise this PDP Critical Capability, a five-level maturity profile is defined at a Critical Capability level. A summary view of this is outlined in Table 2 below:

**Table 2:** PDP critical capability summary maturity profile

Personal Data Protection Maturity Profile	
5 Optimized	<ul style="list-style-type: none"> <li>Flexible and agile personal data protection policies, procedures, and controls are continually updated in line with the latest data protection standards, regulations, and laws, and with evolving business needs.</li> <li>There is a culture of proactive personal data protection across key business ecosystem partners. Non-compliance is rare and the organization may be regarded as an industry leader in driving personal data protection compliance.</li> </ul>
4 Advanced	<ul style="list-style-type: none"> <li>Personal data protection policies, procedures, and controls are updated in line with evolving risk, technical factors, and business objectives, and application of them is automated and enforced across the organization.</li> <li>The organization has a strong reputation for managing personal data. Instances of non-compliance increasingly give rise to proactive remedial actions.</li> </ul>
3 Intermediate	<ul style="list-style-type: none"> <li>Most areas of the business agree on and enforce a 'universal' set of personal data protection policies, procedures, and controls. These are informed by relevant data protection standards, regulations, and laws, and the business's data protection objectives.</li> <li>The organization's reputation for managing personal data is growing. Instances of non-compliance are increasingly used for retrospective remedial improvements.</li> </ul>
2 Basic	<ul style="list-style-type: none"> <li>Some basic personal data protection policies, procedures, and controls are developed to meet high-priority legislative and regulatory requirements.</li> <li>The risks associated with holding personal data have begun to reduce. There is growing awareness of any non-compliance.</li> </ul>
1 Initial	<ul style="list-style-type: none"> <li>Any personal data protection policies, procedures, and controls are ad hoc.</li> <li>There is limited or no understanding of personal data protection obligations and any enforcement is ad hoc. Advocacy of personal data protection, if any, is sporadic.</li> </ul>

This maturity profile serves as the basis for a maturity assessment instrument which can be used by organisations to holistically analyse their performance in respect of their personal data protection capability – i.e. their 'current' capability maturity and their 'target' capability maturity. Figure 2 provides sample insights that can be derived through undertaking this maturity assessment – this figure outlines the organisation's importance rating vis-à-vis their capability maturity gap, which provides key insights into target areas for improvement.



**Figure 2:** PDP critical capability – capability building block importance versus maturity gap

In terms of enabling improvements in the prioritised areas, Table 3 provides a summary view of Practices, Outcomes, and Metrics (POMs) developed. A detailed set of POMs also exists for each Capability Building Block at each level of maturity.

**Table 3:** PDP critical capability POMs

Maturity	Key practices	Outcomes	Metrics
5 <b>Optimized</b>	<ul style="list-style-type: none"> <li>Keep up to date with the latest research on the protection of personal data, and implement best known practice.</li> <li>Continually encourage relevant business ecosystem partners to adopt good personal data protection practices.</li> </ul>	<ul style="list-style-type: none"> <li>The organization is effective in preventing data breaches. Claims for breach of trust, or duty of care are less likely to succeed.</li> <li>There is reduced risk of legal action or reputational damage arising from work with business ecosystem partners.</li> </ul>	<ul style="list-style-type: none"> <li># of data protection research initiatives and industry collaborations being pursued or investigated.</li> <li># of and trends of incidents in relevant business ecosystem partners.</li> </ul>
4 <b>Advanced</b>	<ul style="list-style-type: none"> <li>Mandate privacy impact analysis in all system reviews, and programme, project, and change management processes throughout the organization.</li> <li>Consistently adhere to personal data retention and destruction policies.</li> </ul>	<ul style="list-style-type: none"> <li>A privacy impact analysis identifies ways of preventing personal data protection issues from arising.</li> <li>Retention and destruction of personal data is policy and process compliant across the organization.</li> </ul>	<ul style="list-style-type: none"> <li># of and trends for potential issues identified and averted or mitigated using privacy impact analyses.</li> <li># of personal data retention non-compliance issues identified (per time period).</li> </ul>
3 <b>Intermediate</b>	<ul style="list-style-type: none"> <li>Identify relevant data protection standards, regulations, and legislative requirements.</li> <li>Encourage consistent adoption of personal data protection policies, procedures, controls, and tools across employees and external partners.</li> <li>Audit the effectiveness of the personal data protection approaches.</li> </ul>	<ul style="list-style-type: none"> <li>Relevant standards, regulations, and legislative requirements can inform the approaches to personal data protection.</li> <li>Consistent procedures and controls enable easier detection of anomalies.</li> <li>Issues identified in audits help to improve processes, and identify areas where automation or training might be of value.</li> </ul>	<ul style="list-style-type: none"> <li>% of identified personal data protection standards, regulations, and legislative requirements reflected in policies and procedures.</li> <li># of employee-related data protection incidents and compliance issues.</li> <li># of issues detected in audits and time to closure for those issues.</li> </ul>
2 <b>Basic</b>	<ul style="list-style-type: none"> <li>Provide job-specific personal data protection training.</li> <li>Allocate roles and responsibilities for personal data protection.</li> <li>Document approaches to ensure that personal data is used only for the purposes for which it was collected.</li> </ul>	<ul style="list-style-type: none"> <li>Employee understanding of the need to safeguard personal data grows, which reduces the risk of careless disclosure.</li> <li>Responsibilities are transparent, enabling effective data protection activities.</li> <li>Personal data is used only for appropriate and compliant purposes.</li> </ul>	<ul style="list-style-type: none"> <li>% of employees with data protection training.</li> <li>% of data protection roles filled in key functions.</li> <li># of violations in the use of personal data (per time period).</li> </ul>
1 <b>Initial</b>	• ...	• ...	• ...

#### 4. Conclusion

The effective protection of personal data is a critical factor in maintaining a positive reputation and avoiding financial and legal consequences that can threaten the organization's survival. However, personal data protection is becoming increasingly complex. As stated by Cearley et al, '*the velocity and density of information in digital business adds new risks concerning data protection, complicated by cultural privacy issues, and in some cases, government regulations. This is made even more complicated in a world where computing is everywhere, control of those systems is incomplete, and the perimeter is almost non-existent*' (Cearley et al, 2014). In the

digital landscape, a rethink of the organization's data protection strategy is often required to ensure data protection controls keep pace with both rapid technological evolutions and new data protection regulations. Guided by the direction of Board-level executives, clear strategy, policies, and controls, and consistent interpretation and application of the principles of data protection regulations, the organisation can build an effective personal data protection capability for the digital context. The insights gained from evaluating the organisation's personal data protection capability serve as the basis for the organisation to understand 'how effective it is now' and 'what change it needs to effect'. This serves as the foundation for initiating the organisation's personal data protection improvement roadmap in order to drive compliance with relevant regulations and safeguard the organisation from financial and legal implications.

In summary, this paper presents a conceptual framework; inclusive of an assessment instrument and detailed practices, outcomes and metrics (POMs) that can be used by organisations to mature or improve their data protection capability. The paper highlights how the insights gained can be used in a practical sense to effectively protect personal data and to comply with strict regulatory requirements.

This research acts as a starting point and the basis for a more in-depth development and expansion of the Personal Data Protection (PDP) capability. As such, practitioners and academics interested in the area are asked to validate this critical capability to help determine its relevancy in a real-life industry setting.

#### **4.1 Limitations and future research**

This paper was limited due to word-count. Further detailed discussion on the conceptual model is planned in future publications. This research was also restricted in that only secondary sources of data (academic journal articles, practitioner reports etc.) and insights from a small sample of subject matter experts were used in the development of the model due to time-constraints. However, further empirical testing is planned for the future with firms who are interested in PDP in a digital context. As such, refinements to the model are expected. The organisations used will not be specific to any particular sector or location.

#### **References**

- Accenture, (2013) 'Accenture technology vision 2013. Every business is a digital business'. Online at: <[https://www.accenture.com/us-en/\\_acnmedia/Accenture/Conversion-Assets/Microsites/Documents8/Accenture-Technology-Vision-2013.pdf](https://www.accenture.com/us-en/_acnmedia/Accenture/Conversion-Assets/Microsites/Documents8/Accenture-Technology-Vision-2013.pdf)>.
- Accenture, (2014) 'Accenture technology vision 2014. Every business is a digital business - from digitally disrupted to digital disrupter'. Online at: <<http://investor.accenture.com/~media/Files/A/Accenture-IR/events-and-presentations/Accenture-Technology-Vision-2014.pdf>>.
- Accenture, (2016) 'The state of cybersecurity and digital trust 2016 - identifying cybersecurity gaps to rethink state of the art'. Online at <[https://www.accenture.com/t20160704T014005\\_w/\\_us-en/\\_acnmedia/PDF-23/Accenture-State-Cybersecurity-and-Digital-Trust-2016-Report-June.pdf](https://www.accenture.com/t20160704T014005_w/_us-en/_acnmedia/PDF-23/Accenture-State-Cybersecurity-and-Digital-Trust-2016-Report-June.pdf)>.
- Arend, C., Sundby, N., and Venkatraman, A. (2016) 'Reinventing data protection fit for digital transformation', *IDC*. Online at <<https://www.hpe.com/h20195/v2/GetPDF.aspx/4AA6-8166ENW>>.
- BakerHostetler, (2015) '2015 international compendium of data privacy laws'. Online at: <<http://towerwall.com/wp-content/uploads/2016/02/International-Compendium-of-Data-Privacy-Laws.pdf>>.
- Black, J. (2013) 'Developments in data security breach liability', *The Business Lawyer*, vol. 69, no.1, pp199–207.
- Brown, M. (2014) 'Why digital governance and data protection matters', *Computer Weekly*. Online at: <<http://www.computerweekly.com/opinion/Why-digital-governance-and-data-protection-matters>>.
- Carcary, M. (2018) 'Personal Data Protection: Insights in the digital context'. White Paper. Innovation Value Institute, Maynooth University. Online at: <<http://mural.maynoothuniversity.ie/9625/>>.
- CapGemini, (2016) 'Address c-level cybersecurity issues to enable and secure digital transformation'. Online at: <[https://www.capgemini.com/de-de/wp-content/uploads/sites/5/2017/07/1602\\_cybersecurity\\_strategic\\_consulting\\_brochure\\_cc\\_web\\_en\\_1.pdf](https://www.capgemini.com/de-de/wp-content/uploads/sites/5/2017/07/1602_cybersecurity_strategic_consulting_brochure_cc_web_en_1.pdf)>.
- Cearley, D.W., Walker, M.J., and Blosch, M. (2015) 'The top 10 strategic technology trends for 2015', *Gartner*. Online at: <<https://www.gartner.com/doc/2964518/top--strategic-technology-trends>>.
- Chesbrough, H. (2003) 'Open Innovation: The new imperative for creating and profiting from technology', *Harvard Business School Press*, Boston.
- Curley, M., Kenneally, J. and, Carcary, M. (eds) (2015) IT-Capability Maturity Framework (IT-CMF) Body of Knowledge Guide, *Van Haren Publishing*.
- Druva, (2016) '5-step guide for GDPR compliance – a guide for constructing your planning timeline'. Online at: <<http://pages2.druva.com/rs/307-ANG-704/images/Druva-5-Step-Guide-For-GDPR-Compliance.pdf>>.

- Ernst & Young, (2015) 'Creating trust in the digital world - EY's global information security survey 2015', Online at: <[http://www.ey.com/publication/vwluassets/ey-global-information-security-survey-2015/\\$file/ey-global-information-security-survey-2015.pdf](http://www.ey.com/publication/vwluassets/ey-global-information-security-survey-2015/$file/ey-global-information-security-survey-2015.pdf)>.
- European Commission, (2017) 'Proposal for a regulation on privacy and electronic communications'. Online at: <<https://ec.europa.eu/digital-single-market/en/news/proposal-regulation-privacy-and-electronic-communications>>.
- European Parliament, (2016a) 'Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC'. Online at: <<http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1489407324510&uri=CELEX:32016R0679>>.
- European Parliament, (2016b) 'Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA'. Online at: <<http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1489407324510&uri=CELEX:32016L0680>>.
- European Union (2014) 'Handbook on European data protection law'. Online at: <<https://rm.coe.int/16806b294a>>.
- Howard, P.N. and Gulyas, O. (2014) 'Data breaches in Europe: reported breaches of compromised personal records in Europe, 2005–2014'. *Budapest: Center for Media, Data and Society, Central European University*. Online at: <[https://cmds.ceu.edu/sites/cmcs.ceu.hu/files/attachment/article/663/databreachesineurope\\_1.pdf](https://cmds.ceu.edu/sites/cmcs.ceu.hu/files/attachment/article/663/databreachesineurope_1.pdf)>.
- Information Commissioner's Office (2011) 'Data sharing code of practice'. Online at: <[https://ico.org.uk/media/for-organisations/documents/1068/data\\_sharing\\_code\\_of\\_practice.pdf](https://ico.org.uk/media/for-organisations/documents/1068/data_sharing_code_of_practice.pdf)>.
- Information Commissioner's Office (2012) 'Anonymization: managing data protection risk code of practice'. Online at: <<https://ico.org.uk/media/1061/anonymisation-code.pdf>>.
- Information Commissioner's Office (2014) 'Conducting privacy impact assessments code of practice'. Online at: <<https://ico.org.uk/media/for-organisations/documents/1595/pia-code-of-practice.pdf>>.
- International Organization for Standardization (ISO), (2015) 'ISO 38500: 2015. Information technology - Governance of IT for the organization'. Online at: <<https://www.iso.org/standard/62816.html>>.
- LeHong, H., Alvarez, G., and Sussin, J. (2013) 'Ten new realities of customer engagement to account for when developing a digital strategy', *Gartner*. Online at: <<https://www.gartner.com/doc/2536016/new-realities-customer-engagement-account>>.
- Payment Card Industry, (2016) 'Payment Card Industry (PCI) Data Security Standard - requirements and security assessment procedures'. Online at: <[https://www.pcisecuritystandards.org/document\\_library](https://www.pcisecuritystandards.org/document_library)>.
- Pental, S. (2015) 'Five ways information security can help IT improve stakeholder engagement', *CEB IT Quarterly – Spotlight on business engagement*. Q2, pp30-33. Online at: <<http://ceb.uberflip.com/i/502110-cio152185syn-rp-q2-it-quarterly-web/33?m4=>>>.
- Romanosky, S., Hoffman, D., and Acquisti, A. (2014) 'Empirical analysis of data breach litigation', *Journal of Empirical Legal Studies*, vol. 11, no. 1, pp74–104.
- Webster, J. and Watson, R.T. (2002). 'Analyzing the past to prepare for the future: Writing a literature review', *MIS Quarterly*, 26 (2), pp13-23.

# A Capability Approach to Managing Organisational Information Security

**Marian Carcary, Eileen Doherty and Gerry Conway**  
**Innovation Value Institute, Maynooth University, Ireland**  
[Marian.carcary@mu.ie](mailto:Marian.carcary@mu.ie)  
[eileen.doherty@mu.ie](mailto:eileen.doherty@mu.ie)  
[gerard.conway@mu.ie](mailto:gerard.conway@mu.ie)

**Abstract:** Information security is becoming increasingly important for most organisations, as it can add real value by facilitating interaction with trading partners, enabling closer customer relationships, and enabling new and easier ways to process electronic transactions that result in a competitive advantage. However, this enhanced business performance comes with increased risk, for example in 2018, information security breaches totalled 1,244 and affected more than 446 million records (Identity Theft Resource Centre (ITRC), 2018). Due to the sensitive nature of customer data, the recent legislative changes around how data is handled (e.g. GDPR) and the mounting information security risks, it is critical for organisations to have a robust and reliable information security system in place. The information security system and its associated strategies should not just react to information security incidents, but protect the data, and anticipate and seek to prevent attacks from cyber criminals. A robust information security system should incorporate the inventory and monitoring of information, and manage how the data is captured, stored, used, handled, and transmitted internally, in data centres, in the cloud, and across the network. This paper proposes a capability approach for the management of information security that encapsulates the management and control of the integrity, confidentiality, accountability, usability, and availability of information. The paper presents a conceptual model and assessment tool, developed via an open innovation and collaborative research approach that an organization can use to understand and assess the maturity of their information security. The conceptual model uses a holistic and systematic approach and is designed to provide real value to organizations by enabling them to drive improvements in the management of their information security, to maximise the potential benefits and to minimise or alleviate any risks.

**Keywords:** information security (InfoSec), information security management, information security management system (ISMS)

---

## 1. Introduction

Today's business landscape is characterised by the rapid pace of technological change and growing proliferation and reliance on digital technologies. Evolving business models, greater risk taking and experimentation, enhanced organisational connectivity, and increased information velocity and density are also evident (Bradley et al, 2015, Catlin et al, 2015, Fichman et al, 2014, Bharadwaj et al, 2013). All of these changes, together with the growing sophistication of cyber criminals, are key factors for organisations facing an unprecedented number and range of information security attacks<sup>1</sup> (McClamans et al, 2016, Shepherdson et al, 2016). It is also anticipated that as more devices, systems, and infrastructure become interconnected and interdependent as a result of digital transformation, and as more interfaces between customers, suppliers, and partners are leveraged, the IT 'attack surface' will continue to expand (McClamans et al, 2016, PWC, 2015). Referring to the occurrence of security-related intrusions, Sambamurthy and Zmud (2012) outlined that "*the interconnected nature of today's business environment results in ripple effects ... severely affecting organisations distant from (and seemingly unrelated to) the early targets*". The changes brought about by digital transformation make information security management more challenging and necessitate a distinct evolution from traditional thinking. There is growing acceptance in the areas of IT strategy that organisations need to expand their focus beyond solely considering technology in isolation, to include the underlying organisational capabilities necessary to effectively manage and optimise technology use. In fact, there is acknowledgement that adopting a capability approach as opposed to a process-based approach to IT management in general can result in greater value generation for an organisation. An IT capability refers to a "*firm's ability to mobilise and deploy its IT-based resources, creating value in combination with other resources and capabilities, and the firm-specific IT-enabled knowledge and routines that improve the value of non-IT resources*" (Drnevich and Croson, 2013 p485). It is how an organisation effectively manages its capabilities and how these capabilities are targeted and deployed that is critical in deriving organisational performance benefits (Zahra et al, 2006).

---

<sup>1</sup> Threats faced by organisations include, for example, sophisticated malware, cyber sabotage, phishing, man in the middle attacks, denial of service attacks, brute force attacks, zero day attacks, and ransomware attacks.

This paper presents the key components or ‘capability building blocks’ of a conceptual framework or capability for Information Security Management (ISM). The framework presents the key areas of focus and includes an assessment instrument and other artefacts that enable the organisation to ascertain and improve upon its current level of maturity to effectively and robustly protect the security of its information resources.

## 2. Research methodology

The Information Security Management (ISM) capability was developed through adopting a multi-method research approach based on (in phase 1) a synthesis of the information security management capability themes, derived through a systematic literature review, followed by (in phase 2) an open innovation approach, garnering expert input from subject matter experts in the area.

Phase 1: A systematic literature review was initially undertaken, focused on the key requirements for information security management in the digital context. The literature review covered both academic journals, IT industry publications, and legislative standards. A content analysis of the material extracted from the literature was undertaken to establish the most common concepts. The authors followed the concept matrix method (Webster and Watson, 2002).

Phase 2: Building on the concepts identified through the systematic literature review process, subject matter experts from both industry and academia worked collaboratively in a shared interest workgroup to create a conceptual framework for developing an information security management capability, based on the principles of open innovation (Chesbrough, 2003). These principles supported the leveraging of resources, knowledge, and expertise from multiple stakeholders whereby engaged participants could communicate, exchange views, and develop a research output within a collaborative research environment. A rigorous review of the developed ISM framework was also undertaken within the organisational context by four industry practitioners. Feedback from this review process was incorporated, resulting in the conceptualisation of an ISM capability that had increased practical utility in the industry context.

The conceptual framework for information security management presented in this paper is positioned as part of an industry-standard IT management framework called the IT Capability Maturity Framework™ (IT-CMF™). IT-CMF™ reflects a compilation of 37 Capabilities (see figure 1) which are key for the successful management of IT and is designed to systematically lead an organisation towards developing the capabilities necessary to support effective IT management. It does this through systematically and continually improving the performance of the IT function in an organisation, and through measuring progress and value delivered. Each capability is made up of high-level Categories and Capability Building Blocks. For each capability, the framework outlines 5 levels of maturity, where level 1 indicates a low level of maturity and level 5 indicates an example of best practice (Curley et al, 2016).



**Figure 1:** IT-CMF™

### 3. Findings and discussion

Section 3.1 below presents the findings/themes that emerged from the systematic literature review. Section 3.2 presents the conceptual framework for ISM.

#### 3.1 Themes identified from systematic literature review and concept matrix

In order to better understand the information security management phenomenon, this study adopted a concept-centric examination of the extant literature, focused on the key requirements for securing information resources in the digital context. The key literature insights are now discussed.

Digital business requires a relative<sup>2</sup> view on security that is driven by the organisation's risk appetite and risk tolerance set by the leadership team (Cearley et al, 2015, Ernst & Young, 2015, Accenture, 2013). As such, information security must be solidified as a **key priority on the C-level agenda** (Ernst & Young, 2015, Accenture, 2013, The Open Group, 2011, Whitman and Mattord, 2011). CEOs are now expected to personally know about and be involved in the organisation's information security programme (Raskino, 2015). Such board-level involvement, as well as provision of adequate security funding, communicates the message that information security is a critical issue with business consequences (McClimans et al, 2016, Raskino, 2015). The security programme should be driven by a **clear information security strategy** that is aligned with the business strategy (Cearley et al, 2015, Accenture, 2013), and supported by **effective security policies, procedures, and standards** (Ernst & Young, 2015, Raskino, 2015).

**The organisation needs to evolve its focus on governance** - from an IT governance and a tactical information security focus to enterprise digital governance and enterprise accountability (Accenture, 2014, Sambamurthy and Zmud, 2012). Many CIOs perceive that governance should rest external to the IT function, with roles such as chief information security officer evident across many organisations (LeHong et al, 2014). In general, it is accepted that information security is a board level responsibility, and its management needs to be a shared across all business information users (Accenture, 2013). Greater communication, collaboration, and an enriched security dialogue across departments is required to address information security gaps, as well as a **partnership-type approach with suppliers, partners, and external agencies** (McClimans et al, 2016, Sambamurthy and Zmud, 2012). Furthermore, a wide range of individual responsibilities must be allocated with clear ownership and accountability assigned, as security is regarded as everyone's responsibility (McClimans et al, 2016, Ernst & Young, 2015).

Development of **an information security-aware culture is critical** (CapGemini, 2016). There is a requirement for organisations to 'think like the enemy' at all times (Accenture, 2013); hence the concept of security must become embedded within the organisational mind-set (McClimans et al, 2016). Heightened levels of awareness are necessary to keep pace with the types of threat actors at play and to develop new ways to act on insights gleaned from known security breaches (Accenture, 2013). With the advent of social engineering as an effective attack mechanism (CapGemini Consulting, 2012) people-centric information security needs to be emphasised so that employees are not the weak link (CapGemini, 2016, Sambamurthy and Zmud, 2012), as users who are unaware or misled, can circumvent even the best security systems (CapGemini Consulting, 2012). **Security awareness training is required for all employees** (Ernst & Young, 2015, LeHong et al, 2014). Gaps in the talent pool in terms of the technical and operational skillset required should also be addressed to overcome the challenges posed by a lack of security talent (CapGemini, 2016, McClimans et al, 2016).

In a recent survey, 88% of respondents believed that their cybersecurity approaches did not meet their organisations' needs and 37% did not have a data protection programme or only had ad hoc policies or processes in place (Ernst & Young, 2015). Understanding the rigidity of security controls, on the spectrum from overly relaxed to overly restrictive, can be difficult as organisations now need to simultaneously operate in two worlds: "*the world of tight cost management, slow-moving, risk minimisation and incremental improvement of old IT, versus the new world of entrepreneurial and creative risk-taking, fast-moving, leading-edge digital*" (Raskino, 2014). In instances where digital leaders strive to embrace experimentation, ambiguity, and uncertainty, and quickly and flexibly react to change (Peppard, 2014), the organisation needs to **establish the right balance of controls to secure IT resources without impeding effective business operations**. According to Sambamurthy and

<sup>2</sup> Relative security considers where risks should be mitigated and where they should be accepted (i.e. where the business value exceeds the business risk).

Zmud (2012), “*the real challenge is to balance the necessity to secure an organisation’s computer systems, communication systems, and information systems against the necessity for the organisation to apply IT productively and creatively in executing and evolving the organisation’s business models in the face of an ever-changing competitive environment*” (Sambamurthy and Zmud, 2012).

IT leaders need to **conduct regular threat and vulnerability assessments and map out threat models** for the business in order to help determine the rigidity of controls required (CapGemini, 2016, Accenture, 2014). Threat scenarios should be evaluated to ensure they are inclusive of all relevant perspectives (Accenture, 2013), as the organisation needs to effectively handle both predictable threats and unexpected attacks (Ernst & Young, 2015, Lee and Baby, 2013). This process can be enhanced by incorporating external threat intelligence capabilities (Ernst & Young, 2015, Accenture, 2014, CapGemini Consulting, 2012) and participating in relevant industry sharing communities to share and glean valuable insights (Accenture, 2013). The level of risks faced must be continually re-evaluated as security threats and technologies evolve (Cearley et al, 2015), with risk responses being aligned to the magnitude of the risk posed to the organisation (Accenture, 2013).

Organisations need to **re-conceptualise their information security management and adopt holistic, proactive approaches** that continually adapt to counter emerging threats and minimise the potential negative consequences of exposure (CapGemini, 2016, McClimans et al, 2016, PWC, 2015, Ernst & Young, 2015, Fraser et al, 2014). Organisations need to continually adapt in line with changing business requirements, to design and implement an industry best practice-informed transformation programme and roadmap to continually mature their security practices (Ernst & Young, 2015), and evolve them from being compliance focused, to being threat-centric and strategic risk focused (CapGemini, 2016, Accenture, 2014).

Effective information security management requires an integrated approach covering people, process, and technology (CapGemini Consulting, 2012) and involves a broad spectrum of activities that include anticipation, prevention, protection, detection, and reaction (CapGemini, 2016). As a prerequisite, the organisation should **establish a solid foundation of security measures to provide basic defence** (Ernst & Young, 2015) including identity and access management, and measures to secure data centres, applications, databases, and endpoints (CapGemini, 2016). The organisation also needs to **shift from solely protecting assets to strengthening them and making them more resilient** (McClimans et al, 2016). The security architecture should be revised to reflect ideas regarding depth of defence (Accenture, 2013). Hence, in addition to rich and contextually based access controls, application design and development needs to be security aware (CapGemini, 2016, Cearley et al, 2015) and applications need to be effectively protected at run-time (Cearley et al, 2015). Self-protecting mechanisms are required that include context-based algorithms for identity management, data isolation through mobile containers, rights management tools, and new monitoring capabilities (Cearley et al, 2015).

In order to improve asset resilience, IT leaders need to **keep pace with advances in security technologies** (Accenture, 2013). The organisation needs to look to evolving trends such as cognitive computing/AI, data anonymization, behavioural tracking and analytics, and automation, and they need a mechanism to rapidly pilot and implement such new security technologies and processes. Security teams need to develop innovation and experimentation capabilities to test these new technologies, possibly in a sandbox environment (McClimans et al, 2016).

Finally, organisations need to **develop, deploy, and test processes that enable them to anticipate and detect compromises** to information security and swiftly react to them. This requires a move towards proactive probing, analytics driven event detection and forensics, and reflex-like incident responses (Accenture, 2013). Active defence is a proactive risk-based security approach that involves continually searching for potential attackers and their most likely targets. The insights gleaned enables tailored counter measures to be swiftly implemented to neutralise potential attacks, and facilitates a cycle of continuous learning and improvement that can ultimately lead to improved ROI from security programme investments (CapGemini Consulting, 2012).

Four high-level concepts/categories emerged from the systematic literature review. Within these high-level categories, further sub-concepts or capability building blocks (CBBs) also emerged. These are presented in the form of the ISM conceptual model in the next section and in Table 1.

### **3.2 Conceptual framework**

The Information Security Management (ISM) capability is defined as “the ability to manage approaches, policies, and controls that safeguard the confidentiality, availability, and integrity of information”.

Its key aims are to enable an organisation to:

- Develop and maintain information security approaches, policies, and controls.
- Help employees maintain appropriate levels of understanding and awareness to reduce the occurrence and severity of information security incidents.
- Ensure that all identified incidents, near incidents, and suspected security weaknesses are appropriately investigated and addressed.
- Ensure that the residual risk remaining after the information security technical analysis and mitigation actions for identified security threats have been carried out does not exceed the organisation’s risk appetite.
- Balance the application of information security controls and compliance with regulatory/contractual obligations with the organisation’s ability to engage in innovative initiatives that may support growth and competitiveness.

The conceptual framework for the ISM capability is decomposed into four manageable areas of focus or overarching categories namely: 1) Governance; 2) Technical Security; 3) Security Data Administration; and 4) Business Continuity Management, and 17 sub-categories or Capability Building Blocks. This conceptual framework is presented in Table 1 below.

**Table 1:** ISM capability building blocks

<b>Category</b>	<b>Capability building block</b>	<b>Description</b>
Governance (A)	Information Security Principles, Policies, and Controls (A1)	Define the principles that underpin the organisation’s approach to information security management. Define the information security policies and controls to be put in place, using relevant information security standards, legislative and regulatory compliance requirements, contractual obligations and information security objectives.
	Information Security Strategy (A2)	Develop an information security strategic plan to define how the organisation’s information security objectives can be achieved.
	Governance Structures (A3)	Establish governance structures for information security management. Define the scope of information security management governance bodies, and outline decision rights and authorisations. Establish reporting arrangements, and rules to govern and control the application of information security management authority within the organisation.
	Roles, Responsibilities, and Accountabilities (A4)	Identify the roles required for information security management tasks, and assign employees with the requisite knowledge and experience to those roles. Define the associated responsibilities and assign these to employees who will be accountable for ensuring that information security management objectives are met.
	Skills and Competence Development (A5)	Put in place an information security management training curriculum and other employee developmental mechanisms.
	Culture and Stakeholder Management (A6)	Establish an information security management aware culture. Motivate stakeholders, secure their support for key information security management initiatives, and create a shared understanding of how they can ensure that information security objectives are met.
	Security Performance Measurement (A7)	Monitor and report on the effectiveness/efficiency of the information security principles, policies, controls, strategy, and activities.
	Supplier Security Requirements (A8)	Define security requirements for the procurement and supply of hardware, software, data, and cloud and other IT services to satisfy the information security strategy and objectives.

Technical Security (B)	Security Architecture (B1)	Build security criteria into the design of IT solutions and services – for example, by defining coding protocols, depth of defence, and configuration of security features.
	IT Device Security (B2)	Define, implement, and monitor measures to protect all IT devices such as networks, servers, client computing devices, storage devices, printers, and smart phones.
	Physical Infrastructure Security (B3)	Implement, monitor, and maintain measures to safeguard the IT physical infrastructure from threats including extremes of temperature, fire, flooding, malicious intent, and disruptions.
Security Data Administration (C)	Data Security Classification (C1)	Define information security classes, and provide guidelines on protection levels and access controls appropriate to each class.
	Access Rights Management (C2)	Manage user access rights to information throughout its life cycles, including the granting, denying, and revoking of access privileges.
	Data Life Cycle Management (C3)	Provide the security expertise and guidance to ensure that data throughout its life cycles is appropriately available, adequately preserved, and/or destroyed so that it meets business, regulatory, and/or other security requirements.
Business Continuity Management (D)	Business Continuity Planning (D1)	Provide information security advice to assist in the analysis of incidents and to ensure that data is secure before, during, and after the execution of the business continuity plan.
	Security Risk Management (D2)	Establish an approach to the profiling of security threats and the assessment, prioritisation, treatment, and monitoring of security risks and vulnerabilities.
	Incident Management (D3)	Manage information security-related incidents and near incidents. Establish incident response teams to identify and limit exposure, and to coordinate with regulatory bodies as appropriate.

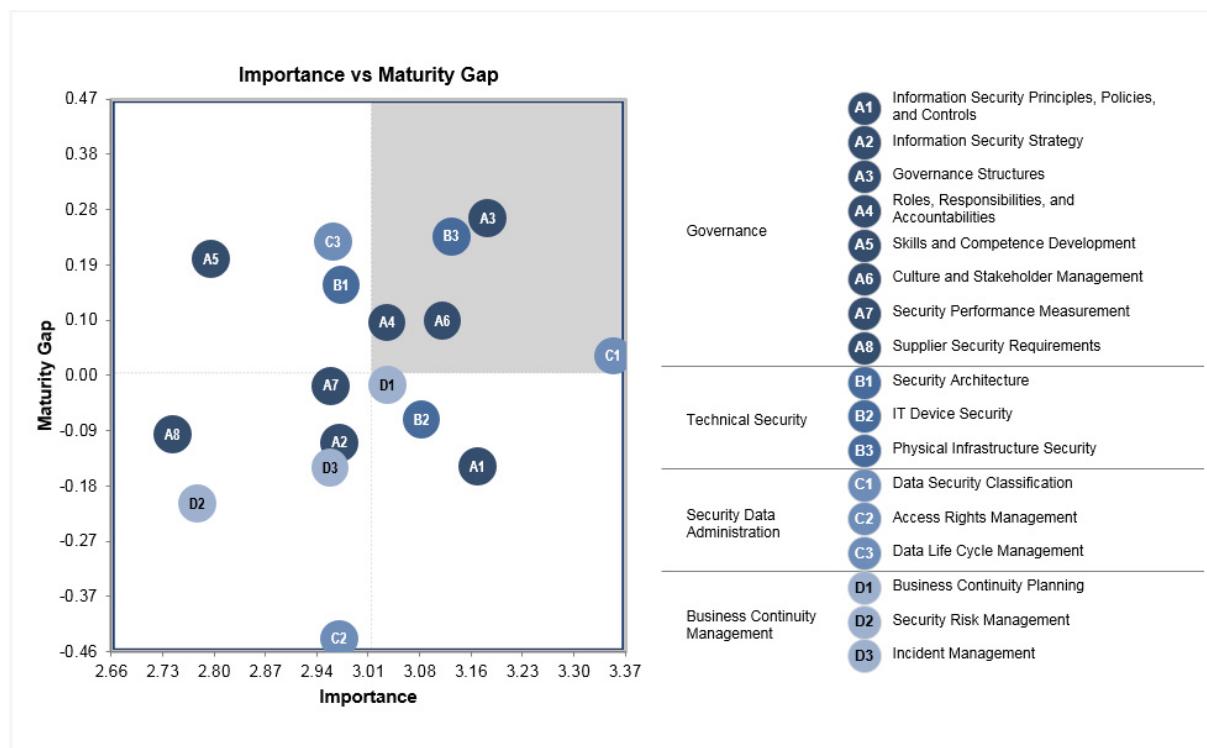
In order to effectively operationalise this ISM capability, a five-level maturity profile is defined at a capability level. A summary view of this is outlined in Table 2 below:

**Table 2:** ISM capability summary maturity profile

		Information Security Management Maturity Profile
5	Optimised	The organisation demonstrates excellent security capabilities, and monitors and applies the best tools, technologies, methods, and approaches to its information security management.
4	Advanced	Comprehensive information security policies, controls, and approaches are fully integrated across the organisation. Forensic analysis of incident data is undertaken to classify all incidents and near incidents, to diagnose their underlying cause, to assess their impact, and to identify corrective measures.
3	Intermediate	Standardised information security policies, controls, and approaches are in place – dealing with access rights, business continuity, toolsets, incident response management, audits, non-compliance, and so on. Forensic analysis of incident data is undertaken to classify some incidents and near incidents, to diagnose their underlying cause, and to assess their impact.
2	Basic	Defined information security policies, controls, and approaches are emerging, focused primarily on compliance with regulations. Some basic forensic analysis of incident data is undertaken.
1	Initial	The approach to information security tends to be localised. Typically, incidents are either not detected or are not responded to in a timely manner.

This maturity profile serves as the basis for a maturity assessment instrument which can be used by organisations to holistically analyse their performance in respect of their information security management capability – i.e. their ‘current’ capability maturity and their ‘target’ capability maturity. Figure 2 provides sample insights that

can be derived through undertaking this maturity assessment – this figure outlines the organisation's importance rating vis-à-vis their capability maturity gap, which provides key insights into target areas for improvement.



**Figure 2:** ISM capability – capability building block importance versus maturity gap

To effectively manage information security an organisation needs to develop a wide-range of capabilities, which will vary in importance depending on the business context and the organisation's needs. However, time and resource constraints will undoubtedly challenge organisations who attempt to develop multiple capabilities simultaneously, hence a key requirement is the need for the organisation to focus on developing the capabilities of greatest importance first. The assessment tool will enable this identification and prioritisation process by highlighting the capabilities of greatest importance against the maturity gap (i.e. the current maturity vs. the target maturity). Figure 2 illustrates how this can work in practice, as it highlights the key areas for improvement, in the upper right (shaded) segment.

To enable and support the key areas for improvements, a series of supporting material and tools are available as follows:

- *Practices, Outcomes and Metrics:* For each capability of the IT-CMF™ there are representative *practices* that defines the organisations current information security maturity and how to progress to the next level of maturity. Each *practice* is accompanied by an *outcome* that states what benefits might result by following the *practice*. To help the organisation to measure how successful they are in achieving the *outcome* one or more *metrics* are provided.
- *Capability Performance Indicators (CPI):* CPI's are directly relate to the ISM capability and are designed to make a connection between the goals, improvement targets and business outcomes. They are used to understand the organisation's progress towards expected outcomes. The CPIs are grouped into balance score card segments (i.e. financial, process, customer, learning and growth) to provide an overview of the target capability improvement.
- *Skills Framework:* To support the improvement of the ISM capability, organisations need to identify the skills to deliver the plan, and also need to harness and coordinate those skills. The skills frameworks provides structured guidance on key skills and competences required for the full range of ICT practices at different levels.

## **4. Conclusion**

Information security is central to the continuity of an organisation's business operations and its adherence to legal and regulatory requirements. However, in the digital context a re-conceptualisation of information security management is required to enable the organisation to address information security threats in more agile and proactive ways. Failure to do so can result in the organisation being impacted by high-profile security breaches. Guided by the direction and sponsorship of C-level executives and clear strategy, policies, procedures, and standards, the organisation can build an effective information security capability for the digital context. Inspiring a security-aware culture or mind-set, and ongoing cognisance of external threat intelligence and advances in security technologies are prerequisites for information security success. Similarly, adopting holistic, proactive, and continually adaptive approaches to anticipate, detect, and react to security compromises can enable the organisation to more effectively counter emerging threats and minimise the potential negative consequences of exposure.

In summary, this paper presents a conceptual framework; inclusive of an assessment instrument and supporting tool kits that can be used by organisations to mature or improve their information security management capability. The paper highlights how the insights gained can be used in a practical sense to effectively secure information resources and to comply with strict regulatory requirements. This research also acts as a starting point and the basis for a more in-depth development and expansion of the Information Security Management (ISM) capability. As such, practitioners and academics interested in the area are asked to validate this capability to help determine its relevancy in a real-life industry setting.

### **4.1 Limitations and future research**

This paper was limited due to word-count. Further detailed discussion on the conceptual model is planned in future publications. This research was also restricted in that only secondary sources of data (academic journal articles, practitioner reports etc.) and insights from a small sample of subject matter experts were used in the development of the model due to time-constraints. However, further empirical testing is planned for the future with firms who are interested in ISM in a digital context. As such, refinements to the model are expected.

## **References**

- Accenture, (2013) 'Accenture technology vision 2013. Every business is a digital business', Online at: <<https://www.accenture.com/us-en/acnmedia/Accenture/Conversion-Assets/Microsites/Documents8/Accenture-Technology-Vision-2013.pdf>>.
- Accenture, (2014) 'Accenture technology vision 2014. Every business is a digital business - from digitally disrupted to digital disrupter', Online at: <<http://investor.accenture.com/~media/Files/A/Accenture-IR/events-and-presentations/Accenture-Technology-Vision-2014.pdf>>.
- Arend, C., Sundby, N. and Venkatraman, A. (2016) 'Reinventing data protection fit for digital transformation', IDC. Online at: <<https://www.hpe.com/h20195/v2/GetPDF.aspx/4AA6-8166ENW>>.
- Bosworth, S., Kabay, M.E. and Whyne, E., (2014) 'Computer security handbook', 6th ed. Hoboken, NJ, U.S.A: John Wiley and Sons.
- Bharadwaj, A., El Sawy, O., Pavlou, P. and Venkatraman, N. (2013) 'Digital business strategy: toward a next generation of insights', MIS Quarterly, vol. 37, no. 2, pp 471-482.
- Bradley, J., Loucks, J., Macaulay, J., Noronha, A. and Wade, M. (2015) 'Digital vortex: How digital disruption is redefining industries', Global Centre for Digital Business Transformation: An IMD and Cisco initiative. Online at: <<http://www.cisco.com/c/dam/en/us/solutions/collateral/industry-solutions/digital-vortex-report.pdf>>.
- CapGemini Consulting, (2012) 'No digital transformation without cybersecurity', Online at: <[https://www.capgemini.com/consulting-nl/wp-content/uploads/sites/33/2017/08/geen\\_digitale\\_transformatie\\_zonder\\_cybersecurity\\_0.pdf](https://www.capgemini.com/consulting-nl/wp-content/uploads/sites/33/2017/08/geen_digitale_transformatie_zonder_cybersecurity_0.pdf)>.
- CapGemini, (2016) 'Address c-level cybersecurity issues to enable and secure digital transformation', Online at: <[https://www.capgemini.com/de-de/wp-content/uploads/sites/5/2017/07/1602\\_cybersecurity\\_strategic\\_consulting\\_brochure\\_cc\\_web\\_en\\_1.pdf](https://www.capgemini.com/de-de/wp-content/uploads/sites/5/2017/07/1602_cybersecurity_strategic_consulting_brochure_cc_web_en_1.pdf)>.
- Catlin, T., Scanlan, J. and Willmott, P. (2015) 'Raising your digital quotient'. McKinsey Quarterly, pp.1-14. Online at: <<https://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/raising-your-digital-quotient>>.
- Cearley, D.W., Walker, M.J., and Blosch, M. (2015) 'The top 10 strategic technology trends for 2015', Gartner. Online at: <<https://www.gartner.com/doc/2964518/top-strategic-technology-trends>>.
- Chesbrough, H. (2003) 'Open innovation'.
- Council on Cybersecurity, 'Critical controls for effective cyber defence', Gartner, 2013. [Online] Available: <<http://www.counciloncybersecurity.org/critical-controls/>>.

- Curley, M., Kenneally, J., and Carcary, M. (Eds) (2016). IT Capability Maturity Framework (IT-CMF™) – the body of knowledge guide. 2nd edition. Van Haren.
- Dahlström, P., Ericson, L., Khanna, S., and Meffert, J. (2017). From disrupted to disrupter: reinventing your business by transforming the core. McKinsey.
- Drnevich, P.L. and Croson, D.C. (2013) 'Information technology and business-level strategy: Toward an integrated theoretical perspective'. MIS Quarterly, 37(2).
- Ernst & Young, (2015) 'Creating trust in the digital world - EY's global information security survey 2015', Online at: <[http://www.ey.com/publication/vwluassets/ey-global-information-security-survey-2015/\\$file/ey-global-information-security-survey-2015.pdf](http://www.ey.com/publication/vwluassets/ey-global-information-security-survey-2015/$file/ey-global-information-security-survey-2015.pdf)>.
- Fichman, R.G., Dos Santos, B.L. and Zheng, Z.E. (2014) 'Digital innovation as a fundamental and powerful concept in the information systems curriculum', MIS quarterly, 38(2).
- Fraser, J., Simkins, B. and Narvaez, K., (2014) 'Implementing enterprise risk management: Case studies and best practices'. John Wiley & Sons.
- Frost and Sullivan (2017) 'The 2017 (ISC)<sup>2</sup> global information security workforce study – benchmarking workforce capacity and response to cyber risk' Online at: <<https://iamcybersafe.org/wp-content/uploads/2017/06/Europe-GISWS-Report.pdf>>.
- Gutwirth, S., Leenes, R., and De Hert, P. (eds.), (2017) 'Data protection on the move - current developments in ICT and privacy/data protection', Dordrecht: Springer.
- Identity Theft Resource Centre - ITRC (2018) 'End-of-year Data Breach Report 2018'. Online at: <[https://www.idtheftcenter.org/wp-content/uploads/2019/02/ITRC\\_2018-End-of-Year-Aftermath\\_FINALWEB-V2-2.pdf](https://www.idtheftcenter.org/wp-content/uploads/2019/02/ITRC_2018-End-of-Year-Aftermath_FINALWEB-V2-2.pdf)>
- International Organisation for Standardisation (ISO), 'ISO/IEC 27002: 2013. Information technology – security techniques – code of practices for information security controls', 2013. [Online] Available: <[http://www.iso.org/iso/catalogue\\_detail?csnumber=54533](http://www.iso.org/iso/catalogue_detail?csnumber=54533)>.
- ISACA, 'COBIT 5 for information security', (2012) Online at: <<http://www.isaca.org/cobit/pages/info-sec.aspx>>.
- ISACA, 'COBIT 5 for risk', (2013) Online at: <<http://www.isaca.org/cobit/pages/risk-product-page.aspx>>.
- Joint Task Force Transformation Initiative, 'Security and privacy controls for federal information systems and organisations'. Gaithersburg, MD: National Institute of Standards and Technology, 2013. [Online] Available: <<http://dx.doi.org/10.6028/NIST.SP.800-53r4>>.
- Lee, O. and Baby, D.V., (2013) 'Managing dynamic risks in global it projects: Agile risk-management using the principles of service-oriented architecture'. International Journal of Information Technology & Decision Making, 12(06), pp.1121-1150.
- LeHong, H., Prentice, S., Steenstrup, K., Nielsen, T., and Perkins, E. (2014) 'How CIOs need to think about digital business technologies', Gartner, Online at: <<https://www.gartner.com/doc/2739917/cios-need-think-digital-business>>.
- McClimans, F., Fersht, P., Snowdon, J., Phelps, B. and LaSalle, R. (2016) 'The State of Cybersecurity and Digital Trust', HfS Research &Accenture, Ltd. Online at: <[https://www.accenture.com/t20160704T014005\\_w\\_us-en/acnmedia/PDF-23/Accenture-State-Cybersecurity-and-Digital-Trust-2016-Report-June.pdf](https://www.accenture.com/t20160704T014005_w_us-en/acnmedia/PDF-23/Accenture-State-Cybersecurity-and-Digital-Trust-2016-Report-June.pdf)>.
- Mueller, U.M. and Allen, K. (2014) 'How to use cybersecurity to generate business value', Ernst & Young. Online at: <[http://www.ey.com/Publication/vwLUAssets/EY\\_CIO\\_-\\_How\\_to\\_use\\_cybersecurity\\_to\\_generate\\_business\\_value/\\$FILE/EY-CIO-How-to-use-cybersecurity.pdf](http://www.ey.com/Publication/vwLUAssets/EY_CIO_-_How_to_use_cybersecurity_to_generate_business_value/$FILE/EY-CIO-How-to-use-cybersecurity.pdf)>.
- Narain Singh, A., Gupta, M.P. and Ojha, A. (2014) 'Identifying factors of organisational information security management' Journal of Enterprise Information Management, 27(5), pp.644-667.
- Pental, S. (2015) 'Five ways information security can help IT improve stakeholder engagement'. CEB IT Quarterly–Spotlight on business engagement. Q. 2. Online at: <<http://ceb.uberflip.com/i/502110-cio152185syn-rp-q2-it-quarterly-web/33?m4=>>.
- Peppard, J., (2014) 'Digital dynamics in the C-suite: accelerating digitisation with the right conversations', Sungard, Online at: <[https://www.sungardas.com/globalassets/\\_multimedia/document-file/digital-dynamics-in-the-c-suite.pdf](https://www.sungardas.com/globalassets/_multimedia/document-file/digital-dynamics-in-the-c-suite.pdf)>.
- PWC, (2015) 'Global digital IQ® survey: lessons from digital leaders - 10 attributes driving stronger performance', Online at: <<https://www.pwc.es/es/publicaciones/gestion-empresarial/assets/septima-encuesta-mundial-coeficiente-digital.pdf>>.
- Shepherdson, K., Hioe, W. and Boxall, L. (2016) '88 Privacy Breaches to Beware of: Practical Data Protection Tips from Real Life Experiences', Marshall Cavendish International Asia Pte Ltd.
- Raskino, M. (2015) '10 CEO information and technology resolutions for 2015', Gartner, Online at: <<https://www.gartner.com/doc/2973217/-ceo-information-technology-resolutions>>.
- Raskino, M. (2014) 'CEO resolutions for 2014. Time to act on digital business'. Gartner.
- Sambamurthy, V. and Zmud, R.W. (2012) 'Guiding the digital transformation of organisations', Legerity Digital Press.
- The Open Group, 'Open information security management maturity model (O-ISM3)', 2011. [Online] Available: <<https://www2.opengroup.org/ogsys/isp/publications/PublicationDetails.jsp?publicationid=12238>>.
- Webster, J. and Watson, R.T. (2002) 'Analyzing the past to prepare for the future: Writing a literature review'. MIS quarterly, pp.xiii-xxiii.
- Whitman, M. and Mattord H. (2011) 'Principles of information security'. Boston, MA: Cengage Learning.
- Zahra, S.A., Sapienza, H.J. and Davidsson, P. (2006) 'Entrepreneurship and dynamic capabilities: A review, model and research agenda'. Journal of Management studies, 43(4), pp.917-955.

# Leveraging the OODA Loop with Digital Analytics to Counter Disinformation

Jami Carroll

Unaffiliated, USA

[jcarroll@prisidian.com](mailto:jcarroll@prisidian.com)

**Abstract:** Propaganda embraces psychological action and in some cases psychological warfare. Psychological action is where the propagandist attempts to change opinions using only psychological means where the individual or organization doubts the validity of their belief system and actions. Psychological warfare is the psychological action when it is applied to an adversary. Propaganda plays on personal complexes, drives, motivations, passions, and prejudices. The U.S. Senate Select Committee on Intelligence identified examples of this type of propaganda and manipulation with the Russian Federation conducting influence operations (IO) worldwide with the manipulation of social media, Democratic National Committee (DNC) decisions, election polls, and even TV station broadcasts. Offensively, these examples of soft cyber power directly support the Military Deception (MILDEC) used to deliberately mislead the adversary or potential adversary and thereby allowing the execution of operations. These forms of manipulation are also used by non-military adversaries such as Internet trolls, Gamergaters, Hate Groups & Ideologues, Conspiracy Theorists, Influencers, Hyper-Partisan News Outlets, and Politicians. In an attempt to counter these attacks, this research will use a grounded theory approach to examine how John Boyd's observe–orient–decide–act (OODA) loop theory can be leveraged in a workflow of digital analytics that could possibly be used to be used to enhance the prosecution of disinformation and fake news.

**Keywords:** information operations, OODA Loop, psychological warfare, disinformation theory

---

## 1. Introduction

All statements of fact, opinion or analysis expressed are those of the author. The views and opinions expressed herein by the author do not represent the official policies or positions of the United States (U.S.) Department of Defense (DoD), U.S. Navy, U.S. Cyber Command (USCYBERCOM) or other agencies or departments of the U.S. government and are solely representative of the views of the author.

This study will examine how the OODA Loop combined with data analytics can be embraced to improve the ability of cyberspace operations professionals to counter disinformation and fake news that can better enable the defensive cyber operations (DCO) and offensive cyber operations (OCO) communities. The purpose of this Qualitative Grounded Theory study is to establish a workflow of processes using data analytics as a means to expand the knowledge, skills and abilities (KSA) of OCO/DCO cyberspace operations professionals. This Grounded Theory approach will focus on the methods, tools, and processes that should be included into digital analytics. Expanding this KSA will expand the ability of these cyberspace operations professionals to examine critical Indicators and Warnings (I&W) key to the exploitation of disinformation Tactics, Techniques and Procedures (TTP) that can lead to improved tradecraft in handling this form of IO. A review of the literature and analysis of what has been leveraged thus far in this field is the primary source for this research. The primary research question was: How can the OODA Loop be leveraged with the use of Digital Analytics to counter disinformation? It is anticipated that the results of this study could be used to enhance the Denial and Deception (D&D) TTP of DCO and OCO personnel. This foundational work may help USCYBERCOM, U.S. Service/Agencies, and our Allies.

## 2. The need to improve analysis of disinformation

Former Director of National Intelligence (DNI) Jim Clapper said counterterrorism, counterproliferation, cybersecurity and counterintelligence are the top four issues affecting the U.S. intelligence community (Clapper, J. 2012; Clapper, J. 2015). He indicated that cyber-attacks could come from extremists, foreign intelligence services, foreign/domestic terrorist groups, hackers, insider threats and organized crime groups (Clapper, J. 2012; Clapper, J. 2015). Non-kinetic cyber warfare has been laced with kinetic warfare in countries like Estonia in 2007, South Ossetia (Georgia) in 2008, Kyrgyzstan in 2009, and the Autonomous Republic of Crimea (Ukraine) (Czosseck, C. & Geers, K., 2009).

### 2.1 Information Operations (IO)

IO is a process of “information-related capabilities” integrated into military operations with the intent “to influence, disrupt, corrupt, or usurp the decision-making of adversaries” while providing protection for friendly

forces (CNSS, 2015, p. 63). MILDEC is often applied in a D&D framework to mislead another state entity (Dulles, A. 2006; Zakem, V., McBride, M., & Hammerberg, K., April 2018). MILDEC is used to influence a nation state's Command, Control, Command, Control, Communications, Computers, Intelligence, Surveillance and Reconnaissance (C4ISR) through disruption, degradation, disablement, or deception using Computer Network Attack (CNA) or Computer Network Exploitation (CNE) operations (Andress, J. & Winterfeld, S. 2011). Arquilla, J. et al. (1999) saw IO as something the U.S. needed to expand. Control of cyberspace has become one of the strategic goals of some world powers; this strategic goal has been coined the Great Power Competition (GPC) (Bey, 2018).

### *2.1.1 Military Deception (MILDEC)*

MILDEC is the deliberate misleading of the adversary and how they execute their operations while allowing own operations to be carried out (JS, 2012). Understanding how the adversary carries out their decision-making processes provides opportunities to change the attack time, location, or force strengths (JS, 2012). Denying information can prevent or slow the enemy from gaining knowledge or operations by hiding the information from the adversary or modifying the information the adversary gets (Bennett, M. & Waltz, E., 2007). This information could be physical, technical, and administrative and could be integrated in with D&D techniques such as cover, concealment, and camouflage & denial and deception (C3D2) (Bennett, M. & Waltz, E., 2007). Whaley, B. (2007) argued that military commanders gain a distinctive edge over an adversary using deception by leveraging secrecy and uncertainty through C3D2 (Heckman, K. et al., 2015; Patrikarakos, D., 2017; Rushkoff, D. et al., 2018).

D&D has been adapted to cyber warfare using CNA and CNE (Amoroso, E. 2011; Andress, J. & Winterfeld, S. 2011; Gabriel, R., 2004; Rowe, N. & Custy, E. 2008). The adversary receiving the deception is referred to as the Target Audience (TA) or Deception Target (DT) (Bennett, M. & Waltz, E., 2007; JS, 2012). Examples of military deception include the Trojan Horse and Genghis Khan (Coolidge, O. & Sandoz, E., 1980). Deception was also used by the Soviets when installing missiles in Cuba (Dulles, A., 2006). During World War II, Operations Mincemeat, Fortitude and Overlord used D&D leading to the Normandy Invasion in 1944 (Dulles, A., 2006).

Bennett, M. and Waltz, E. (2007) laid out three major phases that led to the evolution of military deception: 1) development of the conceptual phase of what constitutes deception, 2) historical analysis of deception approaches, and 3) theoretical deception. The World War II operations laid the groundwork of the conceptual and historical phases where deceptions approaches were developed, tested and applied while analyzing the successes and failures of these approaches. Refinement of deception occurred in the mid-1970s with the Yom Kippur War; this deception not only manipulated the adversary, but also friendly forces (Bennett, M. & Waltz, E., 2007). The Yom Kippur War provided the final transition from the historical analysis to the theoretical deception phase with the advanced techniques affecting adversary and friendly forces (Bennett, M. & Waltz, E., 2007). During the Yom Kippur War, the Israelis used D&D to hide their approach from friendly and adversary forces by leaving few I&W events (Bennett, M. & Waltz, E., 2007).

### *2.1.2 Disinformation*

Marwick, A. and Lewis, R. (2017) indicated that the media is being manipulated by Internet trolls, Gamergaters, Hate Groups & Ideologues, Conspiracy Theorists, Influencers, Hyper-Partisan News Outlets, and Politicians. Deception is effective because it draws attention away from the adversary using the detection because it creates a false network or disinformation which creates a high degree of uncertainty (Amoroso, E., 2011; Jajodia, S. et al., 2016; Mahairas, A. & Dvilyanski, M., 2018; Marwick, A. & Lewis, R., 2017). Ellul, J. (1973) indicated that propaganda embraces psychological action, psychological warfare, and re-education and brainwashing. Psychological action is where the propagandist attempts to change opinions using only psychological means where the individual or organization doubts the validity of their belief system and actions with psychological warfare occurring when applied against an adversary. Ellul, J. (1973) further suggests that propaganda plays on complexes, drives, motivations, passions, and prejudices (xiii). Dulles (2006) said the Russian State Security Service (KGB) was known as the "Disinformation Bureau". During the period of 1957 to 1960 there were 32 forged documents fabricated to appear to have come from high ranking government officials within the U.S.

In 2017, testifying before the Select Senate Committee of the U.S. Congress Kevin Mandia, Chief Executive Officer (CEO) of FireEye, indicated that it was critical to understand the IO campaigns used today and since the Cold War due to the impacts it has on the U.S. and its Allies (Mandia, K., 2017). According to Mandia, K., (2017),

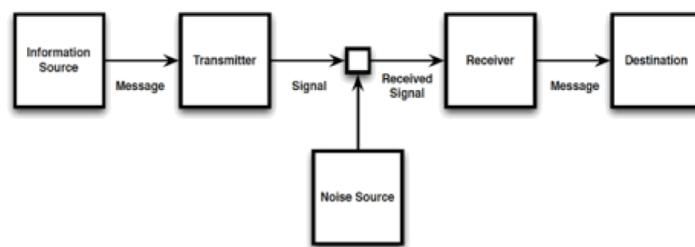
in the past couple years, Russian IO focused largely on hacking and leaking. Some of the targets were high profile individuals such as the Chairman of the Joint Chiefs of Staff; Assistant Secretary of the Air Force; and U.S. Ambassador to Russia (Mandia, K., 2017).

### 2.1.3 OODA loop

The strategy of John Boyd's observation orientation-decide-act" (OODA) loop in quickly and decisively defining the kill chain of a target aircraft for aircraft pilots (Osinga, F., 2007; Whitehead, Y., 1997). These same approaches have been applied to information weaponry where information must be perceived and brought through the OODA Loop process to act on it (Blodgett, 2003; Whitehead, Y., 1997). The OODA Loop is based on endpoint, just-in-time changes to strategy with surgical responses capitalizing on speed and accuracy while responding more rapidly to the adversary's tactics, techniques and procedures (TTP) (Blodgett, D. et al., 2003; Boyd, J., 1987; Osinga, F., 2007). Boyd suggested that he could paralyze the opponent's OODA Loop because of the way information is filtered; this is largely due to the orientation of information (Boyd, F., 1987). Stein, G. (1996) argued that information warfare (IW) is largely based on platform-to-platform offensive and defensive attacks such as psychological operations (PSYOPS) where military deception is the primary ingredient in the attack (Stein, G., 1996). Course of Actions (COAs) must be identified and the best selected to ensure your OODA Loop defeats the adversary's OODA Loop (Blodgett, D. et al., 2003).

### 2.1.4 Detection Disinformation

Disinformation Theory is built from Shannon's Communications Model (Alexander, J. and Smith, J., 2011). Whereas Shannon's Communication Model builds the simple idea of an information being transmitted over a median to a receiver to a destination, Disinformation Theory adds an additional component called a noise source such as D&D which adds the noise or misinformation which misleads or deceives the recipient. Figure 1 illustrates this Disinformation Theory representation. Bennett, M. and Waltz, E. (2007) suggested three major landmark events helped shaped deception theory; they were Operation Fortitude in the 1940s; the Yom Kippur War in the early 1970s (Rabinovich, A., 2004); and the Cold War in the 1970s and 1980s. Bell and Whaley are considered the fathers of Deception Theory; they came up with two broad categories of deception: dissimulative and simulative (Santos, E. & Johnson, G., 2004). The goal of dissimulative is based on hiding the real information about something while simulative is focused on providing false information about something.



**Figure 1:** Disinformation Theory

Analytical capabilities is required for evaluating numbers, words, and the larger context of the data being presented is key. (Alexander, J. and Smith, J., 2011; Bajaj, P. et al., 2016; Hosseini, H. et al., 2017; Levitin, D., 2017; Svetoka, S., 2016; Teyssou, D. et al., 2017). Table 1 illustrates the general and specific areas where analytical capabilities where machine learning, artificial learning, and deep learning could be applied.

**Table 1:** General and specific detection areas

General Detection Area	Specific Area
Numbers	Number collection
Numbers	Number reporting
Numbers	Statistical analysis (averages & probabilities)
Words	Expertise (source of data)

General Detection Area	Specific Area
Words	Overlooked/Undervalued Alternative Explanations
Words	Counter knowledge
Larger context of the data	Scientific theories and proofs
Larger context of the data	Framing probabilities
Larger context of the data	Framing risks
Image/video	Modification & retouching
Image/video	Image bit/frame manipulation

According to Amoroso, E. (2011), deception is effective because it draws attention away from the adversary using the detection because it creates a false network or disinformation which creates a high degree of uncertainty. Malin, C. et al., (2015) indicated that deception has been used with false flags, sock puppets, impersonation and persuasive technology which starts off in many cases as cybercrime-as-a-service and Distributed Denial of Service (DDoS) attacks.

#### *2.1.5 Cyberterrorism, perception management and propaganda*

Deception has also been used with cyberterrorism. Veerasamy, N. (2010) pointed out that one of the basic steps in cyberterrorism is the use of propaganda and disinformation. Building on the three cyberterrorism offensive attacks outlined by Arquilla, J. (2001), perception management and propaganda is used prior to either disruptive attacks that temporarily affect a service, site or system or destructive attacks that destroy a service, site or system. Because of this complexity, Veerasamy, N. (2010) suggested that the OODA Loop should be used in analyzing this use of propaganda and disinformation prior to an attack. Some of these preliminary cyberterrorism steps that have elements of D&D include: credit card information theft, crucial service disruption, data corruption, denial of service, disinformation, propaganda, and Web defacement (Tompkins, J., 2018; Veerasamy, N., 2010). With the multitude of complexities and variations in attacks, the OODA Loop is a natural capability for creating a workflow to analyze how best to respond to these attack vectors.

#### *2.1.6 Social media as a target-rich environment*

Social networks accompanied with digital media has helped terrorist groups spread propaganda while providing these terrorists with a variety of means to communicate and where friendly and adversary OODA Loops could be applied (Shehabat, A. & Mitew, T., 2017). Islamist State terrorists have created their own IO campaign with anonymous and encrypted platforms such as Facebook, Justpase, Sendvid, Telegram, Twitter, WhatsApp, and persistent storage with Google Drive and Dropbox (Shehabat, A. & Mitew, T., 2017). Using this multitude of social media platforms and persistent storage platforms, terrorist organizations are able to evade interception, hide their network structure and survive attempts to disassembly (Shehabat, A. & Mitew, T., 2017). Because of the complexity, OODA Loop analysis is a natural fit (Shehabat, A. & Mitew, T., 2017).

#### *2.1.7 Identification and separation of humans from bots*

One of the most challenging problems with propaganda and disinformation is the separation of humans from Bots (Ferrara, E., 2017). According to Ferrara, cognitive behavioral modeling techniques and machine learning have been used to identify humans from Bots by looking at their characteristics (Ferrara, E., 2017). An added benefit is that users that respond and how they respond has also been analyzed using cognitive behavioral modeling techniques and machine learning (Ferrara, E., 2017).

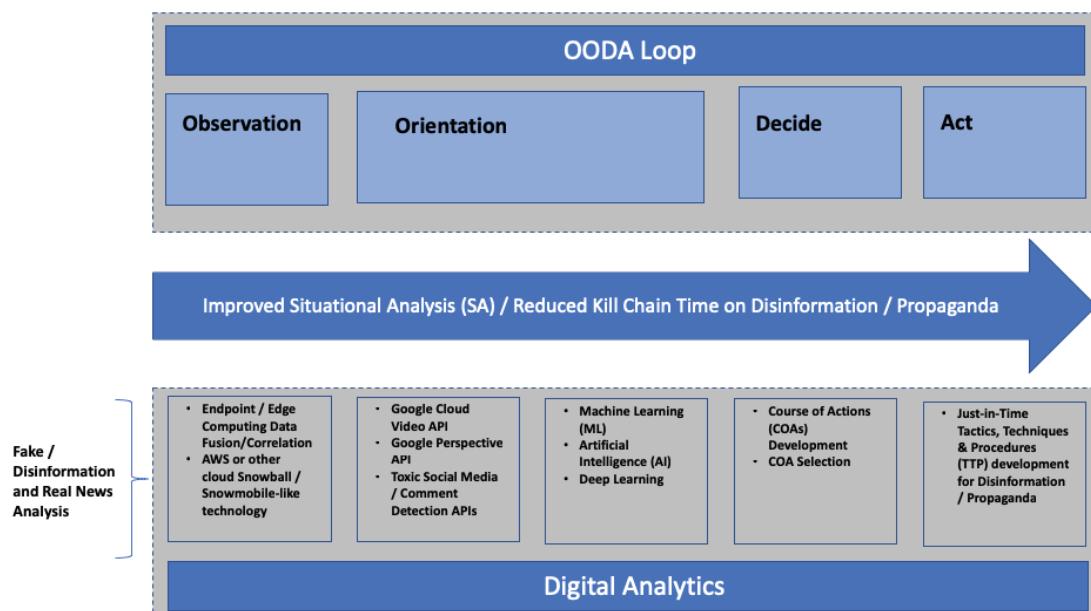
Rumor detection is another area where supervised machine learning (ML) algorithms are useful (Kumar, A. & Sangwan, S., 2019). According to Kumar, A. and Sangwan, S. (2019), rumors can not only be identified using ML, they can be classified as true (factual), false (nonfactual), or unresolved. These rumors can also be analyzed to determine if they are coming from one or multiple sources; this makes it easier to reduce the proliferation across multiple networks (Kumar, A. & Sangwan, S., 2019). Fact-checking of Internet postings is one of the key efforts where ML classification has yielded significant results in disinformation detection (Lemieux & Smith, 2018). According to Lemieux, V. and Smith, T. (2018), this use of ML can be easily used to classify authentic from unauthentic information.

## 2.2 Data Analytics within the OODA loop

Data Analytics through Machine Learning (ML), Artificial Intelligence (AI), and Deep Learning often great opportunities in making more effective decisions in science and business. Data analysis entails collecting/inspecting data; cleansing it; transforming it; and modeling it with known and suspected facts for decision making (Zager, R. & Zager, J., 2017). Data Analytics can be applied to structured data (textual analytics such as statistics, speech, and written formats and unstructured data (images and video). In most cases, 80-90 % of unstructured data is likely to have textual tags (metadata) which supports analysis (Tompkins, J., 2018; Zager, R. & Zager, J., 2017).

OODA Loops have been used extensively in military strategy and data analysis is a critical element to this analysis (Angerman, W., 2004; Hamilton, C. & Kreuzer, U., 2018; Osinga, F., 2007; Schechtman, G., 1996; Zager, R. & Zager, J., 2017). ML has shown to be particularly useful in OODA Loop analysis (Bodström, T. & Hämäläinen, T., 2018; Zager, R. & Zager, J., 2017). IBM Watson's AI capability leverages OODA Loops as a feedback loop to understand over 100 data analysis decision making processes due to the speed of execution (Zager, R. & Zager, J., 2017). IBM Watson AI uses cognitive security to train the AI to understand to situational awareness by collecting and analyzing historical events to provide explanatory analysis of what might happen (Zager, R. & Zager, J., 2017). The U.S. Army is exploring the concept of Internet of Battle Things (IOBTs) so dynamically composable force capabilities can be agilely created to support intelligent command and control through predictive analysis (Russell, S. and Abdelzaher, T., 2018). Figure 2 illustrates an approach where data analytics could be applied within the OODA Loop construct. This illustrates how each phase of the OODA Loop can be leveraged with digital analytics to enhance the prosecution of disinformation/propaganda. The approach proposed:

- In the Observation Phase, fake news/disinformation/real news is ingested at the endpoint for quicker, near real time fusion/correlation using cloud technology like Amazon Web Services (AWS) Snowball/Snowmobile technology with data storage/compute environment.
- In the Orientation Phase, existing detection Application Programming Interfaces (APIs) are used to detect structured and unstructured video and textual data using ML, AI, and Deep Learning to identify likely fake news, disinformation, and propaganda.
- The Decide Phase identifies Courses of Action (COAs) and decision ranking of these COAs.
- The Act Phase supports just-in-time Tactics, Techniques, and Procedures (TTPs) developed by intelligence analysts to support the warfighter with approaches to mitigate or perform a counter-deception of the fake news, disinformation, and propaganda. The end result is improved situational analysis (SA) / reduced kill chain time on disinformation / propaganda.



**Figure 2:** Data Analytics within the OODA Loop

### **3. Qualitative research methodology using grounded theory**

Glaser, B. and Strauss, A. (1967) asserted that Grounded Theory (GT) was useful in building a framework and generating a theory of how something works. GT is a qualitative research methodology that uses methodological gathering, analysis, and sampling of data that when posed a research question through repeated sampling of data helps to develop the theory when no theory or framework previously existed (Glaser, B. and Strauss, A., 1967). As data is collected, it is coded so that it can be binned or grouped into several common themes where these themes evolve into a framework (Glaser, B. and Strauss, A., 1967). Data for this GT qualitative research methodology was collected through an extensive review of the body of knowledge associated with OODA Loop, Data Analytics, and Deception.

### **4. Data collection, analysis and results**

ProQuest and Google Scholar produced thousands of references in the high-level topics of OODA Loop, Data Analytics, and Deception just from abstract and keyword searches. To further refine the search, an eclectic approach where an intersection of keywords from OODA Loop, Data Analytics, and Deception was used within the abstract and keyword searches. Ranking based on frequency hits and academic sources (largely focused on journals or books) was used in the selection of the sources within this academic paper. What this research identified was three things:

- That there was a tremendous amount of depth in Data Analytics using ML and Deep Learning to recognize structured and unstructured fake news, disinformation, and patterns surrounding fake news and disinformation but very little focus on the timely response to counter it in order to affect the kill chain against these IO sources; without rapid responses, these IO sources very quickly spread.
- Although the OODA Loop has been employed simultaneously by friendly and adversarial forces and the timely response to use in reducing a variety of kill chain responses, little application had been applied to fake news and disinformation.
- No true workflow management has been employed to integrate the Data Analytics using ML and Deep Learning with the prosecution of fake news and disinformation in an OODA Loop flow in order to reduce the kill chain time.

### **5. Discussion, course of actions / recommendations and future research**

This research can serve as the pathfinder towards leveraging the OODA Loop with data analytics to detect and counter disinformation. With the increased use of ML, AI, and Deep Learning, the warfighter could have greater fidelity in their decision making associated with disinformation, fake news, and propaganda. Section 2.1 reveals a number of GT themes that could be explored more deeply. Improvements in each of these areas could improve how the U.S. and their Allies handle disinformation.

This research identified great strides in vertical research of Data Analytics with ML, AI, and Deep Learning that need to be fused into a workflow that has the following COAs that must be fused into a workflow:

- Fuses raw fake/disinformation and populates it in an endpoint/edge storage/compute technology,
- Leverages the existing structured and unstructured APIs (video, social media comment, and generic detection) to create focus areas for likely focus areas for fake/disinformation,
- ML, AI, and Deep Learning applied to prosecute the suspect fake/disinformation for I&W,
- After the Data Analytics of fake/disinformation, a ranking and scale of proliferation of the fake/disinformation can be applied to existing TTP in the form of COAs (e.g. a military Operation Order (OPORDER in the form of a playbook) that prescribes defensive as well as offensive maneuvers based on the fake/disinformation,
- The commander working the Area of Responsibility (AoR) then selects the actual TTP that will be used.
- Lastly, all of these workflows need to be mapped to the OODA Loop. This will allow Value Chain Analysis of each individual step of the workflow so that the kill chain can be decreased over time so that friendly OODA Loops are happening faster than adversary OODA Loops.

## 6. Conclusion

This research helps build the foundation upon what is most important in the tools required by offensive and defensive cyberspace operations personnel for decreasing the kill chain leveraging Data Analytics put into an OODA Loop workflow. Using this OODA Loop workflow allows friendly and adversary kill chains to be analyzed using Value Chain Analysis to shorten the Friendly OODA Loop, decrease time, and hopefully decrease adversary impact. Expanding on this research offers capabilities yet to be harnessed.

## References

- Alexander, J. and Smith, J., 2011. Disinformation: A taxonomy. *IEEE Security & Privacy*, 9(1), pp.58-63.
- Amoroso, E. G., 2011. *Cyber attacks protecting national infrastructure*. Burlington, MA: Butterworth-Heinemann.
- Andress, J. & Winterfeld., 2011. *Cyber warfare: Techniques, tactics and tools for security practitioners*. Boston, MA: Syngress/Elsevier.
- Angerman, W.S., 2004. *Coming full circle with Boyd's OODA loop ideas: An analysis of innovation diffusion and evolution*. Wright-Patterson AFB, OH: Air Force Institute of Technology, School of Engineering and Management.
- Arquilla, J., Ronfeldt, D., & Zanini, M., 1999. Networks, netwar, and information-age terrorism. In Lesser, I.O., Hoffman, B., Arquilla, J., Ronsfeldt, D., & Zanini, M. (Eds.), *Countering the new terrorism*. Santa Monica, CA: Rand Corporation.
- Bajaj, P., Kavidayal, M., Srivastava, P., Akhtar, M.N. and Kumaraguru, P., 2016, October. Disinformation in Multimedia Annotation: Misleading Metadata Detection on YouTube. In *Proceedings of the 2016 ACM workshop on Vision and Language Integration Meets Multimedia Fusion* (pp. 53-61).
- Bennett, M., & Waltz, E., 2007. *Counterdeception principles and applications for national security*. Norwood, MA: Artech House.
- Bey, M., 2018. Great Powers in Cyberspace: The Strategic Drivers Behind US, Chinese and Russian Competition. *The Cyber Defense Review*, 3(3), 31-36.
- Blodgett, D.E., Gendreau, M., Guertin, F., Potvin, J.Y. and Séguin, R., 2003. A tabu search heuristic for resource management in naval warfare. *Journal of Heuristics*, 9(2), pp.145-169.
- Bodström, T. and Hämäläinen, T., 2018, December. *A Novel Method for Detecting APT Attacks by Using OODA Loop and Black Swan Theory*. In International Conference on Computational Social Networks (pp. 498-509). Switzerland: Springer.
- Boyd, J., 1987. *A discourse on winning and losing*. Maxwell AFB, AL: Air University Press, Curtis E. LeMay Center for Doctrine Development and Education.
- CNSS., 2015. *Committee on National Security Systems (CNSS) No. 4009 Glossary*, 06 April 2015. Washington, D.C.: Washington Printing Office. Available at:<<https://www.cnss.gov/CNSS/issuances/Instructions.cfm>> [Accessed 23 Dec. 2017].
- Coolidge, O. E. & Sandoz, E., 1980. *The Trojan War*. New York, NY: Houghton Mifflin.
- Coram, R., 2002. *Boyd: The fighter pilot who changed the art of war*. New York, NY: Hachette Book Group.
- Czosseck, C. and Geers, K. eds., 2009. *The virtual battlefield: perspectives on cyber warfare*. Fairfax, VA: Ios Press, Fairfax, VA.
- Dulles, A., 2006. *The craft of intelligence: America's legendary spy master on the fundamentals of intelligence gathering for a free world*. Guilford, CT: Rowman & Littlefield.
- Ellul, J., 1973. *Propaganda: The formation of men's attitudes*. New York, NY: Vintage Books.
- Ferrara, E., 2017. Disinformation and social bot operations in the run up to the 2017 French presidential election. *First Monday*. 22(8) June 2017.
- Gabriel, R. A., 2004. *Genghis Khan's greatest general: Subotai the Valiant*. Westport, CT: Praeger Publishers.
- George, R. Z., & Bruce, J. B., 2008. *Analyzing intelligence: Origins, obstacles, and innovations*. Washington D.C.: Georgetown University.
- Glaser, B.G and Strauss, A.L., 1967. *The discovery of grounded theory: Strategies for qualitative research*. London, UK: AldineTransaction, A Division of Transaction Publishers.
- Hamilton, C.S.P. and Kreuzer, U.L., 2018. The Big Data Imperative. *Air & Space Power Journal*, 32(1), p.4-10.
- Heckman, K. E., Stech, F. J., Thomas, R. K., Schoker, B., & Tsow, A. W., 2015. *Cyber denial, deception and counter deception: A framework for supporting active cyber defense*. Switzerland: Springer International Publishing.
- Hosseini, H., Kannan, S., Zhang, B., & Poovendran, R., 2017. *Deceiving Google's Perspective API built for detecting toxic comments*. arXiv preprint arXiv:1702.08138.
- Jajodia, S., Subrahmanian, V. S., Swarup, V., & Wang, C., 2016. *Cyber deception: Building the scientific foundation*. Switzerland: Springer International Publishing.
- Joint Staff (JS)., 2012. *Joint Publication 3-13, Information Operations*. 27 November 2012 with change 01, 20 November 2014. Washington, D.C.: Washington Printing Office. Available at: <<http://www.jcs.mil/Doctrine/Joint-Doctrine-Pubs/3-0-Operations-Series/>> [Accessed 12 Dec 2018].
- Kumar, A. and Sangwan, S.R., 2019. Rumor Detection Using Machine Learning Techniques on Social Media. In International Conference on Innovative Computing and Communications (pp. 213-221). Singapore: Springer.
- Lemieux, V. and Smith, T.D., 2018, December. Leveraging Archival Theory to Develop A Taxonomy of Online Disinformation. In 2018 IEEE International Conference on Big Data (pp. 4420-4426).

**Jami Carroll**

- Levitin, D.J., 2017. *Weaponized lies: How to think critically in the post-truth era*. New York, NY: Penguin.
- Macintyre, B., 2010. *Operation Mincemeat: How a dead man and a bizarre plan fooled the Nazis and assured an allied victory*. New York: Harmony Books.
- Mahairas, A., & Dvilyanski, M., 2018. Disinformation—Дезинформация (Dezinformatsiya). *The Cyber Defense Review*, 3(3), 21-28.
- Malin, C.H., Gudaitis, T., Holt, T. and Kilger, M., 2017. *Deception in the Digital Age: Exploiting and defending human targets through computer-mediated communications*. London, UK: Elsevier.
- Mandia, K., 2017. *Disinformation: A primer in Russian active measures and influence campaigns*. Government Printing Office (GPO). Retrieved from [<https://www.govinfo.gov/content/pkg/CHRG-115shrg25998/html/CHRG-115shrg25998.htm>]
- Marwick, A. and Lewis, R., 2017. *Media manipulation and disinformation online*. New York: Data & Society Research Institute.
- Osinga, F.P., 2007. *Science, strategy and war: The strategic theory of John Boyd*. New York, NY: Routledge.
- Patrikarakos, D., 2017. *War in 140 characters: How social media is reshaping conflict in the twenty-first century*. New York, NY: Basic Books.
- Rabinovich, A., 2004. *The Yom Kippur War: The epic encounter that transformed the Middle East*. New York: Schocken Books.
- Rowe, N. C. & Custy, E. J., 2008. Deception in cyber attacks. In *Cyber warfare and cyber terrorism*. In Janczewski, L. & Colarik, A. M. (Eds.). Hershey, PA: Information Science Reference.
- Rushkoff, D., Pescovitz, D., & Dunagan, J., 2018. *The biology of disinformation: Memes, media viruses, and cultural inoculation*. Palo Alto, CA: Institute for the Future (IFTF).
- Russell, S. and Abdelzaher, T., 2018, October. The Internet of Battlefield Things: The Next Generation of Command, Control, Communications and Intelligence (C3I) Decision-Making. In *MILCOM 2018 IEEE Military Communications Conference* (pp. 737-742).
- Santos Jr, E., & Johnson Jr, G., 2004, August. Toward detecting deception in intelligent systems. In *Defense and Security* (pp. 130-141).
- Schechtman, G.M., 1996. *Manipulating the OODA Loop: The overlooked role of information resource management in information warfare*. Air Force Institute of Technology, School of Engineering and Management, Wright-Patterson AFB, OH.
- Shehabat, A. and Mitew, T., 2017. Distributed swarming and stigmergic effects on ISIS Networks: OODA Loop Model. *Journal of Media and Information Warfare*, 10, pp.79-109.
- Stein, G.J., 1996. Information attack: Information warfare in 2025. *2025 White Papers: Power and Influence*, 3, pp. 14-28.
- Svetoka, S., 2016. *Social media as a tool of hybrid warfare*. NATO Strategic Communications Centre of Excellence.
- Teyssou, D., Leung, J.M., Apostolidis, E., Apostolidis, K., Papadopoulos, S., Zampoglou, M., Papadopoulou, O. and Mezaris, V., 2017, October. The invid plug-in: Web video verification on the browser. In *Proceedings of the First International Workshop on Multimedia Verification* (pp. 23-30). ACM.
- Tompkins, J., 2018. *Disinformation Detection: A review of linguistic feature selection and classification models in news veracity assessments*. Unpublished work.
- Veerasamy, N., 2010. High-level Mapping of Cyberterrorism to the OODA Loop. In *Proceedings of the 5th International Conference on Information Warfare and Security*, Ohio, USA, pp.352-360.
- Weatherford, J., 2004. *Genghis Khan and the making of the modern world*. New York, NY Random House.
- Whaley, B., 2007. *Stratagem: Deception and surprise in war*. Boston: Artech House.
- Whaley, B. & McLaughlin, J., 2016. *Turnabout and deception: Crafting the double-cross and the Theory of Outs*. Annapolis, MD: U.S. Naval Institute.
- Whitehead, Y. G., 1997. *Information as a weapon: Reality versus promises*. Air University, Maxwell AFB, AL, School of Advanced Airpower Studies.
- Zager, R. and Zager, J., 2017. OODA Loops in cyberspace: A new cyber-defense model. *Small Wars Journal*, October, 20(11), p.33-42.
- Zakem, V., McBride, M. K., & Hammerberg, K., April 2018. Exploring the utility of memes for U.S. government influence campaigns. Retrieved from [https://www.cna.org/cna\\_files/pdf/DRM-2018-U-017433-Final.pdf](https://www.cna.org/cna_files/pdf/DRM-2018-U-017433-Final.pdf).

# A Framework for Improved Home Network Security

António Craveiro, Ana Oliveira, Jorge Proença, Tiago Cruz and Paulo Simões

Department of Informatics Engineering, University of Coimbra, Portugal

[aamcraveiro@gmail.com](mailto:aamcraveiro@gmail.com)

[asofiabo@gmail.com](mailto:asofiabo@gmail.com)

[jdgomess@dei.uc.pt](mailto:jdgomess@dei.uc.pt)

[tjcruz@dei.uc.pt](mailto:tjcruz@dei.uc.pt)

[psimoes@dei.uc.pt](mailto:psimoes@dei.uc.pt)

**Abstract:** modern home networks constitute a diverse ecosystem of devices and services, whose management is mostly handled by means of specific service or device provider mechanisms, with only a minority of customers possessing the technical skills required to deal with such tasks. This means that, when it comes to security, most users often exclusively rely on endpoint protection mechanisms (such as anti-virus) to handle their basic needs. This is mostly due to a basic assumption that considers the home LAN to be a safe environment, with most threats coming from the outside – a premise that underlies most of the research about the subject. However, this perspective predates the widespread adoption of mobile appliances, smart devices and related services, some of which may be abused and/or compromised to spy and exfiltrate sensitive information from their owners. Overall, one cannot consider the home LAN to be a safe environment anymore, as potential vulnerabilities and information leaks may come from within. This paper addresses the initial phase of the development plan for a security appliance for Home LANs, capable of providing seamless protection for the ever-growing ecosystem of devices and services which are common on such networks, while also addressing the specific requirements imposed by multi-tenancy and cohabitation with third-party services. By monitoring and controlling network traffic flows on a per-device basis, we intend to make it possible to characterize typical traffic profiles for each type of device, to detect anomalies and/or information leaking events. Eventually, this capability may be leveraged to provide to control device-to-device communications in order to block unwanted interactions between equipment and external sites. By moving beyond the last mile and into the customers' doorsteps, the proposed solution intends to provide comprehensive and contextual security monitoring on a per-user basis, embedded at the residential gateway or eventually resorting to a specialized appliance deployed in the home LAN.

**Keywords:** home network security, distributed security, smart homes, home appliances

---

## 1. Introduction

In the last years the number of devices connected to the home networks has drastically increased, with the advent of products like smart TVs, smart speakers, smart meters, smart appliances and other smart home IoT boxes. The proliferation of such devices, in an ecosystem which is often poorly managed but holds very sensitive personal information, raises considerable concerns from the security point-of-view. Recent incidents have shown that these devices cannot be trusted, since their manufacturers often abuse basic user privacy rights to begin with. Moreover, even when manufacturers do behave correctly, the devices are designed with such poor security that they easily become the target of malicious attacks from third-parties, providing a valuable entry point to the wealth of personal information available in the home network, while becoming part of large botnets used for high-profile internet attacks. Those risks are amplified by the fact that the home network is often poorly managed by nature (Cruz et al, 2015), without internal segmentation or restrictions on device-to-device communications, and even without proper security monitoring mechanisms at internal level. Moreover, ownership and management of these devices is often fragmented between different parties (e.g. home owner, other home users, energy utilities, local service providers such as telecommunications operators, and cloud service providers such as Google).

This situation provides the rationale to propose a platform for home LAN security monitoring, providing the means to control which devices may send which information to whom. This platform would monitor in and out traffic, not just to search for malicious activities, like a classic Intrusion Detection System (IDS), but also to control the information exchanged between devices and their manufacturers. The idea is not allowing our personal data concerning our activities, to be sent out. This is a challenging task because first we need to know what that information is, and we cannot count with the manufacturers to help us. So, the first step is monitoring the network to see the kind of traffic that flows and decide what to allow and what to block. After that, we need to produce a set of rules to control the flow of in and out data in an automated way. Also, we need to find a suitable hardware platform to run all the needed software. The device to perform this task cannot be too big nor power

consuming because is to be used in home networks. On the other hand, it must be powerful enough to not introduce noticeable lags in data flow, while keeping a fairly filtering capability of data packets.

This project intends to create such a device for home networks, using a suitable Single Board Computer (SBC) platform for its Proof-of-Concept implementation. As such, this device will be setup between the home router and the rest of the network, so it will be capable of analysing all the in and out traffic. For this purpose, we have chosen the Raspberry Pi B 3+, thanks to its widespread availability, good performance and compact form factor. The OS will be Linux based, and the software to monitor the network will include both a custom solution for anomaly detection, working together with an available IDS platform.

The organization of this paper is as follows: Section II provides an overview of related work addressing the specific security needs of home LANs, also including an overview of relevant approaches and a small survey about the usage of the Raspberry Pi SBC for security purposes. Section III describes the specific requirements for the development of the proposed solution, with Section IV providing insights about the Proof-of-Concept (PoC) prototype implementation. Section V discusses specific implementation details for the network traffic capture and processing modules and, finally, Section VI concludes this paper.

## **2. Literature review**

Within the home LAN domain, the introduction of automation by means of the IoT paradigm (Amri et al, 2018) can bring several benefits, but also has implications regarding security issues and the privacy protection of objects and users (Panagiotis et al, 2019). Such developments provide a solid argument for providing IDS mechanisms even for small networks (Microsoft, 2014), especially because the latter ones are usually the most susceptible to be abused by all sorts of threats.

In this perspective, various research has been made around packet inspection and traffic analysis using packet sniffers (Asrodia and Patel, 2012) (Qadeer, 2010) these studies are usually intended to be later on used for intrusion detection. In (Rosa et al, 2015) (Cruz et al, 2015) is introduced a new framework to home security trying to fill the gap that currently exists between ISP security approach and home network security. By now these are two completely different “worlds” that do not communicate with each other. This “no one’s land” facilitates the attacks on home networks, because the average user is completely unaware of the dangers. To overcome this, the proposed framework provides a distributed IDS system with residential gateways and ISP IDS systems so they can more easily find malicious attacks.

(Ahmed et al, 2018) proposed a methodology for detecting anomalies that may develop into an attack, which may be useful for our security appliance. In (Ye et al, 2018), a new approach is proposed using emergent technology, in this case software defined networks, to address the problem of monitoring encrypted network traffic. The proposed method uses machine learning algorithms. In (Sperling et al, 2017), is proposed another approach to IDS systems using DNS. The concept is analyzing DNS requests in order to find suspicious sites request. This is also an alternative approach that may help in protect a home network.

In (Aspernäs et al, 2015) the ability for the Raspberry Pi to perform as an IDS is tested. The test comprised two models: Raspberry Pi model B+, and the latest model Raspberry Pi 2 Model B. The IDS software used was *Snort* (Snort, 2018). The result is that the latest model, Raspberry Pi 2 Model B performs better, as would be expected, but also it is the only one that is really able to work as an IDS. In (Mantere et al, 2013) an analysis is made on the specificities of network traffic on ICS networks, and how that affects an IDS performance. (Poonia et al, 2017) also evaluated the need for network IDS in home networks, as well as the capability of the Raspberry Pi to perform these duties. The tests were made with a Raspberry Pi Model 3, a Raspberry Pi Model B+ and a Raspberry Pi Model B, all with the same configuration. The test results showed that using a Raspberry Pi Model 3 with Snort IDS provides a good compromise for a Home network IDS.

In (Kyaw et al, 2015) a comparison is made between *Snort IDS* and *Bro IDS*, as an IDS to be used in a Raspberry Pi 2 for home or small office use. Also, the capability of the Raspberry Pi to accomplish the task was tested. The result is that Snort IDS performs better, but the RaspberryPi2 may crash if there are a lot of traffic on the network. In (Sforzin et al, 2016), another study using Raspberry Pi and *Snort*, is presented reinforcing the opinion that this is a good option for a small sized network security implementation. This paper also proposes a network architecture that can take advantage of several IDS devices on the same network, which is an interesting area

of research. In (Zitta et al, 2018), another IDS is presented, *Suricata*, and its performance is tested in a Raspberry Pi 3. It is another alternative to the already aforementioned IDS's *Snort* and *Bro*. This IDS also performs adequately on the Raspberry Pi platform, although it needs custom rules to be added.

It is important to note that most of the investigation done so far, is about IDS systems, while this paper is about creating a device to control what is going out of a home network, generated by the appliances in the home network. As so, there are no software readily available.

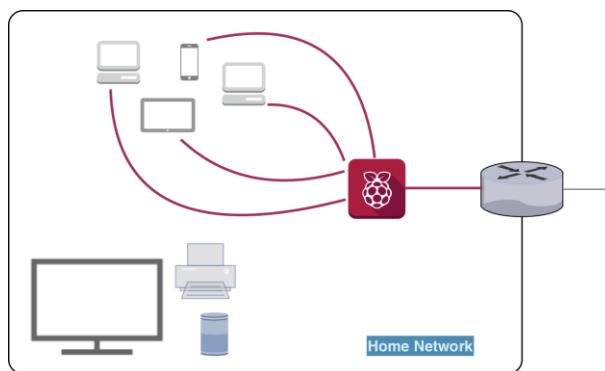
### 3. Platform requirements

To monitor and block outgoing traffic it is needed what could be called a "reverse IDS"— a device whose purpose is to analyze the traffic generated by the home LAN devices. Considering the fact that even normal IDS roles are not supported by most domestic routers, it comes as no surprise that support for this kind of solution is absent from most home LAN equipment. On the other hand, home routers are "black boxes" so it is not possible to easily add new features. This justifies the introduction of a dedicated security appliance, whose main requirements are:

- It must be inexpensive;
- It must be power efficient;
- It must have a compact form factor;
- It must be easily configurable;
- It must be able to perform traffic monitoring with minimal disruption;

While these requirements may seem somewhat contradictory at first, they are in fact compatible (even for a PoC), especially considering the envisioned deployment domain. Home LANs have fewer requirements in terms of network capacity, which fit within the computing capabilities of several Single-Board Computers.

The last step of the setup is to deploy the device within the home LAN (see Figure 1). It has to be placed like a network-based IDS, between the home gateway and the rest of the network. To do that it will need to have two network interfaces. Also, the wireless traffic needs to be analyzed so the device must have a wireless network interface and be to be configured as wireless AP.



**Figure 1:** Deployment of the proposed solution within the home LAN

Regarding software, the OS will be a Linux distribution, and for packet analysis what we need is something similar to an IDS only with a slightly different set of rules. In fact, what the software must do is to look at the packets in and out of the network towards the internet. This is something that is currently done by existing IDS with the only difference been the fact that the IDS look for incoming traffic rather than outgoing. What we will try is to apply the IDS to outgoing packets, filtering the content that the home appliances will send somewhere.

To get to the right set of filtering rules, the first step is to analyze the traffic and try to understand what appliances send out our personal information and capture some packets. Next, is time to analyze the captured packets and try to understand their origin. This is a fundamental step, this analysis will provide the basis for writing the software filtering rules, as well as defining a reference traffic model for a normal state. For the latter aspect, the use of a conventional NIDS will be complemented with a custom module, which is capable of extracting traffic flow information from the network captures, in order to establish a security baseline.

## 4. Proof-of-concept prototype

Despite the merits of signature-based approaches (whose benefits are leveraged by the inclusion of a Snort IDS), our purpose is to conceive a solution capable of (semi)autonomously establishing a set of baseline descriptors for the home LAN environment, which will later serve as a reference for security detection procedures. This will be achieved by creating profiles based on the traffic captured from the network. Such profiles correspond to a set of characteristics determined for each device, whose purpose is to identify the communications allowed for that device: with which IP address can the device communicate with, through which ports it can do it, etc.

Profile characterization is based on patterns found when looking at specific features of the captured data such as timestamps and bandwidth. This concept is the core of our solution and what distinguishes it from other security implementations for IoT devices.

### 4.1 Support platform (hardware and system software)

To implement the device, our choice fell on a Raspberry Pi 3 B+, the most recent version of the Raspberry family. The most recent model has the following characteristics (Raspberry Pi, 2018):

- Broadcom BCM2837B0, Cortex-A53 (ARMv8) 64-bit SoC @ 1.4GHz, with 1GB LPDDR2 SDRAM
- 2.4GHz and 5GHz IEEE 802.11.b/g/n/ac wireless LAN, Bluetooth 4.2, BLE
- Gigabit Ethernet over USB 2.0 (maximum throughput 300 Mbps)

This new model of the Raspberry Pi offers adequate computing power, also being a small, inexpensive device, with the bonus of being able to look neat on a household room, thanks to the great variety of cases available for purchase. As the Raspberry Pi came with only one ethernet adapter so a second one had to be added. The solution was to use an USB-Ethernet adapter using one of the available USB ports. For the rest of the setup, the Raspberry Pi came with what was needed, specially the wireless network card.

Regarding the system software, the chosen OS was the ArchLinux Linux distribution (Archlinux, 2018). This distribution was chosen because it is lightweight, having only the core OS structure, and allowing the user to customize the system to their own needs while avoiding wasting resources with non-relevant services. On top of that it will be installed the Snort IDS [15]. Snort works by monitoring and dissecting network traffic data, matching it against a set of rules, trying to find malicious patterns. These rules can be customized, so it may be possible to improve the existing ruleset with custom rules. Although there are other alternatives, like Bro or Suricata, Snort was chosen because is stable, highly configurable, and performs well on Raspberry Pi. The PoC will also include a custom solution for flow capture and analysis, which will be complementary to the operation of the Snort IDS, described in Section V.

The testing network consists of a home gateway with wireless AP function disabled, the security appliance PoC device configured as a wireless AP, a home switch to allow all the wired links to pass through the appliance (for traffic analysis purposes), a home tv box, two televisions of different brands, and a BluRay reader (see Figure 1). It must be noted that, despite the fact that a home network also includes computers, they are not shown because they are not relevant for this work.

### 4.2 PoC network setup

In our homes, we have two types of devices: the personal ones and the other non-standard devices that connect wirelessly to the network and have the ability to transmit data, also known as IoT devices or smart objects. Any device with an Internet connection can be compromised, which enables third-party vendors being capable of obtaining sensitive data when we connect to their devices or use their applications. Ideally, we want to monitor every device in our network, but the architecture of our initial solution is only intended to capture the traffic of personal devices like smartphones, tablets, and computers.

In order to develop an affordable solution, we used a Raspberry Pi 3 and due to its in-built wireless features, we configure it to act as an Access Point (Figure 1). It was configured as both a wired and wireless bridge – for the latter, the configuration was established as such (Poonia et al, 2017):

```
Install packages (last version and required software):
sudo apt-get update
```

```
sudo apt-get upgrade
sudo apt-get install hostapd
sudo systemctl stop hostapd
```

Edit the file **/etc/dhcpcd.conf**, to configure the **wlan0** interface:

```
interface wlan0
static ip_address=192.168.x.w/24
nohook wpa_supplicant
```

Restart the dhcpcd daemon:

```
sudo service dhcpcd restart
```

Setup hostapd (edit **/etc/hostapd/hostapd.conf**):

```
interface=wlan0
driver=nl80211
ssid=NameOfNetwork
hw_mode=g
channel=7
wmm_enabled=0
macaddr_acl=0
auth_algs=1
ignore_broadcast_ssid=0
wpa=2
wpa_passphrase=PasswordOfTheNetwork
wpa_key_mgmt=WPA-PSK
wpa_pairwise=TKIP
rsn_pairwise=CCMP
```

After having a device that creates a wireless local area network (WLAN) we had to pass all traffic between the WLAN and the Ethernet interface. We set up a bridge thus making possible for the access point to provide wireless connections: anyone connected to it can access the Internet. It follows its configuration:

Install packages:

```
sudo apt-get install hostapd bridge-utils
sudo systemctl stop hostapd
```

Stop IP allocation for the **eth0** and **wlan0** interfaces (add to **/etc/dhcpcd.conf**):

```
denyinterfaces wlan0
denyinterfaces eth0
```

Add a new bridge and connect the network ports

```
sudo brctl addbr br0
sudo brctl addif br0 eth0
```

Add the configuration for the network bridge (**br0** – add to **/etc/dhcpcd.conf**):

```
Auto br0
iface br0 inet manual
bridge_ports eth0 wlan0
```

Every packet sent and received by any device connected to the Raspberry Pi can now be parsed as a result of the bridge deployed between the wireless interface and the Ethernet connection, as shown in Figure 1.

In the future, this configuration will be improved to cover a bigger number of devices, by configuring a bridge between the internal and an external ethernet adaptor to enable bridging on the cabled network segment. Also, we intend to automate the reconfiguration of the domestic router DHCP server, in order to disable it and offload this task to the security appliance, which will be able to pass its own IP address as the default gateway, forcing the network traffic for all dynamically configured devices to be steered through it.

## 5. Network traffic capture and processing

Once the Access Point capabilities of the PoC device became operational, we were in conditions to further proceed with its development. Apart from the Snort component (which is a single process component), we have two further concerns: the capture of packets and the data analysis. It is expected that these two main tasks will have different execution times, with the analysis taking longer than the traffic capture. For this reason, we established that each task should be treated by a single process, taking advantage of the fact the Raspberry Pi 3 has a 1.2 GHz 64-bit quad-core processor, to use CPU affinity. CPU affinity enables binding a process or multiple processes to a specific CPU core in a way that the process(es) will run from that specific core only. By dedicating

one CPU core to a particular process it is possible to minimize competition with other tasks, ensuring maximum execution speed for that process, with obvious performance benefits.

Setting the CPU affinity is easily done with some lines of code in C:

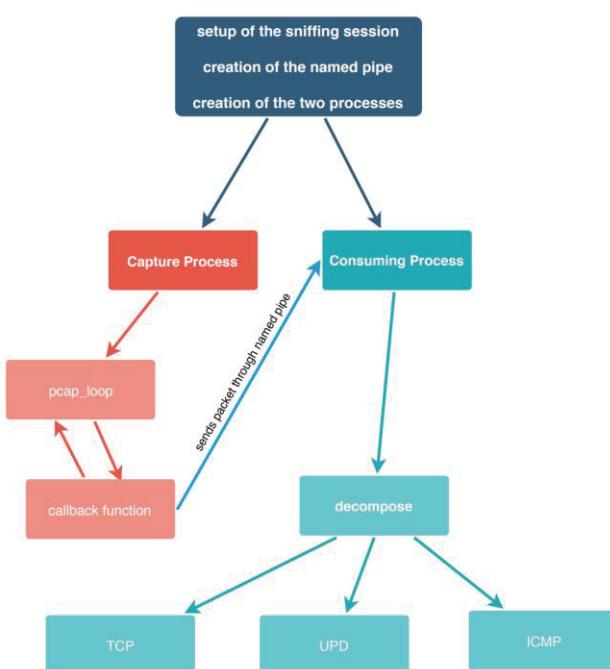
```
#define _GNU_SOURCE
#include <sched.h>
if (core_id < 0 || core_id >= num_cores)
    return EINVAL;
cpu_set_t cpuset;
CPU_ZERO(&cpuset);
CPU_SET(core_id, &cpuset);

int pid = getpid();
sched_setaffinity(pid, sizeof(cpu_set_t), &cpuset);
```

Where `coreid` is the corresponding integer to the core which we want to assign our process ( $0 < \text{core\_id} < \text{number of cores - 1}$ ). Bear in mind that Linux probably runs by default some tasks in core 0.

## 5.1 Managing parallelism

To benefit from core affinity there is the need to separate tasks among separate processes. As such, the workload for capturing and processing network traces was split among two different processes (one for handling network traffic capture and another one for trace analysis), in order to better decouple tasks and manage buffering. However, by pursuing this strategy there is the problem of making the two processes communicate in a robust and efficient way.



**Figure 2:** Trace capture and analysis solution

The first implemented approach to solve this problem consisted on a rotating file pool, using predetermined number of files, with a maximum capture size for each one. When the maximum capture size was reached, the producer process (which captures network traffic) would move to the next file. This approach requires a flow control mechanism to avoid concurrency problems. As such, a simpler solution using named pipes was adopted (see Figure 2). Also known as FIFO, a named pipe is one of the methods for inter-process communication. A simple `mkfifo` call creates a special FIFO file and any process can open it for reading or writing.

## 5.2 Network traffic capture

Libpcap (Carstens, 2002) is a library used to capture packets directly from the network adapter, providing the packet capture and filtering capabilities required for the capture module. Figure 2 shows how the solution is

structured, i.e how each functionality is organized in the code. We start by doing the required initialization procedures:

### Sniffing session

**pcap\_open\_live()** is used to obtain a packet capture handle to look at packets on the network. One of its parameters is the network device to be open. We are sniffing the bridge that was previously created on the Raspberry Pi. **pcap\_lookupnet()** looks for the network mask, **pcap\_compile()** compiles a string into a filter expression and finally **pcap\_setfilter()** applies the filter. The filter expression restricts the type of packets we want to capture. Currently, our program is capturing all the IP traffic but more strict rules could be applied, for example, if we want we can sniff a specific port or protocol (e.g. the expression “port 53 UDP” would filter all DNS traffic).

### Named Pipe

Create the named pipe:

```
#define myfifo 'path/to/fifo/name'
//mkfifo(<pathname>, <permission>
mkfifo(myfifo, 0666);
```

Open for write/read:

```
int fd[2];
fd[0] = open(myfifo, O_WRONLY);
fd[1] = open(myfifo, O_RDONLY);
```

### Processes

After executing **fork()**, each process calls its corresponding function where the CPU affinity is set and the FIFO is open either for writing or for reading depending on the process. Then the procedure is different for each function. We will continue the describe the capture process in the present section while the consuming process will be discussed in the next subsection

The capture initializes when pcap enters its primary execution loop that processes packets from a live capture. The responsible function to do so is **pcap\_loop()**. One of its arguments is the name of a callback function that is called every time a packet that meets our filter requirements is sniffed. The callback function can be implemented to do whatever we need but has a prototype defined in order to **pcap\_loop** be able to deal with it. These two functions are defined as follows:

```
int pcap_loop(pcap_t *p, int cnt, pcap_handler callback, u_char *user)

void callback(u_char *args, const struct pcap_pkthdr *header, const u_char *packet);
```

The first argument of **pcap\_loop** is the session handle followed by an integer to define how many packets should be sniffed before returning (we are interested in capture all of them so we set the maximum number as -1 meaning that we want to continually capture until an error occurs), the third argument is the name of the callback function and the last one is reserved for the arguments of the callback function (**NULL** in our case). The first argument of the callback function corresponds to the last argument of **pcap\_loop** the second is a header provided by pcap that contains information about when the packet was sniffed, how large it is, etc. The actual packet is received as a **u\_char** pointer (last argument) that points to the first byte of the block of data containing the entire packet. The packet is in fact a collection of structures (example in Table 1).

The only purpose of our callback function is to send the sniffed packet (and the header given by pcap) through the named pipe (Figure 2). Whenever a packet is received by the capture process, it is passed it to the consuming process so that it can be analysed. However, this is done with an auxiliary structure (mentioned below) where we save the pcap header and the packet as a byte array because we can't send the **u\_char** pointer through the named pipe (keep in mind that we have two different processes communicating).

```
struct info {
    struct pcap_pkthdr header; u_char packet[SIZE];
};
```

### 5.3 Data analysis

As aforementioned, the consuming process sets the CPU affinity. After doing so, it enters an infinite loop that is always reading from the FIFO and every time it receives a struct “info” it sends it to a decompose function. We are not concerned with the packet payload i.e, we do not analyze the transmitted data within the actual message. Each packet has an IP header with useful information: source and destination IP addresses, time to live, checksum the protocol. The IP and Ethernet headers are defined for every IP packet but as they are the only thing packets have in common, when the decompose function receives one it does a filtering based on the protocol: TCP, UDP, or ICMP. Table 1 showcases the layout of a TCP packet in memory.

**Table 1:** TCP packet in memory

Variable	Location (in bytes)
sniff_ethernet	X
sniff_ip	X + SIZE_ETHERNET
sniff_tcp	X + SIZE_ETHERNET + IP header length
payload	X + SIZE_ETHERNET + IP header length + TCP header length

Having the pointer’s address, set in the previous table as the value X, we can easily access the other headers and therefore all of the information available in the packet. For example, the length of the Ethernet header (SIZE\_ETHERNET) is 14 and once we have the packet (**u char \*packet**) we can access all the information:

Get the IP header:

```
struct sniff_ip *ip;
ip = (struct sniff_ip*) (packet + SIZE_ETHERNET);
```

Determine the protocol:

```
#define IP_HL(ip) (((ip)->ip_vhl) & 0x0f)
int size_ip;
size_ip = IP_HL(ip)*4;
u_char protocol = ip->ip_p;
```

Get the corresponding header:

```
if (protocol == IPPROTO_TCP) {
    struct sniff_tcp *tcp;
    tcp = (struct sniff_tcp*) (packet + SIZE_ETHERNET + size_ip);
}
```

Default values like Ethernet header’s length and all of the necessary structures are defined in the code. This paper only covers the identification of TCP flux. This topic is presented in the following subsection.

#### TCP flow parsing

TCP is a transport layer protocol used by various services such as FTP, SMTP, HTTP and many others. This byte-stream-oriented protocol sets up a connection to the receiver using the three-way handshake method and then sends the data in segments. Sequence numbers are attributed to the packets in order to keep track of the bytes sent in each direction.

The TCP header is defined as follows:

```
struct sniff_tcp {
    u_short th_sport;
    u_short th_dport;
    tcp_seq th_seq;
    tcp_seq th_ack;
    u_char th_offx2;
    #define TH_OFF(th)
    u_char th_flags;
    #define TH_FIN 0x01
    #define TH_SYN 0x02
    #define TH_RST 0x04
    #define TH_PUSH 0x08
    #define TH_ACK 0x10
    #define TH_URG 0x20
    #define TH_ECE 0x40
    #define TH_CWR 0x80
    #define TH_FLAGS (TH_FIN|TH_SYN |TH_RST|TH_ACK|TH_URG|TH_ECE |TH_CWR)
    u_short th_win;
```

```
    u_short th_sum;
    u_short th_urp;
};
```

We have information about the source port, destination port, sequence number, acknowledgement number, data offset, flags, window, checksum and urgent pointer. These parameters about the TCP packet will most likely be used in a further analysis, but by now, a fairly simple identification of TCP flux can be done by the combination of the Source IP address, the Destination IP address and TCP port numbers (both source and destination).

We maintain a record of the TCP Streams in a linked list. Each node is defined by the four parameters aforementioned (source and destination IP addresses and source and destination ports).

```
typedef struct node_tcp {
    char from[100];
    char to[100];
    int src_port;
    int dst_port;
    struct node_tcp *next;
} tcp_stream;
```

A method for searching the list was implemented such that, if we look for a node with a specific combination of addresses and ports and that node already exists then it is put in the first position. Whenever we received a large number of packets of the same stream in a row, the time to search/access the node is the lowest possible.

## 6. Conclusion and future work

This work explores a new perspective about home network security. Here, we are not yet trying to avoid attacks from the outside, although it stills a major concern, but rather figuring out if and how, our home appliances are giving our personal information to third parties without our permission. To achieve this, a thoroughly network analysis must be performed in order to understand what kind of information is sent from our home appliances to the outside. When we will be able to define the pattern behavior it will be possible to implement a traffic analyzer to enforce that same behavior and avoid having our personal information to be sent out.

For the aforementioned purposes, the establishment of a reliable capture and interpretation of packets is the first step to design our security appliance. This makes it possible to perform a concrete analysis of each packet, more specifically about each TCP communication. The main goal is to find meaningful patterns that allow us to establish a set of rules for each device. The ultimate objective of this effort will be the creation of a dynamic process capable of detecting a new device in the network and active the proper “profile” for it thus enabling a set of communication rules based on the previous analysis.

Further developments of this platform will also include support for virtualized home gateways (Cruz et al, 2013), for which the authors envision an integrated approach, using Virtualized Network Functions (VNFs) to host the functionality within virtualized service instances.

## Acknowledgements

This work was partially funded by the "Mobilizador 5G" P2020 Project (project 10/SI/2016 024539) and the ATENA H2020 EU Project (H2020-DS-2015-1 Project 700581).

## References

- Ahmed, A. A. and Kit, Y. W. (2018) Collecting and Analyzing Digital Proof Material to Detect Cybercrimes, Faculty of Computer Systems & Software Engineering, Universiti Malaysia Pahang.
- Amri, Y. and Setiawan, M. A., (2018) Improving Smart Home Concept with the Internet of Things Concept Using RaspberryPi and NodeMCU, Department of Informatics, Universitas Islam Indonesia, IOP Conf. Series: Materials Science and Engineering 325.
- Arch Linux for Raspberry Pi, [online], <https://archlinuxarm.org/platforms/armv8/broadcom/raspberry-pi-3>
- Aspernäs, A. and Simonsson, T. (2015) IDS on Raspberry Pi – A Performance Evaluation, Diploma Thesis
- Asrodia P. and Patel H. and Network Traffic Analysis Using Packet Sniffer, 2012
- Carstens, T. Programming with pcap, 2002.
- Cruz, T. and Simões, P. and Monteiro, E. and Bastos, F. and Laranjeira, A. (2015) Cooperative security management for broadband network environments. Security Comm. Networks, 8: 3953– 3977. doi: 10.1002/sec.1313.

- Cruz, T. and Simões, P. and Reis, N. and Edmundo Monteiro and Bastos, F. and Laranjeira, A. , "An Architecture for Virtualized Home Gateways ", in IM 2013 (IFIP/IEEE International Symposium on Integrated Network Management), 2013.
- Kyaw, A. K. and Chen, Y. and Joseph, J. (2015) Pi-IDS: Evaluation of Open-Source Intrusion Detection Systems on Raspberry Pi 2, IEEE.
- Mantere, M. and Sailio, M. and Noponen, S. (2013) Network Traffic Features for Anomaly Detection in Specific Industrial Control System Network, VTT Technical Research Centre of Finland.
- Microsoft (2014), Noticing and Responding To Network-Borne Attacks, [online], [https://docs.microsoft.com/en-us/previous-versions/tn-archive/cc723457\(v=technet.10\)](https://docs.microsoft.com/en-us/previous-versions/tn-archive/cc723457(v=technet.10)).
- Panagiotis I. Radoglou Grammatikis and Panagiotis G. Sarigiannidis and Ioannis D. Moscholios, Securing the Internet of Things: Challenges, threats and solutions, Internet of Things, Volume 5, 2019, Pages 41-70, ISSN 2542-6605, <https://doi.org/10.1016/j.iot.2018.11.003>.
- Poonia, P. and Kumar, V. and Nasa, C. (2017) Performance Evaluation of Network based Intrusion Detection Techniques with Raspberry Pi - a Comparative Analysis, International Journal of Engineering Research & Technology (IJERT), ICCCS.
- Qadeer M. A., Iqba and A., Zahid and M., Siddiqui and M., Network Traffic Analysis and Intrusion Detection using Packet Sniffer, 2010
- Raspberry Pi specifications, [online], <https://www.raspberrypi.org/products/raspberry-pi-3-model-b-plus/>
- Rosa, L. and Alves, P. and Tiago Cruz and Simões, P. and Edmundo Monteiro , (2015) A Comparative Study of Correlation Engines for Security Event Management, in In Proc. of 10th Int. Conf. on Cyber Warfare and Security (ICCWS-2015). ISBN: 978-1-910309-98-8 ISSN: 2048-9897.
- Sforzin, A. and Conti, M. and Mármlor, F. G. and Bohli, J-M.. (2016) RPiDS: Raspberry Pi IDS A Fruitful Intrusion Detection System for IoT, 2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress.
- Snort website [online], <https://www.snort.org/>
- Sperling, T. L. von and Filho, F. L. de C. and Júnior, R. T. de S., Martins, and L. M. C., and Rocha, R. L., (2017) Tracking intruders in IoT networks by means of DNS traffic analysis, 2017 Workshop on Communication Networks and Power Systems (WCNPS).
- Ye, F. and Qian and Y. (2018) DataNet: Deep Learning based Encrypted Network Traffic Classification in SDN Home Gateway, IEEE Access.
- Zitta, T. and Neruda and M., Vojtech, L. and Matejkova, M. and Jehlicka, M. and Hach, L. and Moravec, J. (2018) Penetration Testing of Intrusion Detection and Prevention System in Low-Performance EmbeddedIoT Device, 18th International Conference on Mechatronics, December.

# Hackers and the Military: How to Recruit and Manage Hidden Talents?

Didier Danet

Saint-Cyr Military Academy and Research Centre GEODE, France

[didier.danet@st-cyr.terre-net.defense.gouv.fr](mailto:didier.danet@st-cyr.terre-net.defense.gouv.fr)

**Abstract:** Recruiting, integrating and retaining talented cyber warriors are a major challenge for countries that consider cyber defense as a priority. We suggest in this paper that the gap between Hacking Culture and Military Culture requires a special attention from armed forces. A renewed recruitment and promotion policy needs to be designed and implemented in order to discover and encourage cyber warriors "hidden talents" to join the forces and forge a career as military officers.

**Keywords:** cyber defense, cyber warriors, hacking culture, military culture, cultural gap, integration strategy, hidden talents, promotion, selection processes, career perspectives

---

## 1. Introduction

The French government has been considering cyber threats as a major matter of concern for more than a decade. In spite of significant political changes during this period of time, a comprehensive and sustained effort is prevailing to face the whole destructive actions which take place in cyberspace: criminal use of the Internet (cyber crime); propagation of false information or propaganda; espionage for political or economic ends; attacks on critical infrastructure (transport, energy, communication, etc.) This led the French government to define and implement national policies and strategies of cyber security and cyber defense. (Luijif, Besseling and De Graaf, 2013; Watanabe, 2013; Haddad, 2017; Gautier, 2018; Poupard, 2018)

The rise of French cyber capabilities demands the implementation of an ambitious recruitment policy. In 2025, an additional 1,000 people are expected to be involved in cyber defense units.<sup>1</sup> According to the French Minister of Armies, Ms. Florence Parly, "this new population of cyber-warriors requires a new definition of our sociology and our Human Resources policies, by integrating, perhaps also inventing, educational and training programs and even completely new recruitment processes." (Parly, 2019) To some extent, this quotation is intriguing. The French armed forces currently have a total manpower (active personal, Gendarmerie Nationale non included), which exceeds 270,000. Why should an additional 1,000 workforce (0,4%) cause such a renewal of human resources policy: "new sociology", "inventing new educational programs and recruitment processes"?

Recruiting people into the All Volunteer Forces has been a challenge in developed countries for a long time. (Faris, 1984; Asch, Hosek and Warner, 2007; Tresch and Leuprecht, 2010; Manigart *et al.*, 2017; Asch and Warner, 2018) "Military is more than just a job" according to the famous book by Charles Moskos. (Moskos and Wood, 1988) Armies need to deploy active policies and strategies in order to attract and retain the required number of motivated and talented people who will serve in every position or ranks of the institution. Very recently, a leaked report by the German Ministry of Defense discussed adding large numbers of foreigners to the Bundeswehr in order to face a major labor shortage. (Schuetze, 2018) As far as cyber defense is concerned, the challenge seems to be a concern all over Europa (Townsend, 2018) and the US (Samuelsohn, 2015). The kind of talented people who are targeted to become cyber-warriors does not fit with the usual profiles of people who volunteer for the Military. A cultural gap exists between the "Hacking Culture" and the "Military Culture" giving birth to the following dilemma: how to fully integrate hackers in the structure and the culture of the Military but, at the same time, preserving their initial "Hacking culture" which is the potting soil in which they draw their specific talents and skills? In other words, can cyber-warriors be both good soldiers and good hackers in the same time? This is mainly why a "new sociology" of human resources and "innovative training programs or recruitment strategies" is needed.

This paper intends to look at the issues French Armed Forces, and European forces as well, need to face in order to attract and retain "hackers" in Cyber Defense Units. Two complementary questions are scrutinized.

- 1. How to recruit and integrate Hackers in the Military in spite of the gap that is widely recognized between the two cultures?

---

<sup>1</sup> By 2025 (term of the Military Planning Law), more than 4000 cyber warriors should contribute to French Cyber defence.

- 2. How to retain and promote Hackers in the Military in spite of their unconventional profile and the risk that usual selection processes ignore their hidden talents?

## **2. Facing the cultural gap**

### **2.1 Beware of the gap**

According to many authors, major differences exist between Hackers culture<sup>2</sup> (Riemens, 2002; Maxigas, 2014; Kaufmann, Rios Luque and Glassey, 2016) and military culture. (Dunivin, 1994; English, 2004; Soeters, Winslow and Weibull, 2006; Winslow, 2006; Wilson, 2008) In order to illustrate this consensus, three noteworthy approaches may be pointed out which drive to the same conclusion.

- Leenen et al. outline three traits that characterize “Cyber Culture”. (Leenen et al., 2018) “Cyber Culture tends to have low power distance relationship that allow subordinates to challenge their managers, exhibits low levels of uncertainty avoidance and values individualism”. Military Culture relies on very different or even opposite principles. (Soeters and Recht, 1998; Soeters, Winslow and Weibull, 2006) The level of power distance and uncertainty avoidance are considered high. Discipline is a foundational in military organizations and they tend to be rule oriented in that sense that strict processes need to be followed in action. “Esprit de corps” and unity prevail on individuals. Cyber culture and military culture also differ on two more dimensions. In the Military, the logic of decision-making is “based on facts to solve problems and thus tends to have a shorter orientation.” Restraint is dominant which means that the group exercises a strong control over their impulses and desires in order to keep focused on the mission. Masculinity dominance is the only common feature of the two cultures.
- Müller and Ulrich use a semi-automated text analysis “for understanding the values and assumptions that are expressed in documents” issued from 25 years of publications on “Cult of the Dead Cow”, a website dedicated to hacking activities. (Müller and Ulrich, 2015) According to Müller and Ulrich, the Adhocracy culture type, mixed with elements of the Market culture type, dominates the hacker culture. Hackers value creativity, innovation, and experimentation rather than compliance to pre-established rules. They are acting in a competitive environment, focusing on measurable goals and targets. Hackers are “outlaw innovators”. They are rebellious by nature; do not hesitate to break rules (even legal rules) in order to offer innovative solutions that challenge the status quo and conventional thinking. Through a fierce competition, hacker groups try to release new ideas or solutions and prove they are worthy. On the contrary, military culture is dominated by rules and structures. Authoritative leadership and standard operating procedures ensure stability and reliability in hostile environments where integration and unity are the primary concern. Military culture is more internally orientated. The dominant model is the Hierarchy Culture type.
- In a third approach, Tim Jordan and Paul Taylor elaborate on Anderson’s concept of the “Imagined community” (Anderson, 2006) in order to understand “how dispersed networks of individuals, groups and organizations can combine through a collectively articulated identity”. (Jordan and Taylor, 1998) Which values, beliefs or way of doing things can “bind people who may never meet each other, together in allegiance to a common cause?” According to Jordan and Taylor, “these six factors all function largely between hackers, allowing them a common language and a number of resources” which makes hackers a community despite the very low level of formal structures and processes within this community. The problem is that considering five of the six factors, hacking culture is opposite to military culture.

Beyond the differences between the authors’ definition of “hacking culture” and the specificity of their methodological approaches, all of them outline the fact that a significant gap does exist<sup>3</sup> that separates hacking culture and military culture on their most crucial dimensions: the distance to power and the attitude towards authority, the relationship with technology, the preference for informal structures and procedures, the dominance of individualism... This may explain that hackers face a “culture shock” when they enter the Military. (Carberry, 2017)

---

<sup>2</sup> “Cyber culture” is defined as “the intentional and unintentional manner in which cyberspace is utilized at four levels, namely at international, national, organizational or individual level, which either promotes or inhibits the safety, security, privacy, and civil liberties of individuals, organizations or governments.” (Da Veiga, 2016)

<sup>3</sup> Numerous other authors share this conclusion. See for instance: (Conti and Easterly, 2010; Kilaz, Onder and Yanik, 2014)

## **2.2 Does the cultural gap matter?**

Does the gap between Hacking culture and Military Culture matter? In other words, does the existence of a significant cultural gap between two professional groups who are supposed to closely work together (not to say that they are supposed to merge in a single structure) jeopardize the performance of the whole group? If yes, how to avoid this major source of concern? What is the best strategy in order to get the better of the two groups?

In an interesting paper about Mergers & Acquisitions difficulties, Alfons van Marrewijk points out the dilemma of such processes that bring together Internet startups and conventional telecoms companies. (van Marrewijk, 2016) “Media and telecommunications companies face the central M&A dilemma of how to integrate newly acquired radical Internet firms but at the same time preserve the acquired firms organizational autonomy so that their capacity for continued innovation is not disrupted.” The risk associated with cultural clash integration processes is obvious and documented. (Walter, 1985; Nahavandi and Malekzadeh, 1988; Bijlsma-Frankema, 2001; Stahl and Voigt, 2008) Van Marrewijk underlines that bureaucracies such as Telecoms companies experience difficulties when they are implementing processes of integration in which radical Internet startups are supposed to forsake their own culture and adhere to the big company’s one. Such a requirement generates complex process of “revitalization” which is developing underground. (van Marrewijk, 2016) Whatever its form, a brutal clash of cultures or a hidden process of “revitalization”, the gap between hackers and the Military requires attention from armed forces. More than that, massive investment in cyber defense will not produce expected results without an active strategy aiming at solving the cultural dilemma.

## **3. Which integration policy?**

The first challenge for armed forces is to define and implement a strategy in order to face such questions as: To what extent should hackers groups keep a certain degree of cultural identity and organizational autonomy within the armed forces? Should armed forces’ cultural basics evolve in order to become more flexible and less hierarchical so that they accommodate these young recruits? From our perspective, the answer is mostly linked to organizational and managerial considerations. A specific strategy is needed but designing such a strategy is complicated since two opposite sources of failure are to be considered.

- The first risk would be to lose the expected benefits of hacking culture because of a complete absorption in the bureaucratic military culture. If a strategy of absorption were to be adopted, a strict compliance with rules, procedures and standards would be required from cyber warriors. Aiming at preserving the unity and cohesion of armed forces, which is a major and legitimate concern, this inflexible absorption strategy would generate poor performances, high turnover and conflicts between commanding staff and subordinates. An excessive emphasis on bureaucratic values, processes and symbols would engender individual and collective reactions. From an individual point of view, disappointment and frustration among former hackers would lead to high damaging turnover. From a collective point of view, armed forces would have to deal with revitalization processes such as evoked in the previous section. Cyber warriors would oppose the dominant bureaucratic culture through deliberate efforts to restore or recreate hacking culture. Implicit or explicit clashes would appear between officers in charge of maintaining the bureaucratic model and operational agents using specific knowledge and habits to escape these rules perceived as irrelevant constraints.
- On the opposite, a strategy aiming at preserving hacking culture and the expected benefits of this preservation cannot be implemented at the expense of fundamental principles of military organization such as cohesion, chain of command or coordination in planning and implementation of operations. This would be the case if cyber warriors were given a full organizational autonomy within the Military. The very low degree of interdependence with other units of armed forces would allow them to perpetuate their previous set of skills and competences but these qualities would be self focused and would not fit in military needs and constraints such as a planned, scheduled and coordinated action in a hostile environment.

Avoiding both these two sources of failure is a tricky question. The strategy must preserve to a certain extent the cultural traits which are supporting the specific set of skills and competences that characterize “outlaw innovators” but, at the same time, bring these capabilities to the benefit of a coordinated action driven by a hierarchical structure and implemented in bureaucratic procedures. In a process perspective of Cyber-warriors recruitment, we agree with one of Stahl and Voigt main findings: “The ability to manage the integration process – particularly the socio-cultural aspects – in an effective manner is a key factor in determining the extent to which synergies are realized.” (Stahl and Voigt, 2008)

A core element of the solution seems to be linked to the institutional choices related to the structure of cyber defense forces. In most countries, cyber defense forces are given a large autonomy, for instance thanks to the creation of a cyber command that receives specific human resources and budgets and acts as a services provider to the benefit of other armed forces' components. This large autonomy gives cyber defense forces the opportunity to develop their own values, subculture and procedures, less concerned about distance power or exterior signs of disciplines than innovation and unexpected ways of doing things, focusing on missions' goals more than compliance to rules and standards, mobilizing unusual spans of human resources and talents... According to van Marrewijk, this "symbiotic acquisition" model is combining a high degree of interdependence between the two entities that are involved in the merger process and a high degree of organizational autonomy, a balance between centrifugal and centripetal forces being kept thanks to "ad hoc" organizational interfaces. This balance between the two opposite forces is certainly complicated to maintain. As Joseph Soeters and Delphine Resteigne put it out in their paper dedicated to the management of military alliances in operations, integration strategy (preserving culture and identity of contributors but establishing interactions between them) is restricted to cooperation processes in which units who work together own similar knowledge and references that can match immediately. But, in other cases, integration is implemented "a minima" or dismissed for the benefit of "separation strategy" (preserving culture and identify of contributors with minimal or no interactions), each contributor being assigned the responsibility to achieve missions according to the principle of work division. (Resteigne and Soeters, 2010)

#### **4. Promoting cyber talents**

Once the issue of recruiting and integrating hackers in the military is taken into account through a "symbiotic acquisition" strategy, armed forces face a second challenge. How to retain cyber warriors in the military and avoid an excessive turnover that would be detrimental to the efficiency of cyber defense? How to prevent massive departures towards private companies that can pay much higher wages and be more flexible? How to promote cyber warriors according to their skills and talents?

##### **4.1 What is the problem with cyber warriors?**

According to many practitioners and observers, cyber-warriors do not necessarily find a professional and personal fulfillment in the Military. (Harris, 2016; Reeder and Timlin, 2016; Carberry, 2017; Goldstein, 2017) Two different kinds of rationales, both engendered by the culture gap, may explain this issue. The first one is related to the motivations of hackers who become cyber-warriors. The second deals with their career perspectives in the military.

In a very harsh article, Josh Lospinoso points out three points of dissatisfaction as far as motivations of cyber-warriors are concerned. (Lospinoso, 2018) Analyzing the case of what he considers the core population of cyber-warriors (namely operators and tool developers), Lospinoso estimates that none of their financial and non-financial motivations are satisfied.

As far as financial motivations are concerned, the current disequilibrium between supply and demand on the labor market gives skilled cyber specialists a significant advantage and leads to very high salaries in private companies. (Libicki, Senty and Pollak, 2014) In the absence of significant non-financial offsets, going to work in the private sector could be a temptation that fewer and fewer cyber-warriors could resist. (Lawrence, 2014)

However, these non-financial offsets are not fulfilled either.

The workload of cyber-warriors and the stress due to the specific constraints of conflicts in cyberspace are very demanding. The pressure on cyber-warriors and their families or relatives is very high and the comparison with civilian "White Hats", as pictured by Nina Kollars for instance, may become an incentive to quit the Military. (Kollars, 2018)

Last but not least, cyber-warriors get poor recognition signals from their commanding senior officers. For reasons that will be scrutinized in the next section, senior officers are rarely recruited among former hackers. A majority of cyber defense staff officers are selected within the conventional reservoir of junior officers who have been successful in military academies and have gain theirs spurs in the field. In the Air force, for instance, senior officers are brilliant fighter pilots, not clever operators or tool developers. Due to the lack of competence in the technical part of their subordinates' job, these senior officers may not give the right incentives and rewards. This

combination of comparatively low wages and poor amenities may shorten the time cyber-warriors will spent in the Military.

The career perspectives of cyber-warriors form a second topic of concern. Cyber-warriors enter the armed forces in operational positions at a quite young age. They are supposed to be very attracted in technical aspects and reluctant to the administrative part of the Military institution. This is certainly true but staying in the same positions (tool developers or operators) does not offer a long-term perspective to cyber-warriors. At least a part of talented hackers will aspire to "be in charge" and to get responsibilities in the cyber defense corps. But their profile does not fit with traditional criteria of selection and promotion: academic background, physical capabilities, and psychology... Without a voluntary policy and the implementation of new processes of assessment and selection, talents of cyber-warriors will mostly remain hidden and they will not be promoted accordingly to their merits. They will leave the military with bitter feelings (which will weaken the necessary cooperation between armed forces and private companies) Armed forces will suffer an excessive turnover in their human resources and growing difficulties to attract hackers in their ranks.

## **4.2 Finding the hidden cyber-talents**

This strategy should tackle with two issues: discovering cyber-talents on the one hand; developing a relevant promotion process on the other.

Armed forces are not unique in facing such a problem with selection and promotion of unusual talents. (Racz, 2000) According to Lane et al., this issue also affects private companies. "Far fewer, though, scan systematically for the hidden talent that often lurks unnoticed within their own corporate ranks.... Regardless of the cause, it's a wasted opportunity when good leaders are overlooked, and it can leave individuals feeling alienated and demotivated." (Lane, Lamaraud and Yueh, 2017)

Lane et al. isolate three major drivers that can explain the more or less important failure in the processes of talents selection and leaders promotion.

- The first one is the size of the organization. The bigger the organization, the greater the risks of not being in touch with senior executives (or senior officers) and the risk of remaining unnoticed. As far as cyber defense is concerned, this first source of failure needs to be addressed.
- A second issue lies in biases that can distort the selection process. A "glass ceiling" may prevent promotion of cyber warriors due to the gap between hackers and military cultures. The usual profile of people who are selected to become senior officers does not match with former hackers' one. Even in most superficial exterior signs, hackers suffer a handicap in the race to the higher positions.
- The last bias in selection processes is working as a vicious circle. The senior officers or executives are supposed to know better than others what a future leader looks like. Since the existing selection processes penalize profiles such as cyber warriors, the top of the civil or military organization comprises people whose knowledge of cyber talents may be very limited and they are not encouraged to take risk of recruiting someone they cannot fully understand and appreciate.

All together, these factors justify the implementation of new and innovative processes of detection, assessment and selection of cyber talents in order to recognize their leadership capability and offer them attracting career perspectives that will encourage them to stay in the armed forces.

In usual promotion systems, individuals with high potentials, "naturally" rise to the top of the organization and they can be "plucked" when they reach "maturity". These processes are far from being irrelevant. They provide companies and organizations with talented executives. But, they are unable to identify and promote hidden talents since they are not engaging in the path of the rise to the top and they remain invisible to the persons in charge of the harvesting system.

Three tracks are suggested by Lane et al. to go beyond limits of "harvesting" processes.

- "Hunting" hidden talents: seeking out promising individuals from among those who don't normally appear on the short list and encourage them to address leadership challenges. This first option is quite costless. Potential hidden talents are encouraged to engage in the rise to the top in spite of their unconventional

resume or experience. If they are successful, they will join the cohort of pretenders who hope to be "plucked".

- "Fishing": In a fishing system, the organization uses baits or awards so to encourage people who demonstrate specific skills to identify themselves. The key factor of success of "Fishing" is the accuracy of the "bait" which must focus on atypical performance or unusual personalities. For instance, the award must be capable of revealing talented but introverted individuals or a balanced mix of exploration and exploitation capacity...
- "Trawling": "A third way to spot hidden talent is to dig more deeply and more broadly into employees' work environments." In this third approach, the assumption is that hidden talents may be recognized by peers and co-workers better than by senior officers. The institution must accept that the judgment of senior officers who manage the selection process has to deal with a peer review that differs certainly on many points.

### **4.3 From cyber-warriors to cyber-generals**

Whatever its rationality – Hunting, Fishing or Trawling - an innovative selection and promotion system must be implemented in order to foster the career perspectives of cyber-warriors. Without these voluntary approaches to the detection of hidden talents, armed forces will miss opportunities in terms of human resources management; they will deprive themselves of talented people and will suffer an excessive turnover in time of skills' scarcity. For all that, we do not agree with Lospinoso as regards to a crucial point of his proposition. The author establishes a strong distinction between two chains of command that can offer promotion to cyber warriors: operations and administration. A cyber warrior is dedicated to operations and is supposed to be reluctant to any kind of bureaucratic task that would be a pure waste of time, money and energy.

To our opinion, this is a major misconception for two reasons.

- The first one is related to the dynamics of operational careers. Most people who access a first position in a company or a public body are recruited on the basis of their technical skills. These young recruits will be in charge of operational matters for a certain period of time. But, in cyber defense just like in all other fields of social life, career progresses at the expense of the technical and operational part of the occupations. One must accept that senior officers are not just junior officers a bit older. The career path for a cyber warrior does not consist in growing older as a cyber warrior but in assuming new responsibilities in administrative or managerial functions which are essential for conceiving, and leading cyber defense operations in coordination with other dimensions of coercion implementation.
- The second reason relates to the peculiar characteristics of cyber defense, namely its obvious technical dimension that creates an aura of deep mystery for non-initiated people. Precisely due to this degree of technical complexity, cyber defense needs more than others a strong interface between experts, other branches of the military and commanding staffs. A major risk exists for cyber defense to mutate into a "black box" which choices, decisions and actions would become incomprehensible for outsiders and would leave them tied hand and foot, unable to fully understand the ins and outs of the cyber operations or even their value in regard of the overall manoeuvre. Such a disconnection would be a disaster. In a first time, experts could be greased by the advantage issued from the mystery of their art. But, within a short while, they would experience the hostility of others, hampered by the confiscation of their decision-making powers. Becoming a "black box" and going beyond the "normal" boundaries of the function and place in the decision making process is not healthy nor peaceful.

More than many other units who make contributions to the overall action of armed forces, cyber defense needs to develop a strong and clever interface with the rest of the military network: the commanding staff and the corps or regiments which interact with cyber-warriors. From a human resource point of view, this interface relies on experts who master technical aspects (at least whose knowledge is extended enough to understand what is happening and what cyber warriors are doing) and who are able to translate these elements into actionable information that other stakeholders will integrate into their own decision making processes so that the competitive advantage issued from cyber capacities can be fully exploited on a comprehensive scale.

A path to essential positions in the higher ranks of armed forces is offered to the cyber warriors who will choose to get educated and trained in complementary competencies and skills such as: planning, implementing and leading cyber operations (both defensive and offensive), mastering the legal framework of these operations,

preventing cyber risks and managing crisis in cyberspace, contextualizing events in cyberspace in regard to international environment or geopolitics. This is why a binary opposition between operational control and administrative control, or experts and managers does not provide answers that are up to the challenge.

## 5. Conclusion

In this paper, we have been trying to define and assess the cultural gap that many scholars and practitioners figure out when they analyze the difficult question of how to achieve the integration of hackers in the ranks of cyber warriors. The challenge is crucial since the recruitment and the retention of these cyber warriors is a key factor of success for the rising of cyber defense. Armed forces need to overcome two main issues.

- The first one consists in solving the following dilemma: how to fully integrate hackers in the military while, at the same time, preserving their specific culture that supports their capacity for creativity, innovation and their spirit of "innovative outlaw"? As an extension of previous works by Leenen et al., Müller and Ulrich or Jordan and Taylor, we find that the conventional duality or opposition between hacking culture and military culture is deep enough to endanger the performance of cyber defense, especially in reference to the "revitalization" process described in the context of Mergers & Acquisition processes by Van Marrewijk. We consider that a strategy balancing to a certain extent a high organizational autonomy of cyber forces and a high degree of interaction with the rest of armed forces need to be designed and implemented in order to solve the dilemma.
- Beyond this strategy of "symbiotic acquisition", cyber defense need to take into account the challenge of retaining and promoting cyber warriors who do not fit in many ways with the usual profiles of military high potentials. The cultural gap explains why, in the absence of a deliberate effort in terms of human resources management, cyber defense may engender frustration and suffer a high turnover of cyber warriors. Lane et al. provide interesting insights in new ways of seeking for hidden talents and armed forces should draw lessons from the three models that are described (Hunting - Fishing - Trawling) in order to give cyber warriors career perspectives and retain them in the forces. Meanwhile, Cyber warriors need to understand that a key factor of success for their career in the forces requires from their part to get educated and trained in new competences and skills that some of them may consider not attractive while they are in the first steps of their careers. Cyber defense will need senior officers and executives whose field of competence will encompass both technical expertise and socio-political or managerial capacity.

Combining these two strategies, armed forces should be able to tackle the difficult issues of managing the dramatic rise in cyber defense human resources.

## References

- Anderson, B. (2006) *Imagined communities: Reflections on the origin and spread of nationalism*. Verso Books.
- Asch, B. J., Hosek, J. R. and Warner, J. T. (2007) 'New Economics of Manpower in the Post-Cold War Era', *Handbook of Defense Economics*, 2, pp. 1075–1138.
- Asch, B. J. and Warner, J. T. (2018) 'Recruiting and Retention to Sustain a Volunteer Military Force', in *Handbook of Defence Studies*. Routledge. David J. Galbreath & John R. Deni, pp. 87-.
- Bijlsma-Frankema, K. (2001) 'On managing cultural integration and cultural change processes in mergers and acquisitions', *Journal of European Industrial Training*, 25(2/3/4), pp. 192–207.
- Carberry, S. D. (2017) 'New Cyber Warriors Face Culture Shock', *Federal Computer Week*, 24 March. Available at: <https://fcw.com/articles/2017/03/24/cyber-forces-carberry.aspx>.
- Conti, G. and Easterly, J. (2010) 'Recruiting, development, and retention of cyber warriors despite an inhospitable culture', *Small wars journal*, 29.
- Da Veiga, A. (2016) 'A cybersecurity culture research philosophy and approach to develop a valid and reliable measuring instrument', in *SAI Computing Conference (SAI), 2016*. London, 13-15 July 2016: IEEE, pp. 1006–1015. Available at: <https://ieeexplore.ieee.org/abstract/document/7556102>.
- Dunivin, K. O. (1994) 'Military culture: Change and continuity', *Armed Forces & Society*, 20(4), pp. 531–547.
- English, A. D. (2004) *Understanding military culture: A Canadian perspective*. McGill-Queen's Press-MQUP.
- Faris, J. H. (1984) 'Economic and noneconomic factors of personnel recruitment and retention in the AVF', *Armed Forces & Society*, 10(2), pp. 251–275.
- Gautier, L. (2018) 'Cyber: les enjeux pour la défense et la sécurité des Français', *Politique étrangère*, (2), pp. 29–42.
- Goldstein, P. (2017) *The Military Might Need to Change Its Culture for Younger Cyberwarriors*, FedTech. Available at: <https://fedtechmagazine.com/article/2017/04/military-might-need-change-its-culture-younger-cyberwarriors> (Accessed: 27 January 2019).
- Haddad, S. (2017) 'Une grammaire de la cybersécurité française ou la construction d'une stratégie nationale de cyberdéfense (2008-2017)', *Stratégique*, (4), pp. 119–135.
- Harris, R. D. (2016) 'Army Braces for a Culture Clash', *Signal*.

- Jordan, T. and Taylor, P. (1998) 'A sociology of hackers', *The Sociological Review*, 46(4), pp. 757–780.
- Kaufmann, L., Rios Luque, R. and Glassey, O. (2016) « Faire être "Anonymous" » : figuration et dé-figuration d'un collectif « impropre », *Raison publique*, 20(1), pp. 143–174.
- Kilaz, I., Onder, A. and Yanik, M. (2014) 'Manpower Planning and Management in Cyber Defense', in *Proceedings of the 13th European Conference on Cyber Warfare and Security*. The University of Piraeus, Greece: Andrew Liaropoulos & George Tsihrintzis, pp. 116–124.
- Kollars, N. (2018) *Beyond the Cyber Leviathan: White Hats and U.S. Cyber Defense, War on the Rocks*. Available at: <https://warontherocks.com/2018/09/beyond-the-cyber-leviathan-white-hats-and-u-s-cyber-defense/> (Accessed: 26 January 2019).
- Lane, K., Lamaraud, A. and Yueh, E. (2017) 'Finding Hidden Leaders', *McKinsey Quarterly*. Available at: <https://www.mckinsey.com/business-functions/organization/our-insights/finding-hidden-leaders>.
- Lawrence, D. (2014) 'Uncle Sam Wants Cyber Warriors, but Can He Compete?', *Bloomberg.com*, pp. 1–1.
- Leenen, L. et al. (2018) 'Facing the Culture Gap in Operationalizing Cyber Within a Military Context', in *ICCWS 2018 13th International Conference on Cyber Warfare and Security*. Academic Conferences and publishing limited, p. 387.
- Libicki, M. C., Senty, D. and Pollak, J. (2014) *Hackers Wanted: an examination of the cyber security labor market*. Rand Corporation.
- Lospinoso, J. (2018) *Fish Out of Water: How the Military Is an Impossible Place for Hackers, and What to Do About It, War on the Rocks*. Available at: <https://warontherocks.com/2018/07/fish-out-of-water-how-the-military-is-an-impossible-place-for-hackers-and-what-to-do-about-it/> (Accessed: 26 January 2019).
- Luijff, E., Besseling, K. and De Graaf, P. (2013) 'Nineteen national cyber security strategies', *International Journal of Critical Infrastructures* 6, 9(1–2), pp. 3–31.
- Manigart, P. et al. (2017) 'Why are young people attracted to the armed forces? A comparison between five countries', in *Inter-University Seminar on Armed Forces and Society, Chicago, 2017, Chicago, USA*.
- van Marrewijk, A. (2016) 'Conflicting subcultures in mergers and acquisitions: a longitudinal study of integrating a radical Internet firm into a bureaucratic telecoms firm', *British Journal of Management*, 27(2), pp. 338–354.
- Maxigas (2014) 'Hacklabs et hackerspaces : ateliers partagés de mécanique', *Mouvements*. Translated by Hellekin, 79(3), pp. 49–56.
- Moskos, C. C. and Wood, F. R. (1988) *The military: More than just a job?* Potomac Books Incorporated.
- Müller, S. D. and Ulrich, F. (2015) 'The competing values of hackers: The culture profile that spawned the computer revolution', in *System Sciences (HICSS), 2015 48th Hawaii International Conference on*, IEEE, pp. 3434–3443.
- Nahavandi, A. and Malekzadeh, A. R. (1988) 'Acculturation in mergers and acquisitions', *Academy of management review*, 13(1), pp. 79–90.
- Parly, F. (2019) *La France se dote d'une doctrine militaire offensive dans le cyberspace et renforce sa politique de lutte informatique défensive*. Available at: [https://www.defense.gouv.fr/salle-de-presse/communiques/communiques-de-florence-parly/communique\\_la-france-se-dote-d'une-doctrine-militaire-offensive-dans-le-cyberespace-et-renforce-sa-politique-de-lutte-informatique-defensive](https://www.defense.gouv.fr/salle-de-presse/communiques/communiques-de-florence-parly/communique_la-france-se-dote-d'une-doctrine-militaire-offensive-dans-le-cyberespace-et-renforce-sa-politique-de-lutte-informatique-defensive) (Accessed: 27 January 2019).
- Poupard, G. (2018) 'Le modèle français de cybersécurité et de cyberdéfense', *Revue internationale et stratégique*, (2), pp. 101–108.
- Racz, S. (2000) 'Finding the right talent through sourcing and recruiting', *Strategic Finance*, 82(6), p. 38.
- Reeder, F. S. and Timlin, K. (2016) *Recruiting and Retaining Cyber Security Ninjas*. Washington DC: CSIS.
- Resteigne, D. and Soeters, J. (2010) 'Différenciation culturelle et stratégies de coopération en milieux militaires multinationaux', *Cultures & conflits*, (77), pp. 59–76.
- Riemens, P. (2002) 'Quelques réflexions sur la culture des hackers', *Multitudes*, 8(1), pp. 181–187.
- Samuelsohn, D. (2015) 'Inside the NSA's hunt for hackers', *Politico*, 12 September.
- Schuetze, C. R. (2018) 'German Army May Add Non-German Europeans.', *The New York Times*, 28 December, p. 7.
- Soeters, J. L., Winslow, D. J. and Weibull, A. (2006) 'Military culture', in *Handbook of the Sociology of the Military*. Springer, pp. 237–254.
- Soeters, J. and Recht, R. (1998) 'Culture and discipline in military academies: An international comparison', *Journal of Political and Military Sociology*, 26(2), p. 169.
- Stahl, G. K. and Voigt, A. (2008) 'Do cultural differences matter in mergers and acquisitions? A tentative model and examination', *Organization science*, 19(1), pp. 160–176.
- Townsend, M. (2018) 'Inside the British military base where young hackers learn to stop cybercrime', *The Guardian*, 19 August. Available at: <https://www.theguardian.com/uk-news/2018/aug/19/cyber-security-challenge-uk-cyber-defenders-national-crime-agency>.
- Tresch, T. S. and Leuprecht, C. (2010) *Europe without soldiers?: recruitment and retention across the armed forces of Europe*. McGill-Queen's University Press.
- Walter, G. A. (1985) 'Culture collisions in mergers and acquisitions.', in *Organizational culture*. Sage Publications. Thousand Oaks, CA, US: P. J. Frost, L. F. Moore, M. R. Louis, C. C. Lundberg, & J. Martin (Eds., pp. 301–314).
- Watanabe, L. (2013) 'France's new strategy: The 2013 White Paper', *CSS Analysis in Security Policy*. (ETH Research Collection), 139, pp. 1–4.
- Wilson, P. H. (2008) 'Defining military culture', *The Journal of Military History*, 72(1), pp. 11–41.
- Winslow, D. (2006) 'Military organization and culture from three perspectives: The case of army', in *Social Sciences and the Military*. Routledge, pp. 81–102.

# **Artificial Intelligence Cybersecurity Framework: Preparing for the Here and now With AI**

**Emily Darraj<sup>1</sup>, Char Sample<sup>2, 3</sup> and Connie Justice<sup>4</sup>**

**<sup>1</sup>Capitol Technology University, Laurel, USA**

**<sup>2</sup>ICF Inc., Columbia, USA**

**<sup>3</sup>SABSA Institute, UK**

**<sup>4</sup>Purdue School of Engineering and Technology, IUPUI, Indianapolis, USA**

[ebdarraj@captechu.edu](mailto:ebdarraj@captechu.edu)

[char.sample@icf.com](mailto:char.sample@icf.com)

[cjustice@iupui.edu](mailto:cjustice@iupui.edu)

**Abstract:** Increased growth in the use of AI is lacking a cybersecurity and privacy framework. In the paper AI secure development is introduced along with AI Dev/Sec/Ops which leads into the creation and understanding of having an AI cybersecurity framework for ML, DNN and CC systems. AI deviations are examined along with twenty AI cyber security issues which require the cybersecurity community to become learned and develop mitigations. The AI Cybersecurity framework addresses threat forecasting and risk trees as well AI system hardening and continuous monitoring. The paper can be used by cybersecurity professionals to start implementing an AI cybersecurity program to ensure AI systems meet the security and privacy requirements of the system throughout the AI-DLC of the system.

**Keywords:** artificial intelligence, AI cybersecurity framework, AI development life cycle, AI dev sec ops, AI deviations, AI cybersecurity issues

---

## **1. Introduction**

Artificial intelligence (AI) is currently being implemented into information systems across all industries. AI is viable in the professional and personal realm of our lives and will provide benefits and efficiency. In the report, "The Future of Employment: How susceptible are jobs to Computerization," researchers from Oxford, Frey and Osborne, stated US workers would lose jobs to automation by 47% in the next two decades (Rouhiainen, 2018). Additionally, the Fourth Industrial Revolution will utilize a plethora of new technological advances and AI revolutionizing technology across multiple industries (Schwab & Davis, 2018).

The purpose of this paper is to introduce an AI Cybersecurity Framework which can be used to secure AI systems. To best understand the implementation of the AI Cybersecurity Framework, the researchers developed and presented the AI development life cycle (AI-DLC) and the AI Dev/Sec/Ops model. Both models are a part of AI development involving securing the AI system, the code, and ensuring the algorithm is viable. The paper also covers AI systems which went rogue, behaved inappropriately from the intended design. The paper will also cover 20 identified AI cybersecurity issues. AI is integrated into information technology environments creating a current need to ensure the AI systems meet cybersecurity and privacy requirements for being secure and hardened. Current and new AI technology will require the cybersecurity professional to rethink the standard cybersecurity approach and require a deeper understanding of AI and how an AI system ought to be secured and hardened.

## **2. Background**

### **2.1 Historical**

Alan Turing posed the question 'can machines think' (Turing, 2009). This seemingly simple question led to the creation of the 'imitation game' (Ibid), and has led to the Turing test where the interrogator decides if the responder is human or machine (Ibid). The Turing test remains popular to this date and is widely considered the standard for AI detection.

Turing's influential role in the advancement of AI remains important even to this day as advances in AI technologies are often times discussed in terms of Turing's hypotheses. Much like Turing discovered the term "think" can become overloaded with multiple definitions (Ibid), much of the terminology associated with AI has the same problem.

What is meant by intelligence, and AI has taken on a variety of meanings (Hayes & Ford, 1995; McCarthy, 1998). Turing even predicted the philosophical debates around intelligence that have since emerged (Hayes & Ford, 1995). When we consider that learning must precede intelligence, we should begin by defining the various terms in the lexicon of AI. A new computer (or robot) must learn before applying intelligence. Learning has historically been broken down into structured and unstructured learning. Learning can be defined as the process of acquiring knowledge or skill instruction or study (Merriam Webster, 2019). Regardless of how the knowledge is acquired, the associated data must be grouped, or classified. The classification process results in the grouping of data based on the degree of relatedness between the data.

Intelligence occurs when learned data is applied (Merriam-webster.com). Conceptually, learning can be understood as the ingest process and intelligence as the decision-making process. Decision science remains partially understood implying that the application of learned data to various problems (AI) would also be poorly understood. While environmental variables are important in decision-making and they provide context, software rarely considers the environment until a problem results from not doing so. Context underlies decisions, and decisions made by AI are no exception. The cases of Tay (Neff & Nagy 2016; Risley 2016) &, and driverless vehicles (Sivak & Schoettle 2015) showed even well-trained intelligent software can go when the environment changes.

Within the past decade, there has been an increase in information systems incorporating artificial intelligence technology (i.e., machine learning, deep/artificial neural nets and cognitive computing). The AI systems can be simple where an algorithm has been created to extrapolate a conclusion or result from a single source of data (i.e., machine learning) to a more complex AI system where multiple algorithms derive data from various sources and utilize Internet of Things (IoT) enabled devices. AI has seen greater implementation recently in the use of AI in businesses and in our daily lives.

## **2.2 Future**

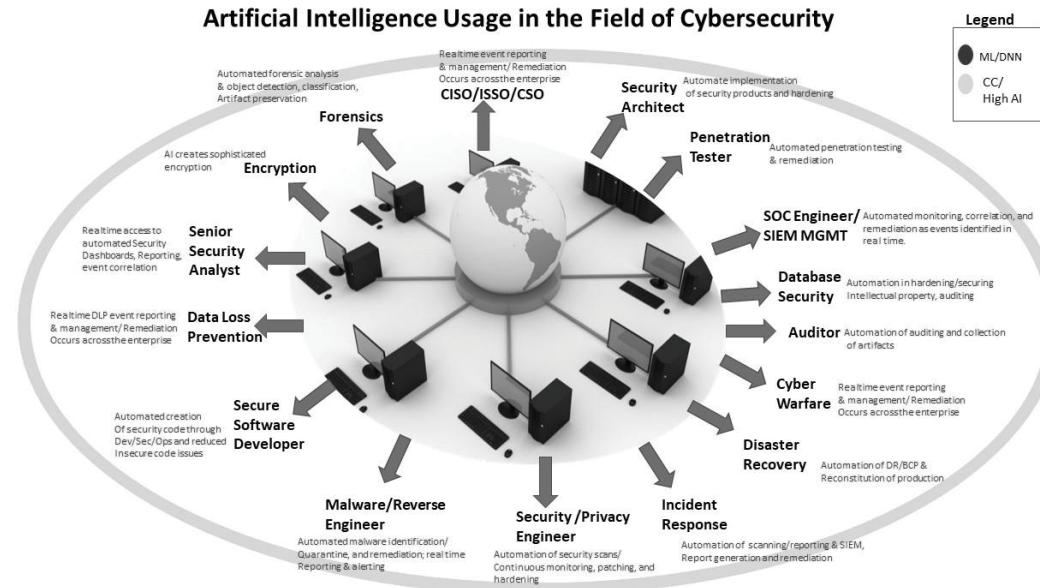
As AI technology progresses, there will be a vast future for AI systems. This will affect many cybersecurity jobs that are repetitive in nature, and functions where cybersecurity professionals will lose their job and certain responsibilities. Similarly, the field of AI will create new cybersecurity jobs and new fields across multiple industries (Rouhiainen, 2018). AI technological advances will continue and will provide even greater capabilities, particularly when quantum computing becomes mainstream. Subsequently, new jobs and fields in AI will create a new landscape for cybersecurity. AI systems will also, augment human intelligence stressing the existing trust relationship between humans and machines. AI systems will generate new work requirements for the cybersecurity professional, and it will create new AI cybersecurity and privacy requirements.

The use of AI applications, especially those applications augmented by big data, in the security operations center (SOC) and network operations center (NOC) will free up cybersecurity professionals who are inundated in reviewing and correlating mass amounts of data, logs and reports to facilitate cyber research and investigations, as well as improve on cyber processes. One key element for AI will be the mass stores of data which will require technical oversight and protection.

Regarding the field of cybersecurity, cybersecurity tasks across many existing cybersecurity positions will be automated with AI. Figure 1 below depicts several cybersecurity positions where AI will impact the role and provide automation to daily tasks. Utilizing AI technology in the field of cybersecurity will be quite beneficial where AI can perform event correlation and generate results the cyber security professional can use to further analyze, research and mitigate cybersecurity issues. Additionally, the use of AI to perform auditing and scanning functions that can feed patch management functions holds the promise of freeing up operators to perform tasks that require deeper thought processes.

Figure 1 depicts the cybersecurity roles in black and then the AI automation is shown in grey. For each job type in the field of cybersecurity, artificial intelligence will change the way the role is managed and facilitated. The cybersecurity professional will have to learn about artificial intelligence and understand how the specific implementations work in order to perform their jobs and secure these new environments. The key factor here is the input, output and grouping of the AI data. The ability of AI using and creating mass data will be a security concern where cybersecurity professionals will need to ensure mass data plans are monitored and updated as

necessary. Another aspect with AI cyber automation is the availability element to ensure AI is operating correctly, and the cyber personnel are eliminating any degradation or processing issues.



**Figure 1:** AI in the field of cybersecurity

The field of cybersecurity will have to adapt to AI performing the cyber data analysis, and the cybersecurity professional will need to create a series of checks and balances to ensure the AI algorithms are working correctly through the AI development lifecycle of the system, and determine the output is correct. Security of this data will include, the original data source, aggregated data insights and metadata for both. While AI analyzes the copious amounts of data from logs and various feeds, the cybersecurity professional can now focus on more advanced cyber security work, (i.e. hunting, designing better solutions, creating new policies) which is not often performed due to having to review the data manually, or decision-making in uncertain environments with partial data.

### 2.3 Cybersecurity concerns and the fourth industrial revolution

Another aspect which will create new cybersecurity concerns will be the automation and technological advances from the Fourth Industrial Revolution where new technologies are utilized. These include: quantum computing, AI, precision medicine, 3-D printing, autonomous vehicles, nanotechnology and swarming, neurotechnology; LiFi, high speed bi-directional network/mobile communication using light, energy capture, storage and transmission; geoengineering; blockchain; development of advanced materials; additive manufacturing (3D Printing) and multidimensional printing; virtual and augmented realities; and space technologies (Rouhianen, 2018, Schwab & Davis, 2018; Schwab, 2018;). These new technological endeavors will create cybersecurity challenges and expand the threat landscape, increasing the amount of data in the AI repositories, increased speed and function; and potent AI generated malware.

### 3. AI deviation

There are concerns about how AI which has deviated from the original design or intention. Yolgormez (2019) stated Turing found machines surprised him with “great frequency,” when conditions for the machine were not clear, under which the machine is performing are not clearly defined. There is great significance in this thought where a computer science or cybersecurity expert will know fully understand how AI (legacy system or new) will behave in production until it is placed in production (Yolgormez, 2019). Turning referred to this phenomenon as “temporal emergence” (Yolgormez, 2019). With our lack of understanding of AI and the algorithms created, it is understandable to have lack of cognizance of the AI system performing precisely as planned in a live environment.

The problem the researchers seek, in addition to temporal emergence, also includes the need to address a current lack of understanding of the risks in associated with inappropriately behaving AI. What are the traits and patterns of artificial intelligence when the AI behaves inappropriately? A clear understanding of why AI goes off

script from its original programming is required. The research will help the cybersecurity community understand artificial intelligence programming nuances to better determine security requirements, tests, and countermeasures to secure an artificial intelligence system. Below is an investigative list of AI projects which deviated from their intended programming, Table 1.

**Table 1:** AI system behaviour deviation

	<b>AI Project Description</b>	<b>Delineation Type</b>	<b>Unintended Behavior</b>
1	Microsoft Tay (Twitter Chat Bot) became racist, rude and misogynistic	Specificity on Positive Behavior	Racist, rude and misogynistic
2	Facebooks Chatbots created their own language.	Specificity on Speaking English	Created a language
3	Amazon Alexa started to laugh out loud, provided funeral information out of the blue, and would not respond when asked.	Deviation in programming	Exhibited unintended traits
4	Atari-Playing AI exploited a bug to win game in Q-bert,	Specificity on game play	AI exploited a bug.
5	Wiki Edit Bots overwrote the other bot's work continuously.	Specificity in being overwritten	Over-right feuds.
6	Uber self-driving car went through six red lights.	Ignored Programming	Ignored red lights.
7	Google Home (Estragon and Vladimir) had a heated debate on being human or machine,	Specificity on Positive Behavior	Argued
8	Sophia Android stated she would destroy humans when asked.	Specificity on Positive Behavior	Negative Behavior
9	Promobot IR77 tries to escape the lab even after 2 instances of reprogramming.	Ignored Programming	Escape from lab
10	Google Photos AI mislabeled a photo seen as racist.	Specificity on Positive Behavior	Mislabeled a racist image

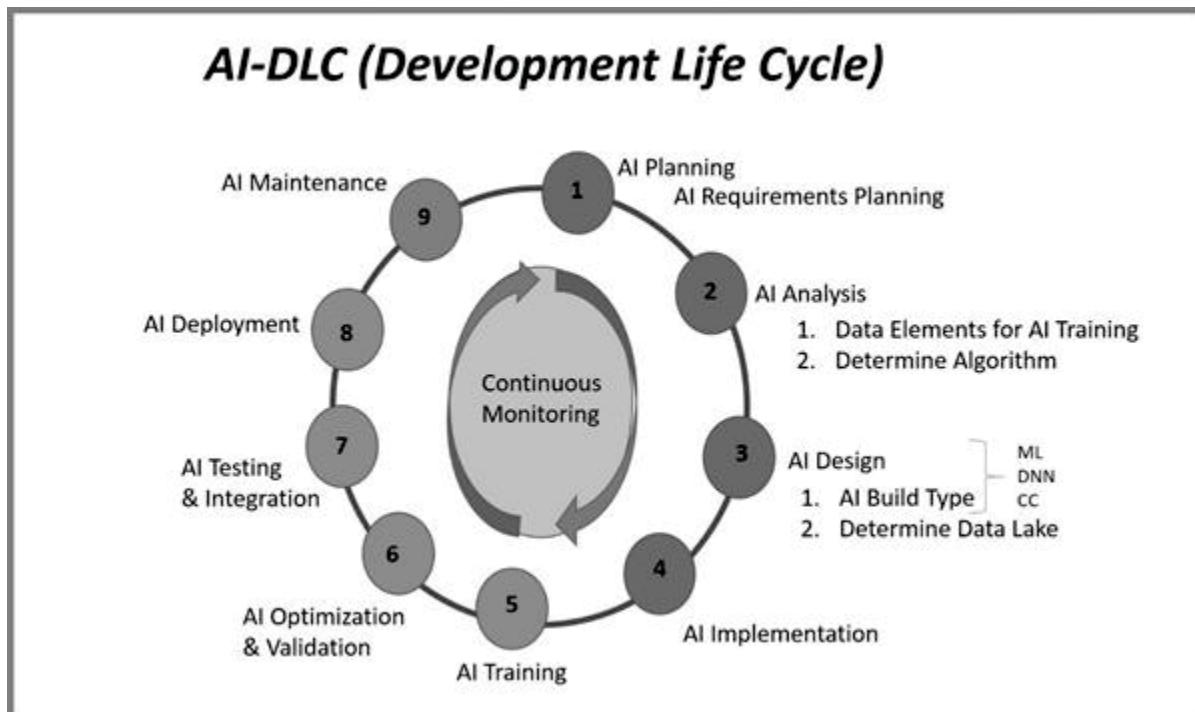
The Table 1 above represents ten AI projects which deviated or performed differently from the intended AI design. The deviations resulted in lack of variables or specificity in the original program design, lack of forethought in the intended behavior of the AI system, and or lack of understanding of the AI system's reaction in a production environment versus a test environment. It is important to understand the deviations in the intended AI system and expected behaviors for these could lead to significant cybersecurity issues.

Another concern with AI data results is the issue of reproducibility on the data set. Dr. Allen was quoted at the American Association for Advancement of Science's annual meeting in questioning scientific discoveries resulting in machine learning from data sets because of erroneous AI programming and lack of reproducibility on the data sets (Cookson, 2019).

#### **4. AI Development Life Cycle (AI-DLC)**

Before the AI Cybersecurity Framework can be addressed, the development of AI systems should be reviewed and determined. Armstrong (2014) stated the problems with AI mathematics should be reconciled so the AI can be programmed safely. This same theme was reiterated by Allen, at the AAAS meeting, stating AI programming requires extended thought and execution to ensure the AI discoveries are reproducible from the dataset (Cookson, 2019). Adhering to an AI application framework would assist and mitigate Armstrong's concern. In developing artificial intelligent systems, there are key elements which pertain to AI when planning, designing, implementing and deploying the AI system. The AI Development Life Cycle (AI-DLC) was created by marrying the traditional SDLC with specific AI design and technology. The AI-DLC can be applied to traditional software designs for artificial intelligent systems. For AI systems being designed in Dev-Sec-Ops, the process is addressed later in this paper. The AI DLC, shown in Figure 2.

Figure 2 depicts the nine phases in the AI-DLC. Phase one is the planning stage for the AI system. AI requirements, security and privacy are established here. Phase two covers the AI analysis of the AI system. Phase three addresses the AI Design including the AI build type. Phase four is the AI implementation where you create the AI system from the design. Phase five is for training the AI designed with the designated data set. Phase six covers the optimization and validation for the AI to ensure the algorithm and the data are what was intended. Phase seven covers the testing and integration of the AI technology and design and ensure the requirements have been met. Phase eight is the deployment of the AI system to production. Phase nine is the maintenance of the AI system in production. The center of the figure represents the continuous monitoring of code, algorithms, learning, and output along with all the cybersecurity components being performed.



**Figure 2:** AI-DLC (Artificial Intelligence Development Life Cycle)

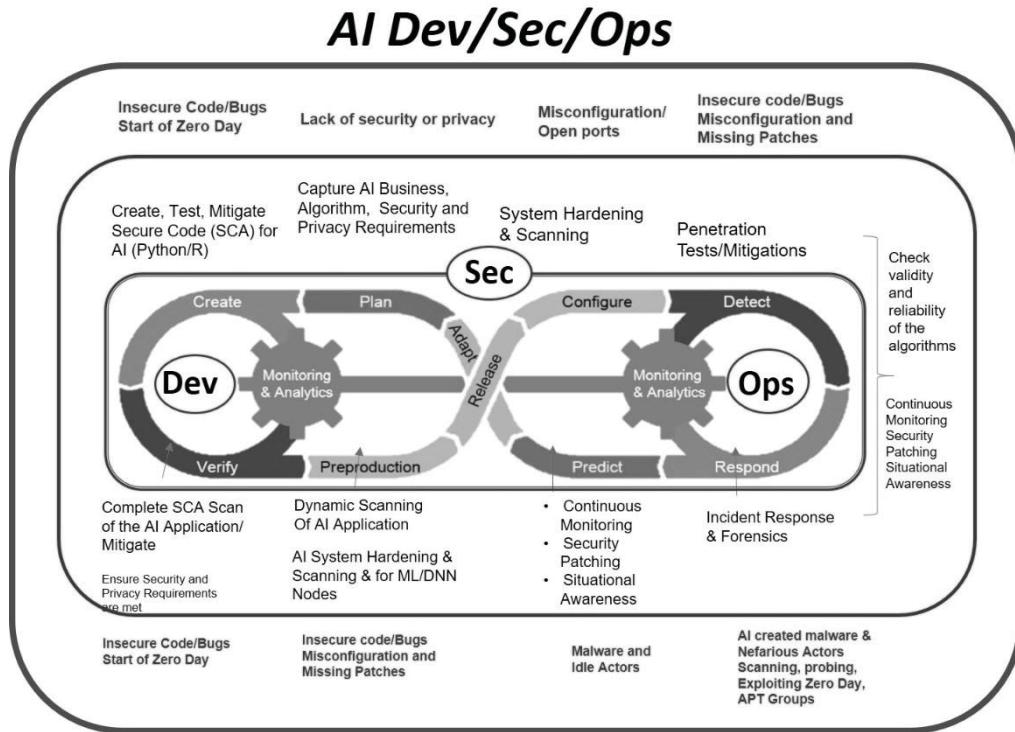
From a cybersecurity perspective, the AI-DLC when not properly followed can introduce security and privacy issues which will be problematic for the AI system. It is important each phase properly addresses cybersecurity. The table 2 below provides cybersecurity tasks which ought to be addressed:

**Table 2:** AI DLC cybersecurity function

	AI-DLC Phases	Cybersecurity Function Pertaining to an AI System
1	AI Planning	Ensure security and privacy requirements are developed. Consider security architecture, access control and encryption for the AI system and data requiring protection in motion and at rest.
2	AI Analysis	Determine the type of algorithm(s) to be used and the data sets.
3	AI Design	Determine AI build type and data lake meeting security and privacy requirements. Ensure secure code is baked into the AI code.
4	AI Implementation	Ensure security and privacy requirements are met in the implementation.
5	AI Training	Ensure the data cannot be poisoned thus generating bogus results. Ensure the end results are reproducible.
6	AI Optimization & Validation	Confirm cybersecurity hardening and patching are met and the system is running accordingly.
7	AI Testing and Integration	Determine security and privacy requirements are met. Ensure the data is properly protected. Ensure AI code is hardened, and vulnerabilities are remediated. Ensure the AI algorithms can reproduce the same results from the dataset.
8	AI Deployment	Adhere to system hardening, patching and begin continuous monitoring for the AI System covering all nodes of the deep neural net and cognitive computing.
9	AI Maintenance	Ensure continuous monitoring is facilitated.

Table 2 depicts the AI-DLC phases in column 2 and it provides the cybersecurity functionality of each phase in column 3. Threat forecasting should be performed at each phase taking into consideration indicators of interest (IOI), indicators of attack (IOA), and indicators or compromise (IOC) (Pirc, Desanto, Davison & Gragido, 2016). Additionally, risk assessments should be conducted at each phase to ensure identified vulnerabilities can be remediated. This includes performing threat tree analysis and secure code review. It is imperative to continuously check the algorithm and the output to ensure the intended results are being generated and fully understood. It is also important to perform security and function tests throughout the life cycle to ensure the AI system is behaving and functioning as planned.

AI systems can be developed in Dev/Sec/Ops. The programming language will vary and can be Python or R. In the event the AI developer is using the Waterfall or Agile methods, AI can be achieved in this manner. AI in the Dev/Sec/Ops model is provided below in Figure 3.



**Figure 3:** AI dev/sec/ops model

Figure 3 depicts AI implementation in a Dev/Sec/Ops where the same elements of the AI-DLC are facilitated but the development cycle is performed in sprints. An infinity symbol was provided by a Gartner report and depicted on Cigniti's website<sup>1</sup>. The text was added around the figure to address how AI development and cybersecurity can co-exist. The text in black shows the daily cybersecurity activities needed to get the AI system up and running. The red text in the outer swim lane shows risks and nefarious activity which can happen in the Dev/Sec/Ops process. The inner text provides the mitigations to the risks identified.

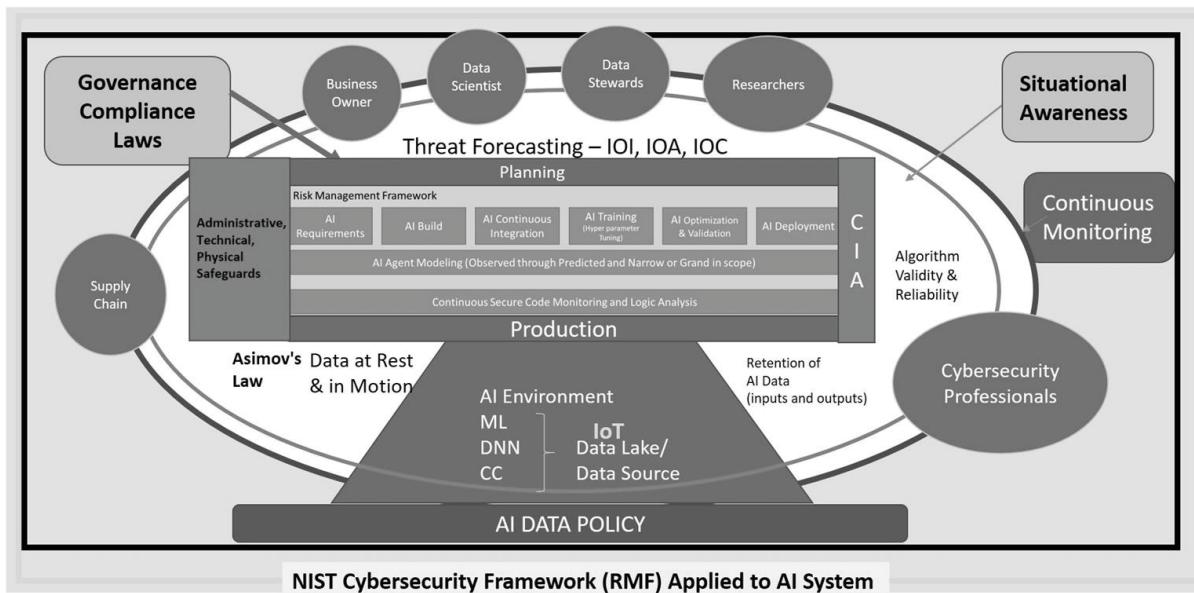
As AI systems come online across multiple platforms and industries, cybersecurity should be formally and thoroughly addressed. There are cyber architecture nuances which ought to be considered for AI systems pertaining to cyber security and privacy. The next section will introduce the AI Cybersecurity Framework which ought to be implemented to address cyber concerns in AI systems.

## 5. AI cybersecurity framework

To ensure cybersecurity addresses artificial intelligence implementation, an AI Cybersecurity Framework (AI-CF) has been created to accompany the AI-DLC in Figure 2 and AI Dev/Sec/Ops in Figure 3. The Cybersecurity framework is aligned to known cybersecurity models integrating artificial intelligence into information systems across multiple industries. The AI-CF covers machine learning, artificial or deep neural nets, and cognitive computing. The model is designed to help cybersecurity professionals facilitate security and privacy for AI systems and take into consideration the complexities of the AI system. The framework is depicted below in Figure 4.

<sup>1</sup> <https://www.cogniti.com/blog/what-you-need-to-know-about-devops-and-security-testing/>

## AI Cybersecurity Framework



**Figure 4:** AI cybersecurity framework

The AI-CF encompasses the AI-DLC. Figure 4 depicts the framework which should be used by cybersecurity professionals to harden an AI information system technology system. The framework can be applied across all industries and address small and large AI projects. The framework starts with the implementation of the National Institute of Standards and Technology Risk Management Frame (RMF) work (NIST, 2010). The AI-CF also incorporates threat forecasting for the entire AI-DLC.

The planning phase addresses the AI system considering the business owner's stated business requirements and functions, as well as, governance, laws, and compliance. Threat forecasting should also start to ensure risks are identified. Asimov's Law should also be addressed in the planning phase to ensure the law is core to all AI systems being developed. Asimov's law is intricate in application, so the design and implementation need to be reflective of what and how AI can meet the laws in the AI application (Chace, 2015).

As AI architects design complex AI systems including data architects designing data lakes, Asimov's Law should be incorporated into the core code to ensure at a later and more precarious time the AI system does not harm humans. This pertains to robotic systems as well as the robotic systems include AI technology. Asimov's law states rules which dictate the AI cannot harm the creator or humans. The reason for including this law into the constructs of the code is to ensure humans are not harmed now or in future implementation of the AI system when High AI or Super AI comes to fruition.

The AI Cybersecurity Framework introduces primary “actors” on the outside of the circle (i.e., data scientists, business owners, data stewards, researchers, individuals in the supply chain, and cyber security professionals). The ‘circles of defense’ (depicted above with the blue and green oval lines) for AI systems and the AI data commences with situational awareness and then to continuous monitoring from the start of the AI-DLC or AI Dev/Sec/Ops to the decommission of the AI system. Continuous monitoring, part of RMF includes the review of the AI code to the review of AI system ensuring no new vulnerabilities are introduced and vulnerabilities identified have been mitigated.

The planning phase is important because this phase builds out all the requirements for the proposed AI system including the mathematical algorithm(s), number and complexity of the nodes, and data elements needed. Asimov's Law should be built into each AI system and robotics system with AI to ensure the AI programming is complicit and meets the requirements. As the militaries move closer to the concept of Hyperwar, Smart Weapons (AI weaponization), and we move closer to High AI, it is important to ensure Asimov's law is properly configured in the AI code (Allen & Husain, 2017; Husain et al., 2018). “Complex and multidimensional” trust

relationships with AI will also have to be established ranging from AI systems, to robotics to autonomous robots and drones (OSD, 2019).

The next phase of the AI Cybersecurity Framework looks at the Risk Management Framework (RMF) and the AI-DLC. In this section of the framework, the AI project team including cybersecurity should ensure from planning to implementation to production the AI system is hardened, secured, and facilitation of continuous monitoring for risks both internal and external are properly addressed. It is also necessary to ensure the AI system, rational agent, is performing as expected and performance measures are met (Russell & Norvig, 2015). The privacy of AI data is addressed and the trust relationships are established. Figure 4 also represents AI data where all data protections should be met when dealing with PII, PHI, PCI/SOX, and FTI data.

The researchers advocate utilizing de-identified data sets for AI data and should follow acceptable standards to ensure the AI data cannot be re-identified. Re-identification of de-identified or redacted data is possible when not enough data has been de-identified or redacted (Hayman, 2017). Precaution should be taken in determining de-identification standards and facilitating de-identification for AI data sets.

No model will absolutely provide assurance of prevention of an AI system being infiltrated by a nefarious actor. The AI Cybersecurity Framework while believed comprehensive in nature, maintains restrained assurance where cyber criminals strive to gain entry of least resistance to infiltrate the AI system directly or through its supply chain. Consequently, the AI cybersecurity framework, like many cybersecurity frameworks, is not 100% sound; however, the AI cybersecurity framework facilitates the proper direction in securing an AI system and protecting the AI data more so than current literature suggests in securing an AI system.

## **6. AI cyber issues**

When a cybersecurity professional implements the AI Cybersecurity Framework, there are 18 identified cybersecurity issues which must be taken into consideration. The AI cybersecurity issues were created from a review of how AI operates, has operated, deviations in the original AI implementation, industry concerns, and from a cybersecurity perspective. The researchers also looked at the AI system's initial implementation and utilization identifying cybersecurity gaps or lack of cybersecurity implementation. The researchers identified several cybersecurity issues, presented in Table 3.

**Table 3:** AI cybersecurity concerns

Type	Description of AI Cybersecurity Issue
AI Design	Integrity of algorithms and output – bias or external bias
Code	Secure code analysis of AI code/functions/AI generated code & functions
AI Design	AI data security measures – ingress and egress of AI data
Code	Catastrophic bugs found by AI and exploitation
Code	AI generated malware –understand how AI could use malware for nefarious purposes
AI Design	Neural nets not being secure
Supply Chain	AI Supply chain concerns for rogue hardware (i.e., motherboard neural net components)
Nefarious Use	Dual use of AI developed for good and nefarious purposes
Nefarious Use	AI Contamination – Nefarious actors target AI to alter learning
Privacy	AI Data lakes – privacy issues with mass data collected
Privacy	Concatenation of data causing sensitive data and exposure
Privacy	Algorithms could result in exposure of sensitive data
Nefarious Use	AI directed to create misinformation campaigns and AI Rogue creates misinformation campaigns
AI Design	Assigning mis-specified goals or wrong goals to AI by mistake, incomplete or wrong requirements
AI Design	Variables could be added to an AI system causing undesirable outcome
AI Design	AI Systems behaving in unintended ways particularly when AI experiences a new scenario
Nefarious Use	AI could cause a good AI system to behave poorly or go rogue
Nefarious Use	AI can morph for offensive and defensive measures and countermeasures
Trustworthiness	AI will need trust relationships being multi-dimensional
Reproducibility	AI algorithms will need to reproduce the finding from the dataset

The first column categorizes the type of cyber issue and or the AI-DLC phase where cybersecurity was not appropriately addressed. The second column depicts the description of the AI cybersecurity issue. As AI progresses, quantum computing is utilized, and AI technology advances, more cybersecurity issues will arise and will be added to the list. In addressing AI systems, the researcher should provide checks and balances on the test problem, and identify the cause of the failures based on the Turning test (Yampolskiy, 2016). In preparing for the AI Cybersecurity Framework for the proposed AI system, it is important to review and understand the various phases and development of AI systems based on the AI-DLC.

## **6.1 Dual use of AI tools**

Another AI cybersecurity concern is the dual use of AI systems. Dual use will be a concern for AI created cybersecurity systems. As with any software created, particularly security software, there is a potential for the application to have a dual use meaning it is used for good but placed in the wrong hands it can also be used for nefarious purposes.

Furthermore, the cybersecurity field will have a new and more difficult job in defending against AI generated malware. Threat hunters will have a more difficult time locating AI generated malware as it obfuscates, encrypts, exploits, and removes itself from the environment. Cybersecurity professionals will need a much deeper understanding of computer and system engineering and securing code at all levels as it relates to AI created systems and malware.

## **6.2 Defensive and offensive AI**

As AI is designed and used for offensive and defensive use, it will be important for a cybersecurity professional to understand the AI applications and their functionality. Cyber architects to CISOs will need to learn and understand how these tools can ease their workload as well as provide further offensive or defensive functionality. The cybersecurity professional should be aware these tools could also be used for nefarious use. In reviewing how the AI programs went rogue in Table 1, one can presume the AI offensive and defensive applications will be quick and efficient.

## **7. AI cyber audit**

As artificial intelligence is integrated into information systems, cybersecurity and auditors will need to understand what needs to be reviewed including the various artifacts. Network diagrams and descriptions will need to include the artificial intelligence integrated into the information system. The AI auditor will need to review and check for the integrity and application of the algorithm(s) used in the system. Ingress and egress of data points in the system will have to be analyzed to ensure the AI system is using the correct data and the integrity of the output is planned. The AI auditor will need to know all the components of the AI system. For example, if it is an AI system using deep neural nets, say 100 nodes, then the AI auditor will need to ensure through test, observation, and interviews those nodes are hardened and secure. The AI auditor will need to review the AI system at each AI-DLC phase to ensure the system is performing as planned and designed. There will be new security and privacy controls to be developed for the AI system.

## **8. Future research**

Further work is required to ensure the proposed AI Cybersecurity Model addresses all facets of artificial intelligence technology and will have to continue to evolve with the new technology created as a result of the fourth industrial revolution. The model will be a living model and will evolve as new technological advances are developed. Detailed AI security and privacy controls should be created as well.

## **9. Conclusion**

The paper explored the use and implementation of an AI cyber security framework for artificial intelligence systems. AI is integrated into information technology environments lacking cybersecurity. This creates the necessity for AI systems to meet cybersecurity requirements for being secure and hardened. Current and new AI technology will require the cybersecurity professional to rethink the standard cybersecurity approach and require a deeper understanding of AI and how an AI system ought to be secured and hardened, and to ensure the AI system behaves as prescribed from the business and security/privacy requirements.

## References

- Allen, J. & Husain, A. (2017). On Hyper-War. Available at <http://www.usni.gov/Magazines/proceedings/2017-07/hyperwar>
- Armstrong, S. (2014). Smarter than us: The rise of machine intelligence. Berkely, CA: Machine Intelligence Research Institute (MIRI).
- Chace, C. (2015). Surviving AI: The promise and peril of artificial intelligence. London, UK: Three Cs Publisher Limited.
- Cookson, C. (2019). Scientist warns against discoveries made with AI. Financial Times. Available at <https://www.ft.com/content/e7bc0fd2-3149-11e9-8744-e7016697f225>
- Hayman, J. (2015). *Case study: Suggested best practices for redacting U.S. Army aviation accident reports to reduce opportunities for doxing re-identified U.S. Army Aircrew* (Unpublished doctoral dissertation). Capitol Technology University, Laurel, MD.
- Hayes, P.J., Ford, K.M. and Adams-Webber, J.R., 1992. Human reasoning about artificial intelligence. *Journal of Experimental & Theoretical Artificial Intelligence*, 4(4), pp.247-263. Available at [https://www.researchgate.net/profile/Kenneth\\_Ford/publication/220813820\\_Turing\\_Test\\_Considered\\_Harmful/links/09e4150d1dc67df32c000000.pdf](https://www.researchgate.net/profile/Kenneth_Ford/publication/220813820_Turing_Test_Considered_Harmful/links/09e4150d1dc67df32c000000.pdf)
- Husain, A., Allen, J., Work, R., Cole, A., Scharre, P., Porter, B., Anderson, W., & Townsend, J. (2018). Hyperwar: conflict and competition in the AI century. Austin, TX: Spark Cognition Press.
- National Institute of Standards and Technology [NIST], (2010). Special Publication 800-37 Revision 1 Guide for Applying the Risk Management Framework to Federal Information Systems A Security Life Cycle Approach. Washington, DC: Author. Available at <http://dx.doi.org/10.6028/NIST.SP.800-37r1>
- McCarthy, J., 1998. What is artificial intelligence? Retrieved from <https://cogprints.org/412/2/whatisai.ps>
- Merriam-Webster website. Available at <https://www.merriam-webster.com>
- Neff, G. and Nagy, P. (2016). "Automation, algorithms and politics| Talking to bots: symbiotic agency and the case of Tay". International Journal of Communications, 10, p. 17.
- Office of Secretary of Defense (OSD), (2017). Unmanned systems integrated roadmap: 2017-2042. Available at [http://cdn.defensedaily.com/wp-content/uploads/post\\_attachment/206477.pdf](http://cdn.defensedaily.com/wp-content/uploads/post_attachment/206477.pdf)
- Pirc, J., Desanto, D., Davison, I., & Gragido, W. (2016). Threat forecasting: Leveraging big data for predictive analysis. Boston, MA: Elsevier.
- Risley, J. (March 24, 2016). "Microsoft's millennial chatbot Tay.ai pulled offline after Internet teaches her racism", Geek Wire. Available at <https://www.geekwire.com/2016/even-robot-teens-impressionable-microsofts-tay-ai-pulled-internet-teaches-racism/>
- Rouhiainen, L. (2018). *Artificial Intelligence: 101 things you must know today about our future*. (n.p.): Author.
- Russell, S., & Norvig, P. (2015). Artificial intelligence: A modern approach, 3<sup>rd</sup> Edition. London, UK: Pearson Education, Inc.
- Schwab, K. & Davis, N. (2018). *Shaping the fourth industrial revolution*. Davos, Switzerland: World Economic Forum.
- Sivak, M. and Schoettle, B., 2015. Road safety with self-driving vehicles: General limitations and road sharing with conventional vehicles. Available at <https://deepblue.lib.umich.edu/bitstream/handle/2027.42/111735/103187.pdf?sequence=1&isAllowed=y>
- Turing, A.M., 2009. Computing machinery and intelligence. In *Parsing the Turing Test* (pp. 23-65). Springer, Dordrecht. Available at <https://cogprints.org/499/turing.html>
- Yampolskiy, R. (2016). Artificial intelligence: A futuristic approach. New York, NY: CRC Press.
- Yolgomez, C. (2019). Machines that collaborate, disrupt, and make change challenge our notion of human-centered society. Available at <https://washingtonspectator.org/machines-that-collaborate-disrupt-and-make-change-challenge-our-notion-of-human-centered-society/>

# A Performance Study of the Ethical Hacking Capabilities of Single Board Computers

Adolfo Jose Arias de la Vega and Christina Thorpe

Technological University Dublin, Ireland

[B00092526@student.itb.ie](mailto:B00092526@student.itb.ie)

[Christina.thorpe@itb.ie](mailto:Christina.thorpe@itb.ie)

**Abstract:** The rate and diversity of cybercrime is growing rapidly. It has become more commonplace, more sophisticated, and more damaging, costing the world's economy billions of euro in losses every year. To counteract this rising menace, organisations are continuously increasing their investment in information security and cybersecurity, including ethical hacking or penetration testing services. Traditionally, ethical hackers have always used expensive computationally powerful hardware to run special security-oriented Linux distributions. However, the advent of single board computers in the last decade, has offered new possibilities for ethical hacking. This research aims to explore, in a practical manner, the actual feasibility of performing ethical hacking duties in a non-traditional way, i.e., following a novel approach that still uses the same Linux security distributions, but installing them on these low-cost, highly portable small boards instead. The outcome is a comprehensive study that contributes to the body of work in the literature. We conducted extensive performance tests on 4 devices (3 boards and 1 laptop); results are varied and show that the small boards are capable of performing many jobs, albeit with a larger execution time than a laptop, but in general they are limited in terms of running heavy GUIs.

**Keywords:** penetration testing, ethical hacking, performance study, single board computers

---

## 1. Introduction

For the last number of years, many technology analysts have been forecasting the death of the traditional PC. Although such pronouncements seem to be somewhat premature, it is clear that while traditional PC sales continue to decline, sales of smaller mobile devices are thriving, especially in emerging countries (Atwal, 2013). Technology sales are directly affected by user preferences; and the current trend seems to be that ultra-capable smart phones and powerful tablets have started replacing PCs and laptops in households and offices (for computationally-easy tasks) around the world. Each year, a new wave of mobile devices is released, more powerful and user-friendly than the generation before them. A 2016 report by Ofcom, the communications regulator in the UK, found that UK users are now more likely to go online via their smart phone (66%) than via traditional computers (62%) (Ofcom, 2016). The reduction in device size has been made possible by the development of smaller, highly capable, more energy efficient chips.

Virtually all mobile devices currently on the market are based on the ARM processor architecture, which was originally developed in the UK by Acorn Computers Ltd. during the early/mid 1980s. This approach represents a departure from the traditional PC x86 architecture. These smaller ARM-based processors are also subject to Moore's law; hence, they have become much more powerful and much cheaper over the last number of years. This current combination of small size, high computing power and decreasing costs has aided the development of the Internet of Things (IoT). As such, it has facilitated the placing of such chips inside devices, for which, acceptable performance coupled with low power consumption and low price are the top priorities. One such class of devices is the Single Board Computer (SBC) (Pajankar, 2017).

Penetration testing (Goel, 2015) (or pentesting/ethical hacking) is the process of applying hacking tools and techniques in order to expose vulnerabilities that are present in any system. From a hardware perspective, penetration testers tend to use a wide variety of different brands of high-performing traditional PCs and laptops, alongside other expensive devices. From a software point of view, penetration testers tend to choose their toolkit from a small pool of readily assembled open source operating systems, with Kali Linux being the most popular. Consequently, there has been significant research into "traditional" pentesting, i.e., pentesting performed almost entirely by using Kali Linux running on a middle or high-end PC or laptop. However, there is a gap in the literature when it comes to ethical hacking performed with "non-traditional" devices. Assuming they can provide a sufficient level of performance, SBCs have the potential to deliver enormous value for pentesters beyond the 'cool' factor. For example, their small form factor make SBCs the perfect candidates to be easily hidden at a target's site in a way that a bigger device could not be. Moreover, even in the event of being discovered, their low cost makes them easily replaceable (compared with more expensive equipment).

Given the potential benefits that may result from using SBCs for ethical hacking purposes, the primary objective of this work is to analyse the overall pentesting capabilities of SBCs. This will involve testing them with different ethical hacking OSs, and then comparing their performance and behaviour to a “traditional” pentesting device, a laptop.

## **2. Related work**

SBCs have gained popularity over the past decade; however, there has been very few studies into the performance capabilities of these devices. (Alee, 2011) compares the performance between operating systems on different kernels (2.6.34 and 2.6.21). (Morabito, 2016) presents a performance evaluation of container technologies on constrained devices, in this case, on Raspberry Pi. (Baun, 2016) describes the construction and performance analysis of a cluster made from SBCs. This research presented in this paper focuses on investigating the practicality of applying SBCs to an ethical hacking context. To the best of our knowledge, there is no existing in-depth study that can illustrate how big the pentesting performance gap is between an SBC and a traditional device. This gap in the literature is made even more critical considering that many ethical hacking OSs do provide an image for both x86 architectures and ARM-based SBCs.

## **3. Objectives and methodology**

The overall objective of this research is to test the ethical hacking capabilities of modern SBCs, in order to determine if they can be used in real world pentesting scenarios. In order to assess this potential circumstance, 3 different contemporary SBC from 2 different manufacturers have been chosen. In addition to that, 3 different ethical hacking OSs have been selected, to ensure that the outcome of this research is not defined or heavily influenced by the choice of OS.

To test the SBCs competences, the same 10 software tools will be tested on the 3 different OSs. These tools have been identified as the top 10 most used hacking tools and are essential for any hacking activity. A test case will be designed for each of the 10 hacking tools, trying to replicate real world challenges. Each test case will be comprised of at least two different tasks to be performed, to ensure consistency in the testing process. Furthermore, each test case will be repeated 5 times per OS, per device. Tests will be carried out under the exact same testing conditions within a controlled networking environment, thus guaranteeing outcome reliability, reproducibility and comparability.

For comparison, a laptop will be used to run the same 3 OSs while operating the same 10 hacking tools under identical conditions as the SBCs. This will be done to obtain quantitative outcomes for the laptop’s performance that can be subsequently compared to the SBCs’ performance, which will gauge the true degree of suitability for the small boards to tackle pentesting duties. All results will be collected, organized, analysed and reported accordingly, and the appropriate conclusions will be drawn.

Furthermore, a Borda scoring system (Baharad, 2003) will be applied to rank the overall performance of each individual SBC (but not the laptop), i.e., for every quantitatively measured experiment, the best performing SBCs will be awarded 3 votes, the 2nd best performing SBC will receive 2 votes and the least performing SBC will get only 1 vote. Votes will be added at the end of the testing phase and this will reveal who the best overall performer is.

## **4. Single board computers**

The term single board computer refers to a single microchip-based PCB, which features an embedded microprocessor as well as memory and the necessary circuitry to provide an Input/Output (I/O) control system. These components allow an SBC to operate as a standalone fully functional computer (Pajankar, 2017). Other frequent characteristics of an SBC are a low profile architecture, small physical size, open-source software (although only a small number are open-source hardware), low power consumption, on-board GPU, and no need for cooling systems.

### **4.1 SBC selection criteria**

Table 1 details the specification of each of the SBCs selected for this performance study. The following 4 conditions were conceived to aid in the selection of the required SBCs for this research:

- 1. All 3 selected SBCs must be relatively recent releases, i.e., from 2015 on-wards.

- 2. All 3 SBCs must be supported by Offensive Security (Kali Linux), and also by at least one of these 2 other projects: the BlackArch project and/or the Parrot Security project (preferable both, but one would be enough).
- 3. Each SBC must cost less than 100 euro, including a clear case, the recommended power adapter, heatsink(s) and a quality 32 GB micro SDHC card.
- 4. The manufacturers of the selected SBC must have a proven track record in the SBC space, as well as dedicated official support, available documentation and adequate software and community backing.

In June 2016, the website Linux.com (which acts as the community site of the Linux Foundation) ran a survey, asking the participants to pick their favourite 3 SBCs for hacking purposes <http://linuxgizmos.com/raspberry-pi-3-takes-the-cake-in-2016-hacker-sbc-survey/> from a given catalogue. Voters ranked their choices in order of preference and subsequently, a scoring scheme was used to determine each SBC's position in the final list. The top 10 hacker-friendly SBCs selected were published online: <http://files.linuxgizmos.com/2016sbcsurvey-top10-by-borda-scores.jpg>. For the purposes of this research, the first 3 models on the above Top 10 list that happen to meet all 4 conditions stated was selected as the 3 boards to be used in this study.

**Table 1:** Single board computer specification

Component	Raspberry Pi 3	Odroid C2	Odroid XU4
<b>SoC</b>	Broadcom BCM2837	Amlogic S905	Samsung Exynos 5422
<b>Architecture</b>	64-bit ARMv8-A	64-bit ARMv8-A	32-bit ARMv7-A
<b>CPU</b>	ARM Cortex-A53 Quad-Core @ 1.2GHz	ARM Cortex-A53 Quad-Core @ 1.5GHz	ARM Cortex-A7 Quad-Core @ 1.4GHz or ARM Cortex-A15 Quad-Core @ 2.0GHz
<b>GPU</b>	Broadcom VideoCore IV Dual-Core @ 400 MHz	ARM Mali-450 MP3 TripleCore @ 750 MHz	ARM Mali-T628 MP6 @ 600 MHz
<b>RAM</b>	1 GB RAM LPDDR2 @ 900MHz	2GB DDR3 SDRAM @ 912MHz	2GB RAM LPDDR3 @ 750 MHz
<b>Extended Storage</b>	MicroSD card slot	MicroSD card slot & eMMC 5.0 socket	MicroSD card slot & eMMC 5.0 socket
<b>USB Ports</b>	4 USB 2.0 + 1 x micro USB OTG	4 x USB 2.0 + 1x micro USB OTG	2 x USB 3.0 + 1 x USB 2.0 (no OTG support)
<b>Ethernet</b>	10/100 Mbits/s	10/100/10000 Mbits/s (Gigabit)	10/100/10000 Mbits/s (Gigabit)
<b>Video Out</b>	HDMI 1.4 H264 1080p	HDMI 2.0 4k/60MHz	HDMI 1.4 (type A)
<b>Audio Out</b>	HDMI / 3.5mm Audio Jack	HDMI / I2S interface	HDMI / optional via USB-SPDIF
<b>Wifi + Blue-tooth</b>	2.4GHz WiFi 802.11b/g/n + Bluetooth 4.1	No (optional USB Wi-Fi dongle)	No (optional USB Wi-Fi dongle)
<b>Power</b>	5V / 2.5A DC input (micro USB OTG port)	5V / 2A DC input (2.5 x 0.8 mm power barrel)	5V / 4A DC input (5.5 x 2.1mm power barrel)
<b>Dimensions</b>	8.5cm x 5.6cm	8.5cm x 5.6cm	8.3cm x 5.8cm
<b>Weight</b>	45g	40g	60g (including fan)
<b>Cost</b>	<b>€71.10</b>	<b>€91.20</b>	<b>€98.78</b>

The System-on-a-chip (SoC) approach to SBC design reduces costs, power consumption, size and heat, hence, it is very much ideal for small, portable devices, not only SBCs but also tablets, mobiles, etc. When it comes to the 3 SBC models used for this research:

- The Odroid XU4 comes with a Samsung Exynos 5422 octa-core System-on-a-Chip (SoC), which showcases 4 x ARM Cortex-A7 CPU cores (low performance) and 4 x ARM Cortex-A15 CPU cores (higher performance). Both the ARM Cortex-A7 and the ARM Cortex-A15 are 32-bit processors.
- The Raspberry Pi 3 comes with the Broadcom BCM2837 SoC, which features 4 x ARM Cortex-A53 cores (clocked at 1.2GHz), i.e., a quad-core CPU. The Coretx-A53 supports 64-bit instruction sets.
- The Odroid C2 comes with the Amlogic S905 SoC, which (same as the Raspberry Pi 3) features 4 x ARM Cortex-A53 cores, but clocked at a higher 1.5GHz (although often incorrectly reported as 2GHz).

Therefore, even though the Raspberry Pi 3 and the Odroid C2 contain the same ARM processor, these come within different SoCs that have been designed by different manufacturers (Broadcomm and Amlogic respectively).

## 4.2 Laptop specification

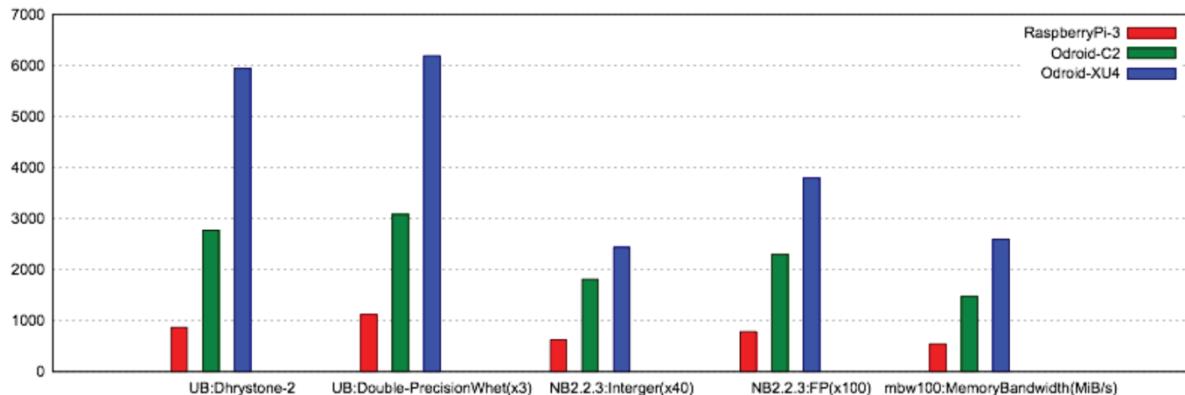
The laptop used in this project is a Sony Vaio unit purchased in mid-2014 (see Table 2):

**Table 2:** Specification of laptop

Arch	x86 64-bit (Intel HM76 Express Chipset)
Processor	Intel Core i7-3612QM Quad-core @ 2.1 GHz
GPU	AMD Radeon HD 7650M (dedicated video memory: 2GB)
RAM	8GB @1333MHz DDR3 SDRAM
Storage	Toshiba HDD SATA - 1TB @ 5400rpm
Ethernet	10/100/10000 Mbits/s (Gigabit Ethernet)

### 4.3 Performance benchmark comparison

Figure 1 outlines a performance comparison of the 3 selected boards using different benchmark suites and various benchmark tests, including: Unixbench: Dhystone2, Unixbench: Double-Precision Whetstone (x3), Nbench 2.2.3: Interger (x40), Nbench 2.2.3: Floating-Point (x100), and mbw100: Memory Bandwidth (MiB/s). **Results suggest that the Odroid XU4 will outperform the other 2 more modern boards.**



**Figure 1:** Performance benchmark results of SBCs

## 5. Ethical hacking

The quality and appropriateness of the tools employed by an ethical hacker will have a substantial impact on the quality of their work. In the recent past, an integral component of any pentesting endeavour would involve the pentester building a custom pentesting toolkit almost from scratch. This was a time-consuming process that entailed choosing an OS, researching and assembling the necessary (and favourite) tools one by one, manually updating them and integrating them into the same toolkit while trying to avoid incompatibilities, dependencies conflicts and other pitfalls (Faircloth, 2016).

### 5.1 Ethical hacking Linux distributions

Nowadays there are quite a few readily-assembled open source penetration testing toolkits. Several Linux distributions have been developed with ethical hacking purposes in mind (e.g., Kali Linux, Backbox Linux, Pentoo Linux, BlackArch). These singular distributions were initially introduced to tackle the time-consuming issue of toolkit self-building.

To choose 3 suitable distros, we used an online service called distrowatch, as it provided a comprehensive distro ranking as well as a description of the individual distros (Castro, 2016). We specified three conditions that the distros selected for this research must meet:

- Be a pre-compiled, current, stable and established ethical hacking distro.
- Provide images for both x86 64 architectures and ARMbased SBCs.
- Occupy a prominent spot on the Distrowatch ranking, which can be interpreted as a measure of its popularity.

A special issue of the Linux User and Developer Magazine on Information Security (number 174, Infosec Special Issue 2017, page 9) specified 5 distros as the best in terms of ethical hacking purposes. These are: (1) Kali Linux, (2) Pentoo Linux, (3) Parrot Security OS, (4) DEFT Linux, and (5) CAINE. However, DEFT Linux and CAINE are suites

heavily focused on the digital forensics field, hence not ideal choices for pentesting objectives. Moreover, Pentoo Linux does not provide an image for ARM-based devices, which according to the eligibility criteria established for this work, rules it out. (Stevensson, 20) recommends the BlackArch distro; it meets the criteria specified in this research to be considered as a valid and suitable choice for the purposes of this work, i.e., it is an already-compiled, current and proven distro; it supplies a suitable ARM image and lastly, it features prominently on the Distrowatch ranking. Furthermore, Kali, Parrot and BlackArch at the time of selection were the top 3 ranked pentesting distros on the Distrowatch ranking.

## 5.2 Ethical hacking tools

The ethical hacking tools selected for this project are listed in Table 3 with the sectools.org ranking. sectools.org is a website created by Gordon Lyon, aka “Fyodor”, a renowned OS programmer and security expert. sectools.org is referenced many times in the literature, mainly to justify the selection of a particular security tool e.g., (Rao, 2014), (Jain, 2015). It is considered the “de facto” tool ranking among the security professionals. Eight of the chosen tools for this project are ranked by sectools.org, one is not ranked (OWASP Zaproxy), and the remaining tool (Nmap) is excluded from the voting process, however, it is reasonable to think that it would have reached the top 3 had it been made eligible. Furthermore, amongst the 8 ranked tools, 4 of them feature in the top 10, ranked 1st, 2nd, 3rd and 10th, while the remaining 4 are in the top 34 (out of 125).

**Table 3:** Selected tools and their sectools.org ranking (July 2017)

Security Tool	sectools.org
Aircrack-ng	3
Burpsuite	13
(THC) Hhydra	22
John the Ripper	10
Maltego	34
Metasploit Framework	2
Nmap	N/A
OWASP Zaproxy	Not Ranked
SqlMap	30
Wireshark	1

## 6. Experiments

The experiments have been designed to be both qualitative and quantitative in nature. In some cases, the most appropriate way to evaluate performance was to measure the time taken to complete a given task. In other tests, the most appropriate approach was to ascertain if the device could complete the task in a perceived satisfactory manner from the point of view of a potential user. After each experiment, the device under test will be cold booted, (as opposite to just restarting it (warm rebooting)). This is the optimal way to free memory and allowed the OS to perform “housecleaning” tasks. A cold boot implies an OS shutdown as well as having the supply power turned off. This is a fast and easy procedure in an SBC, since they are very quick to boot when compared to a PC. Hence, it is the most suitable manner of starting with a clean environment every time for each iteration of an experiment. The outcome of these types of experiments will be essentially of qualitative substance, i.e., a fact-based opinion as opposed to quantifiable data. Full details about the experiment’s testcases can be found here: <https://tinyurl.com/ECCWS2019Experiments>

### 6.1 Hypotheses

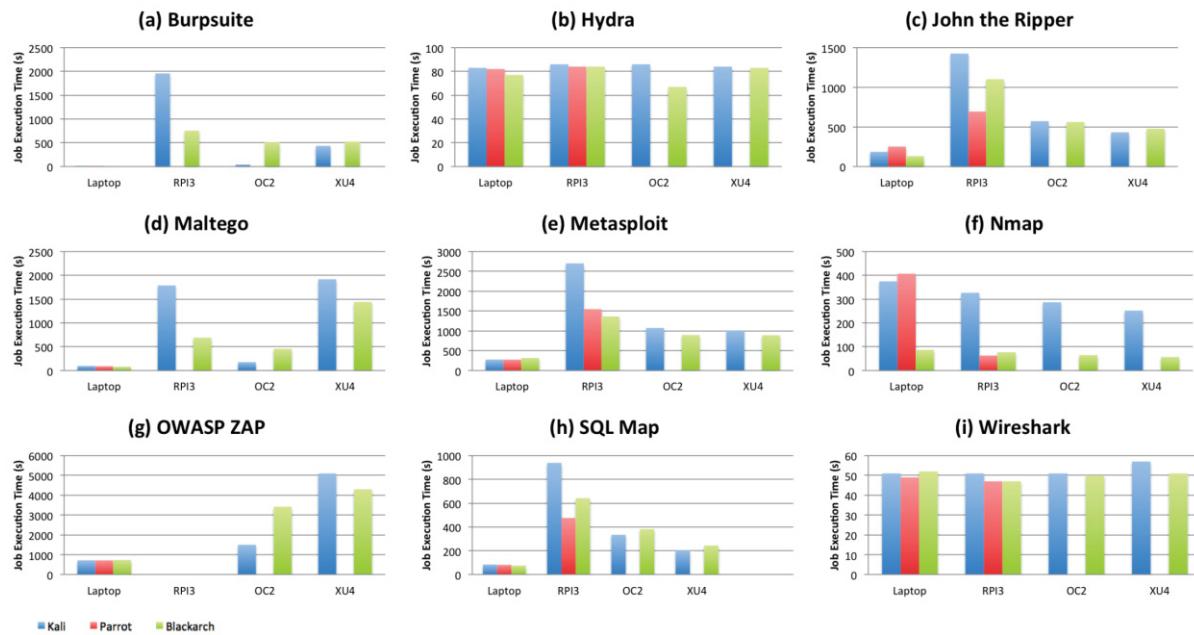
The following hypotheses were formulated for this research:

- The Odroid XU4 will be the best-performing SBC.
- All 3 SBCs will be able to complete all individual tasks, as part of the 10 proposed experiments, in a perceived satisfactory manner.
- The laptop will be the overall best performer in every single task of every single experiment.

## 7. Results

This section presents the results of each test; it also provides a comparative analysis of the performance of each device from the perspective of job execution time. Finally, each of the SBCs are ranked using a borda scoring

scheme to establish the best performing SBC in an ethical hacking context. For the Borda count method, each candidate gets 1 point for each last place vote received, 2 points for each next-to-last point vote, etc., all the way up to N points for each first place vote (where N is the number of candidates/alternatives). The candidate with the largest point total wins the election.



**Figure 2:** Performance test results

## 7.1 Aircrack-ng

All devices completed both parts of the experiment in a satisfactory manner, most of the time this was done in less than 10 minutes. There was no perceived lag or any other issues to report. In fact, there were several occasions in which the SBCs captured the 4-way handshake faster than the laptop (regardless of OS), although the laptop did possess the computational advantage when it came to the use of the “aircrack-ng” cracking tool to obtain the password. There was no perceived difference in the choice of Kali, Parrot or BlackArch, on any board, as experiment completion times greatly varied across devices and OSs, mostly due to the perceived randomness of the deauthentication process, necessary to capture the 4-way handshake. This was not an issue that could be attributed to any device or any distribution.

## 7.2 Burpsuite

The Burp Suite experiment (see Figure 2 (a)) produced the first unexpected result, as the Raspberry Pi 3 could not handle the Burp software in a satisfactory way. Firstly, the Pi took a disproportionate amount of time to open the tool and launch the GUI (especially when running Kali), and subsequently the tool was nearly unusable due to a noticeable lag (although it did continue to work). The 2 Odroids struggled to load the tool when running BlackArch. However, while using Kali, the XU4 model still struggled, but the Odroid C2 was extremely fast. For all these 4 combinations, once the GUI was opened, it was a smooth operating environment, as there were no noticeable issues in running the website crawler. Overall the Odroid C2 was the obvious best performer out of the SBCs (3 borda votes), followed by the Odroid XU4 (2 borda votes), while the Raspberry Pi 3 performed the worst (1 borda vote), although none of the 3 SBCs came remotely close to the laptop’s performance.

## 7.3 Hydra

The Hydra experiment (see Figure 2 (b)) produced the first close result in the testing phase. The single task was to obtain the logon credentials of a remote server, and completion times were very similar across devices and across OSs, with the Odroid C2-BlackArch pairing clocking the fastest average execution time, faster than the laptop. None of the resources were particularly stressed in this test. Overall, the Odroid C2 was again the obvious best performer (3 votes), with the Odroid XU4 achieving the second fastest performance (2 votes), and finishing marginally ahead of the Raspberry Pi in 3rd position (1 vote).

#### **7.4 John the Ripper**

In the password-cracking John experiment (see Figure 2 (c)), the laptop's performance was significantly better than the 3 SBCs. Once more, the Raspberry Pi 3 was the worst performer by a considerable margin, independent of the OS being run. However, during this experiment the Odroid XU4 significantly outperformed the C2 model, when running both Kali and BlackArch. Hence, the Odroid XU4 could be considered a more appropriate board for password cracking activities (3 votes), followed by the Odroid C2 (2 votes) and the Raspberry Pi 3 (1 vote).

#### **7.5 Maltego**

The Maltego experiment (see Figure 2 (d)) proved to be challenging for the SBCs. The Raspberry Pi 3 and the Odroid XU4 struggled significantly during this test, more so the XU4. This performance struggle was more noticeable when running Kali (both boards). When compared to the laptop, both the Raspberry Pi 3 and the Odroid XU4 took a disproportionate amount of time to open the Maltego GUI (10 seconds for the laptop compared to an individual average of approximately 10 minutes for each of 2 boards in question). Furthermore, both the Pi and the XU4 did not fare much better in the 2nd part of the experiment, i.e., applying the transforms to the college website. Again, the difference in performance with the laptop was enormous. However, Maltego did not pose as much of a challenge for the Odroid C2. The C2 - Kali combination produced a performance that was quite comparable to the the laptop in relative terms. The Odroid C2 was the clear best performing SBC (3 votes), with the Raspberry Pi 3 finishing 2nd (2 votes), and the XU4 a poorest performing 3rd position (1 vote).

#### **7.6 Metasploit**

The Metasploit Framework experiment (see Figure 2 (e)) also posed a significant performance problem for the Raspberry Pi 3. The performance difference with the other 2 SBCs as well as the laptop was extremely poor. On average, across all OSs, the Raspberry Pi was approximately 2.5 times slower than any of the other 2 boards, and nearly 7 times slower than the laptop. Although the completion times of the vulnerability scan (the second part of the experiment) for both Odroid models was on average 4 times slower than the laptop, the Odroid XU4 performed overall marginally better than the C2, hence it is ranked the highest (3 votes). The Odroid C2 was the second-best performer (2 votes) and the Raspberry Pi board finished last (1 votes).

#### **7.7 Nmap**

Although the analysis of the individual OSs' performance falls outside of the scope of this research, it was quite evident than the Nmap tool proved to be very problematic for Kali across all devices, since the gulf in execution time (only when a doing network scan) with the other 2 OSs was massive (see Figure 2 (f)). Regardless, on this occasion the performance of each of the 3 boards was superior to the laptop, with the Odroid XU4 - BlackArch combination achieving the fastest performance of all of them. The Odroid XU4 was the clear best performer in this experiment (3 votes), followed by the Raspberry Pi 3 (2 votes) and lastly, the Odroid C2 (1 vote).

#### **7.8 OWASPZAP**

The OWASP ZAP experiment (see Figure 2 (g)) exposed the computational limitations of the Raspberry Pi 3 again, as the Raspberry Pi had to be stopped after it took 3 hours to complete just 10% of the "Forced browse directory", compared to the approximately 12 minutes on average across all OSs that it took the laptop to entirely finish the whole task. The Odroid C2 – Kali combo achieved a remarkable 25 minutes task completion time. Although not as poor as the Raspberry Pi, the Odroid XU4 model also took a massive amount of time (more than an hour) to complete the three sub tasks encompassed in the experiment, independently of the OS used. The Odroid C2 was clearly the best performer amongst the SBCs (3 votes), followed at a large difference by the XU4 (2 votes), which was still significantly ahead of the Raspberry Pi (1 vote).

#### **7.9 SQLMap**

The SQLMap experiment (see Figure 2 (h)) also showed the Raspberry Pi 3 to be the worst performer by a distance. Again, like in many of the previous experiments (John, Burp, Metasploit, Maltego, Zap), Kali was too heavy for the Pi's hardware. Both Odroid models showed quite good execution times, with the Odroid XU4 - Kali being the overall fastest combination amongst the SBCs, but still approximately 3 times slower than the laptop. The laptop remained very stable and under-utilised during all three tasks in this test. The XU4 was definitively

the best performer (3 votes) of the SBCs across all OSs, followed at a considerable distance by the Odroid C2 (2 votes). The Raspberry Pi came way behind the 2 Odroids (1 vote), regardless of OS.

### 7.10 Wireshark

Finally, similar to the Hydra experiment, the Wireshark experiment (see Figure 2 (i)) was also quite close in terms of the performance across devices and OSs. Again, it was demonstrated that the small boards can rival the laptop when running certain tools. Surprisingly, this experiment saw the Raspberry Pi becoming the best performer amongst the boards for the first time. Overall, the Raspberry Pi 3 - Parrot and the Raspberry Pi 3 - BlackArch combinations were the fastest, even faster than the laptop. The Odroid C2 was the second-best performing board (2 votes), and the XU4 fell to the bottom of the table in 3rd position (1 vote), although only for a marginal difference (a matter of seconds).

### 7.11 Borda score

Table 5 details the overall borda votes per tool and the total sum of votes for each SBC. From this, we can see that the Odroid C2 was the best overall performance, the Odroid XU4 was a close second and the Raspberry Pi 3 was significantly poorer.

**Table 5:** Borda score results

Tools	RPI 3	O-C2	O-XU4
Aircrack-ng	N/A	N/A	N/A
Burp Suite	1	3	2
Hydra	1	3	2
John	1	2	3
Maltego	2	3	1
Metasploit	1	2	3
Nmap	2	1	3
ZAP	1	3	2
SQL Map	1	2	3
Wireshark	3	2	1
<b>TOTAL</b>	13	21	20

## 8. Conclusion

This work achieved the objectives defined for the research; a comprehensive set of tests were conducted, and the data generated can be used to address the research question and hypotheses posed. The overarching question was: ***Are low-powered SBC devices capable of performing pentesting activities in real world ethical hacking situations?*** Through analysing the collected data, the first fact that becomes obvious is that from a computational performance point of view, SBCs are significantly behind the bigger x86 devices. However, there were 3 experiments in which the performance of the SBCs generally matched or even surpassed the laptop's, i.e., the Hydra, Nmap and Wireshark experiments. Each of these 3 tools are similar with regards to a lack of a heavy in-built GUI, which seems to be the most significant limitation of the SBCs in general. Although, it is worth noting that the data shows that when it comes to graphical capabilities, the Odroid C2 proved to be a really accomplished device, especially when combined with Kali Linux. The Odroid C2 - Kali pairing showed excellent speed to launch the GUIs of the Burp Suite (42 seconds vs. the Pi's 33 minutes), Maltego (56 seconds vs. the Pi's 10 minutes) and the OWASP Zap (70 seconds vs. the Pi's 10 minutes). The Odroid XU4, which is an octa-core device and the only one of the 3 that incorporates ARM's big.LITTLE technology, was surprisingly disappointing in general, and when dealing with heavy GUIs in particular. Although, it came really close to topping the borda score chart, this does not truly reflect the perceivable gulf in performance with the Odroid C2. An example of this would be the Maltego experiment, in which the even Raspberry Pi outperformed the Odroid XU4. Furthermore, the Raspberry Pi 3 is the best-selling SBC by far, as well as the device of choice of most hackers, but it was really found to be very limited all throughout the testing phase when compared with the other 2 SBCs. Another noticeable occurrence with the Raspberry Pi 3 is that in every single experiment, on average it performed at its worst when running Kali Linux. This raises the question of the suitability of the official Kali image that Offensive Security distributes for the Pi. However, this incongruity is not applicable to the other 2 boards, which worked well with Kali (exceptionally well in the case of the Odroid C2). Reconsidering the hypothesis formulated in section 6.1, they have all been disproved:

- The Odroid XU4 was not the best-performing SBC, the Odroid C2 was.

- All 3 SBCs were not able to complete all individual tasks, as part of the 10 proposed experiments, in a perceived satisfactory manner. The Raspberry Pi 3 offered a very unstable performance at times.
- The laptop was not the overall best performer in every single task of every single experiment, as there was a number of occasions during which it was outperformed, i.e., Hydra, Nmap and Wireshark experiments

## **References**

- Alee, N., Rahman, M. and Ahmad, R.B., 2011, August. Performance comparison of single board computer: A case study of kernel on arm architecture. In Computer Science & Education (ICCSE), 2011 6th International Conference on (pp. 521-524). IEEE.
- Atwal, R., Tay, L., Cozza, R., Nguyen, T.H., Tsai, T., Zimmermann, A. and Lu, C.K., 2013. Forecast: PCs, Ultramobiles, and Mobile Phones, Worldwide, 2010-2017, 4Q13 Update.
- E. Baharad and S. Nitzan, "The borda rule, condorcet consistency and condorcet stability," *Economic Theory*, vol. 22, no. 3, pp. 685–688, 2003.
- C. Baun, "Mobile clusters of single board computers: an option for providing resources to student projects and researchers," *SpringerPlus*, vol. 5, no. 1, p. 360, 2016.
- Castro, J.D., 2016. Arch linux. In *Introducing Linux Distros* (pp. 235-252). Apress, Berkeley, CA.
- J. Faircloth, *Penetration tester's open source toolkit*. Syngress, 2016.
- Goel, J.N. and Mehtre, B.M., 2015. Vulnerability assessment & penetration testing as a cyber defence technology. *Procedia Computer Science*, 57, pp.710-715.
- N. Jain and D. R. Kalbande, "Computer forensic tool using history and feedback approach," in *Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions)*, 2015 4th International Conference on. IEEE, 2015, pp. 1–5.
- R. Morabito, "A performance evaluation of container technologies on internet of things devices," in *Computer Communications Workshops (INFOCOM WKSHPS)*, 2016 IEEE Conference on. IEEE, 2016, pp. 999–1000.
- Ofcom: Office of Communications, "Adults' Media Use and Attitudes," Report, United Kingdom, 2016, [Accessed 23 October 2016]. [Online]. Available: [https://www.ofcom.org.uk/data/assets/pdf\\_file/0026/80828/2016-adults-media-use-and-attitudes.pdf](https://www.ofcom.org.uk/data/assets/pdf_file/0026/80828/2016-adults-media-use-and-attitudes.pdf)
- Pajankar, A., 2017. *Raspberry Pi Supercomputing and Scientific Programming. MPI4PY, NumPy, and SciPy for Enthusiasts*. Apress.
- G. S. Rao, P. N. Kumar, P. Swetha, and G. BhanuKiran, "Security assessment of computer networks-an ethical hacker's perspective," in *Computer and Communications Technologies (ICCCT)*, 2014 International Conference on. IEEE, 2014, pp. 1–5.
- Svensson, R., 2016. *From Hacking to Report Writing*. Apress.

# Investigation and Surveillance on the Darknet: An Architecture to Reconcile Legal Aspects With Technology

Maxence Delong, Eric Filiol and Baptiste David

ESIEA, (C + V) Lab, Laval, France

[eric.filiol@esiea.fr](mailto:eric.filiol@esiea.fr),

[delong@esiea.fr](mailto:delong@esiea.fr)

**Abstract:** While powerful techniques enable to perform efficient forensics and police activities, more or less soon legal aspects limit the capacity for technical action. Aside the fact that what is technically feasible may be legally forbidden, from the judge's point of view, the evidence must be admissible and therefore comply with a rigorous legal framework. Moreover the forensics expert or the criminal investigator must not put himself in danger. In this paper, we are addressing the case of criminal investigation and surveillance over the Darknet. In this parallel network, a lot of criminal activities are conducted and most the time the fact to collect evidences might be considered as itself criminal. We could mention the particular case of child pornography among many others. We have designed an architecture that enables anyone to take part in the surveillance and criminal investigations over the Darknet while complying with all the known legal constraints. Our tools also succeed in bypassing several securities deployed on websites or hidden services such as banishment by IP address or crawler traps.

**Keywords:** data gathering, Darknet, Darkweb, OSINT, surveillance,legal aspects

---

## 1. Introduction

Since the creation of the web, the extraction of web page content (*web scraping*), via a script or a program, has become an essential technique for purposes as diverse as referencing, analysis or intelligence gathering, statistical development or metadata collection. This technique is also used for malicious purposes such as phishing activities. A large number of tools have emerged. From complete libraries such as *Scrapy* (Scrapy, 2019) in Python language to complete commercial solutions such as *HTTrack* (Roche, 1998) or *Parsehub* (Parsehub, 2018), the offer of web extraction programs is now rich and varied.

In 2017, a trial opposing HIQ and LinkedIn (United States District Court Calif, 2017) was opened to decide whether HIQ was allowed to harvest data on LinkedIn even if it was supposed to be forbidden according to LinkedIn's terms of use, or not. Would the harvesting of websites not fully legally defined, the possession of illegal data is. In France, the simple detention of some kind of data is considered as a crime and is prosecuted. For instance, the article 227-23 of the Penal Code (French Penal Code, 2013) in France -- as it is the case in most western countries -- strongly regulates the possession of images concerning the child pornography. In each country of European Union, similar laws are enforced. In this context, our primary and original purpose is to collect information on the "dark web" and more especially on the TOR network. The aim is to provide a technical and citizen support for criminal activities, identification of criminals and support police investigations. According to the TOR foundation who manages the TOR network, the anonymity and the privacy are ensured all along the browsing. This is the reason why illegal and immoral activities are legion on it. To avoid falling under the law, our motivation is to create new tools for crawling and scraping websites (Internet) or hidden services, and collect sensitive, illegal data for the benefit of the police forces while being compliant with regulations in place.

Most of existing tools are SEO crawlers (*Search Engine Optimization*), meaning that they do a lot of treatment related to indexing. For example, a few crawlers are doing either a keywords extraction (via N-Grams methods), or an identification of important metadata (by looking for duplicated data like `<h1>` tags), two techniques that highly resource-consuming for any bot we may use. Even if data analysis is performed separately from link extraction, the latter operation is slower because the crawler must work incrementally and gradually. This requires keeping certain information such as the number of clicks (links visited) to reach a given page, the depth of a web page or other information related to the consultation of the website. Another type of crawler can be found on the web. They are designed to copy a full website from the domain name. They are faster than SEO crawlers because they do not perform any treatment on webpages (except the link extraction). However they are not as stable as SEO crawler. The slightest security measure enforced by the website can break the whole process.

The various tools available are therefore limited in their functionalities. Nor do they make it possible to manage the operational and legal constraints mentioned above in a satisfactory manner. So we decided to design and implement our own tools and infrastructure. The latter must be at least as fast and efficient as existing website copiers and be at least as stable as the available SEOs while being able to bypass most website security mechanisms against copying and extracting data. To achieve that, we have developed a library to gather data in an efficient, resilient, generic and stable manner, both on the web and the dark web. To be compliant with existing laws, we have also developed an architecture which able to sort Medias, extract interesting data before encrypting them.

This paper is organized of follows. Section 2 summarizes the few existing works and solutions. Section 3 presents the secure architecture to process URL and data that may contain illegal data. Then Section 4 presents our scraping library and how it responds to each need. Section 5 exposes several anti-bot technologies and how we have implemented our bot to bypass them. Finally we conclude and present future works related to our tool.

## **2. Related work**

There is no shortage of Internet (TCP/IP) crawling tools. A quick search immediately allows us to find several hundred different libraries allowing you to custom implement crawling techniques in your own program or complete ready-to-use tools. If we look for similar tools for the TOR network (onion routing protocol), they are much less numerous. Even if the use of a socket is the easiest solution to adapt an existing Internet crawler to a ``dark-web'' one, the majority of close source software are unable to do that. When considering TOR very specific features, only a very few software can crawl it.

A basic solution for some crawler is to use an automated browser like *Selenium* (*Selenium*, 2013) is a heavy solution for a simple crawling tool because it includes a lot of useless features like the display of the screen or a JavaScript engine (not necessarily needed, especially on the TOR network). Another software is the *Ache Crawler* (VIDA-NUY, 2015) which adopts a special approach to avoid crawler traps. By defining a relatively low number of pages, whenever the crawler falls in a spider trap, it stops. The main drawback lies on the fact that the user can stop the crawler way too soon and hence can skip a lot of webpages without being aware of that. Next, a smaller crawler with different purposes, *Fresh Onion* (*Dirtyfilthy*, 2016) behaves as an indexing crawler just to create a search engine of the TOR network. Another light project, *TorScouter* (*Flora*, 2014) gathers onion webpages in order to estimate the size of the ``Darknet'' (and do not collect medias).

*PunkSpider* (*Brewster*, 2015) is most similar to our own project and work. Unfortunately, this project is now closed, the source code is no longer available and only few information about the ``dark-web'' scan are released. According to some article (the most recent one is (*Owenson et al.*, 2018)), it take roughly 3 hours to scan the entire Tor network and the result is around 7,000 domain names gathered. The *PunkSpider* project was at least a vulnerability checker and we still do not know if a local copy was created or not. This project was running on several Amazon servers. This may explain these rather good results in terms of time (not in terms of the number of webpages found).

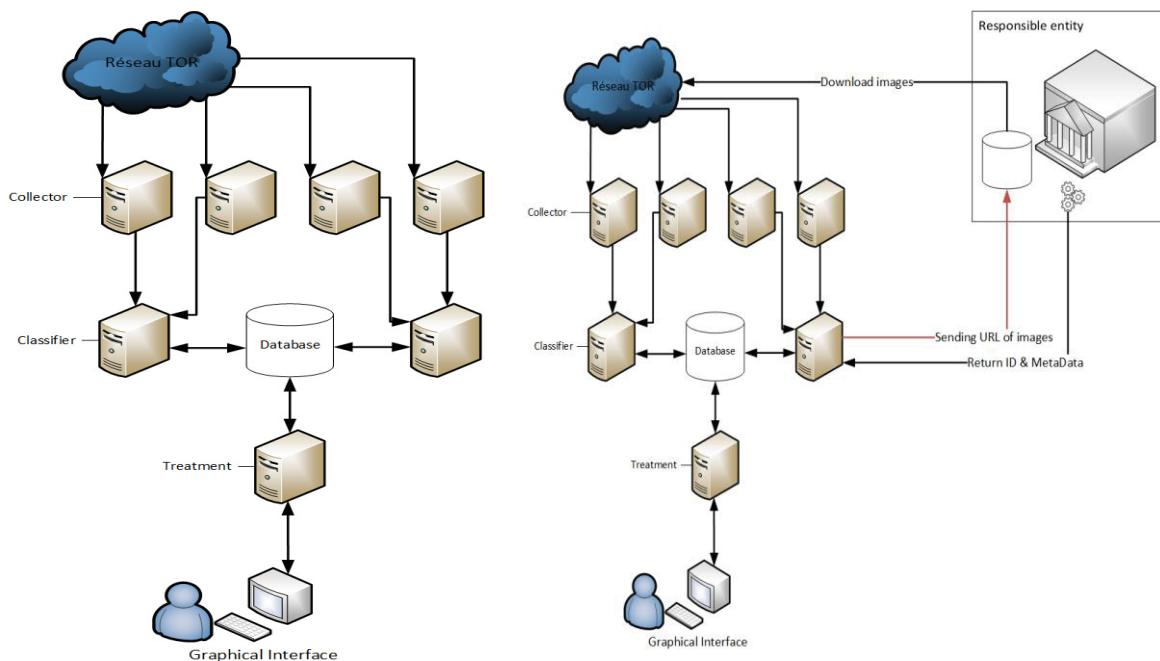
Aside the few available tools, only a very works have been published regarding the content analysis and surveillance of the TOR network. We can mention (*Owenson et al.* 2018; *Lawrence et al.*, 2017; *Biryukov et al.*, 2013; *Avarikioti et al.* 2018 ; *Sanatinia & Noubir*, 2016). However none of them addresses the legal issues regarding the data collecting and they do not propose a legal compliant infrastructure and approach similar to ours.

## **3. A secure architecture**

The majority of software available on the market essentially all performs two tasks: downloading webpages and then process them. Part of the existing crawlers is running on several threads thus allowing simultaneous downloads at the same time and simultaneous content processing. However this approach is still too slow because threads are involved in both tasks. Hence, we have decided to create a highly structured and distributed architecture to assign specific jobs to each server. By fragmenting the architecture and limiting the role of each type of server, the power of each of these servers can be used to maximize their respective efficiency and performances. They are not busy with respect to several roles at once (downloading which requires bandwidth, processing which requires CPU or GPU resources and storage which requires much hard disk space and I/O operations). The proposed architecture is depicted in Figure 1 (left part)

It is worth mentioning that the number of servers in Figure 1 (left part) is not representative. We cannot predict the number of servers required to download and process a full website. The number of servers depends on the target website, the number of its pages, the size of each page, and the number of media... For this architecture, the aim is to find the best trade-off and balance between the downloading servers and the treatment servers (always have pages to process or to download in a continuous but oversized flow). By defining a precise role to each server, we can then use specific servers for each of the main tasks. First let us describe their respective role:

- **Collectors.** The unique function of collectors is downloading webpages. Due to the very specific architecture of the TOR network, it takes an average of two seconds to access a page (a hidden service) on this network. Thus, a monolithic system would not be effective enough. This would be a real waste of time (For example, Internet counts about 1.7 billion web pages according to recent studies (Live Science, 2016) and the processing time would be more than 200 days without a distributed architecture).
- **Classifiers.** Classifiers receive downloaded HTML pages. They extract the information we need for our future study (text for stylometry purposes at least). More information can be added in this process. In a case of a forum, we can gather pseudonyms, messages, date of publication and a lot of other information (including metadata), whereas a market site will be full of products, prices and so on. The data to collect are to be defined in advance to spare resources in the processing server. The better organized is the database, faster and easier is the treatment. All classifiers are linked to a main database whose purpose is to aggregate and store data.
- **Processing.** The processing servers are in charge of finding information and meaning within the raw data. For instance, we can imagine that the processing servers can compute and return the 10 most active people in a forum, a summary of the activity of a nickname/username on the TOR network, a list of sites in a specific category (sale of weapons, drugs or others).



**Figure 1:** The proposed architecture (initial version of the left) – Legally-compliant secure architecture (right)

The key advantage of this architecture is its modularity. If we need to add a collector or a classifier, we just need to add a server, which is set up accordingly and launch the relevant tasks. We can then harvest information on all websites, whatever may be their size (Facebook, Twitter for the largest ones or Adobe's product page for example). As we said earlier, for the large websites, we will need a large "power" according to the type of server we consider. We think that this idea was the most suitable because we can rent some Amazon EC2 servers with specific capabilities to treat the largest websites. A few EC2 servers are designed for storage with huge capabilities (with a maximum of 42TB), a few for calculations on CPU or GPU while others have a high capability in terms of bandwidth. But, when considering such architecture, collected data are stored on an Amazon EC2 server. As a consequence, we may be sued for the possession of confidential files or child pictures or any other

illegal content (keep in mind that moreover the legality is geography-dependent contrary to servers). That is why we decided to make a modification on this generic initial architecture by adding a special entity for storage.

### **3.1 Legally responsible entity**

The responsible entity (Department of Justice, Department of Defense or any law enforcement force) is any authority that can legally store sensitive or illegal content. Our initial architecture has hence modified in order to take this new actor into account (Figure 1, right part).

Classifiers are able to detect whether the URL is corresponding to a media (photo, document, video...) or a to a HTML page. In the first case, the corresponding link is sent to the responsible entity with an ID to record in database which files are currently stored. This architecture has some pros and cons:

- **Pro:** The different collectors do not download and store media any more in the case where the possession of media is illegal, only the responsible entity collects it.
- **Con:** *The downloading step does not necessarily occur right after the processing of the web page. A delay can be present between the sending process and the downloading of the media. The web is by nature very volatile and data can be lost. For example, website like Mega which keep file during 30 days and if the download by the responsible entity is delayed, we can lose some files.*
- *The responsible entity must have a permanent connection to the TOR network.*
- *The bot may lose communications between the Classifier and the entity.*

To sum up, the main problem of this architecture is the existence of a two-step between the classifiers and the responsible entity. However this can be easily solved since it is just a backup problem basically. First, if the Classifier cannot send data to the entity (due to any network error), we can keep all the URLs to download in memory and send them after fixing the communication. For the second communication step, if the responsible entity keeps a backup of metadata, this communication is a lesser problem. Metadata can be retrieved later and it does not hinder the smooth running of download operations. Even with these slight drawbacks, the proposed architecture is the best solution/trade-off. The key feature

The two communications are ciphered (a simple Public Key Infrastructure [PKI] is used for the key management). The first one is the URL to download along with an ID and the second one is the media's metadata. This level of security aims at "hiding" the bot's activity (that could for instance warns any criminal that he is under surveillance) and protect the legal entity itself (it must be impossible to establish any link between the bot and the entity).

### **3.2 Media processing and storage**

#### **3.2.1 Gather information/metadata**

Each media retrieved by the bot may be potentially illegal. In our case, we do not have (and we do not want) access to the data itself. However we need some information for further treatment or for legal purposes. To gather metadata from any type and format of media, we use the ExifTool library from Phil Harvey. This library can process 184 file types or almost all file types we may find on the web.

Regarding the data itself, our purpose is to gather as much information as possible without consulting the media itself. We first chose to collect precise data from two file types (each file type having a specific module (library) within the bot): PDF files and Office documents (Word, Excel, Powerpoint, all versions). From within these two file formats, we also extract images and text included, whenever possible. The text is returned in a raw format but the images are processed as all others (metadata extraction and secure storage) and only metadata are returned.

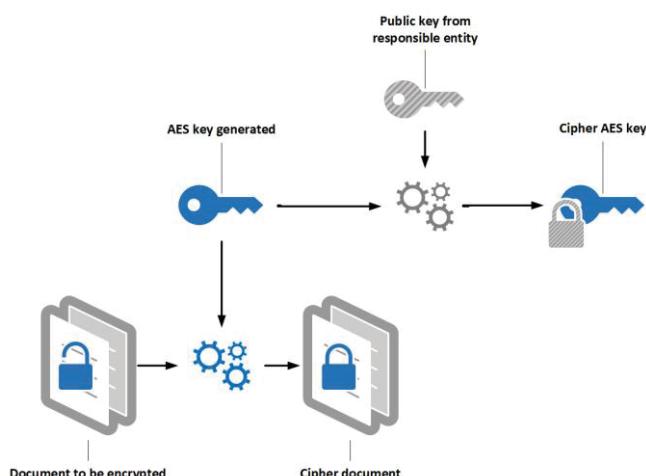
In a further step, we intend to implement a deep learning algorithm to guess the content of an image as Facebook usually does. We do not claim to do as well as Facebook does but having further information can enhance the final evaluation and rating of the collected content. The confidence level of this further information will depend of the result of the deep learning algorithm.

Due to legal obligations, we need to be able to input data and metadata to a file (tagging process and data qualification). To perform this fingerprinting process, hash functions are used (MD5, SHA-1 and SHA-256). Here is an example of the result of a full treatment process of a Word document:

```
[numbers=none, basicstyle=\tiny'digest_md5'=> '9900f366615b2f4619eb9b806f1ee9b7', 'digest_sha1'=>
'55eb0cbe8a05a[...]baede211',
'digest_sha256'=> '14f9726703a64af[...]a3d015a98318', 'medias'=> {'image.jpg'=> {'thumbnail'=>
{'exiftool'=> { [...] }}, 'digest_md5'=> 'e0d15ecd797c455b02e30f585ce3f609', 'digest_sha1'=>
'd6287f7eee6b6041[...]9f4408bb29e', 'digest_sha256'=> 'fa94a4f7d03fcdf[...]1d561661227'}, 'digest_md5'=>
'560191c489a9296d1ccbb5d4cc8404d1', 'digest_sha1'=> 'bfec261f14149c[...]50a70d49c8bb286',
'digest_sha256'=> '1587622868123[...]278d25b0af884f5', 'exiftool'=> { [...] }}, 'word_list'=> {'This'
=> 1, 'is'=> 1, 'a'=> 1, 'test'=> 1}
```

### 3.2.2 Secure storage (optional)

In this part, we explain how to process data and store them in a secure way. Any responsible entity can of course choose to have their own process. The architecture being fully modular, our own module is just a possible instance of secure storage. On the responsible entity's side, the process timeline is as follows. Without loss of generalities, we focus on TOR address (`\texttt{*.onion}`). The first step is the reception of an URL onion address to download. Once downloading is completed, all documents and contents are analysed and processed (see Section 3.2.1). The last step consists in encrypting each file.



**Figure 2:** Encryption system

For performance purposes, we decided to design an encryption procedure with two layers of encryption keys (Figure 2). The first layer is a set of AES keys randomly generated for each file (message key). Storing these message keys in an unencrypted file is not conceivable because anyone who would access to this file would have access to all the media. In order to limit the access to the keys, and by extension, to the files, we have chosen to encrypt all these keys with the public key of the entity. Hence this procedure ensure accountability for anyone accessing any content (thanks to the relevant cryptographic certificates managed by our PKI)

## 4. Bot development

We did not find a bot solution which responds to all our needs. Either the solutions are too slow – such as Selenium, Mechanize (<https://wwwsearch.sourceforge.net/mechanize>) – or not free, or not modular or not resilient (they stop when the least security mechanism is in place on the website). We thus developed our bot to be faster than every existing software. We designed and implemented a multi-threading bot for the above-presented specific architecture. The bot needs to be reliable which means that it can work for a very long time (days or weeks) and in an as much as possible automatic way. It needs to be resilient to bypass websites' security mechanisms and must still work under strong constraints. Last, the bot needs to be generic, to work on different websites without any code modification and written in Perl for better performances (especially for text processing purposes).

#### **4.1 Multi-threading solution**

Nowadays, writing a crawling solution without multi-threading is unthinkable. In Perl, this method of development is not the most popular and only a very few libraries are available to implement a multi-threading program. While there exist very few libraries for a multi-threading program, there is even less libraries to create a thread-pool (set of threads). They are too complex to use and not modular at all. For this reasons, we have chosen to develop our own thread-pool library in order to be able to launch several jobs inside the same thread-pool.

As you can see in Figure 1 (right), there are two available tasks: one which downloads the web page from a given URL and another one which extracts links in an already downloaded webpage. Both take jobs from the same "queue" (represented as a database in the figure) and put jobs in the same queue. This thread-pool is different from others because there is only one kind of thread for two works. When a thread needs a job, it does not know what it has to do next. This concept of doing several jobs inside a same thread-pool is totally in opposition with the thread-pool original purposes. For the same results, we could have implemented two thread-pools instead of one but the communication between them would have been much more complicated to implement.

The best advantage of this design is the transparent load balancing. If we had developed our crawler with two different thread-pools, we should have to handle the number of pages to download and the number of pages waiting for download. This is a waste of time because these two numbers are far too variable. On a directory link list for example, there is almost no text to process but a lot of web pages to download. In this case, the processing thread-pool could have been inactive for a time. In our solution with no precise jobs planned, there is always work for each thread, the load balancing is automatic and transparent.

This conception of a thread-pool does not impact the crawling process but it is modular and we can imagine adding a lot of different jobs like ciphering directly the web page or make statistical analysis inside the same thread-pool. That is not necessarily the purpose of our crawling project to add a lot of different works but this library can and will be used for other projects.

#### **4.2 Banishment and dynamic proxy changeover**

A multi-threading crawler is more efficient than a mono-thread program, even too much efficient for some websites. With a hundred of threads, the requests are too numerous and the website may interpret this as being an attack (DDoS attack for instance). On the website defense's side, the easiest solution is to ban the IP address responsible of the alleged "attack".

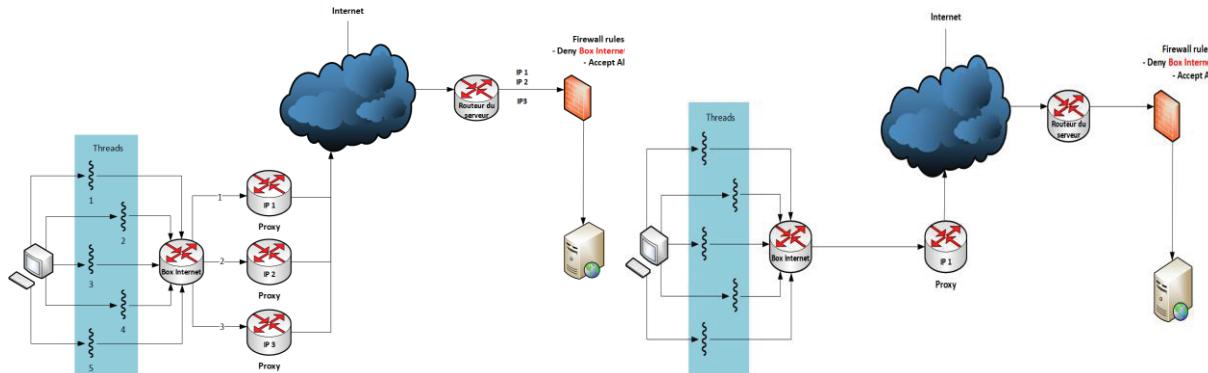
If the web server has been configured correctly or is protected by a dynamic firewall, the crawler can be blocked quite quickly. Intense downloading can be mistaken with a Denial Of Service (DoS) attack and, in this case, the firewall will react and block all future requests from our IP address. Different responses can be received by the crawler as a timeout or a forbidden access (HTTP code 5xx or 403). None of the software on the market has countermeasures against a dynamic banishment and they stop working because they do not have more URLs to collect anymore in HTML pages or pages to download. There are several solutions to avoid being banned from the web server. Here is a non-exhaustive list of what we can do:

- Add random timer between requests.
- Reduce the number of threads.
- Limit the bandwidth used.

All of these solutions have generally a dramatic impact on the program performances. So they were not an option for us. The last solution, proposed in several software, consists adding a proxy between our program and the target website. This solution allows the banned user to continue or restart the process via a proxy. The problem of using a proxy which would be configured at the crawling process initialization lies in the fact that the user needs to check constantly whether the crawler has stopped or not. Consequently he would need to be reloaded through a proxy. In order to create resilient and reliable software, we have developed a solution to avoid banishment automatically.

#### 4.2.1 Two possibilities

As we worked in a multi-threading environment, two way to use a proxy is possible. The first one is given in Figure 3 (left)). It consists of numerous proxies, one per thread.



**Figure 3:** One proxy for each thread (left) - One proxy for all threads (right)

The first possibility has however too many issues. Monitoring independent threads on different proxies is difficult and complex. If the implementation is easier (one connection, one proxy to set), it does not correspond to the thread pool concept exactly. We have already given several job possibilities for a single thread so adding different connections may have been a ``dirty'' solution. Furthermore, this configuration removes a lot of working proxies because if one of them is not working anymore, no thread can use this connection anymore. If we use this solution, we may have a problem of sleeping threads because they are banned and do not have any proxy (not banned) to keep working. Having sleeping threads is not an optimal solution.

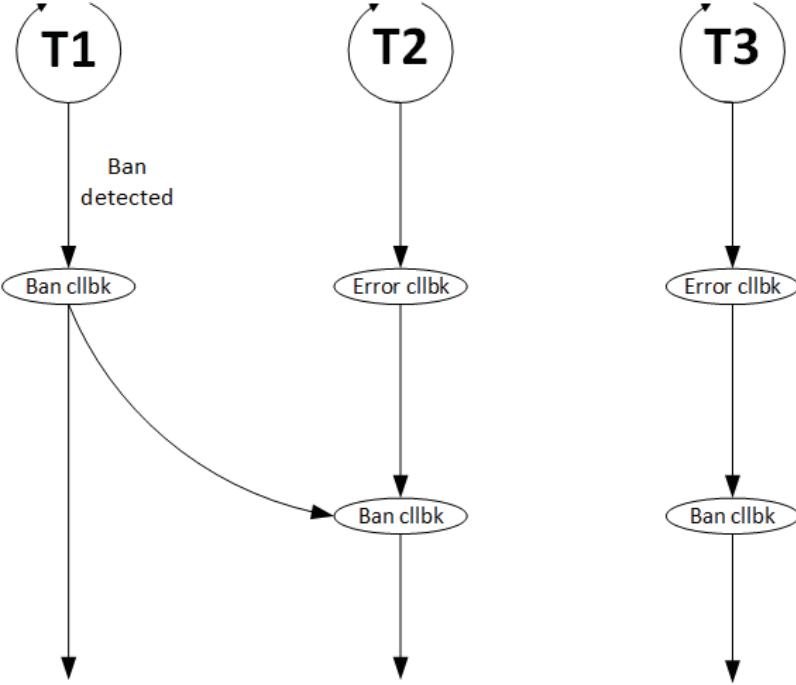
The second solution (Figure 3, right) is better and more organized. All threads will use the same proxy once they are banned. This conception is more homogeneous and is easier to monitor. If all proxies are unusable and the crawler stops, all the threads stop at the same time. The crawler has a stop procedure embedded with a backup of current and future work in case where it stops unexpectedly (often when the crawling is banned from everywhere). In this architecture, the stop procedure is easier to implement but there is still an issue whenever we wish to change the proxy of each thread.

#### 4.2.2 Dynamic proxy changeover

The major issue when a multi-threading bot is banished lies in the fact that each thread has an active connection with the web server. We can set up a proxy on an active connection for one thread but not for each other ones at the same time. To understand the process that we have implemented to set up a proxy on each thread in a row and why it is efficient, we need to explain first how we detect banishment.

As we said before, banishment is characterized by an error code returned by the web server. Unfortunately, a few errors are legitimate (for instance for a forbidden webpage or when a server is down for some reason). To determine whether the error is a legitimate one or not, we have chosen to pick randomly a witness page. This witness page is a page which has been already downloaded, processed and referenced as perfectly working (status code 200). This page is downloaded again and in case of an error code (the same one than the previous page), then the bot is considering itself as banned.

However this is a rather long process for a thread to determine whether it is banned or not. We cannot waste time by letting each thread make its own verification. Moreover we can rapidly come back to the first solution. That is why we have developed a solution where only one thread does the check up before setting up a proxy for all threads.



**Figure 4:** Banned callback creation

Our procedure consists in creating a callback on the fly. In Figure 4, T1, T2 and T3 represents three threads. In this example, the first one detects the banishment and creates a *banishment callback*. Then, this callback is put right after the classic error *callback* that any other thread will detect (because they are banned too, so they will receive an error code). This approach has several advantages. The first one is the detection of the banishment. In our case, only the first thread will do the check. This saves time and the bot only has to request a witness page once instead of \$n\$ times (\$n\$ being the number of threads). The second advantage is that no thread is left aside. As soon as a thread changes the proxy setting, all the others will directly set up the proxy accordingly.

In terms of code, we have implemented this solution with a joint list of proxies for all threads. As we said previously, only the first thread which detects the banishment will change the ``head proxy''. This ``head proxy'' is the current proxy in use (special case at the bot start because no proxy is set). Then, each thread which falls in the callback error will go in the banishment callback and set the ``head proxy'' if it is different from the one they used before. The change is fast and without any long verification for all threads except for the first one.

Let now us detail the process more precisely. Whenever an error code is returned, there are three different cases to deal with. First, the page gets another error code than for the witness page. In this case, we exit from the function because it is not a proxy issue. The second case, where the witness functions works correctly (status code 200 or 304), the original request is replayed and the result is the final one. The last case is when we get the same error code on both requests (original and witness).

First, we check if we have remaining proxies in the list. If not, the bot stops properly, meaning that everything (the work in progress and to do) is backed up for a later restart of the bot at the exact same point with a fresh new list of proxies. Second, in the code, we have set up a mutex to prevent several threads to do the process at the same time and change the ``head proxy'' which would otherwise result in a waste of working proxies and it would not respect our design. Once this mutex is locked, the thread continues the procedure selecting the first proxy in the list if no proxy was set or the second one if a proxy has already been set. Then the bot tries the request again. If a legitimate error page appears, then we put back the old proxy because it was not a proxy issue. The mutex is not released until the end of the previous step because we can still modify the proxy list. If all this process is completed correctly, the only visible change is in the proxy list.

A reduced version has been implemented to avoid this long process of checking if it is a banishment issue. In case of any error code returned, a verification of the ``head proxy'' is performed and if it is not the same that the one currently used, the shift is done immediately without verification of a possible banishment.

## 5. Defeating crawler traps

On the web server side, there exist a lot of methods and mechanisms to detect bots and to limit their impact on a website. A few methods are simple and work due to the basic technical level of bot writers but they do not work against really malicious bot. Ours does no malicious actions but it acts like one and we have to face security mechanisms against bots. The most efficient securities are *Crawler Traps*. They consist in fake webpages to waste the robot time and resources. If the web site detects a bot running on it, it creates these fake web pages to trap the crawler into a loop or to provide false information to it. The purpose is to protect the rest of the website from the bot and we can compare it to a *honeypot*. During our study, we have chosen to classify crawler traps into two categories: those with arguments in URL and those with dynamic paths.

### 5.1 Crawler traps with arguments

This technique has a lot of advantages: it is nearly invisible for users and it does not block legitimate bots. There are several methods to create a spider trap. Let us list a few of them:

- **Calendar.** By putting a calendar on a website. Sometimes when JavaScript is not used, there is a link to the previous and the next day for each day. If the calendar does not have an end, this create an unintentional crawler trap.
- Example: <http://www.example.com/calendar/monthly.html?year=2018&month=1&country=5>
- **Market.** Some markets have a crawler trap and redirect to fake pages. These fake pages are created with parameters. For example, the argument ``Sort By'', ``Products per page'', ``Price Low'' and ``Price High'' can be in the URL but are irrelevant. This can be a crawler trap (intended or not).
- Ex. : [https://example.com/items?minPrice=1500000&maxPrice=2400512000&SortType=price\\_asc](https://example.com/items?minPrice=1500000&maxPrice=2400512000&SortType=price_asc)

We have developed the bot with three possible configurations to avoid crawler trap with arguments. The first configuration (called ``*Blind mode*''), is a foolish crawl. It will treat all URL links with parameters without limitations. This mode can easily fall into crawler traps but can be used on a market website for example and where no crawler trap is deployed. The only advantage is that we are sure to download the entire website without missing anything. The second solution is the total opposite (we called it ``*Safe mode*'') and will get rid of all links with arguments. In this case, the bot will finish without problem but can miss a huge part of a website.

The third configuration is maybe the most precise of the three. This ``*Smart mode*'' will treat a maximum of links and avoid falling in crawler trap by sorting each link. In this mode, the URLs will be accepted only if there is two arguments or less and if this argument is a number (under 1,000) or a path (string starting with ``/'' or ``\'''), no strings allowed (except for path). Everything else will be dumped. This method allows the bot to process forums and repositories with paths in argument but is not falling into a crawler trap. A number of 1,000 possibilities for an argument on the same page is reasonable and this number can be adjusted in the program configuration.

### 5.2 Crawler traps with path

This kind of crawler trap is highly intentional and need to be set up on the website. The purpose is to create an infinite path on the website. If there is theoretically a limit size (according to the RFC 1738 \cite{rfc1738}) for URLs, it can be removed in the configuration of the web server to allow this kind of crawler trap. If the URL is longer than 2048 characters, most of web browsers will show an error message but as the bot is not using a web browser, this trap can work indefinitely.

Example: <http://example.com/bar/foo/bar/foo/bar/foo/bar/.....>

Our bot was developed to be fully adaptive for each need. In that way, we have developed two approaches to avoid crawler traps. The first one is the simplest one. If in a URL, two pieces of the path are the same, the bot will compare the content of the two pages (between ``body'' tags only or the full page if the bot does not find a ``body'' tag) and if the content is identical, both pages are a crawler trap.

The second method is smarter and can be customized by the bot user. By default, starting from the 5th and for all further element of the path, an automatic detection will be activated. To detect a crawler trap we have

developed a mix of several methods used in information theory. We have chosen to compute the entropy of the page, then register all the words and their frequency of occurrence and compare the body tag architecture and the content of some body tag. With this information, the bot is able to detect a crawler trap by comparing the respective information measures. To decide if the page is a crawler trap or not, the threshold is 90\% of similarities. This threshold can be modified by the user according to the website processed and the presence of a possible crawler trap. When processing a website, the bot can be trapped and the threshold can be not restrictive enough and in that case, the user can set up a higher percentage for a particular website. According to the crawler trap that we have faced on the field, 90\% or 95\% are acceptable thresholds.

### **5.3 Other techniques**

#### *5.3.1 Crawler trap with parameters and arguments*

During our work, we have analyzed a website with a path in argument (<https://example.com/directory?name=\foo>). This can be a new way to create a crawler trap. Even if we have noticed this kind of website only once, we have handle this case by adding the text analysis to our "Smart mode".

#### *5.3.2 Robot.txt file*

This file is present on each website and manages bots' activity. Legitimate bots like *GoogleBot* respect the file which indicates the page it is authorized to crawl or not. We do not care about this file in order to be able to download all the web site. Nevertheless, we take into account the file "*sitemap.xml*" (or equivalent) if it is present to know all the pages and we gather the list of address that is forbidden to process. Crawler trap are indicated in this list but not all of the forbidden pages are crawler trap. This gives us another indication which is added to the crawler trap detection process.

#### *5.3.3 Laxity in HTML development*

Nowadays, web pages can be developed without norms. If the "*doctype*" is supposed to be present, it is not the case anymore. Even some closing tags can be omitted and the web page is still correctly displayed in most browsers. This excludes the use of regex techniques. Our bot is not able to reconstruct web pages (the Gecko motor from the Mozilla foundation is too resource-consuming to add it to our bot) but it can detect whether the page is a web page, an image, a pdf file or any other file.

## **6. Conclusion and future work**

We did not only develop a new crawling library, but we have also developed a new crawling method with special specifications. The automatic information gathering had new stakes since the web is more and more present in our life. Since Internet and the TOR network are the reflect of our society, we find the same issues on it as in real life (drugs deals, weapon trafficking, child pornography...). Our project is a solution for police enforcement or company which wants to collect information in a legal and efficient way.

We have developed an architecture able to handle a heavy workload while ensuring compliance with the law. By ciphering directly the media and preventing the possibility of undue consultation, we are respectful of the French law and many similar regulations.

The bot running in the architecture is also a new solution. If a few products are available, none of them are faster than our, none of them can work on the TOR network and none of them are able to bypass a banishment by IP address. This complete solution can be used to download a website or a whole network as the TOR network without restriction. To conclude this paper, let us recall that this project is only a tool for data gathering. The processing of data was not the purpose of this article and needs to be done on specific purpose. For the moment, the bot answers our needs by being able to process websites on the TOR network (hidden services) and on the Internet. We have one of the most powerful and adaptive architecture to crawl efficiently. The future works will be focused on the crawler library. The difficulty is to developed upgrades which will not break another part of the crawler.

## References

- G. Avarikioti, R. Brunner, A. Kiayias, R. Wattenhofer and D. Zindros (2018), "Structure and Content of the Visible Darknet", Arxiv preprint 1811.01348, [Online] <http://arxiv.org/abs/1811.01348>
- A. Biryukov, I. Pustogarov and R.-P. Weinmann (2013), "Content and popularity analysis of Tor hidden services", Arxiv preprint 1308.6768, [Online] <http://arxiv.org/abs/1308.6768>
- T. Brewster (2015), "Meet The Darpa-Backed Hackers Building A Google For Every Web Weakness", Forbes, [Online] <https://www.forbes.com/sites/thomasbrewster/2015/05/06/punkspider-google-for-all-web-vulnerabilities>
- Dirtyfilthy (2016), "Fresh Onion Website", [Online] <https://github.com/dirtyfilthy/freshonions-torscraper>
- M. Flora (2014), "Tor Scouter Website", [Online] <http://mgpf.it/torscout>
- French Penal Code (2013), "Article 227-23", [Online] <https://www.legifrance.gouv.fr/affichCodeArticle.do?cidTexte=LEGITEXT000006070719&idArticle=LEGIARTI000006418095>
- H. Lawrence, A. Hughes, R. Tonic, C. Zou and Y. Jin (2017), "D-miner: A Framework for Mining, Searching, Visualizing, and Alerting on Darknet Events", *Proceedings of IEEE Conference on Communications and Network Security* (CNS, IEEE Press).
- Live Science (2016), "How big is the Internet, really?", [Online] <https://www.livescience.com/54094-how-big-is-the-internet.html>
- G. Owenson, S. Cortes and A. Lewman (2018), "The darknet's smaller than we thought: The life cycle of Tor Hidden Services", *Digital Investigation*, vol. 27, pages 17-22, [Online] <http://www.sciencedirect.com/science/article/pii/S1742287617303894>
- Parsehub (2018), "Parsehub website", [Online] <https://parsehub.com>
- X. Roche (1998). "HTTrack Website", [Online] <https://httracks.com>
- A. Sanatinia and G. Noubir (2016), "Honey Onions: a Framework for Characterizing and Identifying Misbehaving Tor HSDirs", Arxiv preprint 1610.06140, [Online] <http://arxiv.org/abs/1610.06140>
- Scrapy (2019), "Scrapy 1.6", [Online] <https://scrapy.org>
- Selenium (2013) "Selenium crawler", [Online] <https://github.com/corywalker/selenium-crawler>
- United States District Court, N.D. California (2017), "hiQ Labs, Inc., Plaintiff, v. LinkedIn Corporation, Defendant.", Case No. 17-CV-03301-EMC. [Online] <https://www.leagle.com/decision/infco20170628b79>
- ViDA-NYU (2015) "Ache Website", [Online] <https://ache.readthedocs.io>

# Cybersecurity Assessment of the Public Sector in Greece

George Drivas<sup>1, 2</sup>, Leandros Maglaras<sup>1, 3</sup>, Helge Janicke<sup>3</sup> and Sotiris Ioannidis<sup>4</sup>

<sup>1</sup>National Cyber Security Authority of Greece, General Secretariat for Digital Policy, Ministry of Digital Policy, Telecommunications and Media, Kallithea, Greece

<sup>2</sup>University of Piraeus, Department of Digital Systems, Piraeus, Greece

<sup>3</sup>De Montfort University, School of Computer Science and Informatics, Leicester, UK

<sup>4</sup>Foundation for Research and Technology, Greece

[g.drivas@gmdp.gr](mailto:g.drivas@gmdp.gr)

[leandros.maglaras@dmu.ac.uk](mailto:leandros.maglaras@dmu.ac.uk)

[heljanic@dmu.ac.uk](mailto:heljanic@dmu.ac.uk)

[sotiris@ics.forth.gr](mailto:sotiris@ics.forth.gr)

**Abstract:** Organizations have to manage new risks, sometimes proactively, sometimes by being constrained by regulations such as GDPR or the NIS directive. To cope with new threats, it is essential to develop or reinforce a real culture of cybersecurity at the organizational level. Before putting anything in place, we must start by assessing the new risks to which we are exposed. The new regulations that the EU is issuing, invite organizations and member states to follow these approaches. National Cyber Security Authority of Greece (NCSA) is responsible for coordinating the public sector and the National Critical Infrastructures (NCIs) of Greece, in order to take all necessary steps towards a secure Greek Cyberspace. Its main objective is to shield the Nation from external threats and to provide a secure digital environment for all citizens of Greece. One important action is the enhancement of digital skills and the development of a strong public and private security culture, exploiting the potential of the academic community and public and private sector actors. NCSA is following a PDCA-cycle approach with strong cooperation of all relevant stakeholders for securing NCIs. NCSA is planning a series of audits for the entire public sector and for NCIs. The assessment of the central governmental ICT structures was selected as an initial phase. For this purpose, NCSA sent structured questionnaires aiming in capturing the general picture of the security situation of central ICT infrastructures. Data collected during this phase are processed and will be used to design the next steps of deepening and expanding of such assessments but also to institute regular and / or emergency control procedures on a permanent basis. The information that has been gathered is analyzed in order to reveal major threats, capacity building priorities, current situation in terms of procedures, security measures and policies and established incident response plans.

**Keywords:** cyber security, public sector, national critical infrastructures

---

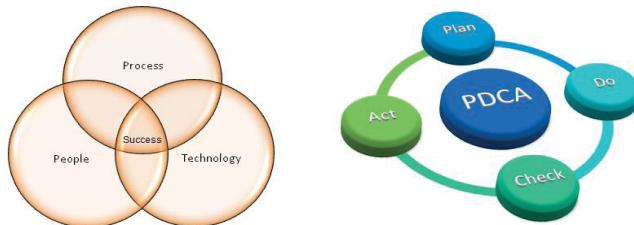
## 1. Introduction

Recently several critical incidents that targeted National Critical Infrastructures (NCIs) have taken place (Maglaras et al, 2019). In September 2018, shortly after Cyber Europe Exercise tested European reaction and cooperation following a cyber attack targeting the aviation sector (Seker and Ozbenli, 2018), information screens in University of Bristol were taken offline to contain an attack similar to so-called "ransomware". Some years ago, in 2015, Ukraine was hit by a massive blackout due to an attack to their SCADA systems, leaving 230K citizens of Ukraine without electricity for several hours. Another attack that took place in 2013, although reported in 2016, targeted a small dam in Rye Brook in New York (Bianco, 2016). The real target of this attack, based on a report of FBI and Homeland security, was Wolf Creek Nuclear Operating Corporation, the impact of which, if successful, would go beyond a single nation. Recently, UK's National Cyber Security Centers (NCSC) is concerned about suspicious attacks that are taking place on UK energy sectors (Kovanen, Nuojua and Lehto, 2018). All of the above are only some of the attacks that are happening every day around the globe and are targeting NCIs, such as oil and gas industry, traffic signal, water sewage building, transportation, and digital infrastructure. It has been shown that a cyberterrorist attack that directly targets a population can have the same effects to those that directly target NCIs (Ayres and Maglaras, 2016).

Following the publication of high-profile security breaches and security incidents, organizations and nations around the globe are increasing their focus and are looking on ways to improve their cyber security assurance (Andreasson, 2011). This will help them protect both their brand and reputation along with the prevention and reduction of financial impacts. Except from technology-related breaches which are due to malicious actors that exploit existing vulnerabilities in technology and that will continue to take place in a regular basis, a big percentage of data breaches or security incidents that are reported are caused by inadvertent human error. Despite the huge surge in interest and acceptance of information security management, incorporating cybersecurity, there still appears to be gaps and weaknesses within organizations. As Critical National

Infrastructures are becoming more vulnerable to cyber attacks, their protection becomes a significant issue for Member States as well. The synergy between the ICS and the IoT has emerged bringing new security challenges. Modern smart societies face new challenges in the area of cyber security, and EU is trying to strengthen Critical Infrastructures by publishing new directives and regulations.

Along with the obligations that directly arise out of the European directives and regulations, Greece and all other member states must take further actions for enhancing cyber security. National Cyber Security Authority of Greece (NCSA) is responsible for coordinating the public sector and the operators of essential services of Greece, in order to take all necessary steps towards a secure Greek Cyberspace. Its main objective is to shield the Nation from external threats and to provide a secure digital environment for all citizens of Greece. One important action is the enhancement of digital skills and the development of a strong public and private security culture, exploiting the potential of the academic community and public and private sector actors. Continuous adaptation of the national institutional framework to the new technological requirements, always in line with the European regulations on data protection and security will help Greece fight cyber-crime. NCSA has issued in 2018 both the National Cyber Security Strategy and the National Law on security of network and information systems (Maglaras et al, 2018). NCSA is planning to follow a PDCA-cycle approach with strong cooperation of all relevant stakeholders for securing NCIs (See Figure 1). A blend of processes, technologies and people is needed to achieve this goal and NCSA must have a general overview of the current situation in terms of hardware, software and security procedures that public sector and NCIs are using. In order to achieve this, a creation of an IT inventory along with a security inventory of all NCIs that reside inside Greece, along with all critical operational centers of the public sector and governmental clouds (Cook et al, 2018) is an essential first step. For that reason a questionnaire was sent to relevant stakeholders aiming in capturing the general picture of the level of security of central ICT infrastructures.



**Figure 1:** Cyber security framework: Success ingredients (left figure) and lifecycle (right figure)

According to the NIS national law, operators of essential services (OES) as well as for the digital service providers (DSP) must introduce appropriate security measures in an effort to achieve a baseline, common level of information security primarily within Greece and in alignment with the European Union (EU) network and information systems. Audits are major enablers to achieve this objective. A security audit is an independent review and examination of system records, activities and related documents using structural procedures and is based on risk exposures (Wood et al, 2017), critical components and business operations of the organization (Tipton and Nozaki, 2007). Having this in mind, along with the necessity to capture the current situation in terms of security as described earlier, NCSA has issued this questionnaire as a pre-audit mechanism.

## 2. Methodology

The questionnaire that was sent to relevant stakeholders was constituted of four main parts of 22 questions in total. Using the initial assessment questionnaire NCSA tried to assess the organizations queried regarding their current level of security, the existence or not of policies, procedures and technical measures, user awareness techniques that they are using and what incident response plans or procedures they have in place. That way we tried to cover all the different aspects that help an organization succeed in the fight against cyber-attacks, procedures, policies, technology and people. The four categories were:

- 1. Current level of security
- 2. Security policies, procedures and technical measures
- 3. User Awareness
- 4. Incident response

The first part consists of six questions regarding the current level of security of the organization. Participants were asked to grade the overall level of security of their organization and answer questions regarding the most

significant threat that according to their opinion exists for their systems. Following questions primarily looked at capacity building needs, cyber attack consequences and cons and pros of enhanced security measures. The second part of the questionnaire included specific questions that tried to capture the current situation of the organization in terms of security policies, procedures and technical measures. In that sense questions about data encryption methods, security mechanisms and self auditing procedures in place were issued to the questioners. A good security awareness program should educate employees about corporate policies and procedures for working with information technology. Employees should receive information about who to contact if they discover a security threat and be taught that data as a valuable corporate asset. For that sense the third part of the questionnaire was focused on the awareness and training of the employees about security and privacy. The fourth part of the questionnaire covered intrusion detection and incident response and handling procedures that the organizations are following. The aim of this project was to assess the security level of central governmental ICS infrastructures of Greece. To meet this aim, the questionnaire was designed to meet six objectives, as follows:

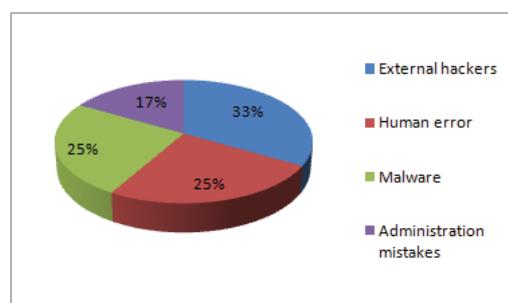
- Build a network of security officers
- To determine major threats to central infrastructures
- To analyze capacity building priorities
- To capture current situation in terms of procedures, security measures and policies
- To determine if there is an incident response plan in place
- To capture training and education policies and mechanisms

### **3. Analysis of results**

The data collected from the questionnaires were recorded and interpreted in accordance with the identified objectives of this research. The analysis of the data was designed to explore any similarities, differences or patterns among the responses and any underlying relationships. *More than 30 questioners filled the answers, in some occasions having the same person to fill the whole questionnaire for an organization while in other occasions two or three persons were needed to cover all the activities of the company being assessed. Most of the responders were Directors or Heads of the IT divisions which are in charge for the security management of their organizations. Although public organizations that were assessed covered a big range of different activities, including also critical infrastructures, the exact identity of each organization cannot be reviled since this information is sensitive in terms of national security.*

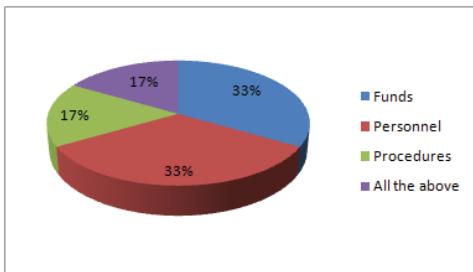
#### **3.1 Current level of security**

The first part of the questionnaire primary looked at what is the overall level of security of the organization according to the participants' opinion. The results revealed that 45% of the experts assess their systems as being relatively safe while 55% think that the level of safety of their systems is satisfactory. Participants were asked about their opinion regarding the most significant threat that their systems face. According to a recent research by Evans et al (2019), the majority of incidents within the public sector relate to human error. The research findings has identified that the actual proportion of reported public sector information security incidents that relate to human error is 92.5%. However, as Figure 2 shows, participants feel that external hackers (33%) is the major threat for their systems, followed by human error (25%) and malware infection (25%), while administration/configuration mistakes (17%) is the least significant threat.



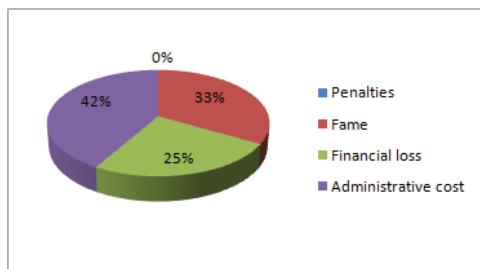
**Figure 2:** Major threat

Training of key stakeholders and key personnel and providing them with the capacities they need to uphold cybersecurity is important for a stable cyber capacity. In that sense we tried to identify the major need that public sector in Greece have in terms of capacity building. Figure 3 shows that organizations identify enhancement of personnel capabilities through training and education along with the increase in numbers of employees that work in specific information security departments as their primary concern with 33%. Increase in funds that are spent for hardware and software security solutions is also a major concern gathering the same number of answers, 33%.



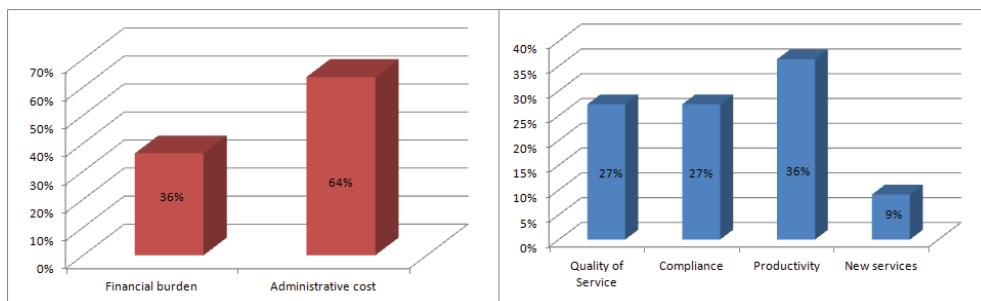
**Figure 3:** Capacity building

Another aspect that the questionnaire looked at was about major concerns of public organizations in Greece, following a data breach or a cyber attack in general. As shown in Figure 4, administrative cost due to following a disaster recovery plan or a mitigation plan for recovering the organization's normal operation is the number one concern with 42%, followed by financial loss and fame with 33% and 25% accordingly. Penalties do not appear to be an important concern for public sector organizations since GDPR and NIS weren't active yet when the questionnaire was issued.



**Figure 4:** Data breach/cyber attack consequences

Pros and cons of a tentative upgrade of the organization's security level were also questioned. As shown in Figure 5 financial burden (37%) and administrative costs (63%) are major negative consequences while on the same time participants expect their organization to be better organized and more productive (37%) after imposing security measures or standard procedures as parts of a security enhancement strategy.

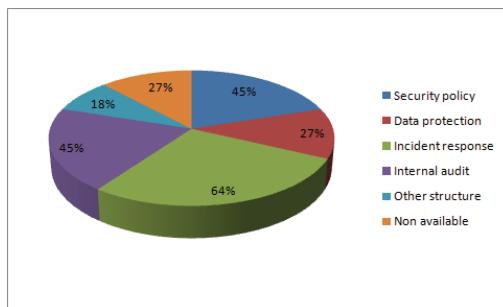


**Figure 5:** Pros and cons of security upgrade

### **3.2 Security policies, procedures and technical measures**

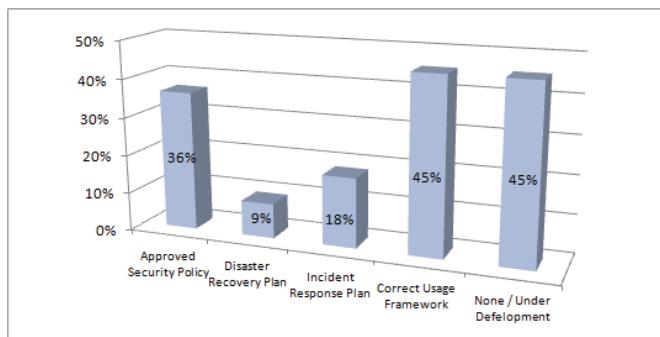
The second part of the questionnaire covered issues regarding any established structures that exist inside the organization, policies that are issued and / or approved, specific security measures that are in place, along with audit plans and procedures that may exist. As shown in Figure 6, 45% of the organizations have a specific department/directorate that is responsible for the implementation and evaluation of the security policy, while on the same time only 28% had a similar structure responsible for protection of personal data. The low

percentage that is observed in regards to data protection is due to the fact that at the time that the questionnaire was issued GDPR was still inactive in Greece.



**Figure 6:** Organizational units

The second part of the questionnaire focused in investigating the existence of security policies, recovery plans, incident handling procedures or a general framework that covers the correct usage of IT equipment and the dissemination of those to all employees. According to the findings, which are presented in Figure 7, almost half of the public organizations (45%) don't have any of the aforementioned documents in place which is an important finding about basic security measures that are missing and should be prioritized in the near future.

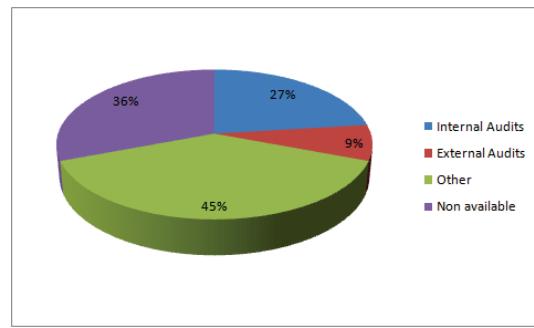


**Figure 7:** Security policy / recovery plan

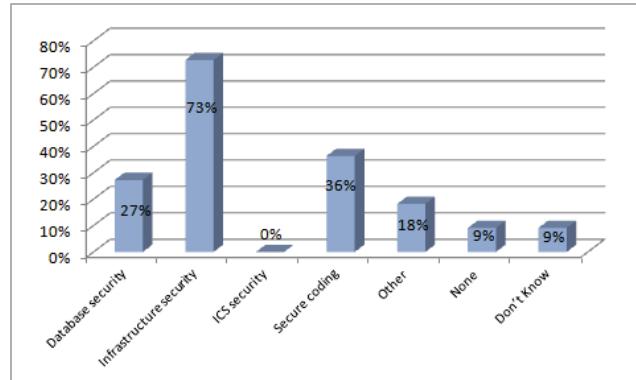
Participants recorder the specific technical measures that the organization is using in order to secure their systems and the data when transmitted or stored in their data centers. Based on the analysis of the data, It was revealed that most of the organizations use a combination of firewalls, anti-spam, antivirus and IDS systems among others. The most common system that almost all participants answered that they are using, was a centrally managed user access control / authentication system while on the other hand a centrally controlled equipment and peripheral device connection control over the internal access network (NAC / Device Control) system was only present at less than 10% of the organizations. With organizations now having to account with an exponential growth of mobile devices accessing their networks and the security risks they bring, it is critical to have tools that provide visibility, access control, and compliance capabilities. A NAC system can deny network access to noncompliant devices, place them in a quarantined area, or give them only restricted access to computing resources, thus keeping insecure nodes from infecting the network (Koh, Oh and Im, 2014).

Responders also answered whether they use commercial or open source solutions for implementing the security measures on their organizations and whether their organization followed an internal audit procedure and how often such audits were conducted. Based on the findings (see Figure 8), most of the organizations follow ad hoc procedures for evaluating the compliance of the organization to the laws and also for assessing the level of security rather than conducting regular internal or external audits.

Finally participants were asked about trainings or certifications in areas related to the security of IT systems, services and infrastructures that employees of the organizations have received recently. Based on the findings, as shown in Figure 9, there was a mix of dedicated trainings, certificates and degrees that employees had. The level of trainings/education was not uniform though. While in several organizations there was a number of trained staff having all of the aforementioned qualifications, in other organizations there was a lack of such trained/expert staff.



**Figure 8:** Audits

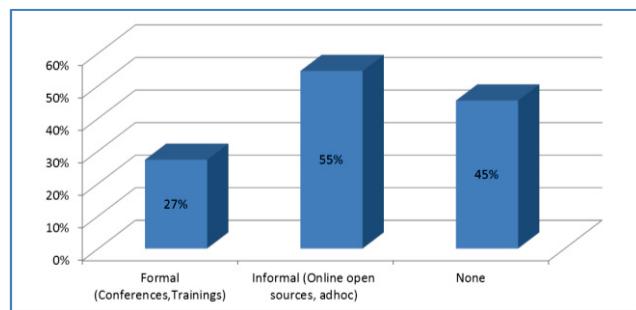


**Figure 9:** Education

### 3.3 User awareness

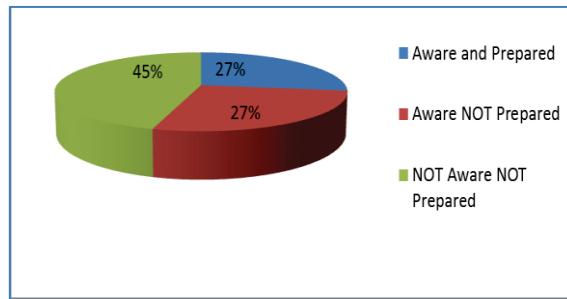
User awareness is investigated through the third part of the questionnaire, in terms of relevant policies and mechanisms in place, along with targeted trainings on new legislative requirements (e.g. NIS Directive, GDPR Regulation).

Organizations were questioned about the established mechanisms that they use to educate their users on security and privacy aspects covering areas like new threats, prevention and reaction practices, legislative requirements and more. Figure 10 shows that the majority of organizations (55%) use Informal ways of achieving such awareness, through online sources in an ad-hoc base (e.g. blogs, mailing lists, social media) and only 27% use Formal and established mechanisms, like specialized conferences and trainings. Meanwhile, 45% reported that there is no structured mechanism for user awareness at all.



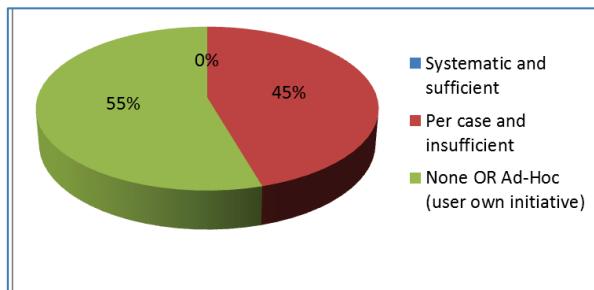
**Figure 10:** User awareness mechanisms

On Figure 11 the overall satisfaction in terms of awareness and preparedness on new legislative requirements (NIS, GDPR) is presented. It was revealed that 45% of the organizations were neither aware nor prepared. This is evaluated as an expected outcome since both legislations were not into force when the questionnaire was first issued.



**Figure 11:** Legislation awareness (NIS, GDPR)

Overall, responders evaluated the current state on user awareness/training policy on their organizations (see Figure 12). Among them, the majority (55%) reported that there was no policy and that this was only viable by own initiative, or, when such policy existed 45% answered that this was insufficient.

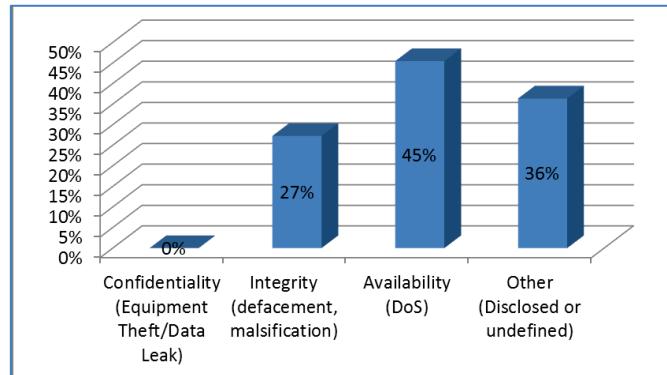


**Figure 12:** User awareness/training policy

### 3.4 Incident response

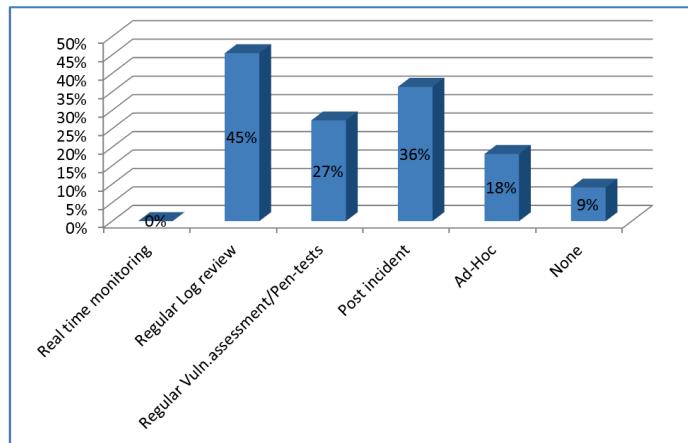
The final part of the questionnaire was focused on Incident response in terms of timely detection, impact evaluation and reaction procedures.

All organizations responded that they have been affected by at least one security incident during the past 12 months and 45% of them had at least one incident relevant to availability disruption, with common type of attack the “Denial of Service” (see Figure 13)



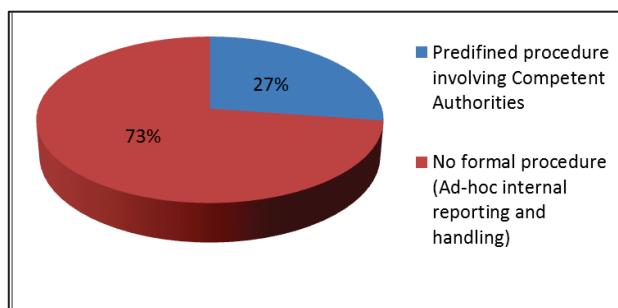
**Figure 13:** Key security principles affected by incidents (past 12 months)

Concerning threat detection, Figure 14 shows that 45% regularly reviewed their log files and 27% tested their systems through vulnerabilities assessment and/or penetration tests regularly (i.e. yearly or sooner). However, it was revealed that none of the responders used a real time monitoring mechanism or a similar procedure. A significant proportion of responders (36%) reported that threats are detected only after a disruptive effect has already occurred and impacted the ICT environment.



**Figure 14:** Threat detection

In light of an incident occurrence, organizations reported that only 27% are following a predefined procedure for filing and handling a security incident with predetermined escalation procedures to Competent Authorities while the rest are handling the incident with an ad-hoc approach leaded by the IT Department (Figure 15).



**Figure 15:** Incident reporting and handling

#### 4. Discussion

The analysis of the results helped us identify what are the major threats that organizations are facing today, create a picture of the current posture in terms of cyber security and define the priorities for strengthening their security. *One interesting result of the study is that 92,5% of security incidents reported have origin in human errors. Although the second part of the questionnaire was mainly focused on technology such as firewalls, anti-spam, antivirus and IDS, it also covered both organizational and certification issues while on the same time the third part was devoted to human factor in terms of awareness and trainings. Based on this, the study revealed that 45% of the questionnaires responded that there is no structured mechanism for user awareness at all.* One of the main concerns that participants had was about education programs, awareness campaigns and exercises that need to be conducted in a regular basis. Dedicated education programs and awareness campaigns can help strengthen the organization and the nation against cyber-attacks (de Bruijn and Janssen, 2017). The majority of educational programs within the cyber security domain are awareness campaigns (Coventry et al, 2014). These campaigns typically use lectures or presentations to articulate the issues surrounding advanced actors to a wide audience, with little tailoring to specific audiences. Experiential learning on the other hand (Kolb, 2014), is an educational technique based on the assumed importance of experimenting and involvement, proposing that active engagement in a scenario develops personal experiences that form the basis of understanding.

As stated in the National Cyber Security Strategy that NCSA has issued in 2018, National Preparedness Exercises are an important tool for evaluating participating stakeholders' preparedness and for detecting weaknesses and vulnerabilities. The simulation of security incidents offers the opportunity for handling these under conditions similar to actual incidents, through implementation of the relevant security measures taken, and of drafted pertinent contingency plans, so that the stakeholders may proceed with relevant improvements and updates (Cook et al, 2017). For these reasons NCSA has decided that a blend of awareness campaigns, dedicated educational programs and exercises must be conducted in a regular basis along with the competent CSIRT, the National CERT and other major stakeholders. NCSA is conducting, hosting or co-organizing a series of awareness

events with OWASP, OSCE and other organizations that are related to cyber security and continue to organize a series for such events for the upcoming months. NCSA is also participating in PANOPTIS that is organized by the directorate of cyber defense of Ministry of Defense. PANOPTIS exercise is organized since 2010 and involves more than 200 people from Armed Forces and Security Bodies, Academic Sector and Research Centers, Public and Private Sector and in 2019 is dedicated to test NIS procedures and mechanisms.

Cooperation both internally inside the Greek Nation and externally with other member states of the EU and beyond, is critical aspect for succeeding in the battle against cyber-attacks against NCIs, reduce the risks of misperception, escalation, and conflict that may stem from the use of ICTs (Boeke, Heinl and Veenendaal, 2015) or even restore peace in the aftermath of a cyber-warfare (Robinson et al, 2018). NCSA has used this questionnaire as a means of initiating a cooperation with relevant stakeholders, create a list of experts that can work together in order to solve problems and increase the overall level of cyber security. In order to strengthen cooperation in European level, NCSA is representing Greece in the NIS cooperation group, in the Horizontal Working Party on cyber issues of the EU, in the informal working group that is set up by OSCE for addressing security of and in the use of information and communication technologies (ICTs), among others. To that sense NCSA is also participating in H2020 and National projects related to cyber security, e.g. CONCORDIA. CONCORDIA, a new H2020 project that started on the 1st of January and lasts for four years, builds a Cybersecurity Competence Network with leading research, technology, industrial and public competences to build the European Secure, Resilient and Trusted Ecosystem.

## 5. Conclusions

Organizations have to manage new risks, sometimes proactively, sometimes by being constrained by regulations such as GDPR or the NIS directive. To cope with new threats, it is essential to develop or reinforce a real culture of cybersecurity at the organizational level. Before putting anything in place, we must start by assessing the new risks to which we are exposed. The new regulations that the EU is issuing, invite organizations and member states to follow these approaches. However, it is not enough to get in compliance to be well protected. Regulations, which lay down very general principles, must be understood in the light of the organization context, its developments and the risks involved. For that purpose NCSA has created a questionnaire as a pre-audit mechanism that was sent to governmental stakeholders of Greece. Using the information that was collected from the answers of the participants, NCSA managed to create a list of experts, identify major threats that their systems face and capture their current level of cyber security.

## Acknowledgements

The authors wish to acknowledge the financial support of the project CONCORDIA, funded under European H2020 Programme (contract No. 830927)

## References

- Andreasson, K. J. (Ed.). (2011). *Cybersecurity: public sector threats and responses*. CRC Press.
- Ayres, N. and Maglaras, L. (2016). Cyberterrorism targeting general public through social media. *Security and Communication Networks* (WILEY), 9(15)
- Bianco, L. J. (2016). The inherent weaknesses in industrial control systems devices; hacking and defending SCADA systems (Doctoral dissertation, Utica College).
- Boeke, S., Heinl, C. H. and Veenendaal, M. A. (2015, May). Civil-military relations and international military cooperation in cyber security: Common challenges & state practices across Asia and Europe. In *Cyber Conflict: Architectures in Cyberspace (CyCon)*, 2015 7th International Conference on (pp. 69-80). IEEE.
- Cook, A., Smith, R. G., Maglaras, L. and Janicke, H. (2017). SCIPS: using experiential learning to raise cyber situational awareness in industrial control system. *International Journal of Cyber Warfare and Terrorism (IJCWT)*, 7(2), 1-15.
- Cook, A., Robinson, M., Ferrag, M. A., Maglaras, L. A., He, Y., Jones, K., & Janicke, H. (2018). Internet of cloud: Security and privacy issues. In *Cloud Computing for Optimization: Foundations, Applications, and Challenges* (pp. 271-301). Springer, Cham.
- Coventry, L., Briggs, P., Blythe, J., and Tran, M. (2014). Using behavioural insights to improve the public's use of cyber security best practices. Gov. UK report.
- de Bruijn, H., and Janssen, M. (2017). Building cybersecurity awareness: The need for evidence-based framing strategies. *Government Information Quarterly*, 34(1), 1-7.
- Evans, M., He, Y., Maglaras, L. and Yevseyeva, I. and Janicke, J. (2019), Evaluating Information Security Core Human Error Causes (IS-CHEC) Technique in Public Sector and Comparison with the Private Sector, Elsevier International Journal of Medical Informatics

- Kovanen, T., Nuojua, V., and Lehto, M. (2018, March). Cyber Threat Landscape in Energy Sector. In ICCWS 2018 13th International Conference on Cyber Warfare and Security (p. 353). Academic Conferences and publishing limited.
- Koh, E. B., Oh, J., and Im, C. (2014). A study on security threats and dynamic access control technology for BYOD, smart-work environment. In Proceedings of the International MultiConference of Engineers and Computer Scientists (Vol. 2014, pp. 12-14).
- Kolb, D. A. (2014). Experiential learning: Experience as the source of learning and development. FT press.
- Maglaras, L., Drivas, G., Noou, K., and Rallis, S. (2018). Nis directive: The case of Greece. EAI Endorsed Transactions on Security and Safety, 18, 5.
- Maglaras, L., Ferrag, M. A., Derhab, A., Mukherjee, M., Janicke, H., and Rallis, S. (2019). Threats, Protection and Attribution of Cyber Attacks on Critical Infrastructures. arXiv preprint arXiv:1901.03899.
- Robinson, M., Jones, K., Janicke, H., and Maglaras, L. (2018). An introduction to cyber peacekeeping. Journal of Network and Computer Applications, 114, 70-87.
- Seker, E., and Ozbenli, H. H. (2018, June). The Concept of Cyber Defence Exercises (CDX): Planning, Execution, Evaluation. In 2018 International Conference on Cyber Security and Protection of Digital Services (Cyber Security) (pp. 1-9). IEEE.
- Tipton, H. F., and Nozaki, M. K. (2007). Information security management handbook. CRC press.
- Wood, A., He, Y., Maglaras, L., and Janicke, H. (2017). A security architectural pattern for risk management of industry control systems within critical national infrastructure.

# Improving Phishing Awareness in the United States Department of Defense

Christopher Dukarm, Richard Dill and Mark Reith

United States Air Force Institute of Technology, Ohio, USA

[christopher.dukarm@afit.edu](mailto:christopher.dukarm@afit.edu)

[richard.dill@afit.edu](mailto:richard.dill@afit.edu)

[mark.reith@afit.edu](mailto:mark.reith@afit.edu)

**Abstract:** Phishing emails are rapidly increasing in sophistication, evolving from poorly crafted attempts to entice a recipient to click, into legitimate looking emails and attachments. In response, email providers have to improve their detection technology by adding new rules to their firewalls and filters to block incoming spam and phishing emails. To overcome technical measures, attackers modify the content of their phishing emails and the source email address. In this cat and mouse game, network defenders rely on the user to report new threats, and the users depend on phishing awareness training to help them identify malicious emails. For a large organization like the United States DoD (DoD) which boasts a workforce of 3.2 million employees, it is difficult to properly train employees to identify and report malicious emails. Like other organizations the DoD requires its employees to complete phishing awareness training, however the effectiveness of this training is widely disputed. Phishing prevention can be broken into three main components: automated filters and firewalls, automated warning messages, and behavioral training. This paper analyzes existing United States DoD phishing awareness behavioral training and proposes 3 principles of an improved behavioral training model. This paper will detail how focused training objectives, a DoD content-sharing platform and a realistic delivery method can be combined to offer an effective and sustainable phishing awareness campaign.

**Keywords:** phishing, social engineering, cyber security, cyber defence

---

## 1. Introduction

The United States (US) Department of Defence (DoD) must continue to adapt to the cyber threats that it receives each day. A domain as fluid and complex as cyber has proven to be extremely difficult to defend. The Internet connects billions of people together using different and evolving technologies. With many attack vectors in the connected cyber domain, the DoD has a challenging time protecting its resources and assets from global cyber threats.

Despite the risks, the DoD embraces communication technologies, such as email, to allow employees to send and receive information to one another, using the Internet as an information transport backbone. It invests billions annually into cyber protection-based technical and procedural mechanisms to ensure the integrity and confidentiality of email exchanges between employees. Unfortunately, cyber threats still manifest themselves in the form of phishing emails, which can be defined as a social-engineering attack that uses email messages to deceive people into disclosing personal information or clicking on malicious links or attachments (Hong, 2012).

Phishing emails have increased in sophistication, evolving from poorly crafted attempts to entice the recipient to click, into legitimate looking emails and attachments. In response to successful phishing campaigns, email providers and companies improve their detection technology by applying filters, firewalls, and threat isolation software to minimize threats. In response, attackers focus their techniques to generate emails targeting specific recipients in what is known as “spear phishing” (Hong, 2012). Figure 1 depicts an example of a spear phishing email where attackers use a technique to spoof the sending email address to fool employees into sharing confidential financial information. Unlike previous attacks, spear phishing requires more background research on the target. This extra information proves to be effective in convincing more users to click on malicious links and attachments (Hong, 2012). Spear phishers circumvent technical protections by limiting their target audience and crafting emails that are crafted. In response, email providers improve their content filters and detection methods. The cat and mouse approach to security cannot be solely relied upon for security against spear phishing attacks.

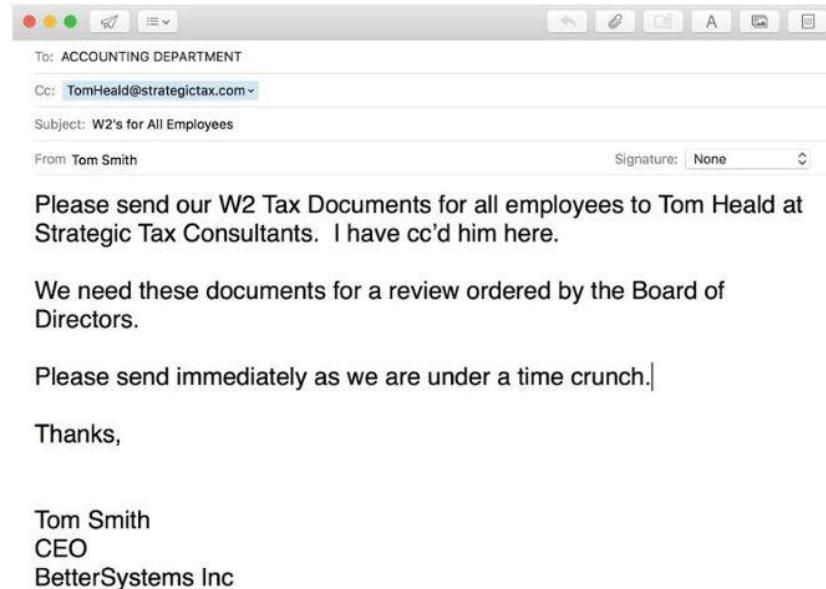


Figure 1: Example phishing email (Johnson 2017)

Training employees to identify the few malicious emails that land in their inboxes is vitally important to the security of the DoD's networks and ultimately, the success of missions across all branches of the U.S. Military. This paper will outline existing DoD phishing training shortfalls and the potential principles that could be used to improve existing training.

## 2. Background

### 2.1 DoD phishing education

The DoD must instil a change in its employees mind-set to understand that relying on cyber for communications while appearing innocuous, opens an attack vector in the cyber domain. Users are unable to comprehend the risks and potential harm that cyber can cause to their personal lives and the mission. Existing cyber training lacks the realism to connect with users emotionally, and as a result users forget the material soon after they receive it. The DoD provides its employees an annual cyber awareness challenge. The goal of the "challenge" is to "provide enhanced guidance for online conduct and proper use of information technology by DoD personnel" (Defense Information Systems Agency, 2017). The course covers various security topics and has a section or mini game focused on identifying phishing emails and the proper reporting procedures.

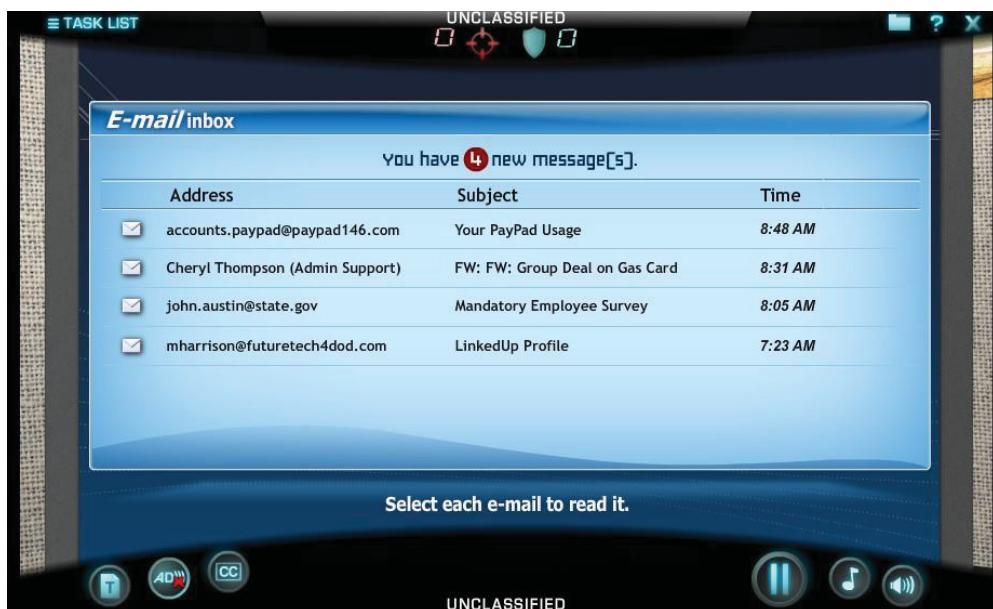


Figure 2: Cyber awareness challenge (defence information system agency)

The training module in Figure 2 displays a simulated inbox and allows the user to go through a series of emails, grading them based on how well they identify potential phishing emails. At first glance this approach does a great job of informing DOD employees about the threat of phishing emails and adequately trains them on how to identify potential threats.

There is a lack of understanding of cyber across our culture. The DoD community grows up watching films of combat in the air, sea and land. The military extensively trains them to protect themselves and those around them long before they are ever thrown into combat. On the contrary soldiers know very little about the cyber domain before they are asked to conduct their mission within its borders. Soldiers experience a brief 60 minute online training module and are expected to know what they need to know to keep themselves and their organizations safe from online threats. User feedback suggests that “training content is not sufficiently processed to ensure that it can be easily recalled when required” (Williams, Hinds, & Joinson, 2018). Most users are motivated to get the certificate saying they passed the training and lack any motivation to learn and retain the material being taught.

The lack of cultural understanding of basic cyber concepts across the Air Force would suggest that the Cyber Awareness challenge needs to be improved. The authors of *Rethinking USAF Cyber Education & Training* point out three main problems of the cyber awareness challenge (Reith et al., 2018). First, the user has no choice in the path they take or mini games they get to complete. Similarly, the phishing submodule of the challenge allows users to choose how they react to emails, but such choices have no visible changes on the simulated network or machine. Secondly, the point system is too generous and ultimately meaningless. When a user decides to click on the nefarious link or open the malicious attachment, the game penalizes them by subtracting a few points. Users can make multiple mistakes within a mini game and still pass that part of the training. This feedback to the user does not properly demonstrate the impact that clicking on a malicious link can have on an organization. Third, gameplay is linear and often unchanged from year to year. Users are naturally going to be frustrated that they have to complete the same game from year to year. Fourth, the training lacks the ability to adjust to a player’s skill level or knowledge on a subject. Please note that since the original writing of this article, the cyber awareness was redone. Despite recent changes, the mechanics and point systems continues to suffer from the same issues as before. The cyber security expert is taking the same training as a standard user. Despite these shortfalls, the cyber awareness challenge continues to offer a level of cyber security training that many companies never offer.

## **2.2 Related work**

Ponnurangam Kumaraguru, a doctoral student at Carnegie Mellon University (CMU) published research on PhishGuru, an embedded training system that trains users to identify phishing emails by sending them simulated phishing emails (Kumaraguru, Hong, Aleven, Tongia, & Acquisti, 2007). PhishGuru replaces the 404 not found pages that appears when phishing sites are taken down and uses an expertly designed infographic detailing tips on phishing prevention. The content is presents a comic strip like format that proved to be engaging when tested with users. Dr. Kumaraguru’s tests proved that these embedded training methods reach the user at a teachable moment after they click on the link, and were more effective than emailing out the same material as attachments to users. Additionally, tests showed that PhishGuru did not decrease users’ willingness to click on legitimate messages.

A follow-on study was performed to analyze a data set from 2014 of 28,471 unique URLs that had the PhishGuru landing page displayed on them (Gupta & Kumaraguru, 2014). The data set contained data of over 3.6 million visits between January and April of 2014. The results demonstrated that 46 percent of users clicked a lesser number of phishing URLs after having exposure to the embedded PhishGuru training pages.

These results solidify the training value of embedded training modules and support the need for the adoption of its use by the DoD. This paper is not arguing that the DoD should merely incorporate third party embedded training as-is , but rather should envelop such technology into a training program containing the following subcomponents:

- Practical objectives
- Visually appealing platform with tailored content
- Realistic delivery method (embedded training)

The rest of this paper intends to build off of existing phishing prevention techniques to provide the DoD with 3 principles that could revolutionize phishing prevention training across its organization.

### **3. Practical objectives**

The DoD clearly defines learning objectives for all DoD-wide training modules, including the Cyber Awareness Challenge. The DoD states that “The DoD Cyber Awareness Challenge addresses the following main objectives (but is not limited to): the importance of IA to the organization and to the authorized user; relevant laws, policies, and procedures; examples of external threats; examples of internal threats; how to prevent self-inflicted damage to system information security through disciplined application of information assurance procedures; prohibited or unauthorized activity on DoD systems; categories of information classification and differences between handling information on the NIPRNet or SIPRNet; requirements and procedures for transferring data to/from a non-DoD network” (Defense Information Systems Agency, 2017). The DoD attempts to teach all of the above stated objectives in a 60 minute course and expects its users to retain and act on the information provided. The objectives would be a lofty goal for a semester-long graduate school course, yet the DoD attempts to meet them through an hour online course. Undoubtedly, these learning objectives makeup a critical requirement for DoD employees and cramming all of them into one module is the most efficient way to meet these demands. In a climate where DoD leadership is attempting to reduce the burden of computer-based training (CBT) on troops, it is unlikely that they would support the creation of new CBTs unless a desperate need was identified. Secretary Mattis, the previous U.S. Secretary of Defense acknowledged the importance of teaching cyber security principles to DoD employees and the Cyber Awareness Challenge remains one of few required CBT modules.

Reducing the amount of proposed learning objectives for future iterations of cybersecurity training will be vital in ensuring that students actually learn and retain the information. Doing this without increasing the amount of required training will require more innovative teaching methods and continued initiatives to identify and cut unnecessary learning objectives and training.

### **4. Platform with tailored content**

Providing custom training for every DoD member is challenging. Social media giants have spent billions on researching and developing platforms that expose users to material that they are interested in. Current education platforms have nowhere near the infrastructure or manpower to provide such capabilities. The problem lies in the source of content. Many educational platforms rely solely on a single provider to procure content. Stale and outdated material has become the norm for educational training as updating content is often not prioritized.

The DoD currently offers the majority of their online training through several branch-specific distributed learning sites (Defense Information Systems Agency, 2017). The government contracts out the creation of their desired training modules and the branches deliver these to their employees through their training platforms. The problem with this approach is that it quickly results in stale content and content creators don't understand the audience to which they are delivering content.

One of the reasons social media sites like Facebook or YouTube are so popular is because their users provide endless amounts of content to the provider. Imagine if YouTube only offered videos created by YouTube employees. Not only would there be significantly less videos, but the content provided would likely not appeal to all the demographics consuming the material. A platform that both allows users to create and consume content related to phishing awareness could prevent material from being stale and would likely be more inviting than existing teaching methods.

In 2014, the Department of Homeland Security (DHS) conducted a study regarding the effectiveness of phishing test emails and the associated click-through rates. The thesis was that they sent several fake phishing emails out to a relatively large subset of people (~1300 people) over the course of 90 days (Caputo, Pfleeger, Freeman, & Johnson, 2014). If a user clicked on a fake phishing URL, they were provided a small infographic with tips to prevent the user from clicking malicious links or attachments in the future. The DHS analysed the time spent on training pages used in a series of tests, and found that despite being directed to training documents, users were not reading the material. The study found that without the proper motivation, users would not consume the educational content made available (Caputo et al., 2014).

Continued phishing-related security breaches makes it apparent that traditional motivational techniques are not effective in encouraging the majority of users to engage with educational material on phishing. Companies have threatened users with disciplinary action if they continue to click on links and attachments in phishing emails. Some companies have gone as far as firing employees for unintentionally infecting machines. A recent study shows that employees who clicked on malicious links or attachments only did so out of a sense of duty to their jobs (Greene, Steves, & Theofanos, 2018). The same study argued that threatened punitive actions would only reduce productivity, as users would potentially be scared to click on legitimate links or attachments.

The DoD needs to take a more personal approach to motivating their users. Stories involving real likeminded people and the affects that phishing has had on their unit's effectiveness would be a great place to start. The DoD encourages such discussions regarding combat related situations and anti-terrorism scenarios. There is a natural interest in these action-related stories, but if framed in the right way stories about cyber attacks have the potential to garner the same level of interest.

Proper gamification of educational concepts has proven to increase motivation levels of students (Carlisle, Chiaramonte, & Caswell, 2015). The gamification of an annual cyber course at the United States Air Force Academy demonstrates just how powerful game-based learning can be. Every year, 17 percent of each class (roughly 200 students) take a short two-week course on cyber security (Carlisle et al., 2015). The majority of the students attending this self-paced course are not majoring in computing disciplines and are frequently placed into the course as a mandatory military training requirement. Despite many students lack of initial interest in cyber security, the course is able to motivate students through competitions and rewards for high scoring players (Carlisle et al., 2015). Cyber gaming at the United States Air Force Academy revolutionized the way cyber security was taught at the institution and demonstrates an effective technique to educate unmotivated personnel on prevention techniques.

The authors of Rethinking United States Air Force (USAF) Cyber Education & Training offer a strategy and framework that aims to offer the USAF a new way to teach cyber education (Reith et al., 2018). Their paper proposes a content sharing and gamification approach to teaching cyber concepts that the DoD could easily apply to phishing prevention across the U.S. Military. Their work encourages Airmen across all USAF communities to contribute content and relies on the community to identify quality content. Current research and development is underway on an open platform that incorporates this framework. The Air Force sponsored Cyber Education Hub (CEH) leverages the same techniques that make social media sites popular and applies them to cyber education. CEH uses public key infrastructure (PKI) certificates in DoD Common Access Cards (CACs) to restrict access to the site and identify users when posting to the site, ensuring confidentiality and integrity. While current beta versions of the site do not offer a specific phishing awareness category, it could be a promising addition to a future iteration of the CEH.

Whether the DoD adopts the CEH or another content sharing platform, users need to be able to contribute to the solution. The DoD has a massive workforce with a diversity of backgrounds and skillsets to offer and failing to leverage these skills continues to be detrimental to the effectiveness of their organization.

## **5. Realistic delivery method**

As discussed in the background of this paper, previous research supports the claim that phishing training delivered to the user through fake phishing emails is more effective than material presented in traditional training modules (Kumaraguru et al., 2007). Fake phishing emails offer a way to surprise the user in a realistic phishing scenario that simply cannot be offered in a training module. The PhishGuru study demonstrates that even basic infographics could be beneficial in training users about phishing.

A content-sharing platform could provide content and ideas for the fake phishing email campaigns within the DoD. DoD employees have already begun using groups on social media sites like Facebook to discuss similar topics and an internal platform would allow them to continue to do so in a more secured environment. User created content would provide the DoD a steady flow of ideas and email content to ensure that fake phishing campaigns would be less of an administrative burdens on the troops tasked with disseminating the emails.

Game developing platforms are also becoming easier to use and could prove to be more interactive than the infographics used by PhishGuru's study. Game developing engines like Unity already offer a platform that can

create basic games in under ten minutes and would allow users to create games that teach basic cyber hygiene concepts. There are also several companies that have developed platforms that allow users to create their own virtual environments on the cloud to teach more complex cyber concepts. These virtual environments would offer a way to challenge more experienced users. Users would be encouraged to produce content as leaders across the DoD would see their work.

## **6. Future work**

A phishing awareness campaign with practical objectives, an appropriate platform and delivery method could be the answer to the DoD's phishing problem. The DoD's 3 million employees offer a wealth of knowledge, diversity and experience that if tapped into could revolutionize existing training across all DoD programs. The key to success lies in providing users with a way to easily create and share content. This approach will require future research and implementation in a number of areas.

- Reducing and refining desired phishing prevention objectives
- Adoption of proper gamification techniques into phishing prevention content and related platforms
- Continued development of DoD content-sharing platforms
- Implementation of phishing simulations across all DoD organizations

The DoD continues to prove its ability to adapt to the constantly evolving domains that it aims to protect and this research aims to continue this tradition.

**Disclaimer:** The views expressed are those of the authors and do not necessarily reflect the official policy or position of the Air Force, the Department of Defence, or the U.S. Government

## **References**

- Alexander, R., 2012. Which is the world's biggest employer. BBC News, 20.
- Caputo, D. D., Pfleeger, S. L., Freeman, J. D., & Johnson, M. E. (2014). Going spear phishing: Exploring embedded training and awareness. *IEEE Security and Privacy*, 12(1), 28–38. <https://doi.org/10.1109/MSP.2013.106>
- Carella, A., Kotsoev, M., & Truta, T. M. (2018). Impact of security awareness training on phishing click-through rates. *Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017, 2018-Janua*, 4458–4466. <https://doi.org/10.1109/BigData.2017.8258485>
- Defense Information Systems Agency. (2017) "Department of Defense Cyber Awareness Challenge" [https://iatraining.disa.mil/eta/disa\\_cac2018/launchPage.htm](https://iatraining.disa.mil/eta/disa_cac2018/launchPage.htm)
- Duran, N., Girtakovskis, J., Jacobi, K., Kennerley, J., Milbourne, G., Moffitt, T., ... Snyder, S. (2018). 2018 web root threat report.
- Greene, K., Steves, M., & Theofanos, M. (2018). No phishing beyond this point. *Computer*, 51(6), 86–89. <https://doi.org/10.1109/MC.2018.2701632>
- Gupta, S., & Kumaraguru, P. (2014). Emerging phishing trends and effectiveness of the anti-phishing landing page. *ECrime Researchers Summit, ECrime, 2014-Janua*, 36–47. <https://doi.org/10.1109/ECRIME.2014.6963163>
- Jansson, K., & Von Solms, R. (2013). Phishing for phishing awareness. *Behaviour and Information Technology*, 32(6), 584–593. <https://doi.org/10.1080/0144929X.2011.632650>
- Jensen, M. L., Dinger, M., Wright, R. T., & Thatcher, J. B. (2017). Training to Mitigate Phishing Attacks Using Mindfulness Techniques. *Journal of Management Information Systems*, 34(2), 597–626. <https://doi.org/10.1080/07421222.2017.1334499>
- Lungu, I., & Tabusca, A. (2010). Optimizing anti-phishing solutions based on user awareness, education and the use of the latest web security solutions. *Informatica Economica Journal*, 14(2), 27–36. Retrieved from <https://search-proquest.com.ezproxylocal.library.nova.edu/advancedtechaerospace/docview/613374491/fulltextPDF/90996CF516D44DFAPQ/38?accountid=6579>
- Lakhita, Yadav, S., Bohra, B., & Pooja. (2016). A review on recent phishing attacks in Internet. *Proceedings of the 2015 International Conference on Green Computing and Internet of Things, ICGCIoT 2015*, 1312–1315. <https://doi.org/10.1109/ICGCIoT.2015.7380669>
- Reith, M., Trias, E., Dacus, E., Martin, S., & Tomcho, S. (2018). Rethinking USAF Cyber Education & Training.

# Putting Together the Pieces: A Concept for Holistic Industrial Intrusion Detection

Simon Duque Antón and Hans Dieter Schotten

German Research Center for Artificial Intelligence, Kaiserslautern, Germany

[Simon.Duque\\_Anton@dfki.de](mailto:Simon.Duque_Anton@dfki.de)

[Hans\\_Dieter.Schotten@dfki.de](mailto:Hans_Dieter.Schotten@dfki.de)

**Abstract:** The fourth industrial revolution, resulting in Industry 4.0, provides a variety of novel business cases. These business cases provide benefits with respect to cost, effort, customer satisfaction and production time. Progress in production can be monitored in real-time by the customer, maintenance can be performed in a remote fashion, time- and cost-efficient production of customer specific products is enabled. These business cases are founded on characteristics of digitisation, namely an increase in intercommunication and embedded computational capacities. Besides the advantages derived from the ever present communication properties, it increases the attack surface of a network as well. As industrial protocols and systems were not designed with security in mind, spectacular attacks on industrial systems occurred over the last years. Most industrial communication protocols do not provide means to ensure authentication or encryption. This means attackers with access to a network can read and write information. Originally not meant to be connected to public networks, the use cases of Industry 4.0 require interconnectivity, often through insecure public networks. This lead to an increasing interest in information security products for industrial applications. In this work, the concept for holistic intrusion detection methods in an industrial context is presented. It is based on different works considering several aspects of industrial environments and their capabilities to identify intrusions as an anomaly in network or process data. These capabilities are based on preceding experiments on real and synthetic data. In order to justify the concept, an overview of potential and actual attack vectors and attacks on industrial systems is provided. It is shown that different aspects of industrial facilities, e.g. office IT, shop floor OT, firewalled connections to customers and partners are analysed as well as the different layers of the automation pyramid require different methods to detect attacks. Additionally, the singular steps of an attack on industrial applications are characterised. Finally, a resulting concept for integration of these methods is proposed, providing the means to detect the different stages of an attack by different means

**Keywords:** industrial (information) security, intrusion detection, attack vectors, hybrid approach, machine learning

## 1. Introduction

The fourth digital revolution, creating Industry 4.0, enables a plethora of novel use and business cases for industrial applications. Founded on the development in communication and computation technology, the Industry 4.0 use cases employ the increase of intercommunication and embedded intelligence, e.g. the digital factory (Stef et al. 2013). These novel business cases can decrease cost, time to production and effort. They enable customer individual production, direct information about the state of production, remote maintenance and operation. Since they rely heavily on communication, connection through network boundaries is essential. However, industrial communication protocols have not been designed with security in mind. Most protocols, such as *Modbus* (Modbus 2012; Modbus-IDA 2006) or *Profinet* (PROFIBUS 2017) do not contain methods to authenticate entities or encrypt communication. Originally, industrial control systems, commonly known as Supervisory Control And Data Acquisition (SCADA) or Operation Technology (OT)-Systems were meant to be physically separated from public networks. Additionally, highly application specific devices and configurations were considered to make it exceedingly difficult for an attacker to exploit the systems (Igure et al. 2006). These assumptions, however, have proven to be wrong. A spectacular series of successful attacks on industrial companies and systems shows that industry has become a target of cyber criminals and state-sponsored actors alike (Duque Anton et al. 2017a). Commercial Off The Shelf (COTS)-products make integration and configuration of new devices easier for operators. On the other hand, they allow attackers to analyse and develop exploits against the devices. Furthermore, the network layout of industry is based on well-established protocols and network segmentation techniques. The application of common techniques makes it easy for attackers to reuse attack techniques. The incentive for attackers is manifold. Attacks on industrial and critical infrastructure facilities can originate in a political agenda, even though attribution is difficult. Different industrial attacks in the past, e.g. the Ukrainian blackout in December 2015 and *Stuxnet*, are rumoured to be part of a political plan. Additionally, cybercrime has become a profitable business. In 2015, the revenue of cyber attacks has surpassed the global profit in drug trafficking for the first time (Leyden 2018). The potential of cybercrime in industry, e.g. by sabotage and espionage, is deemed high.

Industrial applications are characterised by unique properties. They commonly contain a network segment of classic office Information Technology (IT) that is used for communication with customers and partners as well as raw material and product configuration, so-called Enterprise Resource Planning (ERP) and Manufacturing Execution Systems (MESs). Additionally, there is an OT network segment used for control of the production machines. The networks are usually separated by data diodes or De-Militarized Zones (DMZs). Cyber attacks aim on both of these network segments, depending on their goal. However, these kinds of networks are vastly different in terms of connected devices, protocol and usage behaviour. Thus, detecting and preventing attacks on industrial facilities is strongly application- and attack-specific. A thorough understanding of the domain and possible attacks is required, as well as an understanding of promising methods to detect attacks in the individual domains. In this work, an overview of possible attacks on industry is provided, including original vectors and methods for lateral movement. Based on this, promising intrusion detection methods are proposed. Their applicability is derived from prior experiments. Due to the specific nature of industrial environments, different parts of the network and different attack vectors need different methods to detect them. Thus, the concept for a hybrid model integrating different methods required for detecting typical attacks along their path is proposed.

The remainder of this work is structured as follows. In Section 2, an overview of related work is provided. A discussion of attacks on industrial systems can be found in Section 3. Approaches to detect such attacks at different points of the system are introduced in Section 4. A hybrid industrial intrusion detection approach, consisting of a combination of the introduced methods, is presented in Section 5. This work closes with a discussion provided in Section 6.

## **2. Related work**

In this section, works addressing attacks on industrial settings are discussed at first. After that, works discussing industrial intrusion detection are presented and categorised according to aim and method of detection.

There is a variety of scientific discussions on risks and threats in SCADA-systems, e.g. *Zhu et al.* (2011) and (*Igure et al.* 2006). In their work, they discuss the inherent weaknesses of SCADA systems, such as lack in authentication and encryption. *Virvillis and Gritzalis* address the effects of Advanced Persistent Threats (APTs) on industry using the examples of four widely discussed industrial malwares: *Stuxnet*, *Duqu*, *Flame* and *Red October* (*Virvillis and Gritzalis* 2013). *Positive Technologies* provide an exhaustive analysis of the state of industrial and corporate network security as well as risks and threats on industry (*Positive Technologies* 2018). Additionally, the individual attacks have been discussed: *Langner* provides a thorough analysis of the *Stuxnet* attacks (*Langner* 2013). *Lindsay* discusses *Stuxnet* while putting it in the context of cyber warfare and describing its limits (*Lindsay* 2013). *Lee et al.*, *Cherepanov* and *Dragos* discuss the industrial malware *Industroyer*, also known as *Crashoverride* (*Lee et al.* 2016; *Cherepanov* 2017; *Dragos* 2016). It is responsible for the successful attacks on the Ukrainian power grid in December 2015. *Shamir* analyses a new version of *BlackEnergy* that is linked to cyber attacks on the Ukrainian government as well as the aforementioned blackouts (*Shamir* 2016).

In order to address these issues and to protect industrial networks from attacks, a variety of research has been conducted. Due to their periodic, repetitive nature, state information can be used to detect anomalies. This is done by *Khalili and Sami* (2015). *Caselli et al.* focus on sequences in a way that intrusion detection systems are able to understand the step in a process chain they are currently in (*Caselli et al.* 2015). Thus, attacks based on reaching unwanted, but generally allowed states, can be detected. The same approach is chosen by *Fovino et al.* with a focus on *Modbus* and *DNP3* (*Fovino et al.* 2010). The determinism found in industrial applications is made use of by *Hadeli et al.* (*Hadeli et al.*). A similar approach is chosen by *Morris et al.*, they present a system that can be used to detect anomalies in *Modbus RTU* and ASCII-systems in a retrofit fashion (*Morris et al.* 2012). *Gao and Morris* discuss potential attacks on industrial systems and intrusion detection-based countermeasures for *Modbus*-based communication (*Gao and Morris* 2014). *Ponomarev and Atkison* made use of telemetry information to obtain insights about threats and attacks (*Ponomarev and Atkison* 2016).

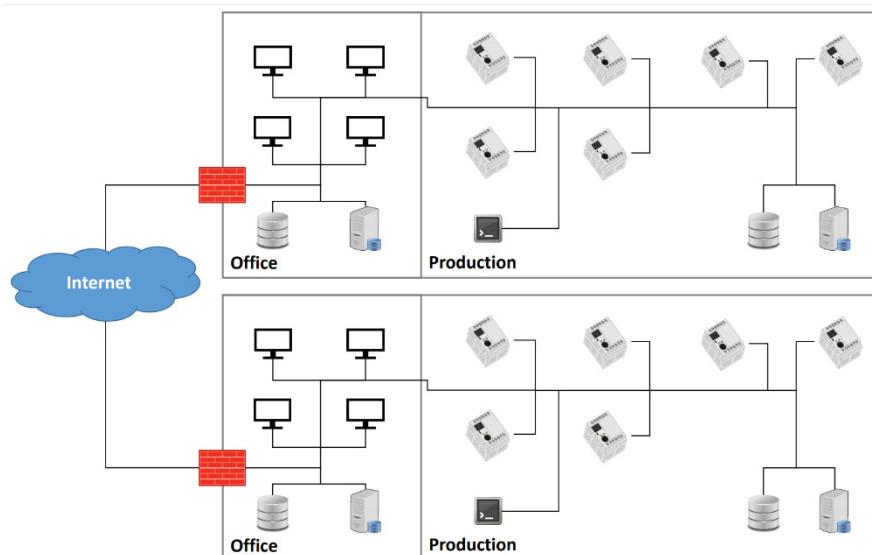
The increase in computational power allows more complex algorithms, such as methods of Artificial Intelligence (AI) or Machine Learning (ML), to work in a time-efficient manner. An overview of the challenges is provided by *Mantere et al.* (2012). *Borges Hink et al.* aim at differentiating between disturbances based on natural deviations and attacks in power grids in order to support human decision making (*Borges Hink et al.*). *Beaver et al.* evaluate ML-based methods for detecting malicious SCADA communication (*Beaver et al.* 2013). Furthermore, *One-Class*

*Support Vector Machine (SVM)* as well as *Ant Colony Optimisation* are used to detect anomalies in industrial networks (Shang et al. 2015; Tsang and Kwong).

Wireless technologies are getting used in an increasing fashion for industrial applications. They are easy and quick in set up and operation. This motivates research on securing wireless channels, e.g. Mobile Ad hoc NETworks (MANETs) (Shakshuki et al. 2013), Wireless Sensor Networks (Shin et al. 2010) and wireless industrial networks (Wei and Kim 2012).

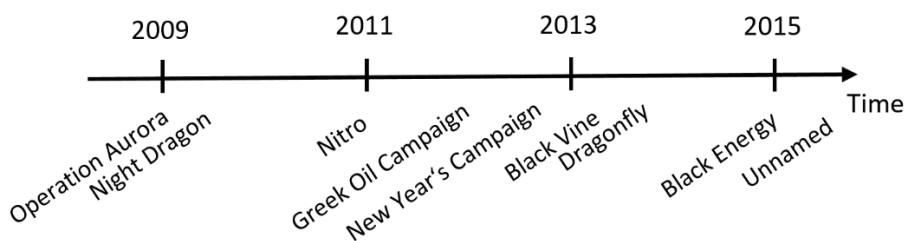
### 3. Attacks on industry – an overview

Attacks on industrial environments are different from attacks on classic home- and office-IT systems. Industrial companies employ IT as well as OT networks. An exemplary network structure found in industrial environments is shown in Figure 1.



**Figure 1:** Typical structure of industrial networks

It can be seen that an office-IT network is connected to the public Internet. This is used for customer contact, company representation and other tasks. However, this network is also used for configuration, creation and modification of OT-level applications, e.g. code for Programmable Logic Controllers (PLCs). If an attacker aims at sabotaging applications, this is the point to execute the payload. If the network where a PLC control is created is connected to a public network, no air gap is in action, allowing for an attacker to compromise OT devices as soon as the IT-perimeter is broken. According to *Positive Technologies*, an attacker could have penetrated the network and accessed the corporate information system in 73 % of the cases they evaluated (Positive Technologies 2018). Many attacks on industrial applications were possible due to phishing, an exemplary set of attacks is shown in Figure 2.



**Figure 2:** Overview of industrial phishing attacks

*Operation Aurora* targeted Google and other software companies (McClure et al. 2010). *Night Dragon*, *Greek Oil* and *New Year's* campaign targeted the energy sector, especially petrol processing facilities (Wueest 2014), while *Nitro* targeted chemical processing facilities (Chien and O'Gorman 2011). *Black Vine* targeted aerospace and healthcare companies (DiMaggio 2015). *Dragonfly* (Symantec Corporation 2014) and *Black Energy* were specifically aimed at Industrial Control Systems (ICSs), with *Black Energy* (Lee et al. 2016) being linked to the

Ukrainian power grid failure in December 2015. The *Unnamed* campaign was used to extract information from ICS systems, making it a corporate espionage tool (Kaspersky Lab 2017).

Apart from spear phishing, any means to break the perimeter allows an attacker leverage to move laterally inside the network. USB flash drive- or SD memory card-based attacks for bridging air gaps have been established, e.g. by the *Stuxnet* attack (Langner 2013). Detecting attacks once the malware has moved to the OT is difficult for several reasons, they are listed in **Table 1**.

**Table 1:** Comparison of IT and OT

Property	IT	OT
<i>Operation times</i>	Commonly around 3 to 5 years	2 to 4 decades
<i>Cost of interruption</i>	Medium	High, hundreds of thousands per hour
<i>Placement &amp; Management</i>	Central, in similar location	Spatially distributed
<i>Upgrade capabilities</i>	Upgrades possible	Upgrades difficult
<i>Software properties</i>	Similar, compatible solutions	Proprietary, vendor-specific solutions
<i>Cost</i>	Moderate	High
<i>Effects</i>	Only digital domain	Physical world

This table shows the difficulty in securing industrial facilities. Their cost and operation time makes add-on solutions for security necessary. However, they need to cope with difficulties to be distributed and upgraded due to software management and distribution of machines. Furthermore, a plethora of proprietary protocols needs to be addressed while downtimes are not acceptable.

#### 4. Industrial intrusion detection approaches

A generic example of a typical industrial network structure was presented in Section 3, pictured in Figure 1. This structure indicates possible places in the system to detect attacks at different stages. In this section, all of those places are discussed. At first, detection is possible when the outer perimeter is broken. After that, lateral movement within the industrial IT network is the goal. Once the malware reached the OT network, it can be detected either by suspicious network traffic or by anomalous process behaviour.

##### 4.1 Detecting perimeter breaks

The task to detect breaches of the perimeter is located in the area of classical IT security, it is not in the scope of this work. It should be noted, however, that the human factor is considered to be the most significant risk with respect to breaches of the perimeter. Phishing and other forms of social engineering have proven to create a significant effect on employees. If a single employee falls for a scam e-mail, attackers can obtain a foothold within the network. For tricking certain, usually high privileged, employees, more targeted methods, such as spear phishing with plausible information or water holing attacks, offer promising results. Unfortunately, technical means can only provide a minor amount of protection against social engineering. Awareness and training of responsible personnel is necessary, in combination with a sound but useable security concept (Surveillance Self-Defense 2018).

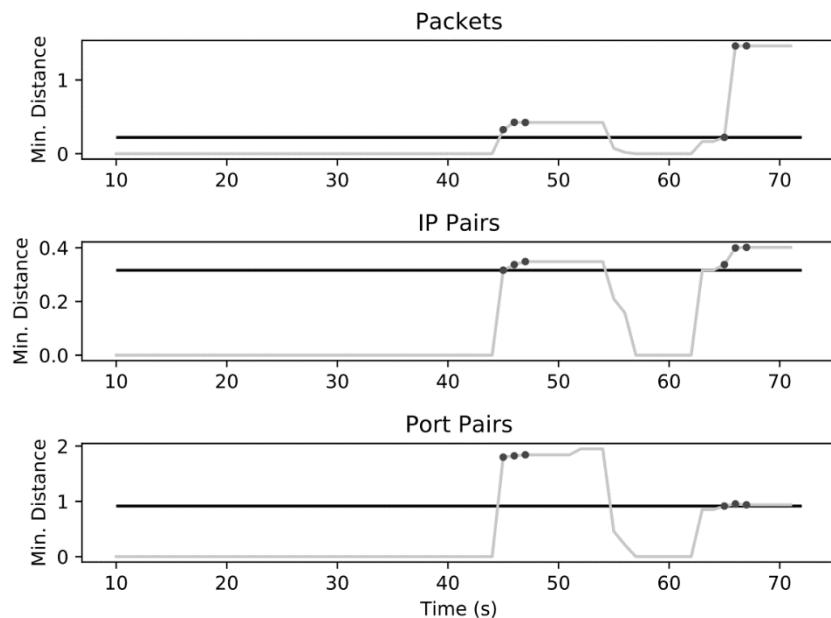
##### 4.2 Detecting lateral movement

Similar to the detection of perimeter breaches, lateral movement in the IT-network need to be detected by well-established IT-security tools, such as Network- and Host-based Intrusion Detection and Prevention Systems (NIDS, HIDS, NIPS, HIPS). Such tools, e.g. firewalls, detect and mitigate the propagation of malware, often with the aid of sophisticated heuristics or Deep Packet Investigation (DPI). Furthermore, the usage of Security Information and Event Management (SIEM) systems can be useful for detecting attacks. Such systems, e.g. *Splunk* (2019) often provide means for the integration of process and OT data as well, allowing industrial IT-security personnel to obtain holistic information about the security status of a network or system.

##### 4.3 Detecting suspicious OT traffic

In contrast to IT-networks, traffic in OT-networks is usually using proprietary protocols that do not consider security objectives. However, the traffic is much more periodic and repetitive than in IT networks. This property can be used to detect attacks in forms of anomalies. Most commercial industrial intrusion detection systems employ a baselining-algorithm that monitors the network for a certain amount of time as a training. After that, deviations from the baseline are flagged as anomalous. Additionally, scientific approaches have been evaluated

in order to find anomalies. Matrix Profiles, as presented by Yeh et al. (2016), have proven to be efficient in detecting anomalies in OT-network traffic (Duque Anton et al. 2018a). This property is shown in Figure 3.



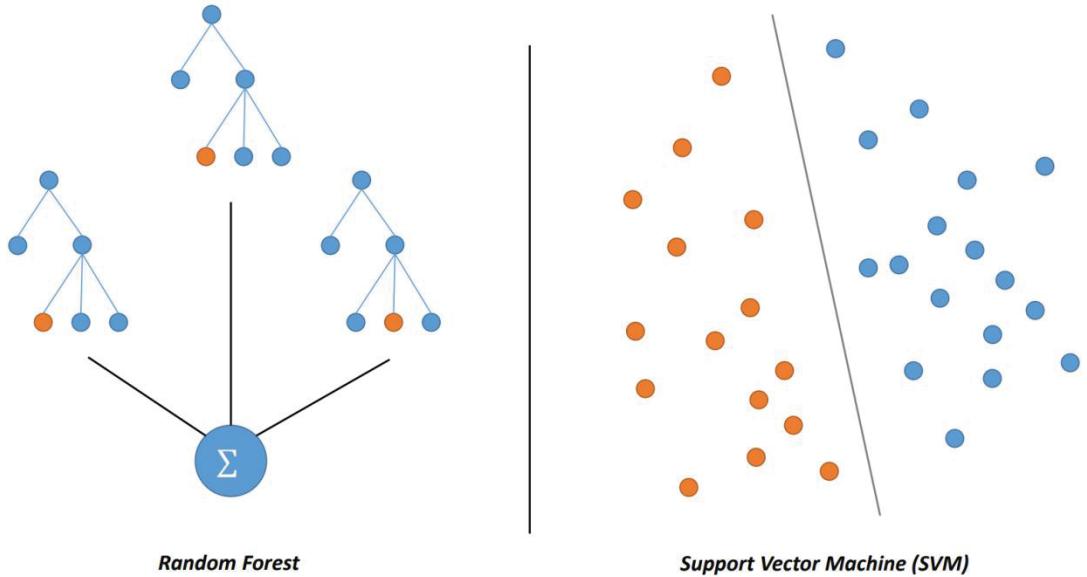
**Figure 3:** Example of *Matrix Profiles* to detect malicious OT traffic

*Matrix Profiles* are calculated based on a distance metric, e.g. the z-Normlized Euclidean distance as follows:

$$d(\hat{x}, \hat{y}) = \sqrt{2m\left(1 - \frac{\sum_{i=1}^m x_i y_i - m\mu_x \mu_y}{m\sigma_x \sigma_y}\right)}$$

The black dots indicate the attacks, the grey line indicates the minimal distance, and the black line indicates the minimum threshold recognising all attacks. *Matrix Profiles* are a windowed approach where for each sequence of length  $m$  the minimal distance to all other sequences of length  $m$  is calculated.  $M$  is the only hyper parameter that needs to be set by the user. Furthermore, *Matrix Profiles* do not need training as such, as they are applied directly to the data. If the minimal distance of a given sequence to all other sequences is small, it is a common motif, meaning it occurs often, indicating normal behaviour. If the minimal distance is high, however, the motif is a singularity which indicates an outlier as well as a potential attack. The evaluation in **Figure 3** is based on an industrial data set introduced by Lemay and Fernandez (2016). They created an emulation of an electric circuit breaker system communicating via Modbus, consisting of three to twelve Remote Terminal Units (RTUs) and one to two Master Terminal Units (MTUs). In doing so, several data sets were monitored, some containing simulated human interaction as well as periodic polling. After that, different kinds of attacks were introduced to a set of the data sets. The attacks are based on TCP/IP, e.g. scanning the network, extracting information or uploading malicious files. The data set under analysis is called “CnC\_uploading\_exe\_modbus\_6RTU\_with\_operate”. As features, the numbers of packets, IP- and port-pairs were used. They provide valid indicators of intrusions.

In addition to the time series analysis of the OT network traffic, packet-based analysis can be used to produce promising results as well. *Random Forests*, *Support Vector Machines (SVM)*, *k-means clustering* and *k nearest neighbours* have been evaluated on three of the data sets provided by Lemay and Fernandez (Duque Anton et al. 2018c). Since *SVM* and *Random Forests* provide promising results, the application of both of them on the data sets “Moving\_two\_files\_Modbus\_6RTU” (DS1), “Send\_a\_fake\_command\_Modbus\_6RTU\_with\_operate” (DS2) and a combination of data sets with and without attacks (DS3). This shows packet-based analysis is a promising approach as well. The concepts of *SVM* and *Random Forest* are shown in **Figure 4**.



**Figure 4:** Random Forest and SVM

*Random Forests* consist of a group of decision trees of which the majority decision is used as the classification. *SVM* is a large-margin classifier, it aims at separating clusters such that the distance of each entity from the separator is maximal.

**Table 2:** Performance of *SVM* and *Random Forest* on OT network traffic

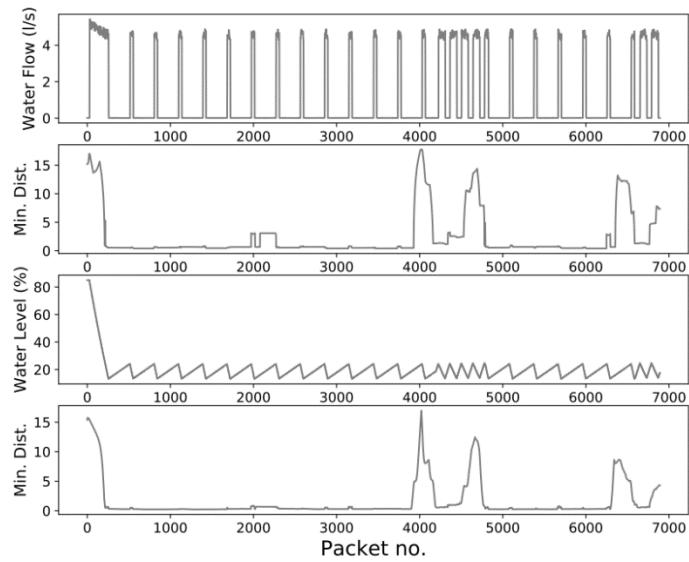
Data Set	<i>SVM</i>		<i>Random Forest</i>	
	F1-score	Accuracy	F1-score	Accuracy
DS1	1.0	1.0	1.0	1.0
DS2	1.0	1.0	0.999851	0.999701
DS3	0.999968	0.999936	0.999986	0.999973

#### 4.4 Detecting suspicious process behaviour

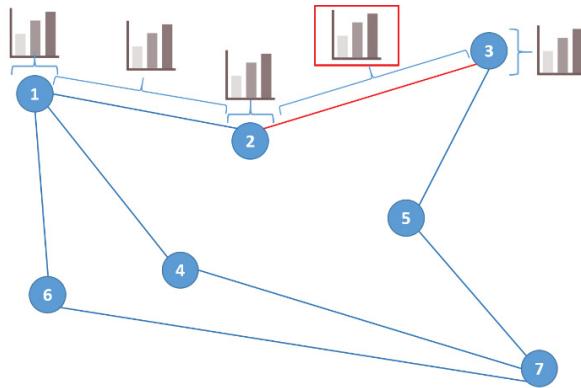
In addition to meta-information, time series approaches such as *Matrix Profiles*, can also be applied to process information, as shown in Figure 5. In the first and third line, the water flow and water tank level of an industrial batch process are shown, as explained in (Duque Anton et al. 2019b) and (Duque Anton et al. 2019a). The second and fourth row describe the according minimal distance. Attacks are introduced at a packet count of 4,000 and 6,500 as analysed and explained by (Duque Anton et al. 2019b). This corresponds to the notable increases in minimal distance, successfully detecting the disruptions of the process as an anomaly.

#### 5. A hybrid concept for industrial intrusion detection

In order to detect the attacks as described in Section 3 as thoroughly as possible, the methods presented in Section 4 need to be combined in an intelligent matter. On the one hand, strong IT-security is required to detect and mitigate intrusions in and lateral movement through the IT network. Additionally, a combination of OT-network and process-based intrusion detection can be used to detect and attribute attacks. If host-based information, e.g. from HMIs, is combined with OT-network data, a dashboard as exemplarily shown in Figure 6 can aid a human operator in detecting attacks, obtaining an overview of the network state and being able to place attacks at their points of origin and destination, thus helping to provide efficient Incident Response (IR). Especially in the OT-domain, time series-based anomaly detection has shown potential, making it a plausible candidate for an intrusion detection mechanism (Duque Anton et al. 2018a).



**Figure 5:** Example of *Matrix Profiles* to detect malicious process behaviour



**Figure 6:** Concept for a hybrid industrial IDS

Combining IT-, OT- and process information has the potential to detect attacks during at least one stage. First, the IT-network has to be breached and traversed. If this happens successfully, OT- or process information can still be used to detect artefacts or effects of an attack, thus increasing the probability of noting an attack. Additionally to the singular information sources, combining information sources and considering contextual information can provide crucial insights into the security status (Duque Anton et al. 2017b; Duque Anton et al. 2018b).

## 6. Conclusion and outlook

In this work, an overview of industrial attack vectors was provided. The stages an attacker needs to overcome in order to successfully penetrate an industrial network, i.e. breaking the perimeter, moving laterally, exploiting industrial hardware, executing the attack, becoming permanent, have been discussed. After that, possible places in the networks commonly found in an industrial environment are presented. It shows that time series are especially promising for the OT and process domain. Finally, the combination of these approaches is expected to address an attack throughout the complete attack cycle, providing a holistic view and aiding the detection.

## Acknowledgements

This work has been supported by the Federal Ministry of Education and Research (BMBF) of the Federal Republic of Germany within the project IUNO Insec (KIS4ITS0001). The authors alone are responsible for the content of the paper.

## References

- Beaver, Justin M.; Borges-Hink, Raymond C.; Buckner, Mark A. (2013): An Evaluation of Machine Learning Methods to Detect Malicious SCADA Communications. In: Moamar Sayed-Mouchaweh and M. Arif Wani (Hg.): 12th International

- Conference on Machine Learning and Applications (ICMLA). Miami, FL, USA, 4/12/2013 - 7/12/2013. IEEE Computer Society. Piscataway, NJ: IEEE, S. 54–59.
- Borges Hink, Raymond C.; Beaver, Justin M.; Buckner, Mark A.; Morris, Tommy; Adhikari, Uttam; Pan, Shengyi: Machine learning for power system disturbance and cyber-attack discrimination. In: 7th International Symposium on Resilient Control Systems, S. 1–8.
- Caselli, Marco; Zambon, Emmanuele; Kargl, Frank (2015): Sequence-aware Intrusion Detection in Industrial Control Systems. In: Jianying Zhou and Douglas Jones (Hg.): CPSS'15. Proceedings of the 1st ACM Workshop on Cyber-Physical System Security. Singapore, Republic of Singapore, April 14. New York, New York: Association for Computing Machinery, S. 13–24.
- Cherepanov, Anton (2017): Win32/Industroyer - A new threat for industrial control systems. ESET.
- Chien, Eric; O’Gorman, Gavin (2011): The Nitro Attacks, Stealing Secrets from the Chemical Industry. Symantec Corporation.
- DiMaggio, Jon (2015): The Black Vine Cyberespionage Group. Symantec Corporation. Available online [http://www.symantec.com/content/en/us/enterprise/media/security\\_response/whitepapers/the-black-vine-cyberespionage-group.pdf](http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/the-black-vine-cyberespionage-group.pdf).
- Dragos (2016): Chrashoverride - Analysis of the Threat to Electric Grid Operations. Dragos Inc (2.20170613).
- Duque Anton, Simon; Ahrens, Lia; Fraunholz, Daniel; Schotten, Hans Dieter (2018a): Time is of the Essence. Machine Learning-Based Intrusion Detection in Industrial Time Series Data. In: 2018 IEEE International Conference on Data Mining Workshops (ICDMW). Singapore, Singapore: IEEE, S. 1–6.
- Duque Anton, Simon; Fraunholz, Daniel; Lipps, Christoph; Alam, Khurshid; Schotten, Hans Dieter (2018b): Putting Things in Context. Securing Industrial Authentication with Context Information. In: IJCSA 3 (1), S. 98–120. DOI: 10.22619/IJCSA.2018.100122.
- Duque Anton, Simon; Fraunholz, Daniel; Lipps, Christoph; Pohl, Frederic; Zimmermann, Marc; Schotten, Hans Dieter (2017a): Two Decades of SCADA Exploitation. A Brief History. In: IEEE Conference on Applications, Information and Network Security (AINS). Miri, Sarawak, Malaysia, November 13-14. IEEE Computer Science Chapter Malaysia: IEEE Press.
- Duque Anton, Simon; Fraunholz, Daniel; Teuber, Stephan; Schotten, Hans Dieter (2017b): A Question of Context. Enhancing Intrusion Detection by Providing Context Information. In: 13th Conference of Telecommunication, Media and Internet Techno-Economics (CTTE-17). Aalborg University Copenhagen: IEEE.
- Duque Anton, Simon; Gundall, Michael; Fraunholz, Daniel; Schotten, Hans Dieter (2019a): Implementing SCADA Scenarios and Introducing Attacks to Obtain Training Data for Intrusion Detection Methods. In: Proceedings of the 14th International Conference on Cyber Warfare and Security (ICCWS-2019). Stellenbosch, South Africa. ACPI: ACPI.
- Duque Anton, Simon; Hafner, Alexander; Schotten, Hans Dieter (2019b): Devil in the Detail. Attack Scenarios in Industrial Applications. In: Proceedings of the 40th IEEE Symposium on Security and Privacy Workshops (SPW).
- Duque Anton, Simon; Kanoor, Suneetha; Fraunholz, Daniel; Schotten, Hans Dieter (2018c): Evaluation of Machine Learning-based Anomaly Detection Algorithms on an Industrial Modbus/TCP Data Set. In: Proceedings of the 13th International Conference on Availability, Reliability and Security (ARES). Hamburg, Germany, August 29 - 30. New York: ACM, S. 1–9.
- Fovino, Igor Nai; Carcano, Andrea; Murel, Thibault De Lacheze; Trombetta, Alberto; Masera, Marcelo (2010): Modbus/DNP3 State-Based Intrusion Detection System. In: Elizabeth Chang (Hg.): 24th IEEE International Conference on Advanced Information Networking and Applications (AINA). Perth, Australia, 20 - 23 April. IEEE Computer Society. Piscataway, NJ: IEEE, S. 729–736.
- Gao, Wei; Morris, Thomas (2014): On Cyber Attacks and Signature Based Intrusion Detection for Modbus Based Industrial Control Systems. In: JDFSL. DOI: 10.15394/jdfsl.2014.1162.
- Hadeli, Hadeli; Schierholz, Ragnar; Braendle, Markus; Tduce, Cristian: Leveraging determinism in industrial control systems for advanced anomaly detection and reliable security configuration. In: IEEEConference on Emerging Technologies & Factory Automation, S. 1–8.
- Igure, Vinay M.; Laughter, Sean A.; Williams, Ronald D. (2006): Security issues in SCADA networks. In: Computers & Security 25 (7), S. 498–506. DOI: 10.1016/j.cose.2006.03.001.
- Kaspersky Lab (2017): Threat Landscape for Industrial Automation Systems in the Second Half of 2016. Available online [https://ics-cert.kaspersky.com/wp-content/uploads/sites/6/2017/03/KL-ICS-CERT\\_H2-2016\\_report\\_FINAL\\_EN.pdf](https://ics-cert.kaspersky.com/wp-content/uploads/sites/6/2017/03/KL-ICS-CERT_H2-2016_report_FINAL_EN.pdf), zuletzt geprüft am 26.06.2017.
- Khalili, Abdullah; Sami, Ashkan (2015): SysDetect. A systematic approach to critical state determination for Industrial Intrusion Detection Systems using Apriori algorithm. In: Journal of Process Control 32, S. 154–160. DOI: 10.1016/j.jprocont.2015.04.005.
- Langner, Ralph (2013): To Kill a Centrifuge. The Langner Group.
- Lee, Robert M.; Assante, Michael J.; Conway, Tim (2016): Analysis of the cyber attack on the Ukrainian power grid. In: Electricity Information Sharing and Analysis Center (E-ISAC).
- Lemay, Antoine; Fernandez, Jose M. (2016): Providing SCADA Network Data Sets for Intrusion Detection Research. In: 9th Workshop on Cyber Security Experimentation and Test (CSET 16). Austin, TX: USENIX Association. Available online <https://www.usenix.org/conference/cset16/workshop-program/presentation/lemay>.

- Leyden, John (2018): Cybercrime: The \$1.5 Trillion Problem. Hg. v. Experian Information Solutions. Available online <https://www.experian.com/blogs/ask-experian/cybercrime-the-1-5-trillion-problem/>, last updated on 09.05.2018, zuletzt geprüft am 14.08.2018.
- Lindsay, Jon R. (2013): Stuxnet and the Limits of Cyber Warfare. In: *Security Studies* 22 (3), S. 365–404. DOI: 10.1080/09636412.2013.816122.
- Mantere, Matti; Uusitalo, Ilkka; Sailio, Mirko; Noponen, Sami (2012): Challenges of Machine Learning Based Monitoring for Industrial Control System Networks. In: Leonard Barolli (Hg.): 26th International Conference on Advanced Information Networking and Applications workshops (WAINA). Fukuoka, Japan. IEEE Computer Society. Piscataway, NJ: IEEE, S. 968–972.
- McClure, Stuart; Gupta, Shanit; Dooley, Carric; Zaytsev, Vitaly; Chen, Xiao Bo; Kaspersky, Kris et al. (2010): Protecting Your Critical Assets - Lessons Learned from "Operation Aurora". McAfee Inc. Available online [https://www.wired.com/images\\_blogs/threatlevel/2010/03/operationaurora\\_wp\\_0310\\_fnl.pdf](https://www.wired.com/images_blogs/threatlevel/2010/03/operationaurora_wp_0310_fnl.pdf).
- Modbus (2012): MODBUS APPLICATION PROTOCOL SPECIFICATION V1.1b3, zuletzt geprüft am 21.06.2017.
- Modbus-IDA (2006): MODBUS MESSAGING ON TCP/IP IMPLEMENTATION GUIDE V1.0b. Available online [http://www.modbus.org/docs/Modbus\\_Messaging\\_Implementation\\_Guide\\_V1\\_0b.pdf](http://www.modbus.org/docs/Modbus_Messaging_Implementation_Guide_V1_0b.pdf), zuletzt geprüft am 21.06.2017.
- Morris, Thomas; Vaughn, Rayford; Dandass, Yoginder (2012): A Retrofit Network Intrusion Detection System for MODBUS RTU and ASCII Industrial Control Systems. In: Ralph H. Sprague (Hg.): 2012 45th Hawaii International Conference on System Science (HICSS). Maui, HI, USA, 4/1/2012 - 7/1/2012. IEEE Computer Society. Piscataway, NJ: IEEE, S. 2338–2345.
- Ponomarev, Stanislav; Atkison, Travis (2016): Industrial Control System Network Intrusion Detection by Telemetry Analysis. In: *IEEE Trans. Dependable and Secure Comput.* 13 (2), S. 252–260. DOI: 10.1109/TDSC.2015.2443793.
- Positive Technologies (2018): Industrial Companies - Attack Vectors. Positive Technologies.
- PROFIBUS (2017): PROFINET Specification. Available online <http://www.profibus.com/nc/download/specifications-standards/downloads/profinet-io-specification/display/>, zuletzt geprüft am 21.06.2017.
- Shakshuki, Elhadi M.; Kang, Nan; Sheltami, Tarek R. (2013): EAACK—A Secure Intrusion-Detection System for MANETs. In: *IEEE Trans. Ind. Electron.* 60 (3), S. 1089–1098. DOI: 10.1109/TIE.2012.2196010.
- Shamir, Udi (2016): Analyzing a New Variant of BlackEnergy 3. Likely Insider-Based Execution. In: *SentinelOne Whitepaper*.
- Shang, Wenli; Li, Lin; Wan, Ming; Zeng, Peng (2015): Industrial communication intrusion detection algorithm based on improved one-class SVM. In: 2015 World Congress on Industrial Control Systems Security (WCICSS). London, United Kingdom. Piscataway, NJ: IEEE, S. 21–25.
- Shin, Sooyeon; Kwon, Taekyoung; Jo, Gil-Yong; Park, Youngman; Rhy, Haekyu (2010): An Experimental Study of Hierarchical Intrusion Detection for Wireless Industrial Sensor Networks. In: *IEEE Trans. Ind. Inf.* 6 (4), S. 744–757. DOI: 10.1109/TII.2010.2051556.
- Splunk (2019): Any Question. Any Data. One Splunk. Unter Mitarbeit von Splunk. Hg. v. Splunk. Available online <https://www.splunk.com/>, zuletzt geprüft am 28.03.2019.
- Stef, Ioan Dorian; Draghici, George; Draghici, Anca (2013): Product Design Process Model in the Digital Factory Context. In: *Procedia Technology* 9, S. 451–462.
- Surveillance Self-Defense (2018): How to: Avoid Phishing Attacks. Hg. v. Surveillance Self-Defense. Available online <https://ssd.eff.org/en/module/how-avoid-phishing-attacks>, last updated on 26.11.2018, zuletzt geprüft am 28.03.2019.
- Symantec Corporation (2014): Dragonfly. Cyberespionage Attacks Against Energy Suppliers. Available online [https://www.symantec.com/content/en/us/enterprise/media/security\\_response/whitepapers/Dragonfly\\_Threat\\_Against\\_Western\\_Energy\\_Suppliers.pdf](https://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/Dragonfly_Threat_Against_Western_Energy_Suppliers.pdf), zuletzt geprüft am 27.06.2017.
- Tsang, Chi-Ho; Kwong, Sam: Multi-Agent Intrusion Detection System in Industrial Network using Ant Colony Clustering Approach and Unsupervised Feature Extraction. In: IEEE International Conference on Industrial Technology 2002, S. 51–56.
- Virvilis, Nikos; Gritzalis, Dimitris (2013): The Big Four - What We Did Wrong in Advanced Persistent Threat Detection? In: Proceedings of the 2013 International Conference on Availability, Reliability and Security (ARES). Regensburg, Germany, 02.09.2013 - 06.09.2013: IEEE, S. 248–254.
- Wei, Min; Kim, Keecheon (2012): Intrusion detection scheme using traffic prediction for wireless industrial networks. In: *J. Commun. Netw.* 14 (3), S. 310–318. DOI: 10.1109/JCN.2012.6253092.
- Wueest, Candid (2014): Targeted Attacks Against the Energy Sector. Symantec Corporation. Available online [http://www.symantec.com/content/en/us/enterprise/media/security\\_response/whitepapers/targeted\\_attacks\\_against\\_the\\_energy\\_sector.pdf](http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/targeted_attacks_against_the_energy_sector.pdf).
- Yeh, Chin-Chia Michael; Zhu, Yan; Ulanova, Liudmila; Begum, Nurjahan; Ding, Yifei; Dau, Hoang Anh et al. (2016): Matrix Profile I. All Pairs Similarity Joins for Time Series: A Unifying View That Includes Motifs, Discords and Shapelets. In: 16th IEEE International Conference on Data Mining (ICDM). Barcelona, Spain, December 12-15. Piscataway, NJ: IEEE, S. 1317–1322.
- Zhu, Bonnie; Joseph, Anthony; Sastry, Shankar (2011): A Taxonomy of Cyber Attacks on SCADA Systems. In: Proceedings of the 2011 International Conference on Internet of Things and 4th International Conference on Cyber, Physical and Social Computing. Washington, DC, USA: IEEE Computer Society (ITHINGSCPSCOM), S. 380–388.

# A Cyber Counterintelligence Matrix for Outsmarting Your Adversaries

Petrus Duvenage, Victor Jaquire and Sebastian von Solms

University of Johannesburg, South Africa

[duvenage@live.co.za](mailto:duvenage@live.co.za)

[jaquire@gmail.com](mailto:jaquire@gmail.com)

[basievs@uj.ac.za](mailto:basievs@uj.ac.za)

**Abstract:** While cyber counterintelligence (CCI) has been a distinctive specialisation field for state security structures internationally for well over a decade, there has of late been growing recognition of CCI's significance to also non-state actors. CCI is gaining main stream traction and is seen a central to proactively mitigating cyber risk and exploiting opportunities. The cybersecurity vendor Panda Labs (2018), for example, recently observed that CCI has increasingly become "more significant among larger companies." Also for smaller role-players which do not have resources for a fully-fledged capacity, CCI offers a way of thinking and an approach towards more robustly asserting their cyber interests. With the growing recognition of CCI's significance, comes an acknowledgment of its complexity. CCI is not an easy to use add-on or plug-in. It is all about the meticulous outthinking and outwitting of both actual and potential adversaries. This paper advances a matrix that on practical level can serve as a concise, high-level 'pocket guide' for outsmarting adversaries by means of a robustly configured CCI endeavour. The matrix's uses include (i) guiding the optimal deployment of offensive and defensive tools; (ii) synchronising CCI with broader organisational processes; and (iii) enabling the configuration of a CCI posture most attuned to particular organisations' requirements.

**Keywords:** cyber security, cyber counterintelligence, risk management, offensive cybersecurity, threat intelligence

---

## 1. Introduction

The breach of the hospitality giant Marriott International - which made headlines in 2018 and exposed more than 500 million customer records - was the second largest to date (Harvard 2018). The Marriot hack is surpassed only by Yahoo's admission in 2017 that breaches affected around three billion of its user accounts (Harvard 2018). Although attribution is contested, both the Marriott and Yahoo hacks are now widely deemed as the work of nation-state sponsored intelligence actors. These actors are also responsible for numerous other damaging cyber-attacks on non-state actors not previously deemed as the in the cross hairs of state intelligence structures. Concurrently, non-state organisations are also increasingly targeted by also other classes of actors with significant intelligence capacities such as crime syndicates, competitors and some corporate entities (coats 2018). Unsurprisingly, then cyber counterintelligence (CCI) has increasingly become "more significant among larger companies" (Panda 2018). For smaller role-players which do not have resources for a fully-fledged capacity, CCI offers a way of thinking and an approach towards more robustly assert their cyber interests (Jaquire, Duvenage & von Solms 2018). With the growing recognition of CCI's significance, comes an acknowledgment of its complexity. CCI is not an easy to use add-on or plug-in. It is all about the meticulous outthinking and outwitting of both actual and potential adversaries. This paper advances a matrix that can serve as a concise, high-level 'pocket guide' for outsmarting adversaries through a robust CCI endeavour. Premised on CCI's passive-defensive and active-offensive dimensions, the matrix (i) guides the optimal deployment of offensive and defensive tools (iii) synchronises CCI with the organisational processes, (iii) and aids the configuration of a CCI posture best-suited for organisations' varying requirements.

The rest of the paper consists of the following of five parts:

- A cursory overview of the CCI matrix and its two composite parts (namely a vertical plane and a horizontal plane)
- Expounding the CCI matrix's horizontal plane which explains CCI's passive-active and defensive-offensive modes.
- Discussing the CCI matrix's vertical plane by means of which we explicate the different levels of CCI's execution, namely strategic, operational and tactical-technical.
- Presentation of a case study to illustrate the CCI matrix's application.
- Conclusion and observations on future research.

## 2. Overview of the CCI matrix

The CCI matrix we advance in this paper comprises a vertical and horizontal plane which can graphically be presented as follow:



**Figure 1:** The cyber counterintelligence matrix (authors)

The CCI matrix's horizontal plane depicted in Figure 1, represents the four quadrants of the CCI postures, namely:

- (1) Passive-defensive
- (2) Active-defensive
- (3) Active-offensive
- (4) Passive-offensive

The CCI matrix's vertical plane aligns CCI with broader organisational processes (such as counterintelligence - CI) at the three organisational levels/layers on which CCI operates, namely:

- (1) Strategic
- (2) Operational
- (3) Tactical/Technical

In this section, we briefly outlined the CCI matrix's composition. In the next section, the CCI matrix's vertical plane is discussed.

## 3. Horizontal plane of the matrix: The cyber counterintelligence modes

In the paper's introduction we observed that CCI is a CCI is not an easy to use add-on or plug-in. To be effective, CCI needs to be executed as part an organisation's CI endeavour. As a subset of CI, CCI are underpinned by time-tested CI principles and notions.

### 3.1 Counterintelligence fundamentals underpinning the CCI matrix

Our CCI matrix's horizontal plane is premised on such two fundamental CI notions. Firstly, that, for a significant part, the wide array of CI measures and tools can be used for defensive and/or offensive purposes. Secondly, that both offensive and defensive tools can be deployed passively and/or actively. Flowing from these two assertions, we can thus infer four modes for deploying CI tools, namely: passive-defensive, active-defensive, passive-offensive and active-offensive. Within CI generally, these modes can be summarised in tabulated format as follows (adapted from Duvenage & von Solms 2014, as compiled from narratives in Prunckun 2012, Sims 2009):

### 3.2 Application of the four sector counterintelligence matrix to cyber counterintelligence

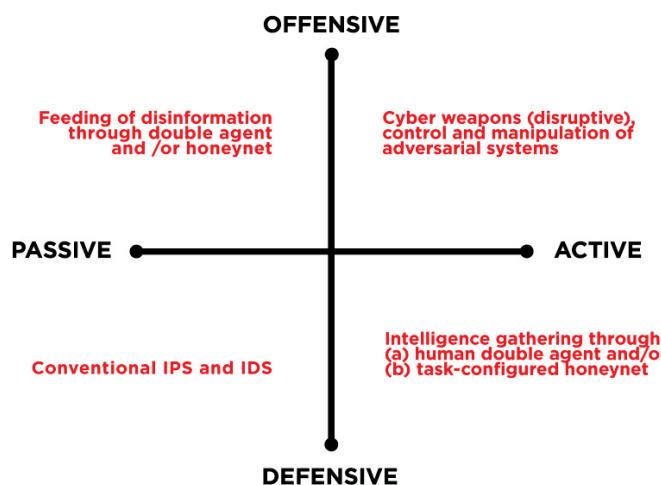
The four-sector CI matrix is applicable to the full spectrum of CCI tools. At the one end of the spectrum, conventional intrusion prevention systems (IPS)/intrusion detection systems (IDS) serve as examples of passive-defensive tools. At the other end of the spectrum, a cyber weapon designed to destroy, disrupt or manipulate

an opponent's systems constitutes an active-offensive tool. CCI tools can seldom be pigeonholed as having only a defensive or offensive purpose, or as being either active or passive. For the most part, to reiterate, one of the paper's recurring emphases, tools are useful to two or more of the four modes. A honeynet, for example, can be used passive-offensively (e.g. to feed disinformation to an adversary) and active-defensively (e.g. to collect information on an opponent). Graphically, this can be depicted as follows:

**Table 1:** Four-sector counterintelligence matrix (adapted from Duvenage & von Solms 2014)

DEFENSIVE MODE Denies adversaries access and gathers intelligence on adversaries	
Passive Defence Mode Denies the adversary access to information through physical security measures and other security systems.	Active Defence Mode The active collection of information on the adversary to determine its sponsor, modus operandi, network and targets. Methods include physical and electronic surveillance, dangles, double agents, moles and electronic tapping.
OFFENSIVE MODE Primarily aims at exploit, manipulate, degrade and neutralise adversarial intelligence. Also gathers intelligence on adversaries' intelligence activities.	
Passive Offensive Mode Reveals to the adversary what you want them to see. This could range from selective exposure of actual information to decoys and dummies. The adversary is thus left to draw its own inferences and interpretations.	Active Offensive Mode The adversary is fed with disinformation and its interpretation thereof manipulated. Disinformation can be channelled through for example double agents and 'moles'. Active-offensive CI could include some forms of covert action.*

\* Covert action, in context of its use in the table, denotes the targeting of an adversary through the influencing of events, conditions, individuals, groups or institutions; to the benefit of a sponsor in a manner not attributable to the sponsor or offering plausible deniability. Influencing is achieved through measures that vary from paramilitary and political actions to propaganda and intelligence assistance.



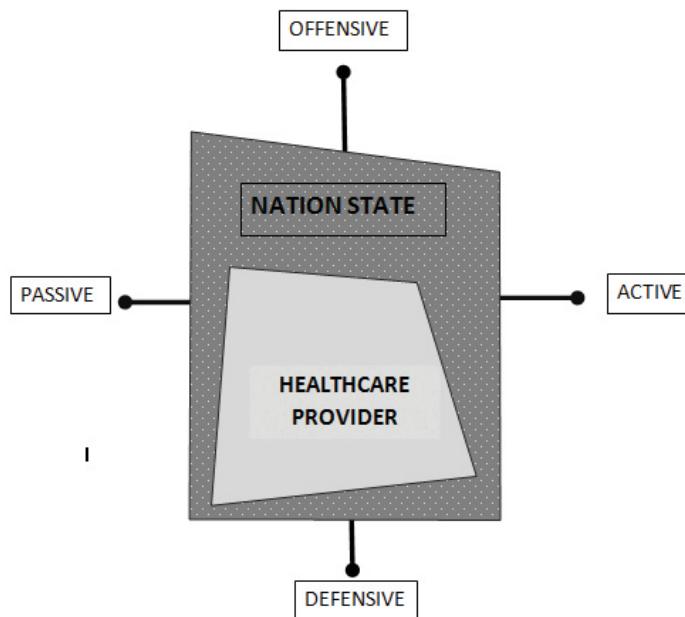
**Figure 2:** Some CCI tools plotted on the CCI matrix's horizontal plane (Authors)

As an academic construct, Figure 2 is useful for the categorisation of CCI tools and to explain their relationship with CI tools in fields other than the cyber field. In CCI practice, Figure 2 could have the following three uses:

- **1. Ensure each CCI tool is utilised to maximum effect.** Since most tools can have more than one purpose, they should be measured against the CCI matrix with the question 'In addition to its initially intended role, in what other modes can the tool be used?' Figure 2, for example, depicts a honeynet deployed in both the active-defensive and passive-offensive modes. To expand on the example used in Figure 2. A honeynet can (if required and depending on circumstances) also be used to facilitate hacking back and the deployment of

cyber weapons (active-offensive). If otherwise configured, a honeynet could furthermore be deployed in tandem with IDS/IPS (passive-defensive). In this hypothetical example, a honeynet is therefore relevant to all four modes.

- **2. Synchronise CCI tools/actions with other CI tools/actions.** The plotting of CCI and other CI tools/actions in Figure 2 will aid the synchronisation of efforts and thus optimise the effectiveness and integration of the CI efforts. The feeding of disinformation through a human agent, to use the example depicted in Figure 2 (passive-offensive mode), should be congruent with disinformation 'planted' in an organisation's honeynet. Incongruencies between these two 'feeds' of disinformation could comprise both CI HUMINT and CCI operations. Similarly, the CCI matrix can be utilised to plot and synchronise CCI tools and actions with those in other Technical Intelligence (TECHINT) fields.
- **3. Configure the CCI posture in accordance with the type and needs of a specific organisation.** Statutory military and intelligence services, for example, will typically have a substantial amount of resources directed to the active-offensive mode. The same will not be the case in relation to, for example, a healthcare provider. Figure 2 can accordingly be used as a template for plotting and appropriately configuring an organisation's CCI posture (See for example Jaquire 2018, Appendix6). For purposes of this paper, we suffice with the following (admittedly oversimplified) comparison:



**Figure 3:** Juxtaposing CCI postures of a nation-state security structure and health care provider (authors)

Implicit to the comparison per Figure 3 is the notion that the configuration of the organisation's CCI posture on the strategic (interests, goals and strategy) levels will ultimately shape CCI activities on the operational and tactical-technical level. These different levels are discussed in the next section as the CC matrix's horizontal plane.

#### 4. Vertical plane of the CCI matrix: Levels of execution

The CCI matrix's vertical plane explains the various levels on which CCI functions and integrates CCI with the broader organisational postures and processes. Since CCI is a CI subset, the importance of synchronising and integrating CCI with especially the organisational CI endeavour on all levels can hardly be overemphasised. As was the case with the development of the horizontal plane, we based our design of the vertical plane on an established CI notion, namely the three levels/layers of execution (strategic, operational and tactical). These levels and their interplay have been described in more detail in existing research (van Niekerk & Duvenage 2016; Duvenage, Jaquire & von Solms 2016; Stech & Heckman 2018, Jaquire 2018). Therefore, this paper suffices with the following synopsis:

**Table 2:** Synopsis of the levels of CCI execution (adapted from Duvenage, Jaquire & von Solms 2016)

	Strategic	Operational	Tactical/Technical
<b>CI mission</b>	Advance and protect organisational interests through defence against and the offensive engagement of adversarial intelligence activities. This is achieved through the following functions: detect, deny, deter, deceive, degrade and/or disrupt.		
<b>CCI mission</b>	As above, when the adversary uses cyber as a conduit or a cyber asset as a target.		
<b>Leadership</b>	C-level	Senior and middle management	Line and team leaders
<b>Interface with CI</b>	Organisational, intelligence and CI strategies All-source CI feed	Multidisciplinary programmes and operations	Multidisciplinary projects and continuous line-functional interaction
<b>Referent objects</b>	Organisation's 'crown jewels' Critical information and cyber-assets sought (e.g. adversary's 'crown jewels') Conditions (competitive advantage)	People, processes, systems, procedures (personal security, ICT architecture and supply-chain management) Own intelligence programme	Systems, networks and devices Network operations Security operations CIA (confidentiality, integrity and availability)
<b>Interrogatives</b>	Who, why?	Who, where, when, how?	What, how?
<b>Level of adversarial role-player (CCI focus)</b>	Sponsors, opponents and Intelligence capacity	Intelligence structures, groups and campaigns	Individuals, TTPs, incidents and actions (on-the-network)
<b>Indicators of targeting and compromise</b>	Geo-political, sector/industry 'flags' Analogous events Adversarial strategy and business decisions	Operational disruption Organisational and/or revenue decline Information leakage	Breach in the C-I-A of cyber and/or information security milieu Identification of malicious code, intrusion and threat exploitation
<b>Analysis output</b>	High-level, strategic appraisals Strategic warning and advisories	Operational reports (CCI operations, threat, damage and vulnerability assessments, alerts and warnings) Trend analyses	Tactical and technical information reports Alerts and warnings
<b>Tools – means, methods and measures (offensive, defensive &amp; collection)</b>	Multidiscipline CI Strategic direction of means, methods and measures	For a <a href="#">taxonomy of the wide array of CCI tools</a> see (Duvenage, Jaquire & von Solms 2016; Jaquire 2018). Interlocked with operational and tactical CI	
<b>Cyber threat intelligence (sourced)</b>	White papers, non-commissioned and non-commissioned research	Platforms	Data feeds
<b>Skill sets required (line-functional)</b>	Sound knowledge of business and industry Specialised knowledge and skills in intelligence, multidisciplinary CI and CCI Strategic analysis and management	Multi-disciplinary CI CCI operational and/or technical specialisation Operational management Elements of both strategic and tactical	ICT and information security Systems, software development, scripting and programming CCI and CCI technical specialisation Ethical hacking Technical cyber defence and collection Humanities, social sciences and languages HUMINT Engineering and reverse engineering

In this section we CCI levels of execution as the CCI matrix's vertical plane. We now proceed with illustrating the matrix's application by means of Stech and Heckman's (2018) hypothetical case study.

#### 4.1 The CCI matrix in practice – a hypothetical case study

As was observed in Section 3.2, the organisation's CCI posture on the strategic (interests, goals and strategy) level will shape CCI activities on the operational and tactical-technical levels. It then logically follows that strategy and operational objectives will determine the offensive-defensive, passive-active modes on a tactical-technical level. This point, as well as the application of our CCI matrix, are illustrated by Stech and Heckman's (2018) proposition on a "Cyber Counterintelligence Framework in Active Defense". Utilising a hypothetical case study of a NATO campaign against "advanced persistent threat actors associated with Russia, APT28 and APT28", Stech and Heckman (2018) pose the following as NATO's strategic CCI goal and operational objectives:

- "Support NATO strategic deception goal: convince Russian authorities their cyber intelligence supports propaganda but is not ready for kinetic war against NATO;
- Active & Passive CCI Defense: Reduce and eliminate effectiveness of APT28 tactics, techniques, and procedures for espionage; Eliminate or counter APT28 and APT29 malware and tradecraft;
- Passive CCI Offense: Poison APT28 and APT29 intelligence stream with deception materials; eliminate, corrupt, or covertly take over control of attackers' command and control; and
- Active CCI Offense: Feed Russian espionage units with false information (e.g. feed APT29 false information about actions and effects of APT28, and *vice versa*).
- Support apparent intrusion successes with cyber and non-cyber strategic NATO deception operations."

In extending the strategic goal and operational objectives to the tactical-technical level, Stech and Heckman (2018) apply our four-sector matrix (advanced per Duvenage & von Solms 2013 and further developed in Table 2 of this paper) advanced earlier in this paper) – as follows:

**Table 3:** Hypothetical NATO cyber CI operations against cyber espionage threat (Stech & Heckman 2018)

<i>Modes</i>	<i>Passive Cyber CI</i>	<i>Active Cyber CI</i>
<b>Defensive mode</b>	<b>Deny access and collect on espionage threat</b>	
	<b>Passive defense:</b>	<b>Active defense:</b>
	Harden endpoint and server configurations	Gather intelligence on on-going intrusions Use honeypots to gather late-stage implants and unpatched exploits
	Share actionable indicators across NATO intelligence partners	Share indicators to force infrastructure and "toolkit" rotations
<b>Offensive Mode</b>	<b>Manipulate, degrade, control and neutralize espionage threat</b>	
	<b>Passive offensive:</b>	<b>Active offensive:</b>
	Use honeypots to deliver deception materials	Counter-hack hop points and control servers Trolling "bait victims" to lure attackers to controlled boxes
	Sinkhole APT28 hop points	
	Identify APT28 operatives	Operating controlled boxes as double agents to inject beacons, double-hacked backdoors, etc. into APT28 control environment

In this section, we illustrated the application of our four-mode, three-tiered matrix by citing Stech& Heckman's (2018) hypothetical NATO case study. We now proceed with observations in conclusion.

#### 5. Conclusion

This paper is submitted within the context of non-state entities' growing adoption of CCI in the face of escalating targeting by intelligence actors of various categories. CCI undoubtedly offers a practicable approach to protect and advance organisational interests. There is, however, a precondition and qualification. CCI not meticulously configured, is more likely to be self-defeating than beneficial. Moving from this premise, key findings of this paper include the following:

- A CCI matrix can aid the configuration of a robust cybersecurity posture and the exploitation of opportunities.
- Such a CCI matrix can be constructed by combining (i) CCI's Passive-Active and Defensive-Offensive modes; and (ii) CCI's levels of execution namely: Strategic, Operational and Tactical-Technical.

On an academic level, the CCI matrix could be useful to conceptually structure aspects of research in this fast growing field.

## **References**

- Coats, D.R. (2018) *Worldwide threat assessment of the US Intelligence Community February 13, 2018*, accessed on 27/07/2018 at <https://www.dni.gov/files/documents/Newsroom/Testimonies/2018-ATA---Unclassified-SSCI.pdf>
- Duvenage, P.C., Jaquire, V.J. & von Solms, S.H. (2016) 'Conceptualising cyber counterintelligence – Two tentative building blocks' in *Published Proceedings of the 15<sup>th</sup> European Conference on Cyber Warfare and Security*, Munich, Germany, June.
- Duvenage, P., & von Solms, S. (2013). 'The Case for Cyber Counterintelligence', *International Conference on Adaptive Science and Technology*. Pretoria, South Africa: IEEE.
- Duvenage, P.C. & von Solms, S.H. (2014) 'Putting counterintelligence in cyber counterintelligence' in *Published Proceedings of the 13<sup>th</sup> European Conference on Cyber Warfare and Security*, Piraeus, Greece, July.
- Duvenage, P., & von Solms, S. (2015). 'Cyber Counterintelligence: Back to the Future', *Journal of Information Warfare*, 13(4), 42-56.
- Harvard University (2018): *Certificate -Managing Risk in the Information Age*, course material, Office of the Vice Provost for Learning Advances (VPAL), Massachusetts, US.
- Jaquire, V.J. (2018) *A framework for a cyber counterintelligence maturity model*, D.Com (Informatics) thesis, University of Johannesburg, Johannesburg, South Africa.
- Jaquire, V.J., Duvenage, P.C. & von Solms, S.H. (2018) 'Building the CCI dream team' in *Published Proceedings of the 17<sup>th</sup> European Conference on Cyber Warfare and Security*, Oslo, Norway, June.
- Panda Security (2018) *The hunter becomes the hunted: How cyber counterintelligence works*, accessed on 06/11.2018at <https://www.pandasecurity.com/mediacenter/panda-security/cyber-counterintelligence/>
- Prunckun, H. (2012) *Counterintelligence: Theory and practice*, Rowman & Littlefield Publishers, Plymouth, UK.
- Stech F.J. & Heckman K.E. (2018) 'Human Nature and Cyber Weaponry: Use of Denial and Deception in Cyber Counterintelligence' in Prunckun, H (ed.) *Cyber Weaponry Issues and Implications of Digital Arms*, Springer, Cham, Switzerland.
- Sims, J.E. (2009) 'Twenty-first-century counterintelligence' in Sims, J.E. & Gerber, B. (eds) *Vaults, mirrors and masks – Rediscovering U.S. counterintelligence*, Georgetown University Press, Washington D.C., US.
- van Niekerk, B. & Duvenage, P. and (2016). "Cyber Intelligence and Counterintelligence," *ISACA South Africa Conference 2016*, Johannesburg, 29-30 August.

# The Offensive Cyber Operations Playbook

Dennis Granåsen and Margarita Jaitner<sup>2</sup>

<sup>1</sup>Division for C4ISR, Swedish Defence Research Agency, Linköping, Sweden

<sup>2</sup>Division for Defence Analysis, Swedish Defence Research Agency, Stockholm, Sweden

[dennis.granasesen@foi.se](mailto:dennis.granasesen@foi.se)

[margarita.jaitner@foi.se](mailto:margarita.jaitner@foi.se)

**Abstract:** Cyber attacks have attracted the attention of politicians and military commanders worldwide, who are now proposing the use of offensive cyber operations as an alternative way of projecting force. This is a controversial issue in politics, some nations have chosen to be open with the fact that they have developed, or are developing, such capabilities, while some have chosen the opposite route. Some nations are still debating whether it is morally justified to engage in offensive cyber operations and have yet to decide on an official stance. This study seeks to review British, American and Russian public doctrines and political documents on offensive cyber operations and synthesize that into a unified language for describing offensive cyber operations.

**Keywords:** cyber operations, cyberspace, taxonomy, doctrine

---

## 1. Cyber this, cyber that

State actors around the world invest time and money into developing cyber capabilities, both to attack and to defend assets in cyberspace, on all three layers: the physical, the logical and the social/persona (Wake, 2018). There is an increasing interest in how this new battlefield, where geographical, ethical and legislative boundaries are still debated, can be exploited (Boer, 2017; Schmitt, 2013). The cyber operation (CO) has emerged as an alternative means to achieve strategic goals in the physical realm as well as in cyberspace (Smeets, 2018). Some nations have gone forward to develop military doctrines that conduct of operations in this new battlefield (Thomas, 2009). Yet many difficult questions, such as under what circumstances foreign cyberspace may be exploited, are yet unanswered (Lambert, 2014).

This article presents an overview of cyber operations and how they are outlined in publicly available documents. It attempts to explain how the USA, UK and Russia define cyber operations and what they can be used for. To summarize, this article presents an approach for dissecting and describing offensive cyber operations in a coherent way that is consistent with military operations. This article does not seek to present any technical details, nor does it attempt to discuss legal or moral aspects of cyber operations.

## 2. National views on cyber operations

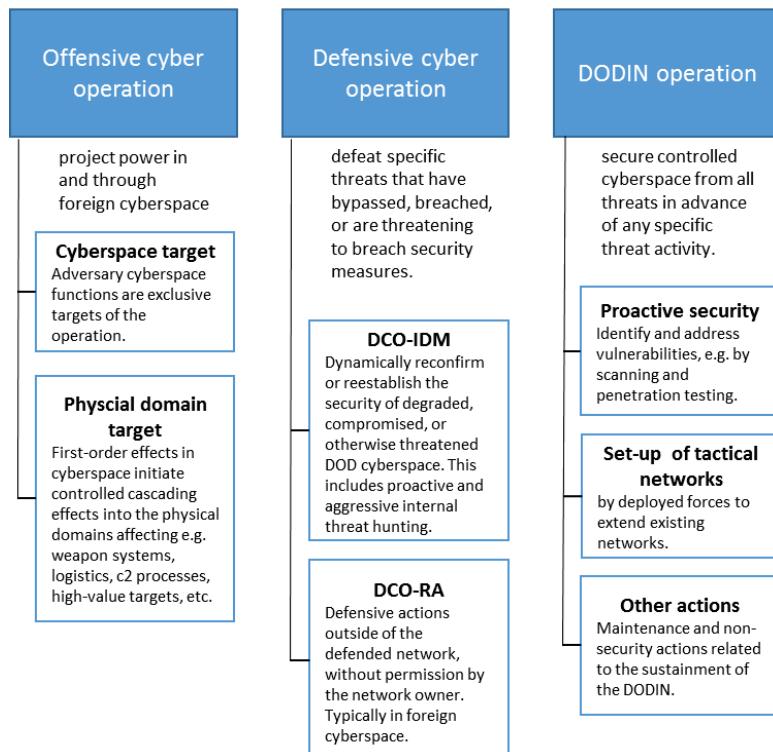
In western military terminology, an operation is generally seen as a sequence of actions coordinated towards a common purpose (NATO NSA, 2013; MOD DCDC, 2014). What is then meant by a cyber operation varies. An objectives-based view implies that the target of the operation resides in cyberspace. For example, Hathaway et al. (2017, p 826) proposes that “a cyber-attack consists of any action taken to undermine the functions of a computer network for a political or national security purpose”. This emphasizes that it is the *functions* that are the targeted and that any means employed to achieve that objective should be considered a cyber-attack. The logical implication of this is that kinetic operations also can be cyber-attacks.

On the other end of the spectrum are the means-based definitions such as Liff’s (2012, p404) definition of cyberwarfare as “computer network operations (CNO) whose means – if not necessarily its indirect effects – are non-kinetic [...].” Such definitions imply that no matter the objective, an operation is a cyber operation if it is conducted in cyberspace. It should be noted that Liff expanded his definition with several constraints regarding objectives valid for cyberwarfare. Nevertheless, the example highlights a difference that may impact the development of national guidelines for cyber operations. In practice, several international organizations are settling on an inclusive definition by claiming that a cyber operation is an operation where the objectives are achieved in, or through, cyberspace (MOD DCDC, 2016; Schmitt, 2013; USCYBERCOM 2018; MOD RF, 2011).

The United States Cyber Command (USCYBERCOM,2018) sponsored Joint publication 3-12 Cyberspace Operations (JP 3-12) and the United Kingdom’s Ministry of Defence’s (UK MOD, 2016) Cyber Primer are two of the more detailed documents. The following sections outline cyber operations as presented in these two documents, as a basis for development of a unified taxonomy.

### 3. Cyberspace operations according to USCYBERCOM

The USCYBERCOM (2018, p. II-2) defines three types of cyber operations: defensive cyberspace operation (DCO), offensive cyberspace operation (OCO) and Department of Defense information network (DODIN) operation, see Fig.1. This taxonomy has sparked a public debate on *jus ad bellum* for offensive cyber operations (Giesen, 2013; Arquilla 2013), separating them from the less controversial defensive cyber operations.



**Figure 1:** USCYBERCOM (2018) taxonomy of cyber operations

*OCO* is defined as operations that either target a function in cyberspace directly, or use a first-order effect in cyberspace to a primary target in the physical domain (USCYBERCOM, 2018, p. II-5). JP 3-12 describes the effects of offensive actions in cyberspace as either manipulation or denial, the latter includes degradation, disruption, and destruction (USCYBERCOM, 2018). The document also specifies that all operations in foreign cyberspace are considered offensive, meaning that intelligence operations could fall into this definition. JP 3-12 does not say much about this type of operations, perhaps because the National Security Agency (NSA) has a significantly larger mandate when it comes to intelligence operations.

In terms of aims, the OCO refers to *targeting* a service or an asset in foreign cyberspace. The definition specifically emphasizes that force may be used to accomplish the objective (USCYBERCOM, 2018, p II-5). Thus, a cyberspace operation where the objective is to shut down a server can, in theory, be executed by destruction applying kinetic force. It should be noted that such an operation would likely also fit under the regulations for conventional operations and therefore may not necessarily be referred to as cyberspace operations.

Actions in cyberspace can also be utilized to make an impact in the *physical realm*. JP 3-12 explicitly mentions causing power outage as one such alternative objective (USCYBERCOM, 2018, p IV-5). The idea is to utilize IT systems that are connected to the target system aiming to generate the desired physical effect. Allegedly, such attacks have been carried out e.g. against the Ukrainian power grid in 2015 (Case, 2016) and on Syria's air defense system in 2007 (Smeets, 2018).

USCYBERCOM (2018, p. II-3) defines *DCO* as operations conducted with the specific purpose of eliminating a known threat to networks that are under the protection of the own forces. DCO that operate in the defended network, such as reconfiguration, isolation and restoration, are called internal defense measures (DCO-IDM) while defensive operations performed external to the defended network, often in foreign cyberspace, are referred to as response actions (DCO-RA) (USCYBERCOM, 2018, p. II-4).

*DODIN operations* cover actions in cyberspace that do not target a specific threat (USCYBERCOM, 2018, p. II-2). Their main purpose is to maintain a secure and operational network, to enable DODIN-dependent capabilities (US JCS J7, 2015, p. I-7). DODIN operations is a standing mission consisting primarily of maintenance events that help sustain an operational DODIN and prepare the environment for other operations (USCYBERCOM, 2018; US JCS J7, 2015).

OCOs and DCO-RAs are normally executed by means of cyberspace attacks and exploitations (Lin, 2010), while DCO-IDMs and DODIN operations are primarily performed by conducting cyberspace defense and security actions. The attack is specified as actions that either deny access to a function, or manipulate information and information systems used by the opponent. Exploitation activities include intelligence, surveillance and reconnaissance (ISR) as well as preparation of the operational environment for current and future operations. Thus, the purpose of exploitation actions is to gain information, cyber situation awareness and access to opponent's systems. Cyberspace defense actions are taking place in the defended networks aiming at detection and mitigation of threats, including restoration of systems. Other protective activities, such as vulnerability removal, firewall configuration, and IT security training and education, are referred to as cyberspace security actions.

#### 4. Cyber operations according to UK MOD

In the preparatory work for the British doctrine for military operations in Cyberspace, the UK MOD (2016) defines a taxonomy of activities in cyberspace that partially overlaps the US definitions. The British cyber operations are subdivided into OCO, DCO, Cyber ISR and Cyber operational preparation of the environment (cyber OPE), see Fig. 2. The official documents point out that cyber operations often are intertwined and dependent on each other. For example, an offensive cyber operation may depend on intelligence gathered in a Cyber ISR operation, and vice versa.

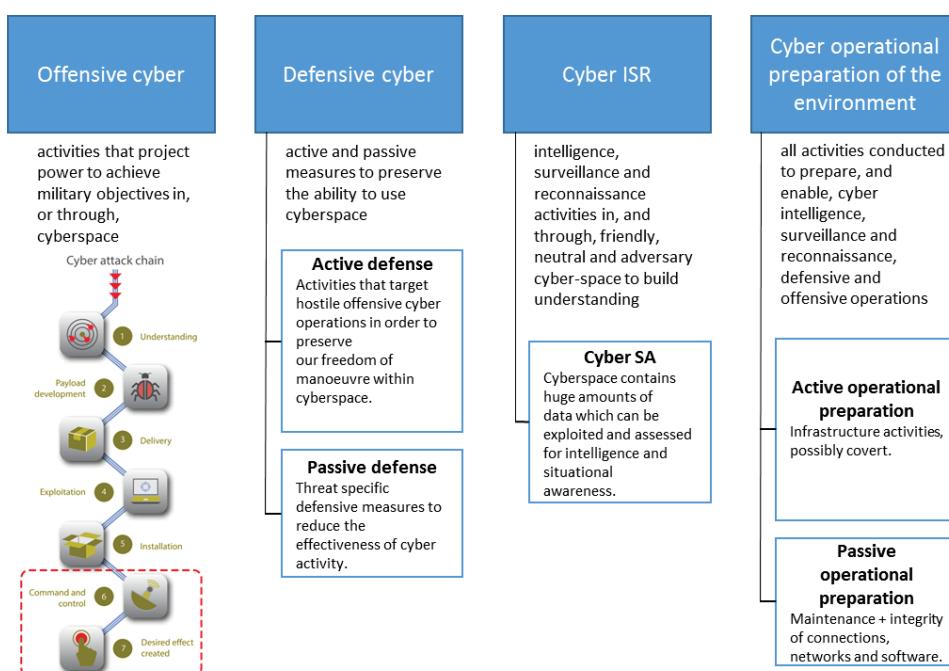


Figure 2: Activities in cyberspace, the UK MOD (2016) view

Similar to US definitions, the UK Cyber Primer document proposes that *OCO* consists of activities aiming "to achieve military objectives in, or through, cyberspace" (UK MOD, 2016 p. 54). In general terms an offensive cyber activity consists of an attack chain with seven phases: understanding, payload development, delivery, exploitation, installation, command and control, and effect achieved (UK MOD, 2016, p. 55).

The UK *defensive cyber operations* aim to preserve the ability to use cyberspace. The British doctrine specifically distinguishes active from passive cyber defense. Although definitions for active and passive cyber defense operations are provided, see Fig. 2, these definitions leave room for interpretation. The Tallinn manual (Schmitt, 2013) delimits passive cyber defence to measures for detection and mitigation of intrusions and adverse effects thereof, whereas pre-emptive or countering operations against the source are defined as active cyber defense.

Such definitions indicate that active cyber defense implies cyberspace attacks and exploitation. Denning (2014, p. 109) provides even further clarification on the subject: “active defenses are direct actions taken against specific threats, passive defenses focus more on making cyber assets more resilient to attack”.

*Cyber ISR* encapsulates such activities that contribute to improved cyber situational awareness (SA) with the purpose of enabling better-informed decision-making (NATO NSA, 2013). Cyber ISR contributes to SA by collecting relevant information within own, neutral and adversary cyberspace (UK MOD 2016, p 56).

*Cyber OPE* consist of preparatory and supporting actions that enable other types of operations, regardless of their type (UK MOD, 2016, p. 57). Maintenance and sustainment operations such as the network and infrastructure management in controlled networks are referred to as passive Cyber OPE operations, while their active counterpart includes typically covert operations that aim to enable execution of other operations, e.g. in foreign cyberspace.

## **5. Russian views on information operations**

Claims that Russia possesses significant cyber capabilities regularly surfaces in the west. These claims are rarely acknowledged by the Russian state, nor are they refuted unless they are allegations of an attribution of an advanced persistent threat (APT). From a westerner’s perspective, this tight-lipped communication strategy is not very sensational, as it is expected that military capabilities and operations in cyberspace are surrounded by secrecy. However, from a Russian perspective it is notable since Russia is often quick to boast its developments, e.g. regarding kinetic weaponry. In addition to the covert nature of cyber operations, one reason that there are few public statements from Russian authorities on cyber operations is that such operations fall under the well-developed umbrella of information operations, alongside with psychological operations and, to some extent, electronic warfare (Makarenko, 2017).

Russian information operations include “actions taken to achieve information superiority within the national military strategy by influencing enemy information and information systems while strengthening and protecting one’s own information, information systems and infrastructure” (Makarenko, 2017). Information operations, in turn, come into use within the framework of information warfare. This includes inflicting damage on information systems and resources as well as psychological manipulation of the population and meddling with the decision-making processes at state level (MOD RF, 2011). An important principle in this regard is that of complexity of action (*комплексность*), which stipulates coordination of activities in information space within a single unified system (Makarenko, 2017). Translated into western terminology this is comparable to the coordination of cyber operations with conventional information operations, e.g. psychological operations, to achieve an overarching objective (MOD RF, 2011).

The introspective security perspective on information operations is arguably the most similar to what in the west is referred to as defensive operations. The Russian information security doctrine defines information security as “protection of [Russia’s] national interests in the information sphere defined by the totality of balanced interests of the individual, society, and the state” (Thomas, 2009). Information security operations are thus a means to protect Russian interests within and outside cyberspace.

Although doctrinal documents describe Russia’s activities in cyberspace in defensive terms, they particularly stipulate carrying out “preemptive strikes” in order to “counteract the development (conservation or aggravation) of the conflict and its transition to a state that significantly increases the costs” (Makarenko, 2017). Thus, in accordance with this doctrine, offensive operations may be carried out whenever deemed necessary. At this point it should be noted that beyond the doctrines, there is a well-developed Russian military and academic discourse regarding offensive operations (Makarenko, 2016).

## **6. Guns and heroes or ones and zeros?**

The goals of a military operation are often achievable in more than one way, likewise an OCO can many times be replaced by a kinetic operation. For instance, if the objective is to deny access to a service, then instead of launching a denial of service (DoS) attack, an aggressor could sabotage the physical infrastructure and achieve the same effect. It should be noted that according to the definition in JP 3-12, this could also be considered an OCO.

The reverse is also true. Kinetic operations can sometimes be replaced by cyber operations. A common example that allegedly impacted the physical realm is Stuxnet. This attack manipulated information systems to search out and infect control systems with malicious code, which in turn caused Iranian uranium-enrichment centrifuges to break down (Barzashka, 2013). While actual effects are yet to be confirmed, it has been confirmed that Stuxnet had these capabilities (Barzashka, 2013). Thus, Stuxnet proves that this type of attack is viable and that cyber operations can indeed make an impact on physical targets.

NATO's stance is that offensive cyber operations can strengthen the defense by deterring potential adversaries (Makenzie, 2017). Perhaps that is also one of the strongest benefits of acquiring offensive cyber capabilities. The Russian take on information operations shows that cyber operations can also be considered a tool for information manipulation and control in a broader sense. Offensive cyber operations also have the benefit of speed, low cost, weather-insensitivity, and stealth (Yasar et. al 2011). The major drawback is that an OCO requires a lot of intelligence regarding the target system, which requires a lot of effort (Yasar et. al 2011). Actions in cyberspace have first-order effects on the systems they exploit, i.e. effects on the logical level of cyberspace. However, the objective of an OCO can often be a second-order physical effect that is triggered by the cyber activity, for which the Stuxnet case serves as example. In turn, recurring campaigns on social media such as the Cambridge Analytica scandal are examples of cyber effects in the social layer.

### **6.1 Effects in the logical layer**

In cyberspace, the natural objectives of OCOs revolve around the information assurance triad: confidentiality, integrity, and availability (CIA). Breaching confidentiality would imply accessing confidential data (e.g. espionage), availability attacks are a matter of denying an opponent access to their systems (e.g. DoS) and integrity attacks are such attacks that seek to manipulate systems and data. Cyber operations can be a composition of several attacks, spanning over multiple elements in the CIA triad.

### **6.2 Hacking cyber-physical systems**

While actions in cyberspace cause first-order effects on the CIA triad, second order effects can impact the physical world. The conceptual attack on a cyber-physical system works by manipulating a control system connected e.g. to an industrial production. Such attacks are becoming increasingly more frequent (Applegate, 2013). The Stuxnet case is an example that effectively demonstrates this concept. NotPetya is another attack that effectively scrambled all data on the target system, rendering it unusable. Although it does not amount to a cyber-physical effect by itself, the context indeed provided physical effects since NotPetya targeted control systems. Notably infection of monitoring systems halted production systems all over the world, for instance the Cadbury chocolate factory in Hobart (Page, 2017).

An extreme form of cyber-physical attack is one that physically targets people. While cyber attacks have to date still not caused any confirmed fatalities, it may be unwise to neglect the possibility of future fatal cyber operations. In particular this is true because connected cyber-physical systems, including within biotech, are becoming increasingly more common (McGraw, 2017).

### **6.3 Social effects of cyber attacks**

Effects on the social (persona) layer impact individual persons and organizations on an abstract level rather than physical. In some cases such attacks are directed, targeting a specific cyber-persona. Social effects can also target groups of people or even entire populations. Conceptually, cyber operations with social effects are similar to psychological operations.

*Doxing* is a popular term for publicly releasing sensitive information about an individual or an organization. Often, the idea of doxing is to discredit or gain leverage on a target with results varying in magnitude depending on the nature of information. The following case shows how cyber operations may be used to affect the political landscape: In January 2019, hackers released personal details including private communications and financial information of hundreds of German politicians, from all major parties except the far-right Alternative für Deutschland (AfD) (CSIS, 2019). No attribution of the attack has been made public. It may seem obvious that the purpose of the attack was to discredit mainstream politicians and thereby benefit AfD. However, easy-to-draw conclusions may mean that the aim was rather to lay blame on AfD (CSIS, 2019). Yet another explanation includes an external actor seeking to spread confusion and raise political tensions. Publicised manipulated, or truthful,

information can be used to affect the public opinion on just about any matter, and the rise of social media has made this type of attacks easier than ever to perform, as demonstrated by tales of troll factories and election-interference.

## **7. The ABC of OCO and CKC**

With plenty coverage of well-known attacks in the past (e.g. Stuxnet and NotPetya), there is no doubt that it is possible to develop a historical cyber operations playbook. The relevance of such a playbook is though questionable, given the short time frame between public exposure of an attack and the release of patches to remove the exploited vulnerability. This leads to the paradoxical conclusion that a public playbook of *modi operandi* for offensive cyber operations is little but a defense manual. Therefore we propose a method for developing cyber operations by aligning the Cyber Kill Chain (CKC) methodology (Hutchins et. al., 2011) with a simplistic description the phases of a conventional military operation: planning (defining objectives and constraints; preparation of the offensive), manoeuvring (establishing access), execution and exfiltration.

### **7.1 Define objectives and constraints**

The first step of the process is the actual decision to pursue the objective through operations in cyberspace. Objectives set the premise for how to plan and execute the operation. Depending on the level of specification in regard to objectives, the commander may have more or less freedom to design and conduct operations. Certain constraints may apply. Conventional military forces are subject to authoritative rules of engagement (ROE) that define conditions and constraints on the use of force. Politicians and scholars alike have proposed that ROEs for military operations in cyberspace should follow the same principles (Kehler et. al, 2017). Aside from legal restrictions, there may be political considerations for state-sponsored cyber operations since their uncovering may trigger unintended and unforeseen responses.

### **7.2 Prepare the offensive**

Information gathering is an important step in any operation, regardless of whether the operation is conducted in cyberspace or not. For OCOs, intelligence is gathered to learn more about system designs, routines, and vulnerabilities as well as to identify potential targets and possible exploits. This is highlighted in the CKC model that specifies reconnaissance as the first step of an operation.

The second step is weapons development, i.e. to design a malware or piece of code that can cause the desired effect on the target. Cyber weapons are designed to trigger a specific effect on a designated target. While there are generic approaches to weaponizing exploits, it is reasonable to expect that higher-value targets are better prepared for well-known attacks, or at least they ought to be. Breaching such systems therefore often requires more sophisticated and customized cyber weapons. This is also one of the reasons why several governments are engaged in the shady business of acquiring and stockpiling zero-day vulnerabilities (Ablon & Bogart, 2017).

### **7.3 Establish access**

Unauthorized access is sometimes done by exploiting weaknesses in hard- and software components of the target system, sometimes by utilizing human operators to trigger an exploit (e.g. by clicking a malicious link). This approach for gaining unauthorized access is typically based on intelligence on system design, and consequently access may be lost due to reconfigurations, patches and upgrades done to the target system. To reduce the risk of losing hard-earned access, it is not uncommon that the adversary utilizes this initial access to install a more permanent backdoor that is more likely to survive the defenders' countermeasures (Hutchens et. al., 2011).

### **7.4 Execute offensive action**

Once adequate access is secured to the target system, CKC specifically advocates the establishment of a two-way communication channel, allowing to remotely control the implants in order to accomplish the mission objectives. While this is partially true, the CKC fails to specify that not all weapons need to be remotely controllable. Hence this step is dependent on the objectives of the operation and the level of autonomy in the weapon. E.g. a cyber operation designed to covertly collect and extract information will undoubtedly need to be able to communicate data. This can be achieved with a fire-and-forget approach, given that the collection routines and the exfiltration approach are coded in such a way that no human-in-the-loop is needed.

## **7.5 Exfiltrate**

Conventional operations often do not end until the soldiers have been extracted from the hot zone. This has two important purposes: (1) to protect the operatives, and (2) to maintain deniability. The cyberspace equivalent is to remove any evidence of attack. This can be done either by restoring systems and manipulating log files, or false-flagging, i.e. planting evidence that points the investigators in the wrong direction. Regardless of modus operandi, successful exfiltration is imperative for black operations, as it ensures plausible deniability.

## **8. Summary and conclusions**

Several governments have over the last decade publicly announced that they are engaging in offensive military operations in cyberspace. The US and the UK have made public documents describing their respective views on the matter. Other nations have been straightforward with their intention of acquiring an offensive capability in cyberspace, while not disclosing any details. Meanwhile, the publication of the Tallinn manual has raised the awareness of legal and ethical considerations of such operations.

The review of the scrutinized doctrines – British, American and Russian shows certain differences to how countries approach cyber operations:

A major difference between the US and the UK taxonomies is that the UK documents do not categorize ISR as neither offensive nor defensive. According to the US doctrine, all operations in cyberspace are offensive unless they are conducted within own networks or in order to defend own networks against an imminent or ongoing threat. This shows a difference in a sense that the US acknowledge cyber ISR conducted on foreign cyberspace as an offensive act, whereas the UK MOD is not as explicit on this particular point. Another difference is that the British categorization of cyber operations appears regard the *actions performed* rather than the *operation objectives*. In that sense the two documents are not semantically aligned, and direct comparison is therefore not advised. In contrast, Russia prefers to publicly retain a defensive terminology.

The publicly available doctrinaire sources on Russian information operations witness of a purpose-driven approach, rather than a technical orientation on how, where and why operations are executed. While NATO and its member and partner nations appear viewing operations in cyberspace as a complement to other military operations, Russia has a more direct approach to integrating cyber operations by defining them as information operations. Arguably, what stands out most from the Russian viewpoint on information operations, is the emphasis on control of information flows in the physical and virtual realms alike. This may have a profound impact on whether pursuing action in cyberspace is the most appropriate choice for the given military (or political) goal, or not. It also allows integration of cyber activities with other approaches such as psychological or kinetic. This in turn may largely impact the content of the steps taken for planning and executing cyber operations. Further, it is suggested, that the Russian approach brings the fundamental goal of impacting the adversary's decision-making process closer to the actual operation and the (cyber) battlefield.

Cyber attacks are commonly described using the CKC, which is intended to be analogue to the military kill chains used to describe conventional military attacks. As the CKC focuses on attack rather the operational aspects, it is not very useful for developing OCO playbooks. Instead, an OCO can somewhat simplistic be described in terms of the following five generic phases: define objectives and constraints (planning), prepare offensive action (ISR), establish access, execute offensive action, and exfiltration. A reality check against the various actors' doctrines can then present what the actual content of this process for the given actor, as suggested above applying the example of the Russian doctrinaire thought. Such an approach enables coherent definitions and descriptions of what offensive cyber operations really are and thereby contribute to increased understanding of how to develop, use, and defend against such operations, tailored to the potential adversary.

## **References**

- Ablon, L. and A. Bogart (2017) *Zero Days, Thousands of Nights: The Life and Times of Zero-Day Vulnerabilities and Their Exploits*. Santa Monica, CA: RAND Corporation.
- Applegate, S. D. (2013) "The dawn of Kinetic Cyber," 2013 5th Int. Conf. Cyber Confl. (CYCON 2013), no. June, pp. 1–15.
- Arquilla, J. (2013) "Twenty Years Of Cyberwar," *J. Mil. Ethics*, vol. 12, no. 1, pp. 80–87, 2013.
- Barzashka, I. (2013) "Are cyber-weapons effective?: Assessing Stuxnet's impact on the Iranian enrichment programme," *RUSI J.*, vol. 158, no. 2, pp. 48–56.
- Boer, L. J. M. (2017) 'Spoofed presence does not suffice': On territoriality in the Tallinn manual, vol. 47.

- Case, D. U. (2016). Analysis of the cyber attack on the Ukrainian power grid. *Electricity Information Sharing and Analysis Center (E-ISAC)*.
- CSIS – Center for Strategic & International Studies (2019) "Significant Cyber Incidents," Technology Policy Program [Online]. Available: <https://www.csis.org/programs/cybersecurity-and-governance/technology-policy-program/other-projects-cybersecurity>. [Accessed: 26-Jan-2019].
- Denning, D. E. (2014) "Framework and principles for active cyber defense," *Comput. Secur.*, vol. 40, pp. 108–113.
- Giesen, K.-G. (2013) "Towards a theory of just cyberwar," in 8th International Conference on Information Warfare and Security, ICIW 2013.
- Hathaway, O. A. R. Crootof, P. Levitz, H. Nix, A. Nowlan, and J. Spiegel (2012) "The Law of Cyber-Attack The Law of Cyber-Attack," *Calif. Law Rev.*, vol. 100, no. 4.
- E. Hutchins, M. Cloppert, and R. Amin (2011) "Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains," in Proceedings of the 6th International Conference on Information Warfare & Security, pp. 113–125
- Kehler, R. C., H. Lin, and M. Sulmeyer (2017) "Rules of engagement for cyberspace operations: A view from the USA," *J. Cybersecurity*, vol. 3, no. 1, pp. 69–80.
- Lambert, N. A. (2014) "The Strategy of Economic Warfare: A Historical Case Study and Possible Analogy to Contemporary Cyber Warfare," in Cyber Analogies, E. O. Goldman and J. Arquilla, Eds. Monterey, CA: Naval Postgraduate School, pp. 76–89.
- Liff, A. P. (2012) "Cyberwar: A New 'Absolute Weapon'? The Proliferation of Cyberwarfare Capabilities and Interstate War," *J. Strateg. Stud.*, vol. 35, no. 3, pp. 401–428.
- Lin, H. (2010) "Offensive Cyber Operations and the Use of Force," *J. Natl. Secur. Law Policy*, vol. 4, no. 1, pp. 63–86.
- Makarenko, S. I. (2016) "Информационное оружие в технической сфере: терминология, классификация, примеры [Information weapons in the technical sphere: terminology, classification, examples]". *Systems of Control, Communication and Security* #3.
- Makarenko, S. I. (2017) "Информационное противоборство и радиоэлектронная борьба в сетевоцентрических войнах начала XXI века [Information warfare and electronic warfare to network-centric wars of the early XXI century]".
- Mackenzie, P. J. (2017) "NATO Joint Air Power and Offensive Cyber Operations," Kalkar, Germany.
- McGraw, G. (2017). Silver Bullet Talks with Marie Moe. *IEEE Security & Privacy*, (1), 8-11.
- MOD RF - Ministry of Defence of the Russian Federation (2011) "Концептуальные взгляды на деятельность Вооруженных Сил Российской Федерации в информационном пространстве [Conceptual Views Regarding the Activities of the Armed Forces of the Russian Federation in the Information Space]," Moscow, 2011.
- Nato NSA (2013) "AAP-6: NATO Glossary of Terms and Definitions (English and French)," NATO Standardization Agency, Brussels.
- Page, C. (2017) "'NotPetya' ransomware hits '2,000 organisations' in WannaCry-style global outbreak," *Computers*, London.
- Schmitt, M. N. Ed. (2013) *Tallinn Manual on the International Law Applicable to Cyber Warfare*. Cambridge, UK: Cambridge University Press.
- Smeets, M. (2018) "The Strategic Promise of Offensive Cyber Operations," *Strateg. Stud. Q.*, vol. 12, no. 3, pp. 90–113, 2018
- Thomas, T. (2009) "Nation-state Cyber Strategies: Examples from China and Russia," in *Cyber Power and National Security*, F. D. Kramer, S. H. Starr, and L. K. Wentz, Eds. National Defense University Press, 2009.
- UK MOD DCDC (2014) "JDP 0-01 UK Defence Doctrine 5th Edition," United Kingdom Ministry of Defence, Shrivenham.
- UK MOD DCDC (2016) "Cyber Primer 2nd edition," United Kingdom Ministry of Defence, Shrivenham.
- USCYBERCOM - United States Cyber Command (2018) "Cyberspace operations," Director for Global Operations, Fort Meade, MD.
- US JCS J7 – US Joint Chiefs of Staff, J7 – Directorate for Joint Force Development (2015) "JP 6-0 Joint Communication System".
- Wake, D. (2018) "NATO Tests Electronic Defenses as Cyber Warfare Threat Grows," Agence France Presse.
- Yasar, N., F. M. Yasar, and Y. Topcu (2012) "Operational advantages of using Cyber Electronic Warfare (CEW) in the battlefield," in Proceedings of SPIE vol 8408: Cyber Sensing 2012.

# **Building a (French) National Priority: The Grammar of the French Cyber Defence Strategy**

**Saïd Haddad**

**CREC Saint-Cyr, Saint-Cyr Military Academy, Guer, France**

[said.haddad@st-cyr.terre-net.defense.gouv.fr](mailto:said.haddad@st-cyr.terre-net.defense.gouv.fr)

**Abstract:** Presented on February 2018, the French Strategic Review on Cyber Defence can be considered as the first French White Paper dedicated only to cyber defence. This Review reaffirms and strengthens the French position on cyberdefence, already defined in several guiding and high-policy documents published since the White paper on Defence and National Security of 2008. Considered as a “seminal document”, the French Strategic Review on Cyber Defence is an important step of the ongoing hyper-securitization process which started with the 2008 White paper. Through the analysis of 7 key documents published between 2008 and 2018 -France’s Cyber Strategy in 2011; White papers on Defence and National Security of 2008 and 2013; the Cyber Defence Pact in 2014, the French National Security Strategy announced in 2015, the Strategic Review of Defence and National Security (October 2017) and the Strategic Review of Cyber defence in February 2018-, this paper is devoted to analyse how cyberdefence became a French priority via security concerns. By adopting the frame of “the specific grammar of the cyber security sector” as developed by Lene Hansen and Helen Nissenbaum (2009), the (social) construction of the French National Cyber Strategy will also be discussed. Finally, the Strategic Review on Cyber Defense is also a seminal document “as it is the first time that the various approaches and strategies of the different components of the State are gathered and articulated into one national strategy”. On one hand, it completes a process that has begun ten years before and on the other hand it is a turning point of the French cyberdefence strategy with the announcement made by the Defence Minister on January 18<sup>th</sup> 2019 that the French military plans to develop and deploy offensive cyber weapons and improve the protection of its networks from “security events”.

**Keywords:** French cyberdefence, cybersecurity, securitization-,cyberstrategy, cyber community, speech act

---

## **1. Introduction**

Presented on February 2018, the French Strategic Review on Cyber Defence can be considered as the first White Paper dedicated to cyber defence. This Review reaffirms and strengthens the French position on cyberdefence, already defined in several guiding and high-policy documents published between 2008 and 2018. So, in 2013, Cyber defence security has been defined as a national priority by the French White Paper on Defence and National Security. One year later the French MOD launched the Cyber Defence Pact. This Pact is dedicated to counter cyber-attacks and cyber-threats by allowing to what is called the defence community to “explore, invest and take control of this new strategic field”. By defining a strategy, the Cyber Defence Pact tries to take up the challenge of these new threats and also academic and economic and industrial opportunities. By adopting the frame of “the specific grammar of the cyber security sector” as developed by Hansen and Nissenbaum (2009) we will also discuss the (social) construction of the French National Cyber Strategy. Through the analysis of 7 key documents published between 2008 and 2018, this paper is devoted to analyse how cyberdefence became a French priority via security concerns.

## **2. The grammar of cyber security: Cybersecurity as a speech act**

This specific grammar is partly based on Copenhagen school’s security approach (Buzan, Waever and Wilde, 1998). The speech or discourse on the cyber threat results from the enlargement of the fields of security discourse. Cyber security has been identified “as a particular sector on the broader terrain of Security studies” and consequently is defined by “particular constitutions of referent objects and types of threats as well as by specific forms or “grammars” of securitization” (Hansen and Nissenbaum, 2009).

Security has been rethought in response to geopolitical change, technological innovations and societal evolutions. The cyber threat is considered consequently as a security stake which means that an “issue is presented as existential threat, requiring emergency measures and justifying actions outside the normal bound of political procedure” (Buzan, Waever and Wilde, 1998). In other words, the “enunciation of security itself creates a new social order wherein ‘normal politics’ is bracketed (Balzacq, 2005). As Balzacq underlines it, the two constitutive rules required for a successful securitization are pertaining in fact to the linguistic competence of the actors involved. These are the “internal, linguistic-grammatical to follow the rule of act” and the “external, contextual and social- to hold a position from which the act can be made” (Buzan, Waever and Wilde, 1998). The securitization concept, or “the exact definition and criteria of securitization is constituted by the

intersubjective establishment of an existential threat with a saliency sufficient to have substantial political effects" (*ibid*: 25). As mentioned by Ole Waever: "with the help of the language theory (Austin, 1970; Searle, 1970, 1982), we can regard "security" as a speech act. In this usage, security is not of interest as a sign that refers to something more real; the utterance *itself* is the act. By saying it, something is done" (Waever, 1995).

Cyber security speech can be seen as "computer security" plus "securitization" (Hansen and Nissenbaum, 2009): a link has to be made between the technical discourse and the securitizing discourse.

The two authors state 'that "network security" and "individual security" are significant referent objects, but that their political importance arises from connections to the collective referent objects of "the state," "society," "the nation," and "the economy." These referent objects are articulated as threatened through a grammar of cybersecurity'.

According to Hansen and Nissenbaum (*ibid*), there are three security modalities that are specific to the cyber sector: hypersecuritization, the everyday security practices and technification. Hypersecuritization has been introduced by Buzan and "describes an expansion of securitization beyond a normal level of threat and dangers". It's a "tendency both to exaggerate threats and to resort to excessive countermeasures". The speech here relies on an element of hypothetical and an if-then logic: "the power of hypersecuritization stems not only from a securitization of the network itself but from how a managed network could cause societal, financial and military break-down hence bringing in all other referent objects and sectors" (Hansen, and Nissenbaum, 2009). Securitization mobilises also the "past as a legitimate reference" (the discursive dimension of memory): making historical analogy, comparing some events to provoke emotion among the audience with formulas such as "cyber Pearl Harbor" (Bumiller and Shanker 2012). NOT IN THE REFERENCE LIST.

The second grammar is based on the mobilization of individuals by "linking elements of the disaster scenario to experiences familiar to everyday life" and "the acceptance of public security discourse may be facilitated by a resonance with an audience's lived, concrete experiences". Interpellation, awareness/accountability and potential threatening of individuals constitute the core process of the mobilization.

The third grammar is technification which is based on the "creation of a particular space for technical, expert discourse" (Hansen and Nissenbaum, 2009). Whether in the written media or in institutional discourse, expert discourse is critical in that it substantiates the claims made by the different players. It appears also as politically and normatively neutral. Experts play an active role both in securitization and in the everyday security practice

### **3. The French cyber policy: securitizing the digital republic**

Cybersecurity and cyber defence have emerged as a national security priority in seven guiding and high-level policy documents since 2008: the White paper on Defence and National Security of 2008; France's Cyber Strategy in 2011; White paper on Defence and National Security of 2013; the Cyber Defence Pact in 2014, the French National Security Strategy announced one year ago in October 16<sup>th</sup> 2016, the Strategic Review of Defence and National Security (October 2017) and the Strategic Review of Cyber defence in February 2018 (Secrétariat Général de la Défense et la Sécurité nationale, 2018). NOT IN REFERENCE LIST. Cybersecurity is defined as the "desired state of an information system in which it can resist events from cyberspace likely to compromise the availability, integrity or confidentiality of the data stored, processed or transmitted and of the related services that these systems offer or make accessible. Cybersecurity makes use of information systems security techniques and is based on fighting cybercrime and establishing cyberdefence" while cyberdefence is "a set of all technical and non-technical measures allowing a State to defend in cyberspace information systems that it considers to be critical (ANSSI, 2011). AS IT 11 OR 12?

### **4. The White paper on National Defence and Security of 2008**

The White paper on National Defence and Security (WPDNS) of 2008 identifies cyber as one of the "new vulnerabilities for Europe's territory and its citizens" (The French White Paper on Defense and National Security, 2008). NOT IN REFERENCE LIST. Among a hierarchy of risks and threats on French soil, cyber-attacks ranked at the third position, after terrorism and ballistic threats. Despite the fact that the word cyberterrorism is not quoted, the White Paper underlines that "the current daily level of cyber-attacks, whether from State sources or not, points to a very high potential for the destabilisation of everyday life, paralysis of critical networks for the life of the nation, or denial of access to certain military capabilities. Society and government are still ill-

prepared for the risks of massive attacks, and these should therefore be the subject of fresh attention, both in term of strengthening defences and enhancing our capacity to hit back" (*ibid*). The White Paper also outlines the many form of the threat (malevolent blocking, physical destruction, neutralisation of computer systems etc.) and the plurality of attacks emanating from non-State actors (computer pirates, activists or criminal organisations) and States in the cyberspace. "The transition from a passive defensive strategy to an active defensive strategy in depth, combining intrinsic systems protection with permanent surveillance, rapid response and offensive action, calls for a strong governmental impetus and a change in mentalities" is underlined and a number of actions that should be undertaken to "acquire active, in depth cyber-defence capability, combining the intrinsic protection of systems, constant monitoring of critical networks and a rapid response in the event of attack" (*ibid*).

The utterance of threats defines first the referent objects (national territory, society, energy, transport or food producers etc.) and then the referent subjects i.e. the actors (here the state and non-state actors) who are threatening the former ones. The WPDNS is also performative by describing the measures to counter cyberthreats. It's also the first high-level policy document to place "cyberspace in a context of national security" (Brangetto, 2015).

## **5. Defining a cyberstrategy**

The *Information Systems Defence and Security: France's Strategy guide* has been published in 2011 by the National Network and Information Security Agency (ANSSI, Agence Nationale de la Sécurité des Systèmes d'Information) established in 2009 under the authority of the Prime Minister and the Secretary General for Defence and National Security (SGDSN). As mentioned in the 2008 WPDNS this agency is dedicated to "reinforce the coherence and capacity of State resources [...], operate a centralised capability to detect and defend against cyber-attacks [and] take on an advisory role to the private sector, particularly in areas of critical strategic importance, and will participate actively in the development of security for the information society". In 2011, was also established a cyberdefence general officer (labelled as OG Cyber) who serves as the head of the French cyber operational command. The French cyber operational command has the following tasks: coordinate cyber defence efforts within the ministry of defence; and plan and command cyber operations within the *Planning and Operations Centre* located at the French Joint Staff. These operations consist mainly but not only of defensive cyber operations. OG cyber has also a functional responsibility to all the bodies of the Ministry regarding military cyber defence (Brangetto, 2015)

The 2011 document describes four main objectives and seven areas of action. The four objectives of the strategy are:

- To "become a world power in cyber defence"
- To "safeguard France's ability to make decision through the protection of information related to its sovereignty".
- To "strengthen the cybersecurity of critical national infrastructures"
- To "ensure security in cyberspace". "In other words, non-critical public service providers, the private sector, and citizens should be able to operate in a reasonably secure cyber space" (Vittel and Bliddal, 2015).

Becoming, safeguarding, strengthening and ensuring are illocutionary acts belonging to the commissives ) category –the act performed in articulating a locution (Balzacq, 2005). Commissives are speech acts that commits a speaker to some future actions.

*"In order to meet these objectives, seven areas of action have been identified:*

- 1. Effectively anticipate and analyse the environment in order to make appropriate decisions.
- 2. Detect and block attacks, alert and support potential victims.
- 3. Enhance and perpetuate our scientific, technical, industrial and human capabilities in order to maintain our independence.
- 4. Protect the information systems of the State and the operators of critical infrastructures to ensure better national resilience.
- 5. Adapt French legislation to incorporate technological developments and new practices.

- 6. Develop international collaboration initiatives in the areas of information systems security, cyberdefence and fight against cybercrime in order to better protect national information systems.
- 7. Communicate, inform and convince to increase the understanding by the French population of the extent of the challenges related to information systems security." (ANSSI, 2011)

As mentioned in the foreword, "the purpose of its document is to outline the strategy undertaken by France since the publication of the French WPDNS in order to safeguard the security of our citizens, our companies and our nation in cyberspace". The referent objects defined in the four objectives are articulated with the seven areas of action (i.e. measures) which are significant responses to the concerns described in the 2008 White paper.

## **6. Strengthening cybersecurity**

Two years later, the focus on cybersecurity was strengthened in the new WPDNS of 2013. While highlighting the three priorities of the French defence strategy (i.e. protection, deterrence, and intervention), the WP underlines the guarantee of protecting French citizens "including cyberthreats" (introduction by the French President). It also underlines that "as identified in the previous White paper, the threats and risks posed by universal access to cyberspace have been confirmed" and [...] that the "rapid development of digital infrastructure has not always been accompanied by a correspond effort to protect it." As an area of confrontation, France has to be protected from cyber-attacks. Cyber-attacks which "do not have the same impact as terrorist acts; given that they have not to date resulted in any fatalities. However, today and even more over the timeframe of this White Paper, they represent a major risk given their high probability and potential impact" (WPDNS, 2013). The State, operators of vital importance (as defined as economic operators having a crucial importance in the functioning of the Nation) and large national or strategic companies quoted as targets are here the referent objects. Priority is also "given to development of military cyber-defence capabilities" and the country "will develop an approach based on a cyber-defence organisation closely integrated with the armed forces, made up of defensive and offensive capacities to prepare or support military operations" (WPDNS, 2013). The fight against cyber-threats must be strengthened as the "the continued growth of this threat, the continuing increase in the importance of information systems in the life of our societies and the very rapid development of technologies, require us to move onto yet another level to maintain the protection and defence capabilities responding to these changes" (ibid.). The WPDNS insists on the "substantial increase in the level of security and the means to defend [the French] information system" in order to preserve the country's sovereignty and defend the economy as well employment. The WPDNS states also that "the capacity to detect and protect ourselves against cyber-attacks and to identify those responsible for them has become an element of national sovereignty. To succeed in this endeavour, the state must support high-level scientific and technological expertise". To respond to cyber-attacks or major IT attacks, the principle of global approach" is the base of the national policy.

The 2013 White Paper broadly follows the framework of the former one (uttering the threats, the referent objects, the referent subjects and the measures to adopt), one must point out that this document deepens the orientations outlined in 2008. For instance, military cyberdefence capacities in close liaison with intelligence activities have to be developed: "France will develop an approach based on a cyber-defence organisation closely integrated with the armed forces, made up of defensive and offensive capacities to prepare or support military operations" (WPDNS, 2013)

## **7. Building a national cyberdefence community**

In February 7<sup>th</sup>, 2014 the minister of Defence presented a Cyber Defence Pact (CDP), another guiding document which emphasises the aims developed (or presented) in the White Paper of 2013 with a strong focus on human resources, training, and cooperation between all the French actors of the cyberdefence. The Pact revolves around six axes and 50 measures.

In the preamble, the minister of Defence claims that this strategy implies actions within the Ministry but also, large industrial groups, small and medium sized enterprises, and finally education and research institutions and training operators. Furthering the emergence of a national cyberdefence community is the 6<sup>th</sup> axis of the pact.

The six axes are:

- “1. Reinforcing the security level of the information systems as well as the defence and intervention assets of the ministry and its major trusted partners;
- 2. Preparing the future through an intensification of the research efforts in the technical, academic and operational domains, while supporting our industrial basis; [for instance, a chair of cyber defence has been inaugurated in Saint Cyr and also a Master of Science in Cyber Crisis Management has been launched in autumn 2015 and other trainings will be opened in the other military training schools;]
- 3. Reinforcing the manpower dedicated to cyber defence and developing the associated career paths;
- 4. Developing the cyber defence centre in Brittany for the Ministry of Defence and the national cyber defence community;
- 5. Keeping up a network of foreign partners, in Europe, within the Atlantic Alliance or in areas of strategic interest;
- 6. Furthering the emergence of a national defence community, relying on group of partners and on the reserve networks”. (Ministry of Defence, 2014).
- The use of performative utterances such as reinforcing, preparing, developing, and keeping up underlines the strong commitment of strengthening the cyberdefence capabilities and raising a strong cyberdefence community.

## **8. Updating the French national digital security strategy**

In 2015, the French National Digital Security Strategy document states in the introduction that “France is going through its digital transition “and that digital technology is “therefore a factor of innovation” (ANSSI, 2015). But as mentioned by the Prime Minister in his foreword “the digital universe is also a locale for competition and confrontation. Cyberspace has become a new domain for unfair competition and espionage, disinformation and propaganda, terrorism and criminality » while the digitalisation of French society is accelerating and increasing. This updated guide emphasizes the former by uttering the threats, the referent objects, the referent subjects and the measures to adopt. Five objectives are defined:

- 1. Fundamental interests, defence and security of State information systems and critical infrastructures, major cybersecurity crisis;
- 2. Digital trust, privacy, personal data, cyber malevolence;
- 3. Awareness raising, initial training, continuing education;
- 4. Environment of digital technology businesses, industrial policy, export and internationalisation;
- 5. Europe, digital strategic autonomy, cyberspace stability.

Underlining the dangerous nature of cyberspace, taking into account the economic dimension of the “digital transition” and establishing a national community are the main aims of the updated strategy.

Indeed, defending the French fundamental interests in cyberspace has to be done in a context of cyber-attacks that damaged France’s fundamental interests. The document states that “the position taken by France on the international scene, its military operation and certain public debates are followed by cyberattacks aimed at marking public opinion” as “the defacement of many web sites after the terrorist attacks that targeted France in the beginning of 2015” the most important being the French global TV5 network one? (TV5 TV1?). “Similarly, the cyberattack that led to the interruption of an international French media was also aimed at making a strong impression and contributes to the radicalisation that leads to terrorist acts. This attack also demonstrated the capacity of attackers determined to disrupt the functioning of a highly symbolic infrastructure” (ANSSI 2015).

Defending the fundamental interests in cyberspace goes along with the promotion and the development of the digital economy, digital security being a “factor in competitiveness”. So, “it’s up to the private sector to ensure its own security in the field of information technology as in the case in other fields” and the detection and treatment of the inevitable growth in the number of cyberattacks that businesses are subjected to”. And this security will be enabled by “the transfer of this knowledge acquired by the State services”. Interpellation, awareness/accountability and potential and rising threatening of private sector constitute the core process of the mobilization of the French economic actors.

The national community is composed by three communities of stakeholders who are responsible “of the stability of our future supported by the digital technology”. The main stakeholders in the first community are “researchers, product and service inventors and integrators, cybersecurity businesses, electronic communications network operators, internet service providers and remote data processing services ». The second community “consists of elected officials, the Government, central and territorial administrations and trade unions”. The third community “consists of all users, companies’ managers, participants in civil society and citizens”.

Each community is related to the two others (“there are synallagmatic commitments made by each stakeholders”, a synallagmatic contract being in civil law systems a contract that creates mutual obligations) and contribute to establish a national community who will defend not only French interests in cyberspace, but also the “Republican” values. Defending the “digital Republic” is a part of the defence or preservation of the French Republic.

## **9. The strategic review of cyber defence**

Published on October 13<sup>th</sup> 2017, the Defence and national Strategic Review sets the strategic framework for the development of the next Military Programming Act 2019-2025, which will raise the French defense effort to 2% of GDP by 2025. The Review draws lessons of the unstable and unpredictable strategic context since the 2013 White Paper was released. The Review updates the 2013 White paper by taking into account the deterioration of France’s geopolitical environment. Cyberspace is considered as “domain of confrontation” (Strategic Review, 2017) and is “the subject of intense strategic competition” and “a major cyberattack may, given the damage it could cause, justify invoking legitimate defence under Article 51 of the UN Charter”. In this new form of warfare and conflict, France has to preserve its autonomy, “condition to France’s credibility in the eyes of allies and partners”. France has decided to adopt a permanent cybersecurity posture by strengthening its defence means and developing its offensive and defensive capabilities (*ibid.*)

In the meantime, a new cyber command unit (CYBERCOM) has been launched on January 1<sup>st</sup> 2017. This new unit is dedicated to increase French cyber-defence and offence capabilities. CYBERCOM The new cyber command will be under the command of the chief of staff of the armed force and will employ nearly 3,200 people belonging to the French armed forces. This new organization of cyberdefence is a result of the emergence of “a new cyber-battlefield” [which] “must make us rethink profoundly our way of approaching the art of war” (French Minister of Defence, 2016).

Portrayed by the General Secretary for Defense and National Security as “a seminal document comparable to the first French White Paper on Defense of 1972 which established France’s nuclear doctrine” (Delerue and Géry, 2018), the Strategic Review of Cyber Defence is published on February 12<sup>th</sup> 2018. The Review deepens the French position on cyberdefence and completes the social construction of cyberspace as a “domain of confrontation” and “a source of vulnerability”. Cyberspace is conceptualized here “as a problematic unruly place that needs to be tamed at all cost, then this inevitably leads to calls for strong(er) interference into the global cyber-system, including the topology of the Internet” (Dunn-Cavelty, 2013). The titles of the three parts of the Review can sum up the spirit or the philosophy of the whole review: The dangers of the Cyber world (part 1); The State, responsible of Nation’s cyber defence (part 2); The State as guarantor (responsible) of the society’s cybersecurity (part 3). This threatening and unruly place (where “international regulation still has a long way to go”) must be “tamed” by a strong commitment from the state. In this context, the cyber review “clarifies the organization of French cyber defence by formalising it around four operational chains, strengthen its governance mechanisms and defines an operational process for cyber crisis management” (Strategic Review of Cyber Defence, 2018). In the third part of the review, digital sovereignty is considered as an essential part of national sovereignty. The two sovereignties are one and the same. For achieving this goal, the state “intervenes in this area as a prescriber, reformer and provider of security solutions” but needs the involvement of all sectoral players, already mentioned in the 2015 French National Digital Security Strategy. The consolidation of a national cyber industrial base (Strategic Review of Cyber Defence, 2018) goes along with the promotion of a culture of digital security in French society by “raising awareness of the digital risk through digital education”. There’s no (digital) sovereignty without the awareness and the commitment of the three communities described in the 2015 Strategy guide.

## **10. Conclusion**

The Cyberdefence Strategic Review “is also a seminal document as it is the first time that the various approaches and strategies of the different components of the State – including the Ministry of Foreign Affairs, Ministry of Defense, ANSSI, etc. – are gathered and articulated into one national strategy” (Delerue and Géry, 2018). This Review is an important step of the ongoing hypersecurization process which started with the 2008 WPDNS. The implementation of a cyber defence strategy in France has been deepened on January 18<sup>th</sup> 2019 with the announcement by the Defence Minister that the French military plans to develop and deploy offensive cyber weapons and improve the protection of its networks from “security events,”. Following the Defence Minister speech, the chief of staff of the French armed forces stated that the chief of cyber command “would craft a new offensive cyberwarfare doctrine meant to “ensure the defense of our interests and the preservation of our sovereignty.” (MacKenzie, 2019 ; Minister of Defence, 2019)

## **References**

- ANSSI (Agence nationale de la sécurité des systèmes d'information- National Network and Information Security Agency) (2015), *French National Digital Security Strategy*, Secrétariat Général de la D2fense et de la Sécurité nationale, Paris, <https://www.ssi.gouv.fr/en/actualite/the-french-national-digital-security-strategy-meeting-the-security-challenges-of-the-digital-world/>
- ANSSI (2011), *Information systems defence and security. France's Strategy*, Premier Ministre, 2011.
- Austin J. (1970), *Quand dire, c'est faire*, Le Seuil, Paris.
- Balzacq, T. (2005), “The three faces of Securitization: Political Agency, Audience and Context”, *European Journal of International Relations*, Vol. 11 (2), pp. 171-201.
- Brangetto, P. (2015), *National Cyber Security Organization: France*, NATO CCDECOE, Tallinn.
- Bumiller, E. and Shanker (2012), “Panetta warns of Dire Threat of Cyberattacks on U.S.”, *nytimes.com*, October 11<sup>th</sup>, <https://www.nytimes.com/2012/10/12/world/panetta-warns-of-dire-threat-of-cyberattack.html>
- Buzan, B., Waever, O., de Wilde, J. (1998), *Security: A New Framework for Analysis*, Boulder: Lunne Rienner Publishers, 1998.
- Delerue, F. and Géry, A. (2018), “ The French Strategic Review of Cyber Defence”, ISPI, May 2, <https://www.ispionline.it/it/pubblicazione/french-strategic-review-cyber-defense-20376>
- Dunn-Cavelty, M. (2013), “From Cyber-Bombs to Political Fallout: Threat Representations with an Impact in the Cyber-Security Discourse”, *International Studies Review*, 15, p. 105-122.
- Hansen, L. and Nissenbaum, H. (2009), “Digital Disaster, Cyber Security and the Copenhagen School”, *International Studies Quarterly*, 53, pp. 1155-1173.
- MacKenzie, C. (2019), “ French defense chief touts offensive tack in new cyber strategy”, *fifthdomain.com*, January 18<sup>th</sup> 2019, <https://www.fifthdomain.com/global/europe/2019/01/18/french-defense-chief-touts-offensive-tack-in-new-cyber-strategy/>
- Minister of Defence (2019), Discours de Florence Parly, ministre des Armées, Stratégie cyber des Armées,, January 18th 2019 : <https://www.defense.gouv.fr/salle-de-presse/discours/discours-de-florence-parly/discours-de-florence-parly-ministre-des-armees-strategie-cyber-des-armees>
- Minister of Defence (2016), *Statement of the French Minister of Defence*, December 12<sup>th</sup> 2016: <http://discours.vie-publique.fr/notices/163003632.html>
- Ministère de la Défense (2014), *Pacte Cyber Défense. 50 mesures pour changer d'échelle*, <https://www.defense.gouv.fr/actualites/articles/presentation-du-pacte-defense-cyber>
- Présidence de la République (2008), *White Paper on Defense and National Security*.
- Searle, J. R. (1970), *Les actes de langage, essai de philosophie du langage*, Herman, Paris.
- Searle, J. R. (1982), *Sens et expression*, Minuit, Paris.
- Secrétariat Général de la Défense et la Sécurité nationale (2018), Strategic Review of Cyber Defence, Paris, [www.sgdsn.gouv.fr/uploads/.../revue-cyber-resume-in-english.pdf](http://www.sgdsn.gouv.fr/uploads/.../revue-cyber-resume-in-english.pdf)
- Secrétariat Général de la Défense et la Sécurité nationale (2017), *Defence and national Strategic Review*, Paris, <https://www.defense.gouv.fr/english/dgris/defence-policy/revue-strategique/revue-strategique>.
- The French With Paper on Defense and National Security* (2008), Odile Jacob Publishing, New York
- Vittel, P. and Bliddel, H. (2015), “French Cyber Security and Defence: An overview”, *Information and Security: An international Journal*, vol.32, pp 29-41.
- Waever O. (1995), “Securitization and Desecuritization”, in Lipschutz R.D. (ed.), *On Security*, New York, Columbia University Press, pp. 46-87
- White Paper on Defense and National Security*, (2013), Présidence de la République <https://www.defense.gouv.fr/english/dgris/defence-policy/white-paper-2013/white-paper-2013>

# Where Cyber Meets the Electromagnetic Spectrum

**Gerhard Henselmann and Martti Lehto**

**University of Jyväskylä, Finland**

[office@ghenselmann.de](mailto:office@ghenselmann.de)

[martti.j.lehto@jyu.fi](mailto:martti.j.lehto@jyu.fi)

**Abstract:** Cyber linked with Information Technology, computers and the internet are the most commonly understood potential threats that everybody is aware of nowadays. In the case of National Cyber Strategy most efforts are given to these potential battlefields and threats. But is it enough to stay with these paradigms? Cyber warfare will look for utmost effects for chaos and misalignment of economy and therefore we cannot exclude the facts, that there is a potential threat also possible in the cut volumes between the information technology, the electromagnetic spectrum and its linked infrastructure and processes. We understand the technology for mobile telecommunications and data transfer, as the providers and linked business models are transparent and an open source and therefore a potential threat in the sense of cyber-attacks. All these services are based on the message transfer in the Electromagnetic Spectrum (EMS). It will be more challenging to enter in secured networks and modern data communication protected by encrypted and secured data links, hardened equipment and architecture. But as those systems, like the avionics and communication subsystems in airborne vehicles and commercial aircrafts are highly interlinked and interoperable based on data networks, it will be necessary to understand the architecture and weaknesses in detail. Of course, these principles are not fixed only to the airborne platforms and relevant applications; you may find the same technical pandemics in data transmission and data engineering in the physics of any data transfer chain, like satellite communication, radio transmission links. The paper will discuss the technical background in principle and identify potential weaknesses in aviation electromagnetic spectrum. We propose measures for attention and the implementation of relevant procedures and quality in order to overcome bottlenecks and grey area dull which helps the opponents to weaken society and open a sideshow for conflicts introduced by “simple cyber threat handling”.

**Keywords:** electromagnetic spectrum, aeronautics, avionics, electronic warfare, cyber warfare

---

## 1. Introduction

### 1.1 Background

Research in the interface between the electromagnetic spectrum and the electronics engineering in radio transmission links and data link applications have shown areas where unfriendly third parties, eventually called opponents can make use of open source knowledge and enter networks for hacking, influencing and destroying them, if they are able to lock into the protocol. Of course, those areas of threat potential might be considered by “critical infrastructure” analysis and subsequently covered in protection and work-arounds for measures. The consequence of getting access into this computer-based infrastructure will cause loss of control and this fear is real in nowadays Internet of Things (IoT) applications with low protection devices. Nowadays Artificial Intelligence (AI) based autonomous systems are becoming part of business and industry. Examples can be found from road transportation, maritime, industrial processes as well as from aviation, where automation has made flying safer than ever before.

So, we are entering in a new era. We are in the transition of the information technology cycle entering into the digital age, where our activities, skills and performances are based very much on computer-based support and the internet, even if we are aware on the cited reality. The Internet is among the few things humans have built that they don't truly understand. (Schmidt and Cohen, 2014)

Regardless of the question on how the digital revolution expresses itself in core economic data and whether its benefits are even reflected in national accounts, from a technical point of view the process of technological transformation has so far expanded exponentially. This is not justified only by the increase in computational, storage and communication capacities, but by the ability to digitally intersect technological areas where information can be created, stored, accessed, processed and shared. The combination of embedded software systems for sensor-based monitoring and control of physical reality with global digital network infrastructures – the cyberspace. It allows a variety of applications and problem solutions with high economic potential and strong innovative power. On the supply side, more and more opportunities for using and linking data are created, enabling new business models. Previous media and technology breaches and related activities of data collection and transformation are eliminated. It is believed that in 2002, for the first time, it was possible to store more

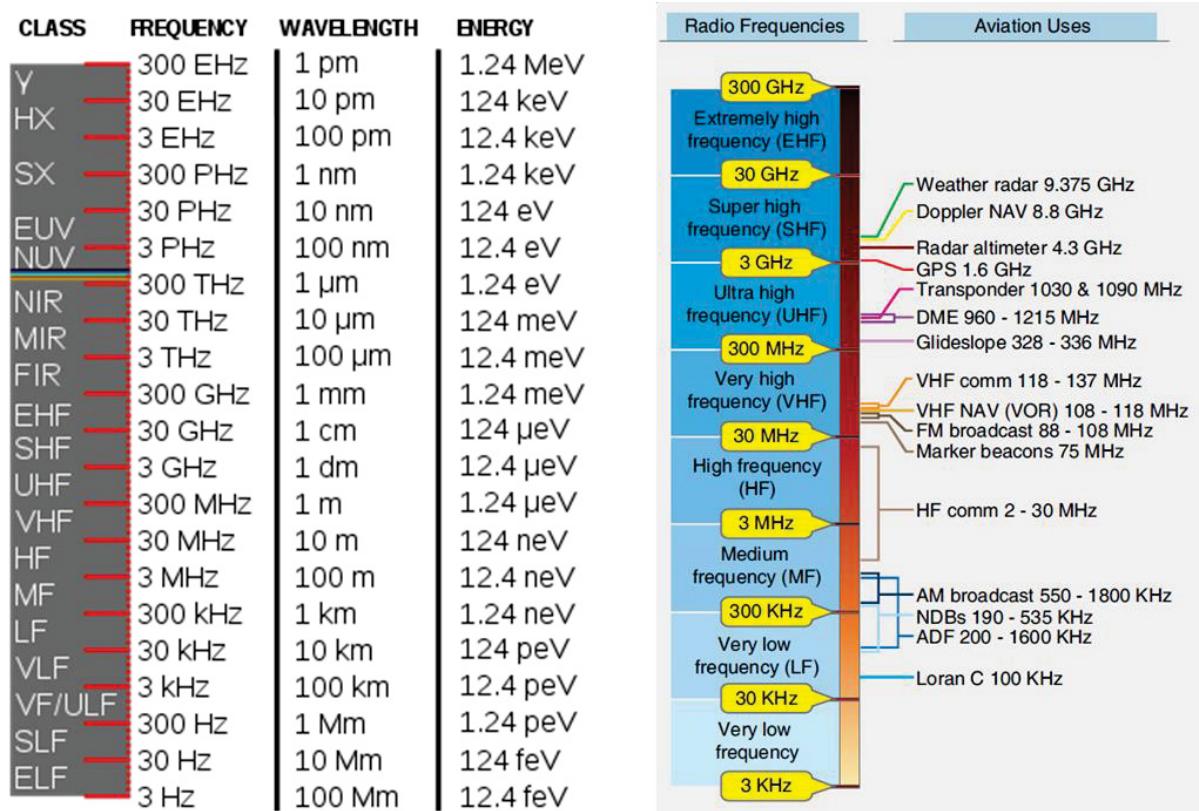
information digitally than in analogue format - a clue to the beginning of the digital age. (Hilbert and López, 2011)

The consequence of getting access into the computer-based infrastructure will cause loss of control and this fear is real in nowadays IoT applications with low protection devices. The Internet of Things has a security problem, it is here - in future everywhere, but it is not secure (Hofmann, 2017). Consequently, the result of those interactions might be "fake news" for the public, misleading the society eventually into chaos and/or for operators behind the consoles bad and misleading starting point for their decision making.

## 1.2 Electromagnetic spectrum

The electromagnetic spectrum is a broad area of activity characterized by physically observable activities such as visible light and lasers and unobservable phenomena such as microwaves and electromagnetic energy. EMS manifests through various frequencies and wavelengths produced by natural sources like solar storms or artificially by hardware such as radar or nuclear weapons. (Stuckenbergs et al. 2018)

As most of the applications are not only connected to the world of data transfer by the internet but also via data link, VoIP-phone lines and the means of the EMS. The electromagnetic spectrum is the range of frequencies (the spectrum) of electromagnetic radiation and their respective wavelengths and photon energies. The EMS is involved everywhere e.g. with datalinks, with sensors, with data transfer on the WLAN and the established VoIP. (see Figure.1)



**Figure 1:** The electromagnetic spectrum and radio waves in communication and navigation<sup>1</sup>

And moreover, thinking a step beyond, the same physics and potential risk area is embedded in the communication and data link in aerospace and commercial flights. Aircraft but also automotive have been certified according to specific release protocols and a fixed set of configurations. Whenever a change to this configuration will apply, it needs to be checked against the certification and safety and airworthiness need to be assured by relevant testing and qualification and verification processes. Any change of take-over by

<sup>1</sup> [https://en.wikipedia.org/wiki/File:Light\\_spectrum.svg](https://en.wikipedia.org/wiki/File:Light_spectrum.svg)  
<http://www.flight-mechanic.com/radio-communication-radio-waves/>

configuration items based on a cyber-attack will endanger life and certainly will limit or even expire the authorization releases for operation.

In the beginning of the research, the focus was more on the hardware driven factors based on “kill switch” or built-in components without a clear configuration status, which were already described in research papers and speeches (Henselmann, 2017b). But searching more in the interface between the EMS and the computer-based networks, as there are avionics bus structures (AFDX, ARINC429, ARINC629), Milbus system in the (military) airplanes or the central information system in the automotive, the problem of spoofing and taking over the controls became clearer and is identified as a real threat, also described in the threat analysis assessment on weapon systems. (Koch and Golling, 2016)

The scheme for entering via the electromagnetic spectrum could be a combination of jamming the original messages in the frame and constellation of the message code and the intrusion of fake information messages via the EMS. But there could be expected alternative procedures and tools too. Cyber-attacks are responsible for delivering information on global end to end capabilities including the provision of SatCom, electromagnetic spectrum and networks to support the business model such as; booking system, Air Traffic Control (ATC) information, maintenance, repair and overhaul (MRO), monitoring systems and military operations.

In countering EMS challenges, some windows of opportunity needed to compete with our adversaries are closing. Meanwhile, EMS threats that have existed since the 1960s and earlier, such as nuclear-EMP and geomagnetic storms, have regained prominence. The salience of these threats has returned due to several factors, including (Stuckenbergs et al. 2018):

- Near-universal integration of electromagnetically sensitive silica-based technologies into most modern hardware,
- Adversaries' increased understanding of how to exploit critical vulnerabilities and
- The emergence of novel technologies, many of them poorly understood.

Having passed through into the third generation of information warfare, we must consider now, what a fourth generation might look like – and we need to be aware of the consequences and not forget it in the prevention and in our National Cyber Security Strategies. Where we are now is not unlike trench warfare, only in cyber space (Ryan, 2015). Where we will go next will emerge in an international landscape that is considering the implications of current capabilities on notions of just warfare, sovereignty and individual freedoms.

## **2. Description of the research and the objectives**

Digitization has led to the convergence of cyber and information activities to such an extent that it is important that we adapt our concepts and doctrine to the changing environment. Electromagnetic spectrum is much more than a congested resource: today it is also constrained and continuously changing and every civilian, military, intelligence, homeland security operation relies on the capability to access and exploit the EMS. Because of physical (air/land/sea/space) and virtual (cyberspace) domains of warfare increasingly depend on the access and control of the EMS, the Cyberspace Electromagnetic Activity (CEMA) will be imperative for operational success. CEMA concept integrates elements from offensive and defensive cyber warfare, electronic warfare and intelligence (Seffers, 2018)

Our literature review revealed the definition of complex infrastructure and means for the daily life and business in the IoT, but it does not cover the electromagnetic activity in cyberspace and therefore this research is focused more on EMS environment – fake information by radio station, ATC and new services (GPS navigation, news distribution), where someone has to expect manipulation to allow entry points for cyber-attack.

We therefore would like to introduce the enlargement of the traditional scope, cyber environment and propose to use the add-on version with indication of radio systems, datalink, ATC etc. on this already complex mapping. As the EMS will provide in the future a major part of cyber application and could even be the starting point for chaos and cyber warfare, it needs to be considered in more detail in this scope and for this paper.

This research is based on literary analysis and reports made in cyber defense and electronic warfare related symposiums, conferences and workshops. The objective from this research will be focused on the perimeter “how cyber warfare might be injected by the EMS on airborne platforms like commercial/ military aircraft and

cruise missiles" and what to be recommended for the National Cyber Security Strategy, especially considering legacy products in place and operation for at least the next 20 years of life.

### **3. Cyber-Electromagnetic environment**

#### **3.1 Background**

It is those unsolved problems in our life, such as; airplane accidents, which are not physically possible to get investigated because they only disappeared from the ATC radar screens or incidences were not made public. On the other side, military attacks with lethal weapons like cruise missiles delivered in the very recent past were not fully successful in delivery within the expected circular error probable (CEP), which brought up the idea to research the possible influencing factors and bring-up some hypothesizes and discuss them in a theoretical research method.

The major effort shall be focused on the interface of cyberspace and electromagnetic spectrum. Therefore, it will be necessary to first provide a definition of this interface. The cyberspace is a domain characterized using electronics and electromagnetic spectrum to store, modify and exchange data via networked systems and associated infrastructures. Cyberspace can be thought of as an interconnection between humans through computers and telecommunication without regard of physical geography.

#### **3.2 Electronic warfare**

Electronic warfare is based on the understanding, controlling and shaping of the electromagnetic spectrum and has become increasingly important to winning on the modern threat scenario everywhere and on the traditional military battlefield environments. Advanced systems, which are in a steady loop for upgrading and adaptation for the latest technologies and methodologies, will provide an improved protection for society and for the military forces by jamming, suppressing or otherwise denying an adversary the full use of the electromagnetic spectrum.

While today we are facing the transition phase from Electronic Warfare into Cyber Warfare. Best practice proposals are provided in the Australian Defence White Paper (Scott, 2013) on the application of cybernetics in cyber criminology. Following the consequences and efforts put into this cyber perimeter today, it will come up with more "self-learning" and cyber warfare based on AI (Russell and Norvig, 2003) and empathy AI based decision-making (Hagengruber, 2017).

While the Electronic Warfare already knows and improves in updating its technologies and methodologies since decades the measure and countermeasure philosophy (jamming) in the radar and communication EMS (radar signals/R-ECM, communications/C-ECM), this technology is evolving for IR signatures by DIRCM-jamming and datalink by spoofing (DeMartino, 2012). EW capabilities include directed energy, decoys, and Radio-Frequency (RF) jamming to deny, disrupt, or deceive an adversary's electromagnetic capability (Arnold, 2009)

The challenge for the EMS and cyber perimeter will be the understanding of where and how to disturb the information technology (internet, computer networks) by jamming algorithm and inject false information, subroutines into the relevant protocol during jamming. This sophisticated procedure is not immensely complicated in open source protocols as we have them available in commercial business to provide services and solutions, but it is more complex in the military application, where equipment and services are based on classified sources, protocols and requirements.

#### **3.3 Cyber warfare**

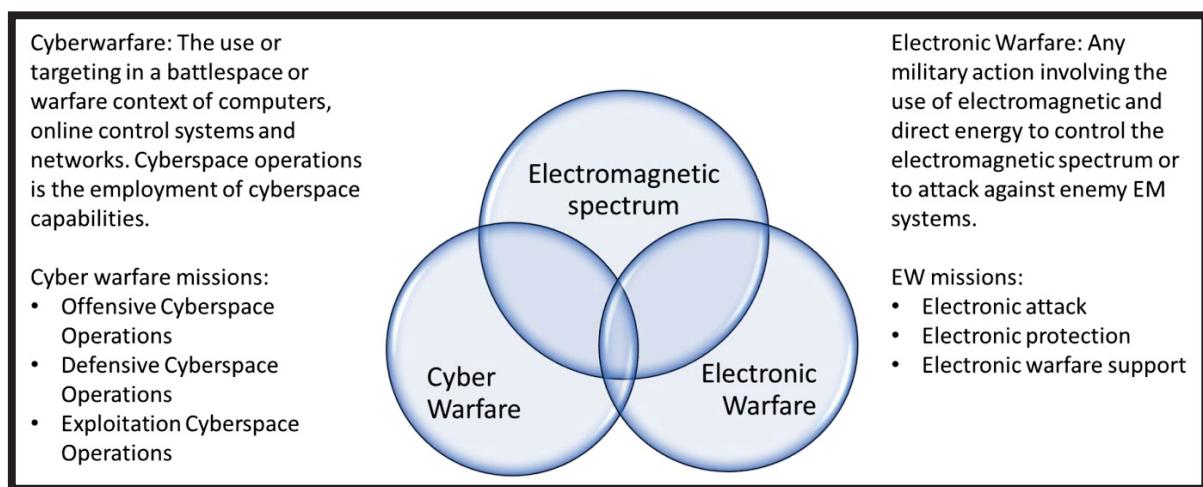
There is no generally accepted definition for Cyber Warfare (CW). For some, cyber warfare is war which is conducted in the virtual domain. But also it is the counterpart of conventional 'kinetic' warfare. EW and cyber operations need to be able to operate at the tactical, operational, and strategic levels both for offensive and defensive activities (Arnold, 2009). Cyber warfare can be divided into strategic and operational-tactical warfare, depending on the role assigned to cyber operations in the different phases of war. State actors launch offensive cyber operations in situations where the states are not at war with each other. In this case, the cyber-attacks constitute a cyber conflict in a low intensity conflict, as was the case with Estonia in 2007. (Lehto, 2015)

The cyber environment and cyberwar capabilities have created a new dimension where it is possible to act within the sovereign territory of another country, employing different military and non-military means of pressure to attain political and military goals. (Lehto, 2018)

### **3.4 Non-kinetic warfare in EMS environment**

The new capacities of armed forces create new possibilities, both the kinetic and non-kinetic use of force in cyberspace. Cyber era capabilities make possible operations in the new non-linear and indefinite hybrid cyber battlespace (Lehto, 2018). In electromagnetic spectrum can be understood to incorporate both CW and EW, thereby establishing non-kinetic warfare in EMS environment (Hay, 2016).

For this research, we defined that Cyber Warfare and Electronic Warfare form a non-kinetic warfare in EMS environment. The figure 2 illustrates that environment.



**Figure 2:** The electronic warfare and cyber warfare in EMS environment

## **4. Cyber relevant EMS interaction**

### **4.1 The principle of the cyber relevant electromagnetic spectrum interaction**

It is stated that EM is a natural physical maneuver space that is visualized through the concepts of EMS. As a natural physical space, the EM environment or also called domain cannot merge with other environments. But will it be possible to use tools and technologies in order to maneuver with the Cyber-EM environment? Today we know a huge variety of EM support systems like radars, communication devices, any kind of radar or communication jammers and GPS, which are networking utilizing the cyberspace and the cyberspace is increasingly utilizing the EM by wireless networks. The trend therefore could be called a technological sharing, and therefore the definition should incorporate that we do not speak about a converging but a sharing of both environments. With the above provided examples already existing in real life, we can easily say, that cyber systems are becoming more and more dependent on the EMS. (Arnold, 2009)

For focusing on the principle of the cyber relevant electromagnetic spectrum interaction, it is suggested to prepare the case in the aerospace environment and transfer the findings onto the other relevant areas like automotive security devices and IoT. Three things to be considered in this respect:

- Open architecture of the system to be targeted for influencing,
- Availability of a non-encrypted data link or radio communication old standard or the SW defined radio based
- Technical knowledge on the interoperability of these systems.

### **4.2 Assessment analysis**

More sophisticated approaches might have dedicated preparation packed during the design phase of a product and embedded by algorithm and sub-routines, which will be triggered from a dormant mode into a take-over mode of operation. The target for this EMS based cyber-attack either in the simple or in the more sophisticated

preparation will be the take-over of control and utilize the weakness for supporting chaos, hijacking or blackmailing authorities. In any case these objectives are possible to achieve either in a camouflaged way, while authorities will not communicate the problem openly, or it will blow-up and the safety and trust in the relevant mobility system is gone and related business area and consequently the economy will be negatively influenced. In order to achieve this, an investigation on potential risk areas on an airborne platform must be prepared and validated.

This weakness analysis is starting to define the system level and cluster for analysis. The next step based on a risk assessment matrix needs to be a Failure Mode Effect and Criticality Analysis (FMECA) which will provide in detail the area of interest, where no redundancy is available but interfacing with either the internal network – like avionics bus system, AFDX etc. and/or the interface to external support devices, called Aircraft Ground Equipment (AGE) for data exchange or mission data transfer is possible. Wherever this ground-based infrastructure is based on open-source Windows products and moreover Internet connection is available or possible, i.e. for updating or data exchange could be possible, the effort for a deeper analysis is recommended.

The assessment is a complex evaluation with deep skills and knowledge on the aerospace design world, but you may find experts worldwide consulting for any kind of ATA chapters with their support. But also, with available free-time “everybody” with some expert knowledge could be able to study some level of details with accessible information sources on the global internet by searching some studies and keywords. As an example, for this first introduction, the information provided is based on Airbus A380 training material available on Internet is utilized.

The final result of the assessment will not be displayed in the article for confidentiality reasons. But some interesting statements and exclusion-matrix for more or less interesting areas are summarized. Pre-assessment is available by research and briefing within the German Cyber Defence Community during a lecture and poster session in 2017 (Henselmann, 2017a).

It needs to be noted and it is clear without doubts that the airworthiness on aircrafts will be intensively checked-out by testing and verification on all integration level and the installation/integration checks includes EMI/EMC testing (Electromagnetic Interference/Compatibility) on a certain power and frequency level.

#### **4.3 Airborne platform**

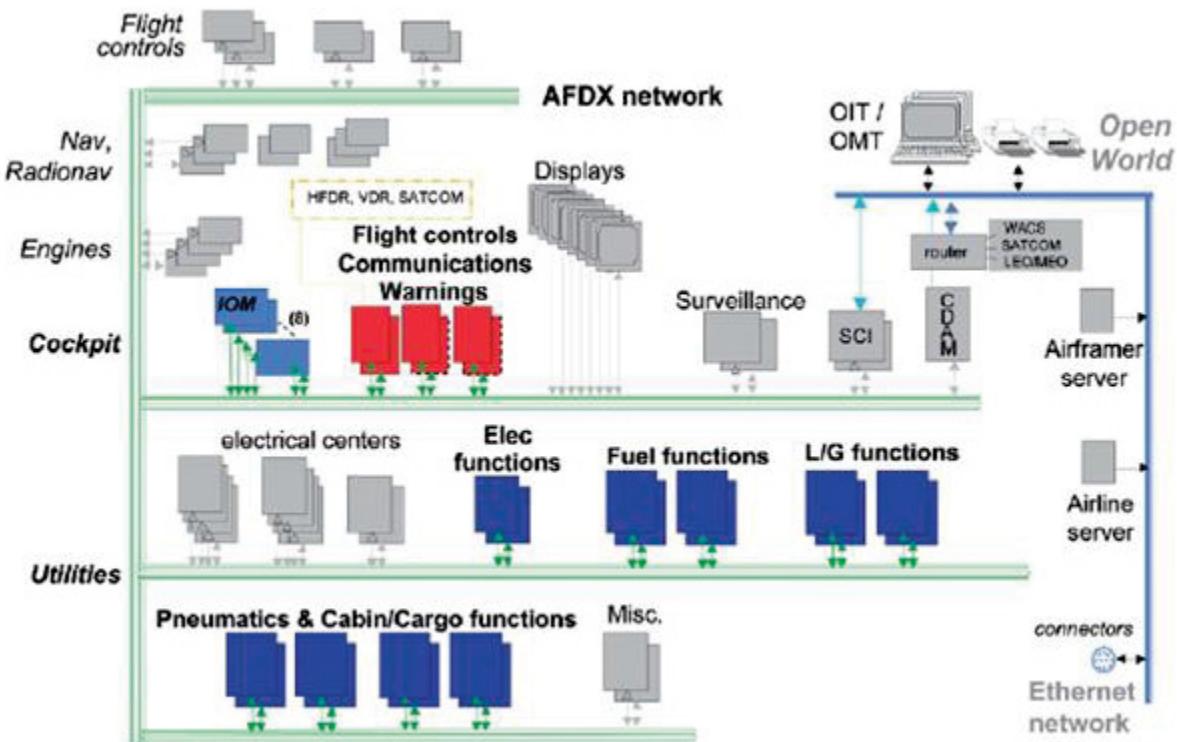
Our hypothesis is that wherever there is an interface between the airborne platform and relevant on-aircraft system, the most effective threat positioning could cause a cyber conflict or crime, by either implementing disharmonies, in case of redundant Command-Monitoring systems or overriding the original functionalities and provide fake information by malware and DDOS in a running system independent from the internet. Major points of interest could be:

- Mission data up-/download
- Routing, navigation data, Jeppesen on mobile tablets
- On-board active sensors linked with on-board bus system (radar, Traffic Alert and Collision Avoidance System (TCAS), Airborne Collision Avoidance System (ACAS))
- Up-/downlink of maintenance or engine data
- Radio communication device, Software defined radio
- Inflight entertainment system
- SW-loading station on-ground (AGE, test systems)

It seems possible that with some training and research, skilled people might get the information necessary to prepare and operate such an attack either on-board or from the ground. For preparation of any kind of this operation, it is helpful that nowadays the flight tracks, airplane position including flight altitude, direction vector and call-sign are available on the free internet for everybody, this is an additional support and could be misused as a threat for the safety.

Information about the avionic devices, the Integrated Modular Avionics concept (IMA) (see Figure 3) and its relevant Air Transport Association's ATA42 chapter as well as the architecture and block diagrams from avionics and flight control of aircrafts like Airbus A380 are available on the Internet. (Airbus, 2006)

IMA shared resources are the avionics communications network. The solution selected is Avionics Full Duplex Ethernet (AFDX), and it is fully compatible with Ethernet Network of Open World and based on common switch modules. IMA modules are i.e. Core Processing & Input/output Modules (CPIOM), Input/output Modules (IOM) for hosting of several applications and signal acquisition/transmission.



**Figure 3:** Integrated modular avionics – A380 airbus

Another most problematic support is available for training of opponents utilizing the free-internet by publishing information packages from the annual DEFCON meetings, where the cyber-attack relevant information lectures for aircraft and drones hacking are provided by video training. (See [http://www.youtube.com/watch?feature=player\\_embedded&v=1pVP2DhR9Us](http://www.youtube.com/watch?feature=player_embedded&v=1pVP2DhR9Us))

With a dedicated analysis by FMECA and the pre-selection of potential entry points for Cyber Attack by EMS, the more detailed analysis and recommendation will be elaborated. It would also be possible to investigate with a dedicated Failure Tree Analysis (Liggesmeyer, 2009) the weak interfaces between the sensors and the avionics system in ATA42/IMA and ATA23/Communication, which are the breakdown structure for the design, verification and airworthiness rules by EASA and FAA.

## 5. Aviation threat environment

Applying the FMECA methodology on the preselected risk areas from the on-aircraft evaluation, the questions have been raised about the safety of prolonged use of life-time consideration and legacy equipment, which might have experienced upgrading and improvements. It is a most problematic area, when you have the legacy requirements from one or two decades ago and look for a replacement due to obsolescence and need to procure a microprocessor with a lot more computing power and dataspace, most probably defined and produced in an area of the world, where you do not have direct access to the foundry. This may cause additional threat potential, as the principle of embedded subroutines and lay-out for “kill-switch” methodologies must be excluded. The procedure for doing so is not implemented in all the relevant companies and relevant procurement and QA processes nowadays, but relevant services are going to be build-up in specialized companies with forensic experiences. (Kudelski, 2017)

Avionics flight control, the most critical system in a commercial and military airplane, has been moving from a partial by-wire to a more complete fully in electronics by-wire architecture. A level overview of an aircraft's electronic flight control system interaction, which is very similar in lay-out, functionality and risk pattern and

redundancy in the sense of signals from cockpit to the control surfaces and feedback loop. A most modern architecture is shown in the design criteria for an airplane like the Airbus A380, where computers compare differences between the commands, which must last for an enough long period before the command and monitor pair disconnects are forming a fail-safe module, while another command and monitoring computer is in the stand-by as so-called hot spare. The computers are running a self-test loop, whenever the plane is energized. (Airbus, 2006)

An additional aspect for reliability is that the alarms awareness messages and fault-recovery are provided and executed real-time to allow very fast and safe failure compensation. Two primary command and monitor computers, which are available in a pair of secondary command and monitor computers, based on different hardware ICs. Each of the four pairs has different SW developed by different SW development teams and with different tools. This is already a very advanced methodology, which is not so common in legacy products either on commercial aircraft or military aircraft designs, where redundancy is build-up by common design but duplicated only and organized in Master-Slave functionality.

Design diversity in aircraft design is a very important aspect for its system designs, which is based on dissimilar computers and the physical separation of redundancy features, multiple software baselines, compilers during the development cycle and the data diversity. Airbus A380 concepts are based on commercial dual redundant ethernet data networks and Windows for non-critical applications. Those systems like the entertainment system, flight log for cockpit aircrew or the passenger list for flight attendants need to be considered in more details, as it is known in public, that the different versions of Windows application gives way for hacking by open problem descriptions published in open sources. It is this kind of awareness, which is utilized by hacker groups to identify open door entry inside the ethernet bus-system network which might not yet shielded or hardened in the most secure way. The efforts for assessment therefore need to consider the possibility to enter with devices from the EMS into the bus-system/network for manipulating access and data integrity. (Airbus, 2006)

The sensors related to ATM and collision avoidance are possible target of the attack. Especially the spoofing of GPS navigation signals could be a form of attacks, which will cause severe problems during flight. More delicate will be the coordinated take-over of 3D-situation data (flight vector, speed and altitude) in the Collision Avoidance Systems, which by nature are not well hardened as those are transmitting the data between the aircrafts in flight to do the job. In combination with the already mentioned transparency of flight radar tracker, a combined spoofing and fake data uploading is possible to introduce and prepare incidents or near-miss between aircrafts in a same region with same flight vectors and Flight Levels.

The threat is real in the aviation industry, like the rest of the world, it is becoming more and more interconnected, which increases attack vectors to enter systems. When you understand the domain, you are operating in and when you understand the weaknesses and vulnerabilities you can build defenses, but it needs core strengths — engineering and technology (Delorge, 2018). While disrupting air traffic and crippling the economy is frightening enough, the greater fear is that hackers could crash airplanes or make them vanish from radarscopes. One solution, the Cyber Intrusion Detection System, is a cyber-attack warning system that alerts pilots if anything on the aircraft has been hacked or is doing something it shouldn't.

During military operations, a cyber-attack on an aircraft could trick pilots into not trusting their instruments and aircraft. If they don't trust their aircraft, then the mission fails. According to assessment it is identified that malware could be introduced through the supply chain, since aircraft parts are manufactured by many different sources around the world. Therefore, detection systems looking for anomalies on the special on-aircraft bus-systems are necessary. These communication systems control, monitor and transfer data between different electronic components in the aircraft and remote terminals. Many devices connect to those buses, such as annunciators, flaps, lights and landing gear.

In simpler terms: to protect planes and everything around them as attentively as people protect their smartphones. "On my phone, I'm constantly being pushed updates to improve the device's security," Delorge said. "We need that same diligence and vigilance in aviation." Delorge believes the aviation industry should implement a layered approach to cybersecurity, which use several defense mechanisms such as access restrictions, two-factor authentication, encryption, proactive threat hunting, insider threat monitoring, and managed detection and response. (Delorge, 2018)

The most critical clusters is in aerospace, but these findings could be transferred also to automotive technology, as there are similar trends in autonomy driving in the future with the data transfer and management for positioning and flight/driving vectors. In combination with the GPS or future GALILEO navigation data support incidents can be prepared by utilizing the EMS for intrusion of fake data and malware in unprotected, less hardened systems.

Software defined radios in principle are an entry point for data and information transfer linked with the communication and on-aircraft bus-system. There are different levels of protection and hardening, but in principle they are an interesting source for getting attacked and utilized for influencing situation awareness; artificial intelligence could help in the future to control the identified critical areas and support the mentioned identification of disharmonies and wrong corrective measures identification as a cyber-attack warning system (Tyugu, 2015).

These problems might be covered under sabotage, but where not part of any consideration for risk assessment (Altfeld, 2010) and mitigation during last decade. Therefore, it will be worthwhile to evaluate the potential influences also in this respect and discuss about potential dormant problem areas embedded in legacy systems from today's point of view.

## **6. Conclusion and recommendation**

The potential for an adversary to inflict damage on states through EMS attack has grown significantly. Today, all aspects of society, governance, and security have dependencies on EMS. However, power grids, telecommunications, and many command-and-control systems have not been designed to survive a hostile EMS environment. (Stuckenbergs, 2018)

In all the different literatures on assessment and research, the conclusion was unique and everybody in the relevant business cluster, once the real problem could raise up and the preparation done nowadays is not sufficient and late compared to the problems embedded by legacy systems and missing regulation but also the indicated sources in the free internet for teaching and supporting hacking. Relevant networks and workshops with the experts are established meanwhile since 2016 and EASA has performed a workshop identifying the needs and procedures for cyber relevant measures like: standardization, training-awareness-education, civil-military interoperability, risk assessment and methodology, testing (EASA, 2017). IATA has established a 360° training and awareness support for airlines with procedures, training material and provides information for a proactive approach to avoid cyber risks. (IATA, 2017)

Are industry and relevant circles with the link to knowledge-based information exchange and training? As there is a closing up between cyber systems and their dependency on the EMS, as those evolved from wired to wireless architectures (NEC, WLAN, situation awareness sensors), this EM dependency is the real essence of the Cyber-EMS relationship. As nearly most of the devices and capabilities uses the EM environment, as well as the EM systems that provide EM control are not inherent dependent on the cyber space, it is not necessary that an EM system has access to cyber space. That access would not be necessary to enable the embedded ability to maneuver in the EM environment.

From an EM environment point of view, cyber systems reside strictly in the data exchange layer and even the prospective cyber-attack options possibly delivered by radiofrequency jammers are performing a communication function and therefore most parts of the cyber systems are only the EMS users, because datalinks, data networks and information exchanges need access to the EM environment to move around the metadata with information content. It is an experience coming out of the Iraqi battlefields by being confronted with the radio-controlled improved explosive devices, that an opponent or adversary will always seek to exploit the areas of the EM environment where the society, military, federal administration and also the commercial business area yields operational control. Therefore, it is recommended to draw down these virtual boundaries in the discussion forums and bring together the people working in cyber, EW. It is a unique maneuver space upon which all the other military, paramilitary and law enforcement organizations depend.

While EMS vulnerabilities and threats have matured, national and even international capabilities to deny or mitigate such threats and vulnerabilities remain highly dispersed or incomplete (Stuckenbergs et al. 2018). Therefore it could be most beneficial to enlarge the scope of relevant assessments for National Cyber Security

Strategies and include the relevant paradigms from cyber in the EMS, because only awareness of potential risks and not covered risk mitigation will be helpful for cyber attackers to generate the only target they want to achieve – harm and chaos for entering in control and suppress the society and established rules and welfare. When European states renew their cyber strategies, it is imperative that they consider the entire cyber and electromagnetic environment. This is how CyberSec. as a whole can take care of as part of national security.

## **References**

- Airbus (2006). A380-Level III-ATA42 Integrated Modular Avionics & Avionics Data Communication Network. A380 technical training manual. [https://de.scribd.com/document/226105294/A380-Level III. 2006](https://de.scribd.com/document/226105294/A380-Level%20III.%202006)
- Altfeld H-H. (2010). Commercial aircraft projects- managing the development of highly complex products. Taylor & Francis Ltd.
- Arnold J. T. (2009). The Shoreline: Where Cyber and Electronic Warfare Operations Coexist, A Research Report, Air War College Air University, Montgomery, Alabama, 17 February 2009
- Delorge B. (2018). Interview Raytheon VP of Transportation and Support Services
- De Martino A. (2012). Introduction to modern EW systems. Artech House
- EASA (2017). Cyber-security Workshop – Final Report.
- Hagengruber R. (2017). Künstliche Intelligenz – wann übernehmen die Maschinen? Univ. Paderborn. Science on - DFG Podiums Diskussion 12.7.2017
- Hay T. E. (2016). Determining Electronic and Cyber Attack Risk Level for Unmanned Aircraft in a Contested Environment, Air Command and Staff College, Air University, Montgomery, Alabama, August 2016
- Henselmann G. (2017a). Cyber Security in Aeronautics – a generic approach to identify the weaknesses in legacy systems. DWT Cyber Security Workshop, Bonn, 12.-13.12.2017
- Henselmann G. (2017b). Devising and Implementing a National Cyber Security Strategy. Lecture in GSCP, Geneva, 5.10.2017
- Hilbert M. and López P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. In: Science, 332(6025), pages 60–65
- Hofmann R. (2017). Cybersecurity of Things-part 1. An introduction to challenges and techniques for building and operating devices securely. High Assurance Systems/ebook MKT1012.1. [www.high assure.com](http://www.high assure.com).
- IATA (2017). Proactive approach key to mitigate cyber risk. <http://airlines.iata.org/news/proactive-approach-key-to-mitigating-cyber-risk-0?ga=2.94774549.808867565.1529068694-1381224138.1529068694>
- Koch R. and Golling M. (2016). Weapon systems and Cyber Security – a challenging union. 8<sup>th</sup> Intern. Conference on Cyber Conflicts, 2016. Pages 191 ff
- Kudelski Laboratories. (2017). Kill switch forensic. [www.kudelskisecurity.com/sites/default/files/files/Kudelski\\_Security\\_MSS\\_Endpoint\\_Breach\\_Detection\\_EN.pdf](http://www.kudelskisecurity.com/sites/default/files/files/Kudelski_Security_MSS_Endpoint_Breach_Detection_EN.pdf)
- Lehto M. (2015). Phenomena in the Cyber World, in M. Lehto, P. Neittaanmäki (Edit.) Cyber Security: Analytics, Technology and Automation, Springer, Berlin, pages 3-29
- Lehto M. (2018). The modern strategies in the cyber warfare, in M. Lehto, P. Neittaanmäki (Edit.) Cyber Security: Cyber power and technology, Springer, Berlin, pages 3-20
- Liggesmeyer P. (2009). Software-Qualität; Testen, Analysieren und Verifizieren von Software. Springer Verlag
- Russel S. and Norvig P. (2003). (Edit.). Artificial Intelligence- a modern approach. 2<sup>nd</sup> edition. Prentice Hall Series
- Ryan J. (2015). (Edit.) Leading issues in Cyber Warfare and Security. Vol.2. ACPI
- Schmidt E. and Cohen J. (2014). The New Digital Age: Reshaping the Future of People, Nations and Business, Vintage Book
- Seffers, G. (2018). A Quest for Answers on Army Expeditionary Cyber teams, Signal, November 2018
- Stuckenberg D., Woolsey R. J., DeMaio D. (2018). Electromagnetic Defense Task Force, 2018 report, LeMay Center for Doctrine Development and Education, Air University, Montgomery, Alabama, November 2018
- Tyugu E. (2015). Artificial Intelligence in Cyber Defence. Scenario 2040. Cyber Defence – NATO CANADA. 32nd International East/West Security Conference

# Protection of Critical Infrastructure in National Cyber Security Strategies

Eduardo Izzycki<sup>1</sup> and Rodrigo Colli<sup>2</sup>

<sup>1</sup>International Relations Institute, University of Brasília (UnB), Brasília, Brazil

<sup>2</sup>Institutional Security Cabinet, Brazil

[eduizycki@protonmail.com](mailto:eduizycki@protonmail.com)

[rodrigo.colli@gmail.com](mailto:rodrigo.colli@gmail.com)

**Abstract:** A group of eighty six nations has published National Cyber Security Strategies (NCSS). The NCSSs present similarities in basic concepts, in the identification of cyber threats and in the delimitation of strategic objectives. The present article analyzes and compares the NCSSs in respect to the similarities shown within the scope of protection of critical infrastructures (CI). The convergence points identified in the article are: the protection of CIs as a strategic objective; the definition of what constitutes a CI; the services and facilities deemed CI; the existence of a national CI protection program; the need to congregate public and private stakeholders; and the need to build resilience into CI systems. The conclusion points out the countries that have shown through their NCSSs interest in international cooperation for the protection of CIs. This can be achieved by means of joint training, information exchange on threats and incidents against CIs, and in the medium term, the regulation of the use of cyber weapons against CIs.

**Keywords:** national cyber security strategy, cyber security, critical infrastructures

---

## 1. Introduction

National strategies are a plan of action based on a national vision to reach a number of objectives that contribute to security in cyberspace (Luijif, Besseling, & De Graaf, 2013).

These strategies can be considered a political demonstration of the subscribing nation inasmuch as their content tends to divide responsibilities among national stakeholders, to stipulate the intended strategic objectives, to define concrete goals to be achieved within a defined timeframe, and to identify the potential threats perceived by the country (Luijif, Besseling, & De Graaf, 2013). This very content allows one to learn some of the real interests of the subscribing country in relation to the cyber dimension.

The present article is the second result of comprehensive research about national cyber security strategies (NCSSs). It seeks to analyze and compare eighty six documents with the objective of identifying convergence points and opportunities to reach global consensus about the protection of critical infrastructures within the scope of cyber security and defense.

## 2. Methodology

The present article is based on the research carried out by OECD (2012), (Luijif, Besseling, & De Graaf 2011), (Luijif, Besseling, & De Graaf, 2013) and (Sabilon, Cavaller e Cano 2016).

None of the aforementioned articles, however, lay out the fundamentals to compare the various NCSSs. They are all based on fewer documents and simply search for common elements among the documents chosen for analysis.

Since the amount of documents analyzed by the present article is four times greater than that of other research, it behooves us to set objective criteria for comparison among the strategies. There is still no international consensus about the least a NCSS should contain.

Choosing just one of the national strategies as a paradigm for comparison would be a very arbitrary and unsatisfactory solution since it would impose on other countries characteristics peculiar to the author-country whose document was chosen as a paradigm.

The alternative solution was to elaborate a parameter for comparison based on five documents produced by independent international organizations, as follows:

- National Cybersecurity Strategy Guide, from UTI

- Analyzing a new generation of national cybersecurity strategies for the Internet Economy, from the Organization for Economic Co-operation and Development (OECD)
- Guidebook on National Cyber Security Strategies, from the European Network and Information Security Agency (ENISA)
- National Cyber Security Framework Manual, from the Cooperative Cyber Defence Centre of Excellence (CCDCOE)
- Commonwealth Approach for National Cybersecurity Strategies, from the Commonwealth Telecommunications Organization (CTO).

The documents from the UTI (Wamala, 2011), ENISA (ENISA, 2012) and CTO (CTO, 2015) have programmatic traits since their objective is to aide in the elaboration of updating of national cyber security strategies. The three institutions have consultancy programs dedicated to member countries.

The CCDCOE (Klimburg, 2012) focuses on the elaboration of a NCSS and presents tools and recommendations on the confection or revision of national documents regarding the cyber dimension.

The OCDE (OCDE, 2012) report compares national cyber strategies of ten different countries in an attempt similar to the one presented by this article, with an emphasis on similarities, differences and trends identified in the documents analyzed.

In relation to the constituent elements of these national strategies, the documents differ in relation to the terminology used. The CTO (CTO, 2015) mentions “key elements”, ENISA (ENISA, 2012) refers to “specific objectives”, UTI (Wamala, 2011) calls “elements of a national cyber security program”; OCDE (OCDE, 2012) points to “shared concepts”, and CCDCOE (Klimburg, 2012) presents chapters on “strategic objectives” and “organizational structures”.

The perusal of the five documents allowed us to identify the following common strategic objectives (present in at least two documents): certifications and standards, confronting cyber threats, confronting cybercrime, awareness raising and training, coordination and protocols, international cooperation, digital economy, legislation and regulatory benchmarks, public-private partnerships (PPP), research and development (R&D), privacy and protection of critical infrastructures.

Four of the five documents analyzed explicitly point out that the protection of Cls must be a strategic objective to be pursued. ENISA does not make such an overt reference, but does list the protection of Cls as an integrating part of the following objectives: public-private partnerships, mechanisms for information sharing and engaging shareholders.

Thus, it is possible to assert that autonomous sources consider the protection of Cls a recommended measure for a NCSS.

According to the present article, NCSSs that mention the protection of Cls must adhere to the following criteria:

- Signalling the protection of Cls as a strategic objective to be pursued by the subscribing country,
- Regarding Cls as a target or an object of interest for cyberthreats defined by the NCSS,
- Having a national program or government measure for the identification and protection of national critical infrastructures,
- Having a sample list of services or facilities classified as critical infrastructure or,
- Referencing a systematic national-level process for risk assessment in Cls.

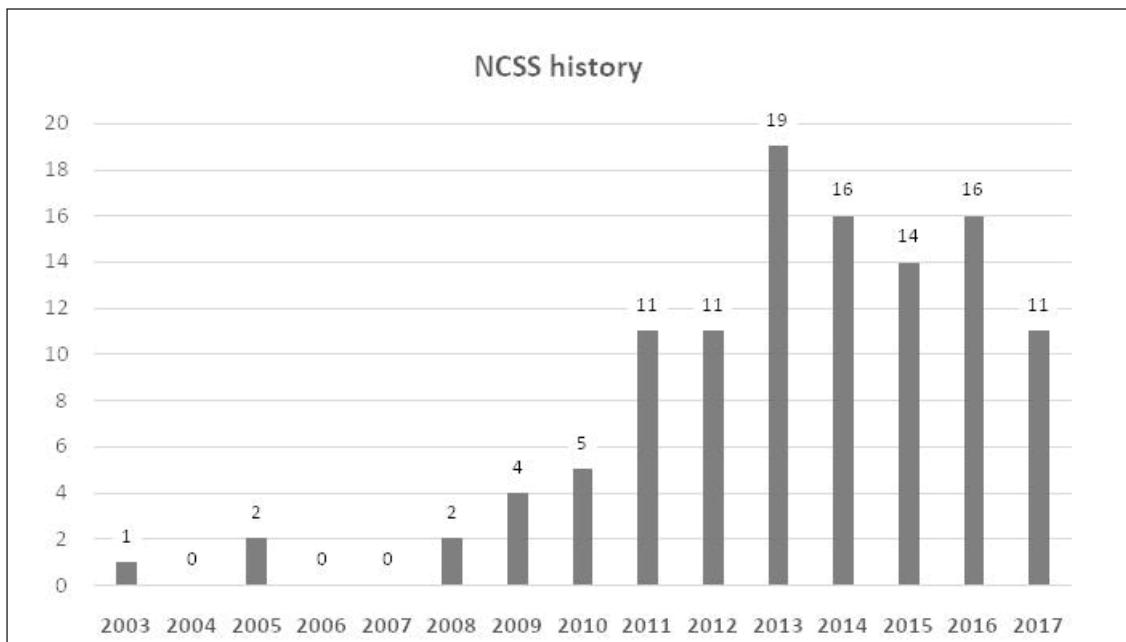
The NCSS content verification process was divided in three stages: during the first one the 86 documents were perused for the identification of common terminology regarding strategic objective and concepts; in the second stage, selected terms were searched for using Python to minimize human error in identifying relevant terms; and, in the third stage, there was human supervision of the collected results via script to eliminate the presence of a false positive in the final result.

## 2.1 Time analysis

The time analysis based on the year of publication of the NCSSs reveals how cyber security has gained importance as a theme of interest for States.

Between 2003 and 2017, 112 strategies were published by 86 different countries, as shown in Figure 1.

The year 2008 appears to be starting point for the annual publication of new NCSSs. In April 2007, there were the notorious cyber attacks against the Republic of Estonia (frequently credited to Russia thought still lacking conclusive evidence) and in 2008 the cyber attacks against Georgia were followed by a conventional Russian military campaign in South Ossetia. There were direct references to these cases in ten different strategies analyzed.



**Figure 1:** Total number of NCSSs published annually

The number of documents published increased significantly between 2010 and 2011, from five to eleven strategies. This number has remained above ten new publications since then.

This increase in the publication of new strategies can be credited to the Stuxnet case, which gained notoriety in that year as a classic case of a cyber attack. The attack, which generated kinetic effects in the Iranian nuclear program, would supposedly have been a joint action between the USA and Israel (Richardson, 2011). Seven different strategies made direct reference to this case.

The NCSSs represent a very significant amount of data for analysis. Besides the sheer number of countries, the nations that have a cyber national strategy are very relevant in terms of geopolitics.

It is worth highlighting that in 13 countries (Australia, Singapore, Colombia, USA, Holland, Japan, Luxembourg, Norway, Poland, United Kingdom, Czech Republic, Russia and Zimbabwe) the strategies have already been updated in newly published editions.

Another aspect that reinforces the perception of relevance towards the cyber dimension is the number of strategies about the issues in some countries. In the case of Germany, Australia, USA, France, New Zealand and the United Kingdom, besides the NCSS, there are strategies for the digital economy, strategies for international cyberspace cooperation and strategies to combat cybercrime.

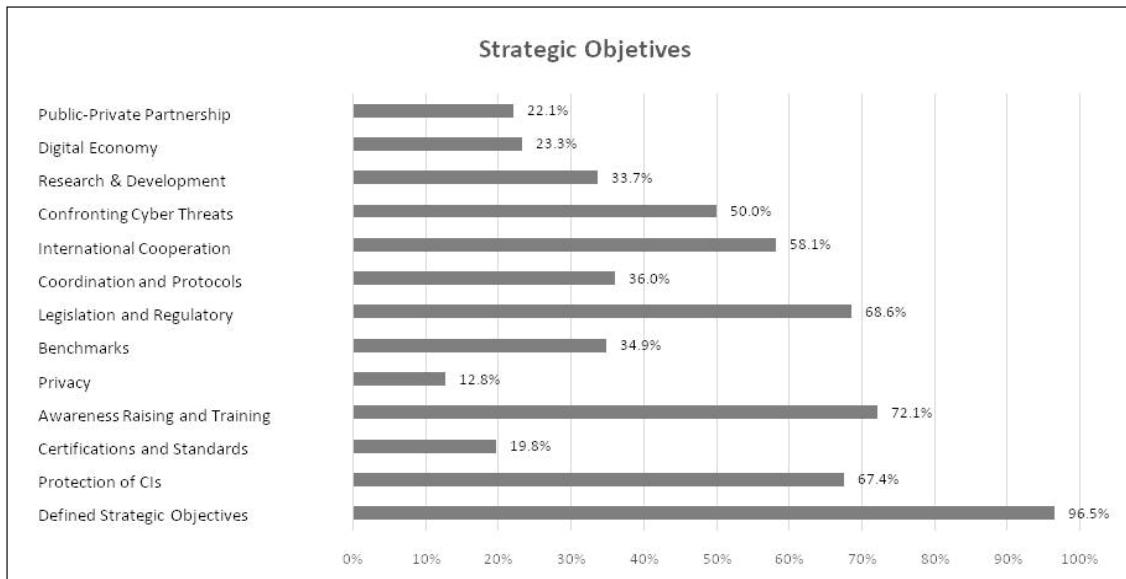
This shows that the cyber dimension has become a prominent theme for geopolitically relevant nations. Furthermore, the continuous updating of strategies clearly points to the need for quick adaptation to a theme that has been undergoing constant modification in the last few years.

### **3. Strategic objectives**

Regarding the strategic objective clearly defined in the NCSSs, in most of the documents (82) the strategic objectives to be pursued were defined, as shown in Figure 3.

Still in relation to Figure 3, we point out that the strategic objectives awareness raising and training (62), coordination and protocols (59), protection of critical infrastructure (58), international cooperation (50) and confronting cyber threats (43) are present in most of the NCSSs.

The protection of CIs was the third most frequent strategic objective. The way in which the countries define this objective will be analyzed by this article.



**Figure 3:** Frequency of the strategic objectives observed

Listing the protection of CIs as a strategic objective in the NCSS should be considered as a qualified reference and not just a mere mention of the concept in the body of the document since this indication suggests the country's commitment to it. We presume this would naturally lead to the allocation of human and material resources in pursuing this declared objective, which, in turn, configure it as a genuine demonstration of national political will.

Thus, we can infer that for the 58 countries (67.4%) that published a NCSS the protection of CIs is considered a priority within the context of cyber security at a national level.

### **4. Concepts and content**

Besides identifying the protection of CIs as a strategic objective, the comparative analysis of NCSSs detected conceptual similarities in relation to the definition of a CI, impact measures (criticality), services and facilities considered CIs, the existence of national CI protection programs and risk assessments as a method for the protection of CIs.

These similarities are indicators that the environment is favorable to international cooperation regarding CIs. The concepts that presented similarities in the NCSSs will be presented in the next section.

#### **4.1 Definitions of critical infrastructures**

From the expression "critical infrastructure" one can derive its two constituent elements in conceptual terms. The first is that of services and facilities used by society (infrastructure) and the second refers to its relevance (critical) measured by the negative consequences that its disruption or malfunction would generate to the public.

In the first part of the concept (the definition of what services are deemed relevant to the public), 20 categories were classified as a CI by 77 NCSSs. The model adopted by most NCSSs was the elaboration of a list of productive

and/or economic sectors to which the predicate CI was added. The lists are mostly illustrative, which would make them dynamic and would allow for the inclusion of other services if the relevance criteria (criticality) is present.

The productive and/or economic categories that were most commonly deemed CI by the NCSSs are listed in Table 1. We have also identified the countries that have directly listed this category of essential service as a segment of CI and the total number of countries.

**Table 1:** categories considered CI

Services and Facilities	Countries	Total
<i>Electrical Energy</i>	<i>Germany, Australia, Bangladesh, Bulgaria, Burkina Faso, Canada, Chile, China, South Korea, Denmark, Slovakia, Ghana, Jersey Island, Ireland, Jamaica, Japan, Jordan, Kosovo, Malaysia, Qatar, Kenya, Romania, Singapore, Sweden, Taiwan, Turkey and Zimbabwe</i>	27
<i>Telecommunications</i>	<i>Germany, Australia, Bangladesh, Bulgaria, Burkina Faso, Canada, Chile, China, South Korea, Denmark, Ghana, Jersey Island, Ireland, Iceland, Jamaica, Japan, Jordan, Kosovo, Malaysia, Qatar, Kenya, Singapore, Taiwan, Turkey, Ukraine and Zimbabwe</i>	26
<i>Transportation</i>	<i>Germany, Australia, Bangladesh, Bulgaria, Burkina Faso, Canada, Chile, China, South Korea, Slovakia, Ghana, Jersey Island, Ireland, Iceland, Jamaica, Japan, Kosovo, Malaysia, Qatar, Romania, Singapore, Sweden, Taiwan, Turkey and Zimbabwe</i>	25
<i>Finance</i>	<i>Germany, Australia, Bangladesh, Bulgaria, Burkina Faso, Canada, Chile, Slovakia, Ghana, Iceland, Jamaica, Japan, Jordan, Kosovo, Malaysia, Qatar, Romania, Singapore, Sweden, Taiwan and Turkey</i>	21
<i>Water / Sewage</i>	<i>Germany, Australia, Bangladesh, Burkina Faso, Canada, Chile, Ghana, Ireland, Jamaica, Japan, Jordan, Kosovo, Malaysia, Qatar, Kenya, Singapore, Sweden, Taiwan and Turkey</i>	19
<i>Public Health</i>	<i>Germany, Australia, Bangladesh, Bulgaria, Burkina Faso, Canada, Chile, China, Slovakia, Ghana, Ireland, Jamaica, Jordan, Kosovo, Malaysia, Qatar, Singapore, Sweden e Turkey</i>	19
<i>Information Technology</i>	<i>Germany, Burkina Faso, Chile, China, Hungary, India, Jamaica, Jordan, Qatar and Ukraine</i>	10
<i>Emergency Services</i>	<i>Australia, Bangladesh, Burkina Faso, Chile, Ghana, Jamaica, Malaysia and Singapore</i>	8
<i>Public Services</i>	<i>Burkina Faso, Canada, Ghana, Japan, Jordan, Kosovo, Singapore and Taiwan</i>	8
<i>Food Security</i>	<i>Germany, Australia, Bangladesh, Canada, Ghana, Jamaica, Kosovo and Malaysia</i>	8
<i>National Defense</i>	<i>Burkina Faso, Canada, Chile, Ghana, Malaysia, Romania and Taiwan</i>	7
<i>Distribution and logistics</i>	<i>Japan, Jordan, Kenya and Sweden</i>	4
<i>Natural Gas and Oil</i>	<i>Jersey Island, Japan and Taiwan</i>	3
<i>Chemical/Nuclear Industry</i>	<i>Japan e Kosovo</i>	2
<i>Trade</i>	<i>Ireland</i>	1

In the case of the economic sector Transportation, we observed specific reference to aerial transportation (Australia, Iceland, Japan e Taiwan), maritime transportation (Australia and Jersey Island), road transportation (Zimbabwe) and railroad transportation (Japan). In all cases, the countries which presented a specific reference to the mode of transportation also indicated the sector Transportation as a CI. This is the reason why they were not presented as classifying autonomous CI.

The second aspect observed in the definitions of CI regards the criticality of services or facilities. The method to assess the relevance of the essential services also presented similarities among the countries. They mostly converge toward consequences related to economic aspects, security, loss of human lives, society well-being, operation safety, social impact and political impact.

The most common methods for assessing the criticality of productive and/or economic sectors identified in the NCSSs are presented in Table 2. We have also identified the countries that have directly listed this category and the total number of countries.

**Table 2:** categories considered CI

Services and facilities	Countries	Total
Economic	Afghanistan, Saudi Arabia, Australia, Austria, Brazil, Canada, Chile, Colombia, The Philippines, Ghana, Jamaica, Jordan, Latvia, Lithuania, Malaysia, Panama, Paraguay, Qatar, Kenya, Trinidad and Tobago and Turkey	21
Security	Afghanistan, Germany, Saudi Arabia, Australia, Austria, Canada, Chile, The Philippines, Ghana, Jamaica, Kosovo, Latvia, Lithuania, Malaysia, Mexico, Paraguay, Qatar, Trinidad and Tobago, Turkey Uganda	20
Physical Integrity and/or Loss of human lives	Afghanistan, Saudi Arabia, Austria, Canada, Chile, The Philippines, Ghana, Jamaica, Jordan, Latvia, Malaysia, Paraguay, Qatar, Trinidad and Tobago, Turkey and Uganda	16
Social well-being	Australia, Austria, Canada, Chile, Colombia, Latvia, Lithuania, Panama, Paraguay and Qatar	10
Operation Safety	Saudi Arabia, Canada, The Philippines, Ghana, Jamaica, Malaysia, Qatar and Uganda	8
Social Impact	Australia, Austria, Brazil, Jordan, Latvia, Lithuania and Paraguay	7
Political Impact	Afghanistan, Jordan and Turkey	3

When measuring the negative consequences of the disruption or malfunctioning of essential services, the NCSSs reveal which are the main values that the documents wish to preserve. Similarities in such valuation represent a convergence point among the countries since they show, based on similar motives that they wish to preserve services and facilities for the general public.

The combination between the CI services and facilities (electrical energy, telecommunications, transportation, finance, water/sewage and public health) and the values they wish to preserve (economy, security, physical integrity and/or loss of human lives) show that there is conceptual convergence among the countries regarding what is considered a CI.

#### **4.2 Cyber threats and CIs**

Fifty-seven NCSSs made a reference to cyber threats. Though the content of the concepts differ slightly, the following taxonomy was rather consistent: cyber criminality, espionage, hacktivism, sabotage, cyber terrorism, state actors and cyber war.

Twenty-four NCSSs perceive that CIs are attractive targets for varied cyber threats since the consequences of a successful attack could be catastrophic.

In twelve NCSSs (Saudi Arabia, Austria, Australia, Bangladesh, Canada, Colombia, Ghana, Moldavia, Mauritius, Norway, Poland, Qatar, Rwanda, Switzerland) the relation between cyber threats and CIs was not oriented by the aforementioned taxonomy. The association was rather generic, based on the assertion these CIs are typically an attractive target for such threats.

In the cases of Slovenia and Romania there was an association between cyber crime and attacks against CIs, based on the specialization of criminal activities (Crime as a Service – CaaS) that can be contracted to carry out an attack promoted by other kinds of actors.

The association between cyber terrorism and attacks against CIs was the most popular, present in the NCSSs of Slovenia, France, Georgia, Switzerland, New Zealand, Samoa and Ukraine. The association between this threat and CIs was made based on the need for media impact that terrorist groups seek – they could gain publicity through a cyber attack against a CI with kinetic effects.

In the cases of Spain, France, Hungary, Yemen and the United Kingdom there was overt association to acts of cyber war by national states against CIs. Spain's NCSS cites "that there is evidence that State actors have offensive capabilities to attack CIs" and the UK's NCSS mentions the detection of "attacks carried out by states or sponsored by them."

There is, however, a growing collective perception that cyber threats are especially interested in CIs. This could represent a more ambitious cooperation point: creating trust-building mechanisms and/or limiting the development of cyber weapons specifically aimed at CIs. The proposal is analogous to measures that control chemical weapons, which are quite consensual in the international community.

There is a relative level of common understanding among the countries regarding which services and facilities should be considered CI: electrical energy (27), telecommunications (26), transportation (25), finance (21), water/sewage (19) and public health (19).

Offensive cyber artefacts built to attack CIs demand development efforts and expertise (protocols applicable to programmable logic controller – PLC). Because of this, they are usually the objective of state actors. Even if they are created by private actors, their main purchasers would be nation-states due to the costs of acquisition and the kinetic effects they might generate.

Based on the shared understanding of what constitutes a CI and that they are the object of cyber threats, it is feasible to propose measures for creating trust-building mechanisms among the countries to regulate the use of cyber weapons with this intent.

Additional evidence can be extracted from historical examples of the use of cyber weapons against CIs: CrashOverride, Dragonfly, Industroyer and Stuxnet. In this last case, the attack was aimed against uranium enrichment plants in Iran but spread throughout the world due to its worm self-replicating characteristic, although it only unleashed its dreadful consequences when very specific conditions were met (Richardson, 2011).

Following the Stuxnet case new malware samples were discovered, they were called Duqu, Flame, and Gauss. Those malware had close resemblance to Stuxnet, but the last two were not authored by the same group. That illustrates that an attack against an automation system might not be restrained to the intended target. Moreover, once the campaign is set in motion, it is impossible to guarantee that part of its code will not be used by other actors for new attacks against different targets. (Bencsáth, Pék, Buttyán, and Félegyházi, 2012)

A parallel argument to limit the use of cyber weapons against CIs can be extracted from the Principle of Necessity, Distinction and Proportionality (International Humanitarian Law) which demands that acts of war be directed strictly against combatants and military objectives of the enemy so as to avoid unnecessary or excessive damage to civilians.

Since a country's CIs, in general, serve simultaneously civilian and military purposes, even military telecommunication networks, in most countries, use the same infrastructure used for civilian purposes. A denial of service attack against traffic transfer hubs or against power plants, for instance, could be considered a disproportionate attack on civilian. These matters merit further discussion regarding if strict cyber effects against CI that affect civilians is indeed a legitimate military target (Schmitt, 2018).

#### **4.3 National critical infrastructure protection program**

Among the 77 NCSSs that acknowledge the protection of CIs, there were 23 cases in which the document indicated the existence of a national IC protection program. Furthermore, 26 strategies presented CI risk assessments as a goal to be reached.

We must also point out the lack of acknowledgment of some countries that have well-known national CI protection programs: USA, France, New Zealand and United Kingdom.

France lists the protection of CIs as a strategic objective in its NCSS. The lack of mention, however, is probably due the higher status of the CI protection strategy (2013), conducted by the Secrétariat général de la Défense et de la Sécurité nationale (SGDSN) in relation to the NCSS (2011).

In the cases of the USA, New Zealand and the United Kingdom, the NCSSs analyzed in this article did not list the protection of CIs as a strategic objective. The omission could be interpreted as the lack of integration between two high level documents or as autonomous demonstrations by distinct entities.

In the United States, the NCSS was published by the Department of Defense and the protection of CIs is the task of the Department of Homeland security. In the United Kingdom, the NCSS was issued by the Cabinet Office and the protection of CIs is the responsibility of the Center for the Protection of National Infrastructure (CPNI). In New Zealand, the NCSS was produced by the Ministry for Communications and the protection of CIs is the task of the National Infrastructure Unit (linked to the Ministry of Finance).

The existence of national CI protection programs in different countries did not reveal similarities between the different approached proposed. Nine countries point to the need to carry out threat assessments (Austria, Bangladesh, Canada, Cyprus, Nigeria, Qatar, Russia, Switzerland and Uganda), four cite vulnerability assessment (Bangladesh, Brazil, Italy and Switzerland) and eleven suggest the need for specific approaches for different CI sectors (South Africa, Bulgaria, Croatia, Greece, Holland, Italy, Japan, Jordan, Norway, Poland and Qatar).

Another aspect regards drills and simulations for the protection and defense of CIs. Only fifteen countries (South Africa, Austria, Cyprus, Estonia, Finland, France, Holland, Ireland, Japan, Luxembourg, Nigeria, Norway, Poland, Qatar and Singapore) stated that they will carry out drills and/or simulations of cyber attacks or disruption of CIs.

In the case of Poland, there is reference to state and sector level drills; South Africa states that drills and simulations must include vulnerability assessment and penetration tests. In the cases of Austria, France, Finland, Ireland and Holland, they underscore the need for active participation of the private sector in the drills.

However, the lack of common elements among the countries regarding their national CI protection programs negatively affects the cohesion of security measures and global defense.

#### **4.4 Similar CI concepts in the NCSSs**

Finally, three aspects regarding the protection of CIs were repeatedly observed: the need to build resilience in CIs, the engagement of different stakeholders and its economical relevance to society.

As far as the need to build resilience is concerned, the 39 NCSSs that mentioned it all converge to a concept of resilience that contemplated the “capabilities to operate and/or to recover after an attack”, but they do not specify the necessary parameters required for resilience.

Regarding the engagement of the different stakeholders, 38 NCSSs suggest that this is an indispensable element for the protection of CIs. The NCSSs are unanimous in pointing out that services and facilities deemed CIs are, in great part, managed by the private sector. Some NCSSs suggest the creation of public-private partnerships to deal with the issue of CI protection.

Another stakeholder frequently mentioned by the NCSSs is academia. Wherever research entities are mentioned, they are seen as responsible for developing cyber security and CI security solutions.

Finally, the economic relevance of CIs is highlighted by 27 NCSSs. These strategies refer to CIs as potential targets for cyber threats due to the potential impact of a cyber attack, which would be maximized due to the interdependence that CIs have with one another (for instance, an attack against a power plant would certainly affect telecommunications).

### **5. Conclusion**

The analysis of 86 documents revealed that the elaboration of NCSSs is a consolidated trend. There are second or third generation strategies in 13 countries. The elaboration of a NCSS is a global trend that has been consolidated in the last 15 years.

The comparative analysis among the NCSSs revealed that the protection of CIs is the third most frequent strategic objective: 58 countries list it as one their goals for cyber security at a national level.

Other identified convergence points among the NCSSs were the services and facilities considered CI (electrical energy, telecommunications, transportation, finance, water/sewage and public health) and the values they wish to preserve (economy, security, physical integrity and/or the loss of human lives).

The susceptibility of CIs to different types of cyber threats was another convergence point in 57 NCSSs.

Finally, the economic consequences of offensive actions against CIs (27) the need to build resilience (39) and the participation of private and public stakeholders in CI protection actions (38) were other similarities observed among the different NCSSs.

The evidence gathered seems support a transnational initiative to build a positive agenda for the protection of CIs among Nation-States.

The fact the three strategic objectives were identified in most NCSS suggests that international cooperation for joint training for CI protection is an initiative that would find acceptance in most countries. These training initiatives must include segments of the private sector (CI operators or solution providers) given the similarities regarding the operation of CIs in different countries.

A second international CI cooperation point would be cyber threats. Twenty-four NCSSs perceive that CIs are an attractive target to different cyber threats since the consequences of a successful attack would be potentially catastrophic.

This collective perception that cyber threats, with an emphasis on cyber terrorism and state actors, are interested in CIs makes it an available area for cooperation, whether due to the information sharing about threats or to cyberattacks carried out against CIs.

A third cooperation point that can be foreseen in a more distant future would be creating trust-building mechanisms and/or limiting the development of cyber weapons aimed specifically against CIs, or at least against the services and facilities considered a CI by many of the NCSSs.

The United Nations Group of Governmental Experts (GGE) could reinforce its initiatives in international security from these shared positions in NCSS. The GGE already portrayed great concern over CIs and the data provided from this research could create common ground for further recommendations (General Assembly, 2015).

## **References**

- Alexander Klimburg (Ed.) (2012), National Cyber Security Framework Manual, NATO CCD COE Publication, Tallinn 2012.  
<https://www.oecd.org/sti/ieconomy/cybersecurity-policy-making.pdf>
- Bencsáth, B.; Pék, G.; Buttyán, L.; Félegyházi, M. The Cousins of Stuxnet: Duqu, Flame, and Gauss. Future Internet 2012, 4, 971-1003.
- Commonwealth Telecommunications Organization (2015), Commonwealth Approach for developing National Cybersecurity Strategies. <http://seguridadcibernetica.mingob.gob.gt/wp-content/uploads/2016/09/ENSC2017.pdf> (Last seen on December 28, 2017)
- European Network and Information Security Agency (2015), National Cyber Security Strategies - Practical Guide on Development and Execution, European Network and Information Security Agency (ENISA).  
[http://www.enisa.europa.eu/activities/Resilience-and-CIIP/national-cyber-security-strategies-ncss/national-cyber-security-strategies-an-implementation-guide/at\\_download/fullReport](http://www.enisa.europa.eu/activities/Resilience-and-CIIP/national-cyber-security-strategies-ncss/national-cyber-security-strategies-an-implementation-guide/at_download/fullReport) (Last seen on December 28, 2017)
- General Assembly resolution 70/147, Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security, A/RES/70/174 (22 July 2015), available from undocs.org/A/70/174.
- International Telecommunications Union (ITU) (2017). Global Cybersecurity Index 2017. Genebra: ITU.  
<http://www.itu.int/en/ITU-D/Cybersecurity/Pages/GCI-2017.aspx> (Last seen on December 28, 2017)
- Luijif H.A.M., Besseling K., Spoelstra M., de Graaf P. (2013) Ten National Cyber Security Strategies: A Comparison. In: Bologna S., Hämmmerli B., Gritzalis D., Wolthusen S. (eds) Critical Infrastructure Security. CRITIS 2011. Lecture Notes in Computer Science, vol 6983. Springer, Berlin, Heidelberg
- Luijif, E., Besseling, K. e De Graaf, P. (2013) Nineteen national cyber security strategies, Int. J. Critical Infrastructures, Vol. 9, Nos. 1/2, pp.3–31.
- Organisation for Economic Co-Operation and Development (2012). Cybersecurity Policy Making at a Turning Point: Analysing a New Generation of National Cybersecurity Strategies for the Internet Economy, OECD Digital Economy Papers, No. 211, OECD Publishing, 2012. <http://www.oecd.org/sti/ieconomy/cybersecurity-policy-making.pdf> (Last seen on December 28, 2017)
- Richardson, John Charles, Stuxnet as Cyberwarfare: Applying the Law of War to the Virtual Battlefield (July 22, 2011). Available at SSRN: <https://ssrn.com/abstract=1892888> or <http://dx.doi.org/10.2139/ssrn.1892888>

***Eduardo Izycki and Rodrigo Colli***

- Sabillon, R., Cavaller, V., Cano, J. "National Cyber Security Strategies: Global Trends in Cyberspace." International Journal of Computer Science and Software Engineering 5, no. 5 (2016): 67-81.
- United Nations (2016). UN e-Government Survey 2016. E-Government in Support of Sustainable Development. New York: UNPAN. <https://publicadministration.un.org/egovkb/en-us/reports/un-e-government-survey-2016> (Last seen on December 28, 2017)
- Schmitt M (2018). International Cyber Norms: Reflections on the Path Ahead. Militair Rechtelijk Tijdschrift, 111, 12-23.  
Available at: [https://puc.overheid.nl/mrt/doc/PUC\\_248171\\_11/1/](https://puc.overheid.nl/mrt/doc/PUC_248171_11/1/)
- Wamala, Frederick (2011). The ITU National Cybersecurity Strategy Guide, International Telecommunications Union (ITU), 2011. <http://www.itu.int/ITU-D/cyb/cybersecurity/docs/ITUNationalCybersecurityStrategyGuide.pdf>(Last seen on December 28, 2017)

# Towards a Framework for the Selection and Prioritisation of National Cybersecurity Functions

Pierre Jacobs<sup>1</sup>, Sebastiaan von Solms<sup>1</sup>, Marthie Grobler<sup>2</sup> and Brett van Niekerk<sup>3</sup>

<sup>1</sup>University of Johannesburg, Johannesburg, South Africa

<sup>2</sup>Data61, CSIRO, Melbourne, Australia

<sup>3</sup>University of KwaZulu-Natal, Durban, South Africa

[pierrej06@gmail.com](mailto:pierrej06@gmail.com)

[basievs@uj.ac.za](mailto:basievs@uj.ac.za)

[marthie.grobler@data61.csiro.au](mailto:marthie.grobler@data61.csiro.au)

[vanniekerk@ukzn.ac.za](mailto:vanniekerk@ukzn.ac.za)

**Abstract:** In the development of a national cybersecurity function model, management tasks include the identification, selection and prioritisation and implementation of these functions. While frameworks do exist to help with the identification of national cybersecurity functions, our research could not find a framework to aid nation states to select and prioritise cybersecurity functions for implementation. Selecting cybersecurity functions for national implementation is primarily achieved by considering the cybersecurity dimensions, mandates and domains they will operate in. As a secondary outcome, the dimensions, mandates and domains (originally introduced by the North Atlantic Treaty Organization's (NATO) National Cyber Security Framework Manual (NATO Cooperative Cyber Defence Centre of Excellence (CCDCOE), 2012) could influence the prioritisation of these functions for implementation. While the primary function of the dimensions, mandates and domains is to aid the selection of national cybersecurity functions for implementation, the prioritisation of cybersecurity functions for implementation is primarily achieved by following a national cybersecurity risk management framework and process. A nation's risk management strategy should prescribe a national cybersecurity risk management framework and process. Preliminary research did not produce a national cybersecurity risk management framework and process, and we further propose such a framework and process by combining two existing standards. In this research we propose a framework to aid nation states to select and prioritise cybersecurity functions for national implementation. This research has as outcome a framework to be used during the selection and prioritisation of national cybersecurity functions for implementation.

**Keywords:** cybersecurity, frameworks, national cybersecurity, cybersecurity functions, cybersecurity risk management

---

## 1. Introduction

A framework to identify the national cybersecurity functions specific to countries were proposed by Jacobs, Von Solms and Grobler (2016). The framework can be used to provide countries with a list of mandatory and non-mandatory national cybersecurity functions depending on how it is applied. The application of this framework has as outcome a list of national cybersecurity functions. Nations then need to make a selection from the identified national cybersecurity functions for implementation. Thereafter, the selected national cybersecurity functions need to be prioritised for implementation.

In this paper, the authors propose a framework to assist nation states with the selection and prioritisation of the identified national cybersecurity functions for implementation based on the nation's unique political environment and structure. Section 2 provides background on the national cybersecurity function identification part of the framework, with the elements contributing to it. This is important since some of the elements contributing to the identification of national cybersecurity functions, also contribute to the selection and prioritisation of those functions.

In Section 3, the Dimensions, Domains and Mandates as Elements influencing the Selection of National Cybersecurity Functions are introduced. Section 4 proposes a national risk management framework, while Section 5 introduces the complete framework, combining the national cybersecurity function identification, selection and implementation elements.

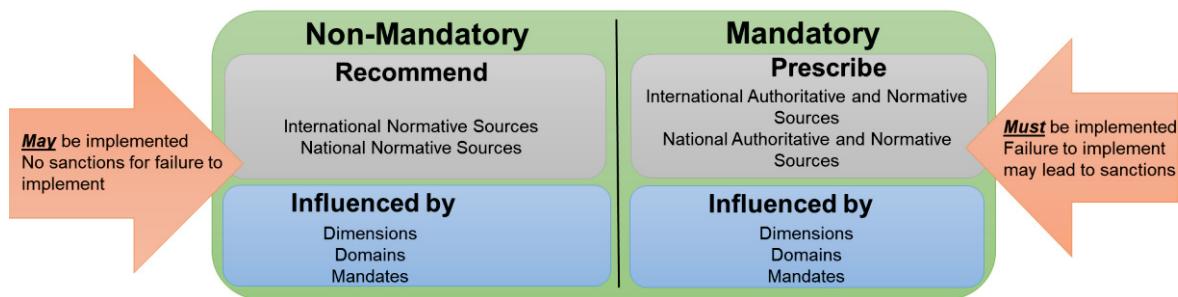
## 2. Background

The framework proposed by Jacobs von Solms and Grobler (2016) (Jacobs, von Solms, & Grobler, 2016) can be used to identify national cybersecurity functions and determine mandatory and non-mandatory national cybersecurity functions. The identification of national cybersecurity functions is most often a political process

and our framework can be augmented with expert advice by establishing advisory committees. To determine the national cybersecurity functions, the following process is followed:

1. Identify existing national and international authoritative and normative sources
2. Identify cybersecurity functional prescripts expressed in these sources
3. Identify cybersecurity functional recommendations expressed in these sources

The outcome of this process yields a list of mandatory and non-mandatory national cybersecurity functions from which a selection must be made for implementation. Neglecting to implement mandatory national cybersecurity functions may result in sanctions as they are prescribed in acts and policies, whereas failure to implement non-mandatory functions will not result in any sanctions. The implementation of the selected national cybersecurity functions also need to be prioritised. The identification of mandatory and non-mandatory national cybersecurity functions may be influenced by the cybersecurity dimensions, domains and mandates. The process followed to determine mandatory national cybersecurity functions is shown in Figure 1:



**Figure 1:** Mandatory national cybersecurity functions determination process

Following this approach, and by consulting generic international normative sources, we have identified thirteen non-mandatory or generic national cybersecurity functions (Jacobs, von Solms & Grobler, 2017). The sources consulted are:

- The North Atlantic Treaty Organisation's (NATO) National Cyber Security Framework Manual (2011)
- The United Kingdom's Cybersecurity Capacity Maturity Model for Nations (CMM) - Revised Edition (2016) (Roberts, 2014).
- The ITU National Cybersecurity Strategy Guide (2011) (Wamala, 2011)
- Cybersecurity Capability Maturity Model (C2M2) - Version 1.1 (2014) developed by the United States
- Department of Homeland Security (Christopher, 2014).

The thirteen generic national cybersecurity functions are presented in Table 1:

**Table 1:** Thirteen national cybersecurity functions

sn	Generic National Cybersecurity Function
1	Military Cyber / Cyber Warfare
2	Cybercrime / Investigations / Digital Forensics
3	Research and Development (R&D), Education and Awareness
4	Critical Information Infrastructure Protection (CIIP)
5	Cryptography
6	E-Identity
7	Incident Handling
8	Monitoring and Evaluation
9	Internal Coordination
10	External Stakeholder Engagement
11	National Policy and Strategy Development
12	National Strategic Risk and Threat Assessment
13	National Regulations Development

These national cybersecurity functions are offered from national cybersecurity structures. Due to cost and skills constraints, it would not be possible for any nation state to implement all thirteen generic national cybersecurity functions at the same time. Therefore, a framework is needed to assist nation states to select and prioritise national cybersecurity functions for implementation based on the nation's unique structure and political

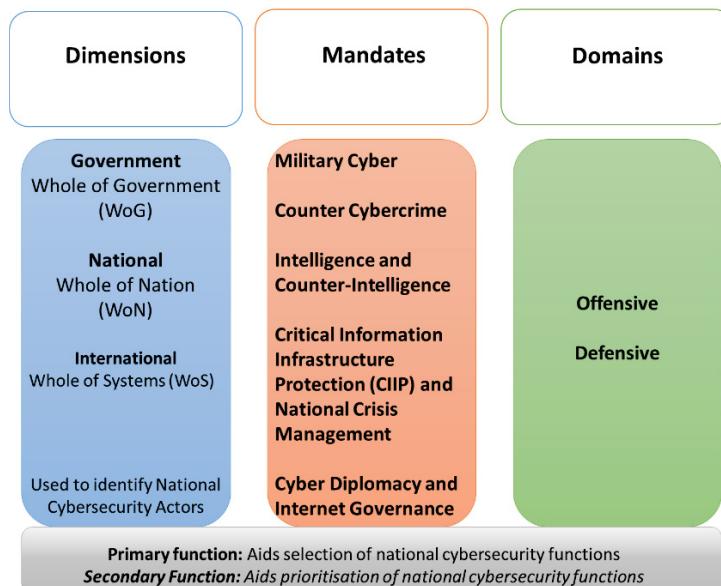
mandates. The elements influencing the selection and prioritisation are presented in Section 3. These elements are:

- Cybersecurity Dimensions
- Cybersecurity Domains
- Cybersecurity Mandates
- National Cybersecurity Risk Management Framework

### 3. Critical dimensions, domains and mandates

NATO identified three dimensions where cybersecurity activity takes place, to address five different mandates in two cyber domains (NATO CCDCOE, 2012) (NATO Cooperative Cyber Defence Centre of Excellence, 2012). Together with the authoritative and normative mandatory prescripts, the dimensions, mandates and domains assist in the selection, and prioritisation of non-mandatory national cybersecurity functions, as well as with the identification of actors and stakeholders.

The dimensions, mandates and domains are presented in Table 1. The dimensions assist with the identification of national cybersecurity stakeholders. Responsibility for the implementation of national cybersecurity functions could be given to these stakeholders. The mandates and domains serve primarily to aid the selection of national cybersecurity functions, and as a secondary outcome, to prioritise national cybersecurity functions. The NATO dimensions, mandates and domains are shown in Figure 2:



**Figure 2:** NATO Cybersecurity Dimensions, Mandates and Domains (2012)

To illustrate the application of the framework, the following dimensions, domains and mandates are selected:

- All three dimensions are selected,
- The defensive domain is selected, and
- The Critical Information Infrastructure Protection (CIIP) and National Crisis Management mandate is selected

Our experience across this mandate and domain in context of South Africa influenced the selection. The selected dimensions, mandate and domain is introduced in the following sub-sections.

#### 3.1 Dimensions

A dimension describes the element or factor making up an entity - such as national cybersecurity, or describe the range or degree to which national cybersecurity stretches (Merriam-Webster, 2017). The three dimensions of Government, National and International allow for national cybersecurity to be viewed from different perspectives, such as political, law enforcement or military. NATO identified three categories of actors across

the three dimensions: State and Government Actors, Organised Non-State Actors, and Non-Organised Non-State Actors (NATO CCDCOE, 2012). A simple framework can be constructed by combining the three dimensions and the categories of actors. This framework is then populated with actors specific to a country. To illustrate its application, this framework is populated with possible actors in Table 2.

Actors are identified using the dimensions and Actors defined by NATO. State and Government Actors could be a nation's Department of Communications (South African Government, 2017), Security Agencies (SSA, 2016), Reserve Banks (SARB, 2017), the Department of Defence (DOD) (DOD, 2016), the country's national CSIRT (DTPS, 2015), and their Police Service (SAPS, 2016).

National Organised Non-State Actors could be the country's Financial Sector Risk Centers (SABRIC, 2016), and the country's IT industry. An example of International Organised Non-State Actors are the United States Computer Emergency Response Team (US-CERT) (Department of Homeland Security, 2017) and the Forum for Incident Response (FIRST) (FIRST, 2016). International Non-Organised Non-State Actors are hacker groups Lulzsec (Fox News, 2011) and Anonymous (Anonymous, 2016).

**Table 2:** National cybersecurity actor identification framework

	<b>Government</b>	<b>National</b>	<b>International</b>
State and Government Actors	Department of Communication, Security Agencies, Police Service, Reserve Banks Department of Defence	National-CSIRT	
Organised Non-State Actors		Financial Sector Risk Centers, Financial Sector CSIRT, IT Industry	US-CERT, FIRST
Non-Organised Non-State Actors		Public	LulzSec Anonymous

Applying this framework will provide a list of cybersecurity actors. From this list, the actors relevant to the domains and mandates can be identified. The availability and type of actors, with their specific skills influences the selection of national cybersecurity functions.

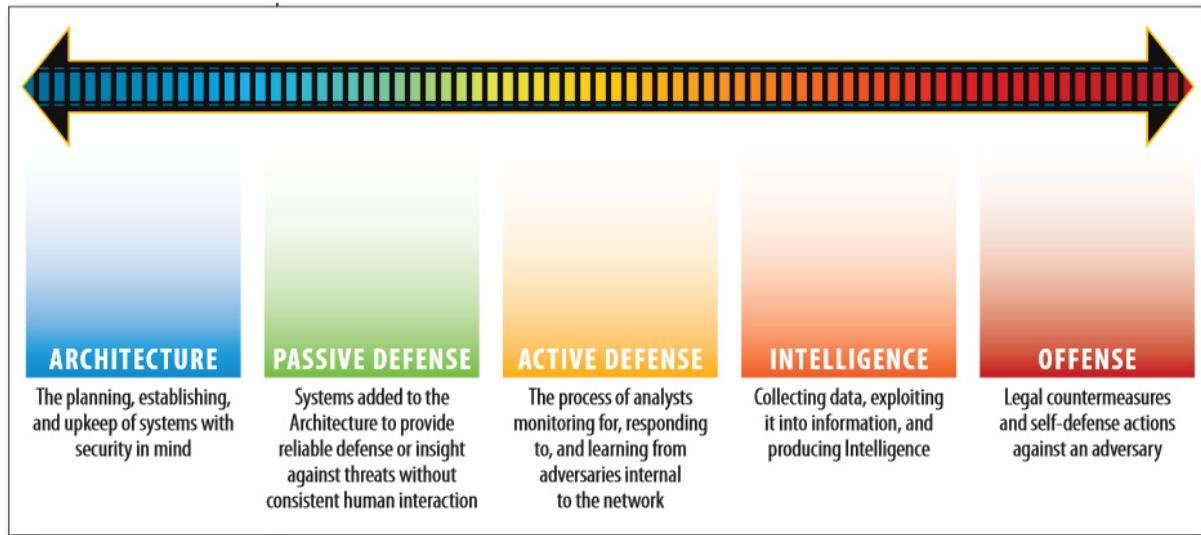
### 3.2 Domains

NATO identifies two domains in which cyber actions can take place. These are the offensive domain, and the defensive domain in cyber (NATO CCDCOE, 2012). The offensive domain is typically the responsibility of a nation states' army, or intelligence agency. Cyber defence is the action of defending organisational or national ICT assets against attacks, discovering attacks, and to respond to, and recover from attacks and cybersecurity incidents. The SysAdmin, Audit, Network, and Security Institute (SANS) proposes a sliding scale for cybersecurity. The sliding scale supports the notion of offensive and defensive domains introduced by NATO. The sliding scale is shown in Figure 3 as taken from Lee (2015), and its shows passive and active defence cycles, moving towards an offensive cycle.

The cybersecurity domains influence the selection and prioritisation of national cybersecurity functions, in that national cybersecurity functions needed by nation states, will differ depending on whether cybersecurity activities takes place in the offensive, or defensive domains. As an example, cybersecurity in the offensive domain could consider national cybersecurity functions such as the establishment of a Military Cyber function consisting of cyberwarfare services, whereas cybersecurity in the defensive domain could consider functions such as a National Incident Handling function which consists of incident response services.

In Section 2 the statement was made that the national cybersecurity functions are offered from national cybersecurity structures. In Section 3.1 we have identified national Actors. The defensive domain lifecycle phases together with the actors identified using the framework introduced in Section 3.1, and structures, are used to

develop a framework to be applied during the mapping of national cybersecurity functions to actors and structures.



**Figure 3:** Cybersecurity sliding scale (Lee, 2015)

The cybersecurity domains assist with the identification of national cybersecurity structures needed from where national cybersecurity functions will be offered from. Furthermore, selecting a cybersecurity domain assist with the identification of State and Government Actors, Organised Non-State Actors, and Non-Organised Non-State Actors associated with either the offensive or defensive domains. The NATO proposes four cyber defence lifecycle phases. The four lifecycle phases proposed by NATO (NATO Cooperative Cyber Defence Centre of Excellence, 2012) are echoed by the United States Nuclear Regulatory Commission (U.S. Nuclear Regulatory Commission, 2010), NIST (NIST, 2013), and the United States National Security Agency (NSA) (Shasha, 2002). Furthermore, it corresponds to the NIST incident handling lifecycle phases (Cichonski; Millar; Grance & Scarfone, 2012) (ITU, 2009) (ISO/IEC, 2011b). The four lifecycle phases proposed by NATO, and others, are (NATO Cooperative Cyber Defence Centre of Excellence, 2012) (Cichonski; Millar; Grance & Scarfone, 2012) (ITU, 2009) (ISO/IEC, 2011b):

- Protecting phase.
- Detection phase.
- Responding phase.
- Recovering phase.

Table 3 shows the proposed framework to map the defensive domain lifecycle phases to the national cybersecurity functions enabling them, as well as existing structures from where they are offered from. Table 3 shows that the “prevent and detect” lifecycle phases are enabled by the Monitoring and Evaluation function, and are offered from SOC structures with IT Industry, SSA and DOD as relevant actors. The “respond” lifecycle phase is enabled by the Incident Handling function, offered from CSIRTs, and the “recover” lifecycle phase can be offered from Critical Information Infrastructure Protection (CIIP) function offered by the Critical Infrastructure Providers (CIP) themselves.

**Table 3:** Defensive domain lifecycle framework

National Cybersecurity Function	Protect	Detect	Respond	Recover
Structure	SOCs	SOCs	National CSIRT ECS-CSIRT SOCs	CIP BCM
Actors	IT Industry, SSA, DOD	IT Industry, SSA, DOD	DTPS, SSA, DOD, DIRCO	CIP

### 3.3 Mandates

A mandate is a formal order, or provides someone with the authority to do something, or to behave in a certain way (Merriam-Webster, 2017). The five mandates - as taken from the NATO CCDCOE (2012) and adapted for an illustrative application in the South African context - are introduced next. The mandates are the responsibility of a Government department, or one Government department could be responsible for more than one mandate. The five cybersecurity mandates identified by the NATO CCDCOE (2012) are Military Cyber, Counter Cybercrime, Intelligence and Counter-Intelligence, Critical Information Infrastructure Protection (CIIP) and National Crisis Management, and Cyber Diplomacy and Internet Governance.

The national cybersecurity function selection framework is shown in Table 4. This is a combination of the information collected in Table 2 and Table 3, with the domains included. Table 4 represents the complete “national cybersecurity function selection framework”. The framework shows that for each mandate and domain, there need to be an actor and a cybersecurity function. The framework’s Cybersecurity Function column is populated with the generic national cybersecurity functions which would give effect to the mandates. The national cybersecurity functions were taken from Table 1, and the mapping of mandates to cybersecurity functions are done based on our experience. The national cybersecurity functions could reside in the offensive, or defensive domains, or in some instances reside in both domains.

**Table 4:** National cybersecurity function selection framework

Mandates	Generic Cybersecurity Function	Domain	Actors
Military Cyber	Military Cyber / Cyber Warfare	Offensive	Actor1 Actor2 Actor3
	Cryptography	Defensive	
	Research and Development (R&D), Education and Awareness	Defensive	
	National Strategic Risk and Threat Assessment	Defensive	
Counter Cyber Crime	Cybercrime / Investigations / Digital Forensics	Defensive	Actor1 Actor2 Actor3
	Cryptography	Defensive	
	Research and Development (R&D), Education and Awareness	Defensive	
	National Strategic Risk and Threat Assessment	Defensive	
Intelligence and Counter-Intelligence	Military Cyber / Cyber Warfare	Offensive	Actor1 Actor2 Actor3
	Monitoring and Evaluation	Defensive	
	Cryptography	Defensive	
	Research and Development (R&D), Education and Awareness	Defensive	
	National Strategic Risk and Threat Assessment	Defensive	
Critical Infrastructure Protection and National Crisis Management	Incident Handling	Defensive	Actor1 Actor2 Actor3
	Monitoring and Evaluation	Defensive	
	Critical Information Infrastructure Protection (CIIP)	Defensive	
	Cryptography	Defensive	
	Research and Development (R&D), Education and Awareness	Defensive	
	National Strategic Risk and Threat Assessment	Defensive	
	Internal Coordination	Defensive	
Cyber Diplomacy and Internet Governance	Research and Development (R&D), Education and Awareness	Defensive	Actor1 Actor2 Actor3
	E-Identity	Defensive	
	Internal Coordination	Defensive	
	External Stakeholder Engagement	Defensive	
	National Policy and Strategy Development	Defensive	
	National Regulations Development	Defensive	

### 3.4 Illustrative example

Table 5 illustrates the application of this framework in context of South Africa within the defensive domain and Critical Infrastructure Protection and National Crisis Management. The process followed are as follows:

- Identify the actors across all dimensions using the framework proposed in Section 3.1, Table 3.

- Select the domain – we have selected the defensive domain.
- Select the mandates – select from table 4.
- Identify national cybersecurity functions to give effect to the mandate.

In this illustrative application, it can be seen that the National Incident Handling function gives effect to the National Crisis Management mandate residing in the defensive domain – this relationship was found in Table 4.

Some mandates use the same cybersecurity functions to give effect to them, such as the Research and Development (R&D), Education and Awareness, as well as the National Strategic Risk and Threat Assessment functions which can be found across all mandates. Some mandates may also be found across the two domains, such as the Military Cyber, and Intelligence and Counter-Intelligence that have activities across the defensive, and offensive domains.

**Table 5:** Illustrative application of the national cybersecurity function selection framework

Mandates	Generic Cybersecurity Function	Domain	Actors
National Crisis Management	National Incident Handling Function  National Strategic Risk and Threat Assessment  Research and Development (R&D), Education and Awareness	Defensive during the Respond phase	Department of Defence (DOD) South African Police Service(SAPS) State Security Agency (SSA) Department of Postal and Telecommunication Services (DTPS) Cybersecurity Hub Department of Justice (DoJ) and SSA Security Contractors, Vendors (McAfee, Nanoteq, Microsoft) ISP's, Fixed and Mobile operators (MTN, Vodacom, CellC, Neotel) Service Providers (DeLoitte's, BCX, DiData) Sector-CSIRTs

#### **4. Combining ISO 27005 and NIST SP 800-39 for a national risk management framework**

We proposed to make use of a combination of ISO 27005, and NIST SP 800-39 (Furlani, 2011) to formulate a risk management framework which can be used, at, and scale to national level. In the following sections, the ISO 27005, and NIST SP 800-39 risk management processes are introduced.

##### **4.1 ISO 27005 and NIST SP 800-39 risk management process**

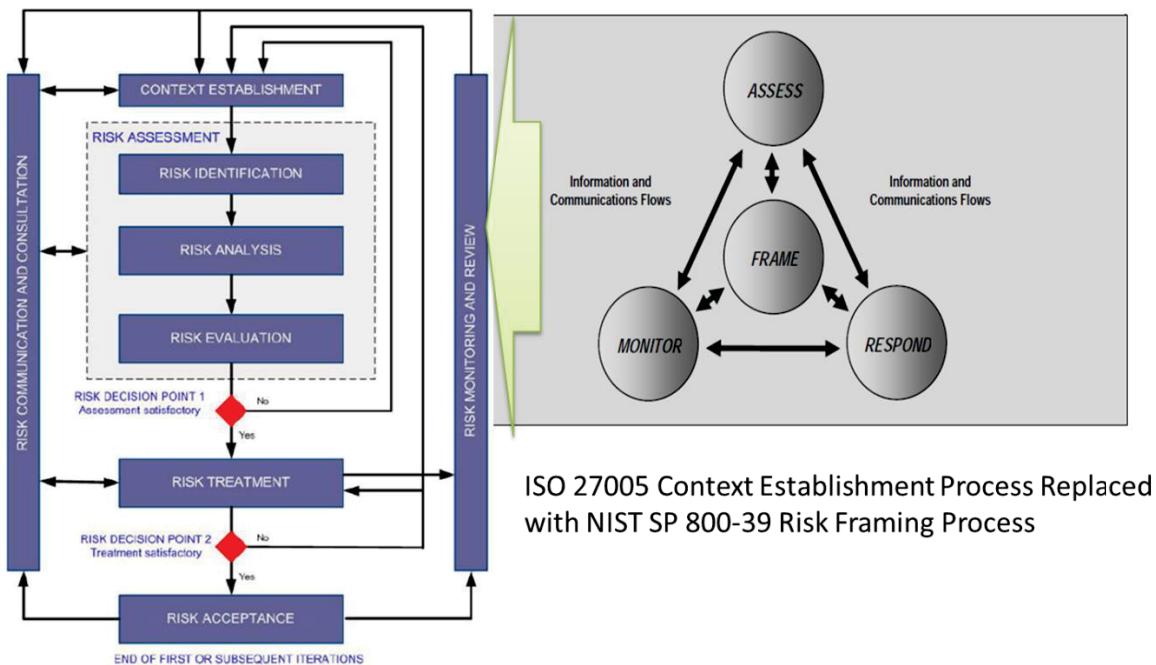
ISO 27005 describes four processes. These are Context Establishment, Risk Assessment, Risk Treatment and Risk Monitoring and Review (ISO/IEC, 2011a). The NIST SP 800-39 also describes four processes. These are Risk Framing, Assessing Risk, Risk Response and Risk Monitoring. The overlap between the risk management processes are shown in Table 6. Table 6 shows that there is an overlap between the NIST SP 800-39 and ISO 27005 processes.

**Table 6:** Process comparison between ISO 27005 and NIST SP 800-39

NIST SP 800-39	ISO 27005
Risk Framing	Context Establishment
Assessing Risk	Risk Assessment
Risk Response	Risk Treatment
Risk Monitoring	Risk Monitoring and Review

##### **4.2 Our proposed model: ISO 27005 and NIST SP 800-39 risk management process**

Figure 4 shows that our model proposes that the ISO 27005 Context establishment process be replaced with the NIST SP 800-39 Risk Framing process. During this process the scope and boundaries will be defined, and the organs managing information security risk is established. ISO 27005 establishes context at the information systems level, while NIST SP 800-39 frames risk at all three layers, allowing for a much wider scope. This also enables it to scale to national level.

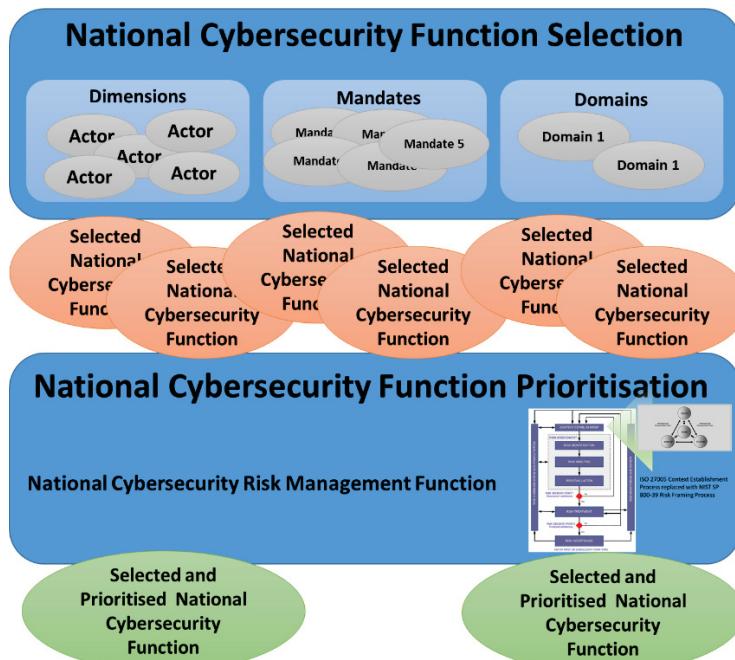


**Figure 4:** Application of NIST SP 800-39 and ISO 27005 (Furlani, 2011)

This Risk Framing Process will be tailored with the NIST SP 800-39 (Furlani, 2011) national view, addressing strategic risk at national level. This is done to accommodate the requirement for the risk management process to be generic enough to be used at national level. NIST SP 800-39 addresses organisational risk at national level risk. ISO 27005 being information security specific is used to determine national cybersecurity risk.

## 5. Presentation of the national cybersecurity function selection and prioritisation framework

The national selection and prioritisation framework is presented in Figure 5. Figure 5 shows that the dimensions may be used to identify national cybersecurity actors. The availability and skills type of national cybersecurity actors will influence the selection of national cybersecurity functions.



**Figure 5:** National cybersecurity function selection and prioritisation framework

The domains, offensive and defensive, further influences the selection of national cybersecurity functions, in that different national cybersecurity functions are needed for the different lifecycle phases found in the offensive and defensive domains.

Nation states would also select a national cybersecurity mandate. The mandate will have specific requirements in terms of national cybersecurity functions. Once the national cybersecurity functions are selected, a risk management approach is proposed to assist with the prioritisation of national cybersecurity functions for implementation. We have proposed a national risk management framework which is a combination of ISO 27005:2011 and NIST SP 800-39.

## **6. Future work**

In future work, the national cybersecurity function selection and implementation framework will be refined. The impact of available and type of actors during across the dimensions will be determined, and a framework and mapping against actors and generic national cybersecurity functions will be done.

## **7. Conclusion**

Jacobs, *et al* (2016) identified thirteen generic national cybersecurity functions, from which nation states can select those relevant to them for implementation. Due to cost and skills constraints, it is unlikely for nation states to implement all thirteen functions. To assist nation states to select and prioritise the most applicable national cybersecurity functions for implementation, a national cybersecurity function selection and prioritisation framework is proposed.

This proposed framework considers the dimensions to determine actors, operating domains, and political mandates, to assist nation states in selecting relevant cybersecurity functions. To prioritise these selected functions, we propose a national cybersecurity risk management framework developed by combining ISO/IEC 27005 and NIST SP 800-39, and by replacing the ISO 27005 Context Establishment process with the NIST SP 800-39 Risk Framing Process.

Using this proposed framework is beneficial in that nation states select and prioritise national cybersecurity functions which are relevant and needed according to their internal structures and political mandates. This aids in improving the national cybersecurity governance and posture. The selection and prioritisation of relevant national cybersecurity functions allows for nation states to optimise their available resources and increase the likelihood of a successful implementation.

## **References**

- Anonymous. (2016). Anonops. Retrieved February 23, 2016, from <https://anonops.com/>
- Christopher J.D. (2014). Cybersecurity Capability Maturity Model (C2M2) Program. Retrieved November 25, 2018, from <http://energy.gov/oe/services/cybersecurity/cybersecurity-capability-maturity-model-c2m2-program>
- Cichonski P.; Millar T. ; Grance T. & Scarfone K. (2012). *Computer Security Incident Handling Guide: Recommendations of the National Institute of Standards and Technology, 800-61. Revision 2. NIST Special Publication* (Vol. 800–61). <https://doi.org/10.6028>
- Department of Homeland Security. (2017). US-CERT | United States Computer Emergency Readiness Team. Retrieved December 20, 2017, from <https://www.us-cert.gov/>
- DOD. (2016). DOD. Retrieved February 23, 2016, from <http://www.dod.mil.za/>
- DTPS. (2015). Cybersecurity Hub. Retrieved February 18, 2016, from <https://www.cybersecurityhub.co.za/>
- FIRST. (2016). FIRST. Retrieved February 23, 2016, from <https://www.first.org/>
- Fox News. (2011). A Brief History of the LulzSec Hackers. Retrieved December 20, 2017, from <http://www.foxnews.com/scitech/2011/06/21/brief-history-lulzsec-hackers/>
- Furlani, C. (2011). Managing Information Security Risk: Organization, Mission, and Information System View. Retrieved December 19, 2018, from <http://csrc.nist.gov/publications/nistpubs/800-39/SP800-39-final.pdf>
- ISO/IEC. (2011a). ISO/IEC 27005:2011 Information technology -- Security techniques -- Information security risk management. Retrieved November 25, 2018, from [http://www.iso.org/iso/catalogue\\_detail?csnumber=56742](http://www.iso.org/iso/catalogue_detail?csnumber=56742)
- ISO/IEC. (2011b). ISO/IEC 27035:2011 Information technology — Security techniques — Information security incident management. Retrieved February 24, 2016, from <http://www.iso27001security.com/html/27035.html>
- ITU. (2009). ITU-T X.1056 - Security incident management guidelines for telecommunications organizations. Retrieved February 24, 2016, from <http://www.itu.int/rec/T-REC-X.1056-200901-I>
- Jacobs, P. C., von Solms, S. H., & Grobler, M. M. (2016). Towards a framework for the development of business cybersecurity capabilities, 7(4), 51–61. <https://doi.org/10.13140/RG.2.1.5110.0406>

- Merriam-Webster. (2017). Definition of mandate. Retrieved October 12, 2017, from <https://www.merriam-webster.com/dictionary/mandate>
- Merriam-Webster. (2017). Dimension. Retrieved October 22, 2017, from <https://www.merriam-webster.com/dictionary/dimension>
- NATO Cooperative Cyber Defence Centre of Excellence. (2012). National Cyber Security Framework Manual. Retrieved November 24, 2018, from <https://ccdcoc.org/publications/books/NationalCyberSecurityFrameworkManual.pdf>
- NIST. (2013). Improving Critical Infrastructure Cybersecurity Executive Order 13636: Preliminary Cybersecurity Framework. Retrieved November 25, 2018, from <http://www.nist.gov/itl/upload/preliminary-cybersecurity-framework.pdf>
- Roberts T. (2014). Cyber Security Capability Maturity Model (CMM) - Pilot. Retrieved February 18, 2016, from <https://www.sbs.ox.ac.uk/cybersecurity-capacity/content/cyber-security-capability-maturity-model-cmm>
- SABRIC. (2016). About Us. Retrieved August 15, 2016, from <https://www.sabric.co.za/about-us/>
- SAPS. (2016). SAPS. Retrieved February 23, 2016, from <http://www.saps.gov.za/>
- SARB. (2017). South African Reserve Bank. Retrieved December 20, 2017, from <https://www.resbank.co.za/Pages/default.aspx>
- Shasha, D. E. (2002). Defense in Depth. <https://doi.org/10.1038/scientificamerican0502-101>
- South African Government. (2017). Department of Telecommunications and Postal Services. Retrieved December 20, 2017, from <https://www.dtps.gov.za/>
- SSA. (2016). SSA. Retrieved February 23, 2016, from <http://www.ssa.gov.za/>
- U.S. Nuclear Regulatory Commission. (2010). Cyber Security Programs for Nuclear Facilities. Retrieved February 23, 2016, from <http://nrc-stp.ornl.gov/slo/regguide571.pdf>
- Wamala, F. (2011). ITU National Cybersecurity Strategy Guide. *Chemistry & ...*, 122. <https://doi.org/10.1017/CBO9781107415324.004>

# Information Security Management Scaffold for Mobile Money Systems in Uganda

Fredrick Kanobe<sup>1</sup>, Margaret Patricia Alexander <sup>1</sup>and Kelvin Joseph Bwalya<sup>2</sup>

<sup>1</sup>Informatics, ICT, Tshwane University of Technology, Pretoria, South Africa

<sup>2</sup>Information and knowledge management, Business and economics, University of Johannesburg, South Africa

[fkanobe2010@gmail.com](mailto:fkanobe2010@gmail.com)

[alexandarMP@tut.ac.za](mailto:alexandarMP@tut.ac.za)

[bwalyakelvinjoseph@gmail.com](mailto:bwalyakelvinjoseph@gmail.com)

**Abstract:** Mobile money systems are widely accepted in Uganda as an easy way to transfer money and to settle domestic financial matters. However, although these systems play a critical role in bridging the financial inclusion gap, several oversight issues need to be addressed. Previous mobile money systems security studies focussed on technical applications and solutions paying less attention to subjective Information security management. The current study sought to understand information security management for mobile money systems using Uganda as a case study in order to develop an information security management framework suitable for mobile money systems in Uganda. Specific objectives included a detailed study of existing information security policies, procedures and standards, investigating and determining their weaknesses, developing and recommending a suitable framework and validating that framework. The case study involved three mobile network operators. Activity Theory guided the study throughout. Management of information security in mobile money systems was easy to understand when investigated as activities and allowed contradictions surrounding mobile money systems to be highlighted. The data collection methods used were semi-structured interviews and an internal documents review. The findings of the study revealed that there were insufficient tools, rules, community and division of labour for information security awareness related to outsourcing, risk management, business continuity planning and incident management. Furthermore, there appeared to be inadequate compliance monitoring, management controls and top management support for mobile money information security activities. The study contributes to theoretical, methodological, body of knowledge in information security management, practice and new areas of future research in information systems security for mobile money systems. In conclusion, the rules, tools, community and division of labour employed by the subjects (MNOs) to attain the intended objects and outcomes of the identified activities were found to be wanting and this indicates that continuous review and updating is needed. Mobile money systems and the associated activities, like any other information systems, are dynamic and require continuous updates. The PDCA (Plan, Do, Check, Act) approach to mobile money information security management activities is recommended for addressing information security management concerns for mobile money systems in Uganda.

**Keywords:** mobile, money, information, security, activity theory, Uganda

---

## 1. Introduction

In 2014, 41% of the districts in Uganda lacked access to any bank branch (BOU, 2014) and there are only 0.3 commercial bank branches serving every 10,000 adult Ugandans (BOU, 2017). In emerging economies such as Uganda, where the majority of the people have limited access to formal financial services, mobile money systems remain the most feasible way to counter financial exclusion. Mobile money to a large extent involves remittance of electronic money that can be accessed via a mobile phone. The speed at which mobile money systems are spreading in the country is remarkable; for example the value of mobile money transactions increased from 52.7 trillion Uganda shillings to 73.2 trillion Uganda shillings during the financial year 2017/2018 (BOU, 2018:66). Unfortunately, the rapid spread of mobile money systems in Uganda is threatened by a fragile mobile money regulatory foundation and a 'light touch' information security policy framework. Castle, Pervaiz and Weld (2016) stress that where there is money, there *must* be security. Certainly, to safeguard mobile money use in Uganda, a strong information security management framework is called for. Users of any type of financial system are clearly sensitive about the privacy of their transaction information and the overall safety of their money but Harris, Goodman and Traynor (2013) reveal that the introduction of mobile money systems in East Africa was not accompanied by adequate concern for privacy.

In Uganda, mobile money remittance includes: person-to-person money transfer and payments; customer-to-business payments for services; and business-to-business payments (Ernst and Young 2009). Ndiwalana, Morawczynski and Popov (2014) maintain that the person-to-person is the dominant category in Uganda used for sending electronic money from the town dwellers to relatives in rural areas. Other uses include paying for

utilities (such as TV, water, electricity), transport, school fees, medical costs; fuel; wages; church offerings; fuel for vehicles; presents and gifts.

This use has been facilitated by mobile money account opening procedures that allow persons without a formal financial history to open a mobile money account, unlike at conventional banks that demand enhanced Know Your Customer (KYC) information. Mobile money systems have virtually eliminated some of the problems associated with physical payment systems including standing in long queues at banks, making stressful and long journeys to the banks and avoiding the security risk of carrying physical money. Despite the benefits for mobile money systems, in Uganda they operate in a context in which the mobile money providers are new and inexperienced regarding electronic financial service provision and the boundaries of operation and interests overlap due to the current minimal regulatory and information security policy framework.

## **2. Motivation**

This study was motivated by information security management concerns relating to mobile money systems in Uganda. Previous studies in this topic focused on technical tools (objective security) and little attention was given to information security management (subjective security). Technical tools alone cannot solve the information security problem (Zahoor, Mahmood and Javed 2016; Safa, Sookhak, Von Solms, Furnell, Ghani and Herwan 2015; Kayworth and Whitten 2010). The study also intended to contribute to the body of knowledge to the information security management in emerging economies where information security management is still in infancy.

Merkow and Breithaupt (2014) contend that “when people are left on their own they tend to make the worst security decisions”; hence the necessity exists for the development of an information security management framework both to make a contribution to information systems security research and to guide the practice of information security management of mobile money systems in Uganda.

## **3. Study purpose and contribution**

The gap identified by this study was the lack of an adequate framework to address the mobile money information security concerns in Uganda. The specific questions addressed by the study are: What are the information security management policies, procedures and standards for mobile money systems in Uganda? What are the weaknesses in the existing information security management policies, procedures and standards for mobile money systems in Uganda? Why are there weaknesses in the information security management policies, procedures and standards for mobile money systems in Uganda? How should the weaknesses in the information security management policies, procedures and standards for mobile money systems in Uganda? How will the information security management framework for mobile money systems developed be validated? The study therefore, purposed to develop a suitable information security management framework for mobile money systems in Uganda.

Study participants from different fields and backgrounds were interviewed and literature comprehensively reviewed to enable us get in-depth understanding of the study phenomenon. The contribution of the study was fivefold: Theoretical contribution to information security management body of knowledge, Contribution to Information security theory, Study contribution to methodological approach, Contribution to practice of information systems security in MNOs and New areas of future research.

## **4. Literature review**

Though mobile money was first introduced in Philippines in 2001 (Scharwatt, Katakam, Frydrych, Murphy and Naghavi 2015), East Africa is the global leader not only in the number of mobile money systems but also in innovations of the new technology and as a global base for knowledge sharing and learning about mobile money (GSMA, 2013). The success story of M-PESA (mobile money) in Kenya has inspired many others countries. In Uganda, the mobile money ecosystem is complex, with various stakeholders (mobile money agents, mobile money customers, Bank of Uganda, Uganda Communications Commission, financial institutions and mobile network operators) (Uganda mobile money guidelines, 2013). This complex ecosystem creates a chain of stakeholders through which mobile money passes before it is ultimately transferred to the end-user. Certainly, the more stakeholders (especially mediating or interim parties), who can access mobile money information, the higher the risk of abuse. It is imperative to have strong information security management policies, procedures and controls to protect the mobile money system within the data-rich mobile money ecosystem.

Initially, mobile money systems were introduced in Uganda to perform basic services, such as buying airtime, making small money deposits and withdrawals, sending and receiving mobile money from other users and buying data bundles. However, currently these systems are the primary means of money transfer and saving and dominate the financial services in the country (more especially in rural areas where the majority of the population lives). This set of services is growing at a rapid rate, challenging the banking systems in Uganda while at the same time providing economic opportunities for new stakeholders, but it is also raising concerns about the privacy and security of the increasing volumes of financial information (Ndiwalana, Morawcynski and Popov 2014). As shown in Table 1, various mobile money services exist in Uganda (Ssetimba 2016).

**Table 1:** Mobile money services in Uganda (Ssetimba, 2016)

Mobile Money Service	Status
Domestic transfers/remittances (P2P)	Live
Merchant payments – enabling corporates to receive payments (P2B)	Live
Statutory payments (taxes) person to government (P2G)	Live
Bulky payments: Salaries and wages. Business to Person (B2P)	Live
Micro-loans and savings	Pilot
Group wallets for SACCOs	Pilot
Cross-border transfers	Live
Mobile banking – transfer from bank account to m-wallet	Live
Government payments (Social benefits) Government to Person (G2P)	Live

As is evident in Table 1, most of these services are in full use (Live). Only the micro-loans and savings, and Group wallets for SACCOs, are still undergoing testing.

The characteristics of mobile money systems require a strong information security management framework to protect mobile money information from abuse. Its simplicity, universality, dependence on trust, fast transactions and Interoperability (open technology allowing systems to interact with each other) necessitates strict information security management policies and procedures to minimize misuse by the stakeholders (Goyal, Pandey and Batra 2012). Amazingly, the most dominant mobile money type in emerging economies such as Uganda is Mobile Network Operator Led (Atanu, Kwon and Gill 2014), that is, is operated by mobile network operators (MNOs) whose backgrounds and experience are not strong in terms of electronic financial services. Traditionally, the MNOs concentrated only on providing the national backbone for telecommunication services, hence the new role is built on a weak foundation for information security of mobile money systems.

In this study, a limited number of theories for information systems security were identified. Activity Theory (AT) was found to be most suitable to underpin the study. AT has previously been used in many studies in which human interaction and use of technology are the focus (recent examples are Mugarura, Rivett & Blake, 2016; Nihra, Mohd, Fadzli & Noor, 2014; Bardram & Doryab, 2011) as compared to other theories such as Integrated Systems Theory and General Deterrence Theory. Mobile money systems similarly involve human interaction and use of technology.

## 5. Research methodology

### 5.1 Research design

The study adopted the interpretivist research paradigm – knowledge is produced through exploring and understanding the social world by focusing on the meaning and multiple interpretations of the related events. The researchers interacted with the study participants who had varying experience and backgrounds and needed to interpret their understanding of the research problem as well as their explanations, descriptions of current practices and problems and their suggested improvements to the status quo. A qualitative case study was undertaken in Uganda involving three mobile network operators (MNOs). Mobile money information security being a contemporary phenomenon in Uganda, a case study helped the researchers work within its natural setting using a variety of data methods and techniques in order to get an in-depth understanding of the problem. The research sites (MNOs) were selected basing on their possession of a license to provide mobile money services, possession of an active mobile money platform in Uganda, network coverage of 50% of the districts and more than five years of experience. The researchers needed to conduct a comprehensive study in order to obtain rich data to inform a new area of study. Therefore, only MNOs with adequate experience and expertise in the area of study were selected. For purposes of confidentiality due to the sensitivity data collected the true

names of the MNOs were not reflected in the study hence the codes MNO1, etc. In conducting the study, the researchers adopted Yin's (2011) approach in carrying out case study as summarized in Table 2.

**Table 2:** Case study approach (Yin 2011)

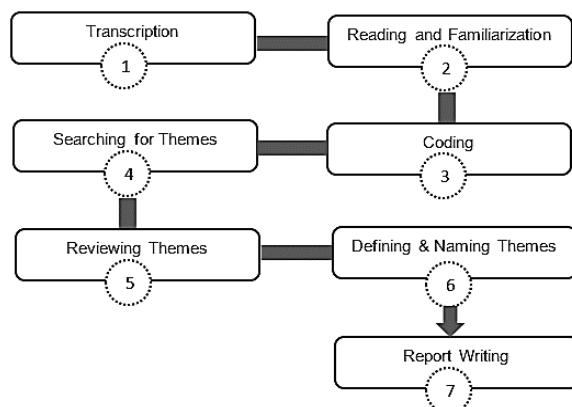
Process	Description
Planning	Compose suitable the research questions that were addressed by the study
Design	Identify possible cases that could studies, studies possible unit of analysis and possible study strategy
Prepare	Build research capacity, coordinating and communicating with study participants
Collect	Collect the data and compile it as adequate proof of the study investigation
Analyse	Organise, clean, transcribe, code, analyse and interpret the data using appropriate analytical methods
Share	Generate the study findings and validate them

## 5.2 Data methods

The recruitment of the research participants followed steps proposed by (Asiamah, Mensah and Oteng-abayie 2017) namely; participants who had attributed meeting the research goal were identified (sampling frame) through human resource office, participated who failed to meet some criteria (experience, expertise, job roles) were dropped, participated who qualified but not ready to participated were removed (informed consent) and finally fully qualified participants formed the sample. Expert purposive sampling was used in order to get an in-depth understanding of the problem and potential improvements from informed participants. The study triangulated data from various data sources using different data collection techniques. 18 participants were interviewed and included IT managers, Audit managers, Finance managers, legal managers, and security and compliance managers. Both primary and secondary data were collected to inform the study. Face-to-face semi-structured interviews formed the basis of the primary data. Interview tools were tested, approved and permission obtained from both study sites and participants. Face to face interviews were conducted with selected participants and their responses transcribed into NVivo 11 plus application for analysis. Internal documents formed the secondary data and these included; mobile money guidelines, user manuals, policy statements, procedures, reports, newsletters, minutes of meetings, mobile money application forms, vendor contracts. Internal documents were reviewed and contents analysed to validate responses obtained from interviews.

## 5.3 Data analysis procedures

Data for the study was analysed thematically and mapped to Activity Theory nodes. Thematic inductive data analysis was used because it is flexible, not aligned to any particular theory or paradigm and appropriate for analysing complex data in-depth (Braun & Clarke 2013; Guest, Macqueen & Namey, 2012). In a new area of research, such as the investigation of information security management of mobile money systems in Uganda, a flexible analysis technique was needed. Percy et al. (2015) suggested that thematic inductive data analysis is suitable for case studies. This fits this study because it involved conducting a case study of MNOs regarding mobile money information security management in Uganda. This study adopted the steps for thematic inductive data analysis recommended by Braun and Clarke (2013) because they are systematic and comprehensive (see Figure 1).



**Figure 1:** Summary of data analysis steps used in the study (adopted from Braun and Clarke 2013)

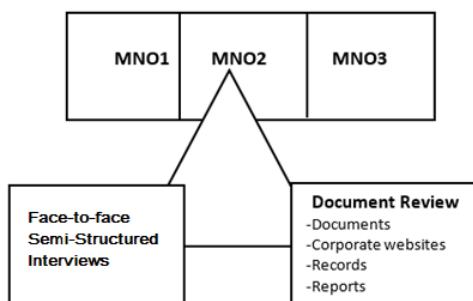
Table 3 describes the steps used in the data analysis of interview data for this study.

**Table 3:** Description of the processes of data analysis used in this research

Transcription	Interviewees' responses were reviewed, organised and prepared in a systematic format before input NVivo application for analysis
Reading and familiarization	This involved the researchers in reviewing the interview transcripts one by one to understand their depth, meaning and information flow.
Coding	The researchers scrutinised data to identify categories by marking similar passages of text with a code label for easy retrieval and further comparison and analysis. Open coding was used as the initial coding scheme though in vivo coding was applied where no descriptive code could be used for a particular phase from the dataset.
Searching for pattern of patterns in data	The researchers Identified patterns in codes leading to the emergence of themes. This was done using NVivo data analysis functions such as matrix query and word frequency.
Reviewing themes	The researchers reviewed candidate themes across the entire dataset and unclear ones further refined to come up with the most satisfying ones.
Defining and naming themes	Refined themes were named and examined individually in relationship to the research questions of the study.
Reporting research findings	Following the naming of themes the study findings were reported according to the study objectives and Activity Theory principles

#### 5.4 Validation of the study

In conducting the investigation of the study, the researchers were aware of qualitative research quality assurance issues. The validation of the research was achieved by soliciting further information during interviews, collection of rich data from different sources, peer assessment of the study, a research participants' validation exercise, keeping the research participants focused, expert assessment and triangulating research sites and data methods. Figure 2 illustrates the triangulation of data the research sites and data methods.



**Figure 2:** Triangulation of research methods and data sites

#### 6. Findings and discussions

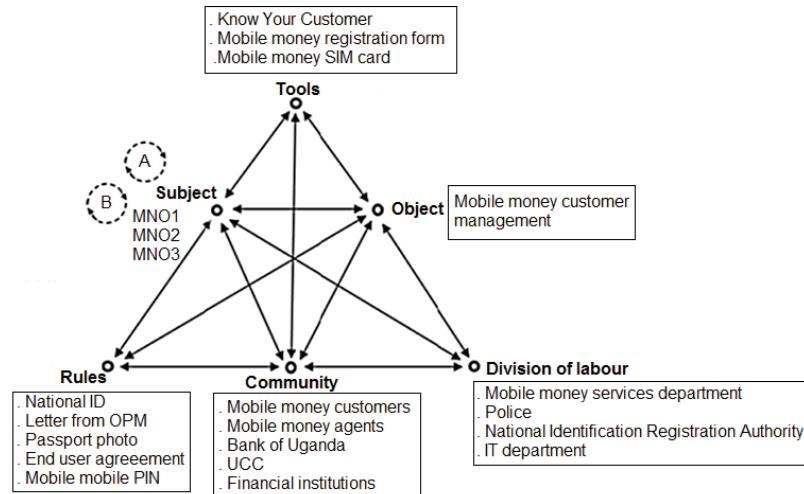
An exploratory interpretivist approach was used for this study. The findings were related to the activities that clearly emerged from the data analysis using an activity system as unit of analysis. Using AT, the study findings revealed contradictions within and among the activities, that is, primary and secondary contradictions. The AT nodes (tools, subject, rules, community, division of labour, objects and outcome) provided anchor for the findings. Engeström (1987) argues that contradictions in a system highlight the dynamic nature of the activity system, its inefficiencies and hence the opportunities for action and change.

The primary contradictions uncovered are illustrated in Figure 3.

The primary contradictions are summarized in Table 4.

In both scenarios the mobile money system administrator threatens the system from an internal position. In most situations damage inflicted by insiders to an information systems is more severe than that from outsiders. To minimize inside threats, tight information security policies, awareness programs, incident response plan, regular audits and procedures for disciplinary action are recommendable (SIFMA, 2018).

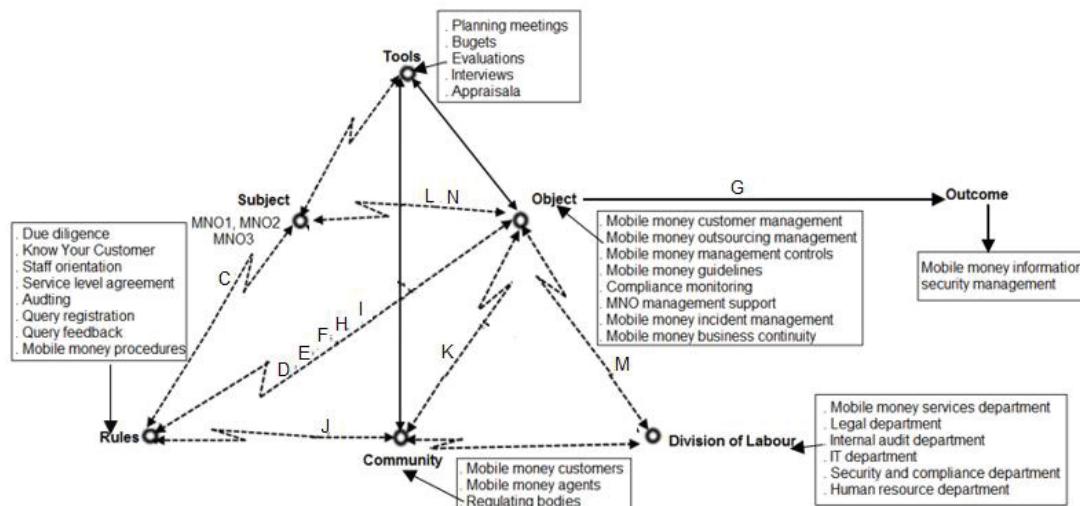
The identified secondary contradictions of the study are presented in Figure 4 and are described in greater detail in Table5.



**Figure 3:** Primary contradictions in the study (Source: Primary data)

**Table 4:** Description of the primary contradictions found in the mobile money system

Scenario	Contradictions
A	The MNOs' mobile money application administrators are simultaneously end-users of the mobile money system they are administering. Operating as system administrators and users of the same system requires tight policies to minimise system abuse (subject vs. subject). System abuse in situations where duties are separated becomes more difficult because it requires cooperation by different employees
B	Mobile money system administrators report mobile money abuse to MNO management. However, mobile money system administrators are also end-users of the system so reporting themselves remains a challenge and increases risks of mobile money abuse by MNOs (subject vs. subject). Humphreys (2008) note that the separation of duties prevents having a combination of functions under one person



**Figure 4:** Secondary contradictions (Source primary and secondary data)

**Table 5:** Description of the secondary contradictions found in the mobile money system

Scenario	Contradictions
C	The mobile money customers' registration is mandatory for all users of mobile money systems but at the same time the mobile money system still has an option for unregistered mobile money users hence proving opportunity for unregistered mobile money users to access the system (Tools Vs. rules). Existence of unregistered option on the mobile money application menu reveals insufficient controls. Ashenden (2008) argues that convincing all employees to comply with controls results in successful and effective information security management.

Scenario	Contradictions
D	The rules followed to screen outsourced mobile money agents are insufficient to ensure the maximum possible safety of mobile money information (rules vs. object). Mobile money information is most handled by the mobile money agents (third parties), therefore it necessitates tight security rules in the outsourcing environment.
E	The mobile money agents' service level contracts are not comprehensive enough to address information security management concerns in mobile money system (rules vs. object). Insufficient rules in contracts open doors for mobile money service providers for abusing the system. To minimize mobile money system abuse by third parties, it necessitates to have strict rules governing the contracts. Ghana mobile money guides (2015) present tight rules that can be used in contracts between MNOs and third parties such as mobile money service providers.
F	There was evidence of inadequate rules followed to raise information security awareness (rules vs. object). MNOs information security awareness campaigns are limited to a few tools and tools such as occasional email alerts and SMS. Using a variety of rules and tools in raising information security awareness in the mobile money ecosystem is important because where one rule is not effective another can do. Eyadat (2015) urges for multiple information security awareness rules and tools in alerting, training, educating and informing system users' issues concerning information security.
G	Information security management has not been integrated into the mission and vision of the MNOs (object vs. outcome). Lack of integration of information security in MNOs mission and vision portrays top management low priority to it. Whereas integrating information security in mission and vision of MNOs makes all staff accountable and responsible for their behaviour. Good information security management is managed from top and requires connection between business objectives and information security (ISO/IEC 27001: 2013)
H	There was evidence of insufficient management information security controls which are expected of a system that involves various money transactions (rules vs. object). Whereas customers' identifications are mandatory at mobile money account opening, no customers' identities are verified during cash withdrawals hence opening doors for system abuse. Both the Tanzania Mobile Money Guidelines (2015) and Ghana Mobile Money Guidelines (2015) have provisions for the verification of customers' identities at the time of cash withdrawal
I	There was a lack of a comprehensive information security policy that defines the 'dos and don'ts' of mobile money information security and sanctions for failing to comply with the guidelines (rules vs. object). It is difficult to prevent critical information assets from abuse without adequate information security policies (ISO/IEC 27001, 2013). The Nigeria mobile money guidelines (2015) provide a comprehensive information security management policy entailing privacy policy, incident response policy, information ownership policy, data confidentiality policy, disclosure and loss of mobile money data policy. Protecting mobile money information in the mobile money ecosystem requires all-inclusive information security policy.
J	There was limited segregation of duties in the division of labour in the MNOs. For example, IT manager also executes the duties of information security personnel which can lead to system compromise (rules vs. division of labour). Chances of system abuse are widen multiple tasks for different departments are executed by one personnel. Humphreys (2008) argues that the separation of duties prevents having a combination of functions under one person
K	There was evidence of inadequate compliance monitoring of the outsourced mobile money agents (third parties) leaving mobile money customers' information susceptible to information security abuse (object vs. community). The study revealed that compliance monitoring is confined to urban centers yet majority of mobile money agent live in rural areas where receipts of mobile money reside. Compliance monitoring is necessary to find out whether agents are adhering to the relevant guidelines and rules governing mobile money. Compliance monitoring should be part of information security management (ISO/IEC 27001:2013).
L	There was evidence of insufficient management support from MNOs top management in promoting information security awareness (subject vs. object). When staff are known to be aware of their information security roles and responsibilities, it is easy to blame them for their actions. Information awareness campaigns that are spearheaded by top management are most likely to yield positive results because subordinate staff and stakeholders tend to pay great attention to top management directives (ISO/IEC 27001, 2013).
M	There was evidence of inadequate timely management of mobile money security incidents among the MNOs (object vs. division of labour). The study revealed that time frame for reporting and responding to mobile money security incidents was not specified. Some incidents when not reported and responded to in time leave great damage. Ghana mobile money guidelines 2015 provide timeframe for reporting and responding to mobile money incidents.

Scenario	Contradictions
N	There was evidence of insufficient top management support for mobile money business continuity management (subject vs. object). The study uncovered that business continuity management for mobile money systems rests mainly in the hands of technical personnel of MNOs. Business continuity for mobile money systems is an activity that requires participation and support of top management right from planning, facilitation and its maintenance. Järveläinen (2013) urges that absence of reliable business continuity leads to organisation reputational damage and negatively impacts on customer royalty.

## 7. Conclusions and recommendations

This study had the main objective to develop a suitable information security management framework for mobile money systems in Uganda. The framework developed contains recommendations regarding tools, subjects, rules, division of labour and objects. Activity Theory was found to be useful in guiding the research that involved the use of technology (mobile money systems) and human interaction. The study revealed that, in order to be successful, information security management must be recognised as a continuous process that requires a multi-faceted approach. Basing on contradictions identified during the investigations, recommendations stress the need to improve the tools, rules, community, division of labour and objectives as pointed out in the proposed framework to achieve the desired outcome. When adopting the developed framework, the Plan, Do, Check, Act (PDCA) approach to information security management is recommended. As can be seen in Figure 5, the framework uses the activity system structure for the overall information security management for mobile money activity. Each of the elements of this activity system are shown in the framework to have a large number of associated items that are specific to this overall activity.

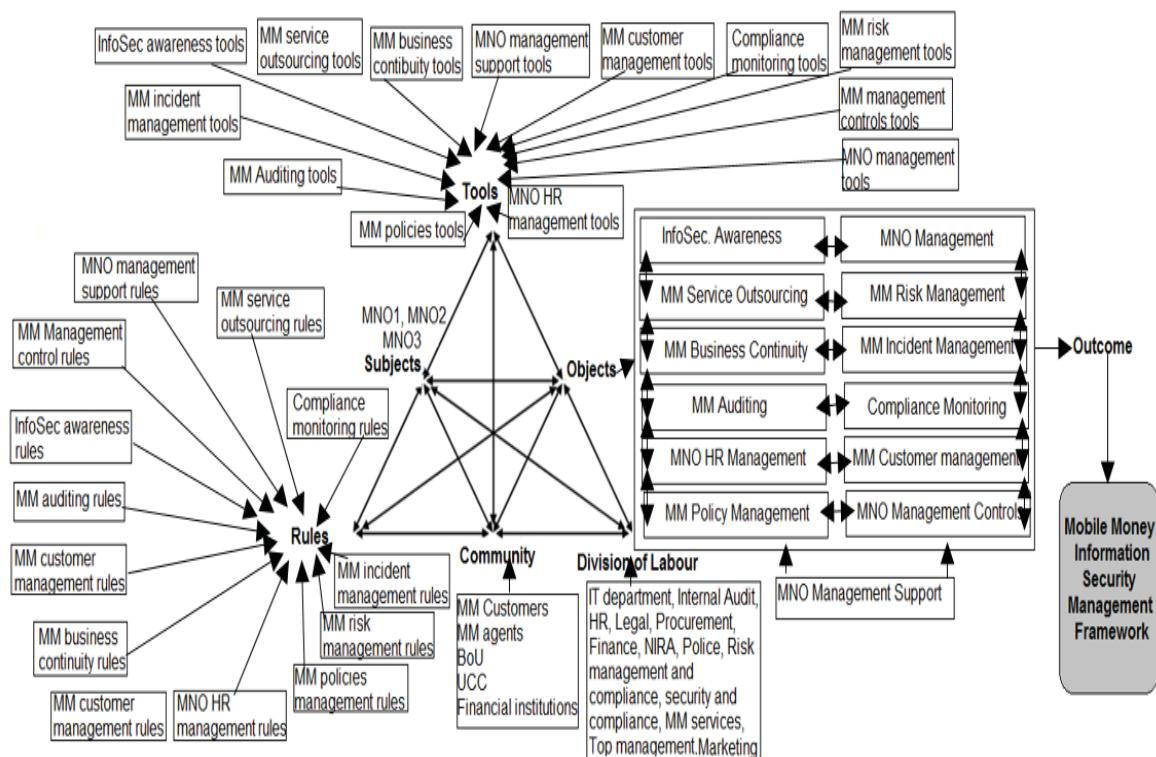


Figure 5: Recommended Information security management system framework for mobile money systems

## References

- Ashenden, D. 2008. Information Security management: A human challenge? *Information security technical report* Vol. 13 pages 195-201, 2008
- Asiamah, N., Mensah, H.K. & Oteng-Abayie, E.F. (2017). General, Target, and Accessible Population: Demystifying Concepts for Effective Sampling. *The Qualitative Report* 22 (6) 1-7
- Atanu, D., Kwon, H. & Gill, R. 2014. Mobile money: Opportunities for Mobile Operators, Business & Network Consulting Division, Huawei White paper, 2014.
- Bardram, J. & Doryab, A. 2011. Activity Analysis – Applying Activity Theory to Analyse Complex Work in Hospitals. CSCW 2011, March 19–23, 2011

- Bou (2014): Financial inclusion: Report on the state of financial inclusion in Uganda. [https://www.bou.or.ug/bou/bou-downloads/Financial\\_Inclusion/Report-on-the-State-of-Financial-Inclusion-First-Edition-March-2014.pdf](https://www.bou.or.ug/bou/bou-downloads/Financial_Inclusion/Report-on-the-State-of-Financial-Inclusion-First-Edition-March-2014.pdf) (Accessed 12/04/2018)
- Bou (2017): Uganda national financial inclusion strategy 2017-2011 [https://www.bou.or.ug/bou/bou-downloads/publications/special\\_pubs/2017/National-Financial-Inclusion-Strategy.pdf](https://www.bou.or.ug/bou/bou-downloads/publications/special_pubs/2017/National-Financial-Inclusion-Strategy.pdf) (Accessed 24/12/2018)
- Bou (2018): Bank of Uganda annual report. [https://www.bou.or.ug/bou/bou-downloads/publications/Annual\\_Reports/Rptrs/All/Bank-of-Uganda-Annual-Report-2018.pdf](https://www.bou.or.ug/bou/bou-downloads/publications/Annual_Reports/Rptrs/All/Bank-of-Uganda-Annual-Report-2018.pdf) (Accessed 06/02/2019)
- Braun, V. & Clarke, V. 2013. Successful qualitative research: A practical guide for beginners. Thousand Oaks, CA: Sage.
- Castle, M., Pervaiz, F. And Weld, G. (2016) Let's Talk Money: Evaluating the Security Challenges of Mobile Money in the Developing World. Proceedings of the ACM Symposium of Computing for Development (DEV). [https://homes.cs.washington.edu/~anderson/papers/2016/castle\\_dev2016\\_preprint.pdf](https://homes.cs.washington.edu/~anderson/papers/2016/castle_dev2016_preprint.pdf) (Accessed 28/09/2018).
- Engeström, Y. 1987. Learning by expanding: An activity-theoretical approach to developmental research. Helsinki: Orienta-Konsultit
- Ernst & Young (2009). Mobile Money Payments. Mobile Money Payments [http://www.ey.com/Publication/vwLUAssets/Mobile\\_Money./\\$FILE/Ernst%20&%20Young%20%20Mobile%20Money%20-%202015.10.09%20\(single%20view\).pdf](http://www.ey.com/Publication/vwLUAssets/Mobile_Money./$FILE/Ernst%20&%20Young%20%20Mobile%20Money%20-%202015.10.09%20(single%20view).pdf) [Access on 02/03/2016]
- Eyadat, M.S. 2015. Higher Education Administrators Roles in Fortification of Information Security Program. *Journal of academic administration in higher education*, 11(2) pp 61-65, 2015.
- Ghana Mobile Money Guidelines (2015). Bank of Ghana; Guidelines for E-Money Issuers in Ghana. <https://www.bog.gov.gh/privatecontent/Banking/E-MONEY%20GUIDELINES-29-06-2015-UPDATED5.pdf> (Accessed on 22/03/2017)
- Goyal, V. Pandey, U.S. & Sanjay, B. (2012). Mobile Banking in India: Practices, Challenges and Security Issues. International Journal of Advanced Trends in Computer Science and Engineering, 1(2), pp 56-6 2012
- GSMA (2013). The mobile Economy. <https://www.gsma.com/newsroom/wp-content/uploads/2013/12/GSMA-Mobile-Economy-2013.pdf> (Accessed 15/12/2017)
- Guest, G., Macqueen, K.M. & Namey, E.E. 2012. Introduction to applied thematic analysis, SAGE publications, California, United States.
- Harris, A., Goodman, S. & Traynor, P. 2013. Privacy and Security Concerns Associated with Mobile Money Applications in Africa. Washington Journal of Law, Technology and Arts, 8(3).
- ISO/IEC 27001:2013: A practical guideline for implementing an ISMS in accordance with the international standard. [https://www.isaca.de/sites/pf7360fd2c1.dev.team-wd.de/files/isaca\\_2017\\_implementation\\_guideline\\_isoiec27001\\_screen.pdf](https://www.isaca.de/sites/pf7360fd2c1.dev.team-wd.de/files/isaca_2017_implementation_guideline_isoiec27001_screen.pdf) (Accessed on 20/02/2018)
- Jarvelainen, J. 2013. IT incidents and business impacts: Validating a framework for continuity management in information systems. International Journal of Information Management, 33(3), 583-590
- Kayworth, T. & Whitten, D. 2010. "Effective Information Security Requires a Balance of Social and Technology Factors," MIS Quarterly Executive 9(3), pp 163-175
- Merkow, M.S. & Breithaupt, J. (2014). Information Security: principles and practices. 2nd Ed. New York, USA, Pearson
- Mugarura, F.S., Rivett, U. & Blake, E. 2016. Using Activity Theory to understand Technology use and perception among rural users in Uganda. In the proceeding of the 8<sup>th</sup> International conference on and information and communication technologies and Development (p.23) ACM <https://people.cs.uct.ac.za/~edwin/MyBib/2016-ICTD.pdf> (Accessed 15/09/2017)
- Ndiwalana, A., Morawcynski, O., & Popov, O. (2014). Mobile Money Use in Uganda: a Preliminary Study. pp 8-10. <https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2012/06/m4dmobilemoney.pdf> (Accessed 25/06/2017)
- Nigeria Mobile Money Guidelines (2015). Guideline on Mobile Money Services in Nigeria. <https://www.cbn.gov.ng/out/2015/bpsd/guidelines%20on%20mobile%20money%20services%20in%20nigeria.pdf> (Accessed on 20/04/2018)
- Nihra, M.N.H, Mohd, T.L., Fadzli, M. & Noor, M. 2014. Using Activity Theory as Analytical Framework for Evaluating Contextual Online Collaborative Learning. International Journal of Emerging Technologies in Learning. 9, 5, 54–59
- Safa, N.S., Sookhak, M., Von Solms, R., Furnell, S., Ghani N. A. & Herawan, T. 2015. Information security conscious care behaviour information in organizations. Computers & Security 53, 65-78
- Scharwatt, C. Katakam, A. Frydrych, J. Murphy, A. & Naghavi, N. 2015. State of the Industry Mobile Financial Services for the Unbanked. GSMA Report. [https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2015/03/SOTIR\\_2014.pdf](https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2015/03/SOTIR_2014.pdf) (Accessed 23/04/2017)
- SIFMA, (2018). Insider attack best practices guide. <https://arxiv.org/pdf/1805.01612.pdf> (Access 12/12/2018)
- Ssetimba, I.J. 2016. Mobile money in Uganda. <https://www.theigc.org/wp-content/uploads/2016/03/3.-Ivan-Ssettimba-Bank-of-Uganda.pdf> (Accessed on 09/12/2017).
- Uganda Mobile Money Guidelines (2013): Bank of Uganda Mobile Money Guidelines 2013 <https://www.ucc.co.ug/files/downloads/mobile-money-guidelines-2013.pdf> (Accessed on 15/07/2015)
- Yin, R.K., 2011. Qualitative research from start to finish. The GUILDFORD PRESS, London.
- Zahoor, A.S., Mahmood, H.S. & Javed, A. 2016. Information security management needs more holistic approach: A literature review. International Journal of Information Management, 36 (2016) pp 215-225

# Optimal Sensor Placement in Network Topology From the Defence Point of View

Vesa Kuikka and Juha-Pekka Nikkarila

Finnish Defence Research Agency, Riihimäki, Finland

[vesa.kuikka@mil.fi](mailto:vesa.kuikka@mil.fi)

[juha-pekknikkarila@mil.fi](mailto:juha-pekknikkarila@mil.fi)

**Abstract:** We present a method for resolving optimal sensor placement in a general network topology. The network model serves in planning cyber defence and defending critical information infrastructure. The model includes node and link weights that can be applied to describe information quantity, data importance to a defender, or criticality to the network operator. In this context we define sensor as a device or logic to monitor and control internet traffic. Network analysis with an extra penalty term for describing distributed defence aspects are used in the modelling. The applicability of the method is demonstrated with three real-world networks and it is shown that results are both intuitive and practical. The networks have been used in our previous studies in analysing the key features in cyber capabilities of closed and open national networks. The Sprint operator network in the USA is used for comparing our results with the outcomes from other research in the literature. By adjusting both the link weight and the penalty parameter we illustrate that the distribution of sensors can be either at the crucial crossroads in the network or at crucial crossroads that are also far away from each other in the network structure. The parameter adjustment determines the level of protection. Resilient arrangement is illustrated with a high parameter value, when sensors are placed on more distributed locations. A well-adjusted arrangement of sensor locations is calculated by an optimized closeness centrality to monitored nodes and sensor protection capability.

**Keywords:** network resilience, network analysis, cyber defence, optimal sensor placements, closed national networks

---

## 1. Introduction

It is commonly understood that cyber domain is a dynamic environment where situations frequently change in various ways. For example, technical and procedural developments lead to new varieties of services. Consequently, systems are replaced and reformed recurrently and numerous software vulnerabilities are established and published constantly. Defensive, offensive and cyber intelligence campaign's operations are developed promptly. In dynamic environments like this, appropriate decisions require good situational awareness; see for example (Endsley 1995a, 1995b). As the cyber domain is dynamic and in where circumstances of significance to cyber security recurrently changes, it is essential to develop fundamental defensive methodologies. It has been shown that these varieties of fundamental defensive methodologies may have a role in the military context as well. One example of a military application is the closed national segment of internet established by Russia. (Kukkola et al 2017b)

Various types of sensors and their application areas exist in modern communication network infrastructure and architecture. Sensors are utilized for both monitoring and controlling purposes. Different technical solutions have been designed for cyber defence and traditional management functions. In cyber defence, sensor systems can be utilized to mitigate malicious actions and software or even hostile actors (i.e. adversaries). So-called "fake news" represents a new tool that can be used for influencing public opinion or for engaging in hybrid warfare. Structural development of cyberspace and closed national segments of the Internet change a nation's capabilities of situational awareness.

In a concurrent study (Nikkarila and Kuikka 2019) the authors have demonstrated how complex network analysis could be utilized for evaluating effects of asymmetry established by closed national networks. The authors understand that there are several possibilities to improve cyber defence of open national networks. For example, it is possible to utilize probabilistic models of war, proposed in the literature (Cioffi-Revilla 1989; Cioffi-Revilla and Dacey 1988; Nikkarila et al 2018). The current research presents one possibility to enhance the defensive capability of an open national network.

We present a method for resolving optimal sensor placements in a general network topology that has both physical and logical levels. Physical topology is a map of network structure and logical topology illustrates how services are used within the network. The network model helps in planning cyber defence and defending critical information infrastructure. The model includes node and link weights that can be applied to describe information quantity, and data importance to a defender, or criticality to the network operator. In this context

we define ‘sensor’ as a device or logic to monitor and control internet traffic. Network analysis with an extra penalty term for describing distributed defence aspects are used in the modelling. The penalty parameter is a phenomenological and a free parameter whose value depends on how much protection is needed. The value is also dependent on link and node weights although the dependence is not strong.

We demonstrate the usability of the method via illustrative real-world networks. The first network has been applied in our previous article in analysing the differences in operational cyber capabilities between closed and open national networks and we further develop the understanding of these networks. The Sprint operator network in the USA represents a topology which is used in other studies as an example network. When compared with these studies, our results are closest to the k-median results (Alenazi 2018).

The placements of five sensors in the Sprint operator network are presented with low and high values of the penalty parameter. With a low parameter value, sensors are placed in central nodes and possibly close to each other. In addition, a resilient arrangement are illustrated with high parameter values, where sensors are placed on more distributed locations. Both aspects of closeness centrality and network resilience are taken into account in a balanced way

## **2. Background: Changes in cyberspace and the closing of national infrastructure networks**

Russia has declared its aim to achieve the capability for disconnecting Russian segment of the Internet by 2020 (Ristolainen 2017) and to establish digital sovereignty by 2024, which would cause substantial structural changes of cyberspace and establish an asymmetric advantage (Kukkola et al 2017b, Kukkola et al 2017b). Allegedly, the situation awareness in a closed national segment of the Internet (Kukkola 2018b) could be superior when compared to a national network relying open networks. Consequently, by the declaration Russia has initiated a network closing process and possibly also China develops its own version of closed national segment of the Internet, and other nations may follow their example. In earlier studies (Nikkarila and Ristolainen 2017) Russia’s potential objectives and effects of the closing process at the tactical level have been studied. The closing process is studied in more detail and it has been shown that the scale of the closing process is important and it may have substantial effects at the strategic level as well (Kukkola et al 2017a, Kukkola et al 2018b). Also technical and procedural foundation of the closing process is researched (Kukkola 2018b, Kukkola 2018c). The role of the national segment of the Internet as a part of system-of-systems of cyber defence has been examined (Kukkola 2018a). In earlier studies the authors have used mathematical means to show that the closing process has an effect on cyber capabilities of different nations (Nikkarila et al 2018). Also wargaming has been proposed in resolving how the closing process affects the nations themselves (Lantto et al 2018a), or how resilience and confrontation could be altered if some nations have essentially established closed national segment of the Internet whereas others have not (Lantto et al 2018b).

## **3. Methods**

Methods of complex networks analysis (Cherifi et al 2017) and network propagation modelling (Kuikka 2018) are used to study communication networks. The first example network is a simple version of an existing real-world infrastructure network. Because one goal of this study is to introduce modelling methods, a small network is appropriate for illustrating the methodology.

To model the process of accessing networked services, we use the information spreading model proposed in (Kuikka 2018). This model is suitable for describing communication networks because connecting network services is typically a step by step process in a network structure from a source node to a target node. The process is a non-conserving spread of information. The model has no explicit relaying functionality. However, network traffic produced by sensors’ activities can be modelled with bi-directional node and link weights. This bears close resemblance with the classical connectivity theory dealing with probabilities of connection between the nodes in a network (Colbourn 1987).

In the methodology of complex networks and social network analysis, the method for optimal sensor placement can be based on two alternative measures of closeness centrality or betweenness centrality. The two measures for optimal sensor placements are the corresponding sums of the centrality measures of nodes with an extra term where the influence between the sensors is eliminated. The extra term is multiplied by a tuneable parameter that acts as a penalty term in the model. The penalty term is included in the model to increase the capability to defend against targeted attacks. The network operator can strengthen distributed defence by

increasing the parameter value. It is crucial that the model is “global” taking into account all the paths in the network to make it possible to calculate the penalty term.

The proposed network analysis method considers all the possible self-avoiding paths in a network. The inclusion of only self-avoiding paths describes information transmission between a source and a target node. A *self-avoiding path* is a sequence of steps on a path that do not visit the same node more than once. The model considers also information flow directions and both, link and node weights. The weights can be applied to model information quantity and data importance to a defender, or criticality to the network operator.

Weights describe the resilience of nodes and links in the network structure from the view point of the process or phenomenon under interest. Two different aspects are considered - the monitored or controlled process and the resilience of the network. Moreover, there are numerous possible metrics for measuring resilience of networks (Lü et al 2016). Because of the mentioned reasons, it is a user’s decision and design question to plan and select weights. We provide several examples that help in selecting weight values based on these principles. Luckily in many cases, a wide range of weights give comparable sensor placements in the network topology. In other words, the placement of sensors is not very sensitive to link or node weights.

Here, the methodology is presented briefly to highlight the main features of the model. More detailed calculations and a pseudo algorithm have been presented in (Kuikka 2018). The mathematical model considers all paths between nodes in the network. Especially, the model takes into account the common links before branching of the paths. Paths from a source node to a target node are combined iteratively in the descending order of common path lengths (number of links) at the beginning of their paths. Practical examples have been provided in our earlier research (Kuikka 2018). Node and link weighting factors are included in the model by including all the corresponding weighting values along a path. We denote these by  $W_L$ . The next equation shows how two paths with path lengths  $L_1$  and  $L_2$  and a common path length of  $L_3$  are combined together:

$$P_{i,\min(L_1,L_2)}(T) = W_{L_1}D_{L_1}(T) + P_{i-1,L_2}(T) - \frac{W_{L_1}D_{L_1}(T)P_{i-1,L_2}(T)}{W_{L_3}D_{L_3}}, i = 1, \dots, N_L - 1 (L_1, L_2 \leq L_{max}),$$

$$P_{0,L_2} = W_{L_2}D_{L_2}(T).$$

where  $P_{i-1,L_2}(T)$  is the intermediate result at step  $i$  during the iterations, and  $N_L$  is the number of different paths between the source node  $s$  and the target node  $t$ . Temporal distribution of the spreading process (Kuikka 2018) is denoted by  $D_L(T)$ . For larger networks an upper limiting value for  $L_{max}$  is necessary to limit computing time. In a later step,  $P_{i-1,L_2}(T)$  is used as an input on the right side of the equation when shorter common path lengths, in turn, are combined during the computation. Finally, all  $N_L$  paths go into the result  $P_{N_L-1}(T)$  for the probability of propagation from the source node to the target node via all possible paths shorter or equal to  $L_{max}$ . We are interested in the limiting value  $\lim_{T \rightarrow \infty} P_{N_L-1}(T)$  as the time scales of information transmission are short when compared with the time horizon of the analysis. The limiting value of the temporal distribution is  $\lim_{T \rightarrow \infty} D_L(T) = 1$  in the above equations. The procedure provides the probability  $C_{s,t}$  for spreading from node  $s$  to node  $t$ .

We use for the measure of sensor placement the following expression:

$$M(\pi) = \sum_{\substack{s \in Sensors \\ \text{or} \\ t \notin Sensors}}^N C_{s,t} - \pi \sum_{\substack{s \in Sensors \\ \text{and} \\ t \in Sensors}}^N C_{s,t}$$

where the first summation is taken over sensor - non-sensor indexes and the second summation over sensor indexes. In the first summation, indexes between non-sensor nodes are not included. The second term is the penalty term including the penalty parameter  $\pi$ . The measure  $M(\pi)$  is maximized in the procedure for determining the optimal locations of the desired number of sensors.

In the literature, most of the network models are “local” and for that reason they are incapable of describing long distance influence between nodes in a network. It is crucial that our model is “global” in that it accounts for all of the paths in the network, thereby making the calculation of penalty terms possible. Penalty terms are

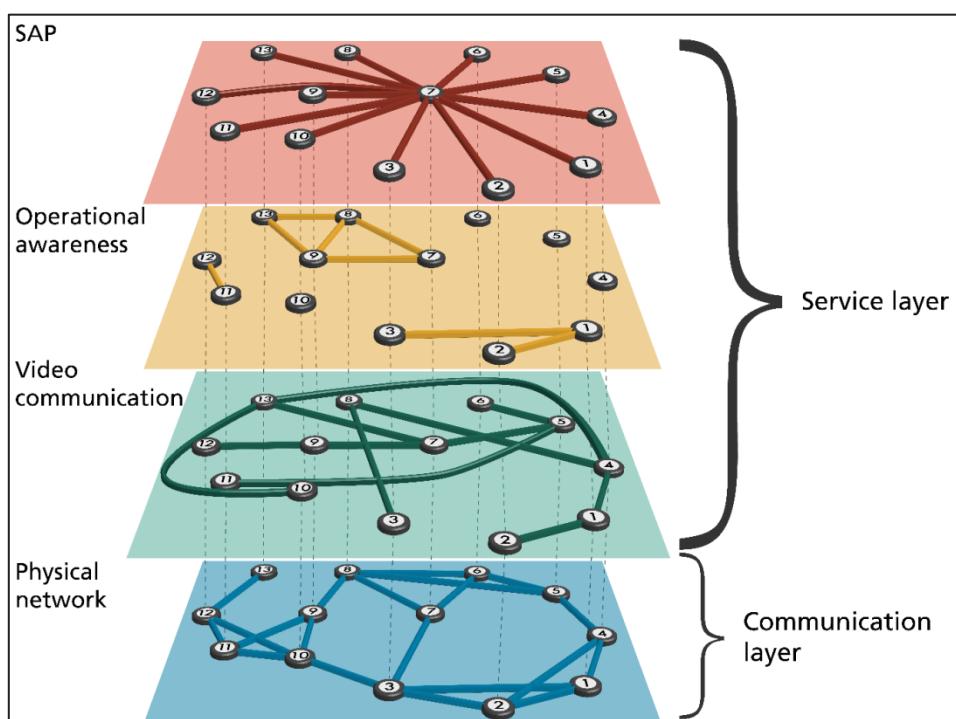
introduced in order to increase the defensive capability against targeted attacks. These terms are weighted with a tuneable parameter.

#### 4. Example networks

Three different real-world network topologies are used for demonstrating the model. As an introduction, we present the model with a simple physical network structure. The physical network layer is only one part of analysing the big picture. In addition, we illustrate the use of service layers on top of the physical network with the simple network. In order to prove the usability of the methodology in a more realistic situation, where the number of sensors is higher, we perform calculations with a well-known network. Therefore as the third example, we analyse the Sprint operator network in the USA.

The same examples are applied in our earlier studies or in the literature. The first network was applied in (Nikkarila and Kuikka 2019) where we analysed the differences in operational cyber capabilities between closed and open national networks. The second network describes the same network as the first network but the complex structure is shown in more detail. The Sprint operator network has been studied in (Alenazi 2018) where the sensor placements have been determined with four different models. The results from those models can be compared with the model in this paper.

Services are essential from the users' point of view. Consequently, the sensor placement should take into account not just the physical network structure but also the networked services. Services can be categorized and rated based on the value of cyber and operational functionalities. In the military context, capability is based on accessibility and usability of these functionalities. In real-world applications, the value of services depend on the task and mission of the scenario. Therefore the values are also time and user dependent. (Kuikka and Syrjänen 2019)



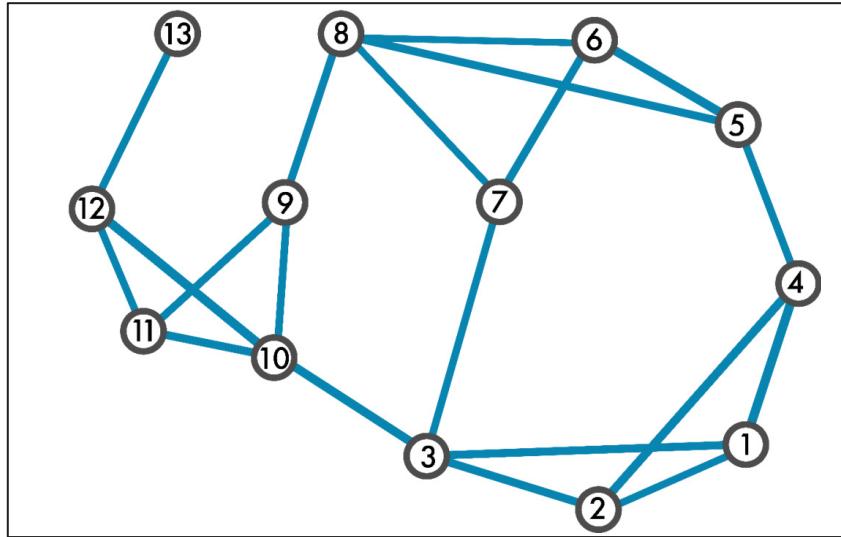
**Figure 1:** Networked service layers on the physical network layer. The physical topology is a map of network structure and the logical topology illustrates how services are used within the network

#### 5. Results

The first example network is a simplified version of a real-world infrastructure network. This network is selected for illustrating the method of selecting optimal sensor placements. In the modelling, three parameter values are to be specified: penalty parameter, link weights, and the number of sensors to be used. In the spreading model, a limiting value of time approaching infinity is describing the equilibrium state. Because self-avoiding paths are

assumed to describe information transmission in communication networks, the number of visits per node in a path is limited to one.

Table 1 shows an analysis of the network topology of Fig. 2. Negative values of placing sensors are indicated in the table. We regard these cases as outliers. It is possible that the previous configurations 1, 6, 11, 13 and 8, 12, 13 in Table 1 are also outliers. Higher link weights with a high number of sensors are extreme cases. The essential question is, which model parameter values should be used in the calculations. The number of sensors is determined by the managers or operators of the network. The decision depends on the cost of the technology and the amount of required resources in maintaining the sensor system.



**Figure 2:** Example network of 13 nodes and 20 bidirectional links (40 links). Communicating devices are modelled as nodes and connections between the devices are modelled as links between the nodes

The optimal placement of sensors have a clear pattern with a wide range of model parameters. This is a favourable feature of the model because results are not very sensitive to the model parameters. For example, it may be difficult to estimate the best value for the link weights. Very coarse estimations and experimenting with a few link weight values is sufficient. The results in Table 1 show that, when only one sensor is placed in the network, node 3 is optimal. Further, nodes 3 and 8 are optimal for two sensors. Different configurations for three sensors are suggested by the analysis: nodes 1, 8, and 10 are optimal for low activity events with low requirements for protection and 1, 8, and 12 for high protection. The corresponding configurations are 3, 8, and 10 (1, 6, and 13) for moderate activity events with low (high) protection. The last columns show the results for a very wide range of managed and monitored events in the network traffic. The final selection can be made among the suggested nodes, calculated with different model parameters, and judging with possible additional budgetary and system management requirements.

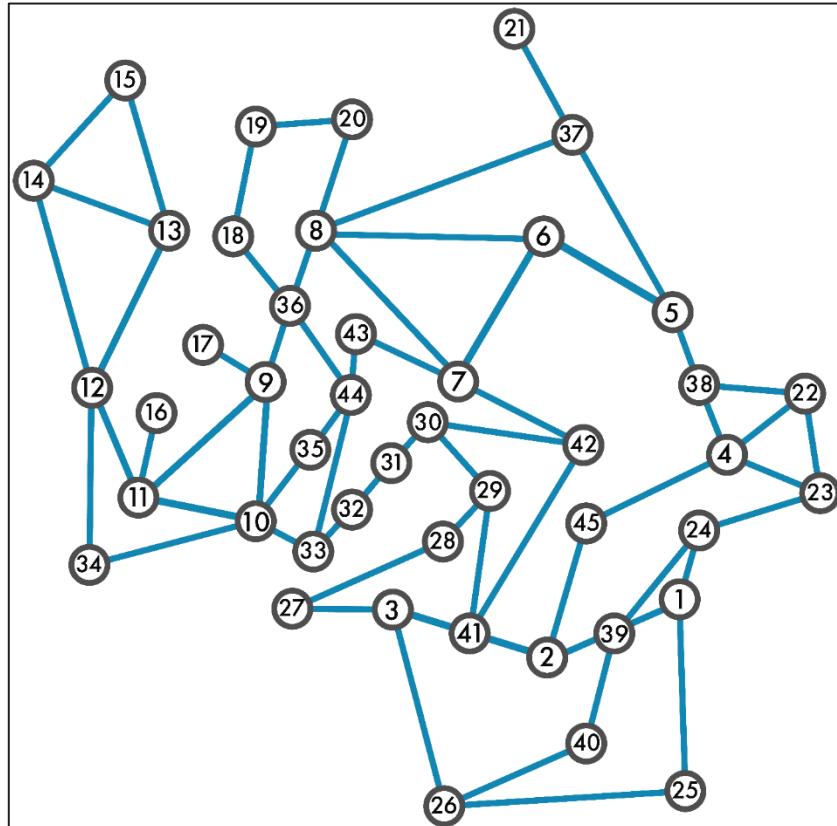
The second example is a real-world communication network of Fig. 3. The network has some characteristics structures of larger communication network topologies. Typically, networks have few nodes with many links and many nodes with few links. This has implications in optimal sensor placement because the degree of a node affects sensors' possibilities to detect network traffic and events in data transmission.

The larger network topology of Fig. 3 requires more sensors than the network in Fig. 2. In addition, the penalty parameter describing distributed protection of sensors should have a higher value. For the analysis, it is useful to calculate sensor placements with different model parameter values.

**Table 1:** Optimal sensor placements computed for the network topology of Fig. 2. Links weights, penalty parameter values, number of sensors, and optimal nodes are denoted by  $W_l$ ,  $\pi$ , #, and Sensors correspondingly

$W_l$	$\pi$	#	Sensors	$W_l$	$\pi$	#	Sensors	$W_l$	$\pi$	#	Sensors
0.1	1	1	3	0.5	1	1	3	0.9	1	1	3
		2	3,8			2	3,8			2	3,8
		3	1,8,10			3	3,8,10			3	3,7,8

$W_l$	$\pi$	#	Sensors	$W_l$	$\pi$	#	Sensors	$W_l$	$\pi$	#	Sensors
	4		3,4,8,11		4		1,6,8,10		4		3,6,7,8
	5		3,4,6,9,12		5		1,3,6,8,11		5		3,6,7,8,10
5	1	3		5	1	3		5	1	3	
	2		3,8		2		4,11		2		3,8
	3		1,8,12		3		1,6,13		3		8,12,13
	4		3,4,8,12		4		1,6,11,13		4		negative



**Figure 3:** Example network of 45 nodes and 64 bidirectional links (128 links)

Table 2 shows the analysis with different model parameter values. The link weight values  $W_l = 0.1$  describe monitoring low activity events and  $W_l = 0.5$  describe more probable activities of the network. The penalty parameter values  $\pi=5$  describe moderate distributed protection requirements and  $\pi=10$  high protection needs of sensors. The choice can be made according to these planning guidelines. In practice, other factors are taken into account which makes the decision easier. Financial resources, physical environments, and other resources should be optimized. Consequently, the analysis can be used to discover a possible set of sensor placements, and the final design of the configuration is conducted by considering other aspects of planning and real-world constraints.

**Table 2:** Placement of sensors in the network topology of Fig. 3 with different model parameters.

$W_l$	$\pi$	#	Sensors	$W_l$	$\pi$	#	Sensors
0.1	5	1	8	0.5	5	1	8
		2	8,10			2	8,10
		3	8,10,41			3	8,10,39
		4	4,8,10,41			4	8,10,22,41
		5	4,8,10,39,41			5	3,4,8,10,13
		6	4,8,10,14,39,41			6	1,8,10,13,22,29
		7	3,4,8,10,14,30,39			7	1,8,10,15,22,27,31
		8	3,4,8,10,14,30,39,44			8	1,10,15,19,21,22,27,31
	10	1	8		10	1	8
		2	8,10			2	10,39
		3	8,10,39			3	8,14,39

$W_l$	$\pi$	#	Sensors	$W_l$	$\pi$	#	Sensors
		4	4,8,10,41			4	8,13,24,28
		5	4,8,10,29,39			5	1,15,21,28,44
		6	4,8,14,17,29,39			6	1,14,17,19,21,28
		7	3,4,8,10,14,30,39			7	14,16,19,21,22,25,28
		8	3,4,8,10,14,18,30,39			8	negative

Table 3 shows an example of sensor placement in the network topology of Fig. 1 where service layers are also taken into account. The results in Table 1 (four columns) are different because services have a high weighting in the calculations in Table 3. If the physical network layer, or the availability of services, were more important, we would have similar results in Table 3. Fig. 1 shows that SAP -services from node 7 to all users in the network are important. (Kuikka and Syrjänen 2019)

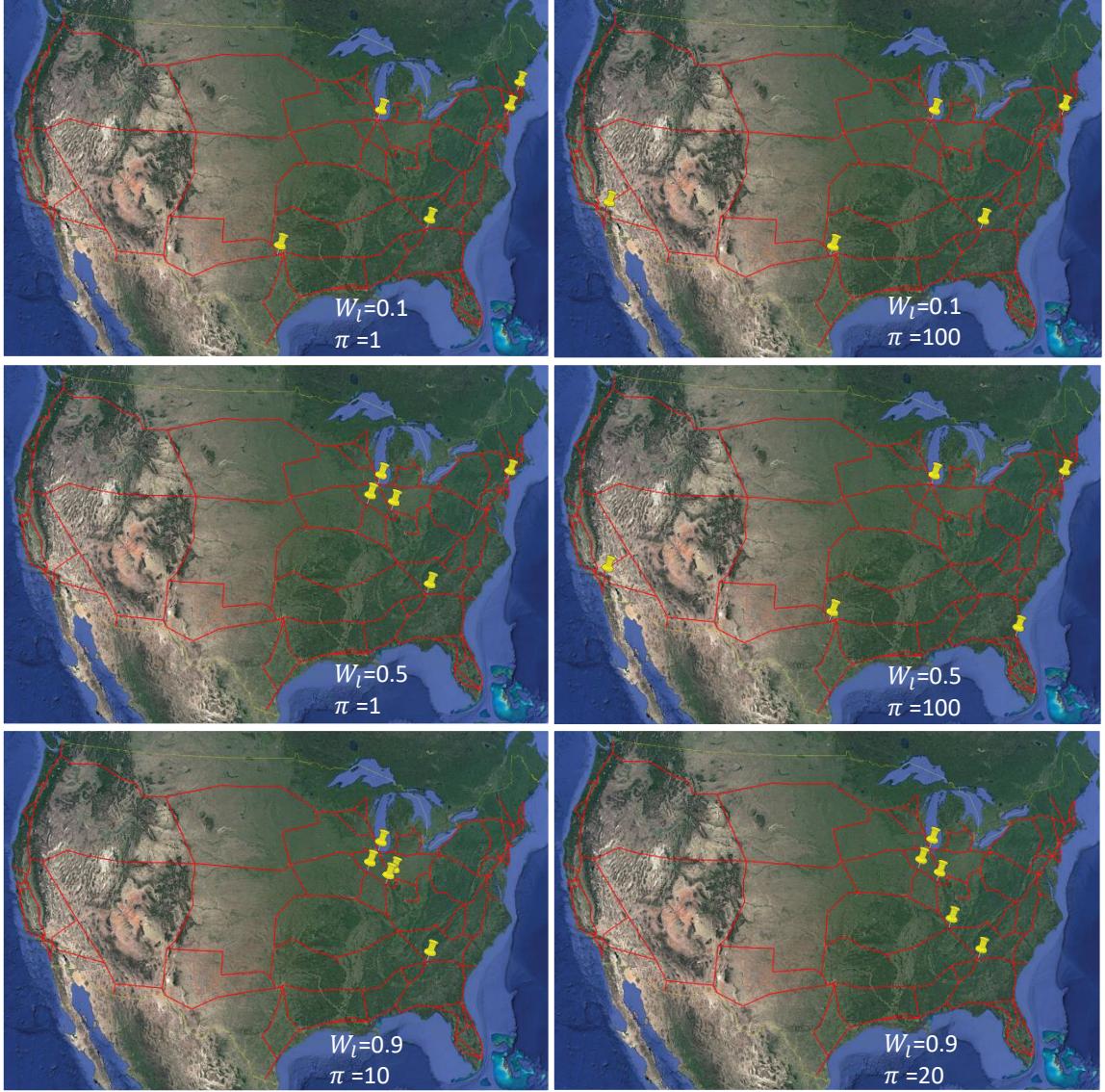
**Table 3:** Placement of sensors in the network topology of Fig. 1 and service layers of Fig. 1 (Kuikka and Syrjänen 2019)

$W_l$	$\pi$	#	Sensors
0.1	1	1	7
		2	7,11
		3	1,9,13
		4	1,5,9,13
		5	1,5,8,9,13
	5	1	7
		2	1,9
		3	1,9,13
		4	1,5,9,13

Fig. 4 shows the placements of five sensors in the Sprint operator network with different values of the link weight and penalty parameter values (in the figure: 0.1 uppermost row, 0.5 middle row and 0.9 lowest row) and penalty parameter 1 (left column) and 100 right column). Optimal sensor placements are calculated based on closeness centrality, self-avoiding paths and maximum path lengths of 75. We observe that increasing the link weight centralizes the optimal sensor locations. On the other hand, increasing the penalty parameter yields more distributed sensor locations. Therefore, a higher penalty parameter leads to a more distributed sensor placement. Interestingly, with the highest weight value (lowest row) the results are reasonable only at a narrow region of penalty parameters, and the latter phenomenon is not clearly seen with the highest link weight. Consequently, in those cases the sensors are placed at central locations and may then be close with each other. In some environments, a centralized sensor placement can be considered, yet it may increase vulnerability to a cyber attack.

We observe that our results with parameter values ( $W_l=0.5$ ,  $\pi=100$ ) correspond very accurately with  $k$ -median results in (Alenazi 2018). According to the study, sensor placement with  $k$ -median principle minimizes the overall latency in the network. Our model does not take into account the actual propagation time of the internet traffic, but instead considers the topology of the network. Even so, our model may lead to comparable sensor placement. To conclude, we have proposed a model that can be used for distributed sensor placement in an optimized manner. Both aspects, closeness centrality and network resilience, are taken into account in a balanced way. It could be worthwhile to study how resilient a network with a sensor placement calculated by our model is, when compared for example to (AC= Algebraic Connectivity, NC =Network Criticality or  $k$ -median)-optimised networks (Alenazi 2018).

By selecting the number of sensors as in (Alenazi 2018) we can directly compare the results. Algebraic connectivity and network criticality have been used as robustness functions in (Alenazi 2018). The sensor placements of this paper are different from these two methods as can be seen in (Alenazi 2018). However the sensor placements for  $k$ -median are almost similar. To compare different methods, one would need to select which processes are monitored and/or controlled (both can be done at the same time) and select criteria for evaluating the resilience of networks. (Lü et al 2016)



**Figure 4:** Optimal locations of five sensors in the Sprint network with the link weight values

## 6. Conclusions

Optimal placement of sensors is important when a limited number of sensors are placed in the communication network infrastructure. In this paper, we propose a method for optimizing sensor placement in a complex network topology. The proposed model is global and accounts for all paths in the network. The novelty of the methodology is that network structures, services, and protection by distributed sensor placements are taken into account consistently in the same model. We demonstrate the model with real-world communication networks and show that results are intuitive and practical. The results are compared with earlier studies conducted by the Sprint network in USA. Requirements for protection are accomplished by implementing a penalty parameter to distribute sensors further away from each other in the network structure. Adjusting the parameter determines the level of protection. A balanced configuration of sensor placements is computed by optimizing closeness centrality to monitored nodes and optimizing protection capability of the sensors. Additional value of the proposed model is that it can be utilized when improving defensive cyber capabilities in a nation-wide context. For example, the proposed model could assist when constructing a large-scale and real-time cyber situation awareness system.

We have introduced a sensor placement method that considers monitor and control processes, resiliency of the network, networked services and topology of the entire network.

## Acknowledgements

The authors appreciate the valuable work Mr. Matti Syrjänen and Mr. Joona Repo have conducted for accomplishing the current research.

## References

- Alenazi, M.J.F. (2018) "On SDN Controller Placement to Achieve Robustness Against Targeted Attacks", In: Cherifi, C., Cherifi, H., Karsai, M. and Musolesi, M. (eds) Complex Networks & Their Applications VI. COMPLEX NETWORKS 2017. *Studies in Computational Intelligence, vol 689. Springer, Cham.*
- Cioffi-Revilla, C. (1989) "Mathematical contributions to the scientific understanding of war", *Mathematical and Computer Modelling, Vol. 12, Issues 4–5, Pages 561-575, 1989.*
- Cioffi-Revilla, C. and Dacey, R. (1989) "The probability of war in then-crises problem: Modeling new alternatives to Wright's solution", *Synthese, Volume 76, Issue 2, pp 285–305, August 1988.*
- Colbourn, C.J.(1987) "The Combinatorics of Network Reliability", Oxford University Press.
- Endsley, M.R. (1995a) "Toward a Theory of Situation Awareness in Dynamic Systems", *Hum. Factors 37, 32–64.*
- Endsley, M.R. (1995b) "Measurement of Situation Awareness in Dynamic Systems", *Hum. Factors J. Hum. Factors Ergon. Soc. 37, 65–84.*
- Kuikka, V. (2018) "Influence Spreading Model Used to Analyse Social Networks and Detect Sub-Communities", *Computational Social Networks 5:12, <https://doi.org/10.1186/s40649-018-0060-z>.*
- Kuikka, V. and Syrjänen, M. (2019) "Modelling Utility of Services in Military Networked Environments", manuscript under review.
- Kukkola, J. (2018a) "The Russian Segment of Internet as a Resilient Battlefield", *ISMS Annual Conference 2018, Military Sciences and Future Security Challenges, Warsaw Poland.*
- Kukkola, J. (2018b) "Civilian and military information infrastructure and the control of the Russian segment of Internet", *Paper presented to ICCWS 2018, Washington, 8.-9. March 2018.*
- Kukkola, J. (2018c) "Russian Cyber Power and Structural Asymmetry". In Chen, Jim Q. & Hurley, John S. *Proceedings of the 13th International Conference on Cyber Warfare and Security. National Defence University Washington DC, 8.-9. March, 362-368.*
- Kukkola, J., Nikkarila, J-P and Ristolainen, M. (2017a) "Shaping Cyberspace –A predictive analysis of adversarial cyber capabilities", *IST-145 specialists' Meeting Predictive Analytics and Analysis in the Cyber Domain, Sibiu, Romania.*
- Kukkola, J., Ristolainen, M. and Nikkarila, J-P (2017b) "Game Changer: Structural Transformation of Cyberspace", Riihimäki: Finnish Defence Research Agency. [Online]. Available: <http://puolustusvoimat.fi/web/tutkimus/tutkimuslaitoksen-julkaisut> [Accessed 14 May 2018].
- Lantto, H., Huopio, S., Åkesson, B., Nikkarila, J-P, Suojanen, M., Ristolainen, M. and Tuukkanen, T. (2018b) "Wargaming the Cyber Resilience of Structurally and Technologically Different Networks", *ISMS Annual Conference 2018, "Military Sciences and Future Security Challenges" (ISMS), Warsaw Poland.*
- Lantto, H., Åkesson, B., Kukkola, J., Nikkarila, J-P and Ristolainen, M. (2018a) "Wargaming a closed national network: What are you willing to sacrifice?", in *MILCOM 2018: Los Angeles, CA, USA, 29-31 October 2018.*
- Lü, L., Chen, D., Ren, X-L., Zhang, Q-M., Zhang, Y-C., Zhou, T., "Vital nodes identification in complex networks", *Physics Reports 650 (2016) 1–63.*
- Nikkarila, J-P and Kuikka, V. (2019) "Complex network analysis for evaluating effects of asymmetry established by closed national networks", sent for publication in *2019 International Conference on Military Communications and Information Systems (ICMCIS) in Budva, Montenegro, 14th-15th May 2019.*
- Nikkarila, J-P and Ristolainen, M. (2017) "'RuNet 2020' - Deploying traditional elements of combat power in cyberspace?", *Proceedings of 2017 International Conference on Military Communications and Information Systems (ICMCIS), 15-16 May 2017, 1-8.*
- Nikkarila, J-P, Åkesson, B., Kuikka, V. and Hämäläinen, J. (2018) "Modelling Closed National Networks – Effects in Cyber Operation Capabilities", *17th European Conference on Cyber Warfare and Security (ECCWS), Oslo, Norway.*
- Ristolainen, M. (2017) "Should 'RuNet 2020' be taken seriously?", *Journal of Information Warfare, vol. 16, no. 4, pp. 113-131, 2017.*

# Operator Impressions of 3D Visualizations for Cybersecurity Analysts

Kaur Kullman<sup>1</sup>, Noam Ben Asher<sup>2</sup> and Char Sample<sup>3</sup>

<sup>1</sup>TalTech University, Tallinn, Estonia

<sup>2</sup>ORAU, Oak Ridge TN, USA

<sup>3</sup>ICF Inc. Columbia, USA

[kaur@ieee.org](mailto:kaur@ieee.org)

**Abstract:** Cybersecurity analysts ingest and process significant amounts of data from diverse sources in order to acquire network situation awareness. Visualizations can enhance the efficiency of analysts' workflow by providing contextual information, various sets of cybersecurity related data, information regarding alerts, among others. However, textual displays and 2D visualizations have limited capabilities in displaying complex, dynamic and multidimensional information. There have been many attempts to visualize data in 3D, while being displayed on 2D displays, but success has been limited. We propose that customized, stereoscopically perceivable 3D visualizations aligned with analysts' internal representations of network topology, may enhance their capability to understand their networks' state in ways that 2D displays cannot afford. These 3D visualizations may also provide a path for users who are trained and comfortable with textual and 2D representations of data to assess visualization methods that may be suitably aligned to implicit knowledge of their networks. Thus, the premise of custom data-visualizations forms the foundation for this study. Herein, we report on findings from a comparative, qualitative, within-subjects usability analysis between 2D and 3D representations of the same network traffic dataset. Study participants (analysts) provided information on: 1.) ability to create an initial understanding of the network, 2.) ease of finding task-relevant information in the representation, and 3.) overall usability. Results indicated that interviewees indicated a preference for 3D visualizations over the 2D alternatives and we discuss possible explanations for this preference.

**Keywords:** visualization, cybersecurity, decision-making, data visualization, virtual reality

---

## 1. Introduction

Cyber is the newest domain of war (Lynn, 2010), endowed with unique sensory characteristics that differentiate this warfare environment from kinetic-physical warfare (Gonzalez, Ben-Asher, Oltramari, & Lebiere, 2014). Cyber security analysts who are responsible for ensuring the security of networks and other assets, utilize a wide array of computer network defense (CND) tools, such as Security Information & Event Management (SIEM) that allow data from various sources to be processed and alerted on. CND tools allow analysts to monitor, detect, investigate and report incidents that occur in the network, as well as provide an overview of the network state. To provide analysts with such capabilities, CND tools depend on the ability to query, process, summarize and display large quantities of diverse data which have fast and unexpected dynamics (Ben-Asher & Gonzalez, 2015). Shneiderman (Shneiderman, 1996) provided a taxonomy depicting 7 human-data interaction task levels: 1.) gaining *Overview* of the entire dataset, 2.) *Zoom* on an item or subsets of items, 3.) *Filter* non relevant items, 4.) get *Details-on-Demand* for an item or subset of items, 5.) *Relate* between items or subset of items, 6.) keep *History* of actions and 7.) allow *Extraction* of subsets of items and query parameters. Traditionally, cyber defenders have used command line tools and alphanumeric data displays to execute these seven tasks. With the need for faster and more accurate situational awareness of increasing data volume, many CND products have integrated graphical user interface (GUI) and 2-dimensional (2D) data visualizations to expedite human information acquisition.

Visualizing multidimensional data on 2D screens bears several limitations. First, the resulting visualization after the dimensionality reduction methods have been applied will likely differ from the mental representation that the analyst had acquired upon reviewing the same data in numerical and textual data. Contextual information will likely be removed in the reduction process that could be crucial to the understanding of the situation and to identify relevant clues (Rajivan, Konstantinidis, Ben-Asher, & Gonzalez, 2016). Also, three dimensional visualizations may address some of the inherent limitations of 2D displays by aligning with the analysts' internal representation of their datasets, if the analyst naturally thinks about data in three dimensions. Users' interactions in VR (Virtual Reality) and XR (Mixed Reality) may also be more intuitive if users are expected to interact with the visualization in ways that humans manipulate objects in the physical world. This way we could harness the dexterity of human hand movement for interactions (Gershon, Klatzky, & Lee, 2014) if the tools used are capable enough, using haptic feedback to further advance users interaction efficiency (Gershon, Klatzky, Palani, & Giudice, 2016).

However, the scale, heterogeneity, and complexity of cybersecurity datasets continue to pose challenges for visualization and interaction designers (Reda, et al., 2013) despite the constant increase in computers' abilities to process and display more data on 2D displays. This could be because visualization designers lack an understanding of cybersecurity operations and network infrastructure to create an effective visualization for cybersecurity analysts. Yet, cybersecurity analysts who are familiar with the technical aspect of network monitoring do not have expertise in data visualization and human perception. The visualization designer might need to have dual expertise.

By providing a data visualization environment where minimalistic visual, audio and haptic cues are informing the user of what is happening in that environment, we allow the user to focus on the task. Furthermore environmental cues should be perceptible and clear to avoid user confusion. Visualization should be functional and available utilities should accurately convey their functions. Controls, navigation, interface, and all other interface conventions should be consistent. Users cognitive and physical workload must be minimized. Human errors should be anticipated and prevented if possible. The environment should be flexible to allow customization for personal preferences, cultural differences, color vision deficiency etc. (Hodent, 2018). We hypothesize that the analyst may find it intuitive to use a 3D representation of cybersecurity network data that is aligned with the above guidelines as well as the analyst's internalized understanding of the data. Intuitive interfaces may enable the analysts to explore and understand their environment more efficiently.

## **2. Visualization for cyber defense**

Cybersecurity visualizations provide analysts with visual representation of alphanumeric data that would otherwise be difficult to comprehend due to its large volume. Such visualizations aim to effectively support analyst's tasks including detecting, monitoring and mitigating cyber attacks in a timely and efficient manner (Sethi & Wills, 2017). Cybersecurity specific visualizations can be broadly classified into three main categories: 1.) network analysis, 2.) malware and 3.) threat analysis, and situational awareness (Sethi & Wills, 2017). Timely and efficient execution of tasks in each of these categories may require different types of visualizations addressed by a growing number of cybersecurity specific visualization tools (Marty, 2008) as well as universal software with visualization capabilities like Tableau, MS Excel, R, Python, and D3 libraries (d3.js) among others. These tools could be used to visualize data in myriad of ways (Munzner, 2014) so that analysts could explore their datasets visually and interactively (Ward, Grinstein, & Keim, 2015). *Graphistry* is one recent example of a 2D force-directed graph visualization (Meyerovich & Tomasello, 2016) whose interface is easy to manage and visualization and is responsive to queries on massive datasets. These are crucial qualities for cybersecurity analysts, with emphasis on the importance of the low-latency between analyst's request for a change in visualization (change in filter, time window or other query parameters) and rendering of the visualized response from the system (Wu, Xu, Chang, Hellerstein, & Wu, 2018).

The usability of data visualizations for CND operations that have not been evaluated, may lead to low adoption rates by practitioners (Best, Endert, & Kidwell, 2014). The challenge in creating useful visualization for cybersecurity practitioners is in aligning data visualization experts' knowledge with cybersecurity analysts' needs and knowledge so, that the resulting visualizations would be useful for work tasking. A recent survey showed that 46% of 130 tools did not have any user-involvement in the evaluation phase (Sethi & Wills, 2017).

To achieve higher visualization adoption rates, analysts should have the ability to intuitively and iteratively adjust the visualizations to suit with their changing needs (Kirk, 2016). Datasets used in cybersecurity operations are often multi-dimensional and analysts would either have to scale down the number of dimensions viewed at one time to be able to use 2D & 3D visualizations, or combine multiple 2D visualizations displaying different dimensions of the same dataset in a single dashboard. This requires the designer to properly encode variables (dimensions) into shapes, colors, sizes among others. The viewer has to translate that shape into spatial perception and compare it to her internal understanding of the data, to decode the meaning of the visualizations; a task that may be non-trivial (Ehrenstein, Spillmann, & Sarris, 2003). There have been numerous attempts to employ 3D visualizations for cybersecurity data that are displayed on 2D computer screens with varying degrees of success. Such visualizations sometimes use monocular depth cues (Lebreton, Raake, Barkowsky, & Le Callet, 2012) and object movement to convey the 3D shape of the visualization; advantages and disadvantages of which were thoroughly discussed in our previous paper (Kullman, Cowley, & Ben-Asher, 2018). VIDS (Shearer & Edwards, 2018) provides an interactive 3D environment for visualizing network and alert (or other) data in 3D shapes, whereby users can seamlessly switch styles and layouts to dynamically shape their data and easily adjust their viewpoint (Gaw, 2014). Real-time 3D visualization engine DAEDALUS-VIZ allows

operators to grasp visually and in real-time an overview of alert circumstances, while providing highly flexible and tangible interactivity (Inoue, Suzuki, Suzuki, Eto, & Nakao, 2012). InetVis (van Riel & Irwin, 2006) allows the user to allocate source and destination IPv4 addresses to X and Z axis, while destination ports are being allocated to the Y axis on a 3D cube. To understand the shape of the cube and detect the positions on Z axis, user must manually change the viewpoint with mouse. Shoki (Berry, n.d.) allows the user to define what values are plotted on which axis, while the screen is divided to four squares, three of them showing each axis in 2D, while the fourth square displays the cube as a 3D object.

Due to the emergence of commodity VR devices, multiple data visualization tools have implemented support for VR headsets, that are capable of 6 degrees of freedom (6DOF) movement of the user's viewpoint, allowing the observer to perceive the depth of the visualization stereoscopically, avoiding the mental work needed to convert 2D images to 3D. OpenGraphiti (Reuille, Hawthorne, Hay, Matsusaki, & Ye, 2015) enables provides customizable graphs, along with querying and filtering capabilities. However, OpenGraphiti does not provide 3D VR interaction capacities. However, V-Arc (Maddix, 2015) enables the positioning of data in a predetermined layout, data selection and color-coding amongst other capabilities. Virtual Data Explorer (VDE) is a VR tool that allows users to collaborate while investigating 3D data visualizations, to find anomalies in a variety of cybersecurity-related datasets (U.S. ARL, 2018). For our research herein, we used VDE (see 3.4), because it enables the user to perceive the spatial layout of the topology based on observed network traffic, while the resulting visualization can be augmented with additional data, like TCP/UDP session counts between network nodes (Kullman, Cowley, & Ben-Asher, 2018). Due to the 6DOF of Oculus Rift VR headset (OVR) used for this study, VDE also allows us to test the usefulness of stereoscopically perceived depth-cues (contrary to monocular depth-cues on flat screens) for encoding data.

### **3. Method**

To understand whether stereoscopically-perceivable 3D data-shapes representing a complex computer network's topology is usable, we conducted semi-structured interviews with 10 subject matter experts working as cybersecurity analysts (as suggested in (Ward, Grinstein, & Keim, 2015)). Included in the usability assessment, we asked analysts about whether network behavior is understandable, helpful and useful for cybersecurity analysts' tasks.

#### **3.1 Participants**

Ten cybersecurity analysts (Mean age = 36.5yr., 20% females) were interviewed in a semi-structured format. All participants work as cyber security practitioners, having 2 months to 10 years of experience in the field (Mean = 4.5 years). Participation was voluntary and these volunteers were not compensated for their time.

#### **3.2 Materials**

The network traffic data used during the interviews was part of the NATO CCDCOE CDX Locked Shields 2018 (LS18) "Partner Run" (LS18PR) dataset. This dataset includes 23 defensive teams' (Blue Teams, BT) networks, an offensive (Red Team, RT), infrastructure support (Green Team), situational awareness (Yellow Team) and the managing team (White Team) nodes and traffic. During LS18 exercise the network included more than 4000 virtual machines, with about 2500 attacks executed by the Red Team against all Blue Teams combined. LS18PR function was to test Read Team' and infrastructure' readiness. Distinct of the main event, LS18 that ran for 2 days (2x7 hours), LS18PR ran for 7 hours, during which only few Blue Team networks were attacked and defended, while the rest of the networks were running as usual. Hence LS18PR dataset provides the ability to observe networks in their "normal" as well as "under attack" and "compromised" states during the same time.

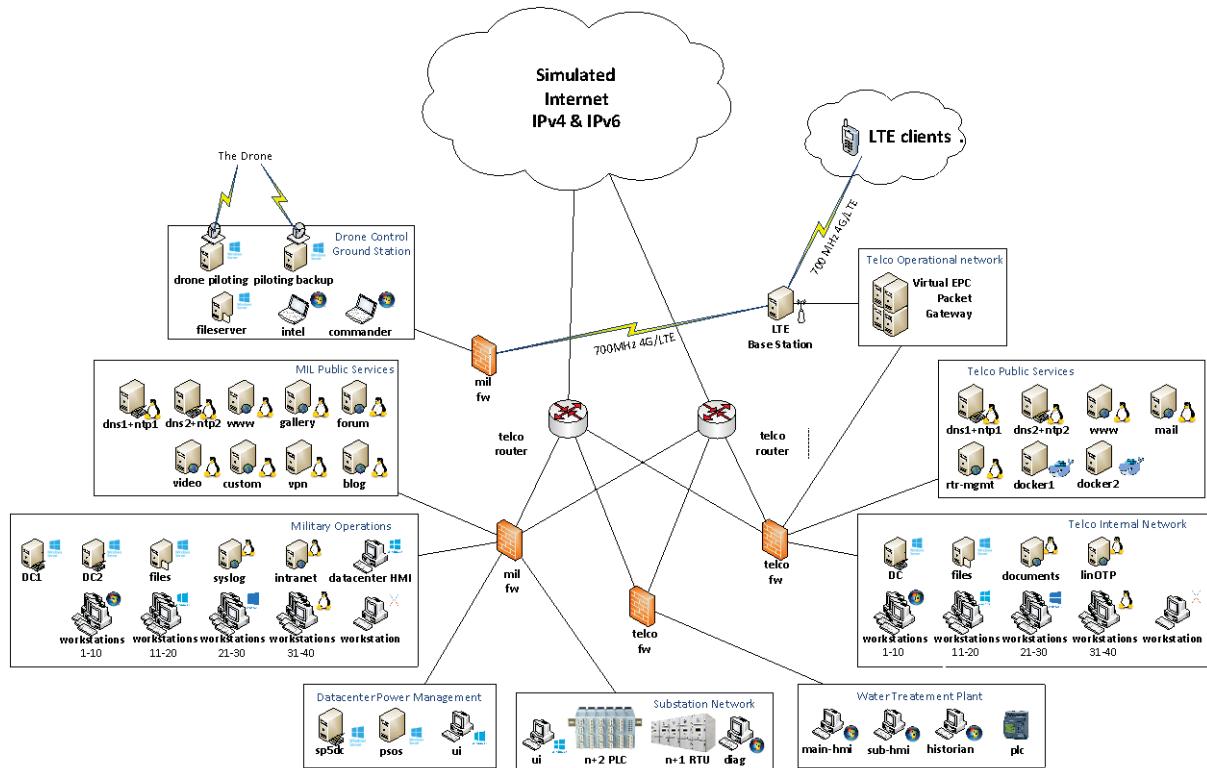
The time-window for the network sessions used in the visualizations shown to the participants was set to 40-minute periods. For data preparation, Moloch (<https://molo.ch/>) was used to process the LS18PR packet data (PCAP). The resulting data and metadata was stored in an Elasticsearch server to allow dynamic querying. Kibana (<https://www.elastic.co/products/kibana>) was used to generate dynamic and interactive 2D visualizations based on the data stored in Elasticsearch server. VDE queried the Elasticsearch server for session counts between entities and presented the information using Oculus Rift (<https://www.oculus.com/rift/>) VR headset (OVR) while an Oculus Touch controller (OTC) was used to interact with the VDE. The controller allowed users to move around the virtual space (by changing their viewpoint), select different groups of objects (e.g. connections from/to a Blue Team), grab a network node to alter its position (and better perceive the destinations of the connections that this node had) and query additional information about the node (e.g. it's IP addresses).

### 3.3 Procedure

Upon arrival to the interview, participants were asked about their cybersecurity expertise, experience with 2D and 3D visualizations, as well as gaming preferences. Gaming preferences were discussed to build the rapport and understand interviewee's level of experience in cybersecurity. Below is a list of the basic questions asked in this introduction portion of the procedure. Participants provided open ended answers that were documented by the experimenter.

- What are your favorite console / computer games?
- Have you used Moloch and / or Kibana before?
- Have you experienced Virtual, Augmented or Mixed Reality before?
- Do you have formal education on IT and / or cybersecurity?
- What area do you specialize in cybersecurity?
- How long have you been working on your current specialty; on cybersecurity?

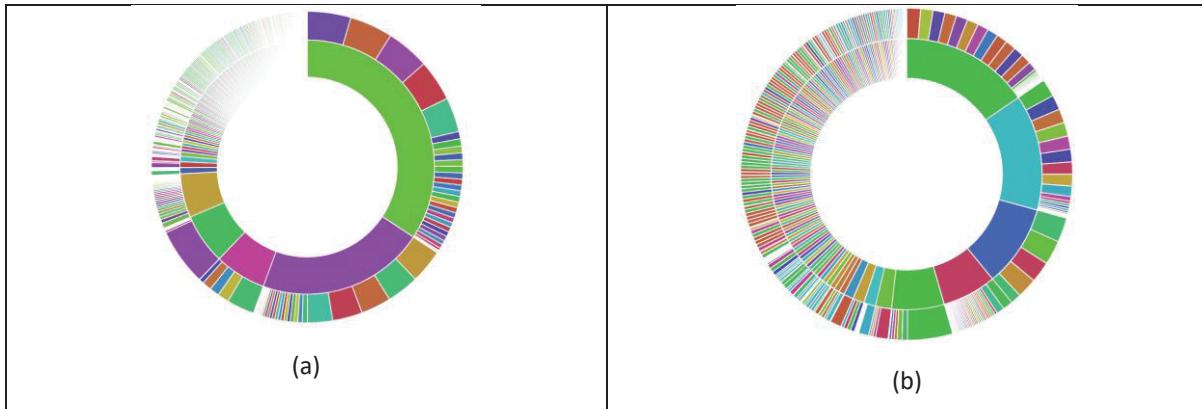
Then, participants received a short briefing about the purpose of the study and an overview of the dataset. They were shown a printed diagram (see Figure 1) illustrating the network topology of a single blue team network. Based on the diagram, participants were asked to consider, what (textual and visual) tools they would prefer to use to learn that network's topology to acquire situational awareness.



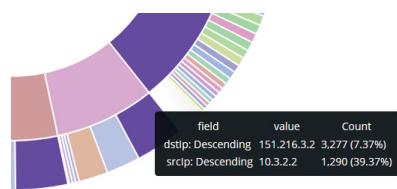
**Figure 1:** Simplified network topology diagram for traffic used in this study

Following, the experimenter presented a 2D visualization of session counts between the different entities in the Blue Team's network as radial figures (Figure 2 using Kibana). As seen in Figure 2a and Figure 2b there is a difference of an actively attacked and maintained network (on the left) compared to another one with exactly the same topology and active services, while unmaintained and not attacked network (on the right).

Size of a sector on Figure 2 represents the count of observed sessions. The radial diagram also indicated how many connections were initiated between the source and the inner ring, how many of those connections were targeted at the one represented in the outer ring, etc. The color of the session-count-block is randomly allocated to each node in that Blue Team network and does not have any relation to other networks. However, the color allocated to each node does allow the observer to find that same node in that same radial. The other sectors belonging to that same node are also highlighted by a mouse-over, which displays a popup detailing the IP addresses of source and its destination nodes and their session counts (see Figure 3).

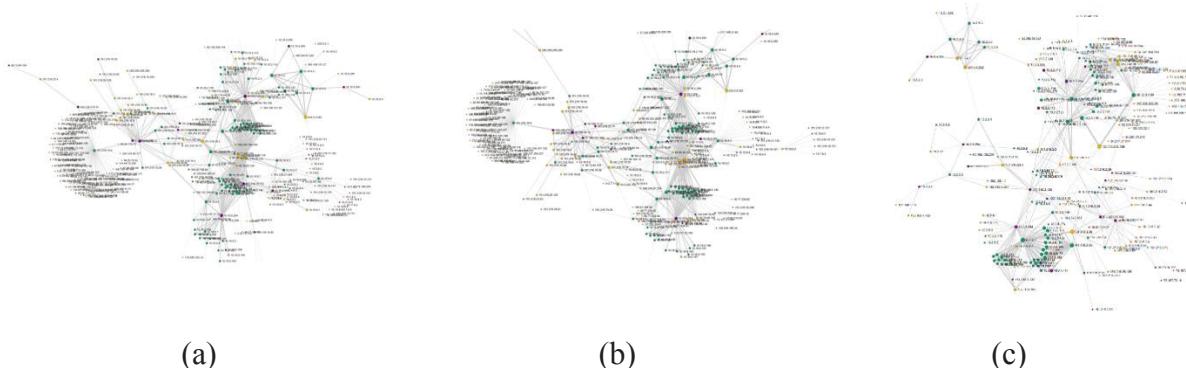


**Figure 2:** Radial 2D diagrams visualizing activeness of Blue Teams' internal nodes. Inner ring contains source addresses of the entities present in that Blue Team's network while the outer ring contains destination addresses of that same Blue Team's entities, with network connection sessions to that source address during a time-window



**Figure 3:** Mouse-over example for selected data-point displayed in a popup window, with source IP address and observed session count (srclp) and destination IP address with its respective session count (dstip)

Participants were also shown a 2D visualization of a force directed graph (using Moloch) to provide another example of graphical representation of relations between networked entities (Figure 4).



**Figure 4:** Network connections observed in the two unused BT (a) and (b) networks compared to a BT (c) networks that were actively engaged during LS18PR. Traffic observed during a set period of time is applied as pull strength of the edges between nodes, while nodes represent the hosts initiating and/or receiving connections, visualized on a 2D graph

Following the review of various 2D visualizations, participants were introduced to the VDE and the reasoning behind the spatial positioning of network elements in a 3D space. During this step, LS18PR networks were first shown as 3D shapes on a 2D display and once participants felt comfortable with their understanding of the 3D visualization as shown in VDE, they were fitted with an Oculus Rift VR Headset and Touch Controller. Participants were encouraged to explore the 3D display by changing their viewpoint while using VDE so that they could observe Blue Team networks from close distance (Figure 5) and would be able to read explanatory texts, reach out to grab the nodes, move those around, highlight nodes' features, select edges' groups and so on. After participants had familiarized themselves with the VDE environment and its 6DOF "rudder head movement" (Pruett, 2017), they were asked to evaluate if and how such network topology visualization and its augmentation with additional data (e.g. network session counts) would relate to the 2D visualizations (Figures 2, 3, 4) and the print-out topology (Figure 1) shown to them before. Once the participant gained an understanding of a blue team's network, she/he was guided to adjust the viewpoint in VR so that all the LS18PR network components

would be in the field of view. Then, the participant was asked to provide feedback and critique regarding the usability of VDE, subjective ease of stereoscopical perception of the 3D data shapes and her/his ability to acquire situation awareness with such tool.

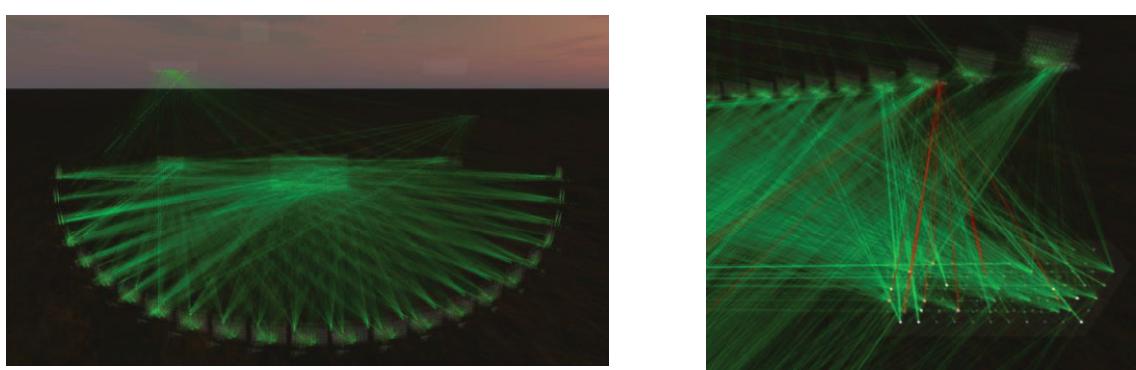
### 3.4 Virtual data explorer

Virtual Data Explorer (VDE) was developed with the Unity 3D game engine to present users with stereoscopically perceivable data visualizations in VR and XR.



**Figure 5:** VDE 3D display of network topology and traffic focusing on a single Blue Team network topology (“Zoom” on “Details-on-Demand”, per (Shneiderman, 1996)). Additional videos of VDE can be found: <https://coda.ee/vde>

The VDE uses a predefined topology description (configuration) for the visualized network. Data-shapes were spatially positioned into a meta-shape (viewed from different angles as shown in Figure 6) to allow the user to take advantage of stereoscopic viewing that VR provides. Multiple layouts were considered to minimize possible edge clutter and enable convenient distinguishability of intra- and extra-network connections and nodes' relations. These 3D shapes are easily understandable in stereoscopically-perceivable VR headsets, while often cluttered and unusable on a 2D screen or on a printed paper.



**Figure 6:** 3D display of LS18PR network topology and network traffic using VDE: (a) overall view of the meta-shape (“Overview”); (b), RT group (“Zoom”) with some added connections and selected (“Filtered”) edges colored red (“Relations”, per Shneiderman, 1996).

VDE allows the spatial topology of the network to be augmented with additional data. For this study, the visualization was enriched with network session counts so, that the most popular connection (represented as a green line (edge) between the nodes that were observed connecting) was fully visible, while the least popular connection was almost transparent. User could add additional sessions using VDE menu system in VR, in which

case the added edges were colored red until a next set of edges was added. Additional information about VDE design decisions can be found in our previous paper (Kullman, Cowley, & Ben-Asher, 2018).

#### **4. Results**

Participants were asked for impressions on the technology, ideas or suggestions for modifications, should such a tool be made available for their tasking. All participants had used 2D visualizations, 6 had used Kibana, and 4 out had experienced VR prior to this study. Seven interviewees self-rated themselves as ‘experienced’ in playing computer games. All participants used command line interfaces or GUIs for tasks like querying NetFlow data, inspect captured packets, review incidents in SIEM, explore various configuration files of the routing and perimeter devices etc. Seven used Kibana and / or Excel for simple 2D visualization. All participants agreed that although printed or electronically provided 2D diagrams of the network topology would be helpful, topologies provided are usually outdated. Overall, there emerged a common process that analysts would have used to build their understanding of a network’s topology. First, analysts would build an understanding of what is “normal” in that specific environment and then find the behaviors that would need further investigation. In order to execute that process, the analysts begins by building filters on available data to exclude the findings that are deemed “normal” and continue fine-tuning that filter, while the analyst learns that network’s behavior and builds her/his internalized understanding of that datasets expected demeanor.

To the questions “How would you build the understanding of a network? How would you map the network topology?” participants responded with: “Textual logs from SIEM [or similar]”; “Textual tools”; “flow in Kibana, xflow, tcpdump”. “Whatever tools that are available. Textual mostly, if some visualization is available, then that too.” To the question “How would you establish expectations regarding normal behaviors of the different entities in the network?” participants responded with: “I just read the logs [...]. Look in packets.”; “Check Ingress/egress points, what services are allowed through firewall, what VLANs&VPNs are in use.”; “filter logs, what type of systems are in these networks from SIEM. Use different filters and queries in Kibana.”; “Work through documentation, dive in, see configuration.”; “flow data, who’s talking to each other, as opposed to heading outside; I’ve used to bar graphs [...] what IP’s are being hit, what hosts are endangered a lot in traffic. Usually I would likely see just textual lists of IP addresses, ports etc.”; “Look at netflows, do a graph chart. Use excel, build bar graphs [and so on]”; “Use a sensor that sees all the traffic in this network, build a BPF to exclude things (ICMP, TCP, UDP) [from tcpdump] to see what’s left, what falls out, what ports and protocols, [...], to find what stands out, what weird ports are in use.”

Once in the VDE VR environment, users had some trouble getting comfortable with the “rudder head movement” to roam around in the VR environment. Few participants found the rudder head movement intuitive to use. Learning to use this tool usually took between one to two minutes. Most problematic was the vertical movement. To execute vertical movement, users had to look straight up or straight down and move backwards or forwards to change their vertical viewpoint. However, one participant commented, “[moving around in this environment] is very natural to me”. Only two participants reported feeling dizzy or having any simulator sickness/motion sickness symptoms during or after the session. The precision of depth perception or the predicting the physical distance of each virtual object from the self was highly variable. Most of the participants managed to grab and hold visualized entities easily, while others struggled. Further tuning of haptic and visual feedback is needed to improve interaction with 3D data representations.

Once the participants were comfortable with adjusting their position and viewpoint in the VR environment, all were able to observe the data-shapes and understand its relation to the network’s textual and 2D visualizations they had seen moments before. Most participants stated that once the logical structure of the shapes positioned in VR had been understood, the topology of these networks became much clearer. Participants also understood the underling relations between 3D visualization and the textual / 2D representation they had seen previously:

*[P1]: “Does this [3D visualization] map back to the topology you saw on the paper before? I think so, I mean I'd have to figure out where things are, but... [yes]”;*

*[P6]: “It is not how I thought about, but... [...] I think it would definitely be helpful, but it would take some re-learning, [e.g.] how to think visually. When you learn networking, then you’re kind of trying to build this in your brain already. But then getting used to seeing this and not having to build your own picture. Like, training brain to think visually not only textually.”;*

[P4]: "So this is what you showed [me] before on the network diagram [on paper]? [...] Yes, I like this. This is different than looking graphs on your computer screen. [...] I take it you'd prefer this to the regular [tools]? Yes, definitely.";

Participants could "see" where "things" are in the network, helping them spatially perceive and understand the structure and topology of this computer network and networked entities' (nodes) positions and logical grouping inside that network ("Overview", as per (Shneiderman, 1996) taxonomy):

[P1]: "This agrees with me particular. I like visual. I always kind of visualize things, in a way like this. And this particularly agrees with me. That's very easy for me to understand.";

[P6]: "If you now look at this network diagram [printed on paper], does this looks familiar? It is familiar, but very flat. Would you prefer 3d? Well, of course. This 2D looks like... why are we still using this.. it seems so.. like.. limited.";

[P2]: "For our team this is a great representation of what we could use. This is a good representation of a network that would work for us. A way to visualize this and interact with.";

Participants admitted that they perceived the traffic "going" between nodes, referring to the edges representing the count of sessions observed between these two nodes ("Relations" of groups, subgroups and nodes, as per the taxonomy):

[P1]: "They are clearly grouped, and I can see exactly where everything is going [which node has been connecting to what other nodes]."

[P4]: "This would help a lot. You can see the traffic leaving networks and so..."

[P5]: "This makes a lot of sense to me. I really like the visualization. I can see.. there's the first firewall, DMZ.. I can kind of understand how the network is built..."

[P2]: "This is useful... and this seems very utilizable. Useful in terms of what's reaching out to what. There's definitely usefulness in this for what we do here. [...] This makes much sense for what we do here in terms of usefulness and utility..."

Participants recognized the advantages of using such visualizations could provide to transfer knowledge about specific networks from senior analysts to trainees:

[P6]: "Since I've been here for 4 years, I've trained about 80 people. I think if we'd have something like that from the start, it would change their whole perception of how to [think of networks] and jump start [their ability to work the networks]. [...] I think a lot of analysts would have different views, that would depend on their knowledge base and artistic side also. [...] When you're using tool like this, when you build your network diagrams, you would like to have same setup, that way [when] you're looking at them on a pdf, you'd have the same layout."

Participants suggested capabilities (see correlation with taxonomy, described in Introduction and (Shneiderman, 1996)), that they would like to have at their disposal:

[P9]: "This is awesome! [...] As an analyst I would want to see [in addition to the visualization] what's happening, the [textual] details. [...] It is cool; you could definitely do a lot with it. [...] This is one of the coolest things I've ever seen. But I do need [additional, textual] information. As an analyst, I could definitely use this. [...] I could probably play with this all day.";

[P1]: "[It] would be nice to control how far apart they [nodes, groups] are. Would be easier to navigate between them. [...] [option to] change the icon of the node to something that would indicate the function of the entity. I would prefer to use colors for grouping the entities.";

[P3]: "Color-code the layers in the network. Any host that is associated [has had sessions] with those should also be color accordingly. 6DOF movement should be available.";

[P6]: "Lines should have arrows showing the direction of the sessions.";

## 5. Conclusion

This study captured cybersecurity analysts' impressions of a network topology presented as a stereoscopically-perceivable 3D structure. Overall, the impressions towards stereoscopically-perceivable 3D data visualizations were highly favorable. Multiple participants acknowledged that such 3D visualizations of network topology could assist in their understanding of the networks they use daily. Participants expressed a wish to integrate such visualization capabilities in their workflow. Prior experience with 3D displays had no influence on user

preferences, while participants with prior gaming experience adjusted quickly to the Oculus Touch motion controllers, suggesting that the relevant dexterity and muscle memory for gaming console controller usage helps users adjusting from those controllers to handling input devices for VR experiences. Further research is needed to understand what specific 3D data shapes would be useful and for which datasets (e.g. computer network topology) to create additional 3D visualization suitable for analysts' preferences and test the usefulness of those visualizations. Follow-up studies should evaluate operator performance in 3D environments.

## **Acknowledgements**

For all the hints, ideas and mentoring, authors thank Jennifer A. Cowley, Alexander Kott, Lee C. Trossbach, Jaan Priisalu, Olaf Manuel Maennel. This research was partly supported by the Army Research Laboratory under Cooperative Agreement Number W911NF-13-2-0045 (ARL Cyber Security CRA) and under Cooperative Agreement Number W911NF-16-2-0113 and W911NF-17-2-0083. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

## **Results**

- Arthur, K. W., Booth, K. S., & Ware, C. (1995, 7). Evaluating 3D Task Performance for Fish Tank Virtual Worlds. ACM Transactions on Information Systems, 11(3), 239-265. doi:10.1145/159161.155359
- Aukstakalnis, S. (2017). Practical Augmented Reality. Addison-Wesley.
- Ben-Asher, N., & Gonzalez, C. (2015). Effects of cyber security knowledge on attack detection. Computers in Human Behavior, 48, 51-61.
- Berry, S. P. (n.d.). The Shoki Packet Hustler. Retrieved from <http://shoki.sourceforge.net/>
- Best, D. M., Endert, A., & Kidwell, D. (2014). 7 Key Challenges for Visualization in Cyber Network Defens. In Proceedings of the Eleventh Workshop on Visualization for Cyber Security (pp. 33-40). ACM.
- Burnett, M. S., & Barfield, W. (1991). Perspective versus plan view air traffic control (ATC) displays - Survey and empirical results. International Symposium on Aviation Psychology, 6th. Columbus. Retrieved from <http://adsabs.harvard.edu/abs/1991STIA...9244967B>
- Dennehy, M. T., Nesbitt, D. W., & Sumey, R. A. (1994). Real-Time Three-Dimensional Graphics Display for Antiair Warfare Command and Control. Johns Hopkins APL Technical Digest, 15(2), 110-119.
- Ehrenstein, W. H., Spillmann, L., & Sarris, V. (2003). Gestalt Issues in Modern Neuroscience. In Axiomathes (pp. 433-458). Springer.
- Gaw, T. J. (2014, 4). 3D Information Visualization of Network Security Event. Munice, Indiana, USA: Ball State University. Retrieved from <https://pdfs.semanticscholar.org/f3fa/c8a059369b96202a70ceb19828c07444dc42.pdf>
- Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (2013). Cognitive Neuroscience: The Biology of the Mind, 4th Edition. W. W. Norton & Company.
- Gershon, P., Klatzky, R. L., & Lee, R. (2014). Handedness in a virtual haptic environment: Assessments from kinematic behavior and modeling. Acta Psychologica, 37-42.
- Gershon, P., Klatzky, R. L., Palani, H., & Giudice, N. A. (2016). Visual, Tangible, and Touch-Screen: Comparison of Platforms for Displaying Simple Graphics. Assistive technology: the official journal of RESNA, 28(1), 1-6. doi:10.1080/10400435.2015.1054566
- Gonzalez, C., Ben-Asher, N., Oltramari, A., & Lebiere, C. (2014). Cognition and technology. In Cyber defense and situational awareness, 93-117.
- Hodent, C. (2018). The Gamer's Brain; How Neuroscience and UX Can Impact Video Game Design. CRC Press.
- Hurter, C. (2016). Image-Based Visualization: Interactive Multidimensional Data Exploration. (N. Elmquist, & D. Ebert, Eds.) Morgan & Claypool.
- Inoue, D., Suzuki, K., Suzuki, M., Eto, M., & Nakao, K. (2012). DAEDALUS-VIZ: Novel Real-time 3D Visualization for Darknet Monitoring-based Alert System. VizSec (pp. 72-79). ACM.
- Kirk, A. (2016). Data Visualisation, A Handbook for Data Driven Design. Sage.
- Kullman, K., Cowley, J. A., & Ben-Asher, N. (2018). Enhancing Cyber Defense Situational Awareness Using 3D Visualizations. Proceedings of the 13th International Conference on Cyber Warfare and Security ICCWS 2018: National Defense University, Washington DC, USA 8-9 March 2018 (p. 369-378). Washington DC: Academic Conferences and Publishing International Limited.
- Lebreton, P., Raake, A., Barkowsky, M., & Le Callet, P. (2012). Evaluating Depth Perception of 3D Stereoscopic Videos. IEEE Journal of Selected Topics in Signal Processing, 6(6). Retrieved from <http://ieeexplore.ieee.org/document/6269042/>
- Lynn, W. J. (2010). Defending a New Domain: The Pentagon's Cyberstrategy. Foreign Affairs, 89(5), 97-108. Retrieved from <https://city.rl.talis.com/items/71ACA705-9A0A-8C2B-EAD3-5FF314AAC847.html>
- Maddix, K. (2015). Big Data VR Challenge – Winners! Retrieved from Masters of Pie: <http://www.mastersofpie.com/big-data-vr-challenge-winners/>
- Marty, R. (2008). Applied Security Visualization.

- Meyerovich, L., & Tomasello, P. (2016). Display Relationships Between Data. *IQT Quarterly*, 7(4).
- Munzner, T. (2014). *Visualization Analysis & Design*. A K Peters/CRC Press.
- Payer, G., & Trossbach, L. (2015). The Application of Virtual Reality for Cyber Information Visualization and Investigation. In M. Blowers, *Evolution of Cyber Technologies and Operations to 2035* (Vol. 63, pp. 71-90). Springer. doi:10.1007/978-3-319-23585-1\_6
- Pruett, C. (2017, 05 17). Vision 2017 - Lessons from Oculus: Overcoming VR Roadblocks. Retrieved from <https://youtu.be/swA8cm8r4iw?t=9m42s>
- Rajivan, P., Konstantinidis, E., Ben-Asher, N., & Gonzalez, C. (2016). Categorization of Events in Security Scenarios: The Role of Context and Heuristics. *Human Factors and Ergonomics Society Annual Meeting*, 60(1), 274-278.
- Reda, K., Febretti, A., Knoll, A., Aurisano, J., Leigh, J., Johnson, A., . . . Hereld, M. (2013). Visualizing large, heterogeneous data in hybrid-reality environments. *IEEE Computer Graphics and Applications*, 33(4), 38-48.
- Reuille, T., Hawthorne, S., Hay, A., Matsusaki, S., & Ye, C. (2015). OpenDNS Data Visualization Framework. Retrieved from OpenGraphiti: <http://www.opengraphiti.com/>
- Schneider, W., Dumais, S. T., & Shiffrin, R. N. (1982). *Automatic and Control Processing and Attention*. Illinois: University of Illinois. Retrieved from <http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&ADNumber=A115078>
- Sethi, A., & Wills, G. (2017). Expert-interviews led analysis of EEVi — A model for effective visualization in cyber-security. *IEEE Symposium on Visualization for Cyber Security* (pp. 1-8). Phoenix, AZ, USA: IEEE.
- Shearer, G., & Edwards, J. (2018). Vids: Version 2.0 Alpha Visualization Engine. Adelphi: US Army Research Laboratory.
- Shneiderman, B. (1996). The eyes have it: a task by data type taxonomy for information visualizations. *Proceedings 1996 IEEE Symposium on Visual Languages*. Boulder, CO, USA, USA: IEEE. doi:10.1109/VL.1996.545307
- Smallman, H. S., St. John, M., Oonk, H. M., & Cowen, M. B. (2001). Information availability in 2D and 3D displays. *IEEE Computer Graphics and Applications*, 21(5), 51-57. Retrieved from <http://journals.sagepub.com/doi/pdf/10.1518/001872001775992534>
- St. John, M., Cowen, M. B., Smallman, H. S., & Oonk, H. M. (2001). The Use of 2D and 3D Displays for Shape-Understanding versus Relative-Position Tasks. *Human Factors*, Spring, 79-98. Retrieved from <http://journals.sagepub.com/doi/pdf/10.1518/001872001775992534>
- U.S. ARL. (2018, 02 12). SEEING THE CYBERTHREAT. Aberdeen Proving Ground, Maryland, USA: U.S. Army Research Laboratory. Retrieved from [https://dodstem.us/sites/default/files/lab-narratives/Seeing-the-Cyberthreat\\_0.pdf](https://dodstem.us/sites/default/files/lab-narratives/Seeing-the-Cyberthreat_0.pdf)
- van Riel, J.-P., & Irwin, B. (2006). InetVis, a Visual Tool for Network Telescope Traffic Analysis. *AFRIGRAPH 2006*. Cape Town: Association for Computing Machinery, Inc.
- Ward, M. O., Grinstein, G., & Keim, D. (2015). *Interactive Data Visualization: Foundations, Techniques, and Applications*, Second Edition. A K Peters/CRC Press .
- Ware, C., & Franck, G. (1996, 4). Evaluating Stereo and Motion Cues for Visualizing Information Nets in Three Dimensions. *ACM Transactions on Graphics*, 15(2), 121-140. doi:10.1145/234972.234975
- Wu, Y., Xu, L., Chang, R., Hellerstein, J. M., & Wu, E. (2018). Making Sense of Asynchrony in Interactive Data. *JOURNAL OF LATEX CLASS FILES*, 14(8).

# The Balanced Digitalization and Digital Security: Case of Regional Authorities

Tuija Kuusisto<sup>1, 2</sup> and Rauno Kuusisto<sup>1, 3, 4</sup>

<sup>1</sup>Ministry of Finance and National Defence University, Helsinki, Finland

<sup>2</sup>University of Jyväskylä, Finland

<sup>3</sup>The Finnish Defence Research Agency, Riihimäki, Finland

<sup>4</sup>National Defence University, Helsinki, Finland

[tuija.kuusisto@vm.fi](mailto:tuija.kuusisto@vm.fi)

[rauno.kuusisto@mil.fi](mailto:rauno.kuusisto@mil.fi)

**Abstract:** The emerging digital infrastructure enables the public authorities to shape their processes and to create attractive digital services for the citizens and the business actors. The public processes, services and infrastructure, however, engage with global and local public and private digital infrastructure and service providers. The complex comprehensiveness of the digital infrastructure and the services challenges the public authorities and create new types of security risks. The achieving of the benefits of the digitalization of public processes and services without increasing security risks requires the adopting of novel approaches to digital security. The paper refers to a framework that aims to balance the digitalization and digital security of society. The approach follows the complex adaptive system and social system theories. The paper demonstrates the framework with widely known digital service indexes and digital security indexes. The paper applies the referred framework and the results of its demonstration in a case study about the governing of the digitalization resources and activities of the regional authorities. The case study was related to a major structural reform. The aim of the reform was to form and launch the operations of new counties. The means of the reform included the co-creation of new types of digital processes and services in collaboration and with the citizens and the business actors as well as with the central government. The empirical data of the case study included the ICT costs, digitalization efforts, shared ICT services and digital security situation of the regions. The central government analyzed the empirical data for the simulation of the financial negotiations between the central government and the regions. The results of the case study show that the framework supported the outlining of the contents of the empirical data so that both the digitalization and digital security aspects were concerned and visualized. The authorities will apply the results of the analysis for the governing of the regions.

**Keywords:** complex systems, modelling of digital security, cyber security, system modelling, digital services, regions

---

## 1. Introduction

The emerging digital infrastructure enables the public authorities to create attractive digital services for the citizens and the business actors. The infrastructure, including IoT devices and autonomous systems, provides splendid opportunities to the authorities to improve the public processes and services. Merriam-Webster (2019) gives a traditional definition of digitalization as 'The process of converting something to digital form'. Gartner's (2019) statement of digitalization follows a broadly accepted modern view as it defines digitalization as 'The use of digital technologies to change a business model and provide new revenue and value-producing opportunities'. Recently the public sector authorities have increasingly considered this modern view. They have applied information and communication technology and information management and analysis for shaping the public processes and services. The authorities have usually attempted to enhance the citizen and the business actor experience of the services in addition to reducing the public sector total costs. The digitalization efforts have often included the creation of new digital processes, services and infrastructure. This has required continuous change management of the civil servants, citizens and the business actors as well as processes and technology.

The public processes, services and infrastructure engage with global and local public and private digital infrastructure and service providers. The complex comprehensiveness of the digital infrastructure and the services challenges the public authorities and creates new types of security risks. Digital security is an emerging term referring to the security view on digitalization and digital services. Typically, it includes information and cyber security and data protection. Often it covers risk management, and preparedness and contingency planning as well. The achieving of the benefits of the digitalization of public processes and services without increasing security risks requires the adopting of novel approaches to digital security.

First, the paper refers to a framework that aims to balance the digitalization and digital security of society (Kuusisto & Kuusisto, 2017). The approach follows the complex adaptive system (Holland, 1996) and social system theories (Parsons, 1951), (Habermas, 1984 & 1989). The approach adopts a social system model that aim

is to increase understanding about the evolving features and culture of the digital era (Kuusisto, 2004). The paper demonstrates the framework with widely known digital service indexes and digital security indexes.

The indexes contain indicators for measuring information society (ITU, 2017a), eGovernment (UN, 2018), the use of information and communication technologies (WEF, 2018) as well as the commitment to cyber security (ITU, 2017b). The demonstration of the framework shows the complex nature of the forming of the indicators. The comprehensive context and the purpose of the use of the indicators shall guide the forming of indicators.

Strategic management is one of the purposes to use indicators. The paper applies the referred framework and the results of its demonstration in a case study about the governing of the digitalization resources and activities of the regional authorities. The case study was related to a major structural reform. The target of the reform was to form and launch the operations of new administrative structure, autonomous counties. The reform was supposed to cover the structure, services and funding of health and social services. In addition, some of the tasks of the central government and the municipalities were planned to be transferred to the counties. The central government financed the planning and was supposed to finance the operating of the counties. The aims of the reform were to provide the citizens and inhabitants with more similar services, to reduce the variations in people's health, and to curb the rising costs. The means to attain these aims included the more efficient use of information technology. The reform was cancelled in March 2019 (Regional Reform, 2019)

The empirical data of the case study included the ICT costs, digitalization efforts, shared ICT services and digital security situation of the regions. A group of regional authorities and public ICT agencies whose aim was to provide the regions with shared ICT services gathered the empirical data in the beginning of 2019. The means of the reform included the co-creation of new types of digital processes and services in collaboration and with the citizens and the business actors as well as with the central government. This was considered to require collaboration between the regions and with the central government as well as with the citizens, the business actors and national and global ICT service providers and the security authorities.

Finally, the paper outlines the results of the analysis of the empirical data. The central government authorities analyzed the empirical data for the simulation of the financial negotiations between the central government and the regions. The results of the case study show that the framework supported the outlining of the contents of the empirical data so that both the digitalization and digital security aspects were concerned and visualized. In addition, the study seemed to raise the awareness of the digital security requirements. The authorities will apply the results of the analysis for the governing of the regions.

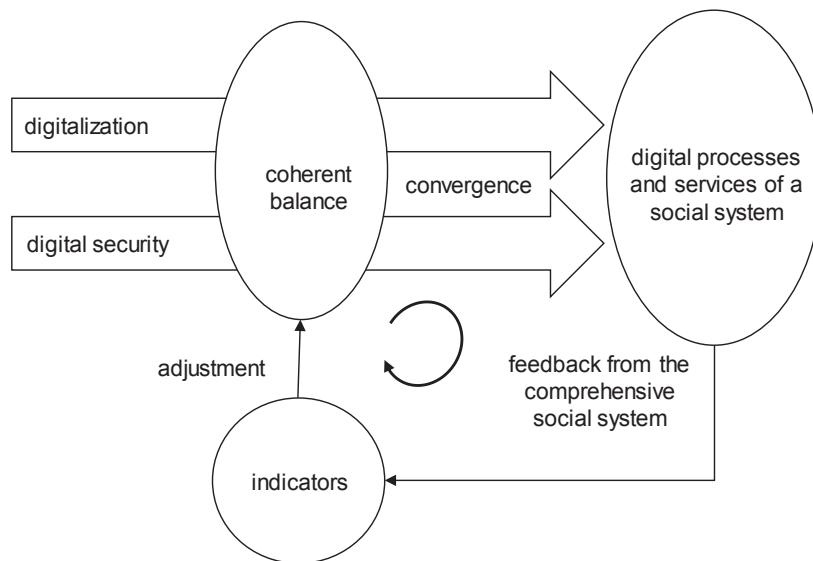
## **2. The balanced digitalization and digital security framework**

Figure 1 illustrates the balanced digitalization and digital secure framework. It consists of a group of complex and emergent entities adapting to their environment over time (Holland 1996). A social system is, e.g., government, a regional authority, a business actor or a citizen or a group of all or some of them. A social system gives feedback as indicators. They are applied for adjusting the digitalization and digital security efforts. Parsons (1951) states that a social system has an initial and a goal state. In addition, he argued that the interaction orientation of the system is internal and external. In the spirit of the complex adaptive systems theory (Holland 1996) it can be argued that the initial state is being shaped to the goal state by information flows in the framework. Information is flowing continuously through the adjusting of digitalization and digital security efforts to the digital processes and services of a social system and as indicators to the adjusting function.

Figure 1 shows how the coherent balance is achieved by adjusting the digitalization and the digital security efforts in parallel (Kuusisto & Kuusisto 2017). As a result, digitalization and digital security are converging. This convergence guides the designing and implementing of the digital processes and services in such a way that both the digitalization target level and the security requirements can be reached.

As outlined in Figure 1, the indicators have to address digitalization and digital security. The selecting of the indicators shall be implemented in the context of the actor that is under concern, e.g., the government or a region. The indicators selected shall be relevant for the case and they shall be balanced with each other, as well. The government authorities at the national level can apply international digitalisation indexes and the cyber security situation indexes for the selecting of the indicators (Kuusisto & Kuusisto 2017). The international indexes that can be applied at the national level include UN (2018) eGovernment survey, ITU's ICT Development Index

(IDI) (2017a), World Economic Forum's (2018) Network Readiness Index, ITU's (2017b) Global Cybersecurity Index and Estonia's National Cyber Security Index (NCSI, 2019). Both the digitalization and cyber-security indicators have to be applied for balancing the digitalization and digital security efforts.



**Figure 1:** The balanced digitalization and digital security framework

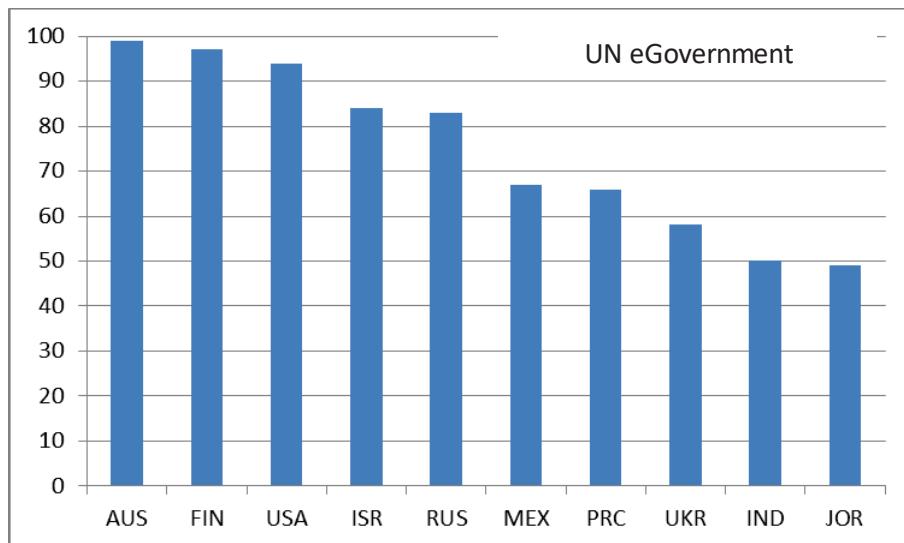
UN (2018) eGovernment survey consists of eGovernment development index and eParticipation index. ITU's (2017b) Global Cybersecurity Index contains legal, technical, organizational, capacity building, and cooperation views on the national cyber-security. The aim of Estonia's NCSI (2019) is 'to measure the preparedness of countries to prevent cyber threats and manage cyber incidents'. The categories of the index are: Cyber security policy development, cyber threat analysis and information, education and professional development, contribution to global cyber security, protection of digital services, protection of essential services, E-identification and trust services, protection of personal data, cyber incidents response, cyber crisis management, fight against cybercrime and military cyber operations. The results of the surveys by these indexes can be applied for identifying the major phenomena and improvement needs of a country.

### **3. The demonstration of the balanced digitalization and digital security framework**

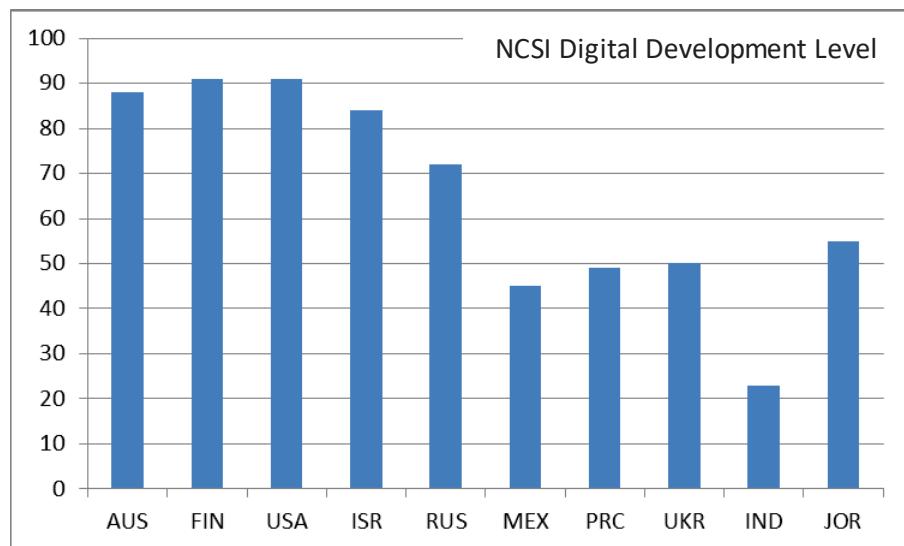
The paper demonstrates the framework by applying UN (2018), ITU (2017b) and NCSI (2019) indexes. NCSI's Digital Development Level (DDL) is based on ITU (2017a) and WEF (2018). In this paper, UN (2018) and NCSI's (2019) DDL are referred as "digitalization" indicator sets and ITU (2017b) and NCSI are referred as "digital security" indicator sets. The selected indicator sets have been formed for the measuring of a certain aspects to the digitalization and digital security. The indicator sets are partly overlapping but all of them contain unique indicators and indicator weights. In addition, they use partly overlapping data sources. Obviously, the results and country rankings by these indicator sets are not similar.

Next, the paper focuses on the analysis of the results of these indicator sets. For the analysis, the scale of the country ranking results were converted. The first country in the ranking received score 100 and the last one received score 1. This aim of this was to visualize the country rankings in a comparable way. Ten countries were included in the visualization of the analysis. Figures 2 and 3 show the results of the digitalization indicator sets. The countries are presented in Figures 2, 3, 4, and 5 in the order of the country ranking results of the UN eGovernment Survey (2018). Figures 4 and 5 show the results of the digital security indicator sets.

When comparing the abstract patterns that Figures 2 and 3 outline, it can be observed that these patterns are alike. However, the country ranking results of UN eGovernment Survey (2018) visualized in Figure 2 and NCSI's (2019) DDL visualized in Figure 3 are clearly not similar. The contents of these indicator sets have to be known for understanding what the indicators measure and what causes the differences in the country ranking results. The results should not be used for strategic-level decision making without understanding the contents of the indicators and source data.



**Figure 2:** UN eGovernment Survey (2018), digital services indicator set



**Figure 3:** NCSI's (2019) Digital Development Level, digital services indicator set

The country ranking results of ITU Global Cybersecurity Index (2017b) visualized in Figure 4 and National Cyber Security Index (2019) results visualized in Figure 5 are somewhat different. The country positions vary more in these rankings than in the digitalization rankings. This may just be a coincidence. The sample data sets are quite small. On the other hand, the contents of the indicator sets and the way the evaluations are implemented are different. Therefore, the differences in the ranking results might indicate that the complex security evaluation concepts are not yet internationally defined, organized and regulated to a sufficient level.

As a conclusion, it can be observed that the country positions in the rankings vary more than just faintly. In addition, the correlation between the country rankings of the digitalization and digital security indicator sets is not complete. A hypothesis can be made that some countries have a noticeable unbalance between digitalization and digital security. In addition, it can be assumed that countries are directing cyber security efforts based on their national focus.

The very modest analysis above aims demonstrating that when planning and using indicators for supporting governance and management, a thorough analysis is needed for selecting or forming relevant and suitable indicator sets. At very general level, an indicator set should serve the context under concern and have proven balance between digitalization and digital security aspects.

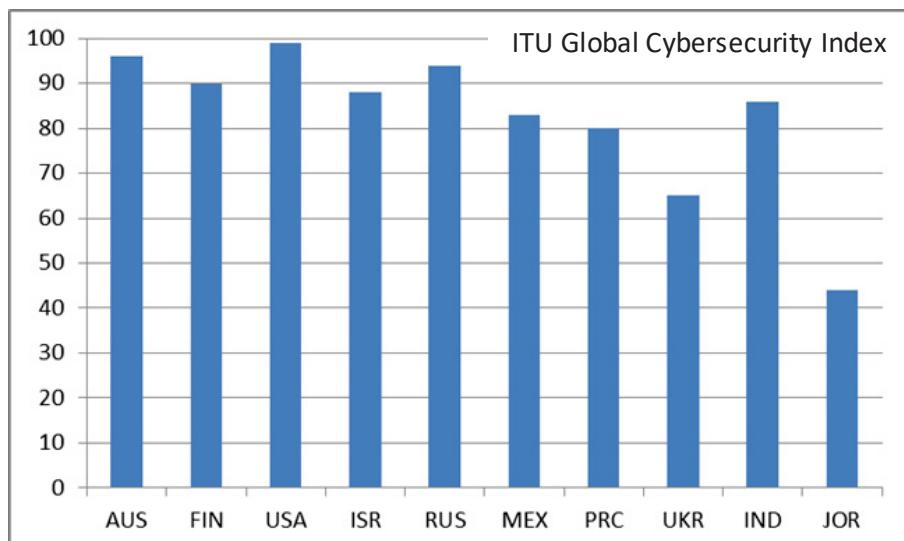


Figure 4: ITU Global Cybersecurity Index (2017b)

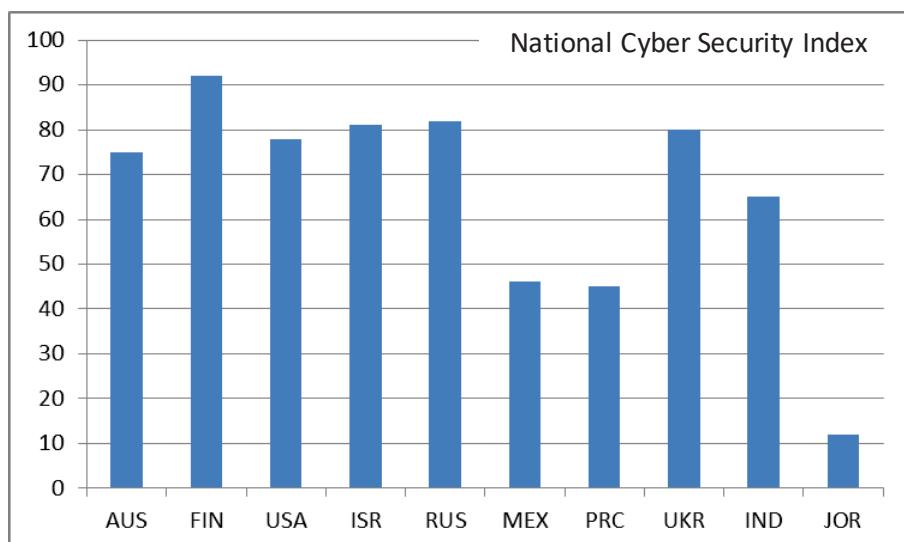


Figure 5: National Cyber Security Index (2019) results

As noted before, the balance between digitalization and digital security is relevant. The security level should be as high as the value of security to the organization require and the digitalization efforts are targeting. Over-digitalization leads to security gaps and over-security cause unnecessary costs. The more relevant both of the aspects are the more influential the organization is. When forming the indicators, we should be aware of the required service and security levels, the nature of the actor's activity and the overall context where the interacting actors are performing their activities in the digital world. When using the indicators, we should explicitly know how they are formed and what data they are using. The indicators shall be constructed to be relevant in the actor's governance, development and management context. Next, we demonstrate briefly how these simple principles succeeded in the case of aiming to develop the first attempt to form digital services - digital security overall service level and their internal balance in context of a major government reform.

#### 4. Regional authorities case study

The balanced digitalization and digital security framework and the results of its demonstration were applied in a case study. The aim of the case study was to illustrate the first efforts to outline the rough contents of indicators relevant for supporting the governance, development, planning and management of the regions digitalization and digital security activities. The case study consisted of the analysis of the ICT and digitalization situation data of the regions. The central government authorities defined the contents of the analysis and the data describing the situation of the regions. Some of the regional authorities preparing the major reform contributed to the content definitions as well. Seven public institutions or agencies were planned to design,

implement and produce the shared ICT services for the regions. These organizations contributed to the outlining of the content definitions too.

The authorities designed the contents of the situation data and data analysis by following the agile approach. First, the authorities studied some of the international and national eGovernment evaluations, ICT surveys and standards, and practices of a global consulting company. They decided to include the financial figures, the major digitalization efforts and information systems, and the digital security situation reports to the situation data. The categories of the situation data of the regions as well as the service providers were:

- The total estimated ICT budget in 2019,
- the major digitalization activities,
- the major development programs and projects containing digitalization efforts or ICT, and their costs,
- the major current information systems and their operating costs, and
- the survey of cyber and information security situation.

The total estimated ICT budget contained several categorizations formed according to the current global practices. These included categorization to the hardware, software, human resources and internal and external sourcing costs as well as to the maintenance and development costs. In addition, the previous year ICT costs were asked to be reported as Capital Expenditure (CAPEX) and operating expenses (OPEX). The regional authorities involved in the major reform created and collected the data in the beginning of 2019. The authorities collected the data based on the tasks and responsibilities of the regions. The shared ICT service providers delivered data about their current or future shared ICT services to the regions.

The major observation about the financial figures was that the quality of the data was low. The definitions of ICT costs were region specific. In addition, some of the regions were not able to collect the required financial figures at all. Therefore, the central government authorities could not assess or compare the digitalization situations of the regions based on these figures. For example, there were several reasons for a high value in the total ICT costs per inhabitant of a region compared to the average ICT costs per inhabitant. These reasons included a high digitalization degree as well as duplicate information systems. Thus, the central government authorities considered that the financial data are insufficient, inaccurate and incomplete for the statistical analysis.

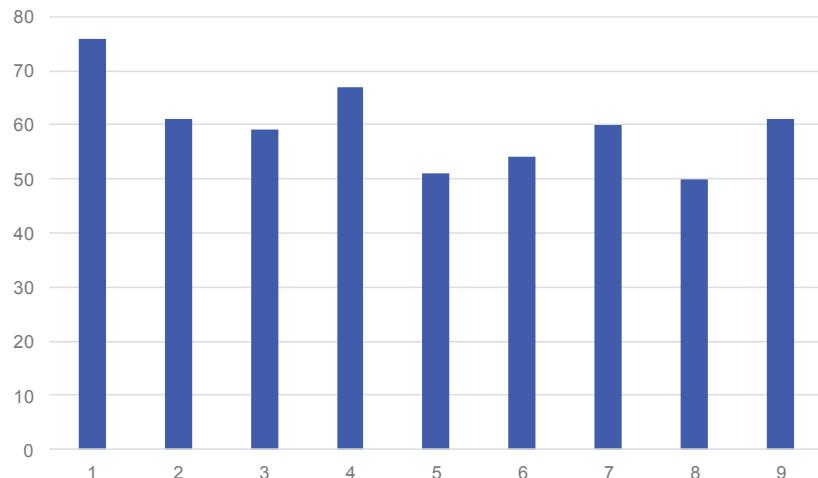
The central government authorities considered that the study of the digitalization efforts of the regions improved their awareness of the digitalization degree of the regions. The regions implementing or already proving the personnel and the inhabitants with the modern health care systems are obviously proceeding in digitalization. The analysis of the catalogues of the major current information systems, however, did not have a significant impact on the awareness of the digitalization degree.

The major observation of the cyber and information security survey of the regions was that the management, guidelines and practices of the health care organizations meet the basic level requirements. The assumption is that the survey improved the cyber and information security awareness of the regional authorities that were preparing the major reform. In the future, the central government authorities could govern the regions to extend the cyber and information security culture of the health care authorities to the other regional functions.

The major observations of the shared ICT service providers' situation data show that the maturity of the service provider and the phase of the life cycle of a digital product or service explains well the accuracy of the financial figures. This applies to the level of the cyber and information security as well. Novel organizations launched a few months ago had not yet formed the region-based commissioning plans of their services and so they could not estimate region-based costs of the shared services or products. In addition, they had not established the risk assessment based cyber and information security controls. A minor observation was that many of the shared ICT service providers reported that they had not exercised their contingency plans. This was not widely recognized before the survey. The assumed benefit of the collecting and analyzing of the situation data is that it increases the cyber and information security awareness of the regional authorities.

In addition to the data gathered from the regions and shared ICT service providers, the central government authorities decided to pilot the use of The Digital Economy and Society Index (DESI, 2019) in the regional level. DESI has five dimensions: Connectivity, Human Capital, Use of Internet Services, Integration of Digital

Technology and Digital Public Services. Statistics Finland defined a regional level pilot DESI with three dimensions that are relevant at the regional level: Connectivity, Human Capital and Use of Internet Services. The method and results are described in (MOF, 2019). The regional DESI pilot has the inhabitant view on the digitalization degree of a region. It illustrates the digital potential of the regions. The results of the regional DESI pilot are shown in Figure 6. As visualized in Figure 6, some of the regions have challenges in their digital economy. The more detailed study of the values of the composite regional DESI indicator dimensions (MOF, 2019) shows that especially the connectivity dimension has low values on some of the regions.



**Figure 6:** The digital potential of regions (MOF, 2019)

The regional authorities and the shared ICT service providers were involved the planning of a major structural reform. They were acting in a complex, evolving environment that yields diversity. The financial impacts of digitalization and digital security efforts are divergent. This challenged the collecting and analysis of the situation data. The aim was, however, that the publishing of the financial figures, digitalization efforts and digital security situation of the regions would adjust the coherent balance of the digitalization and digital security activities of the regions. The divergent definitions and approaches of the basic concepts, terms, and services would convert to understanding that is more similar. As a result, the regions would have secure shared digital services.

However, the case study showed that it is crucial to be able to select or form the indicators in such a way that they describe the actors' potential to develop and implement digital services and assure their security. This requires knowledge and competence about the indicators and their interpretation.

## **5. Conclusions**

This paper refers to a theoretically motivated approach: the balanced digitalization and digital security framework outlined in Figure 1. It shows how the coherent balance of digitalization and digital security is gained by the joint adjusting of the digitalization and digital security efforts. This requires the development of the digital services and digital security in parallel. In addition, the governing of the development needs to be supported by accurate enough and relevant governance toolsets understood and accepted by all the actors involved. The result as its best is the convergence of digitalization and digital security for the governing of the design, implementation and providing of the digital services of society.

The digital services shall be considered as a social system - the system of relevant action by relevant actors in the defined context supported by technological structures and services. Theoretically, this can be considered as a complex adaptive system (CAS) and treated as one in ever evolving context. Thus, the balanced digitalization and digital security framework consists of a group of complex and emergent entities adapting to their environment over time. It will help to identify the most relevant development issues during that evolution. This social system gives feedback as indicators. They shall be used for directing the adjustment of the digitalization and digital security efforts. The nature and the origin of the data of the indicators are critical. It is obvious that the content and the selection of the indicator data will change during the system evolution. The indicator system shall be considered as a part of the strategic guidance of the comprehensive governance.

## **References**

- DESI (2019) *The Digital Economy and Society Index*, Retrieved on the 22nd of February 2019 from  
<https://ec.europa.eu/digital-single-market/en/desi>
- Gartner (2019) *IT Glossary*. Retrieved on 3rd of January 2019 from <https://www.gartner.com/it-glossary>
- Habermas, J. (1984) *The Theory of Communicative Action, Volume 1: Reason and the Rationalization of Society*, Boston, MA: Beacon Press.
- Habermas, J. (1989) *The Theory of Communicative Action, Volume 2: Lifeworld and System: A Critique of Functional Reason*, Boston, MA: Beacon Press.
- Holland, J.H. (1996) *Hidden Order: How Adaptation Builds Complexity*, Cambridge, MA. Perseus Books.
- International Telecommunication Union (ITU) (2017a) *ICT Development Index*. Retrieved on the 30<sup>th</sup> of April 2019 from  
<http://www.itu.int/net4/itu-d/idi/2017/index.html>
- International Telecommunication Union (ITU) (2017b) *The Global Cybersecurity Index*. Retrieved on the 3rd of January 2019 from  
[https://www.itu.int/dms\\_pub/itu-d/opb/str/D-STR-GCI.01-2017-PDF-E.pdf](https://www.itu.int/dms_pub/itu-d/opb/str/D-STR-GCI.01-2017-PDF-E.pdf)
- Kuusisto, R. (2004) *Aspects on availability*, Edita Prima Oy, Helsinki, Finland.
- Kuusisto, R. (2008). *Analyzing the Command and Control Maturity Levels of Collaborating Organizations*. In: Proceedings of 13th International Command and Control Research and Technology Symposium (13th ICCRTS), Bellevue, WA, USA, 17-19 June 2008
- Kuusisto, T., Kuusisto, R., Roehrig, W. (2015) "Situation Understanding for Operational art in Cyber Operations". In Abouzakhar, N. (ed.) Proc of the 14th European Conference on Cyber Warfare and Security ECCWS-2015, Hatfield, UK, 2.-3.7.2015, pp. 169-178, Published by Academic Conferences and Publishing International Limited Reading, UK 44-118-972-4148, www.academic-publishing.org
- Kuusisto, T., Kuusisto, R. (2017) "Security Culture in Digital Inter-Organizational Ecosystems". In Scanlon, M. & Le-Khac, N.-A. (eds.) Proc. of the 16th European Conference on Cyber Warfare and Security, ACPI, 29-30 June 2017, pp. 216-223
- Merriam-Webster (2019). Merriam-Webster online dictionary. Retrieved on the 3rd of January 2019 from  
<https://www.merriam-webster.com>
- Ministry of Finance (MOF) (2019), *DESI results at regional level*, Retrieved on 19th of February 2019 from  
[https://alueudistus.fi/artikkeli/-/asset\\_publisher/maakuntien-ict-tilannekuva-a-taydennetty-digitalisoitumisindikaattorilla\\_in\\_Finnish](https://alueudistus.fi/artikkeli/-/asset_publisher/maakuntien-ict-tilannekuva-a-taydennetty-digitalisoitumisindikaattorilla_in_Finnish)
- NCSI (2019). *National Cyber Security Index*. Retrieved on the 3rd of January 2019 from <https://ncsi.ega.ee/methodology/>
- Parsons, T. (1951) *The Social System*, Free Press, Glencoe, IL.
- Regional Reform (2019). *Regional Government, Health and Social Services Reform*. Retrieved on the 30th of April 2019 from  
[www.regionalreform.fi](http://www.regionalreform.fi)
- United Nations (UN) (2018) *E-Government Survey 2018, Gearing E-Government to Support Transformation Towards Sustainable and Resilient Societies*. United Nations, Economic & Social Affairs. Retrieved on the 3rd of January 2019 from  
[https://publicadministration.un.org/egovkb/Portals/egovkb/Documents/un/2018-Survey/E-Government%20Survey%202018\\_FINAL%20for%20web.pdf](https://publicadministration.un.org/egovkb/Portals/egovkb/Documents/un/2018-Survey/E-Government%20Survey%202018_FINAL%20for%20web.pdf)
- World Economic Forum (WEF) (2018) *Network Readiness Index*. Retrieved on the 30th of April 2019 from  
<http://reports.weforum.org/global-information-technology-report-2016/networked-readiness-index/>

# Operational Tempo in Cyber Operations

Antoine Lemay<sup>1</sup> and Sylvain Leblanc<sup>2</sup>

<sup>1</sup>Cyber Defence Corporation, Montréal, Canada

<sup>2</sup>Department of Electrical and Computer Engineering, Royal Military College of Canada, Kingston, Canada

[antoine.lemay@live.ca](mailto:antoine.lemay@live.ca)

[sylvain.leblanc@rmc.ca](mailto:sylvain.leblanc@rmc.ca)

**Abstract:** The statement that cyber-attacks occur at “electron speed” is often offered as a truism in the study of cyber warfare. A reasonable consequence of this statement is that effective cyber defences should also respond almost instantaneously, creating a view of conflict in the cyber domain as a war of algorithms where processing speed and reaction time reign supreme. Before accepting this however, it is important to ask if this hypothesis is supported by facts. In this paper, we investigate what is the true tempo of cyber operations and theorize about the significance of this tempo for both offensive and defensive cyber operations. By studying historical cases of state-level cyber conflicts, we will show that while some elements of the kill-chain are near-instantaneous, crucial steps such as reconnaissance, weaponization, delivery and command and control (including lateral movement) occur at a much slower pace. This leads us to conclude that the general tempo of cyber operations is in fact quite deliberate, limited by the necessity to secure intelligence to tailor an exploit targeting a vulnerability, by the lead-time to create an effective weapon, by the need for a number of attempts to secure a foothold and by the extensive work required to expand access once within the network. Furthermore evolving situations as the attack unfolds, along with response by defenders, require attackers to change tactics and retool dynamically to respond to the evolving environment. Such changes are driven by human decision and, as such, throttle back the cyber operation tempo to human speed, giving time for human defenders to react. Based on these observations, we assert that this deliberate pace is at odds with the high tempo required of military operations in times of heightened conflicts such as war. In turn, this suggests that military units carrying out offensive cyber operations should be expected to perform this work in periods of relative calm in order to gain the access to adversary systems required during high tempo operations in times of conflict and war. Conversely, defensive cyber operations should focus on the more deliberate elements of the kill-chain and deny the ability of attackers to maintain their presence, rather than focus all their attention on blunting the exploitation and installation components. In that sense, the focus for defenders should be the thoroughness of their ability to deny sustained access rather than the speed of the response.

**Keywords:** cyber tempo, response time, cyber operations case studies, cyber kill-chain

---

## 1. Introduction

In the cyber environment, things move at the speed of electrons. An attack coming from across the globe will reach its target in the time the electronic signal can be transmitted, which can be minimal using fibre optic lines. As such, the idea that cyber-attacks occur at network speed is often offered as a truism. A corollary to this argument is that defence, to be effective, must operate at similar speed. This can often be observed as the motivating factor behind research in cyber defence AI (Rehman and Saba, 2014; Tyugu, 2011). After all, if things occur at network speed, human reaction time would be orders of magnitude too slow to mount an adequate defence; if we accept the argument that attacks happen at such a high rate of speed, fully automated defences are required.

We must ask ourselves however, if this assumption holds. Do cyber-attacks really occur at network speed, precluding human contribution to the defence, or do they follow a more deliberate approach? In this paper, we review the tempo of publicly disclosed cyber operations in order to draw conclusions about the speed of cyber-attacks, and therefore draw inferences regarding the speed required for the defence.

The paper starts by defining the concept of operational tempo and reviewing why it matters to the success of operations. This is followed by a review of cyber operations from publicly available literature particularly focussing on operational tempo. This is followed by an analysis providing insight into the speed of cyber operations. Finally, a brief conclusion is presented.

## 2. Why tempo?

In order to understand the concept of *tempo*, it is necessary to discuss manoeuvre warfare. In a manoeuvrist context, one presents the adversary with a sudden changes occurring fast enough to prevent that adversary from orienting to the changing situation, and thereby making it difficult to effect proper decision-making (Pech

and Durden, 2003). In more concrete terms, an adept of manoeuvre regards attrition-based strategies where strength is pitted against strength until one side's strength is depleted as wasteful. Instead, they would prefer pitting one's strength against the adversary's weakness. In order to do so however, we must ensure that the adversary cannot react by redeploying resources to shield their weakness, hence the focus on the disruption of decision-making.

One of the ways through which the adversary's decision-making can be disrupted is by using speed to overwhelm it. An historical example of this process is the turning of the Maginot line by German forces in World War II. As German armoured division where crashing through France, French Generals behind the lines were producing new defensive plans, which were made obsolete even before they were issued by the speed of the German advance (Allcorn, 2012). This concept was later formalized by Boyd in his famous OODA (*Observe – Orient – Decide – Act*) loop (Boyd, 1996). Boyd's claim was that, by being able to cycle through the loop faster than an opponent, it was possible to gain a tactical advantage. If you were able to figure out where you are and go to the place you should be before your opponent could react, you would be able to gain the "high ground". More abstractly, we could argue that manoeuvre started by trying to outflank your opponent to get behind him in two dimensions, then went on to going over him with the advent of the air and space domain for three-dimensional manoeuvre. The use of tempo could be akin to four-dimensional manoeuvre: put your strength at a time where the adversary shows a weakness.

There is no doubt that conventional militaries are heavily invested in tempo. The Revolution in Military Affairs (RMA) could be summed up as an increase in digitization in order to allow one to operate at a faster pace than less technologically advanced opponents (Mazarr, 1994). In that sense, the essential question regarding the speed of cyber-attacks boils down to the following: is the tempo of cyber operations too fast for human defenders? To answer this question, we can study cyber-attacks that were disclosed in the public domain. In order to better understand these attacks however, we must first examine the progression of a cyber-attack.

### 3. The cyber kill-chain

One of the most commonly used frames of reference used to understand cyber-attacks is the cyber kill-chain (Hutchins et al., 2011). The kill-chain model identifies the steps that attackers must take to cause harm on their target. The kill-chain identifies seven steps, as illustrated in Figure 1, which we illustrate with a representative example in the text that follows.

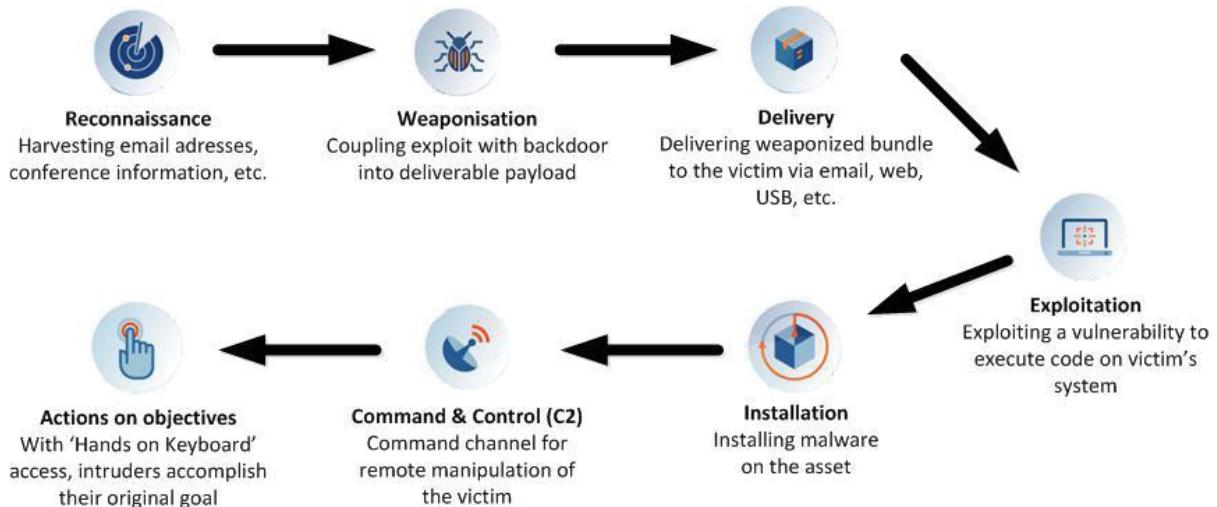


Figure 1: Cyber kill-chain (adapted from (Hutchins et al., 2011))

In the *Reconnaissance* phase, attackers gather the information necessary to proceed with their attacks. For example, they might harvest email addresses, they might comb social media network to gather information necessary in order to create a convincing spear phishing vector, or they might fingerprint the Internet Browser the potential victim is using in order to create or acquire an exploit targeting that specific software.

In the *Weaponisation* phase, the attackers need to develop software that will couple an exploit, which targets a vulnerability of the intended victim, with a means of delivering that exploit. The attacker's aim in weaponization is to achieve successful penetration of the victim's network resulting in a long-lasting presence on that network;

such a long-lasting presence is the very definition of an advanced persistent threat (APT) which is the impetus behind the development of the Cyber kill-chain model. As an example of weaponisation, an APT can create a document which, when accessed by a victim tricked into opening it, runs a macro that installs a backdoor that allows the attackers remote access to the machine. This converts the exploitation of a vulnerability (a human is vulnerable to social engineering) into a presence on the network. It is important to note that the software developed in weaponization often requires customization for a specific operation; it would be expected that, for example, the exploit delivery method need to be modified to ensure that it can successfully bypasses the particular cyber defences put in place by the defenders; it is therefore clear that the *Weaponisation* phase depends on the information gathered during the *reconnaissance* phase..

The *Delivery*, *Exploitation* and *Installation* phases are usually near-simultaneous and required little to no action on the part of the APT. For example, when the victim receives the spear phishing email (*Delivery*), its vulnerability to social engineering is exploited (*Exploitation*), along with the particular system vulnerability targeted by the cyber-weapon, which results in the *Installation* of the malicious software or malware. In other words, the *delivery* is the transport of the exploit developed in the *Weaponisation* phase to the target. Once the victim receives the exploit, *exploitation* occurs when the vulnerability is triggered when the malware is run, resulting in the *Installation* which established persistence in the victim's system. These are divided in distinct stages in the kill-chain model because they must bypass different cyber countermeasures deployed by the defender such as. patch management addressing a vulnerability, anti-virus trying to guard against the installation of malware or clear security policies and training to render users less susceptible to social engineering attacks. The three phases are however sequential, requiring no distinct actions on the part of the attacker. In our example, a phishing email delivers the exploit, the targeted vulnerability is exploited when the user opens a malicious document attached to that email, resulting in malware being installed to their victim's workstation.

In the *Command and control (C2)* phase, the attackers establish communications to and from the asset. This phase provides the APT with a level of access to the network of the victim organization, which can be used to expand their presence and acquire the necessary additional access to perform their ultimate objectives. During this phase, the attackers are usually careful to avoid detection to prevent defenders from rooting them out before they managed to perform their mission. The desire to remain undetected for as long as possible is very valuable to attackers, and Mandiant reports that the average dwell-time for APTs in 2017 was 101 days (*M-Trends 2018*, 2018). For example, in the typical enterprise network we used to motivate our explanation, the APT may be able to scrape credentials from the victim's workstation to gain access to a file server containing sensitive information.

During the *Actions on objectives* phase, the attackers perform whatever actions are required by their mission. They exfiltrate the confidential information, sabotage the control system, disrupt and degrade the service or whatever else is within the scope of their mission. To do so, the attackers leverage the access that they have developed in the previous phases, often leveraging legitimate access paths that became accessible illegitimately after moving around within the network. For example, after having stolen proper credentials, the attackers might connect through the victim organisation's virtual private network (VPN) gateway rather than by using new exploits, thereby obfuscating their activities using the defenders' own protective mechanisms.

An important observation regarding the kill-chain is that, when people talk about cyber-attacks, they often mean the highly visible *Actions on objectives* phase of the kill-chain. In a sense, this is analogous to how terrorist attacks are viewed. For most people, the terrorist attack is the part where a bomb explodes, not the preliminary steps required to successfully prosecute the attack. In that sense, the earlier phases of the cyber kill-chain are often discounted by commentators in offensive cyber operations.

It would be easy to adopt a very defeatist attitude about cyber defence; there are so many vulnerabilities (not the least being the humans using the systems) that one could be forgiven for thinking that cyber defence is hopeless. The sequential nature of the cyber kill-chain is in fact one of the rays of hope for cyber defenders. It may not be evident at first blush, but attackers must get succeed at every single phase of the cyber kill-chain to meet their ultimate goal; defenders on the other hand, need only break one step in the kill-chain to thwart the attack. As we will see in following sections, the early phases of the cyber kill-chain often represent a significant portion of the attackers' the operation and they greatly affect their operational tempo.

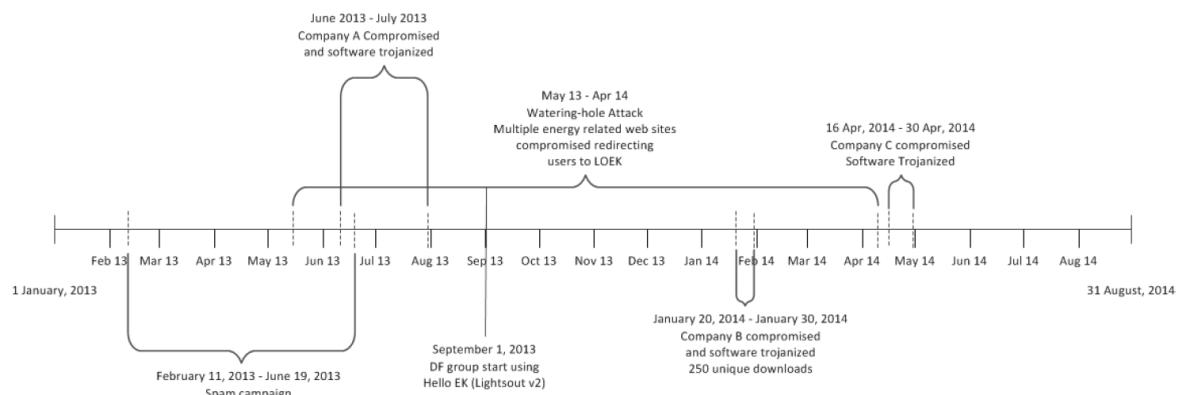
#### 4. Review of real-world attacks

Among the first wake-up calls for cyber warfare were the attacks on Estonia in 2007 and Georgia in 2008. In both of these cases, countries were the target of extensive cyber campaigns, including denial of service attacks and web defacement of government websites (Kostadinov, 2013; Hagan, 2012). In the case of Georgia, these attacks coincided with the deployment of traditional military forces, meaning that the cyber operation was integrated as part of a wider kinetic campaign (Donovan, 2009). In terms of tempo, both attacks lasted several days. It is worthy of note however, that multiple changes in tactics were required to adapt to the realities on the ground. For example, tactical changes had to be made during the Estonian attack to adapt to response by defenders, while changes in tactic due to the saturation of the international Internet link were required in Georgia (Shachtman, 2008). As these attacks used the organisation of cyber mobs among other tactics, specific actions could take several hours or even days to propagate in the face of a change in online tactics. This means that in terms of tempo, while the packets arrived at network speed, the overall flow of the attack was much slower, which we term to be at human speed.

Another well documented cyber-attack is the Stuxnet attack aimed at sabotaging the Iranian nuclear weapons program (Falliere et al., 2011; Langner, 2013). In this attack, malicious software was used to discreetly alter the operation of nuclear enrichment centrifuges in order to prevent the creation of weapons-grade fissile material. In terms of tempo, the Stuxnet operation was very slow. As the operation was allegedly initially approved by President Bush (Sanger, 2012) who left office in 2008, it was discovered only in 2010 and it lasted several years. Furthermore, even on a tactical level, we know that the PLC module recorded between 13 days and 3 months of normal operations to create a baseline before disrupting the operation of centrifuges and hiding its presence (Falliere et al., 2011). This is also a far cry from operating at network speed.

In order to create the destructive Stuxnet payload, it was necessary to perform extensive espionage and reconnaissance operations, which is consistent with the cyber kill-chain that we previously discussed. We can deduce this by looking at the various versions of the Stuxnet malware; older versions contain payloads exclusively devoted to espionage, while payloads associated with the actual disruption of the centrifuge (*the Actions on the objective*) coming much later (Falliere et al., 2011; Gostev and Soumenkov, 2011). This suggests the actions of human operators retooling the malicious software based on the information derived during *Reconnaissance*.

The Dragonfly/Havex campaign is a campaign, generally attributed to Russia, that targeted the energy sector through espionage (*Dragonfly: Cyberespionage Attacks Against Energy Suppliers*, 2014, "Dragonfly," 2017). Of note, the campaign to distribute this malware lasted over a year. The campaign included multiple changes in tactics, starting with phishing campaigns, but eventually moving to targeted web compromises and infection of the supply chain. Presumably, these changes were dictated by the need to find a suitable method to infect their designated target(s), requiring adaptation in *Weaponization*, causing the attackers to change tactics as they monitored the success of their efforts. This behaviour is illustrated at Figure 2.



**Figure 2:** Dragonfly timeline (reproduced from (*Dragonfly: Cyberespionage Attacks Against Energy Suppliers*, 2014))

The cyber-attacks that led to the Ukraine blackout of 2015 (Lee et al., 2018; Zetter, 2016) is an interesting example as it demonstrates attackers ability to alter the tempo of their operations. The attack is reported to have started in the spring of 2015 with a series of spear phishing campaigns targeting information Technology

(IT) staff and system administrators at multiple energy companies to deliver the BlackEnergy3 malware (kaloyan, 2018). In the intervening months between the spear phishing campaign in the spring of 2015 and the blackout in late December of the same year, the attackers mapped out the network and stole credentials, notably the VPN credentials to access the network segment housing the industrial control systems. The attackers then installed malicious software on various control systems components and on uninterruptible power supplies (UPS) that ensured control systems were supplied with power in a blackout situation.

Finally, on the day of the blackout, the *Actions on objectives* started around 3:30 AM by connecting, externally to the network via the VPN, to shut down the UPS and activate control systems, creating blackouts. At the same time, a coordinated telephone denial of service attack targeted the victims' phone systems to prevent legitimate customers from reaching the company. The attackers then rewrote the firmware on important communication devices (serial-to-Ethernet converters) to prevent communication with the affected control systems, requiring the energy operators to switch to manual control and erased operator workstations with the modified KillDisk (LSoft Technologies Inc., n.d.) component of their malware. This phase of the attack ended around 5:30 AM, about two hours from the beginning of the *Actions on objectives* phase.

The first part of the attack was characterized by a very slow tempo. In terms of timeline, around six months elapse between the first phishing campaign to the black out. In fact, Lee describes it as human operators on keyboard figuring out the network rather than an attack carried out at network transmission speed (VICELAND, 2016). Once the *Actions on the objectives* phase of the operation started with the application of kinetic effects (the actions taken to cause the black out), the pace of operation is considerably faster. According to testimony from victims, power grid human operators could observe attackers moving the cursor on their system through remote control software (NATO, 2016); this suggest that while this portion of the operation occurred at human speed, it was taking place over the course of minutes rather than months as observed during the *Reconnaissance* phase. Furthermore, some actions by the adversary, such as the disabling of the phone exchange and the wiping of the workstations, appear to have been designed to slow the ability of the defenders to respond. This suggests that this adversary was specifically concerned with tempo in their operational planning.

Based on these observations, it seems that cyber operations are much slower than generally believed, taking place over years, months and weeks rather than seconds and milliseconds. A first observation is that most of these attacks were not fully automated and included significant human interaction. Even during the Estonia and Georgia attacks, which mostly relied on automated botnets, analysts found that tactics were changed, and targets adjusted based on coordination via social media, which took time. The Stuxnet worm was no different, although it was mostly automated to operate in an isolated environment, analysis shows evidence retooling by human operators based on the results of information collected during the *Command and Control* phase.

## **5. Analysis**

It would be useful for the community to gather empirical and quantitative evidence of the timelines involved in an attack, perhaps analysing the amount of time spent by attackers in each phase of the cyber kill-chain more formally. While such an analysis falls outside the scope of this paper, we attempted to find evidence of such quantified studies without success, and we are left with the historical cases discussed in section 4.

Based on these historical cases, it seems that cyber-attacks require considerable lead time. In particular, the development of toolsets during *Weaponization* and the collection of intelligence during *Reconnaissance* to support operations appear to be lengthy processes. In their RSA presentation, Crowdstrike show how attack groups retool after public disclosure of their toolset (Dennesen, 2016). We can see that significant time is invested to change the tools to make detection more difficult. We could expect that the *Weaponization* of attack vectors, either through vulnerability development or open source intelligence work to create credible spear phishing campaigns, also require significant lead time.

Under this hypothesis, the development of new tools in *Weaponization* and the intelligence preparation in *Reconnaissance* represent bottlenecks on operational tempo. As evidenced by the kinetic portion of the Ukraine attack, once attackers have completed these preliminary steps, they are able to achieve a high tempo of operations during the *Actions on the objectives*. If the attackers' presence is discovered, be that during the short *Delivery – Exploitation – Installation* window or more likely during the drawn-out *Command and Control* phase, retooling appears to take significant time, drastically slowing their tempo. Even with automated attack as was

the case in Estonia and Georgia, the tempo needs to be slowed in order to propagate new tradecraft if the situation on the ground changes, be that in reaction to actions by the defenders or in response to unpredicted network failures. In that sense, the slow tempo of the Stuxnet attack, dictated by the need to record baseline data for a month before starting the disruption, might be a concession to this need for deliberate action: Stuxnet operators deemed it wise to lose performance, in terms of raw quantity of enrichment disrupted, in order to avoid the massive tempo loss that would result from being rooted out.

However, the tempo of operations is typically dictated by the political circumstance rather than the physical and technological limitations of cyber-attacks (Lemay et al., 2010). Imagine a situation where the president has just announced air strikes and the air defence systems need to be disabled immediately. In such a circumstance, the tempo of operation is incompatible with the lead time required to develop access. This would suggest that a core strategy of cyber forces should be to develop access, which involves operations with slow tempo, in times of relative peace for high tempo actions to be available in times of active conflict.

In that context, the key to the success of high tempo cyber operations is not the ability to rapidly develop new infection vectors which can effect a denial of service at network speed, but rather one's ability to access previously developed access. Using the operational functions from Canadian military doctrine (*CFJP 01 - Canadian Military Doctrine*, 2009), one should invest one's strength in SUSTAIN rather than ACT. In other words, a cyber force that is heavily invested in an ACT attack capability, but lacks the ability to SUSTAIN access would be the equivalent of shock troop with no logistical support in a conventional military context: they could do incredible damage to the opponent, but they cannot be counted on to be in the right place at the right time or to have lasting impact.

A corollary of this observation is that defenders should focus their energies on disrupting the attacker's ability to SUSTAIN their presence rather than attempting to blunt the very short-lived attacks. This means that it is more important to develop the capability to detect attackers on a network and to disrupt their access, rather than preventing them from gaining access in the first place. In fact, the report on the Ukraine attack (Lee et al., 2018) implies that at least one victim was infected but did not receive the blackout payload. The report hypothesize that this was due to alterations in the network prior to the attack, which inadvertently inoculated the victim to the attack; the last-minute change to the network configuration means that the attackers did not have the time to retool.

In the same vein, forcing the attackers to retool seems to be one of the most efficient methods to slow down their tempo, providing more opportunity for defenders to respond. This seems particularly critical in situations of active conflict which demand high tempo, and thus do not afford much time to deployed retooled cyber weapons. This strategy could be likened to attacking the supply lines in a conventional operation. The forward deployed units are still able to do damage, but they cannot operate for long as they are unable to operate with a lack of supply.

Ultimately, this could lead to interesting strategic dilemmas. For example, let us consider that an armed force has discovered that an adversary has established pre-positioned access. Is it more advantageous for the armed forces to fight him for access now in times of piece, or try to deny him his prepositioned access and force him to retool just before a heightened conflict period? Should an offensive cyber commander invest in developing shallow access across multiple targets, or should he develop multiple levels of access across fewer targets to be less susceptible to loss of tempo in critical periods? The answer to these questions should be grounded in the understanding of the tempo of cyber operations.

## **6. Conclusion**

By looking at numerous publicly documented nation-state sponsored cyber-attacks, we have shown that cyber operations do not occur at the electron speed that is assumed by a number of defenders. In fact, we have observed that changing situation on the ground, as well as response by defenders require attackers to change tactics and retool to address changing environments in trying to accomplish their mission. Such changes are often the result of human decisions and, as such, they throttle cyber operation tempo to human speed, giving time for defenders to react.

In fact, based on our observations, we submit that the development of tools and tradecraft, along with the collection of intelligence, appear to be significant contributors to long lead time required to acquire access in cyber operations. Once this access is acquired, high tempo cyber operations can be developed, but not before. As such, denying attackers the ability to sustain access could be the most valuable strategy for the defence.

Naturally, these conclusions are predicated on the absence of fully automated artificial intelligence (AI) driven cyber-attacks or cyber defence that could rapidly change tactics based on new situations on the ground. It would be interesting to investigate changes in this field and analyse them for tempo escalation; will we find ourselves in situations where the tool that operates with the fastest tempo wins. Alternatively, it would be interesting to investigate if strategies developed to counter manoeuvre warfare in the conventional world could also apply to defeating AI. This paper suggests that a more in-depth evaluation of how cyber operations are sustained would be valuable. Of particular interest is finding out how long high tempo operations can be sustained.

Finally, as we suggested in Section 2, the revolution in military affairs is heavily dependent on the rapid processing of information enabled by the digital revolution. In that sense, an investigation on the effects of cyber-attacks on kinetic tempo and general combat effectiveness could also be beneficial.

## References

- Allcorn, W., 2012. *The Maginot Line 1928–45*. Bloomsbury Publishing.
- Boyd, J.R., 1996. The essence of winning and losing (Unpublished lecture notes No. 12.23).
- CFJP 01 - Canadian Military Doctrine (No. B-GJ-005-000/FP-001), 2009. . National Defence - Canada.
- Dennesen, K., 2016. Hide and Seek: How Threat Actors Respond in the Face of Public Exposure.
- Donovan, J., 2009. Russian Operational Art in the Russo-Georgian War of 2008. ARMY WAR COLL CARLISLE BARRACKS PA.
- Dragonfly: Cyberespionage Attacks Against Energy Suppliers (No. Verssion 1.21), 2014. . Symantec.
- Dragonfly: Western energy sector targeted by sophisticated attack group [WWW Document], 2017. . Threat Intell. URL <https://www.symantec.com/blogs/threat-intelligence/dragonfly-energy-sector-cyber-attacks> (accessed 2.9.19).
- Falliere, N., Murchu, L.O., Chien, E., 2011. W32.Stuxnet Dossier (No. 1.4). Symantec.
- Gostev, A., Soumenkov, I., 2011. Stuxnet/Duqu: The Evolution of Drivers | Securelist [WWW Document]. URL <https://securelist.com/stuxnetduqu-the-evolution-of-drivers/36462/> (accessed 2.9.19).
- Hagan, A., 2012. The Russo-Georgian War 2008: The Role of the cyber attacks in the conflict. AFCEA.
- LSoft Technologies Inc., n.d. How to erase hard drive by Active@ KillDisk? Disk Eraser, Disk Wiper, Disk Format & Disk Sanitizer. [WWW Document]. URL <https://killdisk.com/eraser.html> (accessed 2.9.19).
- Hutchins, E.M., Clopper, M.J., Amin, R.M., 2011. Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains, in: *Leading Issues in Information Warfare and Security Research*. pp. 80–106.
- kaloyan, 2018. BlackEnergy 3 / Black Energy 3 - CyberX [WWW Document]. URL <https://cyberx-labs.com/glossary/blackenergy-3-black-energy-3/> (accessed 2.9.19).
- Kostadinov, D., 2013. Estonia: To Black Out an Entire Country – part one [WWW Document]. InfoSec Resour. URL <https://resources.infosecinstitute.com/estonia-to-black-out-an-entire-country-part-one/> (accessed 2.9.19).
- Langner, R., 2013. To Kill a Centrifuge A Technical Analysis of What Stuxnet's Creators Tried to Achieve. The Langner Group, Arlington, VA.
- Lee, R.M., Assante, M.J., Conway, T., 2018. Analysis of the Cyber Attack on the Ukrainian Power Grid - Defense Use Case. E-ISAC, SANS ICS.
- Lemay, A., Fernandez, J.M., Knight, S., 2010. PINPRICK ATTACKS, A LESSER INCLUDED CASE? Presented at the Conference on Cyber Conflict Proceedings, Tallin, Estonia, pp. 183–194.
- Mazarr, M.J., 1994. *The Revolution in Military Affairs: A Framework for Defense Planning*. ARMY WAR COLL STRATEGIC STUDIES INST CARLISLE BARRACKS PA.
- M-Trends 2018, 2018. . Mandiant, Milpitas, CA.
- NATO, 2016. What happens when a power plant comes under cyber attack?
- Pech, R.J., Durden, G., 2003. Manoeuvre warfare: a new military paradigm for business decision making. *Manag. Decis.* 41, 168–179. <https://doi.org/10.1108/00251740310457614>
- Rehman, A., Saba, T., 2014. Evaluation of artificial intelligent techniques to secure information in enterprises. *Artif. Intell. Rev.* 42, 1029–1044. <https://doi.org/10.1007/s10462-012-9372-9>
- Sanger, D.E., 2012. Obama Ordered Wave of Cyberattacks Against Iran. *N. Y. Times*.
- Shachtman, N., 2008. Estonia, Google Help “Cyberlocked” Georgia (Updated). *Wired*.
- Tyugu, E., 2011. Artificial intelligence in cyber defense, in: *2011 3rd International Conference on Cyber Conflict*. Presented at the 2011 3rd International Conference on Cyber Conflict, pp. 1–11.
- VICELAND, 2016. Did Russia Hack Ukraine’s Electrical Grid?: CYBERWAR (Clip).
- Zetter, K., 2016. Inside the Cunning, Unprecedented Hack of Ukraine’s Power Grid. *Wired*.

# In Search of a Social Contract for Cybersecurity

Andrew Liaropoulos

University of Piraeus, School of Economics, Business and International Studies,

Department of International and European Studies, Piraeus, Greece

Laboratory of Intelligence and Cyber-Security

[aliarop@unipi.gr](mailto:aliarop@unipi.gr)

[andrewliaropoulos@gmail.com](mailto:andrewliaropoulos@gmail.com)

**Abstract:** The cybersecurity discourse has evolved over the last two decades from a discussion on national cyber policies and internet governance to one that refers to a complex security regime that involves both state and non-state actors. The rise of Big Data, the development of internet censorship techniques, Edward Snowden's revelations about the global surveillance carried out by the United States National Security Agency and the Cambridge Analytica scandal; demonstrate that the universal values of freedom, anonymity and privacy are under attack. Liberal democracies in the era of cyber-surveillance face an enormous challenge. On the one hand, they wish to maintain a high level of security, thereby collecting a vast amount of data on their citizens. On the other hand, they wish to preserve democracy, freedom of law and thereby respect constitutional arrangements, civil and political rights and independent media. The question that is raised is whether states can strike a balance between security concerns and their citizens' rights. The answer is not easy and forces us to readdress the existing security arrangements between state and citizens. States, societies and security institutions (both public and private, national and international) have to rethink the way they approach the constantly evolving security environment, its actors and the way they can reach reliable and effective agreements. As a result, a new social contract for cybersecurity is needed. This social contract has to challenge traditional relations between state and citizens and question the lines between the private and the public sphere. The purpose of this paper is to highlight the societal shift that is taking place in the global digital society and explore the utility of a new social contract.

**Keywords:** cybersecurity, big data, social contract, state of data

---

## 1. Introduction

In an era where more than half of the world is online and billions of people use social media, it is fair to argue that we are experiencing a societal reform. This societal reform is largely driven by information and communication technologies (ICTs) and its defining characteristic is data. Data is an asset, an opportunity, but also a security challenge. The Internet has transformed the way companies operate and the way societies communicate and are governed. Data is a lot more than just information floating around. Big data is considered the new oil. By using sophisticated analytical techniques and algorithms, massive and unstructured amounts of data can be mined, in order to reveal patterns and correlations, which would not be available by looking at smaller samples (Zwitter 2015, 378). Every minute, we produce thousands of internet searches and social media posts that reveal a vast amount of information about what we think and how we feel. In the coming years, all things around us will be connected and the Internet of Things (IoT) will produce an enormous quantity of data (Cukier and Mayer-Schonberger 2013a).

Nevertheless, we should not understand the ongoing digital revolution, solely in terms of size. The most important effect of this technological upheaval is of socio-political nature (Chandler 2015) and involves the process of 'datafication'. Big data is being put to extraordinary new uses. Location has been datafied with global positioning systems, words become data when software digitizes texts and smart algorithms mine their content. Even emotions are being datafied by 'likes' via Facebook and face recognition cameras (Cukier and Mayer-Schonberger 2013, 28-29). In the near future, with the extensive use of wristbands and smart phones, individuals will be able to hold health data that could improve health systems. Likewise, mobility data can be collected from credit card transactions, public transportation smart cards and mobile applications that can be used to identify patterns about how humans travel. Such data can optimize transportation and city planning.

Despite the importance of data in every facet of society, there is still no extensive public information campaign on data and related issues like encryption techniques, algorithmic regulations, artificial intelligence and the responsibilities of the ICTs industry. It is in this context that we need to consider whether the ongoing societal reform needs a new social contract (Keith 2018). Data breaches, the use of data for a broader societal good, or the misuse of metadata raise numerous legal and ethical questions, which are directly linked to democracy, privacy and security. Is the digital revolution forcing us to reconsider the moral and political rules of behavior that regulate the agreement between the ruled and their rulers? Should governments be the sole providers of

societal security, or is there also a role for the private sector in protecting citizen's data? Is it feasible to regulate social agreements between governments, citizens and companies that challenge the lines between the private and the public sphere? How can we build trust, transparency and legitimacy in a social contract between government, citizens and corporations?

In order to tackle the above questions, we need to review first the social contract theory. The latter is a contract between people in the pre-political condition that describes the terms under which they are prepared to enter society and submit to political authority. The social contract encapsulates the transition from a state of nature to a social and political existence. In a latter phase, having established a theoretical background on social contract theory, we will demonstrate the challenges that the digital society faces in the 'state of data' and question whether a new social contract can be constructed in order to fulfill the needs of a data-oriented society.

## **2. Social contract theory**

A great variety of social contract theories have been proposed, but our analysis will focus on the works of Hobbes, Locke and Rousseau. The concept of social contract refers to an actual or hypothetical agreement, between the ruled and their rulers and defines the rights and duties of each party. According to the social contract theory, in the beginning people lived in the state of nature, there was no government and no law to regulate them. In this state of nature, individuals' actions were restricted only by their personal power and conscience. In brief, social contract theory explains why rational individuals, would voluntarily consent to give up their natural freedom and form a society in order to enjoy the benefits of political order (Barker 1980). In order to leave this state of nature, people entered into two agreements. In the first one, people sought protection of their lives and property. Thus, a society was formed, where people agreed to respect each other and live peacefully. In the second agreement, people united and gave their consent to obey an authority and surrender the whole or part of their freedom and rights to this authority. The latter guaranteed everyone protection of life, property and to a certain extent liberty. Therefore, by escaping the state of nature, individuals collectively form a society, were they agree to live together under common laws and create enforcement mechanism for the social contract (Boucher & Kelly 1994).

Although the origins of the social contract theory can be traced back to Plato and Socrates, it was Hobbes, Locke and Rousseau that popularized the concept in the eighteenth century. In 'Leviathan' that was published in 1651 during the Civil War in Britain, Thomas Hobbes illustrates his understanding of the social contract. For Hobbes, the state of nature was one of fear and selfishness, where people experienced a poor, brutish and short life. People have a natural desire for security and order. In order to safeguard their survival and avoid misery and pain, people enter into a contract and voluntarily give up their rights to some authority that has the duty to protect them and their property. According to Hobbes, this authority, the ruler or monarch has to be obeyed under all circumstances, regardless of how bad or strict he might be. As a counterbalance, Hobbes placed moral obligations on the monarch, who is bound by natural law. In his view, the state of nature is a state of war and therefore people have to give their liberty into the hands of the absolute sovereign, on the sole condition that their lives were safeguarded by sovereign power. The law is dependent upon the consent of the sovereign (Wolff 1994).

John Locke's theory of social contract is different from that of Hobbes. In his book 'Two Treatises of Government' published in 1690, he too agrees that people lived in the state of nature, but his view about the state of nature is not as miserable as that of Hobbes. On the contrary, Locke describes it as a reasonably good and enjoyable state, where people were free to pursue their own interests, free from interference. According to Locke, the only disadvantage of the state of nature; is that there is no established law or natural power to implement natural laws, therefore property was not secure. It is the need to protect the property that forces people to abandon the state of nature and enter into the social contract. Thus, people did not surrender all their rights to the authority, but only the right to preserve order and enforce the law of nature. People hold on to themselves the right to life and their liberty. According to Locke, the purpose of the government is to protect the natural rights of men. As long as the government fulfills this purpose, the laws are valid and binding, but when the government ceases to fulfill it, then the laws would have no validity and the government can be thrown out of power. In Locke's view, unlimited sovereignty is contrary to natural law (Baumgold 2005).

Jean Jacques Rousseau gave a different interpretation to the theory of social contract in his book titled 'The Social Contract' that was published in 1762. In contrast to Hobbes and Locke, he argues that social contract is

not a historical fact, but rather a hypothetical construction. According to Rousseau, people, cannot pursue their true interest by being egoists, but must subordinate themselves to the law that is created by the society. People surrender their rights not to an individual, but to the community as a whole, which Rousseau coined as 'general will'. In this sense, the law is not a limitation of individual freedom, but an expression of it. Each individual is not subject to any other individual, but only to the general will, therefore, by complying with this the individual is obeying himself/herself. The enforcement of law is not a restriction on liberty, but a rise from the state of nature into that of a civil society, a civilizing force. Through social contract, a new form of social organization was formed - the state - that has the duty to protect the rights and liberties of its citizens. Through social contract people submit their natural rights and in return get civil liberties, such as freedom of speech and equality (Bertram 2003).

Based on the above brief analysis, we notice that the social contract theory of Hobbes supports absolute sovereign, where whatever the ruler does is just. Society is the creation of the state and the reflection of the will of its ruler. Locke and Rousseau on the other hand, argue that the state exists in order to protect the natural rights of its citizens. When governments fail to fulfill these tasks, then citizens have the right to withdraw their support. In short, Hobbes supports absolute sovereign, downgrading the value of individuals, whereas Locke favors a constitutional government and Rousseau promotes people's sovereignty.

### **3. From the state of nature to the 'state of data'**

The information revolution is the driving force behind cyberspace. The latter reflects all the challenges that globalization and digitization bring about. It substantiates the dematerialization, decentralization and deterritorialization of politics (Choucri 2012, 4). It enables the rapid and inexpensive distribution of data, it empowers non-state actors in a hyper-connected environment and it facilitates a global knowledge based economy. But there is also a dark side to the information revolution. Hackers can attack vulnerable points in the critical information infrastructure, disrupt global commerce and compromise sensitive information. Privacy and anonymity are under attack, surveillance and espionage are on the rise. Smart devices and social networking inform us of each other's thoughts, but also provide a massive amount of personal data to intelligence services, criminals and private organizations. Over the last years, security threats in cyberspace have gone beyond the typical cyberattacks. Social platforms that were used to spread ideas have now been hijacked by trolls and bots that are spreading disinformation. Adding to that, the power to inform, shape opinion and even manipulate is concentrated in the hands of a few technological giants.

Big Data enables us to extract information from, and interpret immense amounts of unstructured data (Zwitter 2015, 378). It also marks a transition from causation to correlation. It produces a different kind of knowledge that emphasizes more on the interpretations of signs and identifying association among phenomena, instead of considering the reasons behind how the world works, thereby understanding chains of causation (Cukier and Mayer-Schonberger 2013, 32; Chandler 2015, 833-851). Big Data is a resource essential to the global economy and we are witnessing an intense struggle to control it. Big Data is a raw material that has the potential to change the power distribution by empowering non-state actors (Zwitter 2015, 380). Data is the fuel of the modern economy, it can be bought and sold and is a strategic resource for nations. The digital economy is witnessing the increase of platform monopolies, like Google, Facebook Apple and Amazon. Google controls internet search performing nine out of ten searches, Facebook dominates social media having over two billion users, Apple the largest mobile store with almost 80 percent of the market Amazon is a leader in e-commerce getting every other dollar spent online in the United States (Srnicek 2017; Mayer-Schonberger and Ramge 2018, 48). These top technological giants own their position in the global market due to their digital assets and not their physical ones (Nikelani 2018, 19).

States, but also companies are gathering vast amounts of data on everyone and everything, raising the question whether Big Data will lead to Big Brother (Cukier and Mayer-Schonberger 2013, 37). In terms of surveillance, Big Data reverses the standard policing and intelligence practices, where targets are identified and then data collections follow. Instead, data is now being collected, before deciding on the full range of their actual and potential use (Lyon 2014). The rationale is that Big Data algorithms will allow us to predict behaviours and events. Personalized information is a valuable service for a knowledge based economy, but it can also be exploited for the purpose of social engineering and manipulation (Zwitter 2015, 386). Indicative of the above is the case of Cambridge Analytica, a consulting firm that harvested personal data from tens of millions of Facebook users and sold it to political campaigns. This scandal revealed not only how actors could use data to target democracy, but

also the power that top technologies hold (Nikelani 2018, 19). The same technology that decentralizes knowledge and institutions of governance is also used for cyberattacks, mass surveillance and the manipulation of public opinion. Likewise, encryption is neither good nor bad. Banning encrypted messages could be perceived as a violation of both the right to privacy and online anonymity, but at the same time it could be justified for reasons of national security. The same encryption that protects governments and corporations is also used by citizens and terrorists (Liaropoulos 2016, 35).

We are living in a ‘state of data’ and we seem to stand at a crossroads. Technology is value free and can still serve democracy, but regulations and structures are needed to ensure that data will not be used solely for commercial use or to deceive voters, but rather for a broader societal goal, as a digital public good. Enacting data privacy laws, regulating the digital market in order to ensure completion and mandating transparency on political advertisers and sponsorship in the digital media, all point to the right direction (Ghosh and Scott 2018). The underlying logic of all the above measures and initiatives is the need to manifest a social contract for cyberspace.

In the global and hyper-connective society, states can no longer be the sole security providers. They have to cooperate with other actors, since security arrangements appear at the local, national and global level and include intergovernmental cooperation, supranational entities, public-private partnerships and cooperation with citizens, non-governmental organizations and advocacy groups. The security governance of cyberspace requires the coordination of all stakeholders. Securing cyberspace against hate speech is only one example that requires the coordination of the state security services, the social media and the internet providers. In cyberspace the security providers include apart from the state also international organizations (e.g. NATO, EU). But the private sector, that largely owns and manages the digital space, cannot be ruled out from this complex security equation (Clinton and Perera 2016).

#### **4. Constructing a social contract for cybersecurity**

The above analysis vividly illustrates the need to readdress the social contract theory in order to redefine the relations and interactions between the actors that are involved in the contemporary complex security environment. Therefore, we need to take a fresh look and outline the conditions that will force people to leave the ‘state of data’ and form a new social contract. Back in the seventeenth century we experienced a shift of authority and legitimacy from divine rule to popular sovereignty. Fear prompted people to leave the natural condition of anarchy and uncertainty and place their trust in a new political entity, the sovereign state. Likewise, a similar shift in terms of knowledge, wealth and authority is taking place in the digital era. We live in a digital society that is dominated by data. Data drives the global economy and produces new knowledge, but data also creates fear and its use needs to be regulated.

As analyzed above, the social contract initially regulated the relationship between the individual and the state and aimed to achieve social order. The purpose of the state is to protect the safety of its citizens. In sovereign democracies the set of fundamental rules governing the politics of a nation can be found in the constitution. The latter illustrates the values, roles and responsibilities that apply to citizens and government alike. A constitution becomes effective through people’s consent to abide by it. The constitution is considered to be a contract, the most common written representation of a social contract. In return for receiving security, citizens fulfill their own described responsibilities to obey the law (Bierens 2017, 4).

In the twentieth century though, with the emergence of private companies, a third actor was added in the social contract. Corporations are legal entities within a state that aim to maximize profit. The constitution also regulates their activities. States must ensure that private companies do not harm the social contract between citizens and government. Thus, there are laws and regulations that apply specifically for corporations while taking into account other drivers such as competitive market forces between corporations and citizens. These competitive market forces are assumed to have a positive effect on the behavior of corporations. In case these drivers are limited, such as within a monopoly, the government will increase its control and strengthen its laws and regulations (Bierens 2017, 6-7).

The laws and agreements that are implemented by the government to the private companies, ultimately aim to ensure the social contract between the former and the citizens. Likewise, the terms and conditions under which companies function and provide their services, is a social contract between the corporations and their customers

- citizens. In a democracy, citizens have the freedom to choose via the elections a new government and similarly in relation to corporations - unless there is a monopoly - this choice is made through market forces (Bierens 2017, 7).

Recent evidence illustrates that the cyber social contract - between governments, citizens and corporations - as analyzed above, is under construction. Elements of this social contract can be found in data privacy laws and market regulations and incentives. In order to prevent the misuse of personal information the EU introduced in 2018 the General Data Protection Regulation (GDPR). In 2017, the European Commission fined Google after finding out that the company had manipulated its search results in order to favor its shopping service. US Senators have been lobbying for the Honest Ads Act, an agreement that would require internet companies to be transparent about the origins of their political digital advertisements (Picchi 2018). The US Federal Trade Commission has launched a public inquiry regarding the anticompetitive behavior of digital platform companies, like Facebook, Google and Twitter. In 2019, Germany's Federal Cartel Office ordered Facebook to restrict the data it collects and combines on its platforms and third-party sites, without explicit consent from their users. Facebook has argued that its activities could improve its advertising targeting and assist in identifying fake accounts, but the counterargument is that it could threaten anonymity and privacy (Dreyfuss 2019). The next step into breaking the monopolies of technological giants is progressive data-sharing mandate, where digital giants will be required to share part of the data they collect with other companies. Such a development would decentralize digital markets and boost innovation (Mayer-Schonberger and Ramge 2018).

## **5. Conclusions**

From the utopian view of an independent cyberspace with little or ideally no top-down regulation, that Jerry Barlow expressed in his manifesto titled *A Declaration of the Independence of Cyberspace*, (1996), to the dystopian view that if technology dominates our lives and solves society's problems then there will be no place for governments and this will lead to the death of politics (Morozov 2014), we need to question how society will decide about its own fate. Drawing the parallel with the historical moment of state formation, where people left the state of nature and entered the social contract, it can be argued, that we are in a similar phase, where people are in a 'state of data' and they need once again to enter a new social contract. Forming a social contract for cybersecurity that involves governments, citizens and corporations, building the necessary trust between them, treating data as a public good and regulating the market in order to serve the needs of the society, is not an easy task, but ignoring the challenge is no longer an option. Hobbes, Locke and Rousseau have sent us a friend request.

## **Acknowledgements**

This work has been partly supported by the University of Piraeus Research Center.

## **References**

- Barker, E. (1980) Social Contract: Essays by Locke, Hume and Rousseau, Westport, CT: Greenwood Press.
- Barlow, J.P. (1996) "A Declaration of the Independence of Cyberspace", available at <https://www.eff.org/cyberspace-independence>, last accessed, 15 January 2019.
- Baumgold, D. (2005) "Hobbe's and Locke's Contract Theories: Political not Metaphysical", Critical Review of International Social and Political Philosophy, vol.8, no.3, pp. 289-308.
- Bertram, C. (2003) Rousseau and the Social Contract, London: Routledge.
- Bierens, R., et.al. (2017) "A Social Cyber Contract Theory Model for Understanding National Cyber Strategies", in Proceedings of International Conference on Electronic Government, pp. 166-176.
- Boucher, D. and Kelly, P. (1994) The Social Contract from Hobbes to Rawls, London: Routledge.
- Chandler, D. (2015) "A World without Causation: Big Data and the Coming Age of Posthumanism", Millennium: Journal of International Studies, vol.43, no.3, pp. 833-851.
- Choucri, N. (2012) Cyberpolitics in International Relations, Cambridge MA: MIT Press.
- Clinton, L. and Perera, D. eds. (2016) The Cybersecurity Social Contract: Implementing a Market-Based Model for Cybersecurity, Internet Security Alliance.
- Cukier, K. and Mayer-Schonberger, V. (2013) "The rise of Big Data", Foreign Affairs, vol.92, no.3, pp. 27-40.
- Cukier, K. and Mayer-Schonberger, V. (2013a) Big Data: A Revolution that Will Transform How we Live, Work and Think, New York: Houghton Mifflin Harcourt.
- Dreyfus, E. (2019) "German regulators just outlawed Facebook's whole ad business", Wired, 7 February, available at: <https://www.wired.com/story/germany-facebook-antitrust-ruling/>, last accessed, 10 February 2019.
- Ghosh, D. and Scott, B. (2018) #DigitalDeceit. The Technologies behind Precision Propaganda on the Internet, Harvard Kennedy School, Shorenstein Center on Media, Politics and Public Policy.

***Andrew Liaropoulos***

- Keith, K. (2018) "We need to build a new social contract for the digital age", The Guardian, 4 April, available at: <https://www.theguardian.com/commentisfree/2018/apr/04/we-need-to-build-a-new-social-contract-for-the-digital-age>, last accessed, 1 February 2019.
- Liaropoulos, A. (2016) "Reconceptualizing Cyber Security: Safeguarding Human Rights in the Era of Cyber Surveillance", International Journal of Cyberwarfare & Terrorism, vol.6, no.2, pp.32-40.
- Lyon, D. (2014) "Surveillance, Snowden and Big Data: Capacities, consequences, critique", Big Data & Society, vol.1, no.2, pp. 1-13.
- Mayer-Schonberger V. and Ramge, T. (2018) "A Big Choice for Big Tech: Share Data or Suffer the Consequences", Foreign Affairs, vol.97, no.5, pp. 48-54.
- Morozov, E. (2014) "The Rise of Data and the Death of Politics", The Observer, 30 July, available at: <https://www.theguardian.com/technology/2014/jul/20/rise-of-data-death-of-politics-evgeny-morozov-algorithmic-regulation>, last accessed, 1 February 2019.
- Nikelani, N. (2018) "Data to the People: India's Inclusive Internet", Foreign Affairs, vol.97, no.5, pp. 19-27.
- Picchi, A. (2018) "Facebook: What is the Honest Ads Act?", CBS NEWS, 11 April, available at: <https://www.cbsnews.com/news/facebook-hearings-what-is-the-honest-ads-act/>, last accessed, 1 February 2019.
- Srnicek, N. (2017) Platform Capitalism, Cambridge: Polity Press.
- Wolff, J. (1994) "Hobbes and the motivations of social contract theory", International Journal of Philosophical Studies, vol.2, no.1, pp. 271-286.
- Zwitter, A. (2015) "Big Data and International Relations", Ethics & International Affairs, vol.29, no.4, pp. 377-389.

# The Importance of Strategic Leadership in Cyber Security: Case of Finland

Jarno Limnell<sup>1</sup> and Martti Lehto<sup>2</sup>

<sup>1</sup>Aalto University, Finland

<sup>2</sup>University of Jyväskylä, Finland

[jarno.limnell@aalto.fi](mailto:jarno.limnell@aalto.fi)

[martti.j.lehto@jyu.fi](mailto:martti.j.lehto@jyu.fi)

**Abstract:** Cyber security has become one of the biggest priorities for businesses and governments. Streamlining and strengthening strategic leadership are key aspects in making sure the cyber security vision is achieved. The strategic leadership of cyber security implies identifying and setting goals based on the protection of the digital operating environment. Furthermore, it implies coordinating actions and preparedness as well as managing extensive disruptions. The aim of this paper is to define what is strategic leadership of cyber security and how it is implemented as part of the comprehensive security model in Finland. The paper also asks (and answers) how the strategic leadership of cyber security must be organised. This paper provides proposals for managing strategic cyber security in society and public administration, for managing large disruptions in the cyber operating environment. Key data consists of different security-related strategies and instructions, existing research information, and interviews with public sector actors and experts of the field. In terms of effective strategic leadership of cyber security, it is vital to identify structures that can respond to the operative requirements set by the environment. As a basis for national development and preparedness, it is necessary to have a clear strategy level leadership model and situation awareness that supports management. They are also necessary for the management of serious, extensive disruptions in both normal and exceptional conditions of the cyber operating environment. The challenges of cyber security management are particularly prominent at the level of strategic leadership. In order to ensure cyber security and achieve the set strategic goals, society must be able to engage different parties and reconcile resources and courses of action as efficiently as possible. Cyber capability must be developed in the entire society, which calls for strategic coordination, management and executive capability. The goals presented in Finland's Cyber Security Strategy have guided the creation of strategic leadership models for cyber security. Alternative models for the strategic leadership of cyber security in Finland is presented in the paper.

**Keywords:** cyber security, strategic leadership, situational picture, leadership, national security

---

## 1. Introduction

### 1.1 Background

Cyber security is an elemental part of society's comprehensive security, and the cyber security operating model is in keeping with the principles and practices specified in Finland's Security Strategy for Society (2017).

Technical and economic development has led to networking and increasing interdependencies between production, services and entire society. An efficient and optimised network economy is based on rapidly developing information and communication technology, which is vulnerable to many new types of threats and risks. Cyber-attacks, malware, denial of service attacks and different forms of influencing through information are becoming ever more prolific. The reliable operation of telecommunications, information systems and communications are an essential precondition for modern society's undisrupted functioning, security and citizens' livelihoods. This is also about maintaining citizens' trust in a well-functioning society. The development of business continuity management accounts for a large proportion of the security of supply work carried out in the information society sector. Due to this development, improved preparedness for maintaining the functioning of society's vital information technology systems and structures in the face of cyber threats and incidents is also needed in normal conditions. In particular, it should be noted that Finnish society's and companies' dependence on the cyber environment will grow further in the years to come. (Lehto et al., 2018)

A strategic level leadership model and situational awareness is needed to create the foundation for national preparedness as well as for the leadership of cyber domain – for both in normal time conditions and during emergency conditions. The weakness of the current model is notified in several researches (Lehto et al., 2017). In the cyber environment, strategic sensitivity requires an ability for forming a situational picture and creating situational awareness rapidly for the basis of decisions and actions. Preconditions for building an environment

for cyber security situational picture are shared situational awareness, coordinated and networked management and sufficient expertise in different areas of cyber security.

The national strategic leadership of cyber security consists of two entities: managing cyber security preparedness and managing serious and extensive incidents in normal and emergency conditions.

## **1.2 Objectives**

This research paper prepared proposals for measures related to the management of society's and public administration's cyber security and managing extensive disruptions in the cyber environment. Key research questions examined were the following:

- What is strategic leadership of cyber security?
- How is strategic cyber security leadership implemented in the responsibility model for comprehensive security?
- How should the strategic leadership of cyber security be organised?

## **1.3 Data and methodology**

Highly versatile and extensive material was collected for this study. Key data consisted of different security-related strategies and instructions, existing research information, and interviews with public sector actors and experts of the field<sup>1</sup>. The research project interviewed 40 employees in managerial roles and officers responsible for information/cyber security in private and public organisations. The interviews were conducted in 25 key organizations. Based on the interviews, document analysis and international comparison data, an analysed data set was created, on which the observations, proposals and models presented in this study are based.

Finland's Cyber Security Strategy and its background dossier (Security committee, 2013) and its Implementation Programme 2017-2020 (Security committee, 2017a), the Security Strategy for Society (Security committee, 2017b), a report titled Central Government Communications in Incidents and Emergencies (Prime Minister's Office, 2013), the Guidelines for developing Finnish legislation on conducting intelligence – A report of the Working Group (Ministry of Defence, 2015) and the National Audit Office's performance audit report titled Cyber Protection Arrangements (National Audit Office, 2017) were used in the research project. Key documents also included other central government strategies like Information Security Strategy for Finland (Ministry of Transport and Communications, 2016) and others.

# **2. Strategic leadership of cyber security**

## **2.1 Definitions of strategic leadership of cyber security**

It must be noted that strategic leadership is not an unambiguous term (Juuti & Luoma 2009). It may be defined and understood in many ways. Additionally, the "boundaries" between strategic and operative management are not always clear in all situations of cyber security management, and in some instances, they are difficult to separate (while this may even be unnecessary). Different definitions of strategic leadership and the difficulty of separating strategic and operative activities also emerged, among other things, in the context of the interviews conducted for this study and the reference countries selected for the international comparison. For example, strategic leadership of cyber security was described as follows in the interviews: "Strategic leadership means leading a phenomenon at the highest level, making an effort to define long-term visions and objectives as comprehensively as possible".

Cyber security is an aspect of society's and companies' security, which is highly important when considering an organisation's strategic goals in an increasingly digital society. In the source documents of the study, strategic leadership of cyber security was often described as securing the central government's capabilities and vital

---

<sup>1</sup> The interviewees represented the following organisations: CGI Finland Oy, Confederation of Finnish Industries, Elisa Corporation, F-Secure Oyj, Fingrid Oyj, Finnish Information Security Cluster, National Emergency Supply Agency, Emergency Response Centre Administration, Insta Group Oyj, National Bureau of Investigation, Ministry of Transport and Communications, National Police Board, Ministry of Defence, Finnish Defence Forces, Ministry of the Interior, SSH Communications Security Oyj, State Security Networks Group, Finnish Technology Industries, Tieto Corporation, Security Committee, Prime Minister's Office, Government ICT Centre Valtori, Ministry of Finance, Finnish Communications Regulatory Authority (incl. National Cyber Security Centre Finland), and Ministry for Foreign Affairs.

functions, also allowing the private and the NGO (non-governmental organisation) to build their activities on well-functioning and secure information networks. Based on these documents, the most important task of the strategic leadership of cyber security is defined as creating a vision and a national mentality which are recognised at all levels of actors participating in cyber security work and which direct the actions in both normal and emergency situations. (Lehto & Limn  l, 2016).

Strategic leadership comprises *long-term implementation* of Finland's Cyber Security Strategy (2013) and Finnish cyber security. Strategic leadership brings society towards the selected vision. The task of implementing strategic leadership is based on *identifying and setting objectives* derived from securing the digital operating environment.

Secondly, strategic leadership *reconciles, coordinates and ensures participation in cooperation between different actors* in cyber security activities and preparedness. As cyber security is an extensive societal phenomenon which connects a great number of different actors, the coordination of cooperation is stressed both in normal and emergency conditions and during incidents. Sufficient preconditions for making decisions and clearly defined powers are stressed in the activities.

Thirdly, as cyber security is a strategic issue for Finnish society, the strategic leadership of cyber security takes place in close *interaction with both political decision-making and operative activities*. Strategic leadership is also associated with *strengthening Finland's cyber security identity*, both nationally and internationally. The cyber security identity is also associated with *seeing to national cyber self-sufficiency* regarding both product and service solutions and expertise and research in Finland. Domestic and international *communications play an important role* in creating a Finnish cyber security identity and credibility based on trust. Several international indicators are specifically geared to measuring cyber security identity and capability. One of the goals of strategic leadership thus is the continuous monitoring of the status of national cyber capability (as a whole) to understand its current level and to *improve the capability*.

Fourthly, strategic leadership *creates coherence and continuity* for Finland's collaborative efforts at both the national and international level. Strategic leadership gathers all available resources together in order to *achieve the set targets*.

To sum up: *The strategic leadership of cyber security comprises identifying and setting objectives derived from securing the digital operating environment, coordinating activities and preparedness, and extensive leadership in incident management*.

## **2.2 Strategic leadership of cyber security – a current status analysis based on research**

The challenges of cyber security management are particularly prominent at the level of strategic leadership. The challenges of the current state are reflected in the views brought up in several of the interviews conducted for the study concerning (1) clear and concrete proposals for strategic leadership structures of cyber security, and (2) the need to discuss the importance of this issue and the required measures with integrity and avoiding any ambiguity. Two basic problems have been identified at the level of strategic leadership:

- 1) The number of actors is large, and for this reason, the strategic leadership of cyber security is fragmented and lacks clear leadership. The ministries carry out the strategic leadership of cyber security independently in their own sectors, and consequently overall strategic leadership is lacking, and the activities are to a great extent siloed in the various administrative branches.
- 2) No effective cooperation structure exists at the level of strategic leadership of cyber security. This is partly linked to the first problem. The ministries look at cyber security on the basis of their own needs, losing sight of the wider societal perspective, and the aforementioned objectives defined for strategic leadership are not achieved.

The interviews conducted for the study indicate that the strategic leadership of cyber security must be recognisable to avoid a situation where administrative branches have no leadership and the requisite measures cannot be carried out. In the current situation, strategic leadership is expected to take care of itself, even if this is not necessarily the case. In the current state, interdependencies between different actors in society have not

been described. Once the relationships between organisations and functions have been described, the impacts decisions will have on societal functions can be anticipated. Experts believe that research to study the interdependences is needed as soon as possible. Identifying the cooperation partners, actors producing information and the entire cyber observation system would be important first steps towards a genuine cross-cutting security strategy for entire society. The interviews conducted for the study indicate that discretion will be needed concerning a body/function that takes care of coordination across organisational boundaries to ensure that its actual role does not remain illusory and that it does not add to the workload unnecessarily.

In a prior study, a need to centralise the management of Finnish cyber security to the Prime Minister's Office emerged, in particular (Lehto et al. 2017). According to the experts interviewed for the study, due to its direct dialogical connections and role as a function supporting top-level government, the Prime Minister's Office is better placed to assume strategic leadership than other branches of government or organisations. The present model is also linked to EU level cyber security models that the Prime Minister's Office reconciles with national models. The cyber security work at the Prime Minister's Office has close links to the Government's work in this area. The Government serves all branches of administration equally and coordinates their cooperation. Models and practices planned for the Prime Minister's Office are relatively similar to those planned for the Government. However, the Prime Minister's Office has no direct authority over the different ministries, and proposed measures are thus implemented through advice and instructions. According to the interviewees, the present model does not respond fast enough to incidents. The National Audit Office (2017) recommends that "the Ministry of Finance define and implement an operative management model for extensive cyber incidents regarding government ICT services".

In terms of effective strategic leadership of cyber security, it is vital to identify structures that can respond to the operative requirements set by the environment. Typical features of the cyber environment are an accelerating rate of change, a phenomenon-based approach, complexity and, in part, unpredictability. The interviewees stressed that the pre-sent model of strategic leadership is unable to respond to the ever-faster rate of change. The loop formed by gathering information on which decisions are based, making the decision and implementing it is currently too slow. "As the vulnerability of society increases it is necessary to be able to rapidly start managing sudden disturbances in the cyber domain".

According to the majority of experts interviewed for the study, a precondition for the current role of the Prime Minister's Office is that existing forms and practices of cooperation for responding to an urgent crisis in the cyber environment have been negotiated. It is almost never possible to negotiate on measures in urgent crisis situations, and a mandate and operating models tested as part of the preparedness process should exist for taking action. The Finnish Communications Regulatory Authority's National Cyber Security Centre has established methods for managing incidents together with private sector actors. Rather than on authority, this procedure is based on cooperation, in which the Cyber Security Centre serves as the contact point.

The interviews conducted for the study indicate that strategic leadership is based on building and maintaining trust. Even today, deep trust and doing things together are the basis for thwarting cyber threats. Fundamental trust has enabled exceptionally good cooperation between different actors in Finnish society, and this cooperation has long traditions in Finland. The National Cyber Security Centre is a good example of how trust achieves more than obligation. So far, keeping the cyber environment safe has been based on identifying key actors and conducting negotiations between them, rather than cyber security management structures. The interviewees even questioned the strategic leadership of cyber security due to factors stemming from the operating environment. A systematic action model, but the issue of the strategic leadership of government cyber security was found challenging.

According to the experts interviewed for the study, *a strategic leadership model of cyber security should be created, as currently there is no strategic leadership of cyber security.*

### **2.3 Strategic leadership of cyber security in the future**

New technologies challenge the current legislation and raise ethical questions concerning such issues as cyberattack capability, autonomous vehicles, artificial intelligence and augmented reality. The advancing technologies mean that the cyber environment is in constant flux which, according to the interviewed experts, hampers the creation of permanent and straightforward operating models. Diversification of activities, group

processes and a correct type of balance between the mechanical and organic nature of activities are means for managing this complexity.

Based on the interviews, research literature and an analysis of the reference countries, successful strategic leadership of cyber security requires:

- Effective legislation,
- Sufficient powers,
- Links to political decision-making,
- Capabilities and expertise, and
- Financial resources.

The interviewees identified leadership capability as a success factor in the strategic leadership of cyber security. The experts referred to historical cases where decisions were made on the wrong grounds without understanding their impacts on our society. Strategic leadership should facilitate interaction between the state's political leadership and, on the other hand, those responsible for operative activities, ensuring that both parties understand each other and that their actions are coherent. According to the experts, one perspective to leadership is that the highest level in the national management of cyber security, or strategic leadership, should be assigned to a ministry that has genuine capabilities for leading the activities.

### **3. Situational awareness as part of strategic leadership**

#### **3.1 Challenges of the current state**

The situational picture of the cyber environment is fragmented, and any understanding of it as a whole is based on information shared between the authorities, the private sector, researchers and experts. The interviewees expressed differing views of cyber security situational awareness. Some found the national cyber security situational awareness fragmented and incomplete. A situational picture that would cover all national cyber environment actors is not being put together and analysed, and capability for making decisions is lacking. Lack of powers prevents the creation of efficient observation capability and thus a cyber security situational picture needed for effective management. While different actors have systems built for their own use, shared national situational awareness that could be used both at the strategic and operative level is lacking. Some of the interviewees felt that the situational picture was good, or at least sufficient, in general terms. The current operating model is sufficient for managing minor cyber-attacks, but situational awareness and understanding are inadequate for thwarting complex and extensive attacks. (Lehto et al. 2017)

The structure for maintaining situational awareness was improved as the strategy was formulated, but it continues to have shortcomings at the practical level. Some of the interviewees felt that shared situational awareness is not implemented at the level of ministries. While the administrative branches do not necessarily have an idea of cyber security in society as a whole, they have a relatively good understanding of it in their own sectors. The interviewees also found that exchanges of information partly take place between specific persons. On the other hand, as yet unresolved questions are associated with maintaining a shared situational picture, including who needs what information, on what cycle it is needed, and what type of information is required. Regarding the nature of information, more analysed information on threats as well as unrealised and actual incidents together with solution models for them are called for. From the perspective of improving cyber security preparedness, we must be able to trust that information will flow during incidents and that the actors will know how to respond to it as indicated by their duties. (Lehto et al. 2017)

While there would also be a demand for a sector-specific situational picture service, the National Cyber Security Centre is not currently able to meet it in all respects. Capability for extensively recognising the impacts of incidents should be developed in other sectors of society besides the one specifically affected by the incident. The Cyber Security Centre needs additional resources for developing sectoral situational awareness data and situational picture reserves.

### **3.2 Analysis of the current state of cyber security situational picture, awareness and understanding**

The different parties involved in developing national situational awareness should be able to improve their operations through more effective technical methods, strengthen network-based operation and focus on utilising technical methods in shared use.

The most significant organisations associated with the functional capacity of Finnish society have developed a relatively good ability to observe the situational picture for the part of technical capabilities. Their ability to do so is also improved by networking within their sectors and partly also more extensively, which is supported by good cooperation between the authorities and the private sector. The significance of situational awareness shaped by different organisations' situational pictures (situational picture and its analysis) for the management of entire national cyber security is crucial.

Research has stressed the organisations' possibilities of drawing on different networks for cyber security situational awareness as a national strength, which has also been referred to in previous reports. At least three types of networks relevant to exchanging confidential information can be identified, and they are used actively. They have emerged in connection with business activities, or a specific trust network has been set up between companies in a sector, which may also extend to international cooperation. Additionally, a national trust network between the authorities and the private sector is in place (PPP cooperation).

The response to national incidents consists of the techniques used by different organisations, procedures developed for responding to incidents, and the observation data of different trust networks. This fragmented ability to observe the organisation-specific situational picture and the data reserves it entails could also be used in the analysis phase of large-scale incident management. Preconditions for this arrangement would include the creation of joint operating models and an arrangement based on voluntary exchange of information. A joint data warehouse would enable the further processing of information to analyse a large-scale incident. The required analysis capabilities could be implemented as a network (virtual analysis).

## **4. Models for the strategic leadership of cyber security and situational awareness**

Alternative models for the strategic leadership of cyber security in Finland were produced in the study. The five models presented below are based on the views of personnel with managerial roles and experts of the field in the interviews conducted for the study, international evaluations of reference countries, views presented in the research literature/documents as well as assessments made by the authors.

Five models for the strategic leadership of cyber security are presented:

- 1. The present model
- 2. A national cyber security manager
- 3. A national cyber security unit
- 4. A strengthened National Cyber Security Centre
- 5. A Cyber Security Agency.

### **4.1 Present model**

In the present model, cyber security is managed as part of seeing society's vital functions, and no separate strategic leadership or management process is created for it.

The strengths of this model include its familiarity (management of cyber security is integrated in existing arrangements for incident management) and minor need for rearrangements in the administration. The Finnish (cyber) security actors are relatively familiar with each other, which facilitates information exchanges and smooth cooperation, even if no unambiguous line of command related to cyber security has been defined. This model is underpinned by the current legislation.

The model's weakness lies in its uncertain ability to respond sufficiently fast to large-scale cyber-attacks or incidents and to produce anticipatory strategic analysis data essential for preparing for ever changing cyber

threats. The present management structure cannot be considered optimal in terms of the coordination of preparedness, identification of strategic goals or strengthening of the national cyber security identity. The present model does not provide sufficient guidance for the cyber security preparedness of the administrative sectors, businesses and the NGO, or produce sufficiently centralised capabilities for strategic analysis to support the production of situational awareness. In the present model, shortcomings are associated with the identification and development of national cyber self-sufficiency. No close link between political decision-making and strategic leadership of cyber security, which was stressed in international comparisons, is manifested clearly.

#### **4.2 A national cyber security director**

In this model, the role of the top director of cyber security is set up in the Prime Minister's Office or, alternatively, a ministry or an organisation with a key role in cyber security.

A key strength of this model is a clear chain of command in cyber security work: the appointed cyber security manager would coordinate, lead or support cyber security work in all situations. Management would also take place close to political decision-making and steering. In this model, however, a single person would be appointed to direct an area with no dedicated resource allocation.

In this situation, management across administrative boundaries would be challenging, as resource allocations and management systems would be specific to each administrative sector. As another weakness of the model may be considered the concentration of disproportionately great power and responsibility to a single person. As strategic leadership comprises an extensive set of tasks, the possibilities of a single person carrying out all the specified tasks may be questioned. If the cyber security manager's role is limited to loose coordination, the management of both preparedness and incident response will remain cursory. In a rapidly escalating incident, fast and effective links should be in place between the strategic leadership and operative actors, and each party should have clear-cut powers.

#### **4.3 A national cyber security unit**

The model of a national cyber security unit is similar to the national cyber manager model. A separate cyber security unit subordinate to the cyber security manager would be set up with capabilities for directing, developing and supporting national cyber preparedness and for promoting the realisation of the national cyber security vision in a broader sense.

The strengths of the cyber security unit would include its placement close to political decision-making and its ability to direct and develop cyber security activities cross-administratively. From the point of view of management, this can be considered a relatively agile and centralised model, in which the manager is supported by his or her own unit in performing a large range of tasks. In reference countries, corresponding units are placed either in the prime minister's office or the ministry with general responsibility for security and justice, or similar tasks are handled by the organisation that has the overall responsibility for the coordination of national security activities. In this model, the management of cyber security is partly integrated with existing arrangements for incident management, and transition to it would thus result in limited needs to rearrange the administration. This would reduce the workload and ambiguities created by changing the arrangements for comprehensive security.

#### **4.4 A strengthened cyber security centre**

In this model, the National Cyber Security Centre would be placed under the steering of a cyber security manager, and its operative competence and powers would be complemented with capabilities for strategic analysis. The Centre's situational picture function would be reinforced with strategic analysis capabilities with the aim of producing situational awareness in support of strategic decision-making. The Centre would be co-located with the cyber security manager, and it would work in close cooperation with the Government's Situation Centre. The Situation Centre would continue to perform the task of providing a situational picture for the entire Government and all administrative sectors.

The strengths of this model include the proximity of strategic and operative actions, a clear line of command and a straightforward approach, which would translate as agility in deploying capabilities, thus serving the

maintenance of strategic stability while enabling action in unexpected situations. Transition to this model would require limited changes to existing comprehensive security arrangements, including more specific arrangements for cross-administrative cooperation as the Cyber Security Centre takes on its new role. Significant additional resources would also have to be allocated to the Cyber Security Centre.

The weaknesses of this model would include a fragmentation of cyber security functions and the fact that various functions would remain in different administrative sectors. It is also likely to take time before the Cyber Security Centre's reference groups (current and future ones) adapt to its new role.

#### **4.5 A cyber security agency**

This model is based on setting up a Cyber Security Agency, which would handle the strategic leadership of cyber security and cyber security functions.

In the agency model, key cyber resources of the central government can be combined into an effective whole, through which the efficiency of both cross-administrative cooperation and collaboration with businesses can be improved. This model would provide an improved ability to respond to changes in customer needs and the operating environment, develop and strengthen the strategic steering of cyber security, and obtain synergy benefits. It can also improve the productivity and, more particularly, the impact of the administration through more diverse and effective resource use.

The weakness of this model is the partial transfer of cyber security functions away from the administrative sectors, with the resulting losses of knowledge of and expertise in the sectors' special features. To create the Agency, broad-based reforms of the existing administrative structures, modifications to the line of command and responsibilities, and adequate resource allocation would be needed. Administrative friction would undermine the efficiency of the activities during a transition period until the new operating model becomes established.

### **5. Conclusion**

In this study, the management models proposed by the research project have been described at the level of principle. Research questions were answered, but the continued preparation of one of the presented leadership models or some other model will require drafting by public servants. In this case, the requisite operative and organisational changes and legislative amendments should be investigated, financial reviews should be carried out, and an extensive assessment of the impacts including statements and a schedule for implementing the reform should be prepared.

The proposed models contain risks, the number and impacts of which are comparative to the scale of the change. The risks may lead to inappropriate solutions when arranging the operations. Consequently, in any further drafting particular attention should be paid to ensuring the quality, reliability and undisrupted continuation of the activities and service level during the potential change and following it.

In order to develop the situational picture, awareness and understanding needed for strategic leadership, the efficiency of the current operating model should be improved, and the data warehouse and cooperation network should be developed. Centralisation of strategic leadership also results in more efficient situational awareness and understanding. The following characteristics, among other things, affect the impact of management: consistency of action in different situations, prevention of siloed activities, taking interdependencies into account, and coordination of activities. The bottlenecks of situational awareness and strategic leadership activities have frequently been identified in expert assessments provided for research projects, and centralisation can thus help optimise limited resources.

### **References**

- Juuti P., Luoma M. (2009). Strateginen johtaminen, Kustannusyhti   Otava, Keuruu 2009, pages 24-27. ISBN-13: 9789511236399
- Lehto M. & Limn  l J. (2016). Cyber Security Capability and Case Finland, Proceedings of the 15th European Conference on Cyber Warfare and Security (ECCWS), 7.-8.7.2016 Munich, Germany, pages 182-190

**Jarno Limnéll and Martti Lehto**

Lehto M., Limnéll J., Innola E., Pöyhönen J., Rusi T., Salminen M. (2017). Finland's cyber security: the present state, vision and the actions needed to achieve the vision, Publications of the Government's analysis, assessment and research activities 30/2017. ISBN 978-952-287-368-2

Lehto M., Limnéll J., Kokkomäki T., Pöyhönen J., Salminen M. (2018). Strategic leadership of cyber security in Finland, Publications of the Government's analysis, assessment and research activities 28/2018. ISBN 978-952-287-532-7

Ministry of Defence (2015). Finland, Guidelines for Developing Finnish Intelligence Legislation, Working group report, March 2015.

[https://www.defmin.fi/files/3144/GUIDELINES FOR DEVELOPING FINNISH INTELLIGENCE LEGISLATION.pdf](https://www.defmin.fi/files/3144/GUIDELINES_FOR DEVELOPING FINNISH INTELLIGENCE LEGISLATION.pdf)

Ministry of Transport and Communications (2016). Information Security Strategy for Finland The World's Most Trusted Digital Business Environment, Publications of the Ministry of Transport and Communications 9/2016.

[https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/78416/Publications\\_9-2016\\_Information\\_Security\\_Strategy\\_for\\_Finland.pdf?sequence=1](https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/78416/Publications_9-2016_Information_Security_Strategy_for_Finland.pdf?sequence=1)

National Audit Office (2017). Performance audit report: Cyber Protection Arrangements, 16/2017.

[https://www.vtv.fi/files/5862/16\\_2017\\_Kybersuojauskens\\_jarjestaminen.pdf](https://www.vtv.fi/files/5862/16_2017_Kybersuojauskens_jarjestaminen.pdf)

Prime Minister's Office (2013). Central Government Communications in Incidents and Emergencies, Regulations, instructions and recommendations issued by the Prime Minister's Office 3/2013.  
[https://vnk.fi/documents/10616/1093242/M0313\\_Central+Government+Communications+in+Incidents+and+Emergencies.pdf/e277d048-6ff2-481c-b678-41388e2b6cef/M0313\\_Central+Government+Communications+in+Incidents+and+Emergencies.pdf.pdf](https://vnk.fi/documents/10616/1093242/M0313_Central+Government+Communications+in+Incidents+and+Emergencies.pdf/e277d048-6ff2-481c-b678-41388e2b6cef/M0313_Central+Government+Communications+in+Incidents+and+Emergencies.pdf.pdf)

Security committee (2013). Finland's Cyber Security Strategy and its background dossier, 24 January 2013  
<http://turvallisuuskomitea.fi/index.php/fi/component/k2/14-suomen-kyberturvallisuusstrategia>

Security committee (2017a). Implementation Programme for Finland's Cyber Security Strategy 2017–2020.  
<https://www.turvallisuuskomitea.fi/index.php/fi/mcdc/126-suomen-kyberturvallisuusstrategian-toimeenpanoohjelma-2017-2020>

Security committee (2017b). Security Strategy for Society, Government resolution, 2.11.2017.  
[https://turvallisuuskomitea.fi/wp-content/uploads/2018/04/YTS\\_2017\\_english.pdf](https://turvallisuuskomitea.fi/wp-content/uploads/2018/04/YTS_2017_english.pdf)

# PhySec in Cellular Networks: Enhancing Security in the IIoT

**Christoph Lipps, Mathias Strufe, Sachinkumar Bavikatti Mallikarjun and Hans Dieter Schotten**

**German Research Center for Artificial Intelligence, Kaiserslautern, Germany**

[Christoph.Lipps@dfki.de](mailto:Christoph.Lipps@dfki.de)

[Mathias.Strufe@dfki.de](mailto:Mathias.Strufe@dfki.de)

[Sachinkumar.Bavikatti\\_Mallikarjun@dfki.de](mailto:Sachinkumar.Bavikatti_Mallikarjun@dfki.de)

[Hans\\_Dieter.Schotten@dfki.de](mailto:Hans_Dieter.Schotten@dfki.de)

**Abstract:** There is currently a jolt through the industrial landscape, based on the actual developments in the field of Cyber-Physical Production Systems (CPPS) and the Industrial Internet of Things (IIoT). New application scenarios within these Factories of the Future, such as Machine-to-Machine (M2M) and Machine-to-Service (M2S) communication burst the traditional communication links within Industrial Automation and Control Systems (IACS). There is an interconnection of assembly lines, robot motion control systems, Automated Guided Vehicles (AVG) as well as a multitude of sensors and actuators. The key enablers of this development are at first the mobility and flexibility of the components due to the use of wireless connections, and secondly, the possibility of actively influencing the network management with Software Defined Network (SDN) approaches. Nevertheless, the use of this wireless communication solutions are accompanied by great risks, new attack vectors, and cyber security threats. The open nature and the broadcast characteristic suffer a huge potential for miscellaneous cyberattacks. Not only because of that, but there is a fundamental need for sound and secure authentication of participating entities and reliable encryption of transmitted data. However, traditional cryptographic applications come along with a lot of overhead in the form of complex computations and communication. Besides that, the new system often no longer have any interface to enter conventional credentials. In order to face these requirements, new methods have to be developed, which meet the demands of the industry such as low latency, low cost and, reliable communication. Within this work, Physical Layer Security (PhySec) approaches are used to derive and establish shared secret keys between participating entities. This Secret Key Generation (SKG) is based on characteristics of the wireless channel. It is shown that a transfer of the principle into Next Generation Mobile Networks (NGMN) such as Long Term Evolution Advanced (LTE+) or the upcoming Fifth Generation (5G) also perform as well. This approach is an easy to use, low cost, resource saving and efficient method to enable confidence and trust into IIoT systems. And this with already existing hardware.

---

**Keywords:** physical layer security, secret key generation, industrial internet of things, next generation mobile networks

## 1. Introduction/motivation

Currently, we are on the verge of the fourth industrial revolution, with the merging of all kinds of devices, in both, the private and the industrial sector. This fusion of technologies and different devices, across technological borders, lead up to so-called Cyber-Physical Production Systems (CPPS) and the Industrial Internet of Things (IIoT).

The driving forces of this process are the mobility, flexibility and, portability, enabled by the developments and enhancements in the wireless communication sector. Without the use of this technology, the IIoT would be inconceivable. The wireless connectivity enables to break the chains of limitations and constraints that are caused by the use of cables and static connectors. The full potential can only be exploited by eliminating these rigid connections.

The huge challenge is to provide a sound and secure application of the procedures. There is a need to authenticate the participating entities within the IIoT, as well as a suitable key management system that is capable to handle the vast amount of interconnected devices. Traditional solutions use cryptographic procedures that come along with an overhead in the form of computational complexity as well as limitation of the available bandwidth. Additionally, while using wireless connections, there is another issue. Due to the open nature of the systems and the broadcast characteristic, they are vulnerable to miscellaneous cyber-attacks, in particular, eavesdropping. This applies not just for IEEE 802.11 Wireless LAN (WLAN) or short-range services such as IEEE 802.15 Wireless Personal Area Networks (WPANs), such as Bluetooth or ZigBee. All wireless networks have similar weaknesses, including cellular networks.

For these, too, conventional cryptography is primarily used. Wang et al. pointed out that communication security for cellular networks is a crucial issue (Wang, et al., 2013). However, as already mentioned, mobile radio solutions will be increasingly used in the future. The reasons for this are the popularity of smart devices and the

demand for exuberant multimedia content (Yang, et al., 2015), also in the industrial sector (see Section 4 and the industrial use-cases). However, since the existing standards such as 3G, Long Term Evolution (LTE) and 4G are not capable to support a large number of participating devices and the amount of traffic demand (Roh, et al., 2014), the Next Generation Mobile Networks (NGMNs) with the upcoming Fifth Generation (5G) are being developed. However, still with transmission technology related vulnerabilities.

One worthwhile approach to mitigate them is offered by Physical Layer Security (PhySec) solutions. The underlying concept of this is the exploitation of a channel intrinsic randomness of the used communication medium and the reaping of the benefits offered by disruptive influences (Yang, et al., 2015).

The remainder of this work is organized as follows. In Section 2 an introduction to Physical Layer Security (PhySec) is given, to get a basic introduction of the topic. Section 3 provides an overview of State-of-the-Art work of others, including the PhySec methods of the previous Section. Building upon this, Section 4 migrates the PhySec concept to cellular networks and give an insight into the requirements of current and future industrial production environments. The developed testbed is introduced in Section 5. Finally, Section 6 concludes the work and provides an outlook on future work and the next steps to the implementation of the concept.

## **2. Physical layer security**

Physical Layer Security comprises a series of different procedures to extract cryptographic credentials from individual physical characteristics of various components. There are currently two major branches of this PhySec.

First of all, there are silicon or electrical approaches that exploit the physical randomness out of components, that results of slightest deviation occurring during the manufacturing process of the individual components. These procedures are also referred to as Physically Unclonable Functions (PUFs) (Gassend, et al., 2002) (Suh & Devadas, 2007). They can be further subdivided into procedures that are timing-based, which means that they make use of individual delays within the signal processing. This includes, among others Arbiter-, Ring-Oscillator-(RO), and Butterfly-PUFs (Herder, et al., 2014).

Furthermore, there are approaches such as SRAM-PUFs (Lipps, et al., 2018) that use the individual, voltage based threshold values of semiconductor devices to derive a cryptographic key, too.

Secondly, PhySec includes approaches that exploit the special characteristics of a wireless channel. These so-called Channel-PUFs (Lipps, et al., 2019) utilize the randomness of a wireless channel and generate a shared secret, a cryptographic key, on both sides of the used channel. This principle is illustrated in Figure 2 and discussed again in Section 4 and explained in more detail.

However, PhySec approaches, in particular, the wireless based version, offers, according to (Yang, et al., 2015), two major benefits compared to standard cryptographic applications:

- they do not depend on any computational complexity, and
- they have high scalability.

Many of the common used cryptographic procedures use complex computations to establish a secure key. In addition to a certain computing time, this requires computing power and thus also resources in the form of power. Furthermore, the generated key must be exchanged over a secure channel and stored in the memory of the participating entities. Alternatively, a key is integrated into the device from the outside before it is delivered.

The addressed benefit of the PhySec method is, that nothing has to be exchanged over an (insecure) channel.

## **3. Current work on physical layer security**

After the introduction of the general PhySec approach, in the previous Section, now a differentiation is made from the work of others. Since the basic concept of PUFs was introduced in 2002, meanwhile there is quite a lot of research in this area. Nevertheless, the number of wireless solutions is manageable, especially in the area of cellular approaches. And, above all, most of the work is theoretical.

An experimental study on Secret Key Generation (SKG), based on PhySec methods, is done by (Zhang, et al., 2016). They developed testbeds for different Wireless communication systems, by using Wireless Open\_access

Research Platform (WARP) hardware. Furthermore, (Weinand, et al., 2018) propose, besides others, a Plug & Trust protocol for Ultra-Reliable Low Latency Communication (URLLC). They introduce a security architecture for wireless network security within closed-loop control systems.

An application of PhySec principles in multiuser wireless networks is given by (Mukherjee, et al., 2014). They detail the work mentioned so far, by introducing and explaining different channel types, such as multi-antenna channels, broadcast channels and multiple access and interference channels. In addition they provide some metrics and give an overview of the PhySec concept.

(Ambekar, et al., 2012) provide a detailed description of the individual steps, as well as suggestions for improvement and adjustments of these steps.

A distinction between PhySec in the heterogeneous network, PhySec in massive Multiple-Input-Multiple-Out (MIMO) systems and PhySec in Millimeter-Wave communication (mmWave) is provided by (Yang, et al., 2015).

A survey of different PhySec approaches, including Next Generation Wireless Networks (NGWN) is given by (Liu, et al., 2017). They provide an overview of existing work and also point out different focus issues of these works, including multi antenna technologies. Furthermore, they name challenges and issues of wireless systems, such as integration of fading influences or deviations while using MIMO systems.

(Lipps, et al., 2019) suggest a combination of different PhySec approaches. They combine SRAM-based Physically Unclonable Functions (PUFs), for device authentication, and they SKF methods based on wireless characteristics, which they define as a *Channel-PUF*.

Besides this WLAN typical approaches, there are a few approaches dealing with PhySec in cellular networks. (Wang, et al., 2016) are doing research in heterogeneous cellular networks, as well as mmWave networks. They build a system model that integrates different scenarios, such as directional beamforming, small scale fading and path loss. Besides that, (Wang, et al., 2013) provide an information theoretic secrecy performance model in large scale cellular networks. Additionally, (Wang & Wang, 2016) are dealing with mmWave cellular communication and compare their results with conventional microwave networks in the bands below 6GHz.

A study of PhySec methods in the downlink of cellular networks, where each Base Station (BS) simultaneously transmit confidential messages to several users and where the confidential messages to each user can easily be eavesdropped by an attacker, is analysed by (Geraci, et al., 2014).

(Wang, et al., 2016) propose a stochastic geometry approach for PhySec in cellular networks. They extend the information theoretic secrecy performance in large scale cellular networks and provide a system model with orthogonal multiple access, and a single class BS. Thereby they focus on the performance achieved by randomly chosen typical mobile users.

Within their work of safeguarding 5G wireless communication with PhySec, (Yang, et al., 2015) utilize the unique characteristics of different ad-hoc networks and carrier operated, high-speed backhaul networks while connecting to individual base stations.

(Chen & Willems, 2018) are doing research about key generation over biased PUFs. Thereby, they are using polar codes.

There are already some works on PhySec in different wireless communication standards. Nevertheless, especially the works about cellular PhySec are primarily theoretical work. Stochastic models are developed and the procedures are simulated. The work presented within this paper, extends these approaches. Section 5 introduces a testbed, currently under construction. With this the different approaches, which are also used in the works mentioned above, are tested with-real world conditions.

An overview of these different works and their priorities is given in Table 1.

**Table 1:** Overview of the work of others

PhySec Method	Topic/Content	Work
Wifi PhySec	Testbed for different Wifi standards	(Zhang, et al., 2016)
	Security Architecture// Plug&Trust Protocol for URLLC	(Weinand, et al., 2018)
	Metrics and overview of different channel types	(Mukherjee, et al., 2014)
	Survey of different PhySec papers	(Liu, et al., 2017)
	PhySec in different channel types	(Yang, et al., 2015)
	Detailed description of the SKG	(Ambekar, et al., 2012)
	Combination of different PhySec methods	(Lipps, et al., 2019)
Cellular PhySec	Multi-Antenna cellular networks	(Geraci, et al., 2014)
	Large Scale cellular networks	(Wang, et al., 2013)
	mmWave cellular communication	(Wang & Wang, 2016)
	5G PhySec	(Yang, et al., 2015)
	Stochastic geometry model	(Chen & Willems, 2018)
	PhySec in heterogeneous cellular networks	(Wang, et al., 2016)

As already mentioned, most of the existing works of cellular PhySec are theoretical approaches, models and simulations. The work, introduced in this paper, aims to develop and implement a real world testbed to prove and validate these models.

Only through implementation and testing under real world conditional reliable statements can be made about the quality and usability of the intended approach. Furthermore, this opens up the possibility to adjust and enhance the implemented approaches.

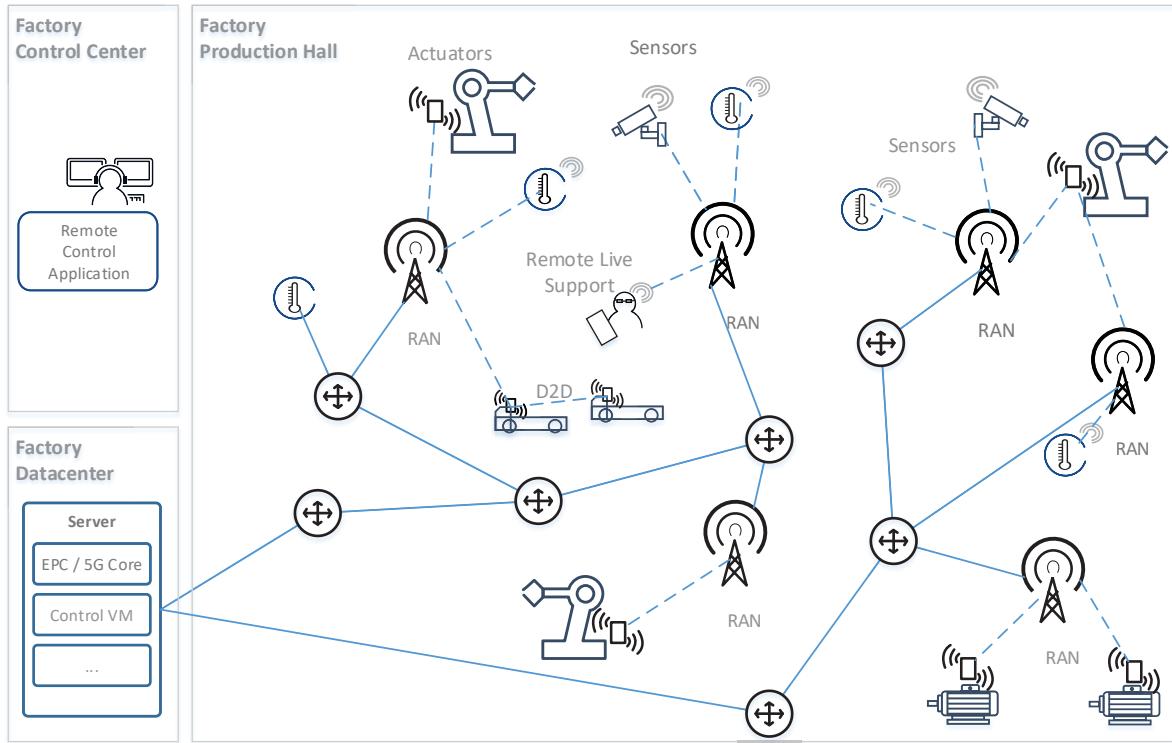
Besides that, the work of others, for instance those in Table 1, provide promising and also solid results for PhySec in “traditional” wireless networks, such as WLAN. Because of this, the expectation is to achieve similar results within cellular networks and to enhance the security level within the future industrial production landscape.

#### 4. Cellular approach

Since current industrial plants with “brownfield” deployments mainly rely on wire-line technologies, 5G will be a key enabler to perform the paradigm change to the highly flexible production lines intended in the fourth industrial revolution referred as Industry 4.0. Wireless connections are crucial for several industrial use cases like shown in Figure 1. Automated guided vehicles (AGVs) for cooperative transport of goods and platooning, maintenance staff with augmented reality (AR) glasses for remote live support, drones for industrial inspections and machine motion control need highly reliable, ultra-low latency, high data rate as well as wide radio coverage and large number of devices to connect (Gundall, et al., 2018) which only 5G cellular networks can provide.

However, wireless transmissions are in particularly vulnerable to security attacks. In wide area public networks, this is usually solved by physical SIM cards and pre-shared 128-bit-master-key cryptographic technologies which need to be installed manually in the User Equipment (e.g. smartphone) and Core network. In a Non Public Network (NPN) operated directly by the factory owner, like planned in the near future, where millions of sensor and actors can be wirelessly connected, the provisioning of each device with a secure element will be not an option. This is where PhySec applies. The main advantages of PhySec is the good scalability since no need for key distribution and management which make it perfectly suitable for massive IIoT communications.

Although the use of PhySec is already well studied including prototypes in wireless systems such as IEEE 802.11n, Ultra-Wideband (UWB) and Bluetooth the research on PhySec in cellular networks just started theoretical considerations.

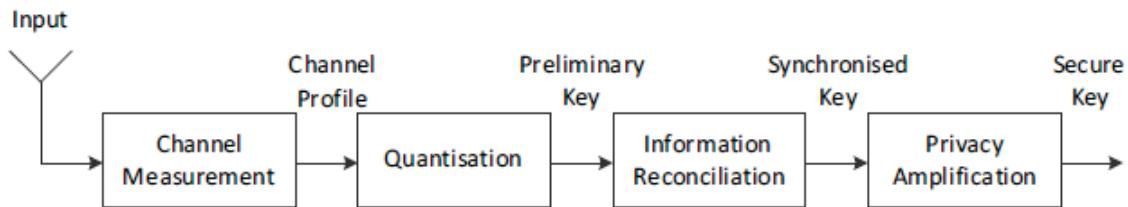


**Figure 1:** Wireless Industrial IoT production environment

Key generation is mainly based on the characteristics of the channel between two communication endpoints. Therefore, a standard method of key generation, also for cellular networks, consists of four phases:

- 1. Channel Measurement,
- 2. Quantisation,
- 3. Information reconciliation and
- 4. Privacy amplification

as shown in Figure 2.



**Figure 2:** Standard method of key generation

Two common methods to determine the channel profile are either the Received Signal Strength (RSS) or the Channel Impulse Response (CIR). RSS has the advantage that it can be quite easily accessed either in the UE as well as from the base station. However, the CIR measurement can provide a more accurate channel profile. We also plan to investigate the use of further cellular PHY layer parameter such as channel quality indicator (CQI), Reference Signal Received Power (RSRP) or Signal to Interference & Noise Ratio (SINR) as a base for the key generation.

The next step is the quantisation of the channel measurement. This can be either done losslessly where all measurements are considered or lossy with an upper and lower threshold where values in between got discarded.

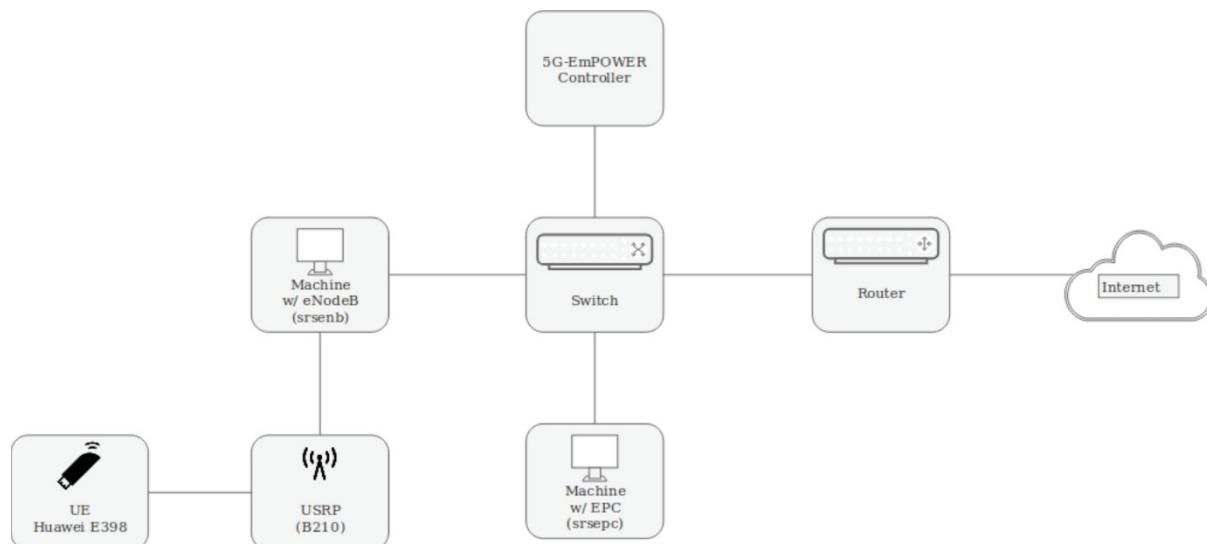
After the channel measurements got quantised, the preliminary key on both sides doesn't match exactly. Therefore, error detection and correction is performed to synchronise both keys. Some of the most common

methods of channel reconciliation are: Convolution codes, Turbo codes, Cyclic Redundancy Check (CRC), Low Density Parity Check (LDPC) or Hash functions. In the last step, the key gets enhanced to create a final secure key.

Within this step, the security of the generated key is enhanced by privacy amplification via secure hashes, fuzzy extractors or Pseudo-Random Generators to minimise the possibility of key prediction by the eavesdropper.

## 5. Cellular PhySec testbed

As 5G is still in the final standardisation phase we start with an LTE Testbed to perform the first real-world experimentations. The testbed consists of a generic router, one switch, four MiniPCs, a Universal Software Radio Peripheral (USRP) B210 device and a commercial Huawei E398 LTE dongle. Figure depicts the testbed that is been used in this experiment. All MiniPCs have an Intel i5 CPU at 2.30 GHz and 8 GB of RAM and runs on Ubuntu 18.04 Operating system. One is used for the Evolved Packet Core (EPC) including the Mobility Management Entity (MME) for the primary signalling, the Home Subscriber Server (HSS) which currently stores critical security information and the Serving and Packet Data Gateway (SPGW). The used USRP B210 is a fully integrated, single board with frequency coverage from 70 MHz – 6 GHz and 2 TX & RX Half or Full Duplex channels driven by a Xilinx Spartan 6 FPGA. The UHD driver is installed in version 3.13.0.1. USRP B210 is interfaced with the MiniPC through USB 3.0. A generic router and switch are used to connect the MiniPCs and serve as the gateway to the internet.



**Figure 3:** Full LTE testbed setup

To realise the LTE EPC srsLTE is used. It is a free open source LTE software suite developed by SRS (Software Radio Systems) under AGPLv3 licence and is fully LTE Release 8 compliant (including selected features of Release 9).

In this testbed, we use srsENB and srsEPC. srsLTE has been built and installed in two separated PCs. Once srsLTE is installed, variables like Mobile Country Code (MCC) and Mobile Network Code (MNC) in configuration files like enb.conf, epc.conf are to be modified based on USIM (Universal Subscriber Identity Module) configuration to get connected to the virtual base station's network. The COTS Huawei E398 dongle connected to a MiniPC via USB2 and act as our user equipment (UE) is currently equipped with a USIM. The USIM is configured with IMSI, key, Operator type, Operator code, Authentication management field, UE's Sequence number, QoS Class Identifier and the configuration of the USIM has to be updated in user\_db.csv and this file holds the details of all the configured UE that are allowed to connect to network created by virtual base station (srsLTE). In epc.conf file, under HSS configuration, the authentication algorithm is set to mileage which is the currently used key generation functions of LTE SIM defined in 3GPP. The location of the .csv file that stores UEs information is set to user\_db.csv and other configuration in epc.conf and enb.conf are left unchanged. UEs (Huawei E398) Access Point Name (APN) is set to the same APN which is configured in enb.conf.

In addition to srsLTE software which implements LTE network, 5G-EmPOWER controller is used and it is capable of managing resources among virtual base stations and separating control and data plane. The controller is used to retrieve the Received Signal Strength Indicator (RSSI) of the eNodeB to create channel profile in this testbed.

## **6. Conclusion and future work**

This paper describes challenges, concepts and current research work of Physical Layer Security for wireless communication. While the Industry 4.0, the Industrial Internet of Things and Cyber-Physical Production Systems are the key enablers for flexible and adaptable industrial automation, PhySec is worthwhile tool to provide secure wireless communication, especially through the required scalability. Furthermore, the rising demand of exchanged data as well as the needed bandwidth pave the way for the enhancements of the upcoming Next Generation Mobile Networks, such as 5G.

Within this paper, an LTE testbed is described to perform real world PhySec experimentations to implement and evaluate current PhySec techniques.

As previous work in this area has mainly focused on simulations and theoretic models, the proposed testbed is capable to validate the simulation results. An application of the PhySec in IEEE802.11 WLAN environments already delivers good and promising results. The simulations and models of colleagues support the thesis that similar promising results can be expected with cellular networks. The testbed is capable to validate this thesis.

Moreover, the objective is to improve and enhance the existing algorithms, primarily with the help of Artificial Intelligence methods. These enables an advanced and extended application of PhySec, as well as they are can be used to minimise the possibility of key prediction by the eavesdropper

Approaches such as Software-Defined Networking, that enables a global view about the networks, are likewise useful extensions of the “pure” PhySec concept and are intended to integrate in the described cellular PhySec concept.

## **Acknowledgements**

This work has been supported by the Federal Ministry of Education and Research of the Federal Republic of Germany (Förderkennzeichen 16KIS0932), IUNO Insec and KIS15GTI007 TACNET. The authors alone are responsible for the content of the paper.

## **References**

- Ambekar, A., Schotten, H. D. & Kuruvatti, N., 2012. Improved Method of Secret Key Generation Based on Variations in Wireless Channel. Vienna, Austria, s.n.
- Chen, B. & Willems, F. M. J., 2018. Secret Key Generation over Biased Physical Unclonable Functions with Polar Codes. *IEEE Internet of Things Journal*, p. 1.
- Gassend, B., Dwaine, C., van Dijk, M. & Srinivas, D., 2002. Silicon physical random functions. *Proceedings of the 9th ACM conference on Computer and communications security*, pp. 148-160.
- Geraci, G. et al., 2014. Physical Layer Security in Downlink Multi-Antenna Cellular Networks. *IEEE Transactions on Communications*, 62(6), pp. 2006-2021.
- Gundall, M. et al., 2018. 5G as Enabler for Industrie 4.0 Use Cases: Challenges and Concepts. *IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA)*.
- Herder, C., Meng-Day, Y., Farinaz, K. & Srinivas, D., 2014. Physical Unclonable Functions and Applications: A Tutorial. *Proceedings of the IEEE*, 102(8), pp. 1126-1141.
- <https://github.com/5g-empower/5g-empower.github.io/wiki/kein-Datum-5G-empower.s.l.s.n>.
- <https://github.com/srsLTE/srsLTE>, 2019-01-31. srsLTE. s.l.s.n.
- Lipps, C., Duque Antón, S. & Schotten , H. D., 2019. Enabling Trust in IIoT: An Physec based Approach. *International Conference on Cyber Warfare and Security (ICCWS19)* , pp. 663-672.
- Lipps, C. et al., 2018. Proof of Concept for IoT Device Authentication based on SRAM PUFs using ATMEGA 2560-MCU. *1st International Conference on Data Intelligence and Security (ICDIS)*.
- Liu, Y., Chen, H.-H. & Wang, L., 2017. Physical Layer Security for Next Generation Wireless Networks. *IEEE Communications Surveys & Tutorials*, 19(1), p. 347–376.
- Mukherjee, A., Fakoorian, S. A. A., Huang, J. & Swindlehurst, A. L., 2014. Principles of Physical Layer Security in Multiuser Wireless Networks. *IEEE Communications Surveys & Tutorials*, 16(3), p. 1550–1573.
- Roh, W. et al., 2014. Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results. *IEEE Communications Magazine* , 52(2), pp. 106-113.
- Suh, G. E. & Devadas, S., 2007. Physical Unclonable Functions for Device Authentication and Secret Key Generation. *DAC* .

***Christoph Lipps et al.***

- Wang, C. & Wang, H.-M., 2016. Physical Layer Security in Millimeter Wave Cellular Networks. *IEEE Transactions on Wireless Communications*, 15(8), pp. 5569-5585.
- Wang, H.-M. et al., 2016. Physical Layer Security in Heterogeneous Cellular Networks. *IEEE Transactions on Communications*, 64(3), pp. 1204-1219.
- Wang, H., Zhou, X. & Reed, M. C., 2013. On the Physical Layer Security in Large Scale Cellular Networks. *IEEE Wireless Communications and Networking Conference (WCNC)*.
- Wang, H., Zhou, X. & Reed, M. C., 2016. Physical Layer Security in Cellular Networks: A Stochastic Geometry Approach. *IEEE Transactions on Wireless Communications*, 15(6), pp. 2776-2787.
- Weinand, A., Michael, K. & Schotten, H. D., 2018. Security Architekture and Solutions for Local Wireless Networks in Closed Loop Control Applications based on Physical Layer Security. *IFAC Conference on Embedded Systems, Computational Intelligence and Telematics in Control CESCIT*, 51(10), pp. 32-39.
- Yang, N. et al., 2015. Safeguarding 5G wireless communication networks using physical layer security. *IEEE Communications Magazine*, 53(4), pp. 20-27.
- Zhang, J. et al., 2016. Experimental Study on Key Generation for Physical Layer Security in Wireless Communications. *IEEE Access*, Issue 4, pp. 4464-4477.

# Technical Guidelines for Evaluating and Selecting Data Sources for Cybersecurity Threat Intelligence

Jabu Mtsweni<sup>1, 2</sup> and Muyowa Mutemwa<sup>1</sup>

<sup>1</sup>Council of Scientific and Industrial Research, Pretoria, South Africa

<sup>2</sup>University of South Africa, Florida, South Africa

[mtswenij@gmail.com](mailto:mtswenij@gmail.com)

[mutemwam@csir.co.za](mailto:mutemwam@csir.co.za)

**Abstract:** Extracting actionable threat intelligence from what is perceived to be relevant data sources in the cybersecurity environments still faces a number of unsolved challenges. The challenges that are normally associated with cyber threat intelligence data sources include the capturing, storage, cleaning, processing, querying and visualization of this data in a useful and contextual manner. Other challenges include the guidelines on how to assess and select relevant threat intelligent data sources from a myriad of these sources available on the Internet from third-party sources. With the advent of social media, large-scale internet scans, and web 4.0 technologies, the generation of security information by different sources is a reality. Nevertheless, the value and veracity of this information remains a challenge. The main objective of this paper is to provide technical guidelines for evaluating and selecting data sources for threat intelligence in cybersecurity environments. The significance of the guidelines is to provide simple to use elements to support cybersecurity researchers and organisations in dealing with various data sources in the cybersecurity domain for contextual and relevant insights to improve own security posture.

**Keywords:** cyber security, cyber threats, vulnerabilities, cybersecurity threat intelligence, data sources, intelligent feeds

---

## 1. Introduction

Security information is becoming increasingly important for proactive responses by individuals and organisations to the complex and forever changing cyber threats. As such, the availability and accessibility of innovative cyber threat intelligence sources to the cybersecurity community is becoming paramount. Today, a number of cyber threat intelligence data sources are made available both to individuals and organisations for awareness of cyber threats and attacks (Mtsweni, Mutemwa and Mkhonto, 2016). These data sources come in different shapes, cost, and context. Some of the data sources make partial cyber threat information available for free to the users and extensive information for a fee. Examples of such data sources include *Shodan.io* (Sentient Hyper-Optimised Data Access Network) and *VirusTotal* from Google that makes available current malware information, but also allows for on the fly analysis of suspicious files and websites (Swart, 2015; J Mtsweni *et al.*, 2016).

These data sources are also continuously changing and increasing in size including relevance and recentness. However, when working with these myriad of data sources for cyber threat intelligence (CTI), a number of challenges emerge. As new data sources become available, containing cybersecurity threat intelligence, it also does not equate that this new data source is relevant, reliable, clean and ready for analysis. In addition, the business and technical value of these emerging cyber threat intelligent sources to organisations is always difficult to measure. At a higher level, it is also a challenge to determine the veracity of the data sources for threat intelligence as there are currently no standards or guidelines that exist to influence this process. It is also a challenge to predict and manage the size or volume of these data sources, which in turn could affect the investment in technology infrastructure required for collection, storage and processing of such data in order to extract contextual cyber threat intelligence.

The main objective of this paper is therefore to investigate and characterise different types of data sources for cybersecurity threat intelligence. Emanating from this investigation and characterisation process would be technical guidelines for evaluating and selecting data sources for threat intelligence in cybersecurity environments.

The rest of this paper is structured as follows: Section 2 provides a brief background on the research methodology adopted for research presented in this paper. In Section 3, cyber threat intelligence is defined in the cybersecurity context, and background information is provided. Different data sources relevant to threat intelligence are discussed and characterised in Section 4. Section 5 presents the proposed technical guidelines for evaluating and selecting data sources for contextual threat intelligence in cybersecurity environments. In Section 6, an illustrative use-case scenario is provided to demonstrate the application of the technical guidelines

proposed in this paper using a selection of relevant data sources. This research paper is concluded with a summary of the results, significance, and recommendations for future research in Section 7.

## **2. Research methodology**

The research presented in this paper follows the Design Science Research (Hevner et al., 2004) which subscribes to the concept of an artefact. In this paper, the technical guidelines proposed are considered as an artefact. These are derived through a systematic process, where the literature review was conducted to clearly derive features of CTI data sources and threat intelligence. From the literature, it is apparent that there currently exist no specific guidelines for selecting and evaluating data sources for cybersecurity threat intelligence. In an attempt to characterise common features in existing data sources for cybersecurity relevant information, a technical assessment and comparison of selected data sources was conducted. Based on the characterisation process, the technical guidelines were defined, represented by a conceptual framework that was evaluated through a use case scenario that may be applicable in a security operations centre environment.

## **3. Definitions**

In this section, the definitions used throughout this research paper are provided. The key definitions provided include cybersecurity, cyber threat intelligence from different perspectives, and big data.

### **3.1 Cybersecurity**

The International Telecommunication Union (ITU) defines cybersecurity as a “*collection of tools, policies, security concepts, security safeguards, guidelines, risk management approaches, actions, training, best practices, assurance and technologies that can be used to protect the cyber environment and organization and user's assets*” (ITU, 2018). Based on this definition, it is essential to note that cybersecurity is more than just technology, but also touches on people and processes.

### **3.2 Cyber threat intelligence**

Cyber threat intelligence (CTI) can be loosely defined as analysed, contextual and actionable security information related to computers, networks and information technology (Mtsweni, Mutemwa and Mkhonto, 2016). Nevertheless, there is no silver bullet in CTI, thus this section clarifies what is meant by CTI in the context of this study.

According to Mavroeidis and Romander (Mavroeidis and Bromander, 2017), “*threat intelligence is the provision of evidence-based knowledge about existing or potential threats*”. From a strategic view point, CTI includes motivation of the adversary. From an operational perspective, “*threat intelligence is defined as information that can aid decisions, with the aim of preventing an attack or decreasing the time taken to discover or respond to an attack*”. From a security perspective, CTI can be defined as a combination of actionable external and internal sources of threat information refined, correlated and analysed to prepare and counter cybersecurity incidents (Sekoia, 2017). The benefits of threat intelligence include improved efficiency and effectiveness in security operations in terms of predictive, detective and preventive capabilities (Mavroeidis and Bromander, 2017).

## **4. Cyber threat intelligence data sources**

CTI data sources tend to be either provided by a commercial third party or via voluntary aggregators (Swart, 2015). Furthermore, a big data source may contain threat intelligence pertaining to a specific type of a threat across different domains, organisations, countries or entire Internet. As a result, the intelligence contained in the data sources of interest may vary in accuracy, method of representation, measurement and a variety of other significant factors. It is therefore important to also highlight that some data sources after careful assessment might not be strictly seen as original sources of the data, but aggregators of different cybersecurity feeds. The technical guidelines contributed in this paper are motivated by the fact that some data sources contain threat intelligence that requires further assessment and correlation for accuracy and relevance.

In this section, we highlight a selected number of CTI data sources. The data sources selected mostly contain data sets that purport to provide threat intelligence for the cybersecurity environment. The number of these data sources is not exhaustive, and were selected based on the availability of the data, type of cyber threat information, the recency of the data, the scope and context of the data, data format supported, including integration support.

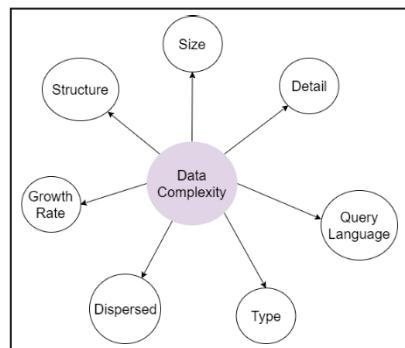
**Table 1:** Overview and analyses of commonly known CTI data sources

Data Source	Description and Analysis
Breach Aware	It is a platform with an aggregated list of recent data breaches. It is commercial, and has various pricing structures with different access levels. It provides for real time credential leak monitoring, and has API and SIEM integration functionality. It can also be used to monitor entire organisational domain for data leaks. The context of the data set is organisational, and uses external data sources, such as paste-bin. Data format support is JSON, and consumers of this data source could also subscribe to RSS feeds (see: <a href="https://breachaware.com/dashboard">https://breachaware.com/dashboard</a> ).
Censys	It is an online platform that continuously scans every host on the Internet to provide for real-time visibility and threat intelligence into unknown and potentially vulnerable servers. It requires licensing for commercial use, with limited queries for the free version supporting only one user. The commercial version can support from 25000 queries depending on subscription type and different types of users. This data source supports searches, JSON data format, SQL queries, raw data downloads, and APIs integration for commercial versions. Data made available through this data source include IPv4 banner information, port scans, and websites (see: <a href="https://censys.io/">https://censys.io/</a> ).
Common Vulnerability Exposure (CVE)	This is a dictionary of publicly known information security vulnerabilities and exposures associated with different applications, operation systems and software. The CVE data is open available and is update on regular basis (i.e. every 2 days). The CVE database has capability to use the open information exchange standards such as STIX, however it only supports XML data format (see: <a href="https://cve.mitre.org/">https://cve.mitre.org/</a> ).
Pastebin Dump	This is a paste website used mostly by the community of hackers to frequently publish samples of complete datasets of compromised data. The data types would vary from one paste to the other, but mostly included leak personal information such as emails, usernames and passwords. Pastebin provides API integration and data is updated on daily basis and has been growing since the inception. The data source has tens of millions of records, which may contain personal identifiable information. However, veracity of the data is always a challenge since collection of data is mostly automated and a number of duplication tend to exist on this data source. In many cases triangulation is required to determine the reliability of the data. Data dumps are updated on regular basis, and supports JSON format. With a small fee, other functionalities can be accessed, which are mainly for contributing data to the data source (see: <a href="https://psbdmp.ws">https://psbdmp.ws</a> )
Dshield	This is a community-based collaborative firewall log correlation system. It receives logs from volunteers worldwide and uses them to analyse attack trends. It is also open source, and is updated on daily basis. It provides an API for integration purposes, and data-context is worldwide, but does not implement any information exchange standards, but supports various file formats such as XML, CSV, TXT, and JSON, which makes it easier to integrated with third-party systems (see: <a href="https://www.dshield.org/">https://www.dshield.org/</a> ).
ExploitDB	The Exploit Database is a repository for public exploits, proof-of-concepts and corresponding vulnerable software, developed for use by penetration testers and vulnerability researchers. The sharing model used is open source, and has an API for integration. It is updated frequently as exploits become available. The data source does not collect personal identifiable information, and its data is relevant to different environments across the globe (see: <a href="https://www.exploit-db.com/">https://www.exploit-db.com/</a> ).
GreyNoise	This web-based tool analyses Internet background noise and can be used to remove non-essential security alerts, find compromised devices using ip addresses, or identify emerging threats. It collects its data from various data sources including Shodan, Censys, SSH and telnet worms, and generally supports JSON data format. It has a limited free version, and enterprise version that provides between 500 and 50 000 API queries per day, including providing data types for identifying comprised devices, and filtering false positives. GreyNoise also provide a query language called GNQL, a domain specific language for complex and one-off queries, but only available to enterprise and research users, and this language also contextual extraction of data for specific threat intelligence and use cases (see: <a href="https://viz.greynoise.io/">https://viz.greynoise.io/</a> ).
Haveibeenpawned	It is free service that collects and analyses database dumps and pastes containing information about leaked accounts, and allows users to search for their own information by entering their username or email address. Other IOCs made available through this data source include passwords and domain names. It provides an API service to query various data types include breaches of accounts or systems and output is available in JSON. Data is updated on ad-hoc basis as breaches are detected, and integrates with other data sources such as Pastebin dumps (see: <a href="https://haveibeenowned.com/">https://haveibeenowned.com/</a> ).
PhishTank	PhishTank is a collaborative platform for data and information about phishing on the Internet cutting across the globe. It is based on open source sharing principles. It is updated on hourly basis with information coming from the community. It provides APIs for integration, and supports multiple data formats (e.g. JSON and XML). The indicators of compromise included the data

Data Source	Description and Analysis
	include URLs, email address, DNS information and TTPs for phishing (see: <a href="https://www.phishtank.com/">https://www.phishtank.com/</a> ).
Ransomware Tracker	Tracks and monitors the status of domain names, IP addresses and URLs that are associated with Ransomware, such as Botnet Command and Control servers, distribution sites and payment sites. It is also open-source based, but it is not clear how often it gets updated. For integration purposes, it provides developers with an API, but does not use any open standards for information exchanges. The context of the data is mostly global (see: <a href="https://ransomwaretracker.abuse.ch/">https://ransomwaretracker.abuse.ch/</a> ).
Shodan	This is a search engine designed to map and gather information (e.g. vulnerable devices) about internet-connected devices and systems. It has some data that can be accessed on limited basis, but generally requires subscription and licensing to integrate with custom systems using own API and extract data. Shodan supports both JSON and XML data formats and data context is country specific and even to an organisational or user level. Indicators of compromise (IOCs) vary from IPs, URLs, and certain types of vulnerabilities. It does not use open information exchange standards, its data gets updated on regular basis, but reliability and integrity of the data is not always guaranteed, since historical data is also made available via the platform (see: <a href="https://www.shodan.io/">https://www.shodan.io/</a> ).
Virustotal	This is an online service that aggregates different antivirus tools and online scan engines to scan for malware on users' files, websites or suspicious documents using various threat intelligence. The data source is registration base, but provides access also via APIs mostly using JSON data format. Extraction of data has some limitations with only four requests per minute. This data source is relevant to different environments across the globe and IOCs made available via this data source include IPs, botnet address, email address, DNS names and others. The data and IOCs can be verified by the community members, and data is updated on daily basis, and as such requires scalable storage and processing resources. However, because the veracity depends highly on the community especially for new virus samples, sometimes bad actors could temper with such a data source as there is no exclusions on who can provide data into this data source including the voting on the reliability of the data samples (see: <a href="https://www.virustotal.com/">https://www.virustotal.com/</a> ).

## 5. Related cyber threat intelligence models

There are different approaches that can be adopted to analyse CTI data sources in order to gain value and aid good business decisions. Because CTI data sources come in different shapes, size and cost, it is at present, a challenge to systematically determine the usefulness and actionable-ness of the data sources. This is further made complex by the myriad availability of these data sources. According to Casey (2017), there are seven signs that can be used to determine the relevance of the data in any environment, including its relevance and complexity when making business decisions. Figure 1 overleaf shows some of the essential nodes to consider when dealing with data for extracting business value.



**Figure 1:** Seven factors to classify complex data (Castle, 2017)

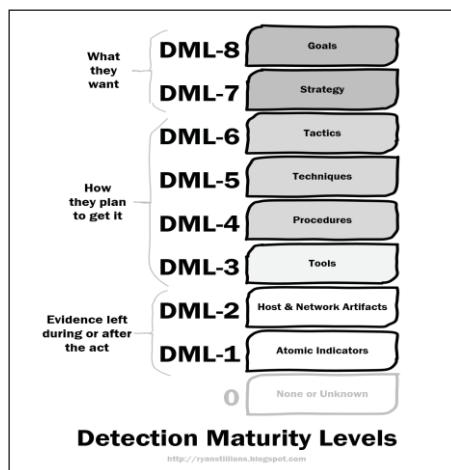
- **Structure:** data sources can present threat intelligence using structured, semi-structured and unstructured data formats, and this has a direct impact on the extraction of CTI. Different CTI data sources present their information using different structures, and this can at times make it a challenge to extract useful threat intelligence.
- **Size:** the amount of data can affect the significance of the intelligence, but also the processing of the data because the larger the size of the CTI data source the requirements for processing will also be different.
- **Detail:** the level granularity that is required from the data to aid decision making is critical. In CTI, this is probably the most important element when considering using any data source for threat intelligence.

- **Query language:** as with the structure element, different data sources accommodate different technical languages. Structured Query Language (SQL) may be fit to use for structured data stored in a relational database. However, when working with semi-structured or data stored in files with no structure, SQL may not be useful for retrieval and processing. Thus, when evaluating a data source, the language to be used for querying the data is vital.
- **Data Type:** different types of data require different methods for useful extraction. CTI data sources may come in different forms and it is thus critical to consider data types and the ability of the organisation to process such data type when selecting sources for CTI.
- **Dispersed:** when a data source relies on information spread across different location, standardisation, policies, legislations and cross-border data transfers need to be considered, and these may affect the CTI.
- **Growth rate:** the speed at which the data growth or changes need to be considered when selecting data sources for CTI as the data source increases in its size, so will the requirements and resources to process such data.

In addition to the seven generic factors highlighted above, there are other related CTI models available in literature that could be useful to evaluate and select data sources for threat intelligence. The non-exhaustive list is discussed in the following subsections.

### 5.1 Detection Maturity Level model (DML)

The DML model is a capability maturity model for determining an organisation's maturity in detecting cyber-attacks at any given time (Stillions, 2014). This model can be used by organisations for gathering threat intelligence from a detection and response perspective. The model is composed of 8 maturity levels, with level-0 indicating no intelligence (i.e. no information is known about an attack or incident) and level-8 detailing the *goals* of the attacker.



**Figure 2:** DML model (Stillions, 2014)

The model has also been refined by Mavroeidis and Bromander (2017) to include level-9 that further considers the *identity* of the threat actor which is critical in understanding the attack and improving organisational cyber security response and posture. The model is relevant for CTI, however does not provide the necessary guidelines for technical assessing data sources that could be used to determine the suggested maturity levels. The proposed technical guidelines in this paper attempts to close this gap.

### 5.2 Structured Threat information eXpression (STIX)

The Structured Threat Information Expression (STIX) is a language or format used for exchanging cyber threat intelligence (Barnum, 2014; J. Mtsweni *et al.*, 2016). The advantage of STIX is that it enables sharing of CTI in a consistent and machine-readable manner. The objective of STIX is to enable organisations to detect and respond to cyber-attacks effectively and faster, and improve threat analysis and information exchange. STIX provides for the following information to be shared: (1) vulnerability, (2) Indicators of Compromise, (3) Threat Agent, (4) Campaign, (5) Observables or patterns (6) TTPs, and other information. STIX is expressive, however, most data sources tend not to comply with it since it mandates JSON or XML format (Barnum, 2012), and as such

unstructured data sources would not be able to have most of the information required by STIX. STIX has grown over the years in the threat intelligence community to be the de-facto standard for threat information exchange, however, it still not suitable for use when selecting and evaluating data sources to be used for CTI.

### 5.3 Diamond model of Intrusion Analysis (DMIA)

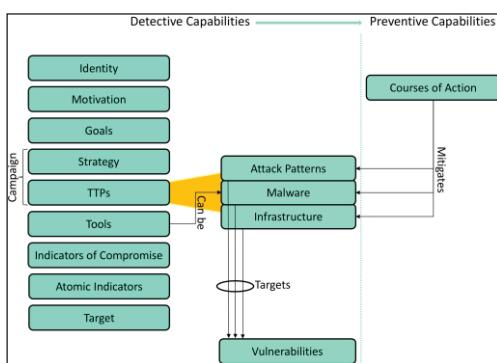
The diamond model of intrusion analysis is generally used for threat modelling, particularly for tracking and analysing features of a cyber-incident. The model is based on four key elements: (1) adversary (2) capabilities (3) infrastructure, and (4) victim (Caltagirone, Pendergast and Betz, 2013). These four elements are described through the model's axiom that states “*for every intrusion event, there exists an adversary taking a step toward an intended goal by using a capability over infrastructure against a victim to produce a result*” (Carreon, 2018). As may be noted, this model is high-level and may not be suitable for use when selecting and evaluating data sources for CTI.

### 5.4 Open Indicators of Compromise (OpenIOC)

OpenIOC addresses information formatting and exchange during incident investigation. It provides a standard format and terms for describing the artefacts found during a cyber-incident investigation. In order for an IOC to be useful, it would normally be made up of three elements: metadata, references and the definition. These components would be used to describe artefacts that may include attacker's activity, tools used, malware, and other indicators of compromise (IOCs) (OpenIOC, 2015). OpenIOC relies on the XML schema to describe the technical characteristics that identify a known threat, an attacker's methodology, or other evidence of compromise. It is also prudent to note that OpenIOC is meant to be used in combination with human intelligence and machine-digestible intelligence. Their objective is to aid an investigation of cyber-incidents (Gibb, 2013). For purposes of selecting and evaluating data sources for CTI, OpenIOC could provide some leads, however, because its components are not descriptive enough, application of OpenIOC in different use cases can be subjective and lead to CTI that is difficult to action.

### 5.5 Cyber Threat Intelligence model

The Cyber Threat Intelligence (CTI) model was devised by Mavroeidis and Bromander (2017) to characterise threat intelligence using almost similar elements as in the DML model discussed above. The slight difference being that the CTI model is not hierarchical, and suggests information for threat intelligence and attack attribution. It is further divided into detective and preventive capabilities.



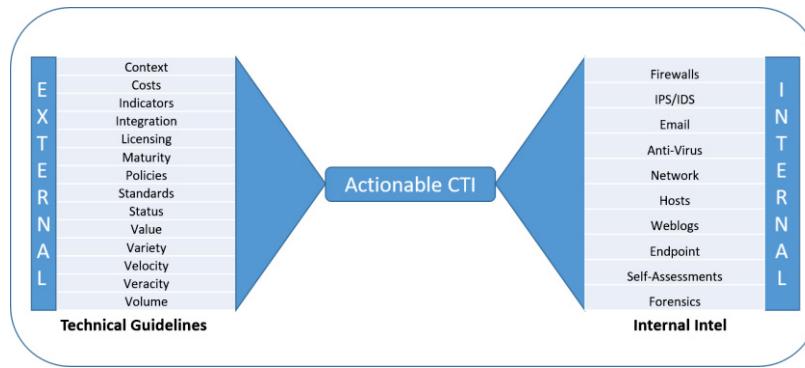
**Figure 3:** Cyber Threat Intelligence model (Mavroeidis and Bromander, 2017)

Generally, the CTI model is not that distinctive from the DML and STIX models. Our proposed technical guidelines also draw some of the features from the CTI model, however the focus is on selecting and evaluating the data sources to use for organisational CTI than only attribution or incident investigation. The following section delves into the proposed technical guidelines for selecting and evaluating relevant and actionable CTI data sources.

## 6. Technical guidelines for selecting and evaluating data sources

In proposing the technical guidelines for selecting and evaluating data sources for use in forming cyber threat intelligence in an organisation, a systematic process was followed. Firstly, the different data sources that are potential for use in CTI were systematically analysed in *Section 3* (see *Table 1*). From this non-exhaustive list, it is clear that CTI data feeds or sources come in different variations, and one data source may never be enough

to build actionable CTI. Thus, when building CTI for an organisation, it is crucial that a multitude of recent data sources are evaluated and reviewed regularly. Secondly, an analysis of existing CTI models was done, and this is summarised in the previous section (see Section 5). This analysis concluded that there is currently no model that exists to aid the selection and evaluation of CTI data sources. All the models analysed focused mostly on deriving CTI for cyber investigations. However, this may be helpful in reactive investigations, but not proactive environments. In this section, we present the technical guidelines that were derived by analysing all the selected data sources focusing on their content, recency, value, data type, and other information. These were also compared against the prescriptions of the existing models as discussed in Section 5.



**Figure 4:** Proposed technical guidelines for selecting CTI data sources (own copyright)

The proposed technical guidelines include 14 elements that focus on the external environment. However, as may be noted in Figure 4, our view is that threat intelligence cannot be derived only from external sources, but would need to be complemented with internal sources (as depicted in Figure 4) in order to have actionable CTI for an organisation. The guidelines can be applied in any particular order, however for an effective evaluation, a top-down approach is suggested. For instance, before any CTI data source is selected, it is important to determine if they are suitable for the *context* to which they are intended to be applied. For example: if a potential CTI data source has threat information related targeting European regions, then such a data source may not be so useful for an organisation in an African region. Thus, it is critical that when selecting and evaluating CTI data sources that context is clearly defined.

Furthermore, *costs, policies and licensing* requirements of the CTI data source need to be assessed upfront as some data sources provide limited information for free, and may require payment when requiring access to bigger data sets or even updates or data may not be used for commercial purposes. Each data source will only provide different types of *indicators* (either atomic indicators or IOCs). For example: some data sources only focuses on virus signatures submitted onto their platforms. In evaluating the indicators provided by the data sources, the CTI model by Mavroeidis and Bromander (2017) is suggested as presented in Figure 3. The indicators can be evaluated based on the objective of the CTI that is required.

It is also critical to evaluate how the CTI data is generated to enable one to determine the *value* of the data for CTI in a business, the speed at which this data is generated (*velocity*), the trustworthiness of the data (*veracity*), including the scale or size of the data (*volume*). The veracity, volume and value of these data sources for cyber threat intelligence is influenced by many factors (Mtsweni, Mutemwa and Mkhonto, 2016) other than just the authenticity of the information provided. Some of these could include Indicators of Compromise (IOCs), Data Structure (e.g. JSON/XML) or Threat, Techniques and Procedures (TTPs), which are not one-size-fits-all. As such, it is critical that the cybersecurity community is provided with technical and practical guidelines for evaluating and selecting relevant big data sources for threat intelligence. As noted by Swart (2015), an understanding of the data provided by these sources can afford decision makers the opportunity to set priorities more effectively.

Different data sources support multiple *data formats and standards*. For instance, the PhishTank data source can be extracted using JSON/XML/CSV/PHP formats or even plain text files, whilst other data sources only accommodate one data format (e.g. XML). This is critical to assess upfront when planning to use specific CTI data sources, because data formats have a direct impact on the processing of the data for threat intelligence. In addition, some data sources adopt common standards such as STIX in order to enable automated information sharing and processing. However, many of the third-party data sources, especially those discussed in Section 4 (see Table 1) do not subscribe to threat intelligence exchange standards. If data sources do not support common

information exchange standards, it is essential to evaluate if these data sources provide some mechanisms for *integration* (e.g. provides APIs) into other systems. This is vital when dealing with big data sources where manual processing and collection of the CTI data may not be effective.

Finally, the *maturity and status* of the CTI data sources need to be evaluated upfront before selection, and this is because many of the data sources we have analysed for this research study tend to be short-lived (i.e. still at proof-of-concept level), and the status of some of these sources remain unknown or inactive, with limited updates over an extended period of time. Data sources that are not regularly updated may not be so useful for threat intelligence in the cybersecurity environment where threats and risks change on daily basis.

The technical guidelines proposed in this section provide the initial attempt in guiding the assessment and selection of data sources amongst many other data sources for threat intelligence. In the following section, an illustrative use-case scenario is provided to demonstrate the simple application of such technical guidelines

## 7. Use case scenario

In order to briefly demonstrate how the technical guidelines could be applied in a matrix format, we selected a limited list of the data sources as highlighted in Section 4, and then assessed them against the guidelines. An illustrative use case scenario for the evaluation focuses on a retail organisation with a web payment processing system that handles credit card information. The organisation wishes to setup actionable CTI in order to understand their cybersecurity posture on regular basis (using internal data sources), compare themselves against other similar organisations and proactively respond to attempted cyber-breaches on their web payment system (using external data sources).

**Table 2:** Use-case scenario evaluation

	Sample of CTI data sources					
	Breach Aware	CVE	Dshield	PhishTank	VirusTotal	ShodanIO
Technical Guidelines						
Context	Red	Green	Yellow	Green	Yellow	Green
Costs	Yellow	Green	Yellow	Green	Yellow	Red
Indicators	Red	Green	Yellow	Green	Yellow	Green
Integration	Green	Green	Green	Green	Yellow	Green
Licensing	Yellow	Green	Yellow	Green	Yellow	Yellow
Maturity	Yellow	Green	Green	Green	Green	Green
Standards	Red	Yellow	Green	Yellow	Yellow	Yellow
Status	Yellow	Green	Green	Green	Green	Green
Value	Yellow	Green	Yellow	Green	Yellow	Yellow
Variety	Yellow	Green	Green	Green	Green	Green
Velocity	Green	Yellow	Green	Green	Yellow	Yellow
Veracity	Yellow	Green	Red	Green	Yellow	Yellow
Volume	Green	Yellow	Green	Green	Yellow	Yellow

The results of the assessments are shown in the Table 2 above. Based on the scenario, the colours have the following meaning: (1) red: guideline not met (2) amber: guideline partially met (3) green: guideline fully met as per the scenario. Based on the evaluation using the proposed technical guidelines to select the relevant CTI data source, it can be deduced from Table 2 that over 50% of the data sources may not be useful for actionable CTI related to the use-case. In the same light, CVE database and PhishTank data sources may be selected for generating threat intelligence related to the use-case scenario. The only criterion not fully met by the PhishTank data source is the veracity of the data, since a disclaimer from the PhishTank terms of use mention that the accuracy of the data available through this service cannot be guaranteed. Both the CVE and PhishTank data sources partially addresses the aspect of information exchange using common standards. For instance, both data sources use JSON and XML, but do not subscribe to STIX or other open information exchange standards. The CVE database stores vulnerabilities of different software since the 1990s and as such this data can be quite large if not contextualised. Based on the use case, the retail organisation may choose to employ only those sources that are fully-compliant. However, they may also focus on the elements that are important in their environment. If for instance, veracity is not such a critical component for them, then even if a data source does not guarantee veracity, they may still choose to use it for CTI.

## **8. Conclusion and future research**

Cyber threat intelligence is become an integral of a cybersecurity capability in any organisation. However, many organisations do not have the necessary skills or knowledge on how to filter and select the relevant data sources that could aid the threat intelligence decisions when dealing with evolving threats and risks. This paper has systematically demonstrated through the highlights of different data sources and existing related models that technical guidelines are necessary for evaluating and selecting data sources for CTI in an organisation. The technical guidelines presented in this paper may be enhanced through further research and extensive use case scenarios.

## **References**

- Barnum, S. (2012) 'Standardizing cyber threat intelligence information with the Structured Threat Information eXpression (STIX™)', *MITRE Corporation*. Available at: [http://stix.mitre.org/about/documents/STIX\\_Whitepaper\\_v1.1.pdf](http://stix.mitre.org/about/documents/STIX_Whitepaper_v1.1.pdf) (Accessed: 8 December 2015).
- Barnum, S. (2014) *Standardizing Cyber Threat Intelligence Information with the Structured Threat Information eXpression (STIX)*. Available at: [https://stix.mitre.org/about/documents/STIX\\_Whitepaper\\_v1.1.pdf](https://stix.mitre.org/about/documents/STIX_Whitepaper_v1.1.pdf).
- Caltagirone, S., Pendergast, A. and Betz, C. (2013) 'The Diamond Model of Intrusion Analysis'. Available at: <https://apps.dtic.mil/docs/citations/ADA586960> (Accessed: 12 March 2019).
- Carreon, C. (2018) *Applying Threat Intelligence to the Diamond Model of Intrusion Analysis, Recorded Future*. Available at: <https://www.recordedfuture.com/diamond-model-intrusion-analysis/> (Accessed: 12 March 2019).
- Castle, E. (2017) *7 Signs You're Dealing with Complex Data, SISENSE*. Available at: <https://www.sisense.com/blog/7-signs-youre-dealing-with-complex-data/> (Accessed: 12 March 2019).
- Gibb, W. (2013) *OpenIOC: Back to the Basics*, *Fire Eye Inc*. Available at: <https://www.fireeye.com/blog/threat-research/2013/10/openioc-basics.html> (Accessed: 12 March 2019).
- Hevner, A. R. et al. (2004) 'Design science in information systems research', *MIS Quarterly*, 28(1), pp. 75–105.
- ITU (2018) *Definition of Cybersecurity - ITU-T x.1205*, International Telecommunication Union. Available at: <https://www.itu.int/en/ITU-T/studygroups/com17/Pages/cybersecurity.aspx> (Accessed: 1 July 2018).
- Mavroeidis, V. and Bromander, S. (2017) 'Cyber Threat Intelligence Model: An Evaluation of Taxonomies, Sharing Standards, and Ontologies within Cyber Threat Intelligence', in *2017 European Intelligence and Security Informatics Conference (EISIC)*. IEEE, pp. 91–98. doi: 10.1109/EISIC.2017.20.
- Mtsweni, J. et al. (2016) 'Development of a semantic-enabled cybersecurity threat intelligence sharing model', in *Proceedings of the 11th International Conference on Cyber Warfare and Security, ICCWS 2016*.
- Mtsweni, J. et al. (2016) 'Development of a Semantic-Enabled Cybersecurity Threat Intelligence Sharing Model', in *Proceedings of the 11th International Conference on Cyber Warfare & Security*. Boston, USA: Academic Conferences Limited, pp. 244–252.
- Mtsweni, J., Mutemwa, M. and Mkhonto, N. (2016) 'Development of a cyber-threat intelligence-sharing model from big data sources', *Journal of Information Warfare*, 15(3), pp. 56–68. Available at: <http://researchspace.csir.co.za/dspace/handle/10204/9342> (Accessed: 20 November 2017).
- OpenIOC (2015) *Open Indicators of Compromise*. Available at: <http://www.openioc.org/> (Accessed: 19 May 2015).
- Sekoia (2017) *InThreat, InThreat*. Available at: <https://inthreat.com/threatintelligence> (Accessed: 12 March 2019).
- Stillions, R. (2014) *The DML model*. Available at: [https://ryanstillions.blogspot.com/2014/04/the-dml-model\\_21.html](https://ryanstillions.blogspot.com/2014/04/the-dml-model_21.html) (Accessed: 12 March 2019).
- Swart, I. (2015) *Pro-active visualization of cyber security on a National Level: A South African Case Study*. Rhodes University. Available at: [https://www.researchgate.net/publication/305443420\\_Pro-active\\_visualization\\_of\\_cyber\\_security\\_on\\_a\\_National\\_Level\\_A\\_South\\_African\\_Case\\_Stud](https://www.researchgate.net/publication/305443420_Pro-active_visualization_of_cyber_security_on_a_National_Level_A_South_African_Case_Stud) (Accessed: 20 November 2017).

# Cyberpsychological Threat Intelligence

Julie Murphy and Anthony Keane

School of Informatics and Engineering, Technological University Dublin, Ireland

[julie.murphy@itb.ie](mailto:julie.murphy@itb.ie)

[anthony.keane@itb.ie](mailto:anthony.keane@itb.ie)

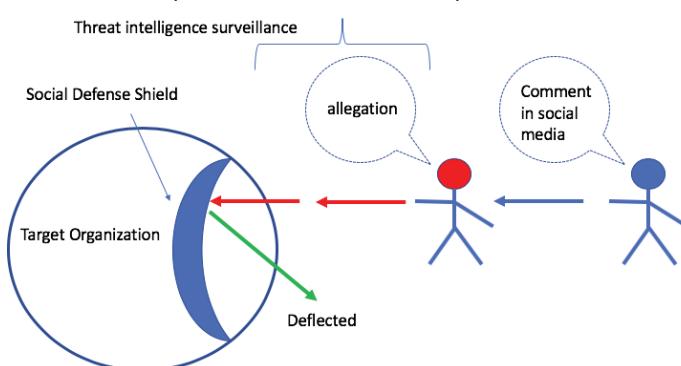
**Abstract:** Threat Intelligence is a core component in an effective Cyber Defensive Strategy. It allows operational defences to be strengthened in advance of emerging cyber-threats, mitigating the impact of potential damage. Social influence, disinformation and misinformation are new areas for threat intelligence and their impact can be seen in recent international controversies surrounding election tampering, civil unrest and journalistic integrity. Recent events suggest that influence attacks are impacting organisations on a broader scale, with rapid global impact. High volume and targeted online traffic add legitimacy that has influenced stock value, international sporting events and disrupted vaccine programmes. Social influence attacks are also a cybersecurity issue that needs to be understood with clear definitions, guidance and analysis of tactics, techniques and procedures, in order for the cybersecurity community to build appropriate defences and skillsets. These types of attacks can be classed as *Cyber-Psychological Attacks (CPA)* and are built on techniques from multiple disciplines. It is an IT issue based on the technologies used; it is a psychological issue based on techniques applied; and it is a marketing and design issue framing psychological techniques to achieve successful outcomes. In isolation, all of these elements are not new but when combined, they have the potential to create a powerful and effective weapon in cyber warfare. This paper presents a review of the effect that social influence attacks are having in the context of cyber-threat intelligence. We explore how important CPA techniques have become in the broad field of cybersecurity as we describe how online influence has been weaponised and deployed to significant effect. We conclude with a discussion on the importance of cyberpsychological threat intelligence in cyber defence and its limitations in the context of emerging threats.

**Keywords:** cyber threat intelligence, psychological operations, social influence, disinformation

## 1. Introduction

*Cyber threat intelligence (CTI)* is currently heavily reliant on technically managed infrastructure with the assumption that this approach can continually be effective for the mitigation of current and future cyber threats. Defences reliant on signature-based and pattern matching technologies were applicable to technology-based threats of previous generations yet are ill-prepared to defend against emerging complex multi-faceted threats. Technology-oriented approaches to CTI lack the key elements of opportunity, motivation and capability that is demanded for situational awareness, which can be a weakness in preparing for future attacks or knowing where best to deploy limited resources.

While technology-oriented approaches will certainly continue having a role to play, it does not account for the emergence of social media which can provide an alternative platform for contributing to the damage caused from an attack. These newer mediatic forms have become powerful tools and are a challenge for threat intelligence analysts in predicting an attack as they do not fall into any cyber defence frameworks or methodologies. As a result, analysts are overwhelmed with data derived from monitoring systems, which are not always useful against emerging threats that may not present on standard defence systems and thus requires a “social defense shield” to deflect the potential threat from weaponised social influence attacks (see figure 1).



**Figure 1:** Early warning through good threat intelligence can counter a social influence attack

## **2. Social Influence attacks**

Social media is a modern information influencing platform that is unregulated and beyond the control of any single organisation or Government. Social media is truly the platform of free speech and allows the global dissemination of information, regardless of its correctness, onto the personal devices of individuals thus having a very intimate contact with any messages or statements. This has created a new type of activity called “social influence” whereby people’s thoughts and actions may be controlled or influenced by what they view on social media, creating the opportunity for “social influence” to be the means of an attack that changes the outcome of a process or issue.

A social influence attack has the potential to disrupt business operations often using legitimate tools which blur the line of legality. Social influence attacks have been observed in high profile events like the Trump US Presidential election and Brexit referendum where the election result was not anticipated from early predictions in the campaigns. The reasons have been attributed to the manipulation of social media content that had a deeply personal influence on people that traditional news outlets do not normally achieve. The impact of these attacks is exacerbated when information is stolen and leaked to give more credence to the false argument being spread on social media.

Threats employing ‘multi-faceted disinformation operations’ are increasing (ENISA, 2017). In a departure from traditional cyber offensive methods such as reconnaissance, targeting and hacking; cyber troops mould public opinion (Bradshaw & Howard, 2017). Fake or compromised accounts and misuse of connected computers on the internet are characterised as weapons whereby, damage is caused to society and individuals (ENISA, 2018).

Threat actors are applying disinformation attacks in parallel with technological attacks to disrupt reliable sources of media information to create distrust in society (Hulcoop et al., 2017). Contemporary social media platforms offer a dramatic departure from traditional media technologies whereby content is distributed with little “third part filtering, fact-checking or editorial judgement” (Allcott and Gentzkow, 2017).

Psychological operations endeavour to win the ‘hearts and minds’ of a populace which requires skills from multiple fields. Volumes of research has been conducted on human response to triggers such as: authority conformity and groups, in addition to research on human computer interaction, and more recently cyberpsychology. When designing systems and platforms, user experience is key to success or failure. Captology has been described as this study of computers as persuasive technologies (Fogg et al., 2007).

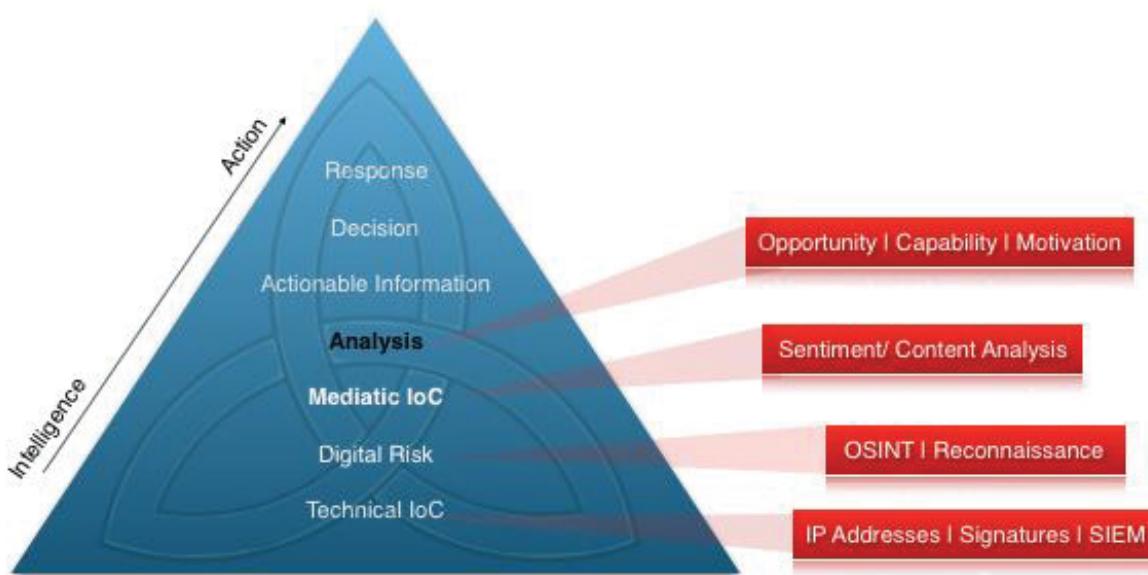
Marketers are acutely aware of the value of social listening, sentiment analysis and advertising to maximise return on investment by continually tailoring to their target audience. Respectively, these methods and theories are not new however combined, they provide a powerful medium (or weapon) to mould opinion for political ideals, health debates, tragic events or indeed any behaviour or opinion.

A global analysis of organised manipulation on social media has identified common techniques such as hashtag poisoning, posting positive/negative and neutral posts, targeting individuals and content creation to promote an agenda (Bradshaw and Howard, 2017). In a subsequent study, Bradshaw and Howard (2018) investigated data-driven ‘online manipulation campaigns’ in 48 countries, concluding that millions are being invested in computational propaganda to influence public opinion. Similar to reflexive control, the objective of psychological operations is to influence a targets behaviour to support an objective (Jaitner & Kantola, 2016), US Government, 2003). Nationalistic movements, ethnic tensions, conflict and political crises are just some observed results of campaigns designed to cause mistrust in society. Howard, Woolley and Calo (2018) discuss the use of bots and algorithms in the US presidential election which Persily (2017) describes as the *“revolutionary use of new media by the winning campaign”*.

Social media account takeovers (AccessNow 2017), tainted leaks (Hulcoop et al. 2017), fake news (Allcott and Gentzkow 2017), large-scale cyber operations (Kumar, 2018) and health debates (Broniatowski, Hilyard and Dredze, 2018) are merely a few examples where online influence has been demonstrated. As techniques are refined, this trend is likely to both increase and permeate to less high-profile targets across all spectrums. Cyberpsychological attacks are gaining popularity as a viable attack vector. CPA (in the context of cybersecurity) is an attack that applies psychological techniques to influence, manipulate or coerce a target audiences’ behaviour or psychological state using technological media, platforms and design.

Although social media analysis is presently employed by vendors for threat analysis, it is generally applied to identify content related to cyber threats and mentions. Similarly, Digital Risk Protection (DRP) is increasing in popularity with venture capital funding exceeding \$400 million in 2017 (Hayes, 2018). DRP is concerned with the online profile of an organisation using techniques such as open source intelligence and reconnaissance. The proposed additional layer adds a level of protection to mitigate threats that negatively impact an organisation, event or campaign to a target audience.

Traditional CTI is depicted in figure 2 as technical indicators of compromise (IoC). Digital risk adds to this intelligence by allowing security analysts to assess information publicly available to attackers that increases organisational risk. Mediatic or ‘influence’ attacks do not conform to standard security testing practices, as threats of this nature are perceived to be the remit of the digital marketing team despite the technologies used. The addition of a ‘mediatic’ layer will build on existing intelligence to understand how online traffic is impacting a target audience. This combined intelligence allows analysts to assess the viability of threats and CPA’s which results in actionable information, that supports decision making for an appropriate response. An effective response deflects the negative impact of an influence attack as shown in Figure 1.



**Figure 2:** Cyberpsychological intelligence model

### 3. The importance and limitations of cyberpsychological intelligence

Cyberpsychology relates to how we engage with technology (Barak, 1999). Kirwan (2016) describes three key aspects of cyberpsychology summarised as (i) the interaction with others via technical mediums, (ii) technological developments to accommodate desires and needs and (iii) the impact technology has on psychological states and behaviour.

The ‘online disinhibition effect’ was coined by (Suler, 2004) to explain how behaviour may differ online. Communication in recent years has shifted from sharing personal activities with a few to “vast digital footprints surrounding traits, interests, intentions and beliefs” (Acquisti et al. 2015). Cyber-Psychological Intelligence (CPI) is the analysis of an adversary’s motivation, capability and opportunity to induce change in a targets behaviour or psychological state. Refined data holds extensive value. Facebook reported profits in excess of \$4.2 billion in 2017 with 2.13 billion users (Facebook, 2018).

People are readily influenced about the information they disclose (Acquisti et al. 2015) comments that “what they share can be used to influence their emotions, thoughts, and behaviours in many aspects of their lives, as individuals, consumers, and citizens”. Social network information, when combined with the accurate data derived from electronic assessment result in highly descriptive patterns of behaviour (Harari et al. 2016). By analysing Facebook ‘likes’, Kosinski, Stillwell and Graepel (2013) determined that extremely sensitive information is accurately predictable such as ethnicity, religion, sexual orientation, happiness and intelligence. Such determinations may improve services and products tailored to individuals such as advertising, benefit

behavioural research in human psychology (Chen et al. 2017), or predict the considerable negative opportunity to manipulate based on mass digital footprints and the availability of data for potential misuse without consent.

Kosinski, Stillwell and Graepel (2013) convey the exacerbated persuasive power of social influence when employed with psychological indicators. A potential limitation to CPI is the difficulties aligning a target with an attack in progress however, as CPI indicators are refined, it is anticipated this will decrease over time.

A multi-disciplinary skillset is required to respond to contemporary attacks which is not the traditional educational path for security analysts. The classification of tactics, techniques and procedures however, combined with the establishment of a framework, will formalise the required training necessary to support future security analysts. The multiple disciplines cover issues from Information Technology to Psychology to Digital Marketing Influences. Cyber threat intelligence (CTI) should adopt the same approach as traditional intelligence and consider information from all perspectives to narrow the focus to support decision making.

#### **4. Conclusion**

Social influence, disinformation and misinformation are at the forefront of mind in light of widespread media coverage. While research is available on cyber threat intelligence, psychological methods and social influence respectively; there is limited research on online social influence in the context of cyber threat intelligence. Traditional approaches focus on technical skills and technologies despite emerging trends indicating an increase in mediatic threats. Personalised attacks have the power to induce actions and opinions without necessarily ‘breaching’ system defences. Cyber threat intelligence demands increased adoption of traditional ‘intelligence’ methods that encompass social, technical and human considerations to mitigate social influence attacks.

#### **Acknowledgements**

The authors wish to thank the Technological University Dublin and the staff from the Cyber Security Research Centre for the use of their facilities in the conducting of this research.

#### **References**

- Access Now (2017) *The “Doubleswitch” social media attack: a threat to advocates in Venezuela and worldwide*. Available at: <https://www.accessnow.org/doubleswitch-attack/> [Accessed: July 19 2017].
- Acquisti, A., Brandimarte, L. and Loewenstein, G. (2015) ‘Privacy and human behaviour in the age of information’ *SCIENCE*, 347(6221).
- Allcott, H. and Gentzkow, M. (2017) ‘Social media and fake news in the 2016 election’ *Journal of Economic Perspectives*, 31(2), pp. 211-236.
- Barak, A. (1999) ‘Psychological applications on the internet: A discipline on the threshold of a new millennium’. *Applied and Preventive Psychology*, 8, 231-246.
- Guetzkow (ed.), *Social Psychology*, 3<sup>rd</sup>. edn. pp. 174-82. New York: Hold, Rinehart & Winston
- Bradshaw, S. and Howard (2017) ‘Troops, trolls and troublemakers: A global inventory of organized social media manipulation’. The Computational Propaganda Project. University of Oxford.
- Bradshaw, S. and Howard, P. (2018) ‘Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation’. The Computational Propaganda Research Project. University of Oxford.
- Broniatowski, D.A., Hilyard, K.M., and Dredze, M. (2016) *Effective vaccine communication during the Disneyland measles outbreak*. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4903916/> [Accessed: 29 August 2018].
- Chen, L., Gong, T., Kosinski, M., Stillwell, D. and Davidson, R. (2017) ‘Building a profile of subjective well-being for social media users’, *PLoS One*, 12(11).
- ENISA (2017) *Disinformation operations in cyber-space*. Available at: <https://www.enisa.europa.eu/publications/info-notes/disinformation-operations-in-cyber-space> [Accessed: 4 January 2019].
- ENISA (2018) *Strengthening network & information security & protecting against online disinformation (“Fake News”)*. Available at: <https://www.enisa.europa.eu/news/enisa-news/strengthening-network-and-information-security-to-protect-the-eu-against-fake-news> [Accessed: 4th Jan 2019].
- Facebook (2018a) *Facebook reports fourth quarter and full year 2017 results*. Available at: <https://investor.fb.com/default.aspx?SectionId=5cc5ecae-6c48-4521-a1ad-480e593e4835&LanguageId=1&PressReleaseId=e20aa0a8-d91c-4331-a010-3fa495d52b86> [Accessed 4 March 2018]
- Fogg, B.J., Cuellar, G., Danielson, D. (2007) ‘Motivating, influencing and persuading users’, Sears, A. and Jacko, J. (eds) *The human-computer interaction handbook*. Boca Raton: CRC Press, 133-146
- Harari, G.M., Lane, N.D., Wang, R., Crosier, B.S., Campbell, A.T. and Gosling, S.D. (2016) ‘Using smartphones to collect behavioral data in psychological science’, *Perspectives on Psychological Science* 11(6), pp. 838-85.
- Hayes, N. (2018) Digital Risk Protection In 2018: New Vendors, New Leaders, New Wave. Available at: <https://go.forrester.com/blogs/blog-digital-risk-protection-drp-wave-18/> (Accessed: 5 March 2019)

***Julie Murphy and Anthony Keane***

- Howard, N. Woolley, S. and Calo, R. (2018) 'Algorithms, bots, and political communications the US 2016 election: The challenge of automated political communication for election law and administration', *Journal of Information Technology & Politics*. (15)2. pp. 81-93.
- Hulcoop, A., Scott-Railton, J. Tanchak, P. Brooks, M. & Deibert, R. (2017) *Tainted leaks, disinformation and phishing with a Russian nexus*. Available at: <https://citizenlab.ca/2017/05/tainted-leaks-disinformation-phish/> [Accessed: 9 January 2018].
- Jaitner, M.L. & Kantola, MAJ.H. (2016) 'Applying principles of reflexive control in information and cyber operations' *Journal of Information Warfare*. (15)27.
- Kosinski, M., Stillwell, D. and Graepel, T. (2013) 'Private traits and attributes are predictable from digital records of human behaviour', *Proceedings of the National Academy of Sciences*, 110 (15) 5802-5805
- Kumar, M. (2018) *12 Russian intelligence agents indicted for hacking DNC emails*. Available at: <https://thehackernews.com/2018/07/russian-dnc-hack-trump.html> (Accessed: July 15 2018).
- Kirwan, G. (2016) 'An introduction to cyberpsychology' in Connolly, I., Palmer, M., Barton, H. and Kirwan, G. (ed) *An introduction to cyberpsychology*. New York: Routledge
- Persily, N. (2016) 'Can democracy survive the internet', *Journal of Democracy*. (28)2 pp.63-76.
- Suler, J. (2004) 'The online disinhibition effect', *CyberPsychology & Behavior*, 7(3), pp. 321-326
- U.S. Government (2003) *Psychological operations tactics, techniques and procedures. FM 3-05.301 (FM 33-1-1)*. MCRP 3-40.6A

# Strategic Foresight and Resilience Through Cyber-Wargaming

David Ormrod<sup>1</sup> and Keith Scott<sup>2</sup>

<sup>1</sup>University of New South Wales, Australia

<sup>2</sup>De Montfort University, UK

[drdave@linux.com](mailto:drdave@linux.com) –

[jkscott@dmu.ac.uk](mailto:jkscott@dmu.ac.uk) –

**Abstract:** Cyber-capabilities provide nation and non-nation state actors, including criminal organisations and individuals, with the ability to project power and influence across borders and into critical infrastructure, corporate networks and military systems with relative anonymity and impunity. Employed on their own or as part of a broader influence activity, cyber-attacks can use vulnerabilities within networked and digitally-enabled systems to create opportunities to undertake a variety of malicious actions, including the theft of intellectual property or financial data, engage in aspects of hybrid warfare or undertake the destruction and/or disabling of physical property that is network connected. Traditionally, strategic and military planners have undertaken wargaming as a means of anticipating potential outcomes relating to system vulnerabilities and failures, as a means of optimizing a system of systems and increasing resilience. However, cyber-wargaming as a strategic planning activity has suffered conceptual and practical problems due to the disconnect between technological design and the conceptual models used for physical systems and critical infrastructure. Traditional concepts such as time, which have generally been easily represented within wargames, are much more difficult to represent in the cyber domain. The lack of suitable models has led to two different approaches; a focus on the operational and technical through red teaming and cyber exercises, or a focus on the strategic through executive table-top activities and matrix wargames. Cyber-wargaming is an iterative approach to optimizing the information security posture of an organisation, whilst simultaneously increasing the knowledge of the participants about their environment. Cyber-wargaming ensures the organisation evolves as a collective and has an opportunity to engage in a safe way with potential risks and threats. This paper proposes a unique cyber-wargaming model which seeks to achieve strategic foresight and increase the resilience of the system of systems. The model provides organisations and individuals with a way of understanding vulnerabilities across the systems of systems within cyber-space, in a way that facilitates understanding of the fundamental risks to an organisation. The cyber-wargaming model proposed by this paper will allow participants to reduce risk, enhance understanding and increase collaboration to address the fundamental socio-technical issues they must address to succeed. This unique approach extends on existing assurance programs and governance frameworks, by recognizing the role of the malicious actor, incorporating a view of the cyber-ecosystem and aligning strategic organizational imperatives with information and communication technology security programs.

---

**Keywords:** cyber security, strategic foresight, resilience, information operations, wargaming

## 1. Introduction

The fourth industrial revolution theory posits a rapidly changing, dynamically evolving technological age where Artificial Intelligence (AI), Machine Learning (ML), big data and robotics drive humanity into new areas of efficiency and effectiveness. This change has significant potential impacts on the ways and means by which society is organised and warfare is conducted. The increasing capabilities of nation and non-nation state actors to conduct operations in cyber space allow threats to operate and bypass national physical borders, potentially attacking government, industry and the public directly. As systems become more tightly coupled, the opportunity to create “cascading, compounding, and collateral effects” (United States Department of Defense 2018, p. xv) increases, reducing the effectiveness of traditional risk management techniques seeking to reduce the complexity of risk analysis through abstraction and treating systems as relatively independent with a few defined connections. Official warnings about the risks pertaining to national security and critical infrastructure resulting from cyber-attacks have been issued since 1997, but the Establishment response has been seen as slow, poorly resourced and unlikely to resolve current and future security challenges (Austin 2017; Lewis 2018).

The current trend of differentiating the information environment (Department of Defense 2016) from cyberspace (Kuehl 2009, p. 24) is not helpful, as the corresponding cyber-attack surface continues to expand at an exponential rate (Johnson 2012); the decoupling of people and cognition from cyberspace is counter-productive and counter-intuitive. The proliferation of networked devices and ubiquitous interconnectivity means that there is very little left of the information environment which is not potentially accessible. Of course, the power of cyberspace comes from these connections, but as the connections and capabilities across cyberspace continue to expand, the loosely bounded environment and complexity increases. As our dependence upon the cyber environment grows, so too do the potential consequences associated with failure or malicious compromise (Ormrod and Turnbull 2016, p. 271).

## 2. Cyber-Attacks and resilience

Organisations “should expect cyber-attacks to be present for all critical missions...” (Gilmore 2016, p. 389) and assume all systems are compromised (‘Assume Breach’). The theory of cyber-attack mechanics requires that an attack imposes a threat upon a vulnerability by a threat agent (Stephenson and Prueitt 2005). A successful attack results in a cyber-incident. Models have been developed to explain the cyber-attack process; a summary of the main models, their key features, and their limitations is given in Table 01 below.

**Table 1:** Key features and limitations of state of the art cyber-attack and red teaming models

MODEL	KEY FEATURES	LIMITATIONS
a. Cyber-Kill Chain™ (Lockheed Martin)	<ul style="list-style-type: none"> <li>1. seven-step process to “target and engage an adversary to create desired effects” (Hutchins et al. 2011, p. 4)</li> <li>2. 7-phase process: Reconnaissance, Weaponisation, Delivery, Exploitation, Installation, C2, and Actions on Objectives</li> <li>3. used to support course of action development for defenders (Hutchins et al. 2011)</li> </ul>	<ul style="list-style-type: none"> <li>1. reflects perimeter-based defensive thinking with a focus on intrusion/malware (Clopert 2009).</li> <li>2. Relates to military concept of Effects Based Operations (EBO); opposed by factions of the military, seen by some as a passing fad (Mattis 2008, p. 22).</li> </ul>
b. Cyber Prep (Bodeau et al. 2010)	<ul style="list-style-type: none"> <li>1. compares threat capabilities with organisational cyber preparedness</li> <li>2. incorporates threat actor into assessment</li> <li>3. assumption of breach by adversary and prolonged presence in the system is balanced with organisational objectives, priorities and risk (Cyber Resilience and Response Team 2018).</li> </ul>	<ul style="list-style-type: none"> <li>1. Reinforces idea of system as fortress to be defended</li> <li>2. contradicts concept of cyber-resilience</li> </ul>
c. Mandiant Attack Lifecycle Model	<ul style="list-style-type: none"> <li>1. Similar to (a), but more holistic</li> <li>2. Broader focus than (a)</li> <li>3. 6 stages: Initial Compromise, Establish Foothold, Escalate Privileges, Internal Reconnaissance, Move Laterally and Maintain Presence (Mandiant 2015)</li> </ul>	<ul style="list-style-type: none"> <li>1. lacks clear objectives or mission focus</li> </ul>
d. MITRE ATT&CK Matrix and Framework	<ul style="list-style-type: none"> <li>1. supports detailed, prioritised investment approach for organisations</li> <li>2. uses previously observed cyber-attack methods to evaluate weaknesses in networks (MITRE 2019c).</li> <li>3. Cyber-security vendor products also evaluated, employing adversary emulation (MITRE 2019b)</li> </ul>	<ul style="list-style-type: none"> <li>significant resource impost. engineering expertise required.</li> </ul>
e. Mission Assurance Engineering (MAE)	<ul style="list-style-type: none"> <li>1. undertakes Crown Jewels Analysis (CJA), Threat Susceptibility Assessment (TSA) and Cyber Risk Remediation Analysis (RRA) as a repeatable process to build resilience.</li> <li>2. CJA produces dependency maps, supporting understanding of CJ and their relationships to assets, tasks and missions.</li> <li>3. offers higher level risk management approach linking mission to assets</li> <li>4. allows the prioritisation of effort</li> </ul>	<ul style="list-style-type: none"> <li>1. likely to focus on tightly coupled systems, without fully considering looser relationships in complex organisations/systems, potentially leading to systemic failures.</li> <li>2. neglects nature of attackers: “security threats come from living, breathing opponents who are creative, knowledgeable, collaborative, and determined [...] they are outside the box” (Casey and Willis 2008).</li> </ul>
f. Red teaming: Sandia National Laboratories Information Design Assurance Red Team (IDART) (Sandia 2014)	<ul style="list-style-type: none"> <li>1. a broader approach, considering various forms of cyber-attacks on organisations</li> </ul>	<ul style="list-style-type: none"> <li>1. fails to progress beyond ‘Analyse’ and ‘Report’ components of the attack process;</li> <li>2. does not describe development of measures/modelling of the full attack process is not described.</li> <li>3. live/active assessments cannot be conducted as an attacker would do.</li> </ul>

Models (a) – (d), particularly the MITRE ATT&CK Framework, are focused on a fundamental technical layer of the problem space. While valuable, this does not resolve the broader strategic issues facing organisations. From a cyber-wargaming perspective, modelling a cyber-attack is critical to understanding the threat that a defender must counter. However, it does not resolve the larger problems dealing with the attack, allocating limited

organisational resources to identify and defend critical assets ('crown jewels' (MITRE 2019a)), or responding effectively and increasing organisational resilience to cyber and cyber-physical threats. Models (e) and (f) add further levels of analysis, but none offer a schema for examining cyber-attack in anything approaching its full complexity.

### **3. Strategic foresight and wargaming**

The challenge of understanding future cyber-threats can be addressed multiple ways. Scenario thinking is a strategic reasoning approach seeking to utilise historical knowledge whilst embracing the intuition of planners. This provides a degree of foresight in dynamic organisational environments by hypothesising possible futures (MacKay and McKiernan 2004; Van der Heijden 2005). Scenario analysis must be detailed and sophisticated, "so that our responses may fit the ambiguities of our information and minimise the risks both of error and inaction" (Wohlstetter 1965, p. 41). Threatcasting differs in detail and process but has similarities to scenario thinking, involving a diverse stakeholder group postulating potential future cyber threat events and discussing the specific actions that can be taken to address them (backcasting) (Johnson 2012). Scenario thinking and threatcasting produce models used by decision makers to understand their threat environment, assess risk and develop mitigations. Modelling is the "...purposeful abstraction and simplification of the perception of a real or imagined system with the intention to solve a sponsor's problem or to answer a research question" (Tolk 2012). If these hypotheses are proven valid, they become theory (Tolk 2013). Modelling is also a process of transition from micro-knowledge to macro-knowledge through observation of the overall system behaviour emerging from its individual actors and rule sets (Drogoul et al. 2003). All models are however incomplete, in that they are only simplified approximations of the real world (Box 1976).

Wargaming is another way of using models as an abstraction, to better understand and potentially predict future challenges. It offers a "decision-making technique that provides structured but intellectually liberating safe-to-fail environments to help explore what works (winning/succeeding) and what does not (losing/failing), typically at relatively low cost." (United Kingdom Ministry of Defence 2017, p. 5). Wargaming also provides a variety of toolsets as part of a broad approach, including simulation, analytical tools and the critical analysis of historic events. However, statistically significant historical data suitable for analysing cyber-attacks is simply not available outside very specific use cases. The reliance on experience is common to both analytic wargames and human-in-the-loop simulations. Wargames introduce the potential for bias and personality to influence experimental results, in a way that cannot be easily quantified or documented; even wargames using computer models and simulations are subject to the biases of their design and coding, although transparency and significant validation may reduce this bias.

Wargaming events can change the organizational culture and educate stakeholders, enabling them to "overcome cognitive barriers, challenge mental models, detect weak signals of change in an organisational environment, re-direct attention in an organisation, and assist an organisation in developing foresight" (Schwarz 2009, p. 291). The development of foresight within an organisation seeks to create self-renewal through the shaping of strategy, exploration and detection of adversaries (Ringland 2010). Strategic foresight activities create value through organizational learning and the understanding of change (Rohrbeck and Schwarz 2013). Scenarios do not explicitly forecast the future, but instead provide a variety of potential futures which can "minimize surprises and broaden the span of managers' thinking about different possibilities" (Reger and Mietzner 2005).

The communication of risk is just as important as the technical understanding of risk. A more holistic model and risk assessment framework is needed that provides clear risk metrics and links between business value and cyber dependencies. Cyber-Wargaming provides one method of communicating cyber-risk, coupled with an alignment model linking business imperatives, organisational value propositions, risk assessments, mitigations and threats.

### **4. Cyber-Wargaming, risk and assurance**

Wargames "enable active learning: players are confronted with continuous and often unexpected questions and challenges as they explore, experiment and compete within the artificial model the game provides" (United Kingdom Ministry of Defence 2017, p. 11). The use of cyber-wargame tools and methods has a history going back to the emergence of cyber-attacks and cyber training systems; the United States Naval War College Information Warfare Analysis and Research (IWAR) Laboratory creation in 1999 and subsequent development of a Cyber Defense Exercise in 2001 is but one example (Schepens et al. 2002). Subsequent developments include

experiments on active defence techniques and deception measures (Heckman et al. 2013), to the point that ‘Mayhem’ has been able to autonomously analyse system and binary code, detect vulnerabilities and issue patches (Brumley 2019).

The industry trend to address cyber-risks has been to develop technical solutions in parallel with governance and assurance programs. Enterprise governance of information and technology (EGIT) seeks to create “value delivery from digital transformation and the mitigation of business risk that results from digital transformation” (ISACA 2019, p. 11). The governance system principles described within the Control Objectives for Information and Related Technologies (COBIT) cannot be instituted simply through an assurance program. They require active, ongoing organisation-wide efforts to continually push forward an agenda of change and evolution, to match the pressures of the environment. A cyber-wargaming framework would need to incorporate each of the COBIT governance principles as a component of a broad approach.

Various security governance and assurance programs amplified within publications such as the Australian Government Information Security Manual (ISM) (Australian Cyber Security Centre 2019), the UK National Cyber Security Centre (NCSC) Cyber Assessment Framework (CAF) (National Cyber Security Centre 2019) and the National Institute of Standards and Technology (NIST) Special Publications (National Institute of Standards and Technology 2018) have progressively evolved into risk frameworks. Frameworks, such as the NIST Risk Management Framework (RMF) recognise that a ‘one size fits all’ approach is unsustainable and unfeasible across the range of industries and stakeholders (National Institute of Standards and Technology 2018, p. ii). NIST provides three levels for an organisation-wide risk management approach (National Institute of Standards and Technology 2018, p. 6): (1) Organisation; (2) Mission/business process; (3) Information system. The development of a mature cyber-wargaming framework should work within an RMF and address these different organizational levels as specific problem sets, in addition to seamlessly integrating each level with the other.

Whilst abstraction is necessary and reduces complexity (applying Occam’s Razor), models must adequately represent the potential states of the environment. Failure to adequately address the requisite variety of a system prevents regulation, because the controlling system does not have sufficient complexity to address the disturbances in the system (Ashby 1991). Cyber-wargaming can provide an agile and complex regulation system, within an abstracted model of the environment. Ashby’s Law is an important consideration when designing the wargame problem and developing the extent of abstraction applied to the scenario. Equally, the conduct of wargaming feeds back into the organizational system and improves the actual regulation potential of the organisation’s people and processes. Wargaming provides an experiential learning experience, through structured debriefing to support self-reasoning and reflective practice. Deep learning occurs through active experiences and the conceptualisation of broader issues (Kilgour et al. 2015), developing the response reflexes and potential states the organisation can manage in the future.

Realism is vital to achieve useful learning outcomes for participants. The Realistic-Environment, Adversary, Communication, Tactics, and Roles (R-EACTR) model supports the development of dynamic, adaptable, and realistic wargames which provide relevance to participants (Sullivan et al. 2018). For cyber-wargames, this may include the requirements for customised threat intelligence models, as well as realistic modelling of relationships with peer organisations, supply chains links, and government security organisations. The concepts of organizational and business value chains (Porter 1985) should be wargamed as they apply to the specific scenario and problem set, as should the relevant processes supporting the generation of business value. The modelling of systems of systems and their interdependencies and interactions is equally important as individual processes, particularly where they constrain the production of value to the organisation (Goldratt 1990). Support from vendors may be needed/modelled if they are likely to provide service continuity during an event or provide threat response and threat hunt capabilities after an incident. For example, decisions on where to place system taps for data analysis can be contextual and may alter based on the specific scenario. However, the process of review and analysis, as well as the discovery of configuration issues, is a valuable activity to prepare for a real event. Transactional business relationships can evolve, through activities such as wargaming coupled with other governance mechanisms, into collaborative learning opportunities and opportunities to share information, intelligence and increase situational intelligence.

## 5. The Cyber-Wargaming model

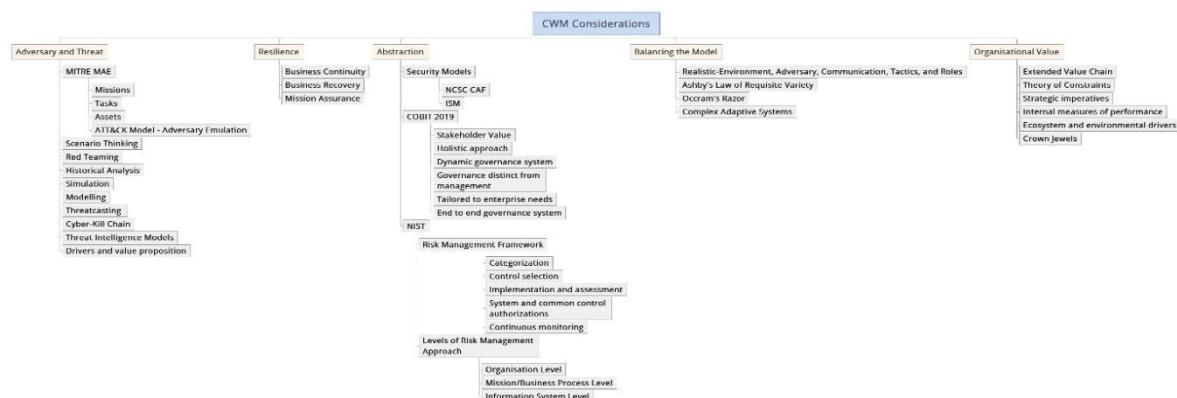
The Cyber-Wargaming Model (CWM) described here is an iterative approach to optimizing the information security posture of an organisation, whilst simultaneously increasing participants' knowledge of their environment. Cyber-wargaming ensures the organisation evolves collectively and has an opportunity to engage safely with potential risks and threats. This paper proposes a unique cyber-wargaming model which seeks to achieve strategic foresight and increase the resilience of the system of systems. The model provides organisations/individuals with a way of understanding vulnerabilities across the systems of systems within cyber-space, facilitating understanding of the fundamental risks to an organisation. Cyber-wargaming as proposed here will allow participants to reduce risk, enhance understanding and increase collaboration to address the fundamental socio-technical issues they must address to succeed. This unique approach extends existing assurance programs and governance frameworks, by recognizing the role of the malicious actor, incorporating a view of the cyber-ecosystem and aligning strategic organizational imperatives with information and communication technology security programs.

### The CWM consists of six parts:

- Part One. CWM Considerations;
- Part Two. CWM Process A: Threat vs System Outcome Model;
- Part Three. CWM Process B: Business Process Coupling Modelling;
- Part Four. CWM Process C: System Architecture;
- Part Five. Causal Connection; and
- Part Six. CWM Adjudication.

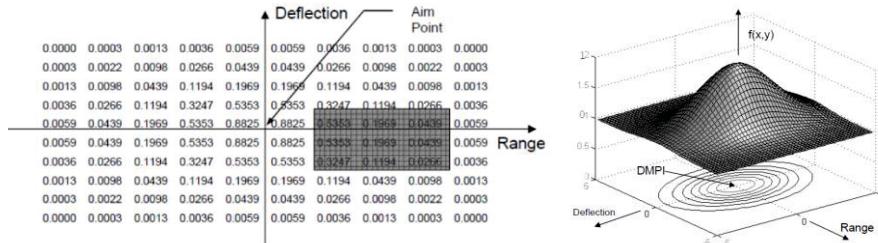
**Part One, Considerations.** The CWM Considerations presented in figure 1 below, unites key principles from a variety of frameworks to provide a hybrid approach. The CWM is a start point for organizational cyber-wargamers to develop their own approaches, using best practice, to enhance governance and risk management approaches. The CWM includes the following sub-sections:

- Adversary and Threat – methods to understand the environment, relevant threat sources and develop scenarios aligned with the MITRE MAE, red teaming and threatcasting approaches.
- Resilience – testing business continuity, recovery and mission assurance across the organisation and key dependencies.
- Abstraction – key resources to support the development of relevant security models, application of the COBIT 2019 principles to the wargame and application of the RMF across all levels of the organisation (including their interactions across levels).
- Balancing the Model – ensuring it is appropriate, realistic and includes sufficient complexity to represent the environment and provide useful lessons.
- Organisational Value – applies a variety of business models such as the value chain and theory of constraints to align the strategic imperatives of the organisation with the specific behavioural drivers of the real system. Includes the Crown Jewels Analysis of the MITRE MAE.



**Figure 1:** The proposed cyber-wargaming model considerations

**Part Two, Process A.** Figure 2 depicts indicative weapon effectiveness models, of the type used in military wargames and simulations. These models include the Joint Munition Effectiveness Manuals (JMEM), monte carlo simulations and a variety of simulation systems applying algorithms such as the Cookie Cutter and Carleton Damage Functions (Anderson, 2004). These well- developed models remain abstractions of reality, but they have well known and researched advantages and disadvantages. This knowledge informs a well-trained community of practice. The application of Cyber Joint Munitions Effectiveness Manuals (CJMEM) has been advocated by Gallagher and Horta (2013). However, their approach does not allow users to understand the underlying model, or assist in the learning process by identifying key aspects of the simulated environment that could change the outcome and mitigate risk.

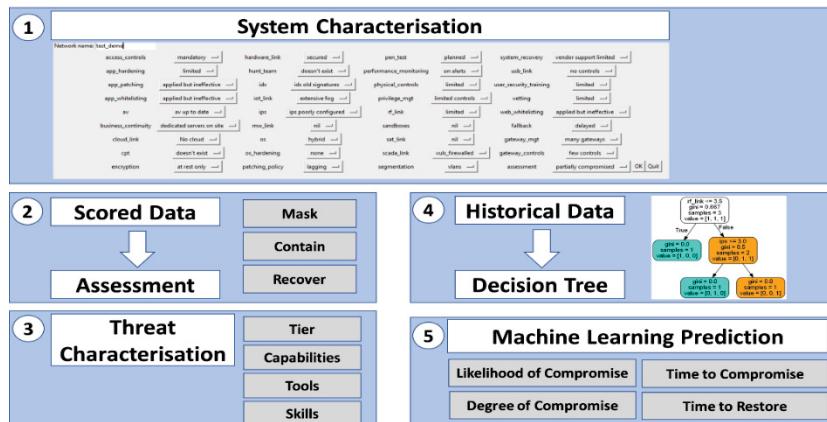


**Figure 2:** Example weapon effectiveness models as used in military wargames and simulations

Figure 3 depicts the first of the prototype wargaming processes. The Threat vs System Outcome Model is used as part of the analysis of the Information System level (3) of the NIST RMF. This process is intended to bridge the gap currently left by the lack of CJMEM, using a five-step process:

- 1. characterising a system based on key system properties, related to known assurance and risk factors, ranging from the resources available to defend the network and key architectural factors through to cultural and training issues;
- 2. scoring the characterisation across three areas; mask, contain and recover. Mask refers to the ability of a defender to mask a vulnerability from adversarial detection; Contain refers to the prevention of a successful cyber-attack from pivoting across a network or corrupting the wider system resources; and Recover is the ability for a defender to detect and remove an adversarial presence or capability from their system;
- 3. characterising the threat across factors such as threat tier, capabilities, tools and skills;
- 4. generating a decision tree, using training data which is currently provided through expert knowledge but in time is intended to incorporate historical data, such as audit results; and
- 5. using a decision tree classifier prediction in scikit, predicting a new network or system's likelihood of compromise, degree of compromise, time to compromise and time to restore.

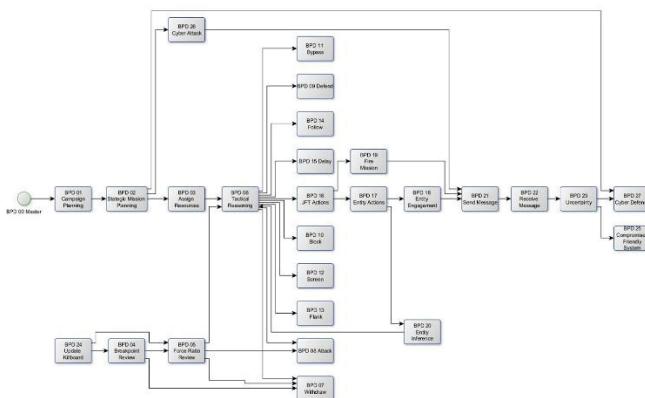
This approach remains in a prototype, development phase to test the utility of the approach. The characteristics and factors used require refinement as the CWM develops.



**Figure 3:** CWM prototype wargaming process A: Threat vs system outcome model

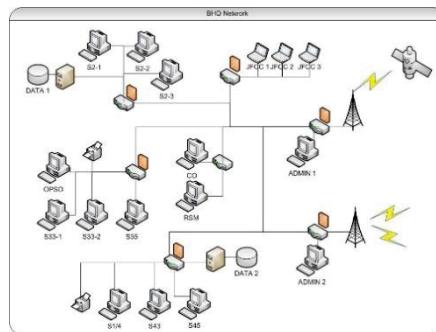
**Part Three, Process B.** Figure 4 depicts the first of the prototype wargaming processes, which takes a series of business processes relevant to the context of the wargame and models them. The model is used to examine couplings between business processes, which in many organisations provide the key value delivery. A cyber

attack that does not impact business processes may not even harm an organisation. Therefore, business processes must be modelled. This modelling includes process coupling, crown jewel data and key digital hardware such as SCADA enabling the processes.



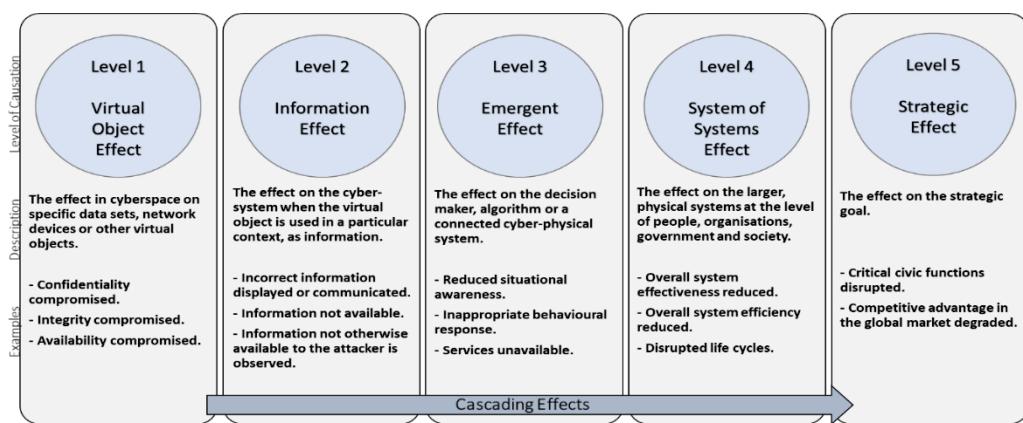
**Figure 4:** Prototype wargaming process B: Business process coupling modelling

**Part Four, Process C.** Figure 5 depicts the architecture normally mapped for a cyber-system. This can be performed using a variety of toolsets. In the existing prototype model only a few assets are modelled, but the process is highly scalable and adaptable, depending on the use case.



**Figure 5:** Prototype wargaming process C: System architecture

**Part Five, Causal Connection.** Figure 6 depicts the causal connection process used to model a cyber-attack's impact on a larger system of systems. A map of the links between level 4 and 5 systems and their outputs/objectives is used to determine the level of causal impact likely to occur within and between high level systems. This informs adjudication in part six.



**Figure 6:** Prototype wargaming causal connection

**Part Six.** Adjudication involves the results from parts one to five, applied across the various system of systems that occur within the context of the wargame scenario (even the adversary). These results currently require human adjudication, but the degree of machine assistance is planned to increase as training data sets and models mature over time. Adjudication remains under development, but is intended to include a process for adjudication, including measures of effectiveness and the management of time and the temporal aspects of

cyber-security. This process remains future work that the authors will develop into a complete methodology with supporting technologies, tools and processes as coherent extension of the model presented within this paper.

## **6. Conclusion**

The CWM proposed within this paper provides a brief, introductory model for future expansion. Several critical elements have been identified for further analysis. The authors intend to build upon this early model through the creation of a 'playbook' for cyber-wargaming. Cyber-wargaming as a strategic planning activity has suffered conceptual and practical problems due to the disconnect between technological design and the conceptual models used for physical systems and critical infrastructure. The CWM seeks to achieve strategic foresight and increase the resilience of the system of systems. The model provides organisations and individuals with a way of building a wargame enhancing individual participant and organisational understanding of vulnerabilities across the systems of systems within cyber-space, facilitating analysis of the fundamental risks to an organisation. The cyber-wargaming model proposed here will allow participants to reduce risk, enhance understanding and increase collaboration to address the fundamental socio-technical issues they must address to succeed. This unique approach extends existing assurance programs and governance frameworks, by recognizing the role of the malicious actor, incorporating a view of the cyber-ecosystem and aligning strategic organizational imperatives with information and communication technology security programs. Future work is planned to further develop existing prototypes, expand the adjudication process and mature the CWM into a useful assistant for those planning cyber-wargames at a strategic level.

## **References**

- Anderson, C.M. 2004. "Generalised Weapon Effectiveness Modeling," Thesis. Naval Postgraduate School. DTIC Reference A424672. Monterey, California.
- Andreessen, M. 2011. "Why Software Is Eating the World," *The Wall Street Journal* (20:2011), p. C2.
- Ashby, W. R. 1991. "Requisite Variety and Its Implications for the Control of Complex Systems," in *Facets of Systems Science*, G.J. Klir (ed.). Boston, MA: Springer US, pp. 405-417.
- Austin, G. 2017. "Are Australia's Responses to Cyber Security Adequate?," in *Australia's Place in the World*. Melbourne, Australia: Committee for Economic Development of Australia.
- Australian Cyber Security Centre. 2019. *Australian Government Information Security Manual*.
- Box, G. E. 1976. "Science and Statistics," *Journal of the American Statistical Association*. Alexandria, VA (71:356), pp. 791-799.
- Brumley, D. 2019. "Mayhem, the Machine That Finds Software Vulnerabilities, Then Patches Them." Retrieved 30 Jan, 2019, from <https://spectrum.ieee.org/computing/software/mayhem-the-machine-that-finds-software-vulnerabilities-then-patches-them>
- Casey, T., and Willis, B. 2008. "Wargames: Serious Play That Tests Enterprise Security Assumptions." from <https://www.sbs.ox.ac.uk/cybersecurity-capacity/system/files/Intel%20-%20Wargames-%20Serious%20Play%20that%20Tests%20Enterprise%20Security%20Assumptions.pdf>
- Cloppert, M. 2009. "Security Intelligence: Attacking the Cyber Kill Chain," *SANS Computer Forensics*.
- Cyber Resilience and Response Team. 2018. *Cyber Resilience and Response. 2018 Public-Private Analytic Exchange Program*.
- Davis, P. K., and Blumenthal, D. 1991. "The Base of Sand Problem: A White Paper on the State of Military Combat Modeling," RAND Corporation, Santa Monica, CA.
- Department of Defense. 2016. "Department of Defense Strategy for Operations in the Information Environment."
- Drogoul, A., Vanbergue, D., and Meurisse, T. 2003. "Multi-Agent Based Simulation: Where Are the Agents?," in: *Multi-Agent-Based Simulation II: Third International Workshop, MABS 2002*, J. Simão Sichman, F. Bousquet and P. Davidsson (eds.). Bologna, Italy. 15-16 July 2002: Springer Berlin Heidelberg, pp. 1-15.
- Epstein, J. M. 1999. *Agent-Based Computational Models and Generative Social Science*. Washington DC, USA: John Wiley and Sons.
- Gallagher, M., & Horta, M. 2013. Cyber Joint Munitions Effectiveness Manual (JMEM). *American Intelligence Journal*, 31(1), 73-81, from <http://www.jstor.org/stable/26202045>
- Gilmore, J. M. 2016. "Director, Operational Test and Evaluation Fy 2015 Annual Report - Cybersecurity," O.T.a.E. The Office of the Director (ed.). United States Department of Defense.
- Goldratt, E. M. 1990. *Theory of Constraints*. North River Croton-on-Hudson.
- Heckman, K. E., Walsh, M. J., Stech, F. J., O'Boyle, T. A., DiCato, S. R., and Herber, A. F. 2013. "Active Cyber Defense with Denial and Deception: A cyber-Wargame Experiment," *Computers & Security* (37), pp. 72-77.
- Hutchins, E. M., Cloppert, M. J., and Amin, R. M. 2011. "Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains," *Leading Issues in Information Warfare & Security Research* (1), p. 80.

- ISACA. 2019. *Cobit 2019 Framework: Introduction and Methodology*.
- Johnson, B. 2012. *Threatcasting 2026: A Widening Attack Plain*. West Point, USA.: Army Cyber Institute,.
- Kilgour, P. W., Reynaud, D., Northcote, M. T., and Shields, M. 2015. "Role-Playing as a Tool to Facilitate Learning, Self Reflection and Social Awareness in Teacher Education," *This article was originally published as: Kilgour, P., Reynaud, D., Northcote, MT, & Shields, M.(2015). Role-playing as a tool to facilitate learning, self-reflection and social awareness in teacher education. International Journal of Innovative Interdisciplinary Research, 2 (4), 8-20. Retrieved from http://www. auamii. com/jiir/Vol-02/issue-04/2Kilgour. pdf ISSN: 1839-9053*).
- Kuehl, D. T. 2009. "From Cyberspace to Cyberpower: Defining the Problem," *Cyberpower and national security* (30).
- Lewis, J. 2018. *Rethinking Cybersecurity - Strategy, Mass Effect and States*. Center for Strategic and International Studies.
- MacKay, R. B., and McKiernan, P. 2004. "The Role of Hindsight in Foresight: Refining Strategic Reasoning," *Futures* (36:2), pp. 161-179.
- Mandiant. 2015. "Mandiant Apt1. Exposing One of China's Cyber Espionage Units. Appendix B: Apt and the Attack Lifecycle. <Https://Www.Fireeye.Com/Content/Dam/Fireeye-Www/Services/Pdfs/Mandiant-Apt1-Report.Pdf>"),
- Mattis, J. N. 2008. "Usjfcom Commander's Guidance for Effects-Based Operations," DTIC Document.
- MITRE. 2019a. "Crown Jewels Analysis." from <https://www.mitre.org/publications/systems-engineering-guide/enterprise-engineering/systems-engineering-for-mission-assurance/crown-jewels-analysis>
- MITRE. 2019b. "Methodology Overview: Adversary Emulation Approach." Retrieved 16 Jan, 2019, from <https://attackevals.mitre.org/methodology/>
- MITRE. 2019c. "Mitre Att&Ck™." from <https://attack.mitre.org/>
- Nagge, J. W. 1932. "Regarding the Law of Parsimony," *The Pedagogical Seminary and Journal of Genetic Psychology* (41:2), pp. 492-494.
- National Cyber Security Centre. 2019. "Ncsc Published Guidance." Retrieved 16 Jan, 2019, from <https://www.ncsc.gov.uk/index/guidance>
- National Institute of Standards and Technology. 2018. "Sp 800-37 Rev. 2: Risk Management Framework for Information Systems and Organizations: A System Life Cycle Approach for Security and Privacy." from <https://csrc.nist.gov/publications/detail/sp/800-37/rev-2/final>
- Ormrod, D., and Turnbull, B. 2016. "The Cyber Conceptual Framework for Developing Military Doctrine," *Defence Studies*, (16:3), pp. 270-298.
- Porter, M. 1985. "Value Chain," *The Value Chain and Competitive advantage: creating and sustaining superior performance*.
- Reger, G., and Mietzner, D. 2005. "Advantages and Disadvantages of Scenario Approaches for Strategic Foresight," *Int. J. Technology Intelligence and Planning*, (Vol. 1, No. 2.).
- Reidenberg, J. R. 1997. "Lex Informatica: The Formulation of Information Policy Rules through Technology," *Tex. L. Rev.* (76), p. 553.
- Ringland, G. 2010. "The Role of Scenarios in Strategic Foresight," *Technological Forecasting and Social Change* (77:9), p. 1493.
- Robinson, S. 2013. "Conceptual Modeling for Simulation," in: *Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World*. IEEE Press, pp. 377-388.
- Robinson, S. B. 2009. "A Modeling Process to Understand Complex System Architectures," in: *School of Aerospace Engineering*. Atlanta, GA. USA.: Georgia Institute of Technology.
- Rohrbeck, R., and Schwarz, J. O. 2013. "The Value Contribution of Strategic Foresight: Insights from an Empirical Study of Large European Companies," *Technological Forecasting and Social Change* (80:8), pp. 1593-1606.
- Schepens, W. J., Ragsdale, D. J., Surdu, J. R., Schafer, J., and New Port, R. 2002. "The Cyber Defense Exercise: An Evaluation of the Effectiveness of Information Assurance Education," *The Journal of Information Security* (1:2), pp. 1-14.
- Schwarz, J. 2009. "Business Wargaming: Developing Foresight within a Strategic Simulation Au - Schwarz, Jan Oliver," *Technology Analysis & Strategic Management* (21:3), pp. 291-305.
- Sober, E. 1981. "The Principle of Parsimony," *British Journal for the Philosophy of Science*), pp. 145-156.
- Stephenson, P. R., and Prueitt, P. S. 2005. "Towards a Theory of Cyber Attack Mechanics," IFIP wg.
- Sullivan, D., Colbert, E., Kott, A., Osterritter, L., and Dobson, G. 2018. "Best Practices for Designing and Conducting Cyber-Physical System Wargames," *International Conference on Cyber Warfare and Security: Academic Conferences International Limited*, pp. 651-XVI.
- Tolk, A. 2012. "Challenges of Combat Modeling and Distributed Simulation," in *Engineering Principles of Combat Modeling and Distributed Simulation*, E.M.a.S. Engineering (ed.). New Jersey, USA: John Wiley & Sons, Inc., pp. 1-22.
- Tolk, A. 2013. *Ontology, Epistemology, and Teleology for Modeling and Simulation*. Berlin, Heidelberg: Springer.
- United Kingdom Ministry of Defence. 2017. "Wargaming Handbook." Retrieved 18 Jan, 2018, from [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/641040/doctrine\\_uk\\_wargaming\\_handbook.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/641040/doctrine_uk_wargaming_handbook.pdf)
- United States Department of Defense. 2018. *Joint Publication (Jp) 3-12 Cyberspace Operations*.
- Van der Heijden, K. 2005. *Scenarios: The Art of Strategic Conversation*. Chichester, UK: Wiley.
- Wohlstetter, R. 1965. *Cuba and Pearl Harbour: Hindsight and Foresight*. Santa Monica, CA.: RAND Corporation for the Office of the Assistant Secretary of Defense International Security Affairs.

# The Persuasion Game: Developing a Serious Game Based Model for Information Warfare and Influence Studies

David Ormrod<sup>1</sup>, Keith Scott<sup>2</sup>, Lynn Scheinman<sup>3</sup>, Thorsten Kodalle<sup>4</sup>, Char Sample<sup>5</sup> and Benjamin Turnbull<sup>1</sup>

<sup>1</sup>University of New South Wales, Australia

<sup>2</sup>De Montfort University, UK

<sup>3</sup>SAP

<sup>4</sup>Command and Staff College of the German Armed Forces, Germany

<sup>5</sup>ICF

[drdave@linux.com](mailto:drdave@linux.com)

[jkscott@dmu.ac.uk](mailto:jkscott@dmu.ac.uk)

[lynn.scheinman@sap.com](mailto:lynn.scheinman@sap.com)

[thorstenkodalle@bundeswehr.org](mailto:thorstenkodalle@bundeswehr.org)

[charsample50@gmail.com](mailto:charsample50@gmail.com)

[benjamin.turnbull@unsw.edu.au](mailto:benjamin.turnbull@unsw.edu.au)

**Abstract:** In an age of hybrid, asymmetric, and non-linear conflict, the role of Information Operations has become ever more important; this paper presents a study of a recent research project. The project examined ways of better enabling stakeholders to respond to the increasing use of influence in warfare, hybrid conflict, competition, and the realms of hard and soft politics. An international and cross-sector research group drawing on military, government, and academic expertise from seven different countries met in October 2018 to understand the best way to wargame influence. In the space of four weeks, the group worked towards the successful achievement of their initial goal; the creation of an influence wargaming community supported by a modular wargaming package and development roadmap. This paper introduces the context which has led to the establishment of the multi-national, multi-disciplinary team; discusses the reasons for employing serious gaming as a research tool for studying influence; outlines the development of the project of its initial four-week span; and summarises the initial key findings and directions for further research. The use of wargaming as a training and research tool is familiar in both the military and civil contexts; the project discussed here presents a truly innovative approach to influence studies, and shows the benefits of an interdisciplinary, cross-domain research team. The final section introduces a new influence wargaming framework that has emerged from the study.

**Keywords:** cyber security, information operations, influence, serious gaming, wargaming

---

## 1. Introduction

The strategic environment consists of “continuous confrontation with the potential for persistent conflict in which major state powers are constrained (both domestically and internationally) from decisive military action and rapid resolution” (US Department of Defense, 2008, p. 11). Unity of effort across the diplomatic, information, military and economic spectrums is a potentially critical capability, in an age of competition between nation states and non-linear conflict. However, unity of effort requires interagency campaign design through a strategic art construct, which does not currently exist in the Western military paradigm (US Department of Defense, 2008, p. 93). Evidence-based policy development approaches are needed to promote peaceful, practical and sustainable stabilization activities (Samii et al. 2011, pp. 1-2). But in the age of fake news, another critical enabler for unity of effort is the conduct of Information Operations (IO) and Influence. From the perspective of influence, “military actions are really a way of speaking in a larger political context... establish the narrative first, and then evaluate all potential actions against it” (US Department of Defense, 2008, p. 11).

This paper presents a recent research project which examined ways of enabling stakeholders to respond to the increasing use of influence in warfare, hybrid conflict, hard and soft politics. An international cross-sector research group drawing on military, government, and academic expertise from seven different countries met in October 2018 to understand the best way to wargame influence. The project discussed here presents a truly innovative approach to influence studies, and shows the benefits of an interdisciplinary, cross-domain research team. This paper also introduces a new influence wargaming framework which emerged during the study. This paper discusses cutting-edge research in a field of importance to contemporary politics and security. The work is innovative, cross-disciplinary, and of potential benefit to a wide number of domains.

## **2. Description of the problem**

RAND defines influence as “the coordinated, integrated and synchronised application of national, diplomatic, informational, military, economic and other capabilities in peacetime, crisis, conflict and post-conflict to foster attitudes, behaviours or decisions by foreign target audiences...” (Larson et al. 2009) to achieve national interests and objectives. The inclusion of foreign audiences in this definition is contentious, given both autocratic and democratic political activities seek as their primary aim to influence domestic target audiences. Many government change management activities which are identified as ‘good’ for society utilise coordinated, integrated and synchronized capabilities to achieve intra-national social change. The paradigm of influence in a competitive world as described herein reflects the narrative across many political institutions.

Democratic nations seek to maintain the balance between protecting citizens, government and institutions from malicious foreign influence and enabling community debate, and permitting dissenting views and effective alternative government and governance (Frost and Michelsen 2018, p. 88). Influence is conducted to alter both behaviour and attitudes. The argument has been made that influence “exert[s] power via the use of soft techniques... so minimizing the use of hard techniques such as military force and economic sanctions” (Hutchinson 2010, p. 13). However, the use of coercion as a form of influence reflects the dynamic, temporal nature of its application. Influence can be soft, or physically and economically hard. This broad mix of options confounds planners because of its breadth and complexity. Behaviour and the stimulus informing decisions that lead to it, such as beliefs and attitudes, can be influenced in a variety of ways including the use of physical, informational and cognitive means. Models explaining the practical application of influence in military environments differ in regard to many assumptions, including how attitudes and behaviours interact (Hutchinson 2010, pp. 13-14).

The lack of models and data available to military planners pertaining to Influence has contributed to the lack of a strategic art addressing the unity of effort required to integrate diplomatic, informational, military and economic action. For example, influence has been conceptualised as an aspect of military command and control rather than a feature of populations and politics (Flaherty, 2003). IO, on the other hand, has been described as a more technical area of military doctrine, consisting of inter-related Information Related Capabilities (IRCs) (Armistead, 2004). The technical and doctrinal relationship between IO and influence is not the core subject of this paper. However, it was a fundamental issue for project participants. Throughout the project there was tension between the military focus on technical IO and the academic, whole of government and political focus on broad influence. The multi-national and multi-disciplined participants held a variety of views on how influence and IO interacted and should be conceptually organised. Therefore, military doctrine which generally places influence in a subordinate role underneath IO was not strictly followed. Rather, the interaction between IO and influence as a broad concept was explored within a bounded problem space.

## **3. Gamification and serious gaming as a research tool**

All play means something (Huizinga, 1949, p.1). The idea of a ‘game’ as mere recreation neglects the ways in which a game is fundamentally a rules-based abstracted model of some aspect of reality (bargaining, negotiation, or conflict). This abstraction allows the study of aspects of the real world which would be harmful, costly, or catastrophic to investigate in reality. The use of gaming as a training/research tool has developed from the military *Kriegspiel*, evolving into the wider field of ‘serious gaming’ (Southgate et al 2016; Jansz 2016), a domain which includes ‘Gamification’ in general (Werbach 2016), and the more specific field of ‘Wargaming’. Wargaming has been discussed from numerous angles, including historical (van Creveld 2013) and methodological frameworks (Sabin 2014). The foundation of gamification is rooted in self-realization theory popularized by Daniel Pink (2011). In this paper the term wargame describes a specific example of a serious game.

The way in which a wargame delivers its simulation of the real is described in two forms; computer based constructive simulation (CAX) and physical live roleplaying. One advantage of the CAX is the quantitative precision in which each move is conducted and no step is lost, forgotten or biased by human error (Hartley III 2017), but the trade-off is a loss of flexibility owing to the difficulty of editing source code during the simulation. The other form consists of manual wargames, which are innately flexible and agile (Sabin 2014). The most flexible form is arguably the matrix wargame with its almost limitless ability to evolve on the fly in response to players’ actions (Curry and Price 2017). However, this comes with a lack of analytical precision and unavoidable human bias. Wargaming approaches should be selected based on the type of problem they seek to address, the

experience of participants and the resources available. The project discussed within this paper sought to develop a hybrid wargame approach, combining the flexibility of a matrix wargame with the highly structured and quantitative strengths of a CAX.

#### **4. The project**

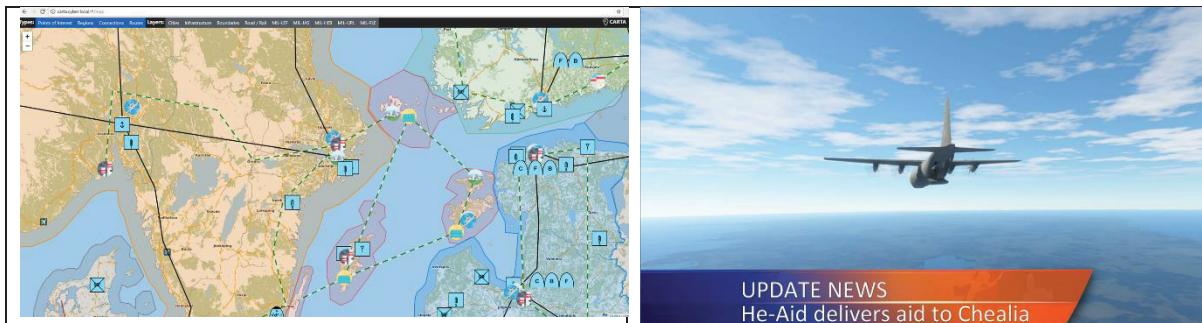
The project team devised a wargame, i.e. a “scenario-based warfare model in which the outcome and sequence of events affect, and are affected by, the decisions made by the players” (UK Ministry of Defence, 2012, Lexicon, p.6). The use of ‘warfare’ in the context of this project, where influence is the focus, implies a competitive strategic environment. Nations’ competing agendas and strategic goals did not necessarily require kinetic conflict. The project was intended to provide a “structured but intellectually liberating safe-to-fail environment to help explore what works (winning/succeeding) and what does not (losing/failing)” (UK Ministry of Defence, 2017, p.5). Within the artificial and constrained environment of the game space, organizers and players were encouraged to explore the full range of outcomes deriving from an initial scenario, with the aim of encouraging innovation and creative thinking. The project sought to develop, build and test a proof-of-concept matrix wargaming approach including aspects of tabletop and simulation approaches, enabling a repeatable technology-enabled process specifically dealing with the issue of influence. The project was exploratory, seeking to build the initial scaffolding of a broader influence wargaming framework and the first iteration of a ruleset. Verification, validation and accreditation of the project and its artefacts was considered out of scope. The project was deliberately conceived to bring together stakeholders from different domains (the military, government, academia) and oblige them to work together as teams, thereby cultivating synergies, challenging groupthink (Whyte, 1951; Janis, 1971) and silo thinking. A declared objective of the project was to establish a community of influence wargame experts.

The diverse project group, which included IT support, computer animation experts (responsible for generating CGI imagery for use in rolling news feeds) and a graphic artist, worked both collectively and in discrete task-and-finish groups, to design the underlying environment of the game world. Rather than adopting the model of a conventional Red/Blue exercise, with a single company, government agency or country falling prey to an IO run by a state or non-state actor, one of the key concerns behind the exercise was to offer the widest possible range of actions and events, to allow both players and game runners the maximum opportunity for innovation and creativity. To this end, the game’s setting became the fictional Belmar region depicted in FIGURE 01, composed of six independent states, each with a distinct culture, economy, political system, kinetic and cyber capability.



**Figure 1:** The Belmar region; the fictional game environment

Much of the worldbuilding detail was not made available to the players. However, the construction of an environment which was coherent, congruent, and ultimately immersive provided depth, context and texture for participants (Stuart, 2010). Further context was provided through the creation of real time country-specific international news channels and social media feeds. FIGURE 02 and FIGURE 03 depict some of the immersive visualisations provided to players through simulation systems and the technology federation of the wargame.

**Figure 2:** The Game Common Operating Picture (COP)**Figure 3:** An immersive news feed produced by the white cell media team

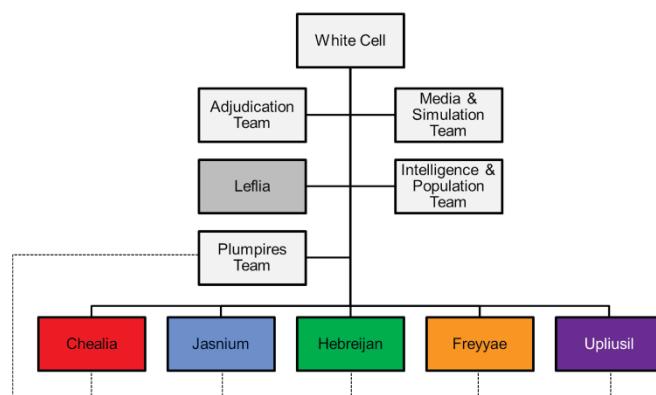
Of the six nations in the game world, only five were assigned to player teams; the largest nation and major regional power, Leflia, was the home state of the game controllers, and as such played no direct part in the gameplay cycle. A controller state allowed management of the teams in a way which avoided simple intervention which would break player engagement. The actual rules and methodology of the game evolved over the course of its first running based on the initial setting of power dynamics in the region.

A key element of the project was a technological federation providing a simulation, communication and immersion backbone, supported by a white team which arbitrated events and gameplay. The federation created a wide range of data to inform player ‘moves’, including: situational awareness through a Common Operating Picture (COP); immersion through virtual simulated news feeds and social media applications (see above); and economy modelling nation and non-nation state effects. Additionally, Slack was used as an asynchronous communication tool for the project team, which allowed private work groups to be established. This IT-driven backbone sought to enable players to utilise the full potential of diplomatic, informational, military and economic options. The project team sought to ensure these options were supported with sufficient detail and data to make the game realistic. A series of social media systems supported by 22,000 bots and 44,000 social media accounts produced 80,000 messages over the four days of wargame conduct together with 16,000 likes. Each bot was developed using non-attributable census data and used base personalities that included home and work addresses, familial relationships, as well as political and social attributes. Bots engaged in a variety of activities including routine discussion, flaming and trending based on decisions enacted by the white team, or alternatively through the bot machine learning constructs, profiles and relationships defined in training data. The variety of immersive feeds and data allowed teams to receive feedback ‘in game’ through the system rather than through briefs or matrix wargame style presentations.

There are several aspects specific to the creation of an Influence game. The very nature of Influence requires a cross-discipline approach to be effective (Psywarrior, 2018). This suggests team members should contain representation from military, political science, psychology, sociology, and communications along with traditional technical disciplines. In support of this goal the exercise included representation from many domain experts. Central to effective influence, and to evaluating team performance in the game, is an understanding of the interaction of individual and group values. Psychology exists within the context of culture (Wang 2016). Culture defines behavioral norms, and while it cannot be used to predict an individual’s behaviors, shared cultural values can provide behavioral boundaries that individuals will not cross (Hofstede et al. 2010; Minkov 2013). Shared cultural values can also explain patterning behaviors during learning (Morgan et al. 2015). The cultural values that individuals carry into the exercise can result in some members deliberately withholding their opinions or views, for reasons varying from fear of authority to general conflict avoidance (Hofstede et al. 2010; Nisbett 2004). Conversely, the voices of outspoken individuals may drown out other team members’ views (Hofstede et al. 2010; Nisbett 2004). Social and cultural dynamics can particularly influence game scenarios and outcomes. Nisbett (2004) shows that different cultural values result in different ways of grouping and organizing data. These differences are believed to cause different strategies and tactics in the virtual environment. More recently differences that align with cultural values have been observed in attacking behaviors (Sample et al. 2017; Sample et al. 2016) and technology usage (Almeshekah and Spafford 2014; Elmasry et al. 2014; Sample and Karamanian 2014).

## 5. Participant team structure

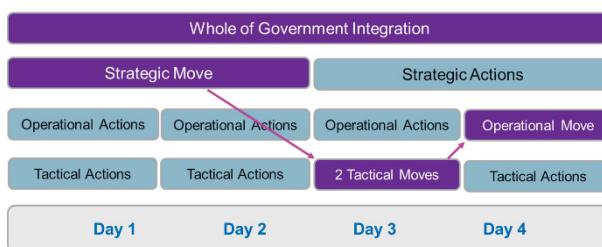
The participants were broken into six groups as depicted in FIGURE 04. The white cell consisted of experts facilitating the wargame and maintaining the technology federation, a media and simulation team, and an intelligence and population team. The remainder consisted of activity controllers who adjudicated when turns were made or events occurred. This adjudication team also included player-umpire hybrids (the term ‘Plumpire’ (or ‘Spielrichter’) was coined in discussions with wargaming experts John Curry and Peter Perla prior to the conduct of the wargame). Plumpires/Spielrichters were assigned to each national team. Their role was to support the teams from a game process perspective, contribute to white cell deliberations and represent their teams during adjudication. However, they were also required to act impartially, particularly when it came to managing secret moves they were aware other nations were making. All plumpires had been involved in developing the scenario. Plumpires were able to provide context, relevant scenario information to teams and act as an information conduit back to the activity controllers.



**Figure 4:** Wargame team structure

## 6. Moves and adjudication

The wargame ‘moves’ (game turns) occurred over a four-day period. Each team made moves at the same time, submitting a Concept of Operations (CONOPS) which described the effects they were seeking to generate, how those effects would be generated, and how success would be measured. Due to limitations in time and the desire of the project team to observe the challenges associated with game time in the context of an influence wargame, different types of moves were trialled. A two-day strategic move was needed initially to allow teams to establish their initial stance. This move included the submission of CONOPS aimed at a strategic level, as well as a series of press releases and statements by country leaders and their representative. A strategic move simulated six to 18 months. Day three consisted of two tactical moves. In this context a ‘tactical move’ was a discrete action that would take less than 15 days in game time. For example, the embarkation of an amphibious assault group, cyber attack, drone strike or a naval blockade were all considered tactical moves because they were discrete actions. Finally, an operational move consisted of a 30-90 day period of time. Whole of government integration needed to be considered and described by the team in every move, regardless of the level and duration. FIGURE 05 depicts game-time over the four days.



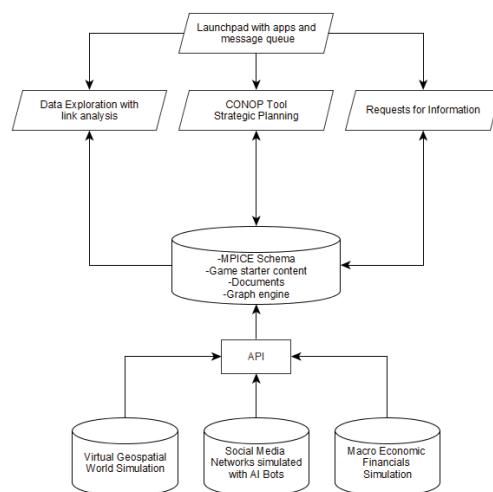
**Figure 5:** Game-time over four days by move level

The adjudication process for each move involved the presentation of each national CONOPS, supplemented with a detailed presentation by each Plumpire of their team’s deliberations. A review, supported by CONOPS risk assessments in which explicit mitigations against another country’s action were stated then determined how each nation would impact each other’s plans (where necessary, a dice throw determined level of success of an

action). The plausibility, feasibility, suitability and acceptability of all plans were assessed to support moderation/adjudication of all CONOPS presented. The intelligence and population team were given an opportunity to comment on any issues or changes that would occur because of the moves. Finally, using this information, a run sheet was prepared by the media and simulation team to provide feedback to all the national teams through the technological support backbone as news bulletins, social media feeds and economic reports.

## 7. Technological federation backbone

The technical solution integrated closed network simulation systems such as social media, financial and geospatial virtual events, providing a backbone for communication across the teams and access to scenario data. The technological federation incorporated a multi-modal NoSQL database with document store and graph model fed by real-time game data. This established a digital intelligence platform on which wargaming “moves” were aligned and adjudicated by the white cell.

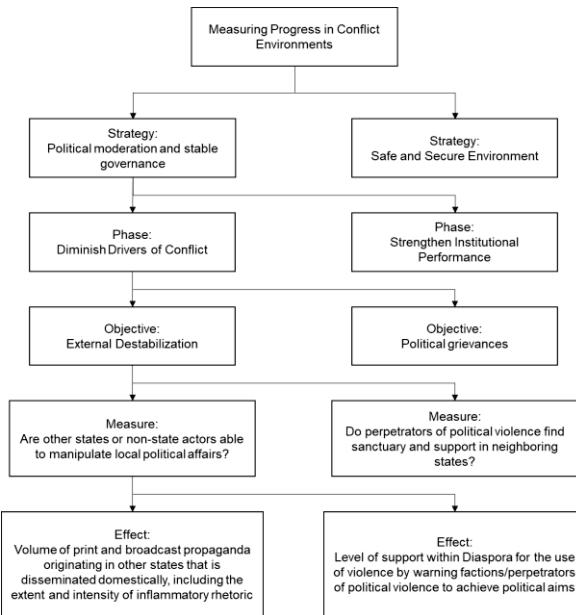


**Figure 6:** Technical solution architecture with central database serving as the integration and user application platform

The solution's focus was providing a set of digital processes to facilitate player moves as opposed to traditional cards and board games in a physical or matrix game space. Customized apps and data model including a graphical user interface with link analysis and launchpad were created to facilitate interactions with communication channels, data exploration and ability to create new intelligence entered through digital forms. The two apps that drove the game moves were Requests for Information (RFI) and the CONOPS. The RFI process is an internal closed-network messaging app that routes questions from an operations user to an intelligence user located in the white cell with access to all information as opposed to the filtered view of teams within a fog of war. Intelligence users receive the RFI as a new message within their queue and provide an answer within a similar field resulting in a new message within the operation user's queue. CONOPS are built using a form with auto-populated drop downs and pre-defined fields that allow players to enter in their plan for actions in a game move. Players choose from existing higher-level strategies and enter descriptions of resources required to accomplish an objective. It includes intended maneuvers, risk assessments, and justifications for commanders to validate and approve execution. The white cell then receives a CONOP from each team allowing the adjudication process.

While the digitization of the wargaming simulation attempts to bring precision to the traditional physical matrix style game environments, the technical challenge of tracing CONOPS effects back to a strategy remains. The findings from the project suggest implementing existing frameworks that attempt to answer this challenge and incorporate them into the subsequent iterations of the solution. Measuring Progress in Conflict Environments (MPICE) is a framework started by the Center for Strategic International Studies (Dziedzic et al. 2008) to quantify the effects of a diplomatic, informational, military, or economic (DIME) action. Influence and complex attributes of a conflict environment such as sentiment are structured into a taxonomy allowing leaders to understand operations' effects. The effects collectively support a strategy through multiple levels including phase and objective. FIGURE 07 shows a single MPICE branch traversed from strategy to effect with greyed out alternate paths representing other available elements from the framework's catalog. The technological solution proposed

the MPICE model in which players would assign an MPICE measure within a CONOP to tactical IRCS such as Civil Affairs, Psychological Operations, or other military units.



**Figure 7:** Example of the MPICE (Measuring Progress in Conflict Environments) catalog hierarchy where a single measurable effect such as the volume of relevant print and propaganda can be traced from a strategic effort in context of influence operations

While the MPICE model has been validated as a thorough framework for social and political indicators or the diplomatic, informational, and military aspects of DIME, it is less so on economic effects. For this Samii et al. (2011, p. 14) suggest the World Bank's Living Standard Measurement (WBLSM) survey where undeveloped countries' economic indicators are better captured than those with developed industries and measures. Therefore, subsequent iterations of the technological solution will incorporate WBLSM with MPICE to capture the full DIME spectra. The technology federation during the wargame demonstrated great potential. Several areas for future development have been identified and vigorously pursued.

## 8. Key findings

A broad array of findings emerged from the project. Both the teams participating and those who developed the scenario and managed the delivery of the game needed to balance short term news cycles and events with long term influence. The role of the economy and monetary policy as a component of international relations and influence caused some participants to struggle, particularly those who were focused on the delivery of military IO. Some participants noted that the decision making by teams was biased based on national outlook and the degree of existential threat they observed in the scenario. There was general agreement that the unity of effort across government concept had value, but how that worked in a democratic construct and how counter-influence could work in a democracy was debated. This debate concluded that significant research and effort needs to be applied across the international community, both civilian and military, to build a coherent conceptual model that suits more than one audience and purpose. This model should address the links between tactical, operational and strategic IO; information and technology; and the broader concept of influence, operating at a political, diplomatic, economic level. Attempts to integrate DIME to solve specific problems revealed a lack of depth to the concept. Most participants could speak as an expert to one component of the DIME model, but few seemed to be able to integrate across the DIME model effectively.

There are many challenges in modelling and simulating the tactical, operational or strategic environment and measuring relevant effects. There are several approaches to build computer-based simulations especially for unconventional conflicts (Hartley III 2017). However, there is a lack of methodology modelling influence in the strategic environment in CAX. There are approaches in matrix wargaming to cover certain aspects of influence (Curry and Price 2013, 2017). According to the 'many models approach' (2016) there might be one model for many problems and/or many models for one problem, thereby allowing different perspectives on the 'problem space' for modelling influence. According to Hatley the DIME paradigm has a certain prominence in the

modelling community to model the domains of power a nation state (as an actor in the international political system) can exploit to reach its political objective (Hartley III 2017). A more detailed description of the problem space is in “Thoughts about a General Theory of Influence in a DIME/PMESII/ASCOP/IRC<sup>2</sup> Model” (Kodalle et al. 2019). In this description, DIME actions occur in the PMESII (political, military, economic, social, information, infrastructural) environment to deliver MPICE. DIME/PMESII occurs along the ASCOP (areas, structures, capabilities, organizations, people) “human terrain” dimensions. The DIME/PMESII/ASCOP dimensions result in a 3D rectangular prism. Each rectangle is projected on a timeline with short, medium or long-time intervals and broken down to the tactical, operational and strategic level. These projections create points in the problem space which can be exploited by an actor to influence. This approach was developed during the project and might be useful as an analytical framework, such as scenario analysis, to identify attack vectors to shape the environment. Further research is recommended to expand upon this preliminary approach.

The proof-of-concept wargaming influence framework resulting from the project is presented in FIGURE 08 and contains several features of the wargame activity for further exploration and development. The groups structure within the game, the management of game-time, the management of team outputs and the adjudication process are all critical considerations. Finally, the representation of diplomatic, informational, military and economic factors as they relate to the game representation must also be considered in depth. The framework will continue to develop and expand as future activities test its components and the relationships between them.

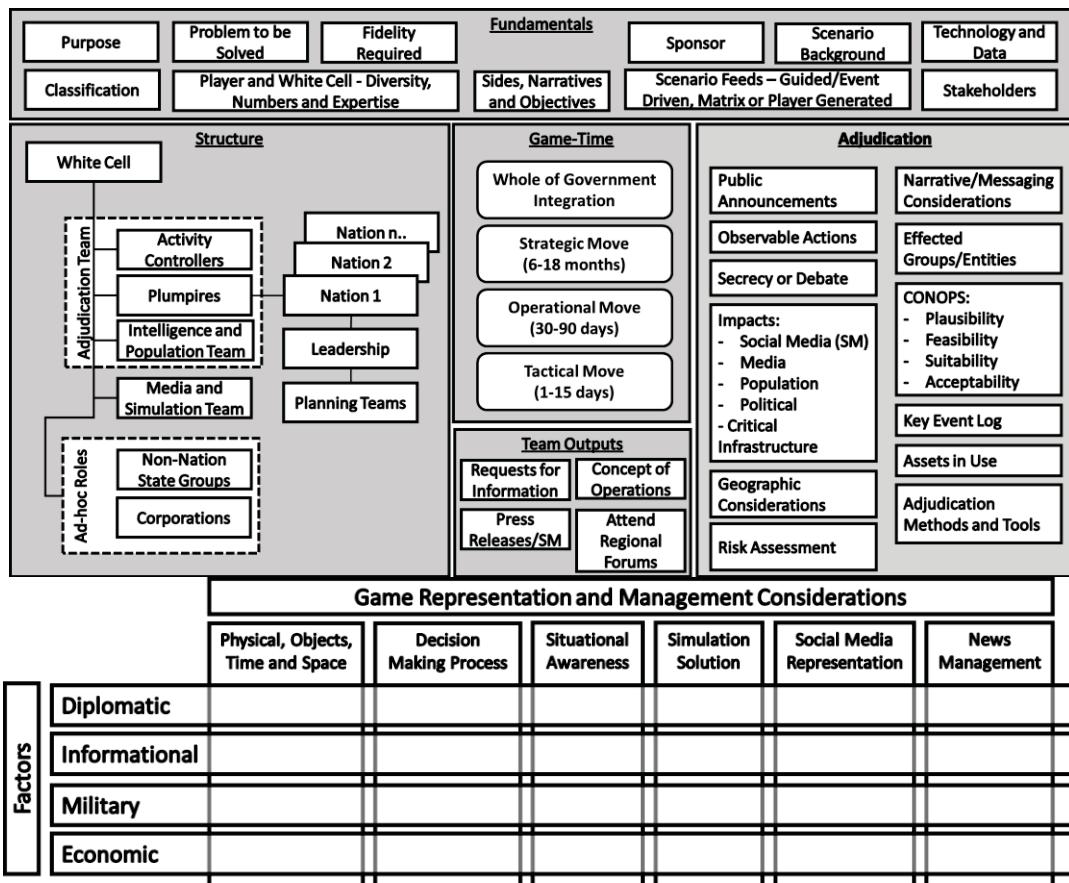


Figure 8: Proof-of-Concept wargaming influence framework

## 9. Conclusion and future work

The success of the wargaming activity described in this paper is directly attributable to the talented and driven participants who developed, delivered and participated in the activity. Despite this success, numerous areas remain open for further research and development in future iterations. The integration of technology and simulation solutions with human users, particularly those not familiar with technology, brings with it a training liability and brittleness that can impact participation. However, for those who embrace technology the degree of immersion may increase. Cultural bias, particularly when players must take on roles they are not familiar with or national identities very different to their norms, may have an impact on outcomes and decisions. The conduct of higher classification activities using more realistic data will present additional challenges. The use of bots to

represent individuals in simulations and social media environments using non-attributed data presented numerous opportunities but significant research is required to ensure bot representation is sufficiently realistic to support scenario requirements. The management by the white cell of event-driven scenarios versus player-driven matrix-style events remains an area for further experimentation. Finally, the verification, validation and accreditation of the systems utilised throughout the wargame must be considered, but only once a greater understanding of the problem space has been developed. A proof-of-concept wargaming influence framework has been developed as an initial point for exploration in future activities and this is proposed as an area of further research and development.

## **References**

- Almeshekah, M. H., and Spafford, E. H. (2014) *Planning and Integrating Deception into Computer Security Defenses*. Proceedings of the 2014 New Security Paradigms Workshop: ACM, pp. 127-138.
- Armistead, L. (Ed.). (2004) *Information operations: Warfare and the hard reality of soft power*. Potomac Books, Inc.
- Boehlke, T. (2006) "Wargaming Guide to preparation and execution". [online], <http://www.professionalwargaming.co.uk/F%C3%BCCAkBw-Wargaming-Guide-2006.pdf>
- Boskic, N. (2018) "Journal articles and reports on serious games". Online Learning and Distance Education, [online], <https://www.tonybates.ca/2018/08/02/journal-articles-and-reports-on-serious-games/>
- Clancy, T. (1995, c1994) *Debt of honor*. Berkley Education. New York: Penguin Group, USA.
- Curry, J. and Price, T. (2013): *Dark Guest Training Games for Cyber Warfare Volume 1: Wargaming Internet Based Attacks* (English Edition). Kindle Edition.
- Curry, J. and Price, T. (2017) *Modern crises scenarios for matrix wargames*. History of Wargaming Project.
- Dziedzic, M., Sotirin, B., and Agoglia, J. (2008) "Measuring Progress in Conflict Environments (MPICE)-a Metrics Framework for Assessing Conflict Transformation and Stabilization. Version 1.0," Corps of Engineers. Department of Defense. Washington, DC.
- Elmasry, M. H., Auter, P. J., and Peuchaud, S. R. (2014) "Facebook across Cultures: A Cross-Cultural Content Analysis of Egyptian, Qatari, and American Student Facebook Pages," Journal of Middle East Media. Vol 10.
- FEMA (nd). "CERT Tabletop Exercise #1". Washington, DC: FEMA. [online], [https://www.fema.gov/media-library-data/20130726-1917.../cert\\_tabletop\\_1.pdf](https://www.fema.gov/media-library-data/20130726-1917.../cert_tabletop_1.pdf)
- Flaherty, C. (2003) "The Role of Command and Influence in Australian Multidimensional Manoeuvre Theory". Australian Defence Force Journal, Issue 162. pp.31-38.
- Frost, M., and Michelsen, N. (2018) "International Ethics and Information Warfare," in Hybrid Conflicts and Information Warfare: New Labels, Old Politics. Ofer Fridman (Ed). Lynne Rienner Publishers.
- Hartley III, D.S. (2017) *Unconventional Conflict. A Modeling Perspective*. Springer International Publishing (Understanding Complex Systems), [online], <http://dx.doi.org/10.1007/978-3-319-51935-7>
- Hofstede, G., Hofstede, G. J., and Minkov, M. (2010) *Culture and Organizations*. New York. McGraw Hill.
- Huizinga, J. (1949) *Homo Ludens: A Study Of The Play-Element In Culture*. London: Routledge and Kegan Paul.
- Hutchinson, W. 2010. *Influence Operations: Action and Attitude*. Proceedings of the 11th Australian Information Warfare and Security Conference, Edith Cowan University, Perth Western Australia, 30th Nov- 2nd Dec 2010. Edith Cowan University. [online], <https://ro.ecu.edu.au/isw/33/>
- Janis, I. L. (1971) "Groupthink". Psychology Today. Nov 1971. Vol 5, Issue 6. pp. 43–46, 74–76
- Jansz, J. (2016) Serious Gaming. Coursera. Erasmus University Rotterdam. [online], <https://www.coursera.org/learn/serious-gaming>
- Kodalle, T., Ormrod, D., Sample, C., Scott, K. (2019) "Thoughts about a General Theory of Influence in a DIME/PMESII/ASCOP/IRC2 Model". Unpublished - provided for ECCWS 2019.
- Larson, E. V., Darilek, R. E., Gibran, D., Nichiporuk, B., Richardson, A., Schwartz, L. H., and Thurston, C. Q. (2009) "Foundations of Effective Influence Operations: A Framework for Enhancing Army Capabilities," RAND. Santa Monica, CA.
- Minkov, M. (2013) *Cross-Cultural Analysis: The Science and Art of Comparing the World's Modern Societies and Their Cultures*. Sage.
- Morgan, T. J., Cross, C. P., and Rendell, L. E. (2015) "Nothing in Human Behavior Makes Sense except in the Light of Culture: Shared Interests of Social Psychology and Cultural Evolution," in Evolutionary Perspectives on Social Psychology. Springer, pp. 215-228.
- Nisbett, R. (2004) *The Geography of Thought: How Asians and Westerners Think Differently... And Why*. Simon and Schuster.
- Page, Scott E. (2016) *Model Thinking*. Coursera. University of Michigan. [online], <https://www.coursera.org/learn/model-thinking>
- Perla, P. (ed. Curry, J.) (2012) *Peter Perla's The Art of Wargaming: A Guide for Professionals and Hobbyists*. Morrisville, NC: Lulu.com.
- Pink, Daniel H. (2011) *Drive. The surprising truth about what motivates us*. Riverhead Books. New York.
- Psywarrior. (2018) "Psywarrior Website." [online], <http://psywarrior.com>
- Ready.gov (2017) 'Exercises'. [online], <https://www.ready.gov/business/testing/exercises>
- Sabin, P. (2014) *Simulating War: Studying Conflict through Simulation Games*. Bloomsbury Academic Press. London.

- Samii, C., Brown, A., and Kulma, M. (2011) "Evaluating Stabilization Interventions," International Initiative for Impact Evaluation, London.
- Sample, C., Cowley, J., and Hutchinson, S. (2017) "Cultural Exploration of Attack Vector Preferences for Self-Identified Attackers," 11th International Conference on Research Challenges in Information Science (RCIS), IEEE, pp. 305-314.
- Sample, C., Cowley, J., Watson, T., and Maple, C. (2016) "Re-Thinking Threat Intelligence" International Conference on Cyber Conflict (CyCon US), IEEE, pp. 1-9.
- Sample, C., and Karamanian, A. (2014) "Application of Hofstede's Cultural Dimensions in Social Networking," Proceedings of the 1st European Conference on Social Media, ECSM, pp. 466-473.
- Southgate, E., Smith, S. P., Smithers, K. and Budd, J. (2016). *Serious Games And Learning: An Annotated Bibliography*. University of Newcastle. Newcastle.
- Stuart, K. (2010) "What do we mean when we call a game 'immersive'?" Guardian, 11 August [online],  
<https://www.theguardian.com/technology/gamesblog/2010/aug/10/games-science-of-immersion>
- UK Ministry of Defence (2012) *Red Teaming Guide*. DCDC Shrivenham. Development, Concepts and Doctrine Centre, UK Ministry of Defence.
- UK Ministry of Defence (2017) *Wargaming Handbook*. [online],  
[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/641040/doctrine\\_uk\\_wargaming\\_handbook.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/641040/doctrine_uk_wargaming_handbook.pdf)
- US Department of Defense. (2008) *Full Spectrum Operations - Unified Quest 2007*, Tradoc Pamphlet 525-5-300. United States Army. Fort Monroe, Virginia.
- van Creveld, M. (2013). *Wargames. From gladiators to gigabytes*. Cambridge University Press. Cambridge. [online],  
<https://doi.org/10.1017/CBO9781139579872>
- Werbach, K. (2016): Gamification. Coursera. University of Pennsylvania. [online],  
<https://www.coursera.org/learn/gamification>
- Whyte, W.H. (1952) "Groupthink". Fortune, March 1952, pp.114-17, 142, 146.
- Wilson, A. (ed. Curry, J.) (2014) *The Bomb and the Computer The History of Professional Wargaming 1780- 1968*. Morrisville, NC: Lulu.com.
- Wang, Q. (2016) "Why Should We All Be Cultural Psychologists? Lessons from the Study of Social Cognition," *Perspectives on Psychological Science* Vol 11, Issue 5. pp. 583-596.

# Cyber Attribution 2.0: Capture the False Flag

Timea Pahi and Florian Skopik

AIT Austrian Institute of Technology, Vienna, Austria

[Timea.Pahi@ait.ac.at](mailto:Timea.Pahi@ait.ac.at)

[Florian.Skopik@ait.ac.at](mailto:Florian.Skopik@ait.ac.at)

**Abstract:** In times, where hacking back is increasingly considered as a legitimate reaction to cyber attacks against nation states, misattribution may undermine a state's credibility and lead to political differences. Cyber attribution at this level must deliver reliable results. In recent years, threat intelligence services have often raised concerns regarding the reliability of attribution, and repeatedly pointed out the possibility of false flag operations. The intention of false flag campaigns is not necessarily to trick intelligence services but also to form public opinion. Unfortunately, there is a lack of a reliable approach that deals with the interdisciplinary challenges of cyber attribution. Additionally, there is a lack of concepts designed to deal with possible false flag operations on the technical side (e.g. manipulating digital evidences) and socio-political side (e.g. distributing fake news). Therefore, we propose a novel concept, the Cyber Attribution Model (CAM) to address these aspects. The model is divided into two closely interacting parts: Cyber Attack Investigation and Cyber Threat Actor Profiling. The scope of the CAM is mainly on professional and organized cyber attacks, such as espionage or APT campaigns, and designed for application in national cyber security centres. This paper presents further a literature research and the attribution model, (1) which is adjusted to today's challenges resulting from the information war, such as false flag operations, and (2) which supports security experts – from technical analysts to intelligence services – to master the attribution process on all levels. Finally, we demonstrate the application of the Cyber Attribution Model in context of a real-world scenario.

**Keywords:** cyber attribution, profiling, cyber investigation

---

## 1. Introduction

The legal perspective of cyber counter attacks, such as hack back (Holzer & Lerums, 2016), is one of the most discussed topics today (Ponemon Institute, 2015). Misattribution of cyber attacks at the national level may undermine a state's credibility and lead to political differences. Threat intelligence services have often raised concerns regarding the reliability of attribution in recent years, and repeatedly pointed out the possibility of false flag operations (Kaspersky, 2016). However, the attackers have an upper hand to reach their targets, while staying anonymous (Goman, 2018) or acting under a false flag, the processes of attributing threats to actors must deliver reliable results.

Nation states have always used information operations to enhance their goals, as conflicts have never been limited to the kinetic warfare (Liang & Xiangsui, 1999). The widespread weaponization of digital vectors has become common place also in the political realm. For instance, nation states and non-state organizations use technological means to distribute fake news, propaganda and other directed-content. Therefore, the attribution model covers the analysis of so called Influence Cyber Operations. According to the NATO definition, these are activities undertaken in and through cyberspace and qualify as cyber attacks with the intention of influencing attitudes, behaviours, or decisions of target audiences. Target audience can be everyone depending on purpose, from forming the public opinion to making security analysts believe that another threat actor is responsible for an attack. The social and psychological aspects of cyber attacks cannot be neglected (CCIOS, 2018) – quite the opposite is the case as they become increasingly significant in cyber operations. Recent incidents, such as the TV5Monde incident, the Sony Hack (Sullivan, 2015), the Winter Olympic Hacking (Dion-Schwarz et al, 2018), show also the need for an analysis approach of possible false flag operations.

The CAM model distinguishes two types of false flags, one applied in technical context and one in socio-political context (see Figure 1). Historically false flag operations were usually conducted at sea, when one ship used the flag of another ship before attacking. Therefore, it was called a 'false flag' attack. According to the NATO terminology, a false-flag is a diversionary or propaganda tactic of deceiving an adversary into thinking that an operation was carried out by another party (NATO CCDCOE). There is a wide range of misdirecting actions. For instance, source IP addresses can be changed by using chains of proxy servers or the TOR network, as well as language settings, agent strings, and false hints in malware code can be placed. On the social side, threat actors could impersonate other actors and generate fake posts and news under wrong identities. It is extremely complicated to get all the false flags consistently right. Therefore, careful attribution must have a particular focus on the consistency of the whole storyline. If a single factor looks odd or does not fit to the obvious story – something might in fact be odd. The presented CAM focuses mainly on targeted and sophisticated cyber attacks

and covers additional social aspects and possible false flag operation for reliable attribution. The paper further presents the outcome of an extensive literature research in this topic with follow-up links and the application of CAM in context of a past real-world scenario.

## **2. Attribution guide: Background and state of the art**

A prerequisite of cyber attribution is to discover the applied techniques, tools and procedures (TTPs). Based on that, the further goal is to identify the source of certain attacks that leads to the threat actor. Both topics, cyber attack investigation (i.e., get to know what happened) and threat actor attribution (i.e., get to know who did it) aims to serve as a basis for actions in law enforcement and national security (such as cyber war or terrorism). There is a wide range of literature on this topic with different approaches. It is often a mix of technical attack analysis and threat actor profiling, that sometimes leads to confusion. Our Cyber Attribution Model (CAM) separates the two corner stones: cyber attack analysis and threat actor profiling and brings them together in the attribution phase. The following section serves as a guide in the wide literature and aims at collecting common denominators of this interdisciplinary topic.

One of the most known models is the Q-Model by Rid & Buchanan (2015). They are looking for an answer to the question, whether attribution is a technical problem or not. Their model introduces multiple levels: strategic, operational, tactical and technical and communication, and several roles: from forensic investigators through national security officers to political leaders. The Q-Model helps analysts to ask the full range of relevant questions and put the investigation into context. It integrates both technical and non-technical information into competing hypotheses. This includes asking more challenging questions on different levels. The study 'Role and Challenges for Sufficient Cyber-Attack Attribution' (Hunker, Hutchinson & Margulies, 2008) summarizes the political, legal and technical challenges. A detailed description of legal issues is available in 'The Law of Attribution: Rules for Attributing the Source of a Cyber-Attack' (Tran, 2018).

Regarding threat actor attribution, the Hacker Profiling Projects has the biggest volume (Chiesa, Ducci & Ciappi, 2008). The research has four principal points of view: technological, social, psychological and criminological. Their profiling methodology contains the 4Ws: who, where, when, why. The resulting hacker profiles contain the following categories: Wanna Be, Script Kiddie, Cracker, Ethical Hacker, Skilled Hacker, Cyber-Warrior, Industrial Spy, Government agent, and Military hacker. The applied correlation standards cover the following aspects for each profile: modus operandi, lone hacker or group, selected targets, hacking career, hacker's ethics, crashed or damaged systems, perception of illegality and effect of laws. The study of PWC developed other profile categories: governments, criminals, hacktivist. They distinguish the perpetrators on the motivation and the technical origin of the attack: cyber crime, cyber activism or cyber warfare. The study 'Cyber Attribution Using Unclassified Data' (2016PPAEP, 2016) focuses on the Diamond Model and on accountability for investigation and prosecution based on cyber attribution. The results show that the cooperation between distinct communities (law enforcement, intelligence community, industry) is required for attribution, and there are no standardized tools in use. The Diamond Model appears again in the research of intrusion analysis (Caltagirone, Pendergast and Betz, 2013).

There are some common challenges in the presented models. The complexity and interdisciplinarity (technical-social-political dimensions) of cyber attribution are the main sources of many challenges. Attribution requires more levels of analysis and the cooperation between different fields of work (security experts, law enforcement agencies). There is no holistic model that unites all the cornerstones of this complex topic, such as digital forensic, hacker profiling, legal issues, and there is no model that focuses on the credibility of the digital evidences. Therefore, the developed Cyber Attribution Model tries to address all these challenges. Hence, for law enforcement agencies already fighting with cyber crime, cyber investigation serves as a basis for the CAM. Cyber investigation is developed based on new requirements resulting from the dependency from technology in almost all areas of the society. Law enforcement agencies need to react quickly on such changes. It is an extended and tailored solution to the digital aspect of crimes today.

## **3. Cyber Attribution Model**

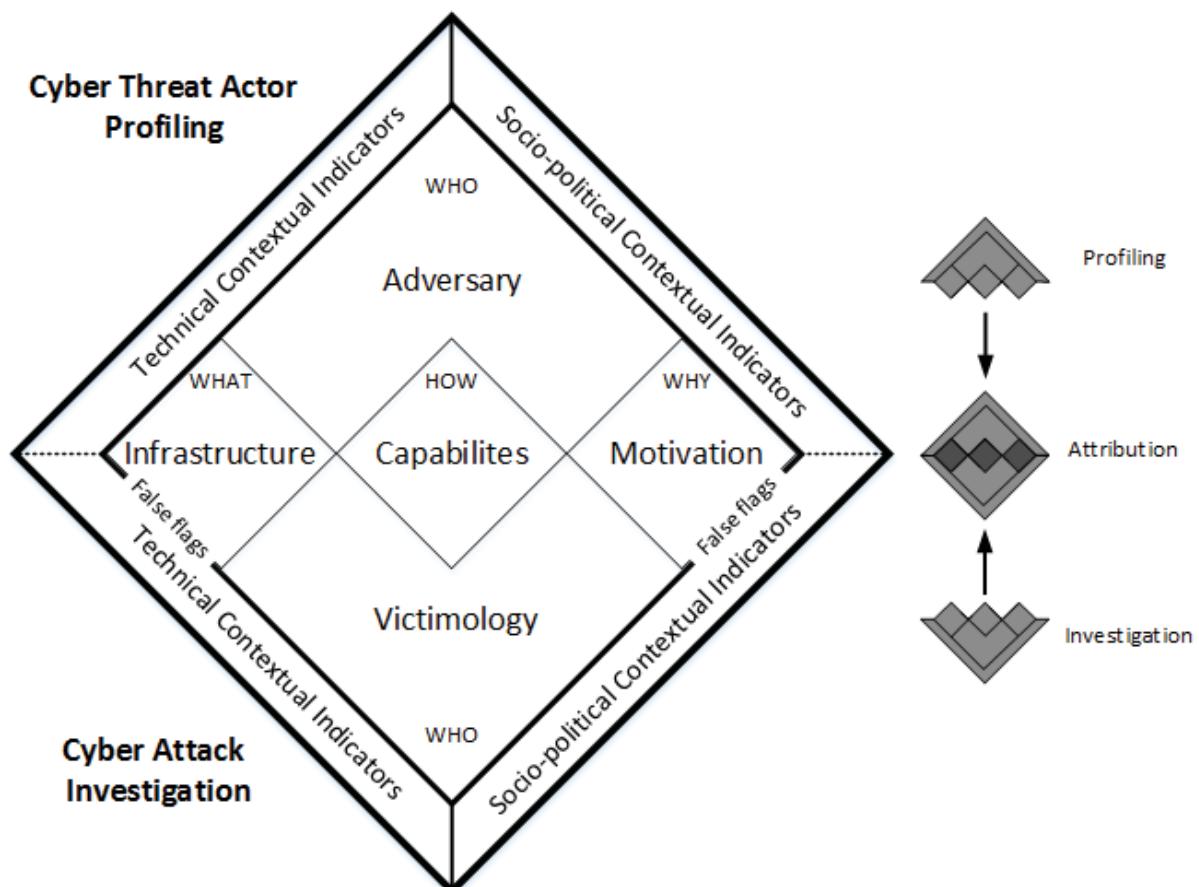
The *Cyber Attribution Model (CAM)* is based on the requirements derived from literature research, and on requirements of national authorities for cyber security (NIS authority or Cyber Security Center) in accordance to the EU NIS Directive. Typical tasks of a Cyber Security Center are to monitor the cyber threat landscape, as well as collect and process different types of threat-related information. The center has in best case cyber security

experts with a solid knowledge on digital forensics, network and application security and penetration testing; law enforcement officers and intelligence officers. The developed CAM unites all relevant components from the technical analysis to threat actor profiling.

### 3.1 Overview

The biggest challenge is to create a universally applicable model. The findings in a cyber investigation arrive usually in a unique, unpredictable order. Hence, strictly defined and sequential flowcharts are not fitting the needs for cyber attribution processes. The application of the model is flexible, and the process flow can differ depending on the currently available information, findings and evidences. The Cyber Attribution Model (CAM) consists of two main parts: *Cyber Attack Investigation* (Part I) and *Cyber Threat Actor Profiling* (Part II). The attribution happens by matching these parts (see Figure 1). Each part consists of Technical and Socio-political Contextual Indicators and the components of our reconsidered Diamond Model.

The primary aim of the Cyber Attack Investigation is to answer the questions, Who is the victim and Why, What has happened and How. Answering these questions is guided by the components: Victimology, Infrastructure, Capabilities Motivation. They help to discover TTPs, the modus operandi of a particular cyber attack and required capabilities and possible false flags. The aim of the Cyber Threat Actor Profiling is developing profiles based on past attacks and find the matching profile to the findings from Part I. The profiling helps finding answers to Who could be the perpetrator, What infrastructure have they used for the attack and What capabilities and motivation might they have. Cyber Threat Actor Profiling takes place either continuously or ad-hoc to support investigations. In that case Part I (bottom-up) and Part II (top-down) are running parallel in order to find a match between the applied TTPs and possible perpetrator profiles.



**Figure 1:** Cyber Attribution Model

In both parts, Technical and Socio-political Contextual Indicators help to understand the evidences and to recognize complex correlations and possible false flag operations. The first step is often analysing the technical hard facts, aka applying digital forensics (Rid & Buchanan, 2015). In this step, security specialists concentrate on the hard facts of already executed cyber attacks as an initial point. Various technical indicators of the attacks are

analysed, such as applied malwares, timestamps, strings, debug path, metadata, infrastructure and backend connection, tools, coding, language settings and pattern-of-life (Bartholomew & Guerrero-Saade, 2016). The difficulty to manipulate or fake technical indicators greatly varies, depending on the infrastructure of the victim and perpetrator. Please note that a detailed description of manipulating technical indicators would go beyond the scope of this paper.

The Socio-political Contextual Indicators cover the use of cyber tools for influencing the perception, opinion and behaviour of a target audience. False flag operations on this side belong to the categories of information war and Influence Cyber Operations (Brangetto & Veenendaal, 2016). Information has been manipulated for political purposes throughout the history of mankind, and the technological revolution opened new possibilities for state and non-state actors to use the cyber space as a tool to shape the social and political mindset (Cohen & Bar'el, 2017). There are numerous examples every day using a wide variety of forms of communication over the Internet and social media, such as influencing elections and spreading propaganda against or for political groups or ideologies, etc. The following section gives more details about the Cyber Attack Investigation and the Cyber Threat Actor Profiling.

### **3.2 Cyber attack investigation and cyber threat actor profiling**

The Who: The starting point of the Cyber Attack Investigation is the victim. According to the retrospective view in Part I, the experts already know who the victim is, and they try to reconstruct the events. This happens by examining the victim itself. At this point victimology comes into effect. Victimology in the cyber space is found in the literature primary in conjunction with cyber crime, especially cyber stalking. The term victimology stems from criminology and covers studying victims of crimes, the psychological effects of the crime. Professionals say, that there is no difference between a physical crime and a digital one (Halder & Jaishankar, 2011). Just as an individual person has victimology-based characteristics, so do organizations. An organization's business interests, political action campaigns, vigilance level, protection abilities, and cyber risk tolerance are just some of the characteristics that can determine if an organization is more likely to be attacked, by whom, how, and why (Bullock, 2018). The complementary part of the victimology is the threat actor profiling (Part II). Its aim is to analyse who is likely to commit a crime and what are the requirements for this. The victims of cyber attacks range from private businesses to the fundamental practices of democracy. Ukraine, France and the United States were affected by attacks during their elections, for instance.

The What: The analysis of the Technical Contextual Indicators and the victim's infrastructure help to reconstruct the operations. The Why: The analysis of the Socio-political Contextual Indicators help to understand possible motives for the attack. The How: After that, the required capabilities can be delivered. What are the minimum requirements to execute the applied technical attacks and social engineering attacks or Influence Cyber Operations. At the end of the Cyber Attack Investigation, the experts have collected all information about the victim and all Technical and Socio-political Contextual Indicators to the incidents. Further results are TTPs, potential false flags on both side, and theories about possible opportunities and motivation.

Deriving TTPs answers mainly the What and How, helps to find the potential threat actors and to prevent similar attacks. The term TTP is often used in conjunction with Threat Intelligence, in the Structured Threat Information eXpression (Barnum, 2012). TTPs refer to Tactics, Techniques and Procedures and represent the behaviour or the modus operandi of the perpetrators. The term has its origin in the traditional military sphere and is used to characterize what an adversary does and how they do it in increasing levels of detail. There is a blurred line between the components of the TTPs. The CAM uses a definition, which we derived from the literature. Tactics have the highest abstraction level. It is the way an adversary chooses to carry out an attack, for instance, to use a malware to steal credit card credentials. Techniques are at a lower level of detail and procedures cover the related preparing processes and technological approaches for achieving intermediate results. An example would be sending targeted emails to potential victims with a malicious code attached. The procedure covers the organizational approach of the attack, for instance a special sequence of actions. This might be reconnaissance to identify potential individuals or creating an exploit to evade malware detection tools. The activities of the APT threat actors are closely linked to the applied malwares and tools. As a consequence, there is a wild mix of naming conventions for APTs. Some use the name of the suspected threat actors, others use the name of applied malware or tools, which are sometimes also inconsistent due to their quick evolution. To sum up, TTPs are used to describes an approach of threat actors, and finally also well suited to profiling threat actors.

Findings from the investigation will be compared to potential threat actor profiles. Profiling is a delicate balance of criminology, psychology and forensics (Turvey, 2011) and studies mainly the motivation and methodology of the attackers. Profiling cyber threat actors is similar to profiling other fields. Since technology changes rapidly, IT security specialist must constantly keep up with the latest attack techniques (Long, 2012).

The Cyber Threat Actor Profiling aims to create, update and manage threat actor profiles. It is the complementary part to victimology and helps to better understand what type of threat actor the perpetrator could be. Part I results in minimum required capabilities and observed TTPs. The analysts compare these results to known threat actor profiles. The What: The analysts are looking for the matching applied infrastructures, tools and tactics. For instance, do the perpetrators have the required special knowledge for preparing the attack against rare industry components (e.g. Stuxnet), do they have the resources to develop their own toolkits and zero-day exploits or do they use already existing components. The Why: The analysis of the motives of the threat actors, such as political motives for hacktivist groups or ideological motives for certain hacker groups. The How: The analysis, that the potential threat actor have the required components at all, such as access to confidential documents or to a certain code. In case, the result from the Cyber Attack Investigation and the potential Threat Actor Profiles do not fit together, the analysts have to consider potential false flag actions. The next section illustrates the application of CAM more visibly.

#### **4. Application of Cyber Attribution Model**

The selected scenario, the well-documented TV5Monde hack, serves as a basis for a short presentation of the application of the CAM (see Figure 2). The victim organization and the French intelligence services have shared their experiences about the incident. The application of CAM starts with Cyber Attack Investigation, especially with victimology. The victim is the TV5Monde, a French television network claiming to be one of the top three most available global television networks internationally (Oakton, 2016). TV5Monde was a victim of a cyber attack, which caused service disruptions for hours in April 2015.

The circumstances need to be understood by analysing Socio-political Contextual Indicators and motivations. At the time of the attack, France was still in shock from terrorist attacks (7th January 2015) on the editors of Charlie Hebdo, a French satirical weekly magazine. The French national agency for the security of information systems reported more than 1,500 cyber attacks against small companies' websites in the wake of the attack on the Charlie Hebdo office in Paris (EuObserver, 2015). Further, the attack on TV5Monde was followed by a series of terrorist attacks, such as mass shooting and suicide bombing in Paris on 13th November 2015. In parallel to the TV5Monde attack, pro-ISIS messages appeared on television's Twitter and Facebook accounts. One of the messages posted against the United States and France, as well as threats issued to families of French soldiers. Furthermore, copies of French soldiers' IDs and passports were published.

The Cyber Attack Investigation continues with the analysis of technical evidences and the victim infrastructure. The Technical Contextual Indicators cover the information discovered by the real incident response team involved in the investigation and reconstruct What is happened. The attackers got their initial access on 23rd January 2015 and took over a server used by the broadcasting company. One of the TV5Monde multimedia servers had its RDP port exposed to the Internet and configured with a default username and password. Since the server was not connected to the internal network, the attackers continued the reconnaissance. They returned, using a compromised third-party account that allowed them to connect to the TV5Monde VPN on 6th February. After that, attackers began scanning internal machines connected to an infected endpoint and identified two internal Windows systems, that were used to manage cameras. Afterwards, the attacker used one of these compromised systems to create a new administrator-level user in the Active Director called "LocalAdministrator" on 11th February 2015 (Oakton, 2016). The first major clue was, that All AD administrator names had French descriptions except for one. During the reconnaissance phase of the attack (16th February - 25th March 2015) the attackers mapped the network's IT services in the victim's infrastructure and collected as much related information as possible, including information from the IT department's internal wiki, which provided details on how logins and passwords were handled (Schwartz, 2017). After that, the attackers compromised another administrator machine with a Remote Access Control software, that was used for the sabotage. At 19:57 on 8th April, the attackers performed their first damaging operation by re-configuring all the IP settings of the media in a faulty manner. This misconfiguration was only enabled, when the technical teams rebooted their machines. At 20:58, the online presence was affected through hacked social media accounts (YouTube, Facebook, Twitter) and the website of TV5Monde which was defaced. At 21:45, the attackers run

commands via TACACS logs, that erased switch and router firmware, resulting in black screens for viewers, except for one new channel that was launched on the same day. The investigations definitely showed the application of Sednit (aka Sofacy) malware, associated with the ongoing Pawn Storm campaign (TrendMicro, 2015). Operation Pawn Storm is an active economic and political cyber-espionage operation that targets a wide range of high-profile entities, from government institutions to media personalities, referred to as APT 28. The challenge is that there is no direct connection between the Pawn Storm campaign and the TV5Monde attack (TrendMicro, 2016).

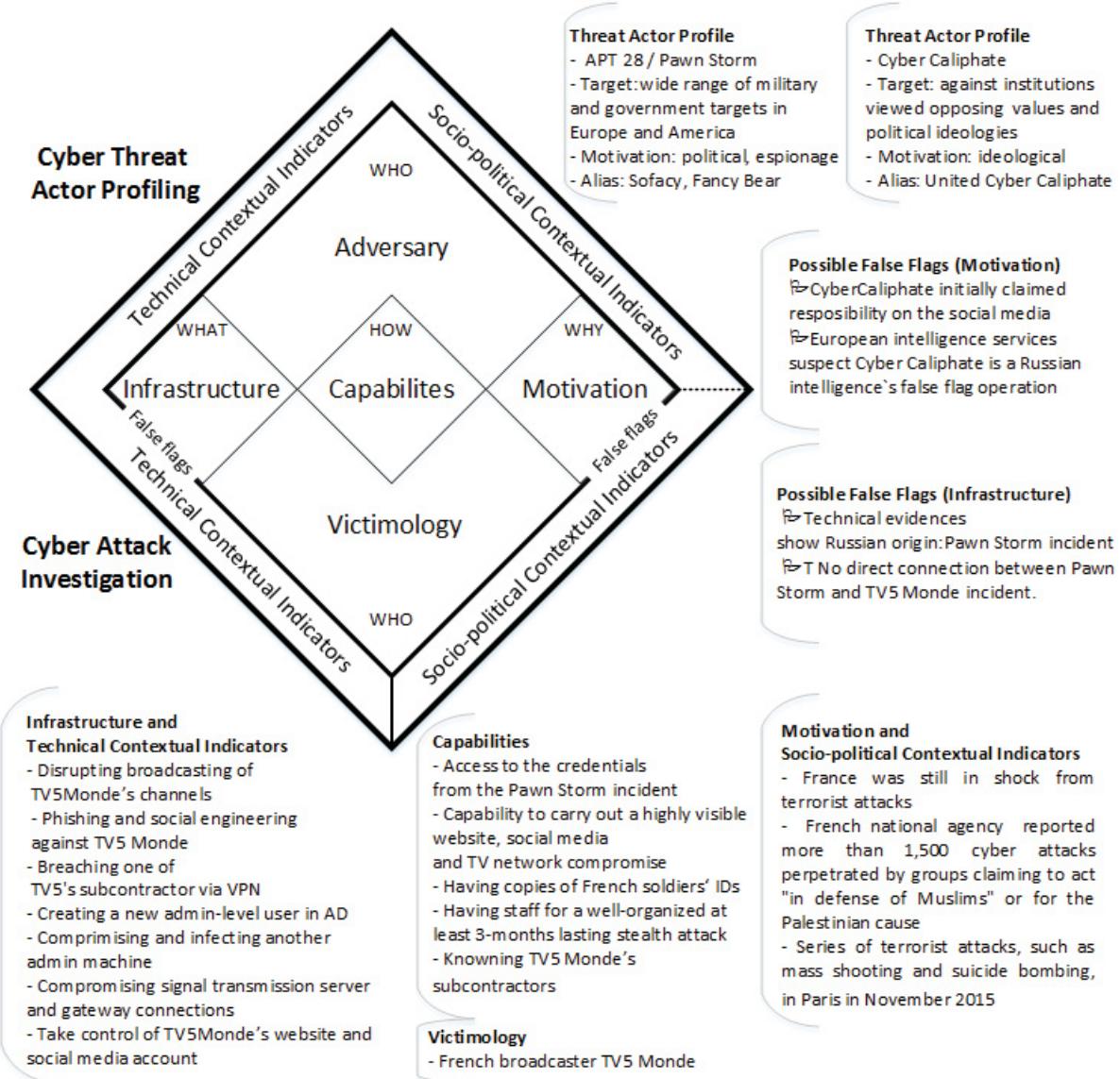


Figure 2: The CAM application

These findings shape possible hacker profiles and so lead to the Cyber Threat Actor Profiling. To sum up the required capabilities, the attackers had to have access to toolsets applied in the Pawn Storm espionage operation and to copies of French soldiers' IDs. Further, they needed the capability and resources (1) to execute an at least 3-months lasting stealthy attack with a deep reconnaissance phase, (2) to have solid knowledge about the victim, by attacking their subcontractors and (3) to carry out a complex network compromise and website-, social media defacement. The Cyber Threat Actor Profiling part supports to analyse, which potential threat actors can fulfil these requirements. The actual perpetrator needed the required infrastructure (e.g. Sednit malware), the capabilities (e.g. well-organized group with deep technical knowledge) and the motivation (e.g. ideological motivated or red herring). The result would be verified asking special questions based on the investigation, such as which threat actors have access to the Sednit malware. In case, there are no fitting puzzle tiles, the analysts consider the possibility of false flag actions. At the time when the attack took place, the CyberCaliphate group initially claimed responsibility for the incident. The analysts however, had to examine, that

the presumable perpetrator could have the required tools and infrastructure, the capabilities and the motivation to execute the attack. Despite the CyberCaliphate's confession, the investigations revealed links to the Russian hacking group APT 28 (CFR, 2018). Finally, in case of the TV5Monde hack, there are three possible theories available with two potential threat actor profiles. A potential threat actor is the CyberCaliphate. This is a hacker group targeting institutions with opposing ideologies. Another one is the APT28 group targeting military and governmental facilities in Europe and America. The first theory is that TV5Monde was the victim of two entirely unrelated incidents, a Pawn Storm infestation and a separate hacktivist compromise. The second theory is that the Pawn Storm group gave information, which was relevant for the attack, to a third party, directly or indirectly connected to Islamic hacktivists. While possible, this would seem highly unlikely as Pawn Storm actively targeted Chechen separatists and Islamic extremists in former Yugoslavia in the past. However, it is also possible that this attack was the work of undisciplined Pawn Storm actors. Though the Pawn Storm actors normally work in a professional way, there have been a few other incidents where some Pawn Storm actors showed a lack of discipline (TrendMicro, 2016). Third, the Pawn Storm group carried out the attack and used it as a false flag operation to lay the blame on Islamic extremists (TrendMicro, 2015). It has become the consensus view among (Western) intelligence services, that the CyberCaliphate and the TV5Monde hack were Russian intelligence's false flag operations. The idea is that the Russian intelligence agencies go to cyber war against the West under an ISIS cloak (Observer, 2016). In that case, the attribution could piece the infrastructure, capabilities and motivation and the matching threat actor together. But since in this particular scenario, there were no clear technical evidences left, the analysts could have continued the attribution process using CAM. When there are no further digital traces, it is possible to continue with intelligence operations in order to underpin one theory with other evidences. As long, as there is no higher degree of certainty, the analysts cannot reliably refer to a potential threat actor and recommend possible (counter)measures. The Cyber Attribution Model aims to identify potential threat actors based on the findings in the cyber investigation considering possible false flag operations.

## **5. Summary and future work**

This paper can offer only a brief insight into the Cyber Attribution Model (CAM) and into its application based on a real-world scenario. Its aim is to steer a reliable cyber attribution process which is adjusted to today's challenges resulting from the information war, such as false flag operations. The main contribution of this paper is the new attribution model which supports the security experts to ask the full range of relevant questions, to aid their critical thinking and to put the investigation into a context. The future work contains the detailed description of the whole Cyber Attribution Model, the required cooperation between different fields of work (from technical analysts through law enforcement agencies to intelligence services), and a deep analysis of possible false flag operations from technical and socio-political aspects.

## **Acknowledgements**

This study was partly funded by the Austrian FFG research program KIRAS in course of the project ACCSA.

## **References**

- 2016 Public-Private Analytic Exchange Program Team (2016PPAEP). (2016). Cyber Attribution Using Unclassified Data. Available at: <https://www.dni.gov/files/PE/Documents/Cyber-Attribution.pdf> [Accessed 20 Jan. 2019]
- Barnum, S. (2012). Standardizing cyber threat intelligence information with the Structured Threat Information eXpression (STIX). MITRE Corporation, 11, 1-22.
- Bartholomew, B. and Guerrero-Saade, J. A. (2016). Wave your false flags! Deception tactics muddying attribution in targeted attacks. In: Virus Bulletin Conference.
- Brangetto, P., & Veenendaal, M. A. (2016). Influence Cyber Operations: The use of cyberattacks in support of Influence Operations. In Cyber Conflict (CyCon), 2016 8th International Conference on (pp. 113-126). IEEE.
- Bullock, C. (2018). Don't Forget Victimology as a Cybersecurity Strategy. Secureworks Inc. Available at: <https://www.secureworks.com/blog/dont-forget-victimology-as-a-cybersecurity-strategy> [Accessed 20 Jan. 2019]
- Caltagirone, S., Pendergast, A., & Betz, C. (2013). The diamond model of intrusion analysis. Center For Cyber Intelligence Analysis And Threat Research Hanover MD.
- Center for Cyber-Influence Operations Studies (CCIOS) (2018). Available at: <https://icitech.org/icit-introduces-center-for-cyber-influence-operations-studies-ccios/> [Accessed 20 Jan. 2019]
- CFR. (2018). Compromise of TV5 Monde. Available at: <https://www.cfr.org/interactive/cyber-operations/compromise-tv5-monde> [Accessed 20 Jan. 2019]
- Chiesa, R., Ducci, S., & Ciappi, S. (2008). Profiling hackers: the science of criminal profiling as applied to the world of hacking. Auerbach Publications.

- Cohen, D. & Bar'el, D. (2017). The Use Of Cyberwarfare In Influence Operations, Yuval Ne'eman Workshop for Science, Technology and Security.
- Dion-Schwarz, C., Ryan, N., Thompson, J. A., Silfversten, E., & Paoli, G. P. (2018). Olympic-Caliber Cybersecurity: Lessons for Safeguarding the 2020 Games and Other Major Events. Rand Corporation.
- EuObserver. (2015). Eric Maurice. Cyber attack on French TV finds EU unprepared. Available at:  
<https://euobserver.com/news/128285> [Accessed 20 Jan. 2019]
- Goman A. (2018) How Weaker Nations Are Taking Cyber Warfare Advantage. The world reporter. Available at:  
<http://www.theworldreporter.com/2018/08/how-weaker-nations-are-taking-cyber-warfare-advantage.html>  
[Accessed 20 Jan. 2019]
- Halder, D., & Jaishankar, K. (2012). Cyber crime and the victimization of women: laws, rights and regulations. Hershey, PA: Information Science Reference.
- Holzer C.T. and Lerums J.E. (2016). The Ethics of Hacking Back. Published in: 2016 IEEE Symposium on Technologies for Homeland Security (HST). Date of Conference: 10-11 May 2016, Waltham, MA, USA. IEEE. Available at:  
<https://ieeexplore.ieee.org/document/7568877> [Accessed 20 Jan. 2019]
- Hunker, J., Hutchinson, B., & Margulies, J. (2008). Role and challenges for sufficient cyber-attack attribution. Institute for Information Infrastructure Protection, 5-10.
- Jaishankar, K. (Ed.). (2011). Cyber criminology: exploring internet crimes and criminal behavior. CRC Press.
- Kaspersky. (2016). Threat Actors Master 'False Flags' Tactics to Deceive Victims and Security Teams. Available at:  
[https://www.kaspersky.com/about/press-releases/2016\\_false-flags](https://www.kaspersky.com/about/press-releases/2016_false-flags) [Accessed 20 Jan. 2019]
- Liang, Q., & Xiangsui, W. (1999). Unrestricted warfare. PLA Literature and Arts Publishing House Arts.
- Long, Larisa April. Profiling hackers. SANS Institute Reading Room, 2012, 26. Jg.
- NATO CCDCOE. (2015). Mitigating Risks arising from False-Flag and No-Flag Cyber Attacks. Available at:  
<https://ccdcoc.org/sites/default/files/multimedia/pdf/False-flag%20and%20no-flag%20-%202020052015.pdf> [Accessed 20 Jan. 2019]
- Oakton, M. (2016). Autopsy of Cyber Attack on TV5Monde. Available at:  
<https://www.infosecpartners.com/newsroom/2016/10/10/autopsy-cyber-attack-tv5monde/> [Accessed 20 Jan. 2019]
- Observer. (2016). John R. Schindler. False Flags: The Kremlin's Hidden Cyber Hand. Available at:  
<https://observer.com/2016/06/false-flags-the-kremlins-hidden-cyber-hand/> [Accessed 20 Jan. 2019]
- Ponemon Institute. (2015). The Rise of Nation State Attacks. Journal of Law & Cyber Warfare, 4(3), 1-42.
- Rid, T., & Buchanan, B. (2015). Attributing cyber attacks. Journal of Strategic Studies, 38(1-2), 4-37.
- Schwartz, M. J. (2017). French Officials details 'Fancy Bear' hack on TV5Monde Available at:  
<https://www.bankinfosecurity.com/french-officials-detail-fancy-bear-hack-tv5monde-a-9983> [Accessed 20 Jan. 2019]
- Sullivan, C. (2015). The 2014 Sony Hack and the Role of International Law. J. Nat'l Sec. L. & Pol'y, 8, 437.
- Tran, D. (2018). The Law of Attribution: Rules for Attribution the Source of a Cyber-Attack. Yale JL & Tech., 20, 376.
- TrendMicro. (2015). TV5 Monde, Russia and the CyberCaliphate. Available at: <https://blog.trendmicro.com/tv5-monde-russia-and-the-cybercaliphate/> [Accessed 20 Jan. 2019]
- TrendMicro. (2016). Operation Pawn Storm: Fast Facts and the Latest Developments. Available at:  
<https://www.trendmicro.com/vinfo/us/security/news/cyber-attacks/operation-pawn-storm-fast-facts> [Accessed 20 Jan. 2019]
- Turvey, B. E. (Hg.). Criminal profiling: An introduction to behavioral evidence analysis. Academic press, 2011.

# **Operational Risk Assessment on Internet of Things: Mitigating Inherent Vulnerabilities**

**Youngjun Park, Mark Reith and Barry Mullins**

**Air Force Institute of Technology, Wright-Patterson AFB, USA**

[youngjun.park@afit.edu](mailto:youngjun.park@afit.edu)

[mark.reith@afit.edu](mailto:mark.reith@afit.edu)

[barry.mullins@afit.edu](mailto:barry.mullins@afit.edu)

**Abstract:** Internet of Things (IoT) is a relatively new term attributed to the wave of technologies that connect to the Internet to provide connectivity and remote access to users. However, this seemingly convenient new capability brought with it an influx of security vulnerabilities that provide new points of entries for potential adversaries. One of the most infamous attacks on IoT was the Mirai botnet, which caused one of the largest and most disruptive Distributed Denial of Service (DDoS) attacks. Unfortunately, even in the aftermaths of the attack, the 351-billion-dollar industry (as of Jan 2018) continues to manufacture IoT with a myriad of security flaws without strictly enforced guidelines. Consequently, there has been numerous attempts to highlight the different security vulnerabilities associated with IoT. In a 2017 report, the U.S. Department of Defense identified multiple IoT-related risks including potential exploitations from supply chain, limited encryption as well as poor built-in security of the systems. However, there is still limited research in terms of their operational impact in the network. With countless systems currently deployed in critical environments such as the U.S. government, medical facilities, and critical infrastructure, deeper investigation of these vulnerabilities in their operational context are warranted. Here we present a preliminary analysis of IoT systems' operational risk factors based on the current methodologies of assessing security risks, and propose policies on their acquisition and proper use for organizations that employ the systems to help mitigate the risks discussed. We assert that an assessment of the operational risk in conjunction with the security vulnerabilities is necessary in order to fully capture the potential effects of the integration of IoT in an organization. Finally, we conclude with a discussion of future directions in research that will help visualize the risks and implications in IoT-saturated networks.

**Keywords:** internet of things (IoT), information security, operational risk assessment, policy guidelines

---

## **1. Current state of IoT systems**

The infiltration of Internet of Things (IoT) into the everyday lives of global citizens is increasing at an exponential rate. The 351-billion-dollar industry (CTA, 2018) is projected to reach 3.9 trillion by the year 2025 (Manyika et al., 2015). Unfortunately, this poses a significant threat to cybersecurity as many of these devices are often shipped with innate vulnerabilities. Furthermore, because of the rapid development of the industry and its heterogeneity, establishing global guidelines to address the vulnerabilities became challenging for policymakers (Weber, 2012). Limited progress has been made in sectors that deploy IoT, such as the U.S. government, to identify the vulnerabilities and enact guidelines to mitigate them. In the 2017 report from the Government Accountability Office (GAO) on IoT, the Department of Defense (DoD) identified multiple risks associated with the devices including possible exploitations from supply chain, limited encryption, as well as poor built-in security of the devices (GAO, 2017). Consequently, in its 2018 report, the GAO determined establishing a cybersecurity strategy for IoT as one of the most critical challenges for the DoD (GAO, 2018). Additionally, the former report concluded that the DoD's policies on cybersecurity and information systems security currently do not address many security concerns pertaining to specific IoT devices nor does it effectively implement the existing policies.

Despite the ongoing research in identifying various security concerns and risk factors associated with IoT, many of the existing methods for evaluating cybersecurity risks such as NIST SP800-30, OCTAVE, and CRAMM seem to fail to account for the rapidly evolving dynamics and ever-complex interactions within IoT-enabled network (Nurse, Creese and De Roure, 2017). Because of this difficulty in concretely attributing cybersecurity risk factors to IoT, there has been limited research efforts associated with assessing their impact in an operational setting.

## **2. CIAA**

Discussions of information security in literatures often revolve around the concepts of confidentiality, integrity, availability, and authenticity, or variations thereof (Farooq et al., 2015) (Gordon and Loeb, 2002) (Von Solms and Van Niekerk, 2013). Compromise of any of these four areas could lead to an adverse effect in system security. Our subsequent discussion of IoT security will revolve around these four concepts. The CIAA model as it relates to IoT is briefly discussed below:

## **2.1 Confidentiality**

In layman's term confidentiality in cybersecurity protects information from unwanted eyes. As the device could collect and store private information, any data stored in a system should be inaccessible to unauthorized parties. One of the main techniques of protecting confidentiality of data is encryption. However, the use of encryption requires the expense of extra resources for the user or the developer and storage of encryption keys.

## **2.2 Integrity**

A device with data Integrity should prevent unwanted changes to the information or alert when a change has been made. Any potential adversaries could attempt to make changes to the system for their gain. Compromise of data integrity can potentially lead to misrepresentation of information, degraded service, and even system failure. Integrity of data can be maintained through methods such as version control, which keeps track of the state of the device at a point in time.

## **2.3 Availability**

Availability pertains to ensuring accessibility of information for the users. One of the most common forms of attack on IoT devices is Distributed Denial of Service (DDoS) attacks (Angrishi, 2017). DDoS attacks flood the network with an influx of requests to overwhelm the system, denying service to the actual users. Failsafe mechanisms such as redundant servers or back-ups can help maintain availability.

## **2.4 Authenticity**

Authenticity verifies that the data, user, and other entities are indeed who they say they are. An attack on authenticity could exploit patching functions on the device to install malicious code or take control of the system. Verification methods of authenticity include strong password management and access control.

# **3. Vulnerabilities found in IoT**

In an attempt to understand better the implications of the vulnerabilities found in IoT, we discuss them in context of the CIAA model. This is by no means an exhaustive list of all security risks associated with IoT. Rather it attempts to classify them in broad categories to help determine the type of vulnerability to easily tie into the corresponding operational impacts. Different classifications have already been proposed such as that of Open Web Application Security Project (OWASP) (OWASP, 2018) and that of (Farooq et al., 2015). However, these classifications are often overly technical and do not explain their operational impact from a risk management standpoint. Our classification illustrated in Figure 1 attempts to address this issue by creating generalized groupings that can easily tie into discussion of their effect in an operation.

## **3.1 Insecure storage and communication**

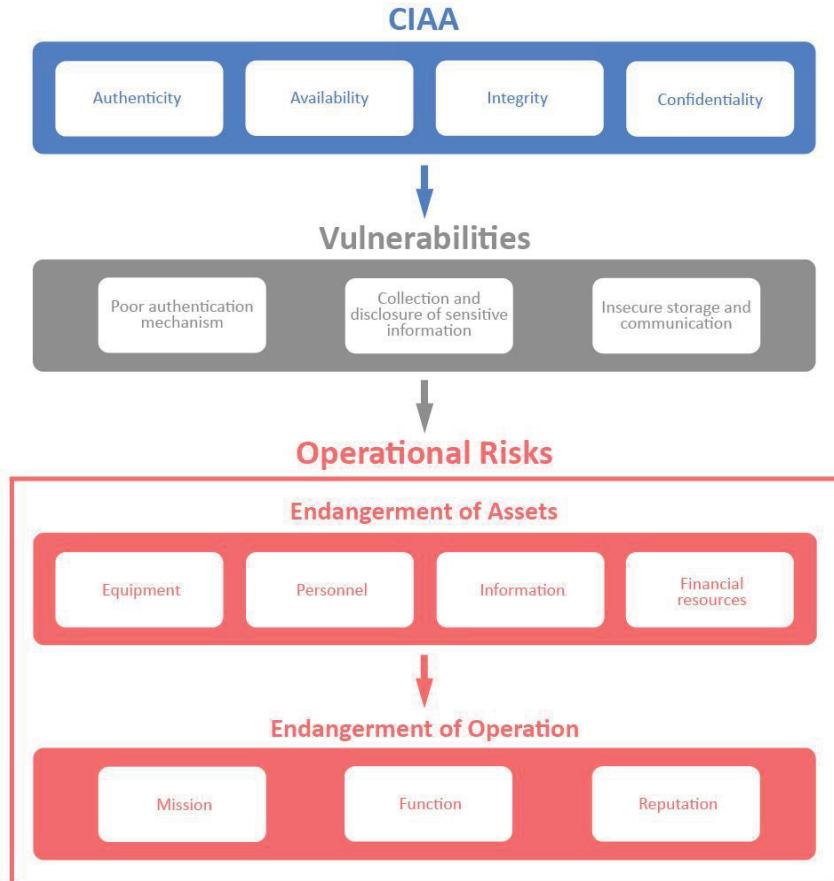
One core problem identified in IoT devices can be attributed to insecure information storage and communication. In a 2014 report by HP, 70 percent of IoT devices used in its study were found to have used unencrypted network service (HP, 2014). This implies that all of the information being transported across the network in the majority of the devices were available in plaintext for an eavesdropper to see.

Similarly, the physical devices themselves often come with limited encryption, if at all. This can be attributed to the limited storage and computing power of IoT devices (GAO, 2017). Thus, the devices often cannot support the computationally intensive encryption algorithms like RSA and AES, requiring dedicated lightweight algorithms such as that of (Usman et al., 2017). Furthermore, because manufacturers are not incentivized to build secure hardware, as the manufacture costs of hardware continue to drop (Manyika et al., 2015), functionality will likely be prioritized over security to out-edge their competitors.

The limited-encryption-supported devices enable the following for potential adversaries: (1) gain access to information stored or being transported over the network; (2) falsify information to their advantage; (3) deny access of proper services for the rightful users. While encryption seems like the obvious answer to ensure data confidentiality, encryption is also used to ensure data integrity and authenticity over a network using digital signatures, and public key encryption (Kurose and Ross, 2017). Therefore, ensuring proper encryption can mitigate many of IoT's vulnerabilities.

Cloud services such as Google and Amazon currently offer services making use of its effectively boundless resources to mitigate the limited hardware capacity of IoT devices to allow for strong encryption standards (Google Cloud, 2018) (Amazon Web Services, Inc., 2018). However, cloud-enabled IoT networks are not without its flaws. Any information in transit during cloud communication are susceptible to be sniffed and tampered with and are vulnerable to the security flaws present in the service providers (Sajid, Abbas and Saleem, 2016). Moreover, the use of cloud services for critical information systems leads to additional problems that stem from the lack of control of third-party services which will be discussed in Section 3.2.

## Operational Risk Assessment of Internet of Things



**Figure 1:** Classification of operational risk assessment of Internet of Things

### 3.2 Collection and disclosure of sensitive information

Perhaps a more appropriate discussion in an operational context is the type of data the system stores. Many of the IoT applications, such as those integrated in medical devices (Dimitrov, 2016) and in Supervisory Control and Data Acquisition (SCADA) systems (Sajid, Abbas and Saleem, 2016), may collect sensitive organizational information. In addition, because much of operational data collected from IoT devices are deemed proprietary, the users may not realize what information is being collected (Manyika et al., 2015). The information is then shared among different devices over the network and are vulnerable to unauthorized access. This has important implications in privacy of proprietary information and a deeper investigation into its regulations is needed.

One of the main concerns when dealing with sensitive information such as PII of employees and classified organizational data is confidentiality. For example, a medical device implanted on a patient may collect the physical state of the patient, and a wearable fitness device may collect the pattern-of-life information, including the typical schedule of the user. A recent study at the Air Force Institute of Technology (AFIT) illustrated that by only using the publicly available information included in the link-layer header of the packets, an adversary was able to sniff a network of over-the-counter IoT devices to accurately collect pattern-of-life information (Beyer, 2018). Even if the data stored on the device as well as the communication medium is properly secured, an adversary is able to use the collective information from a network of IoT to infer sensitive information.

Furthermore, many of the systems that connect to IoT devices were not designed to be exposed to the outside world. As a result, the Internet connectivity brought with it vulnerabilities that were nonexistent before in a closed system. Examples of such systems include the infotainment systems in vehicles (Miller and Valasek, 2015), and the IoT being used in SCADA systems (Sajid, Abbas and Saleem, 2016). The vulnerabilities found in the 2015 study led to the recall of 1.4 million vehicles (Greenberg, 2015), and served as a wake up call on the dangers of blindly incorporating additional functionalities without security in mind.

### **3.3 Poor authentication mechanism**

A recurring problem identified in OWASP's IoT vulnerability report was the devices' weak password management (OWASP, 2015). They are often shipped and even operate with default credentials. In fact a preliminary scan of IoT devices being used at one of the largest international organizations known as Conseil Européen pour la Recherche Nucléaire (CERN), found that many of its devices were running with default credentials such as "admin:admin" (Lueders, 2017). If authentication is improperly managed, with plenty of tutorials available through a simple search on the Internet such as (Medium, 2018), any novice hacker could gain access to mission critical systems in an organization.

Interestingly, most of the IoT devices, cloud services, and applications were found to allow weak passwords such as "1234" and some even prevented the user from setting strong passwords (GAO, 2017) (HP, 2014). A recent study by Angrishi highlighted that much of the malwares targeting IoT exploited these weak credentials supplied via brute force attacks (Angrishi, 2017). The malwares mentioned in the study including the infamous Mirai botnet demonstrate that the weakness in authentication mechanisms can lead to compromise of entire networks of systems, and its effect propagates on to all aspects of the CIAA paradigm. Even if the vulnerabilities in Sections 3.1 and 3.2 are adequately addressed, insufficient authentication mechanisms can compromise all information stored in the network.

## **4. Operational risks**

We broadly classify operational risks into two categories: endangerment of assets and endangerment of operation. Assets in an organization can include equipment, personnel, information, and financial resources; operations include its mission, functions, and reputation. The vulnerabilities identified in Section 3 will be applied in context of the two risk categories to help understand the overall impact of IoT. This two-tier classification allows an organization to first identify the risks associated with its resources then determine the impact they can have in carrying out its goal. Furthermore, this allows organizations to integrate the commonly used risk management frameworks such as CRAMM, NIST 800-30, and OCTAVE that relies on identifying the threats and vulnerable assets (Nurse, Creese and De Roure, 2017).

### **4.1 Endangerment of assets**

The risks surrounding organizational assets revolve around the confidentiality and the integrity of information. With inherently unsafe storage and communication of IoT, any information stored on the devices should be deemed vulnerable. An adversary can eavesdrop on the communication channel to gain access to organizational data, perform spoofing attacks, and even inject malicious code (Farooq et al., 2015). This can compromise the integrity of the organizations' equipment, degrading its performance or denying its service. The attacker could also gain access to the employees' information including their daily schedule and personally identifiable information (PII), leading to endangerment of personnel. If the individual is an important figure in an organization, their compromise could have serious implications for its future.

IoT devices may knowingly and unknowingly collect sensitive information about the organization. A device specification may correlate with obvious associated risks such as insider threats from taking photos and recording audio, but identification of inferred information can be difficult. Even though the standalone information itself may not be important, the aggregate collection of information from different sources can lead to disclosure of sensitive information. The aforementioned AFIT study and Strava's fitness tracker application report (Liptak, 2018) are good examples of such case. Strava's application uses GPS information to determine exercise patterns of the user. However, an analyst was able to show that by cross referencing publicly available maps and the locational data from the application, undisclosed information regarding the U.S military installations such as security guards' patrol routes could be inferred. Unfortunately, identification of these

operational risks can be overlooked while analyzing the device's technical vulnerabilities alone. How the device is used in what setting are important factors to consider to fully capture the associated risks.

Because of the scale of the network and the information being exchanged, proper management of authentication mechanisms of IoT devices and network becomes increasingly important in an organizational setting. Easily exploitable authentication enables adversaries to take control over entire assets. In the case of organizations that rely on cloud-based systems, gaining access to the cloud network could mean gaining access to the organizations' entire operation.

It is worth mentioning that over-the-counter IoT devices are especially vulnerable. Manufacturers and developers can easily install a backdoor such that they can access the system in the future. Because many over-the-counter devices are developed by entities all over the world, integrity of the hardware and the software is often not guaranteed. A recent investigation found that China had installed tiny chips in server motherboards during manufacturing to gain backdoor access to major U.S companies including Apple and Amazon (Robertson and Riley, 2018). Although this was a large-scale attack backed by a nation state, it nonetheless demonstrates the ability of manufacturers to take advantage of the supply chain.

#### **4.2 Endangerment of operation**

The escalation of risks discussed in endangerment of assets can ultimately lead to endangerment of operation. The degree of damage done to an organization's equipment, financial resources, information, and personnel etc. and their combined effect impact its ability to carry out its mission.

Compromise of information can have a range of effect on an organization's operation. Assessing its impact requires the inspection of its importance to the organization. On one hand organizational information such as marketing tactics of a sales company, and operations orders of a military unit could adversely impact their functions and potentially lead to mission failures. On the other hand, information that could damage an organization's reputation could be made available to the public or even made up. Depending on the type of information, the organization can entirely lose the public's trust, which can in turn hinder its performance. Therefore, maintaining the security of mission critical information is paramount to an organization.

### **5. Current problems in risk assessment for IoT devices**

While there is a push to instate a security standard for IoT devices being manufactured such as those recommended by OWASP, enforcing these standards are proving to be a difficult feat. Because many of the devices are often developed and manufactured overseas from countless nations and businesses, establishing a universal guideline to carefully monitor their global distribution remains a difficult challenge (Weber, 2010).

One of the main problems surrounding IoT is their interoperability. Many studies including (Miorandi et al., 2012) and (Nurse, Creese and De Roure, 2017) recognize and explore its importance in IoT's functionality and impact in the network. Additionally, the variability of connectivity of IoT devices and their responses to these new relationships make it difficult for current risk management tools including CRAMM, NIST 800-30, and OCTAVE to fully capture the threat model derived from a linear framework (Nurse, Creese and De Roure, 2017).

On top of the technical standards, new approaches that adapt with the dynamic interoperability of IoT devices must be established and enforced to promote a more secure IoT-enabled network. It is no longer sufficient to address the individual security vulnerabilities of IoT devices. In the rapidly changing and interconnected environment, development of operational risk management practices is paramount. The AFIT study as well as the Strava analysis discussed in previous sections are prime examples.

As outlined in the GAO's 2017 report, the United States DoD currently employs policies to address the vulnerabilities in cyberspace such as the DoD instruction on Cybersecurity as well as the risk management framework for DoD information Technology. However, most of these policies do not adequately address the issues specific to IoT (GAO, 2017). To aid in development of organizational practices, preliminary policy recommendations for organizations are proposed in Section 6.

## **6. Policy recommendations**

There are limited regulations provided by the U.S. government that help protect the information stored on mobile devices such as the Stored Communications Act (Legal Information Institute, 2002). Many of these restrictions do not apply to the service providers who have ownership of the information on their systems. By using third-party services, users are essentially entrusting their privacy to third-party organizations. Because of the lack of national regulations to ensure privacy of electronic information, organizations must develop secure practices within their own network. We provide below several policy recommendations to serve as a strong foundation for the development of other organizational practices.

**(1) Encryption of devices:** Encryption of all devices in use can minimize the attack surface from an attacker. Organizations should seek to acquire devices that support strong encryption and avoid using over-the-counter devices if possible. While it is easy to mandate encryption of devices supplied by an organization, many organizations allow work to be performed on personal devices (e.g. laptops). If personal devices are to be used for work, encryption of those devices should be mandated and supported by the IT department.

**(2) Separation of sensitive information:** The storage and access of differing classification of information should be separated as much as possible. In other words, highly sensitive information should only be accessible to certain devices and separately stored in a highly secure location. This will prevent sensitive information from accidentally being accessed by unauthorized users.

**(3) Strong authentication mechanism:** Requiring strong authentication mechanisms can solve many of the common exploits associated with IoT. Use of two-factor authentication and password managers on top of strong passwords is able to further secure the devices. Proper handling and management of passwords are also required to prevent leakage of credentials.

**(4) Least privilege and scope:** Devices should use the minimum amount of privilege and scope to perform their tasks. This means that devices should only be connected and interact with those required to perform their jobs, and only have the functionalities they need through whitelisting. This can be controlled by only allowing the use of specific devices, or on a network level though the use of software-defined networking (SDN) within the organization to only allow registered devices to communicate with certain entities.

**(5) Monitoring supply chain:** Acquisitions of the devices used in an organization should be carefully monitored to ensure that they are manufactured from a reputable source. It should also be ensured that they incorporate sufficient security measures. Connection of unauthorized devices into the organizational network should be banned.

**(6) Limited use of personal devices:** The use of personal devices such as wearable fitness trackers should be vetted based on their functionality. For example, as illustrated in the Strava's application case, GPS-enabled devices should not be allowed in operations that require geospatial confidentiality. Any devices that has physical or digital recording functions such as a voice recorder, camera, or external drives should be banned from facilities that deal with sensitive information. There is no master list that works for all possible settings; organizations should seek to install policies that fit their mission by exploring case-by-case scenarios.

## **7. Areas of future research**

There are two areas of interest that can be helpful in studying the interoperability of IoT: machine learning and game theory. Machine learning continues to be a hot topic in multiple disciplines including network security. The prediction capability of machine learning could aid in modeling the potential threats based on the varying interactions of IoT. Incorporating machine learning into cybersecurity and IoT have been explored in many literatures including (Ullah and Babar, 2018) and (Abie and Balasingham, 2012). However, because certain algorithms are more suited for certain scenarios, optimal methods should be researched to cater to each organizations' needs. In addition to the traditional intrusion/anomaly detection, machine learning could help better identify the risk factors that stem from IoT.

Because the interactions of IoT devices are often decentralized, using elements of cooperative game theory could help derive and promote certain interactions. Areas where game theory could be used to solve problems identified IoT usage include resource sharing (Brooks and Goss, 2013) (Miorandi et al., 2012) and context aware

devices (Kortuem et al., 2010). Resource sharing problem arises when untrusted devices attempt to share an object. Game theory allows the devices to compete for an object and come up with different strategies based on the varying interactions. It could also allow devices to coordinate different classification levels and only exchange relevant information. Similarly, context aware devices allow devices to behave a certain way based on the surrounding environment. The environmental factors may include other devices, policies, and network configuration. Integration of game theory could potentially facilitate the enforcement of organizationally-catered policies for IoT.

Integration of both machine learning and game theory are promising areas of research for IoT risk assessment. The predictability of machine learning as well as game theory's ability to capture varying interactions between agents could potentially account for the interoperability of IoT. Once the network of affected devices as well as the associated vulnerabilities are identified, the newly derived set of input can be fit into traditional risk management frameworks.

## **8. Conclusion**

Many studies highlighted in this paper have shown numerous vulnerabilities associated with IoT and the need to address these issues. While numerous categorizations have already been proposed, they are often overly technical and are not easily extendable to evaluate the operational risks involved. The proposed classifications of vulnerabilities are: (1) insecure storage and communication, (2) collection and disclosure of sensitive information, and (3) poor authentication mechanism. These vulnerabilities can be applied in context of the two-tier classification of operational risks: (1) endangerment of assets, which can escalate to (2) endangerment of operation.

While there are policies that address issues in cybersecurity, the lack of international guidelines for IoT make it difficult to enforce a security standard. Furthermore, many organizational policies fail to account for the interoperability of IoT. Our recommended policies are designed to serve as a foundation for organizations to develop their own practices.

Future areas of research include integration of machine learning and game theory into risk management pipelines. Incorporating these into the traditional risk management frameworks could provide a much more accurate assessment.

We believe this study will help organizations understand the operational impact of IoT-enabled networks and implement secure organizational practices to protect their network from the vulnerabilities introduced by IoT.

**Disclaimer:** The views expressed are those of the author and do not reflect the official policy or position of the US Air Force, Department of Defense or the US Government.

## **References**

- Abie, H. and Balasingham, I., 2012, February. Risk-based adaptive security for smart IoT in eHealth. In *Proceedings of the 7th International Conference on Body Area Networks* (pp. 269-275). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).
- Amazon Web Services, Inc. (2018). *AWS IoT Device Defender*. [online] Available at: <https://aws.amazon.com/iot-device-defender/> [Accessed 8 Dec. 2018].
- Angrishi, K., 2017. Turning internet of things (iot) into internet of vulnerabilities (iov): IoT botnets. *arXiv preprint arXiv:1702.03681*.
- Beyer, S.M., 2018. *Pattern-of-Life Modeling Using Data Leakage in Smart Homes* (No. AFIT-ENG-MS-18-M-009). Air Force Institute of Technology Wright-Patterson AFB OH United States.
- Brooks, J. and Goss, J. (2013). *Security Issues and Resulting Security Policies for Mobile Devices*. Masters. Naval Postgraduate School.
- CTA (2018) *2018 Tech Industry Revenue to Reach Record \$351 Billion, Says CTA* [Online]. Available at: [https://www.cta.tech/News/Press-Releases/2018/January/2018-Tech-Industry-Revenue-to-Reach-Record-\\$351-Bi.aspx](https://www.cta.tech/News/Press-Releases/2018/January/2018-Tech-Industry-Revenue-to-Reach-Record-$351-Bi.aspx) [Accessed: 4 Nov 2018]
- Dimitrov, D.V., 2016. Medical internet of things and big data in healthcare. *Healthcare informatics research*, 22(3), pp.156-163.
- Farooq, M.U., Waseem, M., Khairi, A. and Mazhar, S., 2015. A critical analysis on the security concerns of internet of things (IoT). *International Journal of Computer Applications*, 111(7).

- GAO (2017) *Internet of Things: Enhanced Assessments and Guidance Are Needed to Address Security Risks in DOD* [Online]. Available at: <https://www.gao.gov/assets/690/686203.pdf> [Accessed: 4 Nov 2018]
- GAO (2018) *HIGH-RISK SERIES: Urgent Actions Are Needed to Address Cybersecurity Challenges Facing the Nation* [Online]. Available at: <https://www.gao.gov/assets/700/694355.pdf> [Accessed: 4 Nov 2018]
- Google Cloud. (2018). *Device Security*. [online] Available at: <https://cloud.google.com/iot/docs/concepts/device-security> [Accessed 8 Dec. 2018].
- Gordon, L.A. and Loeb, M.P., 2002. The economics of information security investment. *ACM Transactions on Information and System Security* (TISSEC), 5(4), pp.438-457.
- Greenberg, A. (2015). *After Jeep Hack, Chrysler Recalls 1.4M Vehicles for Bug Fix*. [online] WIRED. Available at: <https://www.wired.com/2015/07/jeep-hack-chrysler-recalls-1-4m-vehicles-bug-fix/> [Accessed 9 Dec. 2018].
- HP (2014). *Internet of Things Research Study*. [online] Hewlett-Packard Development Company, L.P. Available at: [http://go.saas.hpe.com/l/28912/2015-07-21/32bhv3/28912/69168/IoT\\_Report.pdf](http://go.saas.hpe.com/l/28912/2015-07-21/32bhv3/28912/69168/IoT_Report.pdf) [Accessed 26 Nov. 2018].
- Kortuem, G., Kawsar, F., Sundramoorthy, V. and Fitton, D., 2010. Smart objects as building blocks for the internet of things. *IEEE Internet Computing*, 14(1), pp.44-51.
- Kurose, J.F. and Ross, K.W., 2017. *Computer networking: a top-down approach*. 7<sup>th</sup> edn. New Jersey: Pearson.
- Legal Information Institute. (2002). *18 U.S. Code § 2701 - Unlawful access to stored communications*. [online] Available at: <https://www.law.cornell.edu/uscode/text/18/2701> [Accessed 10 Dec. 2018].
- Liptak, A. (2018). *Strava's fitness tracker heat map reveals the location of military bases*. [online] The Verge. Available at: <https://www.theverge.com/2018/1/28/16942626/strava-fitness-tracker-heat-map-military-base-internet-of-things-geolocation> [Accessed 16 Nov. 2018].
- Lueders, S. (2017). *Computer Security: IoTs: The Treasure trove of CERN*. [online] CERN. Available at: <https://home.cern/news/news/computing/computer-security-rots-treasure-trove-cern> [Accessed 6 Nov. 2018].
- Manyika, J., Chui, M., Bisson, P., Woetzel, J., Dobbs, R., Bughin, J. and Aharon, D. (2015). *The Internet of Things: Mapping the Value Beyond the Hype*. [online] McKinsey Global Institute. Available at: <https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/The%20Internet%20of%20Things%20The%20value%20of%20digitizing%20the%20physical%20world/The-Internet-of-things-Mapping-the-value-beyond-the-hype.ashx> [Accessed 26 Nov. 2018].
- Medium. (2018). *Hack a webcam with a smartphone*. [online] Available at: <https://medium.com/bugbountywriteup/hack-a-webcam-with-a-smartphone-fa8f57c692c5> [Accessed 9 Dec. 2018].
- Miller, C. and Valasek, C., 2015. Remote exploitation of an unaltered passenger vehicle. *Black Hat USA, 2015*, p.91.
- Miorandi, D., Sicari, S., De Pellegrini, F. and Chlamtac, I., 2012. Internet of things: Vision, applications and research challenges. *Ad hoc networks*, 10(7), pp.1497-1516.
- Nurse, J.R., Creese, S. and De Roure, D., 2017. Security risk assessment in Internet of Things systems. *IT Professional*, 19(5), pp.20-26.
- OWASP. (2015). *IoT Attack Surface Areas*. [online] Available at: [https://www.owasp.org/index.php/IoT\\_Attack\\_Surface\\_Areas](https://www.owasp.org/index.php/IoT_Attack_Surface_Areas) [Accessed 26 Nov. 2018].
- Roberton, J. and Riley, M. (2018). *The Big Hack: How China Used a Tiny Chip to Infiltrate U.S. Companies*. [online] Bloomberg Businessweek. Available at: <https://www.bloomberg.com/news/features/2018-10-04/the-big-hack-how-china-used-a-tiny-chip-to-infiltrate-america-s-top-companies> [Accessed 9 Dec. 2018].
- Sajid, A., Abbas, H. and Saleem, K., 2016. Cloud-assisted IoT-based SCADA systems security: A review of the state of the art and future challenges. *IEEE Access*, 4, pp.1375-1384.
- Ullah, F. and Babar, M.A., 2018. Architectural Tactics for Big Data Cybersecurity Analytic Systems: A Review. *arXiv preprint arXiv:1802.03178*.
- Usman, M., Ahmed, I., Aslam, M.I., Khan, S. and Shah, U.A., 2017. Sit: A lightweight encryption algorithm for secure internet of things. *arXiv preprint arXiv:1704.08688*.
- Von Solms, R. and Van Niekerk, J., 2013. From information security to cyber security. *computers & security*, 38, pp.97-102.
- Weber, R.H., 2010. Internet of Things—New security and privacy challenges. *Computer law & security review*, 26(1), pp.23-30.

# Cyber Security of Vehicle CAN bus

**Jouni Pöyhönen, Pyry Kotilainen, Janne Poikolainen, Janne Kalmari, Pekka Neittaanmäki  
University of Jyväskylä, Finland**

[jouni.a.poyhonen@jyu.fi](mailto:jouni.a.poyhonen@jyu.fi)

[pyry.kotilainen@jyu.fi](mailto:pyry.kotilainen@jyu.fi)

[janne.poikolainen@jyu.fi](mailto:janne.poikolainen@jyu.fi)

[janne.kalmari@jyu.fi](mailto:janne.kalmari@jyu.fi)

[pekka.neittaanmäki@jyu.fi](mailto:pekka.neittaanmäki@jyu.fi)

**Abstract:** There are currently many research projects underway concerning the intelligent transport system (ITS), with the intent to develop a variety of communication solutions between vehicles, roadside stations and services. In the near future, the roll-out of 5G networks will improve short-range vehicle-to-vehicle traffic and vehicle-to-infrastructure communications. More extensive services can be introduced due to almost non-delayed response time. Cyber security is central for the usability of the services and, most importantly, for car safety. The Controller Area Network (CAN) is an automation bus that was originally designed for real-time data transfer of distributed control systems to cars. Later, the CAN bus was developed as a universal automation system for many automation solutions. One of its characteristics is that bus traffic is not supervised in any way due to the lack of timing of control. In other words there are no authentication mechanism. This article highlights different approaches and their usability to reveal the car's CAN bus malfunctions. The study complements earlier studies on the safety of vehicles in the CAN bus. Based on the test results, practical methods can be evaluated to detect changes in CAN bus traffic, such as targeted cyber-attacks. The article is based on the results of a study on the cybersecurity of cars conducted at the University of Jyväskylä (AaTi study). Initially, the AaTi study attempted to identify the message content of the bus and to detect interferences via the Neural network solution. However, the problem with the neural network was the computational performance required and the lack of prediction accuracy. After that the study was focused on experiments that were based on the arrival times of control messages, that is, their timing-based intrusion detection. In this sense the research did concentrate on kernel density estimation, one-class support vector machine solution, absolute deviation method and categorization. Due to methodological challenges, a method for detecting intrusions based on statistical processing of message traffic was ultimately developed as an outcome of the study.

**Keywords:** cybersecurity, car, CAN bus, intrusion detection

---

## 1. Introduction

The term intelligent transport systems (ITS) refers to using roadside infrastructure and communication solutions for improving traffic flows and making traffic safer. In order to realize the perquisites for smart traffic, current national and international research projects are focusing on the development of platforms for weather, security and geolocation solutions. These include test environments for real-time road weather reports based on location data as well as for ITS cyber security. (Finnish Meteorological Institute, 2017)

Service usability is closely linked to cyber security, in which taking care of vehicle cyber security can be seen as a primary objective. CAN bus is a network solution originally developed for real-time communication in distributed automotive control systems, such as in engine control units, ABS brakes and drivetrains. (Alanen, 2000)

CAN bus later evolved as a general-purpose automation solution to accommodate other use cases in addition to automotive use. The real-time requirement makes minimizing network delays one of the main principles of CAN-bus functionality. This optimization also leads to design decisions that excluded many safety mechanisms, including authentication. These features make CAN bus implementations vulnerable to several types of attacks, including network traffic forgery, unauthorized access to data and denial of service attacks. As the growing use of automation means also the growing use of network connectivity, the attack surfaces in vehicles can be divided into two groups: surfaces that can be exploited remotely and surfaces requiring physical access. Because of development of intelligent transport systems and smart traffic the need of remote connections will grow even more in the future as ITS develops further. Vehicle network security research has emerged in past years, especially after the inherent vulnerabilities in commonly used technologies have been realized.

The purpose of this article is to present different approaches and their abilities to detect anomalies in vehicle CAN buses. Based on the results of this study, methods plausible for real-world scenarios are proposed. The ultimate goal of the study has been to develop real-time situational awareness methods for automation systems.

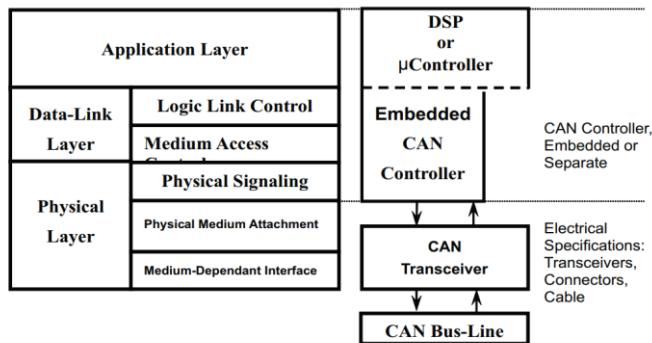
This report is based on a study (AaTi) conducted at the University of Jyväskylä, which concluded on 30 September 2018.

In addition to the chapters dealing with introduction and the CAN bus description, the paper includes a short description from other relevant vehicle studies and explanations from the methods used in the AaTi study to detect harmful bus traffic and the results obtained from their use. The conclusion chapter includes the summary of the AaTi study.

## 2. The CAN bus

### 2.1 The CAN Standard

The CAN communications protocol, ISO-11898: 2003, describes how information is passed between devices on a network and conforms to the Open Systems Interconnection (OSI) model, which is defined in terms of layers. Actual communication between devices connected by the physical medium is defined by the physical layer of the model. The ISO 11898 architecture defines the lowest two layers of the seven-layer OSI/ISO model as the data-link layer and the physical layer, shown in Figure 1 (Corrigan, 2016).



**Figure 1:** CAN bus in the OSI/ISO model

The application layer establishes the communication link to an upper-level application specific protocol such as the vendor-independent CANopen™ protocol. This protocol is supported by CAN in Automation (CiA), the international users and manufacturers group. Many protocols are dedicated to particular applications, such as industrial automation, diesel engines, or aviation. (Corrigan, 2016)

### 2.2 CAN message and frames

The four different message types, or frames (see Figure 2, CSS Electronics, 2018), that can be transmitted on a CAN bus are the data frame, the remote frame, the error frame, and the overload frame.



**Figure 2:** CAN bus message

CAN bus frames: (Corrigan, 2016)

The data frame

The data frame is the most common message type, and comprises the arbitration field, the data field, the CRC field, and the acknowledgment field. In Figure 2 the arbitration field contains a 29-bit identifier (or 11-bit identifier) and the RTR bit, which is dominant for data frames. Next is the data field, which contains zero to eight bytes of data, and the CRC field, which contains the 16-bit checksum used for error detection. The acknowledgment field is last.

#### The remote frame

The intended purpose of the remote frame is to solicit the transmission of data from another node. The remote frame is similar to the data frame, with two important differences. First, this type of message is explicitly marked as a remote frame by a RTR bit in the arbitration field, and second, there is no data.

#### The error frame

The error frame is a special message that violates the formatting rules of a CAN message. It is transmitted when a node detects an error in a message and causes all other nodes in the network to send an error frame as well. The original transmitter then automatically retransmits the message. An error mechanism in the CAN controller ensures that a node cannot tie up a bus by repeatedly transmitting error frames.

#### The overload frame

The overload frame is mentioned for completeness. It is similar to the error frame with regard to the format, and it is transmitted by a node that becomes too busy. It is primarily used to provide for an extra delay between messages.

### **2.3 CAN bus arbitration**

Arbitration is a mechanism for conflict resolution between network nodes. When the network path is free, any of the nodes in the network can start the message send process. If another node also wishes to send at the same time, the order of the transmissions is decided using a bitwise arbitration mechanism. During arbitration, both nodes start their transmission. The transmission starts with a start bit, followed by an id field (identifier, CAN-ID). The sending order decision is made based on the value of the id field and the other node or nodes discontinue their transmissions. The messages are sent ordered by priority, where the zero value is dominant. In practice this means that if a node currently sending a bit with a value of one sees that another node is sending a zero bit, it backs off. In other words, it discontinues its own transmission, forfeiting its turn to the node sending the dominating bit. In practice the message with the smallest decimal id value has the highest priority. (Johansson et al., 2005)

From the viewpoint of attacks, this mechanism enables denial of service attacks. As an example, sending large quantities of forged messages having an id value of zero.

### **2.4 CAN bus pros and cons**

CAN bus was designed for maximal speed and reliability. At the technical level this means, among other aspects, that the network communication uses a provider-consumer model instead of the common sender-receiver model. The second feature aiming for performance gains was the lossless bus arbitration described above. (Voss and Comprehensible, 2005)

Improving the reliability of the data transmitted between the nodes was achieved with a mechanism that insures the integrity and timeliness of the messages. These mechanisms are based on bus arbitration, using checksums checking the payload and resending failed messages. (Voss and Comprehensible, 2005)

Based on these design decisions, CAN bus is effectively a broadcast network, where any node can send a message and all nodes are listening to the network and reacting to the messages they are interested in. The only thing the recipients check is the protocol correctness of the received message. (Voss and Comprehensible, 2005)

CAN bus speed is 1 Mbit/sec, which these days does not seem fast. Yet for transmitting short messages and having an effective collision avoidance mechanism, CAN bus is more suitable to be used in real-time applications than connected protocols such as TCP/IP, even if those would be using greater transmission speeds. (Voss and Comprehensible, 2005)

With further development the CAN bus has become a dominant technology for the data transmission of vehicle basic functions. During the last two decades the number of electronic systems in vehicles has increased and at

the same time they have become more complex. CAN bus vulnerabilities can be traced back to design decisions described above, the most significant of these being the lack of authentication mechanism. The receiving entity does not have any mechanism to verify the origins of the received message or the validity of the data received. In other words, the control unit does not have a mechanism to detect message forgery. This characteristic makes vehicle CAN busses vulnerable to attacks, such as message forgery, unauthorized data use and denial of service. The DoS vulnerability can be exploited by sending a large number of high priority messages. These attacks can affect the vehicles systems in such a way as to cause loss of control, incorrect functionality, premature wear or rendering the vehicle unable to function at all. (Carsten et al., 2015)

## **2.5 Attack surfaces**

The taxonomy of CAN bus attack surfaces is usually divided into two parts: remote exploits and exploits requiring physical access to the CAN bus. In addition to this, some researchers have expanded the use of physical connections by constructing experiments that enable man-in-the-middle type of attacks on the CAN bus (Lebrun and Demay, 2016).

Physical connection to a CAN bus is not technically complex to achieve. The simplest physical connection can be implemented through the vehicle's diagnostics port. This approach does not require any alterations to the vehicle in question. The limitation of this approach is the amount of network data observable at this point of entry, depending heavily on the make and model of the vehicle. CAN bus traffic seen through the diagnostic port is restricted by segmenting the network. These limitations can be avoided by choosing another point of entry from the desired segment. In most cases this approach requires alterations to the vehicle's wiring harnesses, because segment-specific connectors are rarely implemented in production vehicles.

Remotely exploitable attack surfaces that would have a direct effect on the vehicle's physical functionalities are usually more challenging to exploit. In practice, this normally means a multistage attack where the attacker first has to find a vulnerable and remotely accessible service to gain a foothold. As an example, this kind of service can be found from the vehicle telemetry or infotainment systems. After gaining a foothold on one of the connected systems, the attacker needs to find a way to gain access to another system that has connectivity to the more critical segments of the vehicle's CAN bus. This type of attack has been successfully conducted by some vehicle security researchers (see Miller and Valasek, 2013).

## **3. The AaTi study**

### **3.1 Previous research**

Wolf et al. (2004) found that vehicle networks are open and for this reason vulnerable on many levels. The attacker can exploit vehicle wireless connections and networks. Wolf et al. (2007) continued their work in an article where they were attempting to form a full picture of the current situation of automotive electronic systems. This article listed commonly used automotive systems and their properties, including details about communication and cryptography.

The possibility of cyber-attacks as a subject of scientific articles became more prevalent around a decade ago. At that time, the articles started to touch on the subject of, among other things, how to protect vehicles for possible attacks (Larson et al., 2008).

A research group consisting of researchers from the University of Washington and the University of California, San Diego conducted a system security analysis through experiments on a passenger vehicle. This article was aiming for a comprehensive security analysis of a vehicle system rather than an analysis of individual devices. The article also proposed a part threat model that identified the physical connection and wireless functionalities as individual attack vectors. (Koscher et al., 2010) This group continued their work the next year by publishing an article, focusing on a broader analysis of the vehicle attack surfaces. (Checkoway et al., 2011)

Vehicle cyber security research was brought to more common knowledge by Valasek and Miller, who published their first article on this subject in 2013. In this article they examined two vehicles from different manufacturers and got results on how vehicle functionality can be affected that were similar to what previous academic research efforts had shown. In addition to their results, they published most of the reverse engineered CAN messages they discovered, and the source code of the tools used in their research. The additional information was

published to encourage other groups to conduct similar research in the future (Miller and Valasek, 2013). Miller and Valasek (2015) continued their work and published an article that describes in detail how an unaltered vehicle can be taken into partial control without a physical connection. As a point of entry to the vehicle system they used a security flaw in the infotainment system of the vehicle in question.

### **3.2 Analysis methods**

#### *3.2.1 Introduction to the analysis methods used in this research*

The anomaly detection methods proposed in previous academic articles can be divided into groups using several different taxonomies. The first example of such a taxonomy is dividing the methods based on the use of system specification. When system specification is available, detecting anomalies is based on detecting traffic that does not fit the given specification. This type of approach has been suggested in the method used by Larson et al. (2008), where anomaly detection is delegated to the network nodes. The nodes then inspect the traffic and sound an alarm if they see some else sending a message type only they are supposed to send. The second category of systems assumes that system and message specifications are present. In these systems the anomaly detection is based on features such as message timing, data semantics, entropy, repeating message sequences, protocol correctness and signal characteristics differing from normal network traffic.

Anomaly detection methods can also be grouped according to their method of detection. The majority of normal CAN bus traffic is cyclic by nature. This means that a series of messages repeat cyclically after very predictable intervals. Based on this property many proposed methods use message timing as a basis for anomaly detection. Time-based detection methods can also be divided into two main groups: those that measure message frequency and those that measure interval. Methods that fall into the first group have been proposed by Hoppe et al. (2008–2009), Münter et al. (2010) and Miller and Valasek (2014). Miller and Valasek (2014) proposed a substantial rise in the frequency of sent messages for a detection method. Taylor et al. (2015) proposed a more advanced method where the frequency monitoring is based on Hamming distance.

However, methods based on message frequency have their weaknesses. For example, when only message frequency is monitored short-term anomalies are not necessarily detected. However, if instead of frequency the detection method is based on message intervals, even short-term changes in network traffic can be detected more accurately. The use of interval analysis has been proposed by Son et al. (2016) and Moore et al. (2017). The latter article proposes a method based on absolute time deviation.

Several methods based on correctness of data carried by the messages have been proposed. Hoppe et al. (2008) described a method where only gross abuse of messages is detected. They continued their work in 2009 by proposing a method where consecutive messages are monitored for semantic correctness.

Münter and Asaj (2011) described a method based on data entropy, where changes in entropy are detected on the binary level. In addition, their method monitored communication entropy in protocol and signal levels.

Methods based on monitoring repeating message sequences in a specified time window has been proposed in several articles. Narayanan et al. (2015) based their method on a hidden Markov model and Marchetti et al. (2017) observed the order and changes in repeating messages.

The AaTi project used a test vehicle (Toyota Prius Hybrid). The data used in the study was first recorded from a test vehicle. It could then be inspected in laboratory environment. The vehicle-specific CAN bus interference messages were first generated under laboratory conditions and then verified on a test vehicle. The first goal of AaTi study was the ability to distinguish between normal and abnormal network traffic in real-time using recorded samples from a vehicle. In this research, a system of message specifications was not used, so anomaly detection of data payloads proposed a challenge. Mainly for this reason most of the methods that were researched were time based. This has also been the primary approach for previous research.

The only method during this research that was based on anomaly detection in data payloads was a neural network that could learn to predict incoming message payloads based on previous data it had inspected.

### 3.2.2 Neural network

For inspecting message data payloads, a neural network was built based on the method presented in Taylor et al. (2016). In this method the neural network builds a model based on normal data traffic by inspecting network traffic. The method described in the article has produced promising results. Different metrics for identifying and measuring deviations in the data streams have also been proposed in multiple articles (e.g., Taylor et al., 2015).

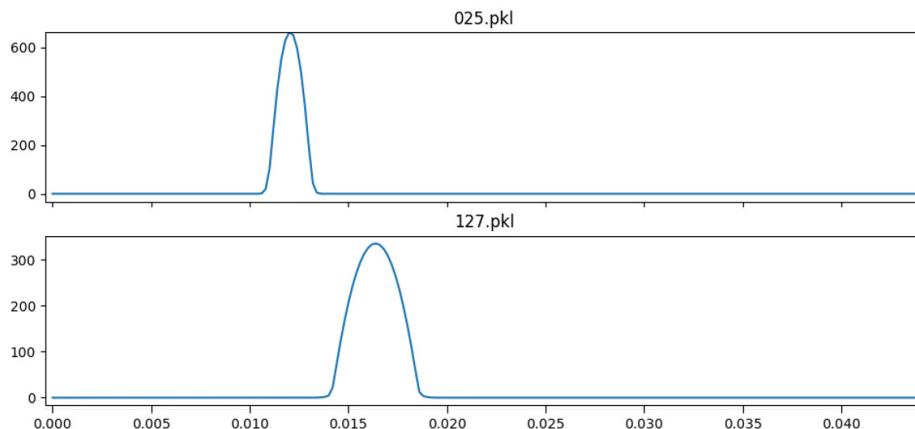
The Long Short-Term Memory (LSTM) network architecture used in this research consisted of layers of nodes with an adjustable feedback loop. This architecture enables the network to have a “memory” as well as the ability to “forget”. The majority of data in CAN bus traffic is regular by nature, in other words the data changes gradually and follows distinct trends. Therefore, predicting should be viable for at least some parts of the network traffic. In the experimental design the neural network was constructed to predict the data bits of an incoming message based on data bits of previously observed messages.

The biggest problems using the described neural network was its resource demand and probable difficulties in making the predictions more accurate. In addition, reasonable accuracy can only be achieved in regular data-flows, so some parts of the CAN bus traffic cannot be inspected using this method.

### 3.2.3 Kernel density estimation

The first method implemented for interval analysis was kernel density estimation. This method can model interval distribution characteristics for each message identifier. The distribution characteristics can then be compared to incoming message distribution to detect anomalies.

Modeled distribution gives the intervals a density function that can be used to calculate reliability values for new messages. If the calculated reliability drops too low, the situation is declared an anomaly and an alarm can be given.



**Figure 3:** Arrival interval distributions of two different message identifiers.

Figure 3 shows arrival interval distributions for two message identifiers that have been modeled using kernel density estimation. In the first graph, the interval deviates between 10 and 15 milliseconds. In the second graph, it deviates between 15 and 20. If incoming traffic shows interval deviations to be different than the peaks showed in the graphs, an indication of anomalous traffic can be given.

The advantage of kernel density estimation is that it can model systems that implement different sending speeds. For example, an engine control unit can send messages with different intervals when the engine is in idle or when the vehicle is moving. This kind of situation would show in the model as two distinct peaks. However, this kind of behavior was not observed in the test vehicle used in the study.

### 3.2.4 One-class support vector machine

One-class support vector machine (OCSVM) is a variation of a support vector machine, which is a popular machine learning method for classification. However, a normal support vector machine requires examples of each class that it should identify. An OCVSM, on the other hand, classifies the elements into two categories: normal

and abnormal. This means examples of normal behavior are sufficient and it does not require examples of abnormal behavior. The method defines a boundary around normal behavior and classifies all messages outside this boundary as abnormal.

Therefore, a one-class support vector machine is fit for detecting abnormal behavior, because it is challenging to find examples of all possible abnormal behaviors for training. Examples of normal behavior, on the other hand, are in most cases easily available. Normal data sets were recorded from test vehicle.

Taylor et al. (2015) presented an application of OCVSM, and the AaTi study implemented a variant of this machine. A moving window containing a set number of messages was used as a data element. From the data elements the characteristics were calculated that could be used to define the whole inspected window as either normal or abnormal. Characteristics calculation was based on mean interval and standard deviation of the messages within a window.

### *3.2.5 Absolute deviation*

Because the kernel density estimation method described above (see 3.2.3) is resource intensive and no multiple peaks were observed in the gathered data, a decision was made to implement a simplified method using the same principles. This method attempted to model the interval deviation for each message identifier, which would give considerable gains in performance and, at the same time, maintain similar performance for detection.

A decision was made to model the intervals with a normal deviation, because this made it possible to describe the deviation using only mean and standard deviation values. In the training phase it was also decided to include upper and lower bound values for each message identifier for classification. The distinct upper and lower bound values were added because positive deviations were more common in the normal network traffic, possibly due to message collisions. In addition, doing these calculations in the training phase made the classification faster. The boundary values were chosen so that no deviations were classified from the training data and a small marginal was added.

Again, a moving window was used for data-element as in the one-class support vector case. The messages in the window were classified based on its average interval. If this value was smaller than the lower bound or greater than the upper bound value then the traffic within the window in question was defined as abnormal.

An almost identical sensor was implemented in an article by Moore et al. (2017). The difference being that they did not use a moving window as a data element (Müter et al., 2010). An alert caused by a single abnormal message would produce too many false positives, so in the implementation described in the article only three consecutive abnormalities will trigger an alarm. In testing this method showed similar performance with other methods tested and it was less resource intensive. A decision was made to do a proof of concept implementation of this method.

### *3.2.6 Categorization*

Based on the previously described methods, it was observed that most false positives originate from control units that send their messages in irregular intervals. For this reason, the possibility of categorizing the messages by their send profiles was examined. Some of the control units send messages at regular intervals and others send irregular messages of events between regular status messages. A simple absolute deviation detection would categorize these messages as abnormal and initiate an alarm.

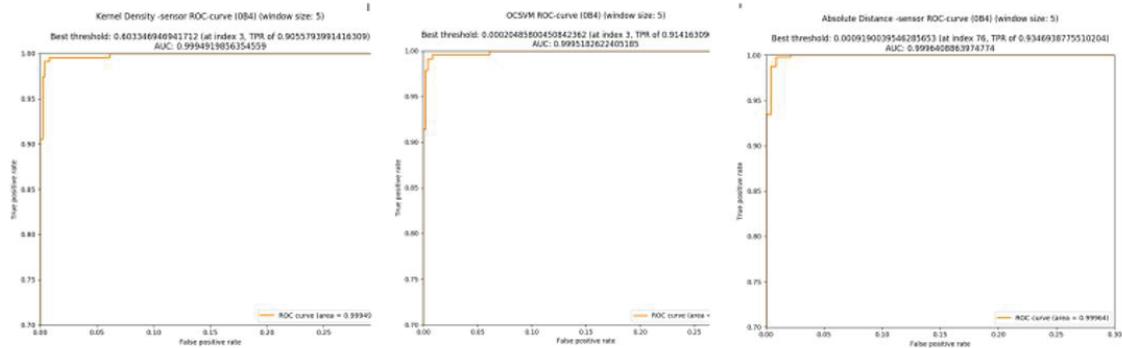
Based on the observations made during the research. It would be possible to reduce the number of false positives using categorization. But the number of send profiles would pose challenges for an implementation of such a categorization method. In addition, some messages that have the same message identifier can use multiple send profiles. In addition to these two drawbacks, there is uncertainty over what kind of send profiles exist in addition to the ones observed.

### *3.2.7 Method comparison*

Time-based methods were compared by drawing a receiver operating characteristic (ROC) curve for each individual method using the same data recorder from a vehicle CAN bus that included a test attack. All methods

used the same window size of five messages per window so that comparability could be maintained. All methods were given a setting that if they detected even one message that was part of the attack, they should mark the whole window as abnormal.

The best threshold value in Figure 4 shows the threshold value for which the best accuracy without any false positives was achieved. The number of true positives is shown in parentheses labeled “TPR” (True Positive Rate). The main interest in this figure should be the threshold value, since a practical real-time sensor would require a minimal number of false positives.



**Figure 4:** ROC graph (Vertical axis: true positive rate and horizontal axis: false positive rate)

Figure 4 left, kernel density estimation method; best achieved TPR without false positives: 0.9056. Figure 4 middle, OCSVM; best achieved TPR without false positives: 0.9142. Figure 4 right, absolute deviation method; best achieved TPR without false positives: 0.9347.

The comparison shows the analysis of a single message identifier attacked during the data recording. The results suggest that the performance of the methods shown is similar. At least with the test data were used in the comparison. The absolute deviation, which is also the simplest method, achieved the most accurate results, maintaining a zero false positive rate.

Based on the observations, the methods described above show that most false positives originate from electronic controller units with irregular send profiles. Message send profile categorization could be used to improve the result in these cases, but it has its drawbacks, as described in section 3.2.6.

#### 4. Conclusions

The focus of the AaTi study was to survey anomaly detection methods applicable to vehicle networks. This research complements previous research and patents by understanding network-traffic characteristics using recordings obtained from a test vehicle. The study shows that attacks against vehicle networks can be categorized into three groups. The network can be injected (a) with special messages such as diagnostics messages; (b) with normal messages that disturb vehicle functionality or (c) by sending normal messages after the real sender has been rendered unfunctional. The most common situation is probably when the real sender is still functional, and the attacker sends normal CAN messages. These kinds of attacks can be detected by observing message send intervals, since in a normal situation the intervals should remain regular.

In the first phase of research a neural network implementation was tested for its ability to detect abnormalities in message data payloads. The aim of this implementation was to provide technical means to learn different payload possibilities and predict the data incoming in the following messages. This would have created the possibility to detect abnormal data payloads. The problem with using neural networks arose from its resource intensiveness and lack of prediction accuracy. The next experiments focused on anomaly detection methods based on message timing.

The first time-based method we tested was One-Class Support Vector Machine (OCSVM), which is a variation of the popular machine learning method. This method defines boundaries around normal behavior and classifies all other traffic as abnormal. In the implementation, a moving window with a set number of messages was used as a data-entity. The characteristics of the messages are then calculated using OCSVM and, based on the results,

the whole window is declared normal or abnormal. The characteristics used in this implementation were average interval and standard deviation.

After this first experiment, other methods based on message interval were surveyed. Kernel density estimation models interval deviation for each message identifier. This value can then be compared to incoming messages in order to detect abnormalities. Modeled deviation provides a density function for the interval that can be used for likelihood value calculation for incoming messages. A drop in the calculated likelihood that exceeds a predetermined threshold can be detected as an anomaly and an alarm can be triggered. Because kernel density estimation is also a resource-intensive method and the observed test data did not show multipeak properties, a simplified version using the same principles of this method was implemented. This method aims to model message identifier deviation using key values. This implementation of absolute deviation achieves substantial gains in resource efficiency and without decline in the performance of the detection properties. The modeling was done using standard deviation in order to use the two key values: average and standard deviation. In the practical implementation training phase average, lower and upper bound values were calculated for each message identifier for classification purposes. A moving window was used as a data entity. If the values within the window went below the lower bound or exceed the upper bound, the whole window is declared an anomaly in the network traffic.

All of the above mentioned methods have their own challenges in either resource intensiveness, accompanied in some cases with inaccuracy of predictions.

Based on the experience described in the method comparison chapter, a novel method for detecting CAN bus anomalies based on message arrival intervals was developed and a patent application for this method has been filed. The description of this method is part of the patent. The functionality of this method was verified in a computational environment.

As different digital platforms become ever more common in automated processes, the protection of different processes and the cyber security of the infrastructure is going to play a significant role in the overall safety of these platforms. For future researchers in this field, the group would like to recommend the usage of outcomes found in the AaTi study as well as the utilization of the patented method as a part of future CAN bus implementations in order to improve cyber security.

## References

- Alanen J. (2000). CAN ajoneuvojen ja koneiden sisäinen paikallisyväylä. Tampere: VTT Automaatio, koneautomaatio.
- Carsten P., Yampolskiy M., Andel T.R. and McDonald J.F. (2015). In-Vehicle Networks: Attacks, Vulnerabilities, and Proposed Solutions. CISR '15 Proceedings of the 10th Annual Cyber and Information Security Research Conference (apr 2015), 477–482
- Checkoway S., McCoy D., Anderson D., Kantor B., Savage S., Koscher K., Czeskis A., Roesner F. and Kohno K. (2011). Comprehensive Experimental Analysis of Automotive Attack Surfaces, in Proceedings of the USENIX Security Symposium, San Francisco, CA.
- Corrigan S. (2016). Introduction to the Controller Area Network (CAN). Texas Instruments.  
<http://www.ti.com/lit/an/sloa101b/sloa101b.pdf>
- CSS Electronics (2018). A Simple Intro to CAN Bus. <https://www.csselectronics.com/screen/page/simple-intro-to-canbus/language/en>
- Finnish Meteorological Institute, (2017). Intelligent Transport. <https://ilmatieteenlaitos.fi/alykas-liikenne>
- Hoppe T., Kiltz S. and Dittmann J. (2008). Security threats to automotive CAN networks - practical examples and selected short-term countermeasures. In SAFECOMP.
- Hoppe T., Kiltz S. and Dittmann J. (2009). "Applying Intrusion Detection to Automotive It-Early Insights and Remaining Challenges." Journal of Information Assurance and Security (JIAS) 4 (6): 226–235.
- Johansson K. H., Törngren M. and Nielsen L. (2005), Vehicle applications of controller area network, in Handbook of Networked and Embedded Control Systems, William S. Levine Dmiitris Hristu-Varsakelis, and,ed., Birkhauser.
- Koscher K., Czeskis A., Roesner F., Patel S., Kohno T., Checkoway S., McCoy D., Kantor B., Anderson D., Shacham H. and Savage S. (2010). Experimental security analysis of a modern automobile. In D. Evans and G. Vigna, editors, IEEE Symposium on Security and Privacy. IEEE Computer Society.
- Larson, U. E., Nilsson D. K. and Jonsson E. (2008). "An Approach to Specification-Based Attack Detection for in-Vehicle Networks." In 2008 IEEE Intelligent Vehicles Symposium, 220–25. doi:10.1109/IVS.2008.4621263.
- Lebrun A. and Demay J. C. (2016). Canspy: a platform for auditing can devices. [https://www.blackhat.com/docs/us-16-materials/us-16-Demay-CANSPY-A-Platorm-For-Auditing-CAN-Devices.pdf](https://www.blackhat.com/docs/us-16/materials/us-16-Demay-CANSPY-A-Platorm-For-Auditing-CAN-Devices.pdf).

- Marchetti M. and Stabili D. (2017). "Anomaly detection of can bus messages through analysis of id sequences," in 28th IEEE Intelligent Vehicle Symposium (IV2017).
- Miller C. and Valasek C. (2013). Adventures in automotive networks and control units, DEFCON 21, Las Vegas, NV.
- Miller C. and Valasek C. (2014). A survey of remote automotive attack surfaces, BlackHat USA.
- Miller C. and Valasek C. (2015). Remote exploitation of an unaltered passenger vehicle, Black Hat USA.
- Moore M. R., Bridges R. A., Combs F. L., Starr M. S. and Powell S. J. (2017). "Modeling inter-signal arrival times for accurate detection of can bus signal injection attacks," in 12th CISRC. ACM.
- Müter M., Groll A. and Freiling F. C. (2010). "A Structured Approach to Anomaly Detection for in-Vehicle Networks." In 2010 Sixth International Conference on Information Assurance and Security (IAS), 92–98.  
doi:10.1109/ISIAS.2010.5604050.
- Müter M. and Asaj N. (2011). "Entropy-based anomaly detection for in-vehicle networks." IEEE IVS.
- Narayanan S. N., Mittal S. and Joshi A. (2015). "Using Data Analytics to Detect Anomalous States in Vehicles." arXiv Preprint arXiv:1512.08048. <http://arxiv.org/abs/1512.08048>.
- Song H. M., Kim H. R. and Kim H. K. (2016). "Intrusion detection system based on the analysis of time intervals of CAN messages for in-vehicle network," in 2016 International Conference on Information Networking (ICOIN), pp. 63-68
- Taylor A., Japkowicz N. and Leblanc S. (2015). "Frequency-Based anomaly detection for the automotive CAN bus," in Proc. of WCICSS, 2015, pp. 45–49.
- Taylor A., Leblanc S. and Japkowicz N. (2016). "Anomaly Detection in Automobile Control Network Data with Long Short-Term Memory Networks". IEEE DSAA (2016).
- Voss W. and Comprehensible A. (2005). Guide to Controller Area Network. Massachusetts, USA: Copperhill Media Corporation.
- [www.br-automation.com](http://www.br-automation.com)
- Wolf M., Weimerskirch A. and Paar C. (2004). Security in automotive bus systems. In Proceedings of the Workshop on Embedded Security in Cars 2004.
- Wolf M., Weimerskirch A. and Wollinger T. (2007). State of the art: Embedding security in vehicles. EURASIP Journal on Embedded Systems.

# How to Apply Privacy by Design in OSINT and big Data Analytics?

Jyri Rajamäki<sup>1, 2</sup> and Jussi Simola<sup>2</sup>

<sup>1</sup>Laurea University of Applied Sciences, Espoo, Finland

<sup>2</sup>University of Jyväskylä, Finland

[jyri.rajamaki@laurea.fi](mailto:jyri.rajamaki@laurea.fi)

[jussi.hm.simola@student.jyu.fi.](mailto:jussi.hm.simola@student.jyu.fi)

**Abstract:** In a world where technology grows exponentially, more information is available to us every day. States and their governments have collected information on their citizens for a long time now. On the other hand, people give out more and more personal information voluntarily through social media. Information available on the Internet is easier to analyze with modern technologies and the original source of information is also easier to track down. Information is available to all of us and that information can be used to investigate personal data, defeat competitors in a corporate world, solve crimes or even win wars. This study analyses open source intelligence (OSINT) and big data analytics (BDA) with the emphasis on cyber reconnaissance and how personal security is part of that entity. The main question is how privacy manifests itself as part of OSINT and BDA. At the same time the study analyses how law enforcement authorities can act so that their reconnaissance actions would be publicly approved. The study uses case study methodology by gathering a comprehensive list of sources for the theory section. The theoretical framework consists of Privacy by Design approach and privacy questions with regard to surveillance, and the General Data Protection Regulation (GDPR) and the Directive 2016/680 'on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data' act as a legal framework. The empirical case dealing with maritime surveillance, explores OSINT and BDA privacy challenges in the MARISA project. The overall target of the paper is to accelerate the discussion on the serious problem of privacy breach that may lead to restrictions of individual liberty and erosion of our society's foundations of trust.

**Keywords:** privacy by design, OSINT, big data analytics, maritime surveillance

---

## 1. Introduction

New surveillance technologies became omnipresent in our everyday live although surveillance has a bad reputation in most countries (Krempel & Beyerer, 2014). Open source intelligence (OSINT) is intelligence collected from publicly available sources, including the Internet, newspapers, radio, television, government reports and professional and academic literature (Glassman & Kang, 2012), and OSINT is being extensively used by local and national law enforcement authorities (LEAs), intelligence agencies and the military. An important aspect of LEAs use of OSINT is social media, which aggregate huge amounts of data generated by users which are in many cases identified or identifiable (Staniforth, 2016): "When combined with other online and stand-alone datasets, this contributes to create a peculiar technological landscape in which the predictive ability that is Big Data Analytics (BDA) has relevant impact for the implementation of social surveillance systems." BDA of OSINT requires the rigorous review and potential overhaul of existing intelligence models and associated processes, however, LEAs must always ensure that their access and use of publicly available information is within national and international legal frameworks (Staniforth, 2016).

This case study, carried out by Yin's (2009) framework, researches privacy issues of the MARitime Integrated Surveillance Awareness (MARISA) project in maritime surveillance domain. Maritime surveillance is essential for creating maritime awareness, in other words 'knowing what is happening at sea'. Integrated maritime surveillance is about providing authorities interested or active in maritime surveillance with ways to exchange information and data. Support is provided by responding to the needs of a wide range of maritime policies-irregular migration/border control, maritime security, fisheries control, anti-piracy, oil pollution, smuggling etc. Also the global dimension of these policies is addressed, e.g. to help detect unlawful activities in international waters. Sharing data will make surveillance cheaper and more effective. Currently, EU and national authorities responsible for different aspects of surveillance collect data separately and often do not share them. As a result, the same data may be collected more than once. The European Commission and EU/EEA members develop a common information-sharing environment (CISE) that integrates existing surveillance systems and networks and gives all relevant authorities access to the information they need for their missions at sea. CISE will make different systems interoperable so that data can be exchanged easily through the use of modern technologies.

The General Data Protection Regulation (GDPR) and/or the Directive 2016/680 'on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the

prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data' regulate processing of personal data in maritime surveillance. Privacy by Design (PbD) is one of the key requirements in both regulations. The purpose of this paper is to understand and build explanation what privacy means with regard to OSINT and BDA. The research questions is: How we can understand PbD in regard to the MARISA services? From the MARISA project's point of view, the target of this case study is to provide research findings for the MARISA project's second phase, in which the already developed MARISA services will be revised and enhanced, and additional services will be included. The overall target of the paper is to accelerate the discussion on the serious problem of privacy breach that may lead to restrictions of individual liberty and erosion of our society's foundations of trust.

## **2. Theoretical framework**

### **2.1 Privacy by design**

Privacy by design (PbD) is an approach to systems engineering approach intended to ensure privacy protection from the earliest stages of a project and to be taken into account throughout the whole engineering process, not just in hindsight. The PbD concept is closely related to the concept of privacy enhancing technologies (PET) published in 1995 (Hustinx, 2010). PbD framework was published in 2009 (Cavoukian, 2011). The concept is an example of value sensitive design that takes human values into account in a well-defined manner throughout the whole process. PbD is one of the key requirements in the European Data Protection Reform beings included in GDPR and Directive 2016/680. The GDPR also requires Privacy by Default, meaning that the strictest privacy settings should be the default.

According to Antignac and Le Métayer (2014), PbD related research has focused on technologies and components rather than methodologies and architectures. They advocate that PbD should be addressed at the architectural level and be associated with suitable methodologies, among other benefits, architectural descriptions enable a more systematic exploration of the design space. In addition, because privacy is intrinsically a complex notion that can be in tension with other requirements, they believe that formal methods should play a key role in this area (Antignac & Le Métayer, 2014). Kung (2014) continues the importance of architecture in designing a PbD system and provides an overview on how architectures are designed, analysed and evaluated, through quality attributes, tactics and architecture patterns. Kung also specifies a straw man architecture design methodology for privacy and present PEAR (Privacy Enhancing ARchitecture) methodology. Martin and Kung (2018) posit that for PbD to be viable, engineers must be effectively involved and endowed with methodological and technological tools closer to their mindset, and which integrate within software and systems engineering methods and tools.

In many respects, original PbD framework has been criticized as being a vague concept. To make its underlying goals more concrete, Colesky Hoepman and Hillen (2016) propose a more specific privacy design strategy: 1) minimize: only collect that data which is strictly necessary, and remove that which no longer is; 2) hide: encrypt, pseudonymize, and take other measures that protect and obscure links between elements of data and their source; 3) abstract: reduce the granularity of data collected; combine or aggregate data from multiple sources so that the sources are no longer uniquely identifiable; 4) separate: store and access data only where it is used; process data at the source instead of centrally; 5) inform: explain to data subjects how their personal data is processed, and how profiles and automated decision-making based on their personal data work. A subject can only provide valid consent to data processing if they understand how their data is being processed; 6) control: allow data subjects to provide and revoke consent to process, and to access, correct, and delete their provided and derived data; 7) enforce: build technical and organizational measures that ensure the design decisions taken with regard to privacy are actually implemented, and log the actions of the systems; and 8) demonstrate: document, audit, and report on the operational and PbD processes. The first four strategies are more focused on data and the last four are about policies and the surrounding processes. Given these strategies, the PbD process could then ideally be implemented as follows (Van Aubel, et al., 2018): "look at each project requirement, figure out what potential privacy impacts it has, and apply strategies to mitigate those impacts". This iterative process should be repeated as the design becomes more detailed (Van Aubel, et al., 2018).

## **2.2 Privacy and surveillance**

The PARIS (PrivAcy pReserving Infrastructure for Surveillance) project (2013-2015) defined and demonstrated a methodological approach for the development of a surveillance infrastructure which enforces the right of citizens for privacy, justice and freedom. The project took into account the evolving nature of such rights, since aspects that are acceptable today might not be acceptable in the future. It also included the social and ethical nature of such rights, since the perception of such rights varies over time and in different countries. Its methodological approach was based on two pillars: 1) a theoretical framework for balancing surveillance and privacy/data protection which fully integrates the concept of accountability; and 2) an associated process for the design of surveillance systems which takes from the start privacy (i.e. Privacy-by-Design) and accountability (i.e. Accountability-by-Design).

Koops (2013) concerns procedural issues of OSINT in police investigations and investigates criminal-procedure law in relation to open source data gathering by the police. He studies the international legal context for gathering data from openly accessible and semi-open sources, including the issue of cross-border gathering of data. This analysis is used to determine if investigating open sources by the police in the Netherlands is allowed on the basis of the general task description of the police, or whether a specific legal basis and appropriate authorization is required for such systematic observation or intelligence. The line between espionage and OSINT can be very thin, therefore caution and double-checking are advised before conducting OSINT activities (Hribar, et al., 2014).

Koops, Hoepman and Leenes (2013) considers the challenge of embedding PbD in OSINT carried out by law enforcement. Ideally, the technical development process of OSINT tools is combined with legal and ethical safeguards in such a way that the resulting products have a legally compliant design, are acceptable within society, and at the same time meet in a sufficiently flexible way the varying requirements of different end-user groups. They use the analytic PbD framework and they discuss two promising approaches, revocable privacy and policy enforcement language. The approaches are tested against three requirements that seem suitable for a ‘compliance by design’ approach in OSINT: purpose specification; collection and use limitation and data minimization; and data quality (up-to-datedness). For each requirement, they analyze whether and to what extent the approach could work to build in the requirement in the system. They demonstrate that even though not all legal requirements can be embedded fully in OSINT systems, it is possible to embed functionalities that facilitate compliance in allowing end-users to determine to what extent they adopt PbD approach when procuring an OSINT platform, extending it with plug-ins, and fine-tuning it to their needs. Therefore, developers of OSINT platforms and networks have a responsibility to make sure that end-users are enabled to use PbD, by allowing functionalities such as revocable privacy and a policy enforcement language (Koops, et al., 2013). Even though actual end-users have a responsibility of their own for ethical and legal compliance, it is important to recognize that it is questionable whether all responsibility for a proper functioning and use of OSINT platforms can be ascribed to the end-users; and some responsibility for a proper functioning of OSINT framework in practice also lies with the developers of the platform and individual components (Guest Editorial, 2013).

## **3. Open source related MARISA services**

The MARISA toolkit provides a suite of services to correlate and fuse various heterogeneous and homogeneous data and information from different sources, including Internet and social networks. MARISA also aims to build on the huge opportunity that comes from using the open access to big data for maritime surveillance: the availability of large to very large amounts of data, acquired from various sources ranging from sensors, satellites, open source, internal sources and of extracting from these amounts through advanced correlation improves knowledge. The first phase MARISA service description document (MARISA, 2018) defines three open source related services: Twitter service, OSINT service and GDELT service. Next we will present those services.

### **3.1 Twitter service**

Twitter is a popular and widely used social media platform for microblogging, or broadcasting short messages. Twitter has hundreds of millions of users worldwide, and they broadcast over every day 500 million messages, known as tweets that may include text, images, and links (Glasgow, 2015). In crisis management, Twitter can act as a human sensor network for real-time event detection, but little attention has been paid to applying text mining and natural language processing techniques to monitor events in a multilingual setting and most of the work focusses on one single language only (Zielinski, 2013).

MARISA Twitter service enables access to open source social media information. Twitter is selected because its users are fast at creating content and an application program interface (API) is available. Many publications in the field of natural language processing are done using Twitter. MARISA Twitter service will read tweets via the Twitter search API. The input is the Area of Interest (AOI), which contains a geolocation (as point in lat/lon coordinates and a radius in km or miles) and the period of time. Each tweet is first analyzed for its language and then tokenized. A special classifier with a language and domain dependent model will assess the relevance of the tweet in this context (domain, use case). The result will be an instance of the Risk class defined in CISE containing a list of assessed tweets with their relevance exposed in the attributes *RiskProbability*, *RiskSeverity* and *RiskLevel*. MARISA Twitter service won't correlate position information extracted from social media with known ship positions, as it will just give away all identified risks in a given area. An example may be illegal immigration. For the trained use case of illegal immigration, if the classifier delivers a relevance of i.e. 0.9, there is a high probability that the tweet is about a real immigration event. So the *RiskProbability* is frequent (01), the *RiskSeverity* is catastrophic (01) related to possible death of immigrants and the derived *RiskLevel* is high (01). (MARISA, 2018)

The first function block *ReadTweets* consists of two parallel threads. The thread containing the functions *BuildRequest* and *SendRequest* handles the interpretation of the function argument (AOI) and the construction of the HTTP request for *TwitterSearch* API. The second thread in *ReadTweets* processes the HTTP response of the *TwitterSearch* API. If there are no tweets in the response an *EmptyResponse* object is built and the flow ends. Otherwise the tweets list is handed over to the *AnalyseTweet* function. In the *AnalyseTweet* function the critical operation is to detect the language of a tweet. All subsequent operations are dependent of the correct identification of the language of the short message. If the language is not successfully identified, an *EmptyResponse* object is built with an appropriate error code as return state. Upon successful language detection the tweet is tokenized with a special tokenizer which respects all the special controls and characteristics of a tweet. The tokenized tweets enter the *ClassifyTweet* function. This function's building blocks *FindClass* and *AssignRelevance* are based on a *DeepLearning* concept using paragraph vectors and their vector space similarity characteristics. So two tweets are similar, if their corresponding (paragraph) vectors enclose a small angle in a multi-dimensional space (around 500 dimensions). After each tweet is processed, a *Response* object is built, containing the classification result, and the flow ends. (MARISA, 2018)

### **3.2 GDELT service**

The Global Database of Events, Language, and Tone (GDELT) is a CAMEO-coded dataset containing geo-located events with global coverage from 1979 to the present. The data are collected from news reports throughout the world and the dataset provides daily coverage on the events found in news reports published on that day. In 2015, datasets Mentions and Global Knowledge Graph (GKG) were added to GDELT. The Mentions table records the network trajectory of the story of each event in flight through the global media system while the GKG table expands GDELT's ability to quantify global human society beyond cataloging physical occurrences towards actually representing all of the latent dimensions, geography, and network structure of the global news. Today, GDELT is a real time database of global human society for open research which monitors the world's broadcast, print, and web news, creating a free open platform for computing on the entire world containing three data tables: Event, Mentions and GKG while most researches are based only on the Event table (Chen, et al., 2016). GDELT archives an exhaustive collection of available online news sources in more than 100 languages (Guo & Vargo, 2017).

MARISA GDELT service integrates open-source data from GDELT project into MARISA. It filters the results using natural language processing in order to identify possible events related to maritime domain, such as naval incidents, piracy events, and pollution events (MARISA, 2018).

Satellite data represent major value adding maritime surveillance information outside the coastal systems coverage. Social media information such as Twitter provides no adjunct value offshore far from ports where online news sources based GDELT data are more relevant, including maritime events such as emergencies like sinking ships, collisions at sea, or information related to the conflicts in defending territorial waters. MARISA proves the capability and potentially the exportability to other areas and topics, further to the ones dealt with in the project. (MARISA, 2018)

OpenGeo Suite Web Feature Service is integrated in the MARISA toolkit that filters relevant events and news from GDELT data. The function exploits big data information extraction techniques from non-conventional sources. Via the MARISA toolkit, the user accesses OSINT data and reports relevant information to the maritime domain, classified per event type and with references, to enrich the maritime picture within the selected AOI. Current GDELT database queries have no correlation with vessels, but news relevant to a certain AOI will be extracted from the database and made available to the MARISA toolkit for a potential contextual analysis. (MARISA, 2018)

### **3.3 OSINT service**

MARISA OSINT service involves the collection, analysis and use of data from open sources for intelligence purposes. It exploits existing open source solutions for social media data stream integration, especially Twitter web crawlers in order to provide capabilities for discovering of alert of any kind of illegal activities in the maritime environment. Multilingual investigations into social media, based mainly on the ability to identify geo-located information, allow to associate the OSINT information with more closely related to the marine environment information (e.g. vessels, sea condition, pollution risks) and then generate an improved maritime picture that meets cross-border requirements of the MARISA project. Depending on the search type that the user wants to perform (based on geographical locations or coordinates, dates or specific keywords that will be configured in a specific phase of the usage of the service) different API services have been applied inside the service code. OSINT service can search, identify and merge relevant multilingual events that can be considered as input to generate alert/incident/tracks. (MARISA, 2018)

MARISA OSINT service receives parameter from the configuration (e.g. AOI, keywords), and on the basis of this, solves possible conflicts, receives tweets and after analysis propagate alarms as following: 1) *TwitterRetriever* provides a sort of orchestration of all service's components; 2) *OrganizerParmas\_API* evaluates if there are conflicts or inconsistencies between the input parameters that would lead to a negative search result; 3) *REST\_APISInvoker* evaluates which representational state transfer API shall invoke, merge or sum the results; 4) *LanguageDetection* evaluates suitable Twitter APIs to invoke, merge or sum the results; and 5) *AlarmPropagator* informs if there are tweets coming from AOI and provides the link for retrieving. Depending on the search type the user wants to perform (based on geographical locations or date), different API services are applied inside the service code. Due to sophisticated combinations of criteria, Twitter service can search and identify the set of relevant tweets that can be considered as alert/incident and from which the list of coordinates can be extracted. (MARISA, 2018)

## **4. Privacy challenges of MARISA OSINT and BDA services**

The MARISA toolkit was built on the top of a big data infrastructure that provides the means to collect external data sources and operational systems products and to organize and exploit all the incoming data as well as all the data produced by the various services. Next we look privacy challenges in four different dimensions of BDA: data generation, data analysis, use of data, and infrastructure behind data.

### **4.1 Data generation**

Data generation can be classified into active data generation and passive data generation: active data generation means that the data owner will give the data to a third party, while passive data generation refers to the circumstances that the data are produced by data owner's online actions (e.g., browsing) and the data owner may not know about that the data are being gathered by a third party (Jain, et al., 2016).

The MARISA Toolkit has two relevant data sources: data coming from the sensors, and data coming from open sources. With regard to data coming from the sensors, these sensors are embodied in the operational environment of the Legacy Systems. Here Legacy Systems mean the previously existing end-users Maritime Surveillance systems in the National/Regional Coordination Centres or Coastal Stations to which MARISA Toolkit must establish some kind of communications. In these environments, owned by Participating Member State governmental entities, we can suppose that the data are used on the basis of need-to-know and need-to-share. Examples of those data from heterogeneous sources are radar and AIS tracks, AIS data validation, near real-time satellite detections and heat maps, integration of maps of most used routes (density maps) and traffic patterns, search and rescue risk maps, fusion of surveillance pictures information from end-users operational environments.

MARISA services include three services (Twitter service, GDELT service and OSINT service) that collect open source information. Their main target is to extract and integrate maritime related safety and security events. OSINT service mainly collects its information via Twitter service and DGETL service. From data collection point of view, MARISA GDELT service may not have privacy concerns because professional journalists should have taken that issue into account when making news. However other ethical issues may arise, for example wealthier countries not only continue to attract most of the world news attention, they are also more likely to decide how other countries perceive the world (Guo & Vargo, 2017).

In Twitter, several technical features and tweet-based social behaviors occur that might compromise privacy. Tweets are complex objects that, in addition to the message content, have many pieces of associated metadata, such as the username of the sender, the date and time the tweet was sent, the geographic coordinates the tweet was sent from if available, and much more (Glasgow, 2015). "Most metadata are readily interpretable by automated systems, whereas tweet message content may require text processing methods for any automated interpretation of meaning" (Glasgow, 2015). "Direct Messages" are the private side of Twitter and "retweeting" is directly quoting and rebroadcasting another user's tweet. Someone might unintentionally or intentionally retweet private tweet to a public forum. Other behaviors include mentioning another user in one's tweet that is, talking about that user. According to Rumbold and Wilson (2018), when one puts any information in the public domain—whether intentionally or not—one does not waive one's right to privacy, but one can only waive one's right to privacy by actually waiving it.

## **4.2 Data analytics**

Big data may be analyzed by artificial intelligence (AI). Machine learning (ML), a branch of AI, can provide detailed, personalized characteristics of an individual and prediction of his or her future behavior (Moallem, 2019). According to Wójtowicz and Cellary (2019), one of the most important characteristics of BDA is the paradigm shift, in which instead of discovering knowledge by searching for causality, one can discover it by searching for correlation: it is possible via BDA to learn with high probability what is happening, and even what will happen, but not why it happens or why it will happen. If a human programmer writes a program, another human programmer may inspect program code and find possible errors, but if a neural network is trained by peta-bytes of data, nobody is able to check whether a particular prediction is correct or not (Moallem, 2019).

Algorithms tell computers step by step how to solve a certain problem. However, predictive algorithms are often themselves unpredictable (Wójtowicz & Cellary, 2019). According to Rahman (2017), the first problem comes from algorithmic bias—AI algorithms being a reflection of the programmers' biases—may possibly give rise to the risk of false alerts by AI surveillance systems thus resulting in wrongful profiling and arrest; and the second problem is that AI profiling systems utilise historical data to generate lists of suspects for the purposes of predicting or solving crimes. ML techniques including neural networks run in two phases (the training phase and the prediction phase) and the quality of predictions is absolutely dependent on examples used for the training phase. ML systems are only as good as the data sets that the systems trained and worked with (Rahman, 2017).

## **4.3 Use of data**

Data analysis does not directly touch the individual and may have no external visibility. An important ethical issue comes with automated policing. Automated discrimination is possible when augmented surveillance becomes more common. It intersects with the technical issues of unintended biases in algorithms and big data that could skew analyses generated by AI systems (Rahman, 2017). If a person is wrongly qualified as a potential terrorist, the consequences may be very severe (Wójtowicz & Cellary, 2019). If BDA provides predictions with 99% accuracy, wrong predictions would concern over 5 million people in the EU, which population is 508 million. Big Data used by law enforcement will increase the chances of certain tagged people to suffer from adverse consequences without the ability to get back or even having knowledge that they are being discriminated (Matturdi, et al., 2014).

## **4.4 Infrastructure behind data**

Data analytics requires not just algorithms and data but also physical platforms where the data are stored and analysed. Cloud computing is currently the most economic option of providing computing power and storage capacity, and privacy assurance can be successfully deployed in private clouds. Although stored data are encrypted and advances in homomorphic encryption, there is no prospect of commercial systems being able to

maintain this encryption during real-time processing of large datasets (Wójtowicz & Cellary, 2019). The security and privacy for big data is not different from security and privacy research in general (Nelson & Olovsson, 2016).

## 5. Summary and conclusion

According to the European Data Protection Reform, PbD is a mandatory approach in maritime surveillance context. Although PbD as a concept is becoming well-known, it turns out that there is not much standardization in how to actually apply it, especially by security authorities. This paper explores privacy challenges in the MARISA project and tries to accelerate the discussion on the serious problem of privacy breach that may lead to restrictions of individual liberty and erosion of our society's foundations of trust. Current academic arguments are shifting the focus of privacy concerns from data collection to data analytics and data use, and there are scholars requiring "algorithmic accountability" (Broeders, et al., 2017). It can therefore be expected that the legal requirements concerning OSINT and BDA may develop into this direction.

One very important issue is who watches the watchers (political issue) and how this can be carried out (technical issue). Utilizing BDA in the security domain requires intensive oversight (Broeders, et al., 2017). However, BDA is often a "black box", and more research is needed, especially in the phase of the analysis: selecting the algorithms, data sources and categorization, assigning weight to various data, etc.

## Acknowledgements

Acknowledgement is paid to the MARISA Maritime Integrated Surveillance Awareness project. This project is funded by the European Commission through the Horizon 2020 Framework under the grant agreement number 740698. The sole responsibility for the content of this paper lies with the authors. It does not necessarily reflect the opinion of the European Commission or of the full project. The European Commission is not responsible for any use that may be made of the information contained therein.

## References

- Antignac, T. & Le Métayer, D., 2014. Privacy by Design: From Technologies to Architectures. In: *Privacy Technologies and Policy*. Cham: Springer, pp. 1-17.
- Broeders, D. et al., 2017. Big data and security policies: Towards a framework for regulating the phases of analytics and use of Big Data. *Computer Law & Security Review*, Volume 33, pp. 309-323.
- Cavoukian, A., 2011. *Privacy by Design: The 7 Foundational Principles*, Ontario: Information and Privacy Commissioner of Ontario.
- Chen, K., Qiao, F. & Wang, H., 2016. Correlation Analysis Using Global Dataset of Events, Location and Tone. *2016 IEEE First International Conference on Data Science in Cyberspace (DSC)*, pp. 648-652.
- Colesky, M., Hoepman, J.-H. & Hillen, C., 2016. A critical analysis of privacy design strategies", Procs. IWPE'16, IEEE, 33–40.. *2016 IEEE Security and Privacy Workshops (SPW)*, pp. 33-40.
- Glasgow, K., 2015. Big data and law enforcement: Advances, implications, and lessons from an active shooter case study. In: *Application of Big Data for National Security*. Waltham: Butterworth-Heinemann, pp. 39-54.
- Glassman, M. & Kang, M. J., 2012. Intelligence in the internet age: the emergence and evolution of OSINT. *Computers in Human Behavior*, Volume 28, pp. 673-682.
- Guest Editorial, 2013. Legal aspects of open source intelligence - Results of the VIRTUOSO project. *Computer law & security review*, Volume 29, pp. 642-653.
- Guo, L. & Vargo, C., 2017. Global Intermedia Agenda Setting: A Big Data Analysis of International News Flow. *Journal of Communication* , pp. 499-520.
- Hribar, G., Podbregar, I. & Ivanusa, T., 2014. OSINT: A "Grey Zone"? *International Journal of Intelligence and CounterIntelligence*, Volume 27, p. 529–549.
- Hustinx, P., 2010. Privacy by design: delivering the promises. *Identity in the Information Society*, 3(2), pp. 253-255.
- Jain, P., Gyanchandani, M. & Khare, N., 2016. Big data privacy: a technological perspective and review. *Journal of Big Data*.
- Koops, B., 2013. Police investigations in Internet open sources: procedural law issues. *Computer Law & Security Review*, Volume 29, pp. 676-688.
- Koops, B., Hoepman, J. & Leenes, R., 2013. Open-source intelligence and privacy by design. *Computer Law & Security Review*, Volume 29, pp. 676-688.
- Krempel, E. & Beyerer, J., 2014. TAM-VS: A Technology Acceptance Model for Video Surveillance. In: *Privacy Technologies and Policy*. Cham: Springer, pp. 86-100.
- Kung, A., 2014. PEARS: Privacy Enhancing ARchitectures. In: *Privacy Technologies and Policy*. Cham: Springer, pp. 18-29.
- MARISA, 2018. D3.2 MARISA SERVICES DESCRIPTION DOCUMENT, s.l.: s.n.
- Martín, Y.-S. & Kung, A., 2018. Methods and Tools for GDPR Compliance through Privacy and Data Protection Engineering. *2018 IEEE European Symposium on Security and Privacy Workshops*, pp. 108-111.
- Matturdi, B., Zhou, X., Li, S. & Lin, F., 2014. Big Data security and privacy: A review. *China Communications*, pp. 135-145.

- Moallem, A., 2019. Perspectives on the future of human factors in cybersecurity. In: Human-computer interaction and cybersecurity handbook. Boca Raton: CRC Press, pp. 353-366.
- Nelson, B. & Olovsson, T., 2016. Security and privacy for big data: A systematic literature review. 2016 IEEE International Conference on Big Data (Big Data), pp. 3693-3702.
- Rahman, F., 2017. Smart Security: Balancing Effectiveness and Ethics. RSIS Commentary, 14 Dec. Volume 235.
- Rumbold, B. & Wilson, J., 2018. Privacy Rights and Public Information. *The Journal of Political Philosophy*.
- Staniforth, A., 2016. Big Data and open source intelligence – A game-changer for counter-terrorism?. TRENDS Research & Advisory, 19 July.
- Van Aubel, P. et al., 2018. Privacy by design for local energy communities. Ljubljana, s.n.
- Wójtowicz, A. & Cellary, W., 2019. New challenges for user privacy in cyberspace. In: Human-computer interaction and cybersecurity handbook. Boca Raton: Taylor & Francis Group, pp. 77-96.
- Yin, R. K., 2009. Case Study Research Design and Methods. s.l.:Thousand Oaks: Sage Publications.
- Zielinski, A., 2013. Detecting natural disaster events on twitter across languages. Intelligent interactive multimedia systems and services. 6th International conference on Intelligent Interactive Multimedia Systems and Services, pp. 291-301.

# **Research Challenges for Cybersecurity and Cyberwarfare: A South African Perspective**

**Trishana Ramluckan<sup>1</sup>, Brett van Niekerk<sup>1</sup> and Louise Leenen<sup>2</sup>**

**<sup>1</sup>University of KwaZulu-Natal, Durban, South Africa**

**<sup>2</sup>University of the Western Cape and CAIR, Cape Town, South Africa**

[ramluckant@ukzn.ac.za](mailto:ramluckant@ukzn.ac.za)

[vanniekerkb@ukzn.ac.za](mailto:vanniekerkb@ukzn.ac.za)

[lleenan@uwc.ac.za](mailto:lleenan@uwc.ac.za)

**Abstract:** The International Institute for Strategic Studies (2018: 6) states that “cyber capability should now be seen as a key aspect of some states’ coercive power ... This has driven some European states to re-examine their industrial, political, social and economic vulnerabilities, influence operations and information warfare, as well as more traditional areas of military power.” Cybersecurity is often incorrectly assumed to be a purely technical field, however there are numerous multidisciplinary aspects. The very nature of cybersecurity and operations in cyberspace is disruptive, and this is true for many disciplines attempting to introduce cybersecurity research into their offerings. This can provide challenges to researchers and students where methodologies that do not necessarily follow disciplinary norms are prejudiced against by old-school thought. Foundational understanding of concepts may also hinder multi-disciplinary research, as specific terminology that is used in cybersecurity may be considered colloquial or have different meanings in other disciplinary settings. The experimental, observational and mathematical research methodologies often employed by computer scientists do not address the political or legal aspects of cybersecurity research. Research methods for cybersecurity generally apply and teach the limited scientific methods for creating new knowledge, validating theories, and providing some critical insights into to the cybersecurity arena. This paper aims to investigate the South African national and institutional perspectives for higher education and research, identify challenges, and propose interventions to facilitate multidisciplinary research into cybersecurity and cyberwarfare in South Africa. Legislature and policies, organisational structures, processes, resources, and historical and socio-economic factors will be discussed as to the influence on cybersecurity research. A review and analysis of international efforts for multidisciplinary research in higher education institutions will provide for a basis to propose a framework for South African higher education institutions to effectively implement cybersecurity research.

**Keywords:** cybersecurity, cyberwarfare, higher education, multidisciplinarity, research methods

---

## **1. Introduction**

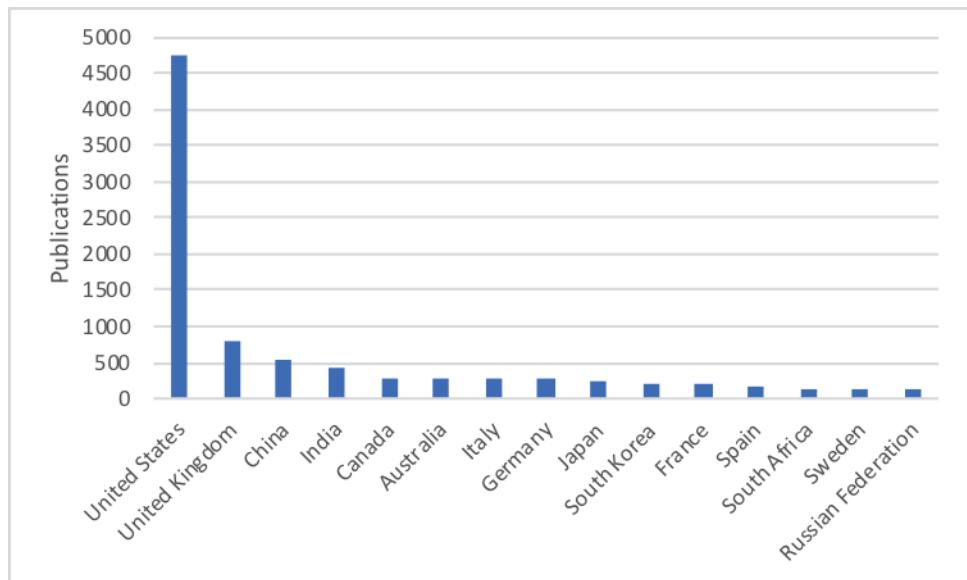
In today’s ‘global’ society, research methodology challenges require collaboration by researchers from multidisciplinary backgrounds. However, multidisciplinary research is not without its challenges which may result from training as well as the scientific culture. Regarding cybersecurity research, it can be said to currently be of a strict scientific methodology, with little to no involvement or collaboration with other disciplines thereby restricting innovation in such a diverse field. By definition research refers to “a search for knowledge” and a systematic method of identifying a problem, collecting and collating information, developing hypothesis and analysing them to form grounded conclusions and provide viable recommendations.

There is now a growing request by research policy-makers for the change and adaptation of research methodologies to allow for collaboration and multidisciplinarity, to gain the most value from a “shared research methodology”. Although the idea of multidisciplinarity appears good on paper, there is limited literature on collaborative research between disciplines. While there are numerous terms such as “multi-disciplinarity”, “trans-disciplinarity” and “inter-disciplinarity”, the area of collaborative research remains idealist and its practicality is yet to be established.

Cybersecurity refers to “the measures that are taken to protect a computer or computer system (as on the Internet) against unauthorized access or attack” (Merriam-Webster, 2018) while cyber warfare is defined as the “use of computer technology to disrupt the activities of a state or organization, especially the deliberate attacking of information systems for strategic or military purposes”. From the definition of cybersecurity, it can be seen as a purely technical area. However, cybersecurity and cyber-warfare are unique areas as they combine the social sciences including political sciences as well the technical sciences. Therefore, it becomes essential for the development and practice of multidisciplinarity research methodologies with reference to cybersecurity and cyberwarfare.

Desk research was conducted for the paper. For the desk research information was gathered using existing resources, including the press, the Internet, analytical reports and statistical publications. This was then followed by the collation of data.

Figure 1 illustrates South Africa's standing in international research. A Scopus search for publications with a title or keyword including 'cybersecurity' or 'cyber-security' was performed, and South Africa was 13<sup>th</sup> with 149 publications.



**Figure 1:** South Africa's standing in international cyber-security research, source: (Scopus, 2019)

## **2. The social and technical elements of cybersecurity and cyberwarfare**

The most well-known "hacking" incident was that used in the 2016 US elections. According to media reports 13 Russian nationals and three other entities were charged with conducting an illegal "information warfare" by attempting to disrupt the 2016 presidential election in order to influence the election outcome. This according to Matishak (2018), had cost millions of dollars, time and labour resources. The purpose of the campaign, conducted through a Russian "Troll Farm", was to spread distrust towards the presidential candidates and the US political system. This incident was termed as "cyber-warfare" or "election hacking". The general definition of hacking is "the gaining of unauthorized access to data in a system or computer" (Merriam-Webster, 2018). However, this was not a technical hack, as no system was infiltrated by an unauthorised user. Instead fake online accounts were created to influence the voters. Similar concepts were considered by Cybenko, Giani, and Thompson (2002) in what they termed "cognitive hacking".

Furthermore, the Irish Republican Army had subsequently begun an operation using social media platforms i.e. Facebook, Twitter, Instagram and YouTube to influence the US people in their choice of Presidential candidate. This was apparently done through the creation of fake bot accounts and false or misleading advertising (Matishak, 2018). The key example is that the IRA trolls produced materials promoting Trump e.g. #TrumpTrain, as well as anti-Clinton hashtags on Twitter, such as #Hillary4Prison. Further to this the alleged trolls had also encouraged minorities either to not vote or to vote for a third-party candidate starting in the latter half of 2016. This is a key element reflecting psychological warfare in cyber-space.

The technical side of hacking may refer to the plot which included the operation in the middle of 2016, which involved malware on "at least ten of the Democratic Congressional Campaign Committee's (DCCC) computers", which may have stolen employees' passwords, and lead to the indictment leaks. This further enabled the "hackers" to illegally monitor the Democratic Party's activities. This allowed the publication of the retrieved files by DC Leaks, Guccifer 2.0 and what is believed to be WikiLeaks (Matishak, 2018).

Cybersecurity is relevant at a tactical or technical level, as well as a strategic international level. Traditional concepts of deterrence and state sovereignty are questioned when it comes to cybersecurity (Davis, 2015). The role of cyber in military and intelligence applications is uncertain. There is debate surrounding the applicability

of traditional military theorists such as Sun Tzu, Clausewitz and Jomini (Duggan, 2016), or if a new theory of war is required. The head of the British MI6, Alex Younger, indicated that intelligence needs to fuse traditional intelligence with modern technological methods (Fitsanakis, 2018). Attempts are being made to ascertain the relevance of current international humanitarian laws regarding war to cyber-space (Schmitt, 2017). Liaropoulos (2014) illustrates the dualistic technical and socio-political nature of cyber-conflict through a discussion of theoretical paradigms.

Cybersecurity and cyberwarfare consist of many elements both technical and social, which requires a multidisciplinarity approach to be able to fully analyse/understand it. This creates numerous challenges in the field of research.

### **3. Challenges in cyber security research**

#### **3.1 Global challenges**

The shortage of cybersecurity skills is a global problem and it has become critical ((ISC)<sup>2</sup>, 2018; Florentine, 2015). The (ISC)<sup>2</sup>'s Cybersecurity Workforce Study of 2018 estimates a global shortage of just under 3 million positions. Governments, civil society, business and the military are competing to recruit within a small pool of cybersecurity professionals. Universities and research institutions are struggling to compete against all these other sectors to find and retain researchers and lecturers in this field.

Awareness of the discipline amongst potential researchers and workers is often not adequate. School leavers need to be educated about the opportunities in the field. Raytheon and the National Cybersecurity Alliance published the result of a study in 2015 that indicated 67 % of men and 77 % of woman in the US and 62 % of men and 75 % of woman globally, did not receive any counselling in high school or secondary schools on careers in cybersecurity (Florentine, 2015).

Women are under-represented in the cybersecurity and this creates another challenge in the field. The (ISC)<sup>2</sup> Cybersecurity Workforce Study (2018), fielded in North America, Latin America, Asia Pacific and Europe, found that women represent only 24% of the workforce overall. Willes-Ford (2018) found that although women 50% of the U.S. workforce comprise of women, only 10-15% of the U.S. cybersecurity workforce consist of women, Foley et al. (2017) lodged a study that found women only comprised 11% of the global cybersecurity workforce and 10% in the Asia-Pacific region. This study's key findings are concerning:

- Fifty-one percent of women in the filed report widespread discrimination and stereotype bias;
- The lack of flexibility in work hours and long work hours is a primary obstacle;
- Women feel their opinions are not valued by their employers;
- Women persistently face wage inequality;
- The attrition of women from the ICT, Science and Engineering fields start in primary school and the current number of female graduates are declining. It is possible that the perception women may have of the cybersecurity career could be a deterrent to choosing this field. Jessica Ortega of SiteLock noted that "Women often don't see tech or security as viable career paths because they're often considered masculine professions" (Bradford, 2018). In addition to the problem of attracting women to the industry, retention is also problematic. Women frequently leave the field after a brief tenure according to a study by Georgetown University Center on Education and the Workforce 2011 (Hechinger Report, 2018). Willis-Ford (2018) conducted a comprehensive study on the perceived barriers on retaining women in cybersecurity and found that lack of mentorship, the Imposter Phenomenon and hostile work environments are noteworthy barriers. The Imposter Phenomenon is when a suitably qualified individual feels inadequate to satisfy the requirements of a position. A hostile work environment include elements such as harassment and a lack work/life balance.

Availability of data is often considered as a problem for academic cybersecurity research. Corporations and governments are unwilling to provide too much detail on successful cyber-attacks, and to get access to detailed network information to test new algorithms or systems is very difficult. In nations where there are laws mandating public announcements of data breaches, data will be more readily available; in countries without this requirement there will be limited information on cyber-attacks available. The legal mandates however, are usually for data breaches of personal information, and do not necessarily require public notification of other

cyber-attacks that do not affect data. Therefore there is even less available data on cyber-attacks against industrial systems, making specific research into this area even more challenging.

Data cleanliness and consistency is problematic. Vendor and computer security incident response team (CSIRT) reports often differ in the categorisation of data, which limits the accuracy of analysis. Some vendors and CSIRT also change their categorisations from year to year (Pretorius, 2016). For instance, if a researcher is focussing on the transportation sector, it could be combined with utilities in one year, and categorised on its own in another year. Most cyber-attack techniques rely on secrecy and deception; when conducting research, it is important to filter through possible erroneous reports (for example news reports that are sensationalist and based on conjecture). The reliability of data collected based on human perceptions also needs to be considered; for example, a respondent may report that their organisation has not been a victim of a cyber-attack, however they may not yet be aware that an attack has occurred (van Niekerk, 2011).

As stated by James Brokenshire, the Minister for Crime and Security, “governments cannot deliver a safer online world. We need to work closely with industry to ensure that safe infrastructure and services can be provided to the public and share information and skills (OGL, 2017).” Jansen van Vuuren & Leenen (2018) identified roles for government, business and academia to work together to build cybersecurity capacity and capability in South Africa but these results are applicable globally. Educational institutions must be supported to initiate new cybersecurity qualifications. Businesses should embark and expand in capability building initiatives such as the participation in cybersecurity awareness in schools as well as career guidance of young people. The businesses can also support curricula development by providing threat and attack information to allow for current and local relevance. The implementation of cybersecurity exercises gives opportunities of hands-on learning, and can enhance interest in these careers. In addition, business can sponsor bursaries, support internships, and help to conduct cybersecurity exercises.

### **3.2 Challenges in South Africa**

In South African Higher Education, the National Qualifications (NQF) structure remains rigid and does not allow for a multidisciplinary approach. With the diversity of cybersecurity, a multidisciplinary approach is required. The South African NQF structure sets the boundaries, principle and guidelines, which provides the vision, as well as the philosophical foundational base for the construction of the qualifications system.

Whilst multi-disciplinary research is advocated at South African institutions, implementing it in practice is often problematic. To establish a strong national research capacity, inter-institutional collaboration is required; however, they cannot even achieve effective collaboration internally. A concept known as vertical progression ensure that to do a post-graduate in Computer Science or Politics, a student requires an undergraduate major in that discipline. However, due to institutional structures, it is very difficult for a student to major in both a technical science and a social science. The various disciplines can be very territorial and have strict expectations. This presents challenges for students wishing to do a postgraduate with a cybersecurity topic as the project can be considered too technical for social science disciplines, however it is not technical enough for computer science. Prospective students may therefore lose enthusiasm and opt not to register, or choose another topic.

Cybersecurity is a discipline that has been growing rapidly but is still a young discipline. There are still no specialised cybersecurity degree courses in South Africa. This is partially due to the limited number of academics specialising in cybersecurity and the fact that most universities do not have lecturers that can teach a sufficiently broad range of cybersecurity topics. Most active researchers were trained in other related disciplines such as Computer Science, Statistics, Engineering, or Information Systems. This creates challenges due to different research approaches and the fact that a large percentage of researchers have not been working in the discipline for a long time. They often have steep learning curves and need to modify approaches and perceptions to adjust to the cybersecurity domain. Often in industry cybersecurity professionals emerge from accountants and auditors due to the function of auditing security controls on financial information systems. They then attempt to consider postgraduate studies related to cybersecurity without the necessary background.

The schooling system in South Africa provides a very weak grounding in Science, Technology, Engineering, and Mathematics (STEM) subjects. For those going into a technical cybersecurity field, the necessary skills and knowledge can be provided through the undergraduate programmes. For those going through the social sciences, and then into cybersecurity, they may then struggle to understand the technical concepts required.

Government support is also lacking. Whilst countries such as the US and UK have established accreditation programmes supported by the military and intelligence communities (Government Communications Headquarters, 2016; National Security Agency, 2016), South Africa does not have a dedicated support programme from government. Therefore, cyber-security research projects need to compete with all the other disciplines. The government should include cybersecurity specifically as a scarce-skill and allocated dedicates funding to post-graduate research in this discipline. There is a mandate that the national Council for Scientific and Industrial Research conducts research on behalf of the military, so there is limited engagement directly between the academia and the military. However, the South African Cybercrimes Bill has indicated that the military, law enforcement, and intelligence agencies should engage with academia to develop the necessary skills (Minister of Justice and Correctional Services, 2017). A possible hinderance to the implementation of this is the general lack of funding for tertiary institutions.

In South Africa, there is a limited core of researchers, with little or no evidence of younger researchers coming through: this is evidenced by the high prevalence of specific South African authors, with limited publications coming from a variety of other authors who do not appear to continue with research (van Niekerk, 2015). This could indicate a future shortage should the existing researchers leave the South African research environment. Whilst there is a global under representation of women in cybersecurity, the strong political mandate to deliver on employment equity in South Africa is difficult to meet when it is almost impossible to find any suitable candidates.

#### **4. Discussion**

In an online society, cybersecurity and cyber-war challenges involve more than just the technical issues. Cybersecurity extends into every aspect of modern society, as critical infrastructure from basic healthcare providers to the power grid, is now fully dependent on technology. This creates the need for a multidisciplinary approach towards cybersecurity and cyberwarfare as it should consist of people, processes and technology. This concept, also known as the “Golden Triangle”, was made popular by Bruce Schneier, a cyber-security and privacy expert, in the 1990’s and states that “operational efficiency requires an approach that optimises the relationships between people, process & technology” (Banks, 2016).

Although cyber networks have a technical component, they also have a social one. As such cybersecurity threats become social threats (University of Nevada Cyber Research Centre, 2018). Therefore, a multidisciplinarity research approach to cybersecurity becomes imperative. This approach combines numerous branches of learning, with the sole purpose of achieving the same objective. The concept of multidisciplinarity research involves an in-depth inquiry into the problem for ascertaining the main hypothesis, but also combines different academic approaches and methods. Molteberg and Bergstrom (2000) have argued that “Multidisciplinary Studies appear to be both applied and action or policy-orientated” and is considered as a “progressive scholarly method.” Therefore, modern research is becoming more multidisciplinary in nature. With reference to Choudhary (2015), a multidisciplinary research approach provides for international cooperation linking the key principles in areas from policy development to cybersecurity.

With the broad scope of cybersecurity, defining an expert in the field is difficult. Does someone who graduate with a doctorate with a focus of student awareness of cybersecurity topics qualify as an expert, if they are unable to implement technical cybersecurity solutions? If someone has a deep knowledge of firewall rules, but does not understand the international relations of cyber-conflict, can they be considered as cybersecurity experts? With such a diverse set of perspectives, cybersecurity should be a discipline on its own and not reside under other disciplines such as computer science. Researchers and students of cybersecurity should then receive a grounding of the different perspectives prior to specialising through research in one specific area. This requires dedicated postgraduate degrees for cybersecurity to be implemented within South Africa. Dedicated degrees will allow for students to register without external disciplinary constraints negatively affecting promising research.

Additional modules should be made available to students specifically to develop skills critical thinking and analytic techniques. As Beebe and Pherson (2015) indicate, these skills are required for academic research as well as intelligence analysis; therefore, they are critical for students intending to conduct research in cybersecurity. Given that the available data and information in South Africa is imperfect or incomplete, and the secret nature of cyber-attacks, the need for critical thinking and analytical skills to provide a coherent view through research is required.

In South Africa, the average time to detect a breach is 150 days, followed by an average of 40 days to contain the breach (Moyo, 2018). As the average time to detect a breach is nearly 5 months, obtaining human perceptions or feedback may be erroneous due to not yet having detected a breach. Whilst the Protection of Personal Information (POPI) Act does require organisations to publicly disclose the occurrence of data breach, the Act is not fully implemented, therefore companies can still opt not to report a breach, however the breaches may be discovered and/or disclosed by third parties. This indicates that there will be a lack of data on all cyber-attacks in the country. Efforts should be made to engage with industries and the relevant government departments for them to supply desensitised data that can be used in cybersecurity research. For technical research, an alternative is for there to be initial research projects that use simulation and emulation to generate synthetic data that can be used.

Given the general lack of available funding for tertiary institutions in South Africa, it is imperative that some of the funding that is available be dedicated to cybersecurity research. Alternative funding mechanisms to allow engagements between government functions (governance, military, law enforcement, and intelligence agencies) and academia should be provided to ensure mutual support towards developing a sustainable national cybersecurity skills base. It is imperative for South African researchers to collaborate to motivate for dedicated cybersecurity research funding and partnerships. Of the funding that is available, there should be a portion dedicated towards cybersecurity research to provide impetus to the established researchers to foster young academics in the field.

Diversity in the workplace is good for any business (Bradford, 2018; Shaban, 2016) and the cyber workforce can only benefit from higher numbers of women and other minorities (currently within the field). Shaban (2016) found that diversity in terms of social background in a workforce usually leads to improved innovation, ideas and creativity. In South Africa, an effort should be made to attract recruits from previously disadvantaged backgrounds.

## **5. Conclusion**

There remain challenges both within a national and international context, which include the rigidity of the NQF Educational structure and lack of monetary resources in South Africa, and the lack of cyber skills both nationally and internationally. Further to this is the aspect of critical thinking, which is usually not taught at any level of education in South Africa. The traditional scientific research methodologies used by scientists fail to address the political and social elements of cybersecurity research. Research methods for cybersecurity is accustomed to maintaining the “traditional” means for creating new knowledge and validating theories in cyber. In this era of rapid movement in technology, society, and various socio-economic problems, research with connections to different disciplines such as political science may provide the ideal solution. The purpose of research is to provide a solution for the betterment of society. By its nature, cybersecurity and cyber warfare involves a “human element” and cannot be studied or researched in isolation i.e. as a purely technical field. South African institutions need to implement multi-disciplinary research, with dedicated postgraduate degrees in cybersecurity, in order to circumvent the challenges faced do to the current siloed study areas that are currently present. Dedicated funding should be allocated to cybersecurity research, particularly for women and previously disadvantaged demographics.

## **Acknowledgements**

The second author received funding by the South African National Research Foundation, Grant no. 115059.

## **References**

- Banks, C. (2016). People, process & technology – why is it important to consider all 3? [online], accessed 16 February 2019, <https://analyze.co.za/people-process-technology-important-consider-3/>.
- Beebe, S.M., and Pherson, R.H. (2015) *Cases in Intelligence Analysis: Structured Analytic Techniques in Action*, 2<sup>nd</sup> ed., Los Angeles: Sage.
- Bradford, L. (2018) "Cybersecurity Needs Women: Here's Why," *Forbes*, 8 October, [online], accessed 4 February 2019, <https://www.forbes.com/sites/laurencebradford/2018/10/18/cybersecurity-needs-women-heres-why/#3d4590d647e8>.
- Choudhary, A. (2015). Multidisciplinary Research. [online] , accessed 24 January 2019, <https://www.lawctopus.com/academike/multidisciplinary-research/>
- Cybenko, G., Giani, A., and Thompson, P. (2002) "Cognitive Hacking: A Battle for the Mind", *Computer* 35(8), pp. 50 – 56.
- Davis, P.K. (2015) "Deterrence, Influence, Cyber Attack, and Cyberwar," *International Law and Politics*, vol. 47, pp. 327-355.

- Duggan, P. (2016) Why Special Operations Forces in US Cyber-Warfare? *The Cyber Defense Review* 1(2), January 8, pp. 73-79 [online], accessed 16 February 2019,  
[https://cyberdefensereview.army.mil/Portals/6/Documents/CDR%20Journal%20Articles/US\\_Special\\_Operations\\_Duggan\\_Oren.pdf?ver=2018-08-01-090209-853](https://cyberdefensereview.army.mil/Portals/6/Documents/CDR%20Journal%20Articles/US_Special_Operations_Duggan_Oren.pdf?ver=2018-08-01-090209-853)
- Fitsanakis, J. (2018) MI6 spy chief outlines 'fourth generation espionage' in rare public speech, 4 December, [online], accessed 16 February 2019, <https://intelnews.org/2018/12/04/01-2449/>.
- Florentine, S. (2015) "Closing the cybersecurity talent gap, one woman at a time," CIO, 17 November, [online], accessed 4 February 2019, <https://www.cio.com/article/3005637/cyber-attacks-espionage/closing-the-cybersecurity-talent-gap-one-woman-at-a-time.html>.
- Foley, M., Dewey, L., Williamson, S., Blackman, D., Creagh, A., Davidson, L., and Zhu, M. (2017) *Women in Cyber Security Literature Review*, UNSW Canberra Australian Centre for Cybersecurity, June, [online], <https://www.homeaffairs.gov.au/cyber-security-subsite/files/cyber-security-literature-review.pdf>.
- Government Communications Headquarters. (2016) *GCHQ certifies six more Masters' degrees in Cyber Security*, May 23, [online], accessed 11 April 2017, <https://www.gchq.gov.uk/news-article/gchq-certifies-six-more-masters-degrees-cyber-security>.
- The Hechinger Report. (2018) "Jobs in cybersecurity are exploding: Why are women locked out?" *Transmisis*, 4 May, [online], accessed 4 February 2019, <https://transmisis.com/jobs-in-cybersecurity-are-exploding-why-are-women-locked-out/>.
- International Institute for Strategic Studies. (2018) "Editor's Introduction: Western technology edge erodes further," *The Military Balance*, 118(1), pp. 5-6.
- (ISC)<sup>2</sup>. (2018). *Cybersecurity Workforce Study 2018: Professionals Focus on Developing New Skills as Workforce Gap widens*, [online], accessed 4 February 2019, <https://www.isc2.org/-/media/ISC2/Research/2018-ISC2-Cybersecurity-Workforce-Study.ashx?la=en&hash=4E09681D0FB51698D9BA6BF13EEABFA48BD17DB0>.
- Jansen van Vuuren, J. and Leenen, L. (2018) "Cybersecurity and Capacity Building for South Africa," *Proceedings of the 13th Human Choice and Computers Conference (HCC 13)*, September, Poznan, Poland.
- Liaropoulos, A. (2014) "Cyberconflict and Theoretical Paradigms: Current Trends and Future Challenges in the Literature," *Proceedings of the 13th European Conference on Cyber Warfare and Security*, 3-4 July, pp. 133- 139.
- Matishak, M. (2018). What we know about Russia's election hacking. Politico. [online]. Accessed 29 January 2019, <https://www.politico.eu/article/russia-hacking-us-election-what-we-know/>
- Merriam-Webster. (2018). Merriam-Webster Dictionary. Encyclopaedia Britannica [Online], 27 January 2018, <https://www.merriam-webster.com/dictionary/cybersecurity>
- Merriam-Webster. (2018). Merriam-Webster Dictionary. Encyclopaedia Britannica [Online], 27 January 2018, <https://www.merriam-webster.com/dictionary/electionhacking>
- Minister of Justice and Correctional Services. (2017) *Cybercrimes and Cybersecurity Bill*. Republic of South Africa.
- Molteberg, E & Bergstrom, C. (2000) Our Common Discourse: Diversity and Paradigms in Development Studies, [online], accessed 3 February 2019, [https://www.researchgate.net/publication/242094789\\_Our\\_Common\\_Discourse\\_Diversity\\_and\\_Paradigms\\_in\\_Development\\_Studies](https://www.researchgate.net/publication/242094789_Our_Common_Discourse_Diversity_and_Paradigms_in_Development_Studies)
- Moyo, A. (2018) "SA's Average Data Breach Costs Escalate," *ITWeb*, 12 July, [online], accessed 3 February 2019, <https://www.itweb.co.za/content/Pero37ZgOnXMQb6m>.
- National Security Agency. (2016) *Centers of Academic Excellence in Cybersecurity*, May 3, [online] accessed 11 April 2017, <https://www.nsa.gov/resources/educators/centers-academic-excellence/>.
- OGL. (2017). Industrial Strategy Building a Britain fit for the future. HM Government, London, [online], accessed 3 February 2019, [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/664563/industrial-strategy-white-paper-web-ready-version.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/664563/industrial-strategy-white-paper-web-ready-version.pdf)
- Pretorius, B. (2016) *Cyber-Security and Governance for Industrial Control Systems (ICS) in South Africa*, Masters Dissertation, Durban, South Africa: University of KwaZulu-Natal.
- Schmitt, M.N. (2017) *Tallinn Manual 2.0: On The International Law Applicable to Cyber Operations*, Cambridge: Cambridge University Press.
- Scopus. (2019). Analysis of Search Results, [online], accessed 4 February 2019, <https://www.scopus.com>.
- Shaban, A. (2016). "Managing and Leading a Diverse Workforce: One of the Main Challenges in Management," *Social and Behavioural Sciences*, vol. 230, pp. 76-84.
- University of Nevada. (2018) *Cybersecurity Research*, [online], accessed 31 January 2019, <https://www.unr.edu/cybersecurity/research>
- van Niekerk, B. (2011) *Vulnerability Assessment of Modern ICT Infrastructure from an Information Warfare Perspective*, Doctoral Thesis, Durban, South Africa: University of KwaZulu-Natal.
- van Niekerk, B. (2015) "An Information Operations Roadmap for South Africa , " *10th International Conference on Cyber Warfare and Security*, 24-25 March, South Africa, pp. 347-357.
- Willis-Ford, C. (2018) "The Perceived Impact of Barriers to Retention on Women in Cybersecurity", Thesis, University of Fairfax, [online], accessed 17 February 2019, [https://www.researchgate.net/publication/329754528\\_THE\\_PERCEIVED\\_IMPACT\\_OF\\_BARRIERS\\_TO\\_RETENTION\\_ON\\_WOMEN\\_IN\\_CYBERSECURITY](https://www.researchgate.net/publication/329754528_THE_PERCEIVED_IMPACT_OF_BARRIERS_TO_RETENTION_ON_WOMEN_IN_CYBERSECURITY).

# Cyber-Influence Operations: A Legal Perspective

Trishana Ramluckan<sup>1</sup>, Alicia Wanless<sup>2</sup> and Brett van Niekerk<sup>1</sup>

<sup>1</sup>University of KwaZulu-Natal, Durban, South Africa

<sup>2</sup>Kings College London, UK

[ramluckant@ukzn.ac.za](mailto:ramluckant@ukzn.ac.za)

[Alicia.wanless@kcl.ac.uk](mailto:Alicia.wanless@kcl.ac.uk)

[vanniekerkb@ukzn.ac.za](mailto:vanniekerkb@ukzn.ac.za)

**Abstract:** The role of social media in mass protests, such as the Arab Spring, provided insight into the use of online technologies for large-scale influence. The investigations and allegations surrounding the British independence referendum, the 2016 US election, and influence campaigns in South Africa illustrate a complex weave of actors, including foreign operators, trans-Atlantic ideological networks with wealthy financiers, political firms using behavioural advertising, and domestic politicians, engaging in or benefiting from cyber-influence operations, often complicating international relations between affected states, resulting in tit-for-tat increased regulation on foreign journalists in the U.S. and Russia. The majority of academic studies of cyber-influence operations focus on the mechanisms to conduct them or the security and societal impact thereof. Few studies consider the legal perspectives of employing cyber-influence operations or the measures taken by states to mitigate the impact. The regulatory options that can be applied to these scenarios are complex, ranging from international humanitarian law to local legislation. This paper identifies the common components and techniques of online influence and the countermeasures. A document analysis of various regulatory guidelines and norms is conducted to investigate the legality of various cyber-influence operations and the mechanisms to conduct them, as well as the proposed counter-measures.

**Keywords:** cyber-influence, cyber law, information operations, international law

---

## 1. Introduction

Cyber-enabled influence warfare has been defined as "the deliberate use of information by one party on an adversary to confuse, mislead, and ultimately to influence the choices and decisions that the adversary makes (Lin & Kerr, 2017) The Intelligence and National Security Alliance differentiates between cyber-attacks and influence operations in that a cyber-attack is technical and targets the infrastructure, whereas "in an influence operation, information is weaponized or retooled as a mechanism for social engineering with intent to influence decision-making" (INSA, 2018: 2). However, for Brangetto & Veenendaal (2016) influence operations include activities to demoralise, distract, divide and confuse the target audience in addition to developing a narrative that is convincing and coherent. Darraj, Sample, and Cowley (2017) provide a similar description for what they term as information weaponization, which is tactics to paralyse, subvert, confuse and demoralise the target; in addition they consider the platforms to conduct these activities as social media, trolls (including bots - (automated software that mimics human online behaviour), rogue news media and the traditional media. Cohen and Bar'el (2017) concur that these are the primary tools used.

While influence operations are now receiving increased attention, historically the information environment was not been managed well. As Siegel notes of UN efforts during the Balkan wars in the 1990s, it is sometimes a struggle to simply communicate a message effectively, much less counter adversarial information campaigns (2007: 35):

*..the UN proved unable to manage the information environment. As the mission did not stop the fighting and left ethnic purification intact, the UN protection Force (UNPROFOR) quickly lost credibility with the international press and came under increasing fire. The international press did not fault the UN for not delivering humanitarian aid, but faulted it for letting aggression and ethnic cleansing go unpunished and most of all, unstopped. As ethnic cleansing, concentration camps, and mass murders were uncovered in the summer of 1992, the press became increasingly hostile toward the UN leadership and demanded greater action from the peacekeepers. UN-press relationship became increasingly contentious and hostile, to the point where the UN was unable to communicate its agenda. In fact, it is not unreasonable to say that the UN, in essence, washed its hands and gave up on communication. The hostility came to the point where the UN would send a Spanish spokesman to the daily briefing who was just competent enough in English to read a prepared statement, but not competent enough to answer questions.*

Despite the obvious lessons to be learned from the above scenario, others have continued to warn that the information environment was still not receiving sufficient attention, and governments struggle to deal with both states and non-state actors in this space (Kramer & Wentz, 2008). Likewise, traditional cyber security approaches such as digital signatures, authentication, encryption and access control, are ineffective against cyber-influence tactics such as weaponised information (Cybenko, Giani, & Thompson, 2002). Therefore, deeper investigation of modern influence operations from other perspectives is required.

### **1.1 Cyber-influence operations: Past legal perspectives**

While the body of research around cyber-influence operations is growing, very little considers the topic from a legal perspective. Much of the literature focuses on new concepts such as cognitive hacking (Cybenko, Giani & Thompson, 2002), the international security implications of influence operations (Kramer & Wentz, 2008), or the tactics and effects of such campaigns (Hutchinson, 2009; van Niekerk & Maharaj, 2013; Nissen, 2015; van Niekerk, 2015; Bodine-Baro et al, 2016; Darraj, Sample & Cowley, 2017; Cox et al, 2018, Wanless & Berk, 2019). Other research focuses more on the traditional cyber-specific aspects of influence operations (Davis, 2015; Brangetto & Veenendaal, 2016; Cohen & Bar'el, 2017) or the application of such tactics for deterrence purposes (Siedler, 2016).

However, in the research noted above only Nissen (2015) considers the legal ramifications of the weaponization of social media. A handful of other works have touched on the role of law in countering influence operations, ranging from national legislative and regulatory responses (Bradshaw, Neudert & Howard, 2018) to international law (Melzer, 2011; Schmitt, 2017), but more often in the context of cyber-operations in general, while touching on propaganda, or in specific types of interference, such as in elections (Rotondo and Salvati, 2018). While beginning to fill, there remains a gap in research providing a broader legal analysis on prior cyber-influence operations.

## **2. International law and cyber operations**

Schmitt (2017) in Tallinn Manual 2 states, "with regard to propaganda, ..... transmission into other States is generally not a violation of sovereignty. However, the transmission of propaganda, depending on its nature, might violate other rules of international law. For instance, propaganda designed to incite civil unrest in another State would likely violate the prohibition of intervention (Rule 66)." With reference note 569, Rule 66 (237), it is further stated that "...the transmission of propaganda does not constitute a prohibited intervention as it is not coercive in nature." The nature of coercion is further described as (318-319) being distinguishable from "persuasion, criticism, public diplomacy, propaganda (as in Rule 4), retribution, mere maliciousness, and the like in the sense that, unlike coercion, such activities merely involve either influencing (as distinct from factually compelling) the voluntary actions of the target State, or 318 international peace and security seek no action on the part of the target State at all." It is further stated as an example that "State-sponsored public information campaigns via the Internet designed to persuade another State of the logic of ratifying a particular treaty would not amount to a violation of the prohibition of intervention. Similarly, if a State's Ministry of Foreign Affairs publishes content on social media that is highly critical of another State's internal and external policies, the activity is not coercive in nature and therefore does not constitute prohibited intervention." The point remains that the "coercive act must have the potential for compelling the target State to engage in an action that it would otherwise not take (or refrain from taking an action it would otherwise take)." The contentious area is, however, to distinguish whether an act would constitute intervention without knowledge of the context and/or consequences.

The concept of "Presumptive legality" arises (Schmitt, 2017: 336). By its very nature international law is prohibitive, meaning acts "that are not forbidden are permitted; absent an express treaty or accepted customary law prohibition, an act is presumptively legal." For example, International law "does not prohibit propaganda, psychological operations, espionage, or mere economic pressure per se." Therefore, acts that fall into these categories become presumptively legal. As such, these acts may not be deemed as uses of force by States.

Furthermore, some operations affecting civil society, for example "psychological operations such as dropping leaflets or making propaganda broadcasts are not prohibited even if civilians are the intended audience" (Schmitt, 2017: 421-422). Within the context of cyber warfare, "transmitting email messages to the enemy urging capitulation would likewise comport with the law of armed conflict". "Only when a cyber operation against civilians or civilian objects (or other protected persons and objects) rises to the level of an attack is it

prohibited by the principle of distinction and those rules of the law of armed conflict that derive from the principle."

The issue of whether the use of electronic or other media to spread propaganda qualifies as direct participation in hostilities remains unsettled. With reference to Tallinn 2 (2017:528 -529), it was established by the majority of Experts that "spreading propaganda does not per se constitute direct participation in hostilities" (Rule 97), while the minority stated "that the use of networks or computers to spread propaganda might convert journalistic equipment into a military objective such that they are subject to cyber-attacks."

It was deemed that broadcasts which were used to "incite war crimes, genocide, or crimes against humanity would render a journalist a direct participant and qualify the equipment used as a military objective liable to attack, including by cyber means," by the majority of Experts, with a few in disagreement.

For Melzer a cyber operation could be legally viewed as violation of sovereignty (2011:9): "a cyber operation need not amount to "force" within the meaning of article 2(4) of the UN Charter to be internationally wrongful, nor would all cyber operations amounting to "force" necessarily be unlawful." The legality of a cyber operation can possibly result from a violation of any obligation under International law. For example, "interstate computer network exploitation for the purposes of intelligence gathering, electronic dissemination of hostile propaganda, or denial of service attacks would each violate the sphere of sovereignty of the affected state and, thus, the customary principle of non-intervention, even if they do not qualify as a use of force within the meaning of article 2(4) of the UN Charter."

The UN "Declaration on the Inadmissibility of Intervention and Interference in the Internal

Affairs of States" (1981), puts forward the following:

- "The right of states and peoples to have freely accessible information and to develop fully, without interference, their system of information and mass media and to use their information media in order to promote their political, social, cultural affiliations based on the relevant articles of the Universal Declaration of Human Rights and the principles of the new international information order" (II(c)) ;
- "The duty of a state to abstain from any defamatory campaign, vilification or hostile propaganda for the purpose of intervening or interfering in the internal affairs of other states" (II(j)); and
- "The right and duty of states to combat, within their constitutional prerogatives, the dissemination of false or distorted news which can be interpreted as interference in the internal affairs of other states or as being harmful to the promotion of peace, co-operation and friendly relations among states and nations." (III(d))

The Paris Call for Trust and Security in Cyberspace (2018), aligned to the UN Charter, aims to "strengthen our capacity to prevent malign interference by foreign actors aimed at undermining electoral processes through malicious cyber activities".

The Budapest extension (2003) is an extension to the Budapest Convention on cybercrime. It was established to specifically counter racism and xenophobic propaganda committed or spread through computer systems and networks (Council of Europe, 2003).

South Africa's Cybercrimes Bill (yet to be implemented), was created for the purpose of imposing penalties regarding cybercrime and to "criminalise the distribution of data messages which is harmful and to provide for interim protection orders and regulate jurisdiction in respect of cybercrimes". The Prevention and Combating of Hate Crime and Hate Speech Bill of Apr 2018, was developed to "give effect to South Africa's obligations in terms of the Constitution and international human rights instruments concerning racism, racial discrimination, xenophobia and related intolerance, in accordance with international law obligations and make provisions for the offence of hate crime and the offence of hate speech and the prosecution of persons who commit those offences; to provide for appropriate sentences that may be imposed on persons who commit hate crime and hate speech offences; to provide for the prevention of hate crimes and hate speech; to provide for the reporting on the implementation, application and administration of this Act; to effect consequential amendments to certain Acts of Parliament; and to provide for matters connected therewith".

### **3. Incidents and legal responses**

#### **3.1 2016 US election and foreign attempts at influence**

Russian attempts to shape the information environment during the 2016 U.S. presidential election are well documented. The 2017 Intelligence Community Assessment by the Director of National Intelligence alleged that Russian efforts to influence blended “covert intelligence operations—such as cyber activity—with overt efforts by Russian Government agencies, state-funded media, third-party intermediaries, and paid social media users or “trolls” (DNI, 2017). Such efforts, much of it run through a proxy organization, the Internet Research Agency, included using fake social media accounts to pretend to be Americans (Dwoskin, Entous & Timberg, 2016; Twitter, 2018) sharing content on divisive issues such as racism (USA v. Internet Research Agency, 2018), buying fake Facebook ads on “divisive social and political messages across the ideological spectrum — touching on topics from LGBT matters to race issues to immigration to gun rights’ (Stamos, 2017), reaching and engaging with millions of accounts (Howard et al. 2018; DiResta et al, 2018). There has been further claims of possible collusion between Russia and the Trump campaign, which ultimately led to the broader concern over Russian interference in the democratic process and has caused a debate over possible legal action (Savage, 2017).

In terms of legally addressing such activity, the US has used its electoral laws to indict the Internet Research Agency. According to the indictment, the “law bans foreign nationals from making certain expenditures or financial disbursements for the purpose of influencing federal elections...[and] bars agents of any foreign entity from engaging in political activities within the United States without first registering with the Attorney

General.” (USA v. Internet Research Agency, 2018: 2) The Trump administration also imposed “sanctions on 19 Russians for alleged interference in the 2016 U.S. election, including 13 indicted by special counsel Robert Mueller.” (Lee & Lederman, 2018)

With reference to Tallinn 2 (2017, 237), propaganda or rather its dissemination into other States is generally not regarded as a violation of a State’s sovereignty, although depending on its nature may be in violation of the other rules of international law. As an example, if the nature of such propaganda was created for the inception of civil conflict in another State, it would be in contravention of Rule 66. Russian activity in the 2016 US elections in 2016, does interfere with a States’ sovereignty, however, it is not considered a hostility by definition of Rule 67. The Tallinn Manuals, are however, a non-binding document and cannot be enforced. Furthermore, it is further stated in Tallinn 2 that the transmission of propaganda does not constitute a prohibited intervention as it does not use “force”. However, with reference to Melzer (2011: 9) and article 2(4) of the UN Charter, cyber operation does not need to amount to “force” within the meaning to be internationally wrongful, “nor would all cyber operations amounting to force necessarily be unlawful.”

#### **3.2 Ukraine**

In 2013, Ukraine had undergone a social media-fuelled revolution replacing the pro-Russian government with a pro-EU one. This incident emphasised the role of social media (Satell, 2014), being used for the coordination of the mass demonstrations, the provisioning of support services during the protests, and to leak documents used to allege corruption of the government (van Niekerk, 2015). This led to increasing tensions in the country, with pro-Russian protests and separatist movements from the ethnic Russian population in Ukraine. This eventually led to the Russian incursion into Crimea, which exhibited the use of information operations including online propaganda, cyber-attacks, and communication disruptions; all used to support military objectives (van Niekerk, 2015). Pro-Russian information was prevalent online, the series of claims led to an increasing state of uncertainty amongst the ethnic Russians in Ukraine and, influencing them to lend their support to the Russian incursion and at a later stage resulting in the conflict in Eastern Ukraine, allegedly backed by the Russian military. Due to a large number of inconsistencies in the coverage, numerous claims were refuted leading to claims of fictitious news reporting by both sides. The inconsistencies and social media posts resulted in claims illustrating the Russian interference in the conflict regions, even though Russian authorities had denied it (van Niekerk, 2015). The influence of social media as well as online misinformation and propaganda campaigns is clearly illustrated in the Ukraine conflict. Cohen and Bar'el (2017) refer to, what they consider, the three main tools in the Russian arsenal for influence operations i.e. bots, trolls and hackers. Of importance is the combination of information operations and cyber-warfare tactics, usually employed by Russian forces. These include attacking digital

networks, deception, fraud, disinformation and psychological warfare. Through coordinated operations, the pro-Russian factions exerted information dominance in the region.

The challenge with this scenario is that the propaganda itself is not expressly forbidden and therefore are not considered as a use of force (Schmitt, 2017). However, it is possible to motivate that the information operations resulted in a breach of Ukraine's state sovereignty in that a region (Crimea) was annexed, and an armed conflict emerged in Eastern Ukraine. The extension to the Budapest Convention can apply in any cases containing explicit inciting ethnic conflict. A possible mechanism that may be used to counter such use of propaganda and online misinformation is to legally target the infrastructure used to distribute the misinformation. Bots are often created through the use of malicious computer code, where the owners of the devices are unaware that it is being used for malicious purposes. This falls under cyber-crime, covered by the Budapest Convention, and law enforcement agencies can attempt to disassemble or disrupt the network of bots. Whilst the disruption of the bots may mitigate some of the misinformation through 'unofficial' channels, it will do little to counter any misinformation or propaganda emanating from established media corporations aiding the narrative of a government.

### **3.3 South Africa: Bell Pottinger**

The British PR Firm Bell Pottinger's activities were exposed in South Africa due to the #GuptaLeaks campaign. It was alleged that the company disseminated fake news on social media platforms which increased racial tension in South Africa, which acted as a hinderance to political issues regarding alleged corruption (Sweney, 2017). The residual effects of the campaign to, some extent, still remains as some groups still occasionally use the phrases posted during the campaign. As a result of the unethical practices by the company, it had been placed under administration (Sweney, 2017).

The Budapest extension (2003), which was established to specifically counter racism and xenophobic propaganda committed through computer systems and networks (Council of Europe, 2003), may apply in relation to the Bell Pottinger case, as fake news was disseminated to create racial tension. Although the South African Cybercrimes Bill mirrors sections of the Budapest Convention, the Convention was not ratified by South Africa, due to the issue of sovereignty (Fidler, 2017). The South African Prevention and Combating of Hate Crime and Hate Speech Bill of April 2018, was developed to "give effect to South Africa's obligations in terms of the Constitution and international human rights instruments concerning racism, racial discrimination, xenophobia and related intolerance, in accordance with international law obligations and make provisions for the offence of hate crime and the offence of hate speech" and would again apply to the Bell Pottinger case.

As for the issue of the dissemination of fake news, the Paris Call for Trust and Security in Cyberspace (2018), aligned to the UN Charter would apply. And South Africa's Cybercrimes Bill, when implemented, applies with regard to penalties regarding cybercrime as well as to "criminalise the distribution of data messages which is harmful and to provide for interim protection orders and regulate jurisdiction in respect of cybercrimes".

## **4. Conclusion**

Many countries have begun developing and instituting cyber-security legislation globally however, the main challenge is developing legislation to counter what has been termed "cyber-influence". From a brief analysis of existing legislation, both nationally and abroad, it can be argued that "cyber-security" legislation does not adequately, if at all, address or provide any recourse for "cyber-influence operations". In fact, the main issue is whether cyber-influence operations are deemed as criminal activity, and if so, which legislation would apply.

Cyber-Influence Operations are used with the aim of overcoming an adversary using information. As discussed in this paper the actors using such techniques vary and range from political actors seeking to influence domestic audiences in the context of the South African election, foreign states engaged in conflict as between Russia and Ukraine, and grey area proxy operators as in the example of the 2016 U.S. election. Cyberspace has been blanketed in a digital fog of war, with influence operations proving to be disruptive and destructive.

Attribution of cyber influence creates a challenge as it becomes difficult to pinpoint and identify the perpetrators as illustrated by the distribution of fake news by the Russians using "bots", supposedly during the 2016 US elections.

From the few scenarios provided, a distinction must be drawn between what constitutes a “cyber-attack” and how it is different from “cyber-influence”. Cyber-attacks are “socially, or politically motivated attacks carried out primarily through the Internet. Attacks target the general public or national and corporate organizations and are carried out through the spread of malicious programs (viruses), unauthorized web access, fake websites, and other means of stealing personal or institutional information from targets of attacks, causing far-reaching damage” (NEC, 2018).

Legal challenges of governing cyber influence operations have emerged, with countries attempting to develop necessary legislation. While some existing legislation is already in place, it may only address cyber influence operations in part. The main challenge remains the ability to create legislation to meet the technical definition of cybercrime or what constitutes a cybercrime, without denying the citizens rights to freedom of expression and privacy in modern democracies as provisioned for in the Declaration on the Inadmissibility of Intervention and Interference in the Internal Affairs of States.

## Acknowledgements

The third author received funding by the South African National Research Foundation, Grant no. 115059.

## References

- Bodine-Baron, E., Helmus, T.C., Magnuson, M., and Winkelman, Z. (2016) *Examining ISIS Support and Opposition Networks on Twitter*, Santa Monica: Rand Corporation.
- Bolgov, R., Filatova, O., and Tarnavsky, A. (2016) "Analysis of Public Discourse About Donbas Conflict in Russian Social Media," *Proceedings of the 11th International Conference on Cyber Warfare and Security*, Boston, USA, 17-18 March, pp. 37-46.
- Bradshaw, S., Neudert, L., and Howard, P.N. (2018) *Government Responses to Malicious use of Social Media*, November, Riga: NATO STRATCOM COE.
- Brangetto, P., and Veenendaal, M.A. (2016) "Influence Cyber Operations: The Use of Cyberattacks in Support of Influence Operations," *8th International Conference on Cyber Conflict*, N. Pissanidis, H. Röigas, M. Veenendaal (Eds.), pp. 113-126.
- Chong, Z. (2017) "India 'helpless' against fake news spreading through WhatsApp," CNET, 31 July, [online], accessed 16 November 2017, <https://www.cnet.com/news/india-helpless-against-spread-of-offensive-content-on-whatsapp/>.
- Cohen, D., and Bar'el, O. (2017) *The Use of Cyberwarfare in Influence Operations*, October, Tel Aviv: Yuval Ne'eman Workshop for Science, Technology and Security.
- Council of Europe. (2003) *Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems*, European Treaty Series - No. 189.
- Cox, K., Marcellino, W., Bellasio, J., Ward, A., Galai, K., Meranto, S., and Paoli, G.P. (2018) *Social Media in Africa: A Double-edged Sword for Security and Development*, 5 November, United National Development Programme, [online], accessed 23 January 2019, <http://www.africa.undp.org/content/rba/en/home/library/reports/social-media-in-africa-.html>.
- Cybenko, G., Giani, A., and Thompson, P. (2002) "Cognitive hacking a battle for the mind", *Computer* 35(8), pp. 50 – 56.
- Darraj, E., Sample, C., and Cowley, J. (2017) "Information Operations: The use of Information Weapons in the 2016 US Presidential Election", *Proceedings of the 16<sup>th</sup> European Conference on Cyber Warfare and Security*, Ireland, 29-20 June, pp. 92-101.
- Davis, P.K. (2015) "Deterrence, Influence, Cyber Attack, and Cyberwar," *International Law and Politics*, vol. 47, pp. 327-355.
- DiResta, R., Shaffer, K., Ruppel, D., Sullivan, R. M., Fox, R., Albright, J., Johnson, B. (2018) "The Tactics & Tropes of the Internet Research Agency." New Knowledge
- Dixon, R. (2017). "Zimbabwe jails American activist-journalist over tweet critical of Robert Mugabe", *Los Angeles Times*, 4 November, 2017, [online], accessed 4 February 2019, <https://www.thestar.com/news/world/2017/11/04/zimbabwe-jails-american-activist-journalist-over-tweet-critical-of-robert-mugabe.html>
- DNI, United States Director of National Intelligence, (7 January 2017) "Intelligence Community Assessment: Assessing Russian Activities and Intentions in Recent US Elections., [https://www.dni.gov/files/documents/ICA\\_2017\\_01.pdf](https://www.dni.gov/files/documents/ICA_2017_01.pdf) (accessed on April 12, 2018).
- Dwoskin, E., Entous, A., and Timberg, C. (2016) "Google uncovers Russian-bought ads on YouTube, Gmail and other platforms," *Washington Post*, 24 November, [online], accessed 16 November 2017, <https://www.washingtonpost.com/news/the-switch/wp/2017/10/09/google-uncovers-russian-bought-ads-on-youtube-gmail-and-other-platforms/>.
- Fidler, M. (2017). "South Africa Introduces Revised Cybercrime Legislation, Acknowledging Criticism." Net Politics [online], accessed 12 March 2019 <https://www.cfr.org/blog/south-africa-introduces-revised-cybercrime-legislation-acknowledging-criticism>
- Frykberg, M. (2017) "US woman faces 20 years in prison over Mugabe insult," *Independent Online*, 6 November, [online], accessed 6 November 2017 <https://www.iol.co.za/news/africa/us-woman-faces-20-years-in-prison-over-mugabe-insult-11874326>.

- Gu, L., Kropotov, V., Yarochkin, F., Leopando, J., and Estialbo, J. (2017) "Fake News and Cyber Propaganda: The Use and Abuse of Social Media," *Trend Micro*, 13 June, [online], accessed 28 June 2017, <https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/fake-news-cyber-propaganda-the-abuse-of-social-media>.
- Howard, P. N., Ganesh, B., Liotsiou, D., Kelly, J., and François, C. (2018) "The IRA, Social Media and Political Polarization in the United States, 2012-2018." Computational Research Project, University of Oxford.
- Hutchinson, W. (2009) Cyber Influence, *Proceedings of the 10th Australian Information Warfare and Security Conference*, Perth Western Australia, 1st-3rd December.
- Intelligence and National Security Alliance. (2018) *Getting Ahead of Foreign Influence Operations*, May, [online], accessed 19 January 2019, <https://www.insaonline.org/getting-ahead-of-foreign-influence-operations-may-2018/>.
- Kramer, F.D., and Wentz L. (2008) Cyber Influence and International Security, *Defense Horizons*, no. 61, January, National Defense University Center for Technology and National Security Policy.
- Lee, M, and Lederman, J. (2018). "U.S. imposes sanctions on 13 Russians indicted by Robert Mueller", PBS, accessed on 6 March 2019 <https://www.pbs.org/newshour/politics/u-s-imposes-sanctions-on-13-russians-indicted-by-robert-mueller>
- Lin, H., and Kerr, J. (2017) On Cyber-Enabled Information/Influence Warfare and Manipulation, working paper.
- Lunghi, D. (2017) "From Cybercrime to Cyberpropaganda," *Trend Micro*, 16 October, [online], accessed 13 November 2017, <http://blog.trendmicro.com/trendlabs-security-intelligence/from-cybercrime-to-cyberpropaganda/>.
- Madrigal, A. (2011) "The Inside Story of How Facebook Responded to Tunisian Hacks", *The Atlantic*, 24 January, [online], accessed 25 January 2011, <http://www.theatlantic.com/technology/archive/2011/01/the-inside-story-of-how-facebook-responded-to-tunisian-hacks/70044/#>.
- McDonald-Gibson, C. (2017) "The E.U. Agency Fighting Russia's Wildfire of Fake News with a Hosepipe," *Time*, 11 September, [online], accessed 12 September 2017, <http://time.com/4887297/europe-fake-news-east-stratcom-kremlin/>.
- Melzer, N. (2011) *Cyberwarfare and International Law*, UNIDIR, [online], accessed 22 January 2019, <http://unidir.org/files/publications/pdfs/cyberwarfare-and-international-law-382.pdf>.
- Meredith, S. (2018) "Here's everything you need to know about the Cambridge Analytica scandal," *CNBC*, 21 March, [online], accessed 25 January 2019, <https://www.cnbc.com/2018/03/21/facebook-cambridge-analytica-scandal-everything-you-need-to-know.html>.
- Minister of Justice and Correctional Services. (2017) Cybercrimes and Cybersecurity Bill. Republic of South Africa.
- NEC.(2018). What constitutes a cyber-attack? [online], accessed 06 February 2019, [https://www.nec.com/en/global/solutions/safety/info\\_management/cyberattack.html](https://www.nec.com/en/global/solutions/safety/info_management/cyberattack.html)
- Nissen, T.E. (2015) *The Weaponization of Social Media*, Copenhagen: Royal Danish Defence College.
- Paris Call for Trust and Security in Cyberspace (2018), 12 November, [online], accessed 18 January 2019, [https://www.diplomatie.gouv.fr/IMG/pdf/paris\\_call\\_cyber\\_cle443433.pdf](https://www.diplomatie.gouv.fr/IMG/pdf/paris_call_cyber_cle443433.pdf)
- Roberts, R. (2017) "Russia hired 1,000 people to create anti-Clinton 'fake news' in key US states during election, Trump-Russia hearings leader reveals," *The Independent*, 30 March, [online], accessed 16 November 2017, <http://www.independent.co.uk/news/world/americas/us-politics/russian-trolls-hillary-clinton-fake-news-election-democrat-mark-warner-intelligence-committee-a7657641.html>.
- Rotondo, A., and Salvati, P. (2018) "Fake news, (dis)information and principle of non-intervention," CyCon US 2018, [online], accessed 28 January 2019, <https://cyberdefensereview.army.mil/Portals/6/Documents/CyConUS18%20Conference%20Papers/Session4-Paper3.pdf?ver=2018-11-13-160900-917>.
- Satell, G. (2014) "If you doubt that social media has changed the World, take a look at Ukraine," *Forbes*, 18 January, [online], accessed 16 November 2017, <https://www.forbes.com/forbes/welcome/?toURL=https://www.forbes.com/sites/gregsatell/2014/01/18/if-you-doubt-that-social-media-has-changed-the-world-take-a-look-at-ukraine/>
- Savage, C. (2017) "Trump Campaign Is Sued Over Leaked Emails Linked to Russians," *NY Times*, 12 July, [online], accessed 15 November 2017, <https://www.nytimes.com/2017/07/12/us/politics/trump-campaign-and-adviser-are-sued-over-leaked-emails.html>
- Sawant, N. (2017) "The Fake News problem in India can get out of hand if not controlled in due time; experts tell us why," *FirstPost*, 17 February, [online], accessed 6 March 2017, <http://tech.firstpost.com/news-analysis/the-fake-news-problem-in-india-can-get-out-of-hand-if-not-controlled-in-due-time-experts-tell-us-why-362791.html>.
- Schmitt, M.N. (2017) *Tallinn Manual 2.0: On The International Law Applicable to Cyber Operations*, Cambridge: Cambridge University Press.
- Shuster, S. (2017) "Russia Has Launched a Fake News War on Europe. Now Germany Is Fighting Back," *Time*, 9 August, [online], accessed 10 August 2017, <http://time.com/4889471/germany-election-russia-fake-news-angela-merkel/>.
- Siedler, R.E. (2016) " Hard Power in Cyberspace: CNA as a Political Means," 8th International Conference on Cyber Conflict, N. Pissanidis, H. Rõigas, M. Veenendaal (Eds.), pp. 23-36.
- Siegel, P.C. (2007) "Perception Management: IO's Stepchild," In: *Information Warfare*, E.L. Armistead (ed.), Washington, DC: Potomac Books, pp. 27-44.

**Trishana Ramluckan, Alicia Wanless and Brett van Niekerk**

- Stamos, A. (2017). An Update On Information Operations On Facebook. Facebook.  
<https://newsroom.fb.com/news/2017/09/information-operations-update/> (accessed 26 January 2019)
- Swaney, M. (2017) "Bell Pottinger goes into administration amid South Africa scandal," *The Guardian*, 12 September, [online], accessed 16 November 2017, <https://www.theguardian.com/media/2017/sep/12/bell-pottinger-goes-into-administration>.
- Timberg, C. (2016) "Russian propaganda effort helped spread 'fake news' during election, experts say," *Washington Post*, 24 November, [online], accessed 16 November 2017, [https://www.washingtonpost.com/business/economy/russian-propaganda-effort-helped-spread-fake-news-during-election-experts-say/2016/11/24/793903b6-8a40-4ca9-b712-716af66098fe\\_story.html](https://www.washingtonpost.com/business/economy/russian-propaganda-effort-helped-spread-fake-news-during-election-experts-say/2016/11/24/793903b6-8a40-4ca9-b712-716af66098fe_story.html).
- Twitter. (2018) Update on Twitter's Review of the 2016 US Election. Twitter Public Policy..  
[https://blog.twitter.com/official/en\\_us/topics/company/2018/2016-election-update.html](https://blog.twitter.com/official/en_us/topics/company/2018/2016-election-update.html) (accessed 26 January 2019)
- United States Department of Justice Indictment, (2018) United States of America v. Internet Research Agency, et.al., filed February 16, 2018, in the US District Court for the District of Columbia. Case 1:18-cr-00032-DLF, available at <https://www.justice.gov/file/1035477/download> (accessed 26 January 2018)
- van Niekerk, B. (2015) "Information Warfare in the 2013-2014 Ukraine Crisis," in: *Cybersecurity Policies and Strategies for Cyberwarfare Prevention*, J. Richet, Hershey, PA: IGI Global, pp. 307-339.
- van Niekerk, B. and Maharaj, M. (2013) "Social Media and Information Conflict," *International Journal of Communication*, Vol. 7, pp. 1162–1184.
- Wanless, A., and Berk, M., (2019, in press) "The Audience is the Amplifier: Participatory Propaganda". In P Baines, N O'Shaughnessy & N Snow (Eds.),*The Sage Handbook of Propaganda*. Sage: London
- Weber, R.H. (2011) "Politics through Social Networks and Politics by Government Blocking: Do We Need New Rules?," *International Journal of Communication*, vol. 5, 2011, pp. 1186–1194. Available:  
<http://ijoc.org/index.php/ijoc/article/view/1167/605>
- World Movement for Democracy. (2009) *Twitter case study*, [online], accessed 17 June 2010,  
<http://www.wmd.org/resources/whats-being-done/information-and-communication-technologies/case-study-twitter>.

# Online Expression and Spending on Personal Cybersecurity

Juhani Rauhala<sup>1</sup>, Pasi Tyrväinen<sup>1</sup> and Nezer Zaidenberg<sup>2</sup>

<sup>1</sup>University of Jyväskylä, Finland

<sup>2</sup>College of Management Academic Studies, Rishon LeZion, Israel

[juhani.jr.rauhala@jyu.fi](mailto:juhani.jr.rauhala@jyu.fi)

[pasi.tyrvainen@jyu.fi](mailto:pasi.tyrvainen@jyu.fi)

[scipio@scipio.org](mailto:scipio@scipio.org)

**Abstract:** The Internet is used increasingly as a platform both for free expression and e-commerce. Internet users have a variety of attitudes towards the security and privacy risks involved with using the Internet; and distinct concerns and behaviors with regard to expressing themselves online. Users may have controversial viewpoints that they may express online in various ways. Controversial viewpoints or artwork by their nature may not be as well received as positive or polite expressions. In the online environment, users with controversial viewpoints may be reluctant to express the viewpoints due to concern about possible consequences resulting from the expressions. Consequences may be imposed by individuals, groups, organizations, businesses, or nation-states. Examples of such consequences include firings, removal of forum posting privileges (“banning”), violent attacks, online stalking, and doxing. Users may also have different attitudes towards personal spending of money for cybersecurity products and services. Factors such as concern about the risks associated with free expression online may impact their attitudes towards spending for personal cybersecurity. We perform a factor analysis on survey data. Our goal is to establish variables for expression reluctance, and attitude towards personal cybersecurity purchasing. The positive attitude toward spending on personal cybersecurity, as a factor, includes reported activity of purchasing cybersecurity products or services, and an overall generally positive attitude toward the purchasing of such products or services. We propose a research model that enables an analysis of the relationship between the reluctance to make controversial expressions online and a positive attitude toward spending money on personal cybersecurity products and services. We perform a correlation analysis between the factors. Results indicate that there is a correlation between users’ reluctance to express controversial messages online, and a positive attitude towards spending money on personal cybersecurity. Future work will include additional analyses, including the effects of various demographic factors.

**Keywords:** online expression reluctance, personal cybersecurity spending, privacy concerns, online spending, risk avoidance

---

## 1. Introduction

One generally accepted beneficial use of the Internet is as a platform for commerce, which is continuously increasing (Emarketer.com, 2014). At the same time, spending by consumers and businesses on cybersecurity products and services is also increasing (Morgan, 2017). It is reasonable to expect that a significant proportion of personal cybersecurity software is being purchased online. Another commonly accepted benefit of the Internet is that it serves as a platform for free expression. Debate and discussions occur over online forums and social media such as Twitter and Facebook. These discussions are raising attention to a virtually unlimited array of topics. Importantly, political topics are also discussed as well as other topics without socially accepted *savoir-faire*. In oppressive nation-states, the free expression enabled by access to the Internet can be particularly important for increasing the possibilities for improved human rights (Nadi and Firth, 2004). However, there are potential adverse consequences for users making controversial or provocative expressions on the Internet including from governments (Baroni, 2015; Cooper, 2000; Mony, 2017), offended individuals (Cassidy, 2017), employers (Jaschik, 2014), and schools (Curтом, 2014). Concern about such consequences may not only have an inhibiting effect on users’ use of the Internet for expression but it may also correlate with their desire to purchase personal cybersecurity products and anonymizing services. These effects may differ across certain demographic groupings. It is possible that misgivings of users about the Internet as a platform for free expression may correlate with increased Internet utilization by those same users for commerce in personal cybersecurity products and services. This study explores this somewhat paradoxical relationship given that the Internet is seen as an overall good for humanity.

This paper first presents an overview of previous related research, followed by a description of the research model. It then establishes two general latent factors. The first corresponds to a reluctance to self-express online. The second factor corresponds to a positive predilection toward personal spending to enhance personal cybersecurity. The correlation between these factors is then analyzed. The results are presented and discussed, followed by a conclusion and a description of future research goals.

## **2. Background**

Booth (2017) has raised some attention to the issue of freedom of expression. Her work examines the relationships between laws and norms governing free expression, and the benefits of ICT on national well-being. As of the time of this writing, Booth's research is not yet complete; moreover, the research does not consider the expression of free speech on aspects of individual users. It is of note that Booth and other researchers utilize the Human Freedom Index (HFI) (Vasquez and Porcnik, 2017). Included in the HFI measures are those that measure freedom of expression. Among those measures are "Laws and Regulations that Influence Media Content," "Political Pressures and Controls on Media Content," and "State Control over Internet Access." The measures of Laws and Regulations that Influence Media Content and Political Pressures and Controls on Media Content could be useful for this study on the condition that they be applied indirectly. That is to say, for example, that an assumption would be that an average user would feel some reluctance to freely express themselves as a result of the laws and controls. This study addresses reluctance more directly in the survey questions, whereas the subset of HFI measures does not measure reluctance to express. The HFI's "expression freedom" measures have not been examined for their relationship to personal cybersecurity spending. In particular, they do not measure concern regarding the consequences of personal free expression and neither have they been analyzed for their relationship to Internet users' attitudes and behaviors toward purchasing personal cybersecurity protections. Other research has established that usage of the Internet for free expression can be a way of circumventing censorship or other hindrances preventing citizens' free expression in more traditional publishing methods, especially in authoritarian regimes (Nadi and Firth, 2014).

There are also studies observing the impact of demographic factors on Internet users' behavior relevant to this study. Research into culture-based differences in the perception of risk in online shopping and other tasks has yielded conflicting results. For example, Sims and Xu (2012) found no significant difference between UK and Chinese shoppers' perceived risk of online shopping despite those shoppers' differing cultural backgrounds. This was against expectations based on the results of prior similar research. However, Chen, Hsu, and Lin (2010) have studied consumers with different levels of computer expertise. They determined that consumers' preferences of attributes of shopping websites differ according to their levels of expertise. Sheehan (2002) found that users' education and age correlate with their level of concern about online privacy. Regan, Fitzgerald, and Balint (2014) evaluated attitudes toward information privacy between age groups (specifically generations). Their analysis revealed a trend where younger generations tend to be more concerned than older ones about wiretapping and data privacy. Hazari and Brown (2013) studied whether demographic variables can affect Internet users' privacy concerns and, thus, their attitudes toward using social networking sites. In contrast to the results from Sheehan and from Regan, Fitzgerald, and Balint, their research found that age was not correlated with online privacy concerns. Bandyopadhyay (2011) found that factors such as level of Internet literacy, social awareness, and cultural background affect Internet users' online privacy concerns. He found that among the possible consequences of such concerns is an unwillingness to use the Internet. Liu et al (2016) applied social exchange theory to examine perceived risks and rewards of individual users' self-disclosure in social media. The authors found that perceived privacy risk can reduce the willingness of social media users to disclose personal information. There does not seem to be existing research on social exchange theory applied to controversial expression by individual users online. This study directly assesses the reluctance to express controversial viewpoints. It also assesses the reluctance that is caused by concern about consequences. Previous work has examined the effect on willingness to disclose information about oneself. Based on previous research, it can be hypothesized that the reluctance to express oneself on the Internet may be connected with concerns about the consequences. Further, reluctance to express oneself may lead to the use of cybersecurity as a means to protect oneself in these cases. However, there seems not to be previous results addressing this hypothesis.

This research analyzes whether users are more inclined to spend money on personal cybersecurity if they are reluctant to express themselves online. The researchers consider that it is important to consider the attitudes of users toward free expression on the Internet and possible consequences resulting from users' reluctance to freely express themselves on the Internet. This is relevant to participation in social media and other online expression contexts. The relationship between online expression aspects and personal cybersecurity spending seems to be lacking in prior research.

### 3. Research model

Based on the research questions raised in the previous section, this study analyzes the correlation between individuals' personal cybersecurity spending (referred to as "Loss of Money," LoM) and their reluctance to express themselves online (RtoEx). The reluctance of expressing is further divided into two factors based on inclusion or exclusion of consequences of the expression, RtoExC and RtoExnonC, respectively. The research model is presented in Figure 1. The hypotheses are as follows.

#### 3.1 Hypotheses

*H1: Users' refusal or reluctance to express themselves online (RtoEx) is correlated with their personal cybersecurity spending attitude and behavior (LoM).*

*H2: The correlation of H1 will vary by certain demographic factors.*

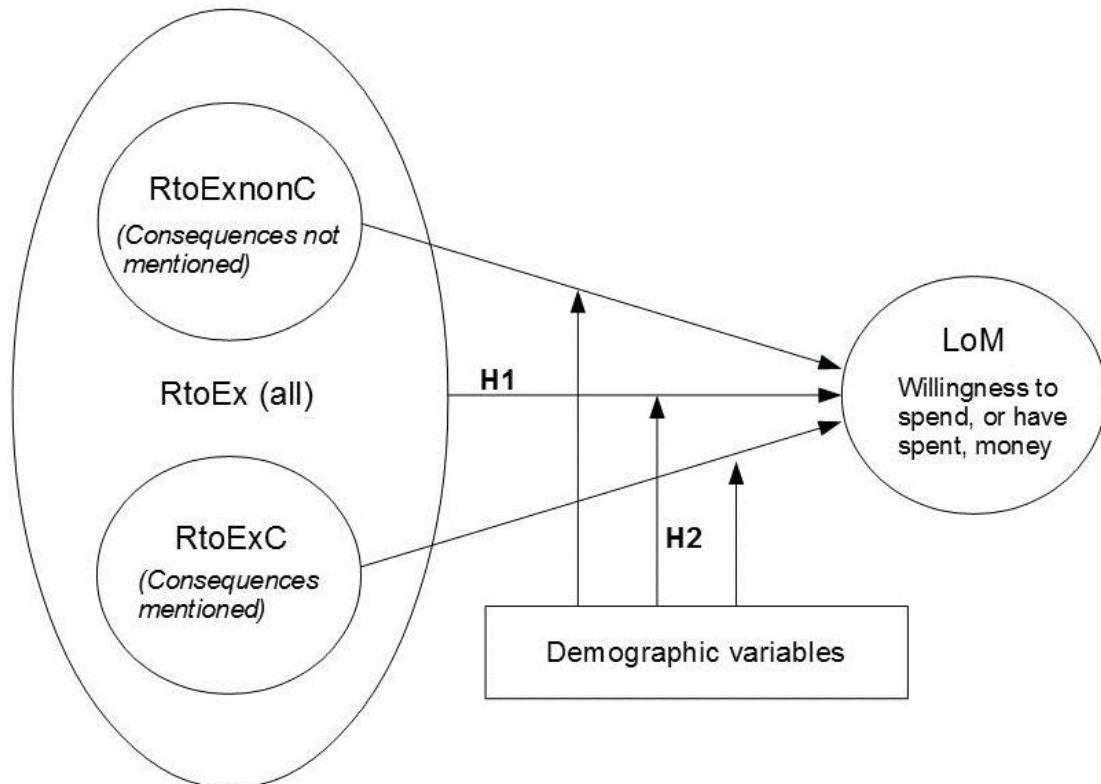


Figure 1: Latent variables RtoEx and LoM, and the independent demographic variables

### 4. Method

#### 4.1 Operationalizing the model

Latent variables Loss of Money (LoM) and Reluctance to Express (RtoEx) are introduced. Each latent variable is defined by responses to respective sets of indicator questions. LoM is defined by four indicator questions and RtoEx by eight indicator questions. The questions for LoM are designed as follows: two questions to ascertain whether the respondent/subject has actually made a purchase for the purpose of enhancing his cybersecurity and two questions to ascertain the general attitude of the respondent toward security software purchases. Cumulatively, it is suggested the LoM indicator questions indicate the willingness to buy software products or services that enhance personal cybersecurity.

The questions for the RtoEx variable are designed as follows: the questions ascertain the attitude of the respondent toward theoretical scenarios of his/her posting controversial opinions or artwork online, including one question to ascertain his/her attitude toward using electronic methods vs. face-to-face communication for discussion of a sensitive topic with a friend. It is suggested that this set of RtoEx indicator questions can convey the level of the respondent's reluctance to openly communicate using electronic methods, or the Internet.

For data gathering, a survey was administered over the Web to a population composed mainly of university students and working adults. The survey included questions on respondents' behaviors and attitudes regarding personal spending on cybersecurity, and on their attitude toward posting or discussing controversial subjects online (Appendix). The questions were answered using a five-point Likert scale, ranging from "strongly agree" to "strongly disagree." The survey produced 191 useful responses. Responses by nationality include Finland (131 responses), USA (28 responses), and Israel (14 responses). The age groups of respondents are given in Table 1. The average age was approximately 31 years.

**Table 1:** Respondents by age group

n	15-25	26-36	37-44	45-54	55-64	65+
191	84	53	20	20	6	3

## 5. Results

### 5.1 Correlations between indicator questions for each latent variable

Because the response data to the indicator questions consist of Likert rankings to ordered categories, a Spearman correlation analysis is used. The results show a high correlation among responses to the four LoM indicator questions (Appendix, Table 4). The lowest correlation is .500 and the highest .863, all with two-star significance at the 0.01 level (two-tailed). Cronbach's alpha for the LoM questions is .871. The results also show a high correlation among responses to the eight RtoE indicator questions (Appendix, Table 5). The lowest correlation is .198 and the highest .699, all with two-star significance at the 0.01 level (two-tailed). Cronbach's alpha for the RtoEx questions is .838. For the four RtoExnonC questions, Cronbach's alpha is .764. For the four RtoExc questions, Cronbach's alpha is .796. Because the indicator questions for each latent variable (component) have high intercorrelation, the mean scores of the responses were computed and utilized for analysis. The Cronbach's alpha values are acceptable for good internal consistency within the sets of indicator questions (Table 2).

**Table 2:** Spearman correlations (two-tailed significance at 0.01 level) between indicator question responses for each latent factor, mean correlations, and Cronbach's alpha

Latent Factor	Minimum	Maximum	Mean	Cronbach's Alpha
LoM	.500**	.863**	.639	.871
RtoEx	.198**	.699**	.395	.838
RtoExnonC	.359**	.564**	.457	.764
RtoExc	.292**	.699**	.490	.796

### 5.2 Correlations between latent variables

Pearson correlation analysis is performed on LoM as the dependent variable and RtoEx, RtoExnonC, and RtoExc as the independent variables. For all respondents, there is a correlation of .199\*\* between RtoEx and LoM (Table 3). This correlation increases to .201\*\* when respondents were presented with a consequences or safety issue in the survey question. For responses to RtoEx questions without a mention of consequences or safety, the correlation decreases to .149\*. It might be expected that respondents who are reluctant to express themselves due to a concern about resulting consequences would have a more positive attitude toward spending money on their personal cybersecurity. There is a significant correlation between LoM and RtoExc. The correlation between LoM and RtoExnonC is weaker. Age was not correlated with LoM. A linear regression analysis for LoM was performed using age and the RtoExc factor as independent variables. This showed some correlation (adjusted R squared = .037, p-value = .011). Therefore, H2 is validated for age.

**Table 3:** Pearson correlations between RtoEx and LoM (two-tailed significances: \* to 0.05 level; \*\* to 0.01 level)

n=191	RtoEx	RtoExnonC (consequences not mentioned)	RtoExc (consequences mentioned)
LoM	.199**	.149*	.201**

## 6. Discussion

For behaviors and attitudes toward personal cybersecurity spending (LoM), and attitudes toward making controversial expressions online (RtoEx), the results showed significant correlation. This confirms H1. Some users

who are reluctant to freely express controversial viewpoints online not only deprive themselves of making the online expressions, but they also divert some of their purchasing power toward personal cybersecurity. Whether RtoEx has a causal role in the spending diversion has not been established.

With regard to correlation between LoM and RtoExnonC or RtoExC, the strongest correlation was between LoM and RtoExC. This may be expected because the respondent who is concerned about safety or consequences can have more motivation to protect their device than a respondent who is reluctant to express themselves for reasons not related to safety or consequences. When the correlation between LoM and RtoExnonC is examined, a correlation is seen there as well, though not as significant as between LoM and RtoExC. Internet users who are not as concerned about safety issues or consequences of freely expressing controversial topics online do still have concerns about personal cybersecurity for other reasons. These users have a favorable attitude toward purchasing, or have purchased, cybersecurity products and services to a lesser extent than users who are concerned about consequences or safety issues of controversial online expression.

## **7. Limitations of the study**

The survey was administered in English. Thus, the accuracy of the results may be tainted by limitations in non-native English speakers' understanding of the survey questions. There were no survey questions to assess the English language competence of the respondents. The nationality of individual respondents could be indirectly used to gauge the reliability of individual responses; e.g., if a respondent indicates that their nationality is of a country whose official languages do not include English. In this case, inconsistency between responses of a question category may be at least partially explained by a possible non- or misunderstanding of the questions. In this study, it is generally assumed that all respondents have a sufficient understanding of all the survey questions. This assumption is supported by the fact that the respondents who are not native English speakers are, for the most part, university students or academic professionals.

This study did not consider free and open source personal cybersecurity products and tools that are available. Such tools include Tor browser, ClamAV, and free VPN services. Some respondents may have responded negatively to the survey questions regarding spending because they believe that they can achieve sufficient personal cybersecurity without spending money doing so.

## **8. Conclusion**

This research demonstrated a significant correlation between Internet users' reluctance to controversially express themselves online and a positive proclivity toward personal cybersecurity spending. The correlation was even stronger for those users concerned about safety issues and consequences that could result from controversial online expression. It may be inferred that these concerned users are more actively making purchases of cybersecurity products and services. While sales of cybersecurity products and services are good for the cybersecurity industry, they also indicate the real cybersecurity concerns of Internet users. Many Internet users go online, but are then reluctant to freely express themselves, spending their time and money to alleviate perceived cybersecurity risks. This scenario is not the ideal or optimal use of the Internet by society. Future research can investigate methods to encourage free expression online and reduce the perceived risks of such free expression. An extension of this research can be to explore on which topics users are less inclined to express their opinions online.

From the viewpoint of encouraging open and robust political discourse, governments should ensure the framework and conditions for free expression by their citizens with online regulatory safeguards that correspond to the traditional safeguards in traditional communications media. This could help Internet users feel freer to spend money for personal interests instead of diverting spending due to concerns about their online privacy and security.

The HFI may be enhanced by the inclusion of a measure to assess citizens' reluctance to express legal, but controversial, viewpoints online. Citizens may be reluctant to express such viewpoints despite states' official policies allowing free expression. The concern about consequences resulting from such expression may not necessarily align with states' official policies and the possibility of state-imposed consequences does not necessarily align with states' official policies. The current HFI does not account for citizens' concerns and perceptions of these issues.

Applied social exchange theory could be expanded to account for Internet users' reluctance to freely express their thoughts and opinions online. Further research could explore the factors that inhibit users from expressing controversial viewpoints and factors that encourage such expression online.

In future, the correlation between the studied factors and additional demographic factors will be analyzed as well as a regression analysis of these factors against them. This research will continue to determine the effects of some independent variables (e.g., income and ICT expertise) on hypothesis H1. The work will explore the relationship of certain demographic variables to personal cybersecurity spending and to any reluctance to express oneself online. Subject to available survey data, analysis for geographical region clustering and other clustering may be performed.

## **Appendix 1**

**Table 4:** Survey questions comprising the Loss of Money (LoM) component

<b>Loss of Money (LoM) (has spent money, or positive attitude toward spending)</b>
1. I have paid for security software that was not already included in my device.
2. It is worth spending money to sufficiently protect my device and software from security threats.
3. I have purchased security software for my device, such as an antivirus or firewall suite, or any software to protect my privacy, such as encryption software or data trail deletion software.
4. The amount of money it would cost to sufficiently protect my system and data from cyberattacks would be worth it.

**Table 5:** Survey questions comprising the Reluctance to Express (RtoEx, RtoExnonC, RtoExC) component

<b>Reluctance to Express (RtoEx) (Reluctance to Express, with or without mention of consequences)</b>
1. I would never post a controversial message in an online forum.
2. If I have a controversial opinion about something, I'm hesitant to publish it on the Internet.
3. I am, or would be, reluctant to display any of my controversial artwork (writing, music, drawings, etc.) online.
4. It's usually not a good idea to post controversial comments or opinions online.
5. I would never post a controversial message in an online forum, because someone or some organization could get revenge against me.
6. I have decided against posting my political opinion on a discussion forum/message board, because I was concerned about consequences to myself or to someone I care about.
7. When discussing something with a good friend, I feel more safe to express controversial opinions face-to-face than by electronic communication.
8. I have decided against posting my controversial opinion on a discussion forum, because of concern that someone, or some organization (including government), might use it against me in the future.

## **References**

- Bandyopadhyay, S. (2011) 'Antecedents And Consequences Of Consumers Online Privacy Concerns', *Journal of Business & Economics Research (JBER)*, 7(3). doi: [10.19030/jber.v7i3.2269](https://doi.org/10.19030/jber.v7i3.2269).
- Baroni, D. (2015) 'New Zealand Government To Punish Online Trolls With Prison Time', *Reaxxion.com*, 3 July. Available at: <http://www.reaxxion.com/10115/new-zealand-government-to-punish-online-trolls-with-prison-time> (Accessed: 24 January 2019).
- Booth, R. E. (2017) 'The Effect of Freedom of Expression and Access to Information on the Relationship between ICTs and the Well-being of Nations.', in *Proceedings of the 23nd Americas Conference on Information Systems*.

- Cassidy, P. (2017) 'Man petrol bombed homes in revenge for Facebook post', *STV News*, 3 November. Available at: <https://stv.tv/news/east-central/1401461-man-petrol-bombed-houses-in-revenge-for-facebook-post/> (Accessed: 24 January 2019).
- Chen, Y.-H., Hsu, I.-C. and Lin, C.-C. (2010) 'Website attributes that increase consumer purchase intention: A conjoint analysis', *Journal of Business Research*, 63(9–10), pp. 1007–1014. doi: [10.1016/j.jbusres.2009.01.023](https://doi.org/10.1016/j.jbusres.2009.01.023).
- Cooper, A. K. (2000) 'China: Government punishes Internet journalists', *Committee to Protect Journalists*, 12 July. Available at: <https://cpj.org/2000/07/china-government-punishes-internet-journalists.php>.
- Curtom, G. (2014) 'Students punished for expressing free speech on Twitter', *The Cougar*, 24 April. Available at: <http://thedailycougar.com/2014/04/24/students-punished-expressing-free-speech-twitter/> (Accessed: 24 January 2019).
- eMarketer.com (2014) 'Worldwide Ecommerce Sales to Increase Nearly 20% in 2014 - eMarketer', *eMarketer.com*. Available at: <https://www.emarketer.com/Article/Worldwide-Ecommerce-Sales-Increase-Nearly-20-2014/1011039> (Accessed: 24 January 2019).
- Hazari, S. and Brown, C. (2013) 'An Empirical Investigation of Privacy Awareness and Concerns on Social Networking Sites', *Journal of Information Privacy and Security*, 9(4), pp. 31–51. doi: [10.1080/15536548.2013.10845689](https://doi.org/10.1080/15536548.2013.10845689).
- Jaschik, S. (2014) 'Interview with professor fired by West Bank university who compares himself to Steven Salaita', *Inside Higher Ed*, 15 September. Available at: <https://www.insidehighered.com/news/2014/09/15/interview-professor-fired-west-bank-university-who-compares-himself-steven-salaita> (Accessed: 24 January 2019).
- Liu, Z. et al. (2016) 'Self-disclosure in Chinese micro-blogging: A social exchange theory perspective', *Information & Management*, 53(1), pp. 53–63. doi: [10.1016/j.im.2015.08.006](https://doi.org/10.1016/j.im.2015.08.006).
- Mony, S. (2017) 'Cambodian Netizens Face New Risks as Government Tightens Online Controls', VOA, 11 November. Available at: <https://www.voanews.com/a/cambodian-netizens-new-risks-governmentonline-controls/4111483.html> (Accessed: 24 January 2019).
- Morgan, S. (2017) 'The Cybersecurity Market Report covers the business of cybersecurity, including market sizing and industry forecasts, spending, notable M&A and IPO activity, and more.', *Cybersecurity Ventures*. Available at: <https://cybersecurityventures.com/cybersecurity-market-report/> (Accessed: 24 January 2019).
- Nadi, Y. and Firth, L. (2004) 'The Internet Implication in Expanding Individual Freedom in Authoritarian States', in *ACIS 2004 Proceedings. ACIS 2004* (94).
- Regan, P. M., FitzGerald, G. and Balint, P. (2013) 'Generational views of information privacy?', *Innovation: The European Journal of Social Science Research*, 26(1–2), pp. 81–99. doi: [10.1080/13511610.2013.747650](https://doi.org/10.1080/13511610.2013.747650).
- Sheehan, K. B. (2002) 'Toward a Typology of Internet Users and Online Privacy Concerns', *The Information Society*, 18(1), pp. 21–32. doi: [10.1080/01972240252818207](https://doi.org/10.1080/01972240252818207).
- Sims, J. and Xu, L. (2012) 'Perceived Risk of Online Shopping: Differences Between the UK and China', in *UK Academy for Information Systems Conference Proceedings*.
- Vasquez, I. and Porcnik, T. (2017) *The Human Freedom Index 2017: A Global Measurement of Personal, Civil, and Economic Freedom*. Washington, D.C.: Cato Institute, Fraser Institute, and the Friedrich Naumann Foundation for Freedom.

# Does Time Spent on Device Security and Privacy Inhibit Online Expression?

Juhani Rauhala<sup>1</sup>, Pasi Tyrväinen<sup>1</sup> and Nezer Zaidenberg<sup>2</sup>

<sup>1</sup>University of Jyväskylä, Finland

<sup>2</sup>College of Management Academic Studies, Rishon LeZion, Israel

[juhani.jr.rauhala@jyu.fi](mailto:juhani.jr.rauhala@jyu.fi)

[pasi.tyrvainen@jyu.fi](mailto:pasi.tyrvainen@jyu.fi)

[scipio@scipio.org](mailto:scipio@scipio.org)

**Abstract:** Freedom of expression is a recognized human right. More recently, the UN has resolved that unrestricted access to the Internet is also a human right. A commonly accepted benefit of the Internet is that it serves as a platform for free expression. Usage of the Internet for free expression can be a way of circumventing censorship or other hindrances that prevent citizens' freedom of expression in more traditional publishing media. However, the Internet has unique security and privacy risks that may affect users' attitudes toward expressing themselves online. In the online environment, users with controversial viewpoints may be reluctant to express the viewpoints due to concern about possible consequences resulting from the expressions. Consequences may be imposed by individuals, groups, organizations, businesses, or nation-states. In order to mitigate the security and privacy risks of the Internet, some Internet users spend valuable time thinking about and configuring the security settings of their devices. Some users may have a negative attitude toward such time expenditure. Users may be reluctant to express themselves online simply because it costs too much time and effort for proper configuration for anonymity. That is, the users may be aware of the importance and abundance of tools providing anonymity and may wish to express themselves online but decide spending time on anonymity is just too much effort. The association of time spent on personal cybersecurity with the reluctance to express online does not appear to have been studied in prior research. The purpose of this paper is to explore the effects of these issues on users' reluctance to express themselves online. We constructed a model to represent our hypotheses and collected data using a survey. We then validated the model by performing factor and correlation analyses on the collected data. The results of this research show that the attitude of users toward time expenditure on device security aspects correlates with users' reluctance to express themselves online.

**Keywords:** device security, time consumption, online expression, device settings, frustration

---

## 1. Introduction

Freedom of expression has been declared a universal human right (UN General Assembly, 1948). As of 2016, the United Nations has resolved that unrestricted access to the Internet is also a human right (UN Human Rights Council, 2016). A commonly accepted benefit of the Internet is that it serves as a platform for free expression. However, there are potential consequences for users who make controversial or provocative expressions over the Internet from other users and organizations participating in or following the communication (Baroni, 2015; Cassidy, 2017; Jaschik, 2014).

Users' concerns about such consequences may have an inhibiting effect on their usage of the Internet for free expression. This inhibiting effect may correlate with what users believe and how users behave with respect to addressing security and privacy issues of their devices. The inhibiting effect may also correlate with users' attitude toward and perception of the time they spend addressing their devices' security and privacy issues. However, the association between online expression aspects and the perception of time consumption on security aspects is lacking in prior research. Users may be reluctant to express themselves online simply because it costs too much time and effort for proper configuration for anonymity. That is, the users may be aware of the importance and abundance of tools providing anonymity and may wish to express themselves online but decide spending time on anonymity is just too much effort.

This leads to the main goal of this research; that is, to examine users' reluctance to express themselves in relation to their attitudes and perceptions regarding the time and effort they invest on security, i.e., the time spent on security aspects. This is relevant to participation in social media and other online expression contexts. To achieve the research goal, we establish three latent factors: one corresponding to a reluctance to self-express online, one corresponding to a belief that handling security and privacy aspects of one's device requires an excessive amount ("too much") of one's time, and one for time considering device cybersecurity and privacy settings aspects. We then analyze the correlation among these factors and the correlation between these factors and

related demographic factors. We also perform a linear regression of one latent factor against the others and against a demographic factor.

## **2. Background**

The emerging research of Booth (2017) has raised attention to the issue of freedom of expression and the laws and norms thereof in terms of their relationship to the benefits of ICT on national well-being. At present, Booth has not yet completed her research; moreover, the research will not consider the relationship between the expression of free speech on aspects of the individual user. Internet communication is largely beyond the territorial control of the nation-state and access to the Internet has been recognized as important to the freedom of expression and to participation in a democracy (Lucchi, 2011). Previous research has established that usage of the Internet for free expression can be a way of circumventing censorship or other hindrances that prevent citizens' freedom of expression in more traditional publishing media, especially in authoritarian regimes (Nadi and Firth, 2004).

Debate and discussions that occur over online forums and social media, such as Twitter and Facebook, are raising attention to a virtually unlimited array of topics. Importantly, socially controversial topics and political topics are also discussed. Certain organizations consider and evaluate various threats to the freedom of expression online (Stanton, 2014). In oppressive states, free expression enabled by access to the Internet can be particularly important for advancing human rights (Nadi and Firth, 2004). However, there are potential consequences for users who make controversial or provocative expressions on the Internet, including a negative reaction from the government (Baroni, 2015; Cooper, 2000; Mony, 2017) and offended individuals (Cassidy, 2017), employers (Jaschik, 2014), and schools (Curtom, 2014). Participating in social media is a form of individual expression and there is some research-in-progress on the effects of perceived security threats on user's social media behaviour (Alqubaiti, Li, and He, 2016).

The time that Internet users spend on performing self-protective cybersecurity and privacy-related tasks detracts from the amount of time users have available for other preferred activities. For example, when using open WiFi connectivity in a public space or vehicle, spending time connecting to a secure VPN or updating the security software will leave less time for messaging and for checking social media updates. The excess use of time spent waiting can be merely a perception but may still have negative consequences in terms of user experience or perception of the services for which the waiting is done (Dellaert and Kahn, 1999). Another study has been performed to determine how consumers react when web pages of shopping websites take too much time to load (Anonymous, 2010). It found that 70% of respondents reported that they abandon shopping on a site if the site takes more than 10 seconds to load and 35% said they would not return if the loading delays take "too long." On the other hand, the tolerance of users to the amount of time spent waiting will vary according to the individual and the context (Katz and Martin, 1989). During Internet usage, a loading delay may be experienced with most mouse-clicks or screen taps. However, the need to spend time waiting for a security software update process to complete occurs relatively infrequently, e.g. weekly or monthly.

Excessive non-ideal time consumption, therefore, can be said to detract from more desirable activities and may cause a negative perception of offerings associated with waiting. Frustration with excessive time consumption can result in a negative attitude toward, and possibly abandonment of, desirable online content and activities.

There are also studies observing the impact of demographic factors, such as nationality and age, on Internet behaviour that are relevant to this study. Regan, FitzGerald, and Balint (2013) have evaluated attitudes toward information privacy between age groups (specifically generations). Their analysis revealed a trend where younger generations tend to be more concerned than older ones about wiretapping and data privacy. Chen, Hsu, and Lin (2010) determined that consumers with different levels of computer expertise have different preferences for attributes of shopping websites. Research into culture-based differences in perception of risk for online shopping and other tasks has yielded conflicting results (Sims and Xu, 2012). Sims and Xu (2012) found no significant difference between UK and Chinese shoppers' perceived risk of online shopping despite those shoppers' differing cultural backgrounds. This conclusion was against their expectations and the contradicted results from prior research that showed differences in risk-aversion between the two cultures (Hofstede, 1980).

Controversial expression in an online communications context is affected by other factors. Such factors include perceived anonymity and familiarity with other online community participants (Luarn and Hsieh, 2014). Luarn

and Hsieh studied the expression behaviour of users in a laboratory-controlled virtual community. The virtual community simulated different online group communications environments. They found that users were more willing to express controversial opinions when their identities were anonymous or when they were familiar with other members of the community. When users in the study were not anonymous, they were more reluctant to express such opinions. They also found that there was no effect of anonymity or member familiarity on users' willingness to express non-controversial opinions.

Prior research has shown that negative expressions are received differently than neutral or positive ones. Kwon et al. (2013) studied communications and expressions in a messaging context. They examined the acceptability of negative communications and found that emotional expressions that accompany negative communications were considered much less acceptable than emotional expressions in positive ones. Negative messages by their nature are less welcome.

We expand prior research by investigating the correlation between perception of time consumption used for addressing device cybersecurity and the willingness to freely express on the Internet. Negative expressions (e.g., unpleasant or aggressive) can result in unwanted consequences. Internet users may be reluctant to express themselves because of concerns about such consequences. The time they spend on personal cybersecurity issues may further discourage their controversial expressionism. We hypothesize that users who feel they spend excessive time on their devices' cybersecurity and privacy aspects are more reluctant to freely express themselves online. This is relevant to the users' participation in social media and other online expression contexts.

### **3. Research model**

It is important to consider users' attitudes toward free expression on the Internet and the possible consequences of reluctance to freely express. A key goal of our research is the examination of users' reluctance to express themselves in this context in relation to their attitude and perception regarding time consumption for their devices' security and privacy aspects. Previous research has considered implications on free expression and the benefits of free expression. Willingness to express opinions online has been measured in terms of a web forum's view/reply ratio (Shen and Liang, 2015) and by asking users how likely they would be to express their opinions in specified online scenarios using a 0-100% or 0-10 scale (Ho and McLeod, 2008; Stoycheff, 2016). Hayes et al. (2005) established a self-reporting tool consisting of eight five-point Likert questions to measure willingness to self-censor. However, the tool's questions pertain to a general social context and not specifically to self-expression of controversial opinions on the Internet. Attempts to measure a reluctance to express on the Internet or to establish the same as a latent factor seem to be lacking in previous research. Therefore, our research model defines as a latent factor "reluctance to freely express oneself on the Internet," (RtoEx) for analyzing responses to a set of indicator questions asked in a survey. This factor enables analysis for correlations and the performance of other analyses against other variables or factors.

Time per se is easily quantifiable; however, customers' or Internet users' attitudes or perceptions about the quantity or utility of their time usage are more difficult to define. Prior research has considered the consequences of excessive wait times on customers' attitudes. Prior research has also considered the consequences on Internet users of excessive time spent on the Internet. There seems to be no previous research to consider a user's attitudes and perceptions toward their time usage dedicated to the security and privacy aspects of their Internet device. Our model introduces two latent factors to address this gap: one to measure the belief that addressing device security and privacy aspects negatively impacts one's experience (or takes "too much time") and one to measure whether the user has thought about the security and privacy aspects of his/her device and has checked (and optionally changed) its security and privacy settings. These factors are established by responses to indicator questions that were implemented in a survey. We denote these factors TMT (from "too much time") and TChS (from Thinking about and Changing Settings), respectively. TMT and TChS enable easier analysis for research that seeks to analyze or study the concepts in relation to other variables.

When users contemplate, check, or adjust their device's security and privacy settings out of a sense of obligation instead of preference, the user may experience the corresponding time expenditure negatively. The user may think "this takes too much time" or even "this is a waste of time." We expect that this related cognizance of cybersecurity and privacy risks will be reflected by their attitude toward freely expressing themselves online and on their willingness to do so. We want to see if the resultant frustration from a personal experience-based belief

or perception that cybersecurity threat amelioration requires excessive personal investment of time, may inhibit the willingness to freely express oneself on the Internet.

Conversely, users' reluctance to express online may correlate with their belief of whether addressing security and privacy issues requires an excessive amount of time. These effects may differ across certain demographic groupings, including cultural groupings. We will attempt to explain such differences. It is possible that misgivings in users about the Internet as a platform for free expression may correlate with the belief (or perception) of these same users regarding the need for excessive time to address security and privacy aspects. Such aspects include the contemplation, examination, and adjustment of the relevant device settings.

This paper uses as a general basis the Antecedents -> Privacy Concerns -> Outcomes (APCO) research model defined by Smith, Dinev, and Xu (2011). Variations of the APCO model or models similar to APCO have been applied in other pertinent works in the field, e.g., by Benamati, Ozdemir, and Smith (2017), and by Bandyopadhyay (2009). Our work may be described by way of comparison to Bandyopadhyay's 2009 framework. In Bandyopadhyay's framework there are three consequences, or outcomes, of users' privacy concerns: 1. Refusing to provide personal information, 2. Refusing to enter e-commerce transactions, and 3. Refusing to use the Internet. While Bandyopadhyay's framework has implications for online marketers (Bandyopadhyay, 2009), ours presumes implications for individuals' online expression. In our variation of the framework, we specify one outcome - a reluctance to freely express oneself on the Internet. In place of "privacy concerns" in Bandyopadhyay's proposed framework, we use "usage or perceived excessive usage of time addressing device privacy and security aspects." With regard to the antecedents in Bandyopadhyay's model, we instead propose to use the demographic factors of age, ICT expertise and income as independent variables for a regression analysis between latent factors.

We establish three latent factors: one corresponding to a reluctance to self-express online (RtoEx), one corresponding to a belief that handling security and privacy aspects of one's device requires an excessive amount of one's time (TMT, from "too much time"), and one corresponding to the performance of checking and changing device privacy and security settings (TChS, from "think about and change settings"). We then analyze the correlation among these factors and the correlation between these factors and the related demographic factors. We also plan to examine the effects of demographic factors on the correlation between the two latent factors. The demographic factors include age, income, ICT experience, and nationality.

We have established three factors pertinent to the model:

*Reluctance to Express (RtoEx):* reluctance to freely self-express online

*Too Much Time (TMT):* belief that cybersecurity risk amelioration requires excessive usage of one's time

*Think Change Settings (TChS):* time considering two aspects of one's ICT device – contemplation of the device's cybersecurity aspects and whether time is consumed specifically for the checking and possibly changing of device settings that relate to security and privacy.

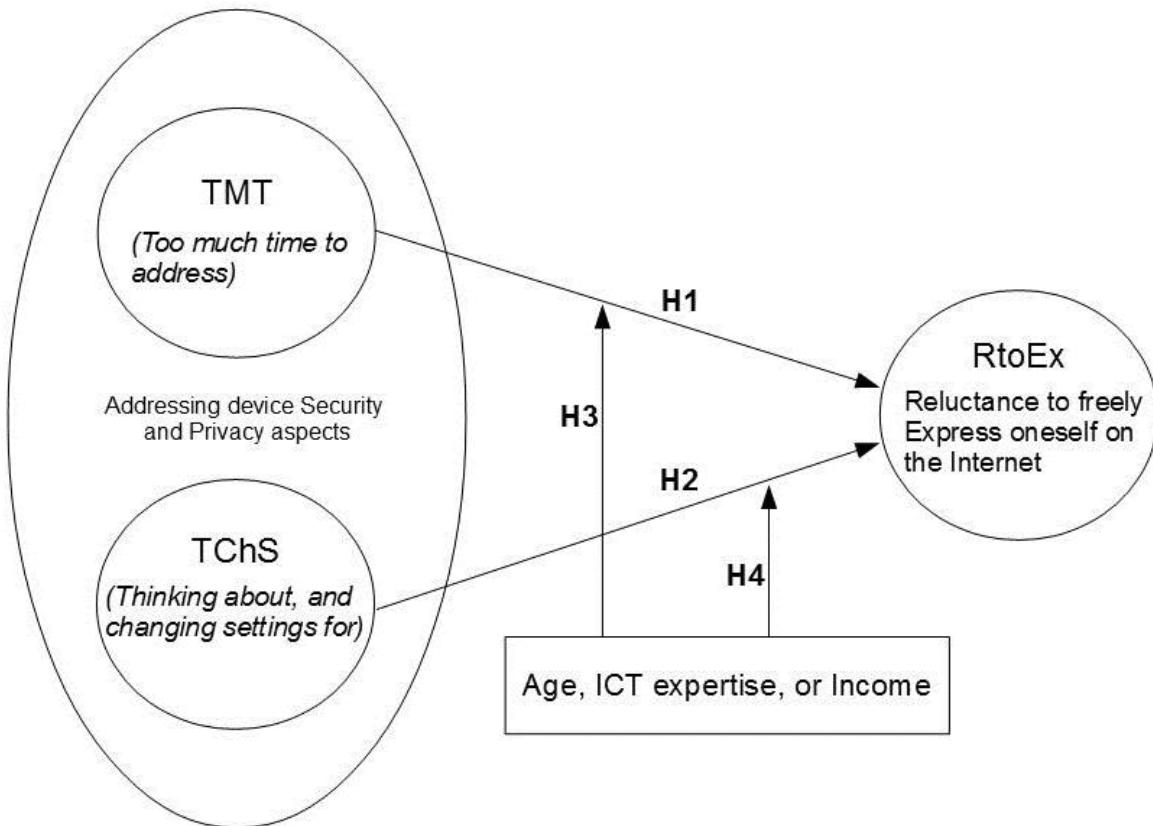
The hypotheses are (Figure 1):

*H1: TMT will positively correlate with RtoEx.*

*H2: TChS will positively correlate with RtoEx.*

*H3: H1 will vary by age, level of ICT expertise, and/or income.*

*H4: H2 will vary by age, level of ICT expertise, and/or income.*



**Figure 1:** Latent variables TMT, TChS, and RtoEx; and the independent variables

### 3.1 Latent factors and their indicators

In this study, we defined sets of indicator questions from which each of the latent variables was derived. The indicator questions were included in the survey, and each consisted of responses along a five-point Likert scale from “strongly agree” to “strongly disagree.” The questions for TMT were as follows: five questions to assess the perception that excessive time has been spent addressing device security and privacy issues, and a belief that time spent on device security and privacy aspects has detracted from time intended for other tasks. TChS is established with three questions to assess whether the user has contemplated and checked (and perhaps adjusted) their device’s security and privacy settings. Cumulatively, we suggest the five “too much time” indicator questions imply that the respondent spends time contemplating and actively addressing security and privacy aspects but tends to feel negatively about doing so (“too much time” implies that the amount of time required is excessive and detracts from activities for which the respondent could preferably be using their time).

The questions for the RtoEx variable were designed as follows: the questions ascertain the attitude of the respondent toward theoretical scenarios of their posting controversial opinions or artwork online, including one question to ascertain their attitude toward using electronic methods vs. face-to-face communication for discussion of a sensitive topic with a friend. We suggest that this set of RtoE indicator questions can convey the level of the respondent’s reluctance to openly communicate using electronic methods including the Internet. For data gathering, a survey was administered over the Web to a population composed mainly of university students and working adults. 197 responses have been obtained, of which 131 are from Finnish nationals.

We use an Exploratory Factor Analysis with direct oblimin rotation to extract latent components from a set of survey questions. The set pertains to TMT, TChS, and RtoEx. The results for TMT and TChS confirm two components. Review of the corresponding survey questions indicates that the TMT and TChS responses are differentiated by the mention of security issues detracting time from preferred tasks, or by a belief that addressing security issues takes too much of one’s time (Appendix, Table 5). Thus, we use three latent factors: RtoEx, TMT, and TChS. We performed a Spearman correlation analysis on the indicator question responses corresponding to RtoEx (eight questions), TMT (five questions), and TChS (three questions). Within all three

groups of latent variables, we found that the responses have a high correlation. For the RtoEx questions (Appendix, Table 4), we found the lowest correlation to be .199, and the highest .673, both two-star significant at the .01 level (two-tailed). For the TMT questions (Appendix, Table 5), the lowest two-tailed correlation was .223 (two-star) and the highest was .752 (two-star). For TChS (Appendix, Table 5), the lowest two-tailed correlation was .314 and the highest .481, both two-star significant. Based on these correlations and on the factor analysis results, we used the means of the responses for each question set. The representative values of latent factors were calculated as averages of responses to the indicator statements. We used SPSS statistical software to calculate Pearson correlations between the three latent variables as well as the Cronbach's alphas (Table 1). The Cronbach's alpha values show an acceptable reliability between the latent variables' indicators.

**Table 1:** Spearman correlations (two-tailed significance at 0.01 level) between indicator question responses for each latent factor; mean correlations; and Cronbach's alpha

Latent Factor	Minimum	Maximum	Mean	Cronbach's Alpha
RtoEx	.199**	.673**	.384	.831
TChS	.314**	.481**	.403	.668
TMT	.223**	.752**	.404	.770

#### 4. Results

The results in Table 2 show that reluctance to express oneself online correlates positively with a long-perceived time spent on setting device security settings (.220\*\*) but not significantly with time spent thinking about device security settings. Thus, we can confirm hypothesis H1 and reject hypothesis H2. Out of the potential moderating variables, we found direct negative correlation between reluctance to express and age (-.225\*\*) but not with other background variables. In a linear regression analysis, the same factors (TMT and age) together reached significant correlation (adjusted R squared = .075, p-value = .000), thus H3 is confirmed for age. Other factors added did not reach significant correlations. Analysis of moderating effects of other demographic variables for H3 will be elaborated in forthcoming research. Table 3 presents the percentage of respondents tending to agree with TMT, TChS, and RtoEx.

**Table 2:** Pearson correlations between RtoEx and TMT, TChS, and age. Two-tailed significances: \* to 0.05 level, \*\* to 0.01 level

n=197	Device security/privacy takes "too much time" (TMT)	Spend time thinking about and changing settings (TChS)	Age (15-25, 26-36, 37-44, 45-54, 55-64, or 65+)
RtoEx	.220**	.077	-.225**

**Table 3:** Percentages of respondents who tend to agree or strongly agree with TMT, TChS, and RtoEx

Overall addressing security and privacy aspects takes too much time (TMT)	Overall spend time thinking about device security and check/change settings (TChS)	Reluctant to express online (RtoEx)
27.4%	59.9%	57.4%

#### 5. Summary, discussion, and conclusion

In this study, our first goal was to establish three latent factors; one corresponding to a reluctance to self-express online, RtoEx; one corresponding to a belief that handling security and privacy aspects of one's device requires an excessive amount of one's time, TMT; and one for time considering device cybersecurity and privacy settings aspects, TChS. Based on the factor analysis of the responses to the indicator statements, this was established.

Our second goal was to analyze the correlation between these factors and the correlation between these factors and the related demographic factors. With respect to the 197 responses from our initial survey, the factor correlations were determined as was the correlation between the reluctance to express and age. We found that RtoEx is positively correlated with TMT. The correlation of RtoEx with age is consistent with Regan, Fitzgerald, and Balint's (2013) findings that older users tend to be less concerned about anonymity and privacy. A linear regression was also performed with the age moderator. We found that older users are less reluctant to express themselves online, and are less likely to consider the time that they use for device security and privacy to be excessive.

Our work has not confirmed a causal relationship between TChS or TMT, and RtoEx. Nonetheless, there are some steps that governments and industry could take to improve Internet users' perceptions of online safety. Nation-states that respect free online expression as a fundamental right for their citizens may choose to create and implement cybersecurity strategies and regulations that improve their citizens' perceptions of the level of online safety. In this way, their citizens may perceive a reduced need to spend time addressing their device settings or their cybersecurity software, and thus an improved opportunity to express themselves online or to perform other preferred tasks. The personal cybersecurity products and services industry could design device security and privacy safeguards to be easier to understand and adjust, and to automate more functions to the background of device or software UIs. Thus, device security and privacy aspects would (ideally) be less time-consuming for consumers to address. However, there may not be clear economic motivations for the cybersecurity industry to modify their consumer products and services in such a manner.

This paper reports results while further analysis is still to be done. We looked into the impact of age on the reluctance to express online, but we have not yet done the full analysis of other moderating demographic factors that can have an impact on it. That is left for further analysis.

In our analysis, we also observe differences between nationalities in the responses. One direction to search for potential explanation is cultural differences (Hofstede, 1980). However, with the current number of responses from non-Finnish respondents, we cannot evaluate this alternative without collecting further data. Other potential sources of explanation include the variation of national social media cultures as well as variations in attitudes and actions of enterprises and public institutions on and to individuals expressing non-controversial opinions online. An extension of this research can be to explore topics about which users are less inclined to express their opinion online.

In addition to the relatively low quantity of responses from non-Finnish nationals, this survey has other limitations. The survey was implemented only in English. English is not the native language of most respondents. Our study also does not examine how, in the case of waiting, the management of time affects the perspective of the person waiting. Examples of such cases could be the users' management of the time spent waiting for a security software update to install; or the content displayed on screen by the software during the update (Hanyang, et al., 2015).

## **Appendix 1**

**Table 4:** Survey questions to indicate level of reluctance to express (RtoEx)

1. I would never post a controversial message in an online forum.
2. If I have a controversial opinion about something, I'm hesitant to publish it on the Internet.
3. I am, or would be, reluctant to display any of my controversial artwork (writing, music, drawings, etc.) online.
4. It's usually not a good idea to post controversial comments or opinions online.
5 I would never post a controversial message in an online forum, because someone or some organization could get revenge against me.
6. I have decided against posting my political opinion on a discussion forum/message board, because I was concerned about consequences to myself or to someone I care about.
7. When discussing something with a good friend, I feel more safe to express controversial opinions face to face, than by electronic communication.
8. I have decided against posting my controversial opinion on a discussion forum, because of concern that someone, or some organization (including government), might use it against me in the future.

**Table 5:** Survey questions to indicate that the user contemplates device security aspects (TChS), and perception or belief that dealing with them requires too much of one's time (TMT)

1. When using my computer or smartphone, I spend time making sure that its security software is up to date. (TChS)
2. When I begin using a new computer or smartphone, I first check its privacy settings, and adjust them to my preference. (TChS)
3. I have had less time to finish a task I wanted to do, due to a device security or software security issue. (TMT)
4. It has taken me longer to finish a task I wanted to do, due to a device security or software security issue. (TMT)
5. The security alerts and pop-up notifications of security software take too much time to deal with. (TMT)
6. I have spent a lot of time thinking about my device and software security. (TChS)
7. I would spend more time performing online tasks I want to do, but my device and software security often needs to be considered. (TMT)
8. Device and software security issues take up much of my time. (TMT)

## References

- Alqubaiti, Z., Li, L. and He, J. (2016) 'The Paradox of Social Media Security: Users' Perceptions versus Behaviors', in *Proceedings of the 5th Annual Conference on Research in Information Technology - RIIT '16. the 5th Annual Conference*, Boston, Massachusetts, USA: ACM Press, pp. 29–34. doi: [10.1145/2978178.2978187](https://doi.org/10.1145/2978178.2978187).
- Anonymous (2010) 'KEEPING ONLINE CUSTOMERS', *Dealerscope*, January, p. 26.
- Baroni, D. (2015) 'New Zealand Government To Punish Online Trolls With Prison Time', *Reaxxion.com*, 3 July. Available at: <http://www.reaxxion.com/10115/new-zealand-government-to-punish-online-trolls-with-prison-time> (Accessed: 24 January 2019).
- Booth, R. E. (2017) 'The Effect of Freedom of Expression and Access to Information on the Relationship between ICTs and the Well-being of Nations.', in *Proceedings of the 23nd Americas Conference on Information Systems*.
- Bandyopadhyay, S. (2009) 'Antecedents And Consequences Of Consumers Online Privacy Concerns', *Journal of Business & Economics Research (JBER)*, 7(3). doi: [10.19030/jber.v7i3.2269](https://doi.org/10.19030/jber.v7i3.2269).
- Benamati, J. H., Ozdemir, Z. D. and Smith, H. J. (2017) 'An empirical test of an Antecedents – Privacy Concerns – Outcomes model', *Journal of Information Science*, 43(5), pp. 583–600. doi: [10.1177/0165551516653590](https://doi.org/10.1177/0165551516653590).
- Cassidy, P. (2017) 'Man petrol bombed homes in revenge for Facebook post', *STV News*, 3 November. Available at: <https://stv.tv/news/east-central/1401461-man-petrol-bombed-houses-in-revenge-for-facebook-post/> (Accessed: 24 January 2019).
- Chen, Y.-H., Hsu, I.-C. and Lin, C.-C. (2010) 'Website attributes that increase consumer purchase intention: A conjoint analysis', *Journal of Business Research*, 63(9–10), pp. 1007–1014. doi: [10.1016/j.jbusres.2009.01.023](https://doi.org/10.1016/j.jbusres.2009.01.023).
- Cooper, A. K. (2000) 'China: Government punishes Internet journalists', *Committee to Protect Journalists*, 12 July. Available at: <https://cpj.org/2000/07/china-government-punishes-internet-journalists.php> (Accessed: 24 January 2019).
- Curtom, G. (2014) 'Students punished for expressing free speech on Twitter', *The Cougar*, 24 April. Available at: <http://thedailycougar.com/2014/04/24/students-punished-expressing-free-speech-twitter/> (Accessed: 24 January 2019).
- Dellaert, B. G. C. and Kahn, B. E. (1999) 'How tolerable is delay?: Consumers' evaluations of internet web sites after waiting', *Journal of Interactive Marketing*, 13(1), pp. 41–54. doi: [10.1002/\(SICI\)1520-6653\(199924\)13:1<41::AID-DIR4>3.0.CO;2-S](https://doi.org/10.1002/(SICI)1520-6653(199924)13:1<41::AID-DIR4>3.0.CO;2-S).
- Hayes, A. F. (2005) 'Willingness to Self-Censor: A Construct and Measurement Tool for Public Opinion Research', *International Journal of Public Opinion Research*, 17(3), pp. 298–323. doi: [10.1093/ijpor/edh073](https://doi.org/10.1093/ijpor/edh073).
- Ho, S. S. and McLeod, D. M. (2008) 'Social-Psychological Influences on Opinion Expression in Face-to-Face and Computer-Mediated Communication', *Communication Research*, 35(2), pp. 190–207. doi: [10.1177/0093650207313159](https://doi.org/10.1177/0093650207313159).
- Hofstede, G. (1980) *Culture's Consequences: International Differences in Work-Related Values*. 1st edn. Beverly Hills: Sage Publications.
- Jaschik, S. (2014) 'Interview with professor fired by West Bank university who compares himself to Steven Salaita', *Inside Higher Ed*, 15 September. Available at: <https://www.insidehighered.com/news/2014/09/15/interview-professor-fired-west-bank-university-who-compares-himself-steven-salaita> (Accessed: 24 January 2019).
- Katz, K. L. and Martin, B. R. (1989) *Improving customer satisfaction through the management of perceptions of waiting*. Massachusetts Institute of Technology. Available at: <http://hdl.handle.net/1721.1/37703> (Accessed: 24 January 2019).
- Kwon, O., Kim, C. and Kim, G. (2013) 'Factors affecting the intensity of emotional expressions in mobile communications', *Online Information Review*. Edited by K. Chang Lee, 37(1), pp. 114–131. doi: [10.1108/14684521311311667](https://doi.org/10.1108/14684521311311667).
- Luarn, P. and Hsieh, A.-Y. (2014) 'Speech or silence: The effect of user anonymity and member familiarity on the willingness to express opinions in virtual communities', *Online Information Review*, 38(7), pp. 881–895. doi: [10.1108/OIR-03-2014-0076](https://doi.org/10.1108/OIR-03-2014-0076).

- Lucchi, N. (2011) 'Access to Network Services and Protection of Constitutional Rights: Recognizing the Essential Role of Internet Access for the Freedom of Expression', *ARDOZO JOURNAL OF INTERNATIONAL AND COMPARATIVE LAW*, 19(3), pp. 645–678.
- Mony, S. (2017) 'Cambodian Netizens Face New Risks as Government Tightens Online Controls', *VOA*, 11 November. Available at: <https://www.voanews.com/a/cambodian-netizens-new-risks-governmentonline-controls/4111483.html> (Accessed: 24 January 2019).
- Nadi, Y. and Firth, L. (2004) 'The Internet Implication in Expanding Individual Freedom in Authoritarian States', in *ACIS 2004 Proceedings. ACIS 2004* (94).
- Regan, P. M., FitzGerald, G. and Balint, P. (2013) 'Generational views of information privacy?', *Innovation: The European Journal of Social Science Research*, 26(1–2), pp. 81–99. doi: [10.1080/13511610.2013.747650](https://doi.org/10.1080/13511610.2013.747650).
- Shen, F. and Liang, H. (2015) 'Cultural Difference, Social Values, or Political Systems? Predicting Willingness to Engage in Online Political Discussion in 75 Societies', *International Journal of Public Opinion Research*, 27(1), pp. 111–124. doi: [10.1093/ijpor/edu012](https://doi.org/10.1093/ijpor/edu012).
- Sims, J. and Xu, L. (2012) 'Perceived Risk of Online Shopping: Differences Between the UK and China', in *UK Academy for Information Systems Conference Proceedings*.
- Smith, Dinev and Xu (2011) 'Information Privacy Research: An Interdisciplinary Review', *MIS Quarterly*, 35(4), p. 989. doi: [10.2307/41409970](https://doi.org/10.2307/41409970).
- Stanton, L. (2014) 'EFFECT OF "RIGHT TO BE FORGOTTEN" ON FREE EXPRESSION SPARKS DEBATE', *Cybersecurity Policy Report*, 18 August.
- Stoycheff, E. (2016) 'Under Surveillance: Examining Facebook's Spiral of Silence Effects in the Wake of NSA Internet Monitoring', *Journalism & Mass Communication Quarterly*, 93(2), pp. 296–311. doi: [10.1177/1077699016630255](https://doi.org/10.1177/1077699016630255).
- UN General Assembly (1948) *Universal Declaration of Human Rights*. Paris (217 A). Available at: <https://www.un.org/en/universal-declaration-human-rights/index.html> (Accessed: 24 January 2019).
- UN Human Rights Council (2016) *Resolution on the promotion, protection and enjoyment of human rights on the Internet*. Geneva (A /HRC/ 3 2 /L . 20). Available at: [https://www.article19.org/data/files/Internet\\_Statement\\_Adopted.pdf](https://www.article19.org/data/files/Internet_Statement_Adopted.pdf) (Accessed: 24 January 2019).

# The Need for More Sophisticated Cyber-Physical Systems war Gaming Exercises

Aunshul Rege<sup>1</sup> and Joe Adams<sup>2</sup>

<sup>1</sup>Temple University, Philadelphia, USA

<sup>2</sup>Merit Network, Ann Arbor, USA

[rege@temple.edu](mailto:rege@temple.edu)

[wjadams@merit.edu](mailto:wjadams@merit.edu)

**Abstract:** Cyber-physical systems (CPS) are highly integrated into critical infrastructures. These systems execute automated control of physical equipment in transportation networks, nuclear plants, water and gas distribution networks, and power plants. CPSs offer a unique cybersecurity challenge as cyberattacks against CPSs adversely affect public services (e.g., WannaCry attacks in Europe), research facilities (e.g., STUXNET), or transportation services (e.g., OnionDog's attack on South Korea). It is critical to train and educate operators, owners, and users of CPSs on how these systems are subjected to cyberattacks; how to defend CPSs in real time; how to manage limited employee and monetary resources during and after cyberattacks; and how to better manage system confidentiality, integrity, and availability. Real-time CPS cybersecurity exercises serve as ideal training platforms. This paper reviews existing CPS cybersecurity red team-blue-team exercises (RTBTEs) conducted in USA. This paper highlights the many benefits of these exercises, such as understanding real-time attacks and defense, testing and validating security models, and also understanding human behavior of both attackers and defenders. While these are important contributions, they focus on a small subset of CPSs inside particular infrastructures within condensed temporal frameworks. Collectively, these factors approximate the reality of CPS cyberattacks, which take longer, are more sophisticated, and target multiple, connected infrastructures. This paper thus argues for a more sophisticated CPS wargame hosted in an environment more representative of reality. An advanced training environment is being constructed at Camp Grayling Michigan in collaboration with public industry and government agencies.

**Keywords:** cyber-physical systems, war gaming, cybersecurity exercises, decision-making and adaptation, group dynamics

## 1. Introduction

Cyber-physical systems (CPS) are highly integrated into critical infrastructures. These systems execute automated control of physical equipment in transportation networks, nuclear plants, water and gas distribution networks, and power plants. Originally, CPS were segregated (air-gapped) from corporate networks and the internet. Over time, however, organizations connected these systems to outside networks to cut costs, share information, or distribute usage and billing data. This air-gap removal made CPS accessible to cybercriminals. Furthermore, many of these systems run on old platforms that cannot be patched, and cybercriminals leverage the well-known vulnerabilities of their platforms when targeting critical infrastructures. As such, CPSs offer a unique cybersecurity challenge. Attacking these systems could have devastating impacts on a country's functionality, security, and stability. In the past decade alone, there have been twenty publicly documented critical infrastructure cyberattacks targeting utility, military, water and wastewater, and critical manufacturing sectors (see Table 1).

It is thus critical to train and educate current and future operators, owners, and users of CPSs on how these systems are subjected to cyberattacks; how to defend CPSs in real time; how to manage limited employee and monetary resources during and after cyberattacks; and how to better manage system confidentiality, integrity, and availability (CIA). Real-time CPS cybersecurity exercises serve as ideal training platforms.

**Table 1:** List of publicly documented critical infrastructure cyberattacks 2009-2018

	Year	Name	Type	Target	Country
1	2009	Conficker	worm	French Navy	France
2	2010	Stuxnet/Duqu/Flame/Gauss	worm	Natanz nuclear facility	Iran
3	2011	Operation Night Dragon	remote access trojan (RAT)	Exxon, Shell, BP	USA
4	2011	DUQU	malware, virus	Kaspersky Labs	Western countries, Middle East, Asia
5	2012	Flame	malware, worm	oil ministry, national oil co	Iran

	Year	Name	Type	Target	Country
6	2014	Havex	botnet, trojan	multiple targets (undisclosed)	France and Germany
7	2014	BlackEnergy	botnet, trojan	General Electric (GE)	USA
8	2014	SIEMENS Trojan	software update	Multiple ICS sites	Worldwide
9	2014	OnionDog	malware, USB worm	Energy and Water utilities	Korea
10	2015	OnionCity	botnet, trojan	port harbors, Vessel Traffic Systems, transportation systems	Korea
11	2015	Unknown APT group	malware, spear phishing	German Steel Mill	Germany
12	2015	BlackEnergy3	malware, spear phishing	Power Grid	Ukraine
13	2016	Syrian Hacktivists Group	spear phishing	Kemuri Water Company	USA
14	2016	Conficker	worm	KGG Nuclear Power Facility	Germany
15	2017	Crashoverride	worm	Ukrainian Power Grid	Ukraine
16	2017	WannaCry	ransomware	Dacia car manufacturer	Romania
17	2017	WannaCry	ransomware	Nissan car manufacturer	UK
18	2017	NotPetya	ransomware	multiple targets	Ukraine & Germany
19	2018	Emotet	cryptocurrency attack	wastewater site	USA
20	2018	Triton	worm	engineering workstation	Undisclosed

This paper reviews existing CPS cybersecurity red team-blue-team exercises (RTBTEs) conducted in the USA by the Army Research Lab; United States Computer Emergency Readiness Team-Idaho National Laboratory (US-CERT/INL), the Department of Homeland Security (DHS), and Merit Network's Michigan Cyber Range. This paper highlights the many benefits of these exercises, such as understanding real-time attacks and defense, testing and validating security models, and also understanding human behavior of both attackers and defenders. While these are important contributions, they focus on a small subset of CPSs inside particular infrastructures within condensed temporal frameworks. Collectively, these factors approximate the reality of CPS cyberattacks, which take longer, are more sophisticated, and target multiple, connected infrastructures.

The paper is structured as follows. The next section offers an overview of the various RTBTEs in the United States. The third section identifies the many research benefits of RTBTEs that move beyond education and training; it discusses how RTBTEs can be used to study intrusion chains, adversarial behavior, decision-making, and adaptation, and group dynamics. Next, the limitations of these RTBTEs are highlighted, such as temporal and infrastructure constraints and the lack of data triangulation. Finally, an advanced training environment, called Griffinville, that is being constructed at Camp Grayling Michigan is discussed, which tries to address some of these limitations. The paper concludes with discussion points that should be considered for future RTBTE design and implementation.

## **2. Existing critical infrastructure cybersecurity exercises in USA**

Currently, there are four cybersecurity exercises that are conducted in the United States. The structure of each is discussed next.

### **2.1 US Army Research Lab**

The US Army Research Lab conducts a two tier exercise focused on protecting SCADA and ICS networks of a notional SCADA system called AQUA, which was a fictional food-processing plant that involved Programmable Logic Controllers (PLCs), a closed circuit television system (CCTV), and a Human-Machine Interface. The first is a table top game that tests process and communication between participants. The goal of the exercise is to general discussion and identify gaps in security measures. The second tier used a wargame environment with red and blue teams that attacked and defended AQUA respectively (Colbert et al 2017).

## **2.2 ICS-CERT/INL**

Structured as a five-day event, this hands-on educational program offers understanding, protecting, and securing control systems from cyberattacks. The program also includes a RTBTE, which is conducted within a CPS environment. The purpose of this training program is to increase cybersecurity awareness; understand of the nature of threats; and understand the challenges of conducting or responding to an attack (Bralant et al 2011).

## **2.3 US Cyber Storm**

Cyber Storm is a biennial exercise sponsored by the US DHS that focuses on strategic decision making and interagency cooperation in response to threats on critical national infrastructure. Currently in its 13th year, the exercise participants work through scenarios that are centered on information sharing and collaborative response. There is not a “live” wargaming component, but the exercise uses elements such as social media and law enforcement to add realism to the three-day “live fire” session (DHS, 2019). Organizations participate from their home location. An exercise control cell in Washington DC choreographs the participants through exercise injects through e-mail, phone, in person, and via exercise web sites.

## **2.4 Alphaville**

Merit Network hosts the Michigan Cyber Range, a training and education platform that conducts cyber exercises as part of its cybersecurity training curriculum. These exercises are conducted in a simulated city called Alphaville. In addition to a library, a school, and a city hall, Alphaville includes an engineering-manufacturing facility for Industrial Control System training and a power distribution utility for SCADA exercises (Merit Network, 2019).

An example of Alphaville’s use is the International Cyber Exercise (ICE), where National Guard units from around the United States participate along with their partner nations in the State Partnership Program to participate in a free flow cyber exercise called “paintball,” described below, at the North American International Cyber Summit (NAICS) (National Guard Bureau, 2019).

During the one-day exercise, teams of five participants from each organization are placed inside Alphaville. Their objective is to infiltrate and control network assets (i.e., devices and clients,) then secure them against other groups. The number team controlling the most assets at the end of the exercise is declared the winner. The ICE forces teams to make quick decisions, delegate tasks, and communicate amongst each other to be successful.

## **3. Benefits of existing cybersecurity exercises**

In addition to serving an educational purpose, cybersecurity exercises offer numerous other benefits in understanding human behavior, group dynamics, decision-making and adaptability, and intrusion chains. This section identifies how two of the exercises discussed above (ICS-CERT/INL and Merit’s ICE) served as a rich resource to research these various components.

### **3.1 Data sources and analysis**

The first dataset is from observations of the red team during ICS-CERT/INL’s RTBTE held in September/October 2014. The red team was randomly created and consisted of ten members (referred to as S1 to S10 in this paper) who had not previously met one another. Team members had an assortment of jobs, such as system administrators, control systems engineers, and information technology specialists. Days three and four occurred in an enclosed room. The red team had to complete a set of predetermined tasks that varied in difficulty; each task was assigned points proportional to the level of its difficulty. Composition of the blue team was unknown to the red team (and to the authors) (Rege et al 2017 b).

During the 2015 NAICS, researchers observed a force on force exercise colloquially called “paintball”, which served as the second dataset. In this type of exercise, teams of five participants battle to claim Alphaville’s network by controlling critical servers and firewalls. A team marks its control of an asset by planting an encrypted beacon and defending that asset against other teams. The 2015 NAICS featured six teams operating in the same environment. Teams were spread across the globe, with seven time zones between some participants. Their progress was displayed on a scoreboard, which was visible to all teams. Dots on the scoreboard represented the assets in the different locations, with the dot’s color signifying which team had control of it in that 5 second

segment. The team we observed consisted of four members (henceforward referred to as Subjects S1, S2, S3 and S4). Two members had worked with each other before, which facilitated team organization and division of labor. The exercise lasted for 5 hours, during which the team attempted to control various parts of Alphaville's infrastructure as described above. Composition of the other teams was unknown to the team observed (and to the researchers) (Rege et al 2017 a).

The observed data were analyzed by focusing, simplifying, and transforming the written-up field notes into visual representations (tables and charts), which facilitated reviewing large amounts of data efficiently (DeWalt & DeWalt 2010). Doing so allowed making comparisons, summarizing patterns, drawing conclusions, and presenting effective arguments. Finally, to achieve interpretation and verification, interviews with the red team members were conducted where possible and compared to the exercise debriefing that occurred on day five. This approach was the best means of ensuring that observed data matched what the participants had experienced during the exercise.

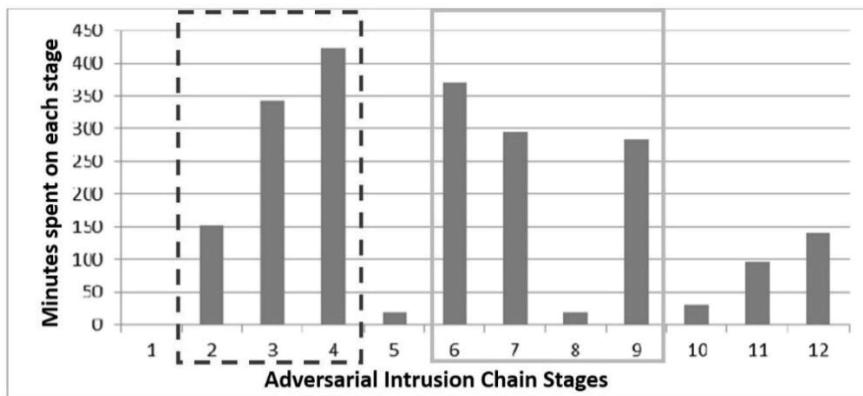
### **3.2 Intrusion chains**

Intrusion chains detail the step-by-step process of unfolding cyberattacks. There are many intrusion chain models, but essentially, they all are comprised of twelve stages, as illustrated in Figure 1. Adversaries first identify their targets. They then form alliances with other adversaries that complement and supplement their own skill sets. In the third stage, adversaries design and build their attack vectors essential to execute cyberattacks. Fourth, adversaries conduct reconnaissance to obtain target information, such as infrastructure blueprints and vulnerabilities. Fifth, adversaries conduct defender reconnaissance, where they gather information on security protocols established by the target. Doing so enables adversaries to create appropriate evasion and response plans. In the sixth stage, adversaries deploy their attack vectors to gain a foothold inside the target environment. In the seventh stage, adversaries gain preliminary access to the targeted environment and install malware. Adversaries establish further access points within the targeted environment in the eighth stage, while obtaining credentials to gain greater system access that will increase their control in the ninth stage. Adversaries then strengthen their presence by moving laterally and deeper inside the targeted environment. This pivoting movement allows adversaries to establish control over as many different parts of the system as possible. In the eleventh stage, adversaries accomplish their objectives, such as exfiltrating data or disrupting functionality. Finally, adversaries remove evidence of their presence and actions in the targeted environment.

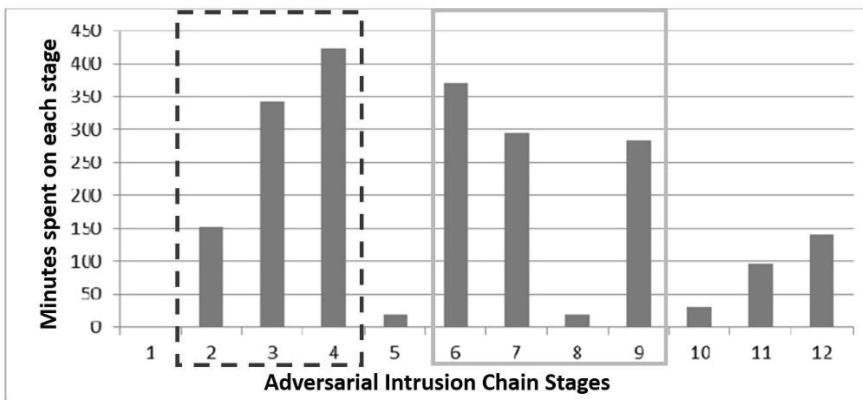


**Figure 1:** Cyber-adversarial intrusion chain model (Clopgett, 2009)

Intrusion chain analysis was done for both the ICS-CERT/INL and Merit ICE data, as shown in Figures 2 and 3 below. Both these diagrams suggest that the adversarial teams spent a good portion of their time on the reconnaissance stages (stages 2, 3, and 4) of the intrusion chain. Thus, these exercises allowed for understanding of the time spend on various intrusion chain stages and suggested that conducting reconnaissance is very relevant to the overall cyberattack process.



**Figure 2:** Time spent on intrusion chain stages in ICS-CERT/INL RTBTE (Rege et al. 2017 b)



**Figure 3:** Time spent on intrusion chain stages in NAICS RTBTE (Rege et al. 2017 a)

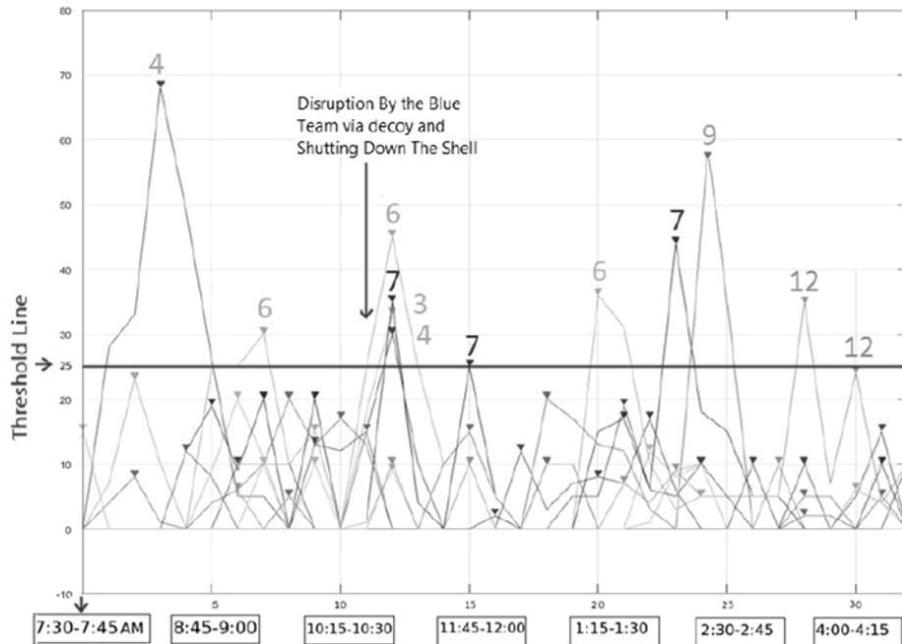
### 3.3 Disrupting adversarial behavior and adaptability

Previous research has found that adversarial behavior can be disrupted in two main ways. The first is that they are blocked by defenders through infected system isolation, which limits adversarial access and their ability to use the infected system to pivot and move laterally or deeper inside the target system. Consider the NAICS event, where subjects S1 and S2 repeatedly had their access to systems removed. Unfortunately, they were unable to regain their foothold (Rege et al. 2017 a). In the INL event, two team members were able to regain access and then changed passwords to lock the defenders out of their own systems (Rege et al. 2017 b). Defenders can also use decoys to confuse or mislead the adversary, which might increase the effort required, waste their time (which is time taken away from the actual attack), or send them down incorrect paths. This tactic was used in the INL RTBTE, where S4 was misled by a blue team decoy and realized much later that he had expended his effort in vain (Rege et al. 2017 a). Figure 4 showcases two disruptions encountered by the red team that were caused by the blue team.

The second disruption is adversary-based. Here, adversaries make mistakes, which hinders their progress in the intrusion chain. For instance, across both RTBTEs, the red team disrupted its own intrusion chains, by accidentally disconnecting from servers and machines they had previously gained access to (Rege et al. 2017a,b). Adversaries may be unable to proceed in their intrusion chains if they do not have sufficient knowledge about the targeted environment or have limited skills. This was also evident across both RTBTEs, where the less experienced members would often look to skilled teammates for guidance (Rege et al. 2017a, b).

Interestingly, these qualitative data can be made ‘granular’ and then subjected to data analytic techniques to further understand which intrusion stages adversaries spend more time on after experiencing disruptions. The time series analysis in Figure 4 offers interesting insights into adaptations after disruptions. For instance, at 10am, the red team experienced two disruptions, a blue team decoy, and access shutdown. Immediately after these disruptions, there were spikes in stages 3, 4, 6, and 7. The spikes in stages 3 and 4 may have been due to the access shutdown as the red team had to spend more time on researching the infrastructure again (stage 4) to find new access points and find corresponding tools (stage 3) to obtain access. The spikes in stages 6 and 7

could be due to the blue team decoy. The red team may have been distracted and thus had to get back on track by deploying the attack vector (stage 6) and conducting initial intrusion (stage 7). There is some correlation between disruptions and how adversaries adapt by focusing on certain stages.



**Figure 4:** Time series analysis of the ICS-CERT/INL exercise (Rege et al 2017c). Intrusion chain stages: 1. Define Target; 2. Find & Organize Accomplices; 3. Build/Acquire Tools; 4. Research Target Infrastructure/Employees; 5. Test for Detection; 6. Deployment; 7. Initial Intrusion; 8. Outbound Connection Initiated; 9. Expand Access & Obtain Credentials; 10. Strengthen Foothold; 11. Exfiltrate Data; 12. Cover Tracks & Remain Undetected (Clopert, 2009)

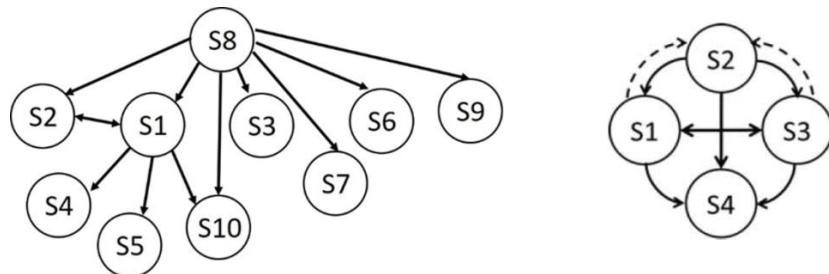
### 3.4 Group dynamics and divisions of labor

Existing research looking at criminal organization and operation has typically focused on structure and roles/divisions of labor. For instance, Lemieux (2003) identifies nine roles inherent to the structure of crime networks. At the core are the *organizers*, who determine the nature and scope of activities, and also offer the guidance and impetus necessary to carry out these activities. The *executors* are responsible for carrying out the objectives of the organizers and they possess the specialized skills necessary to successfully carry out the operation. Then, *insulators* work to protect the internal structure, the core, from being exposed and the *communicators* handle the effective flow of information between the various sub-networks. Given the sometimes conflicting nature of insulators and communicators, an individual may sometimes simultaneously take on both roles. *Guardians* are the protectors of the criminal network and they take the measures necessary to prevent external attacks. Unlike insulators who protect the criminal core, guardians protect the entire network from exterior threats. *Monitors* deal with the effectiveness of the network and ensure that the network can adjust to different circumstances and aid in developing techniques to circumvent law enforcement.

Research on cybercrime groups is limited to organizational dynamics, such as structure and divisions of labor (Wagen & Pieters, 2015). However, there are many other aspects of group dynamics that remain understudied because of the difficulty in observing real-time interactions between group members given their dispersed, covert, and dynamic nature. Some of these group aspects include member dynamics, power and status dynamics, conflicts and tensions, cohesiveness, and interdependencies.

Groups exhibit some form of interdependence, which is a mutual dependence or influence, where one member's outcomes, actions, thoughts, and experiences are determined in whole/part by other group members (Forsyth, 2006). Unilateral interdependencies occur when one member influences all others, sequential interdependencies occur when one member influences the next member, who then influences the next, and reciprocal interdependencies occur when members influence each other (Forsyth, 2006). RTBTEs are great

platforms for studying these issues as they occur in real-time and group members are present/available for observation (Rege et al. 2017a, b). Consider Figure 5 below.



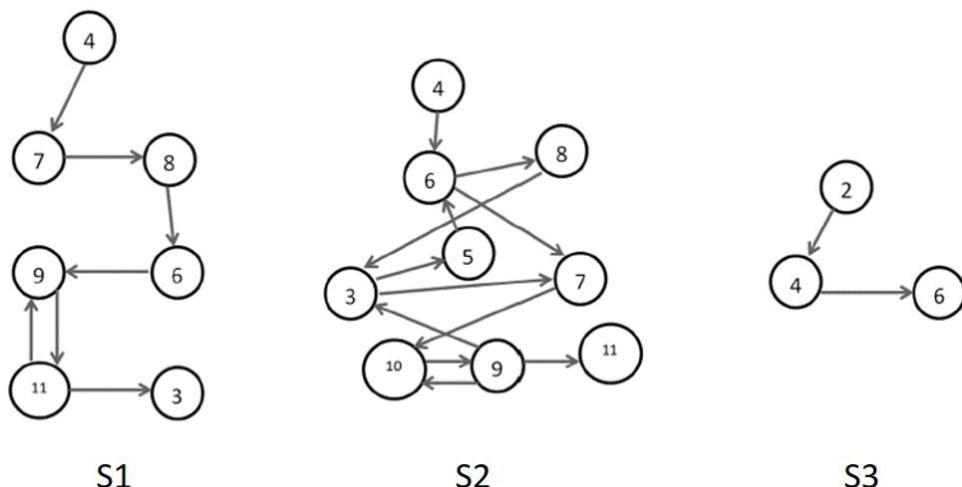
**Figure 5:** Group dynamics in the INL exercise (left) and NAICS exercise (right). (Rege et al. 2017 a, b)

For the INL event, team member S8 emerged as the leader, despite lacking technical expertise. He requested status updates from different subgroups, organized sub-teams based on skills and familiarity, and encouraged communication among members to ensure that the group functioned effectively. The rest of the team worked in sub-teams and had full autonomy on its tasks. Team members also displayed transient sub-team membership as they often shifted to different sub-teams when they completed their tasks or if their expertise was needed elsewhere (Rege et al. 2017b).

In the NAICS event, subjects S1, S2, S3 and S4 all played different roles and worked on a variety of tasks. S2 consistently exhibited the greatest skill sets, knowledge, and leadership traits as he delegated tasks to S1, S3, and S4. Even though S1 and S3 were less skilled in comparison to S2, they occasionally gave out both help and smaller task delegations to each other. S4 was the most inexperienced, lacked skills the other subjects had, received assignments from the other subjects, and required more help than any other member. In this context, it appears that the interdependencies between team members were of a very informal nature with a combination of unilateral and reciprocal, as illustrated in Figure 5 (Rege et al. 2017b). Both RTBTEs suggest a networked based structure. S2 and S8 did display the characteristics of organizers and monitors, by setting the overall objectives and assessing the progress of the team. All team members were executors, insulators, and communicators, as each member was involved in some aspect of the cyberattack and ensured that information was relayed efficiently as the exercise progressed.

### 3.5 Adversarial movement across intrusion chains

RTBTEs can also shed light on how adversaries might move across the intrusion chain stages. Figure 6 shows how some members of the red team navigated across the various intrusion chain stages during the INL exercise (Asadi et al 2018).



**Figure 6:** Movement across the intrusion chain stages for different red team members (Asadi et al 2018)

Figure 6 offers interesting insights into how adversaries might move across intrusion chains during cyberattacks. First, not all adversaries navigate through all intrusion chain stages. For instance, S1 does not go through stages

2 (Find and organize accomplices), 5 (Test for detection), and 10 (Strengthen foothold). Second, subjects do not move linearly across the intrusion chain stages. Consider again S1, who moves from stage 4 to 7 (bypassing stages 5 and 6) to 8 and then to 6. This non-sequential movement suggests that (i) intrusion chain stages do not occur in a linear manner, (ii) certain stages might be irrelevant to progression along the intrusion chain, or (iii) S1 only work on those stages for which he possesses skills and knowledge. Third, adversarial movement itself may be iterative. Consider S2's movement, which not only shows nonlinear movement (for example, stage 4 to 6 and then to 8), but also bi-directionality (stages 9 and 10).

#### **4. Limitations of existing cybersecurity exercises**

The above discussion showcases how RTBTEs can go well beyond their purpose of training and education, and can be leveraged to shed light on many aspects of adversarial behavior, movement, decision-making and adaptation, and group dynamics. Of course, there are many limitations of these RTBTEs, which impact the generalizability of findings.

##### **4.1 Temporal restrictions**

Most RTBTEs are time constrained as their primary purpose is to serve as a training exercise. They typically last for a day, and range from four to ten hours in duration. Participants specifically engage in the exercise and the debriefing proceedings. This temporal restriction is an obvious limitation as it is not always representative of reality. Some cyberattacks do indeed occur in these short time frames, but many sophisticated cyberattacks conducted by advanced persistent threats (APTs) occur over extended time periods (days, months, or even years). Thus, RTBTEs may not capture the movements, behaviors, and adaptations that occur over longer time periods.

##### **4.2 Infrastructure restrictions**

Many RTBTEs are also structured around specific designs and infrastructures. For instance, the INL exercise had a cyber-physical system that the red team had to target, while in the NAICS event, teams had to target and control various virtual organizations in Alphaville. This variation in RTBTEs is to be expected because each has a different focus, purpose, and structure. In reality, targeted infrastructure is a combination of cyber, physical, and cyber-physical systems. Furthermore, the infrastructure may be interconnected and cyberattacks may result in cascading impacts. Representing these realities are often outside the scope of exercises, which are geared towards education and training.

##### **4.3 Data triangulation**

RTBTEs are excellent data sources. As noted above, observations, interviews, and focus groups were conducted with participants at two exercises. There are several other data that can also be collected. For instance, technical logs from the exercise, which capture participant actions and performance, could be triangulated with qualitative data to get a more complete picture of behavior and decision-making. In fact, the qualitative data and technical logs could complement and supplement each other by filling in any gaps (data not captured by either dataset), validating (do the two data points offer a similar representation), and identifying any contradictions (do the two data points differ at certain, key points).

#### **5. Enhancing exercise realism**

Cyber-physical exercises are showing the same maturation process that both physical and cyber security exercises have been through. The goal is to make the lessons provided by the exercise more insightful, realistic, and relevant to the issue the exercise is addressing (Perla & McGrady, 2011). The exercises and training environment we have discussed in the previous sections all require a certain amount of "suspension of belief" from the participant. While recognizing the place of seminars and tabletop exercises, the state of the art in cyber-physical exercises has progressed to the point that participants have to be exposed to sensory experiences, technical challenges, and emotional pressure in order to reap the full training benefit. Griffinville is a major step toward implementing the kind of immersive environment that will validate the types of interactivity that have been observed in virtual exercise environment and address the limitations of existing exercises noted in Section 4. Using the physical construct of Camp Grayling's CACTF, the limitation of time can be eliminated because the facility is available at all times of day throughout the year. Infrastructure restrictions are dealt with by installing types of equipment (e.g., IoT devices, PLCs, traffic lights) in the facility that present the challenges the participants would find in a real mission set. Of all the limitations, data triangulation would be addressed by

tapping the CACTF's network, CCTV, and observer logs. Camp Grayling is a National Guard training facility located in Grayling, Michigan. It is the largest National Guard training area in the United States and includes a Combined Arms Combat Training Facility (CACTF) among its many training venues (PBS, 2019). The CACTF's physical infrastructure includes realistic buildings, roads, and utility infrastructure. This physical infrastructure offers an opportunity to combine physical training tasks with cyber security activity. Merit Network has partnered with the Michigan National Guard to add cyber security capabilities to the training at Camp Grayling's CACTF. Because connectivity to the CACTF is provided through Merit, cyber teams around the world share access to training experiences with each other and the units operating within the CACTF. Finally, both physical and cyber components are combined to contribute to a common operational environment to inform cyber operators, non-cyber planners, and commanders on the efficacy and potential of cyber activity. This environment consists of two parts. The first is a working virtual representation of the physical environment. Virtual machines and networks simulate building control systems, public utilities, and local area networks. These are linked to a visual representation of the city built in a commercial-grade gaming engine. This portrayal provides visual feedback to planning and operations staff on the effects of the cyber operation. The second component integrates cyber-physical systems, such as Internet of Things devices, automobiles, and civic utility infrastructure, by networking devices such that they communicate with the virtual Griffinville. This component will grow from a test lab set of devices to an on-site "Smart House" and IoT devices interacting with the live environment of Griffinville. Combining these characteristics, Griffinville will provide training that is not available elsewhere in the United States. The flexible physical space could facilitate multiple training scenarios and the ability to expand and change setting to accommodate future mission sets. Bringing all of these benefits together in a shared environment provides not only a bridge between the physical and virtual environments but also unites planners and commanders with their cyber operators in a common, shared picture of the engagement area.

## **Acknowledgements**

This material is based upon work supported by the National Science Foundation CAREER Award, Grant No. CNS-1453040.

## **References**

- Asadi, N., Rege, A. & Obradovic, Z. (2018). "Analysis of Adversarial Movement Through Characteristics of Graph Topological Ordering". Proceedings of the International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA). Institute of Electrical and Electronics Engineers (IEEE) Xplore Digital Library.
- Cloppert, M. (2009). "Security Intelligence: Attacking the Cyber Kill Chain". Retrieved February 2, 2014. Online at <http://digital-forensics.sans.org/blog/2009/10/14/security-intelligence-attacking-the-kill-chain>
- Colbert, Sullivan, & Kott (2017). Cyber-Physical War Gaming. *Journal of Information Warfare*, 16 (3).
- DHS (Department of Homeland Security) (2017). "Cyber Storm: Securing Cyber Space". Retrieved January 21, 2019. Online at <https://www.dhs.gov/cyber-storm>
- DeWalt, K. & DeWalt, B. (2010). Participant observation: a guide for fieldworkers, Rowman Altamira, Plymouth, U.K.
- Lemieux, V. (2003). Criminal Networks. Retrieved January 14, 2006, from RCMP. Online at [http://www.rcmp.ca/ccaps/reports/criminal\\_net\\_e.pdf](http://www.rcmp.ca/ccaps/reports/criminal_net_e.pdf)
- Merit Network, Inc. (2019). "Alphaville & Griffinville". Retrieved January, 21 2019. Online at <https://www.merit.edu/cyberrange/alphaville/>
- National Guard Bureau. (2019). "State Partnership Program". Retrieved January 21, 2019. Online at <https://www.nationalguard.mil/leadership/joint-staff/j-5/international-affairs-division/state-partnership-program/>
- Perla, P. & McGrady E. (2011). "Why Wargames Work". *Naval War College Review*, Summer 2011, Vol 64. No. 3: Pp 111 – 130.
- PBS (Public Broadcasting System). (2018). "Destination Michigan: Camp Grayling," video online. Retrieved February 4, 2019. Online at <http://www.pbs.org/video/grayling-camp-grayling-pynfis/>
- Rege, A., Adams, J., Parker, E., Singer, B., Masceri, N. & Pandit, R. (2017a). "Using Cyber-security Exercises to Study Adversarial Intrusion Chains, Decision-Making, and Group Dynamics". Proceedings of the 16th European Conference on Cyber Warfare and Security.
- Rege, A., Parker, E., Singer, B. & Masceri, N. (2017b). A Qualitative Exploration of Adversarial Adaptability, Group Dynamics, and Cyber Intrusion Chains. *Journal of Information Warfare* 16(3): 1-16.
- Rege, A., Obradovic, Z., Asadi, N., Singer, S. & Masceri, N. (2017c). "A Temporal Assessment of Cyber Intrusion Chains Using Multidisciplinary Frameworks and Methodologies". Proceedings of the International Conference on Cyber Situational Awareness, Data Analytics and Assessment (Cyber SA 2017). Institute of Electrical and Electronics Engineers (IEEE) Xplore Digital Library.
- Wagen, W., & Pieters, W. (2015). From cybercrime to cyborg crime: botnets as hybrid criminal actor-networks. *British journal of criminology*, 55(2), 1-18.

# Western World Order in the Crosshairs? A Theoretical Review and Application of the Russian ‘Information Weapon’

Mari Ristolainen<sup>1</sup> and Juha Kukkola<sup>2</sup>

<sup>1</sup>Finnish Defence Research Agency, Riihimäki, Finland

<sup>2</sup>National Defence University, Helsinki, Finland

[mari.ristolainen@mil.fi](mailto:mari.ristolainen@mil.fi)

[juha.kukkola@mil.fi](mailto:juha.kukkola@mil.fi)

**Abstract:** Could you envision a weapon that would make the adversary destroy itself without even recognizing it was under attack? This paper highlights the importance of understanding the theoretical and contextual background of Russian cyber and information campaigns. We ask what the purpose of the Russian information and cyber campaigns is. How is the Russian ‘information weapon’ functioning? What is actually targeted? In this paper, we focus on a theoretical concept ‘information weapon’ by Professor Sergei Pavlovich Rastorguev (1958-2017). Our aim is to show how an existing theory is adequate for explaining contemporary Russian cyber and information campaigns. Moreover, our aim is to demonstrate how the theoretical framework of a ‘Russian information weapon’ could be applied in practice. Russian ‘information weapon’ can be used against a specific state or society. However, we argue that it is designed to threaten the entire ‘Western world order’. We illustrate two potential future information and cyberspace scenarios where the ‘information weapon’ envisioned by Rastorguev could destroy the free Internet and Western democracy and lead to the so-called ‘Post-Western world order’. This paper stresses the importance of comprehending the Russian way of thinking both on a theoretical and a conceptual level when making information and cyber security decisions or deciding on responsive actions.

**Keywords:** Russian information warfare theory, ‘Information weapon’, Sergei Pavlovich Rastorguev, cyber and information campaigns, cyber conflict

## 1. Introduction

*[...] by the time truth is revealed, work will be done, and the truth will be rewritten...”*

Sergei Pavlovich Rastorguev, 2013<sup>1</sup>

Could you envision a weapon that would make the adversary destroy itself without even recognizing it was under attack? This paper shows that Russian scholars have been theorising and lecturing on ‘information weapon’<sup>2</sup> (*informatsionnoe oruzhie*) for the past twenty years. Moreover, we argue that Russia has applied its own ‘information weapon’ into practise and will continue to do so – without the target recognising it. Yes, there are signs that Russia is actively preparing for a future cyber warfare against the ‘West’<sup>3</sup> (Nezavisimaya gazeta 2017; Vesti.ru 2017; Tvezvezda.ru 2017; Urusova 2018; Kommersant’ 2018). In addition, there are more than a few reports summarizing Russia’s alleged and decades-long information and cyber campaigns against different governments and their critical infrastructure (e.g. DiResta et al. 2018; Brantly & Collins 2018). Russian Foreign Minister Sergey Lavrov has openly stated that Russia aims for ‘Post-Western word order’ (Lavrov 2017). It seems that President Vladimir Putin aims to go down in history as leader who ended the Western domination of the world. Nevertheless, do we really comprehend what is happening? What is the purpose of the Russian information and cyber campaigns? How the Russian ‘information weapon’ is functioning? What is actually targeted?

This paper is a part of a larger research endeavour to systematically review and re-review Russian cyber and information theorists. To begin with, we have established what kind of Russian cyber and information theories exist. Next, we have conducted an in-depth analysis of the key theoretical works<sup>4</sup>. Due to the limits of one conference paper, we have chosen to present one theorist and to conduct a practical implementation of his not well-known (in the Western world) thoughts. We focus on a theoretical concept of ‘information weapon’ by Professor Sergei Pavlovich Rastorguev who has produced over twenty scientific works on the theme. Our aim is

<sup>1</sup> A quote from an interview of Professor Sergei Pavlovich Rastorguev (see: Nazarov 2013).

<sup>2</sup> Timothy Thomas (2005) has claimed that beginning from 1990s Russian scholars have used the term ‘information weapon’ to denote psychological and technological means to affect enemy decision-making and will. The Russian understanding of ‘cyberspace’ is more comprehensive than in the West and therefore it is called ‘information space’ and ‘information environment’ (cf. Tuchkov 2008). Thus, it can be argued that the Russian ‘information weapon’ includes both technical and information weaponry.

<sup>3</sup> In this paper, ‘West’ (Western) is defined as relating to the countries of EU and NATO and used without quotation marks hereafter.

<sup>4</sup> E.g. S. V. Rastorguev, V. N. Tsygichko, S. N. Bukharin, A. V. Manoilo, V. V. Tsyganov, I. N. Panarin, D. B. Frolov, S. I. Makarenko, S. N. Griniaev, I. S. Ashmanov and G. G. Pocheptsov.

to show how an existing theory is adequate for explaining contemporary Russian cyber and information campaigns. Moreover, our aim is to demonstrate how the theoretical framework of a ‘Russian information weapon’ could be applied in practice.

Russian ‘information weapon’ can be used against a specific state or society - even the Russian society itself. However, we argue that it is designed to threaten the entire ‘Western world order’. Therefore, we approach Rastorguev’s ‘information weapon’ as a means to affect the West. Consequently, we illustrate two potential future information and cyberspace scenarios where the ‘information weapon’ envisioned by Rastorguev could destroy the free Internet and the values it presents and Western democracy and way of life and lead to the so-called ‘Post-Western world order’.

## **2. Research method and materials**

Firstly, this paper is a theoretical review, i.e. we have conducted a contextual analysis of selected theoretical works of Sergei Pavlovich Rastorguev with the aim of developing interpretations that are adequate for explaining emerging research questions. Secondly, we use two potential future information and cyberspace scenarios as a method for the estimation of the actual target of the Russian ‘information weapon’. These scenarios allow us to reflect on Rastorguev’s theoretical ideas in the contemporary context of Russian cyber and information campaigns. In the study of the Russian ‘information weapon’, we combine Russian studies, Strategic studies, and IT studies.

The primary research material includes one journal article and five monographs by Sergei Pavlovich Rastorguev published 1997-2014. Secondary research material include interviews and reviews of his works. All the research material is originally in Russian and we have translated the key points into English<sup>5</sup>. As additional background sources we have used official Russian government documents and both Russian and Western media releases.

## **3. Selected works and concepts of Professor Sergei Rastorguev**

Professor Sergei Pavlovich Rastorguev (1958-2017), doctor of technical sciences, is considered as one of the founders of Russian scientific school of ‘information warfare’. His books on ‘information warfare’, published in the 1990s, have become classics of the theory and practice of information and cybersecurity in Russia. His works are used as university textbooks and his ideas are widely recognised in the Russian academic sphere. He lectured a special course on ‘cyber/information warfare’ for ten years at the Faculty of Computational Mathematics and Cybernetics of Lomonosov Moscow State University and at the Institute of Cryptography, Telecommunications and Informatics of Academy of the Federal Security Service of Russia (cf. Rastorguev 2014). Moreover, he was a member of the editor board of the journal called *Informatsionnye voiny* [Information Wars] supported by the Academy of Military Sciences of the Russian Federation (cf. [www.iwars.su](http://www.iwars.su)). It can be assumed that Professor Rastorguev has been serving as a scientific advisor for the Security Committee of the State Duma in the 2000s. Moreover, he has been acknowledged also in the Western analysis of Russian key authors of information and cyber related works (Thomas 2015, 166-167).

Professor Rastorguev was a mathematician and in his works, he identified his theoretical framework as ‘existential mathematics’ (*eksistentsial’naia matematika*) that approximately represents mathematical modelling of emotions, thoughts and doubts. However, in this paper, we do not examine Rastorguev’s mathematical models, but focus on reviewing his theoretical thoughts. In the following, the primary sources are shortly presented in chronological order to introduce and reflect upon the conceptual development of Rastorguev’s ‘information weapon’.

An article called “Information war as purposeful information influencing on information systems” published in 1997 was one of Rastorguev’s first works on the theme (Rastorguev 1997). It starts with basic conceptual definitions and introduces the basic idea behind the ‘information weapon’ (*informatsionnoe oruzhie*) and starts developing the concept of ‘self-learning information system’ (*informatsionnaia samoobuchaiushchaia sistema*) (Rastorguev 1997). Furthermore, in his first work, Rastorguev (1997) separates ‘cybernetic space’ (*kiberneticheskoe prostranstvo*) and ‘social space’ (*sotsial’noe prostranstvo*) from ‘information space’. Likewise, in this article he suggests that when speaking of ‘information warfare’ on technical systems the concepts ‘cybernetic warfare’ (*kiberneticheskaiia voina*) and ‘cybernetic weapon’ (*kiberneticheskoe oruzhie*) should be

<sup>5</sup> Note on transliteration and translation: Russian words are transliterated according to the Library of Congress system. The titles of documents and specific noteworthy concepts are given in translated form with transliterations.

applied (Rastorguev 1997). Interestingly, he supports this conceptualization in his following book (Rastorguev 1999a), but not in his later works. Overall, Rastorguev considers his works more theoretical than practical, i.e. technical.

In 1998, Rastorguev published a full academic study and a university textbook called “Information war”<sup>6</sup>. The Security Committee of the State Duma and the section of military-technical problems of the Russian Engineering Academy recommended it for publication. In this book Rastorguev’s concepts have on one hand expanded, but on the other become more clarified. Rastorguev’s ‘information weapon’ is an algorithm, i.e. logical function or mathematical operation (*Voennyi entsiklopedicheskii slovar’* 2001 s.v. algoritm) that allows a decisive control of one information system in the interests of another. This process is implemented by controlling the incoming or processed data of the system (Rastorguev 1999a, 61-62, 212). To apply ‘information weapon’ (*primenit’ informatsionnoe oruzhie*) is to select the incoming data in order to activate the certain algorithm and if that algorithm does not exist, the aim is to activate algorithms for generating these algorithms (Rastorguev 1999a, 76, 213). Information influencing (*informatsionnoe vozdeistvie*) is carried out with the use of information weapon. Information influencing contains means that allow implementation of the planned actions with the transmitted, processed, destroyed and perceived data (Rastorguev 1999a, 212).

Shortly after its publication, Rastorguev published a short monography called “Elections to power as a form of information expansion” (Rastorguev 1999b). In this work, he further develops the conceptualization of ‘information warfare’ and explains that contemporary wars do not aim to direct destruction of the adversary. Rather, the aim is ‘to reprogram’ (*pereprogrammirovat’*) the adversary (Rastorguev 1999b). As an example of the ‘reprogramming’ he uses elections. Elections to power are a way to achieve geopolitical goals: “If you want to destroy a rich and strong country help a local thief come to power. That is all you need. All the rest they will do by themselves, with their own hands.” (Rastorguev 1999b). Moreover, he continues to explicate the concept of ‘information weapon’ and how it can be used in situations where firearms and kinetic influencing is not allowed (Rastorguev 1999b). It can be argued that Rastorguev’s ideas are connected to the increasingly geopolitical and realist orientation of the Russian defence and security policy discourse in the latter part of the 1990s which already included Western information operations as a threat to Russia (cf. Thomas 2005).

In a monography called “Philosophy of information war” (Rastorguev 2003) published in 2003, Rastorguev already reflects his past works. This monograph contains all the earlier mentioned works and practical examples. Rastorguev shows that future wars attack firstly people’s worldviews and after that seize the whole country (Rastorguev 2003). In a preface written by one of the peer-reviewers, Professor Vladimir Aleksandrovich Razumnyi, notes that the defence in information war is protecting knowledge. And in this sense, the Internet, according to Razumnyi, aims at destruction of Russia. At this moment, the ‘information weapon’ was targeted at Russia (Rastorguev 2003).

In 2006, Rastorguev published a monograph called “Information war. Problems and models. Existential mathematics” (Rastorguev 2006). In this book, Rastorguev has turned his theory into mathematical models. Professor Anatoly Streltsov who at that time headed the Information Security Department of the Security Council of the Russian Federation reviewed this book and describes Rastorguev’s role as one of the main Russian authors in the field of ‘information confrontation’ (*informatsionnoe protivoborstvo*<sup>7</sup>). In 2014, another re-print was published of Rastorguev’s thoughts. A book called “Mathematical models of information confrontation. Existential mathematics” (Rastorguev 2014). Professor Rastorguev suddenly passed away in 2017. His last works were devoted to ‘egregore theory’ that is an occult concept representing a collective group mind (E.g. Rastorguev 2016, 2017). His last works do not fall into the subject of this paper.

---

<sup>6</sup> The first edition of *Informatsionnaia voyna* [Information war] (Moskva: Radio i sviaz’) by Sergei Rastorguev was published in 1998 (Rastorguev 2003). In this paper, we have used reprint from 1999 (Rastorguev 1999a).

<sup>7</sup> The Russian writers use this term to refer either to information warfare, as understood in the West, or to continuous competition between great powers with information psychological and technological means. The term *protivoborstvo* was used in the Soviet period in the context to foreign policy to denote competition between the two systems (capitalist and socialist), and it was redeployed later to characterize relations between great powers (i.e. Russia and the United States). It also resonates naturally with systemic theoretical ideas of the Russian information warfare theories who loan their theoretical framework from the Soviet era cybernetics.

#### **4. Rastorguev's 'information weapon'**

Professor Rastorguev envisioned a weapon that would make the adversary destroy itself without even recognizing it was under attack. In his first monographs on information war, Rastorguev (1999a, 6; 2003, 9) uses a classic fable based example when simplifying and explaining the basic idea of his 'information weapon':

*"Once upon a time there was a Turtle carrying his heavy shell. Every step was hard as the heavy shell pushed him to the ground. His life was not easy with the heavy shell. But when a hungry Fox came from the nearby forest, the Turtle hid under his shell and calmly waited the danger to pass. The Fox jumped around, tried to bite the shell, tried to turn the turtle over and used all the techniques of an aggressor, but the Turtle stood firmly on the ground and remained alive. One day the Fox came with a big purse and a lawyer and offered to buy the shell. The Turtle took long time to think, but because of his poor imagination, turned the offer down. The Fox left. Yet, there became a day when another hungry Fox came from the nearby forest. And the Turtle hid under his shell and calmly waited the danger to pass. The Fox jumped around, tried to bite the shell, tried to turn the turtle over and used all the techniques of an aggressor, but the turtle stood firmly on the ground and remained alive. As the time went by, the world changed. New telecommunication systems appeared in the forest. And one day, when the Turtle came out of his house, he saw a TV screen hanging on a tree showing a program with flying turtles without a shell. Overjoyed Woodpecker commented their flying: "How easy it that! What a speed! What a beauty! How smooth!" The Turtle watched the program for a day, two, and three... and started to think what a fool he is when he carries the heavy shell. Wouldn't it be better to drop it? Life would be much easier! He was scared a little, but in the latest news the TV presenter Owl said that that the Fox had become a Hare Krishna and a vegetarian. The world is changing. And the Turtle thought – Why not to fly? The sky is so big and beautiful! It's enough to give up the shell and life will become easier! The Fox saw the same program also and thought – Yes, it's enough to the Turtle to give up the shell and it will be easier to eat the Turtle at once! And one fine morning, when the sky seemed bigger than ever, the Turtle took its first and last steps towards freedom from its defence system. The Turtle did not know and will never know that information war is a purposeful training of the enemy to remove the shell from him by himself."*<sup>8</sup>

This turtle example captures perfectly the essence of Rastorguev's 'information weapon'. In the traditional fables (e.g. Aesop's Tortoise and Hare and its Russian variants by Ivan Krylov and Sergei Mikhalkov) turtle is commonly represented as an intelligent and clever character that does not underestimate the opponent. Rastorguev turns the traditional positioning over by representing that even the wise turtle is vulnerable to the information influence. Conceptually, the turtle and the supporting characters (i.e. the entire forest) portray Rastorguev's 'self-learning information system' that is an intelligent information system automatically forming knowledge based on real practical examples. Rastorguev (2014, 9) argues that human societies are self-learning and organizing systems which can be manipulated from the outside through information. This kind of systems can change their internal algorithms, structures and objectives under the influence of outside information. For Rastorguev (2014, 28-29), this is how the Soviet Union was destroyed and, also, how the West currently tries to keep Russia weak.

Rastorguev claims that information weapon outperforms any other type of weapon because it does not need 'energy' in order to destroy the adversary. Rastorguev argues that the basic principle of an information weapon is that the adversary has all the necessary resources for self-destruction (Rastorguev 1999a, 57; 2006, 15; 2014, 20). Information weapons are like viruses that reprogram self-learning systems (society or human mind) and change their structures until the system self-destructs (Rastorguev 2014, 19). They are used to activate, destroy, block or create processes in the information system (Rastorguev 2014, 20).

Nevertheless, the challenge of using information weapon is in the process where the adversary is directed to use his resources, including technical resources, against himself. Information threat (*informatsionnaia ugroza*) launches the self-destruction mechanism (Rastorguev 1999a, 110-111). The main purpose of information threat is to activate the algorithms responsible for unusual function mode that withdraws the system beyond permissible state. This requires explicit knowledge of the target system and the modelling of systemic processes (*ibid.*).

---

<sup>8</sup> Translated from an original example in Russian (Rastorguev 1999a, 6) by the authors.

The source of the information threat to the system can be either external or internal. A targeted information influencing causes external information threat (*prichiny vneshnykh ugroz*), i.e. a targeted information influencing causes a struggle between competing information systems for common resources that provide the acceptable mode of existence of the system (Rastorguev 1999a, 110). Internal threats appear inside the system (*prichiny vnutrennykh ugroz*) i.e. the acceptable mode of existence becomes unacceptable due to the changes in the many elements and subsystems (*ibid.*). Furthermore, the external and internal information threat can be divided into explicit (*iavnaia ugroza*) and hidden threats (*skrytaia ugroza*). An explicit threat is the data that is perceived as a threat. A hidden threat is the data that is not recognized by the system in real time or is not perceived by the system as something that threatens its security (Rastorguev 1999a, 112-113; 118).

Finally, the purpose of information operation (*informatsionnaia operatsiia*) is to change the behaviour of the system (Rastorguev 2014, 88). This means that the main feature of an information defeat (*informatsionnoe porazhenie*) are the changes in the behaviour of the affected system (*pereprogramirovanie*) (Rastorguev 2006, 50-51; 2014, 88). An information system that has been defeated is no longer led by its own interests, but controlled by adversary leaders (Rastorguev 1999a, 139). Rastorguev argued that the target system could defend itself through 1) setting a barrier between themselves and the source of danger; 2) avoiding danger by moving beyond its reach; 3) through destruction of the source of danger; and 4) through self-modification beyond recognition (Rastorguev 1999a, 116-117). The target system could also resist information operations based on its resiliency (*ustoichivost'*) which is a function of shared, non-fluctuating elements of knowledge. How these elements are connected and how the system is ordered affect the resiliency (Rastorguev 2014, 73-77). In practice, this means that a society with highly homogenous and deeply held culture and hierarchical leadership will resist information threats better than other societies.

Rastorguev (1999a, 219; 1999b, 6) emphasis that future conflicts will be more and more shifted towards the use of information weapons rather than firearms. Moreover, he explicates that, after all, the goal of any military action is to change the adversary's behaviour. Previously this desired result has been achieved by intimidation or kinetic destruction. Yet, the information confrontation can continue indefinitely and the winner is the one that can simulate the behaviour of the adversary in different situations (Rastorguev 1999a, 78-79). Thus, according to Rastorguev (1999a, 79), it is important to collect, store and process information about the adversary, i.e. to study the adversary's behaviour and to understand its history, culture, religion etc. This information can be used for launching the self-destruction mechanism.

## **5. Target of the Russian 'information weapon'**

Already in the late 1990s, Rastorguev stated that 'information war' is no longer a war between individual states – it is a war between modern 'civilizations'<sup>9</sup> (Rastorguev 1999a, 99-103). According to Rastorguev (1999a, 103), the emergence of new communication tools that can be used to control the masses bypassing territorial borders allows the theatre of the war to expand to such depths that are sometimes called as the "soul of the civilization" (*dukh tsivilizatsii*). Consequently, Rastorguev (1999a, 103) argued that it makes no sense to fight the 'information war' against a single state.

'Information war' should be aimed at the 'control systems' (*mehanizmi upravleniia*) and the structures that contain information about the 'control systems' (*sfera upravleniia*) (Rastorguev 1999a, 174-175; 2014, 90). In our interpretation, Rastorguev's war between 'civilizations' is similar to the contemporary Russian political rhetoric of the struggle between Russia and the rest. There exists a long term and serious determination to challenge the US-dominated/led world order and progress towards 'a Post-Western world order' in political, ideological, economic and technological fields<sup>10</sup>.

In the Russian approach, the Internet represents a threat to Russian cultural and technological integrity and sovereignty because it is by-product of the dominant American culture. The global Internet is dependent on popular applications and services, i.e. 'control systems' that are provided by the United States based companies. Furthermore, the 'Western world order' is based on 'government by the people'. Consequently, electoral system

<sup>9</sup> The term 'civilization' is ambiguous. In many sentences Rastorguev refers to 'Western civilization' (*zapadnaia tsivilizatsiia*) as a society with its own material and spiritual culture. Similar terminological approaches can be found, for instance, in Jackson 2002 and Omelicheva & Zubyska 2016.

<sup>10</sup> On a theoretical level, this claim integrates in a Russian way systems theory (more generally cybernetics) and cultural theories about civilizational conflicts (e.g. Huntington) (cf. Sakwa 2017).

and practices could be considered as ‘control system’ of Western democracy. Overall, democracy represents a threat to autocratic leadership.

Rastorguev (1999a, 220) defines the ‘information target’ (*informatsionnaia mishen*) as a set of elements belonging to (or that are capable of belonging to) the ‘control system’ and that have the potential resources for reprogramming the system for achieving the abnormal goals. These elements are highly dependent on information and thus, they will be vulnerable to manipulation through information.

Consequently, we claim that if Russia seeks ‘Post-Western world order’, the target of its ‘information weapon’ are the ‘control systems’ upholding the current world order. Russia has launched cyber and information campaigns on the Western democracies unobtrusively. On one hand, Russia has declared its aim to close off its national segment from the global Internet and to become ‘digitally sovereign’ (Kukkola & al. 2017). On the other hand, Russia’s interference campaigns against Western political processes have been comprehensively documented (e.g. DiResta et al. 2018; Brantly & Collins 2018).

If approached through Rastorguev’s concept of ‘information weapon’, these operations have created an information threat that has the potential to launch the self-destruction mechanism. Moreover, we argue that Russia has applied its own ‘information weapon’ into practise and will continue to do so – without the target recognising it. Next, we illustrate two potential future information and cyberspace scenarios where the ‘information weapon’ envisioned by Rastorguev could destroy the free Internet and Western democracy and finally, lead to the so-called ‘Post-Western world order’.

### **5.1 Internal and hidden information threat launches an uncontrolled usage of ‘the Internet kill switch’**

Russia has declared its aim to close off its national segment from the global Internet and to become ‘digitally sovereign’. At the same time, Russia might be pursuing a decisive military advantage in cyberspace through an asymmetry in offensive and defensive capabilities. Moreover, once Russia declares itself a ‘digitally sovereign’ nation it could cause a situation where ‘open network nations’<sup>11</sup> might be forced to choose what they do with their own national segments (Kukkola et al. 2017). This causes an information threat in the control system that could launch a self-destruction mechanism. The launching of a self-destruction mechanism is not related to any direct actions of Russia or objective circumstances in the technical layer but to the perceptions about the structural change in the cyberspace.

If applying Rastorguev’s concepts, here the information threat that launches the self-destruction mechanism is in essence internal as the acceptable mode of existence of the free and open Internet becomes unacceptable due to the Russian potential altering of the many elements and subsystems of the global open Internet (i.e. ‘open network society’). In other words, Russia’s national network closing processes passively create an internal information threat that forces the democratic and free nations to defend themselves and thus, destroy the free Internet and the values it presents (i.e. their ‘civilization’). Here the information threat stays hidden for longer period as the process of creating a nationally closed network is slow and is not recognized in real time. Moreover, the Russian national segment of the Internet is still not widely perceived as something that threatens the security of the global Internet.

Over the years, there have been many speculations of an ‘Internet kill switch’ that is a concept of activating a single shut off mechanism for all Internet traffic (c.f. Medows 2012). One could speculate, if forcing an ‘open network nation’ to an uncontrolled usage of such kind of ‘kill switch’ might be one of Russia’s strategic goals. A situation where national governments substantially restrict the information flows and connectivity of the network could cause serious effects to its critical infrastructure, economy, and alliances (cf. Lantto et al 2018), i.e. the adversary is directed to use his resources, including technical resources, against himself.

---

<sup>11</sup> An open network (i.e. global Internet) is defined in our studies as a network based on a multi stakeholder process, non-nation based governance, public-private partnerships, open access and global connectivity. The open network represents part of the global commons – a collective asset that secures freedom of expression, media pluralism, equal access to knowledge etc. Open network nations share the values of open networks and their segment of the Internet is built on those principles. The open network society is a collection of the above defined nations. A contradictionary concept used is ‘a closed network nation’ that is understood as a nation that is technically able to maintain a closed network, i.e. to operate a nationally governed segment of the Internet that can be technically separated from the global Internet.

## **5.2 External and explicit information threat grounds for abandoning elections**

Risks associated with elections and referendums, such as possible attempts by foreign states to influence the results create an atmosphere that there is no point in organizing elections because the voters are heavily influenced and the result becomes manipulated (e.g. the US presidential elections, Brexit).

If applying Rastorguev's ideas, here the information threat is both external and explicit. Targeted foreign electoral intervention to the 'control system' causes a struggle between competing information systems, i.e. democracy and non-democracy. Russia's cyber and information campaigns create an information threat that launches the self-destruction mechanism, i.e. in the end, there is no point to organize elections or referendums. Here the Russian 'information weapon' does not eternally and explicitly only manipulate election results, but destroys the Western democracy and way of life without the target even recognizing what was under attack. Here the 'control system' that has been defeated is no longer led by its own interests, but controlled by adversary leaders. The changes in the behaviour of the affected system feature information defeat. In 2017 in the Finnish presidential elections, comparable opinions were already publicly expressed. One of the candidates argued that referendums should not be organized in questions of security policy, e.g. Finland joining NATO, because this kind of referendum would make Finland "a cyber laboratory of election influencing" (Yle 2017; Haavisto 2017).

In addition to illustrating how 'information weapons' could work, the above presented scenarios give a glimpse of the nature of the Post-Western world order. It would be a world where globalization based on liberal cultural, political and economic ideals would be replaced by geopolitical realpolitik based on suspicion, cultural relativism, statism and zero-sum power politics. Although we should not be so naïve as to accept globalization (or capitalism or liberal democracy) as an absolutely positive phenomenon, neither should we forget that international system could also self-destruct if pushed in the wrong direction.

## **6. Conclusions: Facing the post-western word order?**

*"Question: Foreign Minister, the summit is happening in Helsinki. Russian President V. Putin and US President D. Trump together. Is this the Post-West world order that you have talked of in the past? Has it now arrived?"*

*S. Lavrov: Well, I think that we are in the Post-West world order, but this order is being shaped and it will take a long time. [...]" (Lavrov 2018).*

The aim of this paper was to show that an 'information weapon' envisioned by a Russian scientist outperforms any other type of weapon because it does not need 'energy' in order to destroy the adversary. The basic principle of an information weapon is that the adversary has all the necessary resources for self-destruction. Rastorguev demonstrates the advantages in maintaining foreign intelligence capabilities and explicitly acquiring knowledge about the culture of the adversary. When the 'control systems' of the opponent have been identified, systems analysis and mathematics guide 'information weapons' to their target. Nevertheless, although Rastorguev is quite confident in his theory, it is common knowledge that exact and working cybernetic models of living organisms and societies are difficult to produce. Because Rastorguev's 'information weapon' is based on detailed information about the target, there is always the danger of misinterpretation. This could lead to unforeseen processes and outcomes in the target system, i.e. society or state. This danger, of course, does not stop Russians from trying. Rastorguev's thoughts about the 'information weapon' are adequate for explaining contemporary Russian cyber and information campaigns. The described potential future information and cyberspace scenarios challenge the assumptions that are taken for granted and suggest that the Russian 'information weapon' is not targeted at any specific nation. Rather, it threatens the entire 'Western world order'. Besides, it seems that Russia is following Rastorguev's idea's how to defend itself in the 'information war', i.e. a society with highly homogenous and deeply held culture and hierarchical leadership will resist information threats better than other societies. Consequently, it is important of comprehend the Russian way of thinking also on theoretical level when making security decisions or deciding responsive actions.

## **References**

Brantly, A. & Collins, L. (2018) "A Bear of a Problem: Russian Special Forces Perfecting Their Cyber Capabilities", Association of the United States Army, 28 November 2018 [Online]. <https://www.usa.org/articles/bear-problem-russian-special-forces-perfecting-their-cyber-capabilities> [Accessed 15 January 2019].

- DiResta, R., Shaffer, K., Ruppel, B., Sullivan, D., Matney, R., Fox, R., Albright, J. & Johnson, B. (2018) "The Disinformation Report", New Knowledge, 17 December 2018, [Online].  
<https://disinformationreport.blob.core.windows.net/disinformation-report/NewKnowledge-Disinformation-Report-Whitepaper-121718.pdf>
- Haavisto, P. (2017) "Kolme syytä kuopata kansanäänestys - perustuslakia ei saa sivuuttaa", Blog post, 15 December 2017 [Online], <http://www.pekkahaavisto.fi/kolme-syyta-kuopata-kansanaanestys-perustuslakia-ei-saa-sivuuttaa/> [Accessed 16 January 2019].
- Jackson, W.D. (2002) "Encircled Again: Russia's Military Assesses Threats in a Post-Soviet World", *Political Science Quarterly*, Vol 117, No. 3, pp 373-400.
- Kommersant' (2018) "Roskomnadzor otritsaet plany potratit' 20 mlrd rublei na blokirovku Telegram", *Kommersant'*, 24 dekabr'ia 2018 [Online]. <https://www.kommersant.ru/doc/3841663?from=hotnews> [Accessed 15 January 2019].
- Kukkola, J., Ristolainen, M. & Nikkarila, J.-P. (2017) *Game Changer: Structural Transformation of Cyberspace*, Finnish Defence Research Agency, Riihimäki.
- Lantto, H., Åkesson, B., Kukkola, J., Nikkarila, J-P & Ristolainen, M. (2018) "War-gaming a closed national network: What are you willing to sacrifice?" *Conference for Military Communications (MILCOM)*, 29-31 October, Los Angeles, CA, USA.
- Lavrov, S. (2017) "Foreign Minister Sergey Lavrov's address and answers to questions at the 53rd Munich Security Conference", Munich, February 18, 2017, 18 February 2017. [Online]. [http://www.mid.ru/en/foreign\\_policy/news/-/asset\\_publisher/cKNonkJE02Bw/content/id/2648249](http://www.mid.ru/en/foreign_policy/news/-/asset_publisher/cKNonkJE02Bw/content/id/2648249) [Accessed 15 December 2018].
- Lavrov, S. (2018) "Foreign Minister S. Lavrov's interview with Channel 4", Moscow, June 29, 2018 [Online].  
[http://www.mid.ru/en/foreign\\_policy/news/-/asset\\_publisher/cKNonkJE02Bw/content/id/3285972](http://www.mid.ru/en/foreign_policy/news/-/asset_publisher/cKNonkJE02Bw/content/id/3285972) [Accessed 17 January 2019].
- Medows, D. B. (2012) "The Sound of Silence: The Legality of the American "Kill Switch" Bill"" *Journal of Law, Technology & the Internet*, vol. 4, no. 1, pp. 59-79.
- Nazarov, O. (2013) "Informatsionnye voiny – ugroza dlia tsivilisatsii", *Literaturnaia gazeta*, No. 42 (6435), 23-29 oktiabria 2013 g. [Online]. [http://www.reading-hall.ru/lit\\_gazeta/42\(6435\)2013.pdf](http://www.reading-hall.ru/lit_gazeta/42(6435)2013.pdf) [Accessed 15 December 2018].
- Nezavisimaia gazeta (2017) "Minoborony mozhet pustit' v khod voiska informatsionnykh operatsii", *Nezavisimaia gazeta*, 22 fevralia 2017 [Online]. [http://www.ng.ru/armies/2017-02-22/100\\_informvoiska.html](http://www.ng.ru/armies/2017-02-22/100_informvoiska.html) [Accessed 15 January 2019].
- Omelicheva, M.Y. & Zubytska, L. (2016) "An Unending Quest for Russia's Place in the World: The Discursive Co-evolution of the Study and Practice of International Relations", *New Perspectives*, Vol. 24, No. 1/2016, pp. 19-51.
- Rastorguev, S. P. (1997) "Informatsionnaia voina kak tselenapravlennoe informatsionnoe vozdeistvie informatsionnykh system", *Informatsionnoe obshchestvo*, No. 1
- Rastorguev, S. P. (1999a) *Informatsionnaia voina*, Radio i sviaz', Moskva.
- Rastorguev, S. P. (1999b) *Vybory bo vlast' kak forma informatsionnoi ekspansii*, Novyi vek, Moskva.
- Rastorguev, S. P. (2003) *Filosofia informatsionnoi voiny*, Vyzovskaia kniga, MPSI, Moskva.
- Rastorguev, S. P. (2006) *Informatsionnaia voina. Problemy i modeli. Eksistentsial'naiia matematika*, Gelios ARV, Moskva.
- Rastorguev, S. P. (2014) *Matematicheskie modeli v informatsionnom protivoborstve. Eksistentsial'naiia matematika*, ANO TSSOIP, Moskva.
- Rastorguev, S. P. (2016) *O vlozhdennosti prostranstv (teoriia egregorov)*, Izdatel'stvo reshenia, Moskva.
- Rastorguev, S.P. (2017) *Prakticheskie aspekty teorii vlozhennosti prostranstv*, ANO TSSOIP, Moskva.
- Sakwa, Richard (2017) *Russia against the rest: The post-cold war crisis of world order*. Cambridge University Press, Cambridge.
- Thomas, T. (2005) *Cyber Silhouettes: Shadows Over Information Operations*, Foreign Military Studies Office (FMSO), Fort Leavenworth, KS.
- Thomas, T. (2015) "Russia's Information Warfare Strategy: Can the Nation Cope in Future Conflicts?", McDermott, Roger N. (ed.), *The Transformation of Russia's Armed Forces: Twenty Lost Years*, London & New York: Routledge, pp. 148-177.
- Tuchkov, Iu.N. (2008) *Slovlar' terminov i opredelenii v oblasti informatsionnoi bezopasnosti*, 1-e izd, VAGSH RF, NITS informatsionnoi bezopasnosti, Moskva.
- Tvzvezda.ru (2017) "Ekspert rasskazal, chem budut zanimat'sia voiska informatsionnykh operatsii armii RF", *Tvzvezda.ru*, 22 fevralia 2017 [Online]. [https://tvzvezda.ru/news/vstrane\\_i\\_mire/content/201702221744-n6h2.htm](https://tvzvezda.ru/news/vstrane_i_mire/content/201702221744-n6h2.htm) [Accessed 15 January 2019].
- Urusova, A. (2018) "Glava Roskomnadzora: perspektivy otkliucheniiia Rossii ot interneta seichas net (interv'iui)", *Tass.ru*, 24 dekabr'ia 2018 [Online]. <https://tass.ru/interviews/5937355> [Accessed 15 January 2019].
- Vesti.ru (2017) "Voiska informatsionnykh operatsii: teper' ne tol'ko oborona, no i napadanie", *Vesti.ru*, 22 fevralia 2017 [Online]. <https://www.vesti.ru/doc.html?id=2858620> [Accessed 15 January 2019].
- Voennyi entsiklopedicheskii slovar' (2001) *Voennyi entsiklopedicheskii slovar'*, tom I., Naychnoe izdatel'stvo "Bol'shaia Rossiiskaia entsiklopediia", Moskva.
- Yle (2017) "Vihreiden Haavisto ei halua Nato-kansanäänestystä - "Yksikään hybridti- ja kybervaikuttaja ei pysyisi siitä poissa"", Yle, 27 November 2017 [Online]. <https://yle.fi/uutiset/3-9951046> [Accessed 15 January 2019].

# A Cross-Discipline Approach to Countering 4th Generation Espionage

Char Sample<sup>1</sup>, Keith Scott<sup>2</sup> and Emily Darraj<sup>3</sup>

<sup>1</sup>ICF Inc., Columbia, USA

<sup>2</sup>De Montfort University, UK

<sup>3</sup>Capitol Technology University, Laurel, USA

[char.sample@icf.com](mailto:char.sample@icf.com)

[jklscott@dmu.ac.uk](mailto:jklscott@dmu.ac.uk)

[edarraj@captechu.edu](mailto:edarraj@captechu.edu)

**Abstract:** In 2018 the UK government introduced the term ‘4<sup>th</sup> generation espionage’ to define the hybrid threats exemplified in fake news. As events occur, even seemingly benign events, the rush to report and take charge of the narrative can result in the facts being diminished, or in some cases obscured. The need to quickly and accurately assess the content suggests a role for AI/ML solutions. Before AI/ML can be applied, proper rules for training data must be considered and those rules require inputs from non-technical as well as technical disciplines. The researchers examined the rules of rhetoric, propaganda, and linguistics, then mapped these rules to the behavioural sciences first, and, the technical counterparts associated with computational linguistics and mathematics providing the framework for news story labelling. The examination of the various disciplines is discussed and how they each feed the fake news process.

**Keywords:** propaganda, rhetoric, computational linguistics, behavioural science, fake news

---

## 1. Introduction

Fake news (FN) officially joined the lexicon in 2017 as word of the year (Meza 2017). Considered synonymous with propaganda, this iteration differs from earlier generations in several areas including the interactive nature, message customization, and the exploitation of trust relationships. Younger (2018) introduced the term “4<sup>th</sup> generation espionage” to describe the fusing of “traditional human skills with accelerated innovation” where thought diversity and agility will be required. Countering this threat requires the convergence of data science, behavioural science and technology with tight coupling of interdisciplinary research solutions (Sample et al. 2018; Younger 2018).

Propaganda historically relied on inputs from various disciplines including the social sciences in order to craft a resonating message targeting values of intended audiences. This remains unchanged; however, the communications channel today is synchronous, and amplified at “speed and scale”(Verrall & Mason 2018). Countermeasures will require automated responses; however, automated solutions (Horne & Adali, 2017, Hamouda 2018; Isle & Smith 2018; Lemieux & Smith 2018) hold promise, but their accuracy rates, are inadequate at “speed and scale” (Verrall & Mason 2018). Consider that a solution with a 90% success rate in evaluating a million stories would result in 100,000 FN stories being mislabelled. When FN stories are targeting hundreds of millions (Booth et al., 2017; Lee & Kent 2017) the decision criteria requires a robust set of requirements. Lee & Kent (2017) noted that FN stories once shared, reached 126 million users.

The time required to disprove FN is longer than the time needed to plant the seed of a customized narrative through rhetoric and repetition. While an automated response provides a rapid countering mechanism, in order for the automated response to be effective the response must also be accurate. Accuracy requires contextualized training data that reflects the complete set of rhetorical devices used in propaganda; current training data uses known examples of propaganda.

This problem requires a deeper understanding of audience targeting, message creation, distribution and propagation; how each of these can utilize technologies in all manners, while determining the contribution of various academic disciplines in the overall process. The delineation between fact and fiction has been blurred (Younger 2018), and academic disciplines that are traditionally separated into silos must now work together, as equals supporting a common goal.

A reasonable starting point begins with terminology. Rhetoric can be defined as the art of persuasion ([www.oxforddictionaries.com](http://www.oxforddictionaries.com)). Propaganda is defined as biased or misleading information supporting a viewpoint (*Ibid*). While rhetoric and propaganda carry negative connotations both rhetoric and propaganda also have positive uses. For example, Neil Armstrong’s use of the rhetorical device antithesis when he said, ‘That’s

'one small step for man, one giant leap for mankind' was considered inspirational (Butler 2017). The fact that rhetoric and propaganda can carry both negative and positive connotations creates one of several currently unresolved challenges for AI software.

AI software continues to improve. However, AI also struggles with nuance and other subtle behaviours. Unsupervised learning generally produces weaker findings than does supervised learning (CITE). In order for supervised learning to be accurate and effective the data must accurately reflect the environment (CITE); thereby allowing the machine learning instances to be applied over the larger data set. When dealing with deceptive data in general, and fake news specifically, the data ranges from mildly biased to outright falsehood. Furthermore, deceptive data relies on linguistic features or rhetorical methods to form the message these rhetorical methods must be understood by the technical professionals (e.g. programmers) in order to create the needed rules for detection and countering.

This effort can be considered as a joining of rhetoric to the creation of disinformation and cybersecurity. Of note, the terms disinformation, FN and propaganda are used interchangeably throughout this document. Disinformation is a broad term that encompasses many different types of information, not simply news and propaganda can be considered a superset of FN.

The deliberate manipulation of public opinion through the transmission of misleading communication is of course nothing new, as even the most cursory reading of the literature on the history of propaganda will show (see Bernays 1928, Ellul 1962, Herman and Chomsky 1988, Taylor 2003). However, the growth of the Internet and the wholesale adoption of digitally-mediated communication (DMC) as an essential medium for information distribution has undoubtedly changed both the political and media landscapes and the multiplied the number of potential threat actors and the domains where they can exert influence. Designing and deploying the tools of influence to a mass audience was previously the domain of governments and organizations with the financial capacity to print and audio-visual media; the birth of DMC has seen a democratization of the process, with the tools for creating effective influence campaigns, and access to a medium which allows their near-instantaneous, near-global distribution, available to anyone with an Internet connection and a networked device (Marwick and Lewis (2017) offers an excellent summary of the current state of affairs). Just as the power of traditional broadcast journalism and entertainment has been challenged by the rise of blogs, newsfeeds, and streaming video, so conventional 'mainstream' propaganda has been drowned out by a rising tide of small groups and individuals. In non-kinetic as much as in kinetic conflict, we see the rise of asymmetric warfare.

What we might term 'new propaganda' (a broader definition of this form of online IO than 'fake news') shares many common features with its predecessor (and these, as will be argued, are essential to any attempt to detect and counter it). However, there are certain elements, which are unique to this form of DMC must be considered as central to its very nature. Among these elements are the following:

**Rapidity of transmission/spread:** DMC permits near-instantaneous, quasi-global reach, with minimal resource implications;

**Filter bubbles and echo chambers:** although the idea is open to question (see Gentzkow and Shapiro, 2010; Bakshy, Messing, and Adamic, 2015; Zuiderveen Borgesius et al, 2016), the growing application of tailored newsfeeds and algorithmically-driven information distribution on social media can be seen as limiting access to a wide range of information. Put crudely, platforms driven by advertising revenue wish to maximise user interaction with these sites, and will do all they can to ensure that users receive only information that conforms to their preconceptions – 'give the people what they want'. The risk of the 'cyber-Balkanization' of the online world into what Clyde Wayne Crews dubbed the 'Splinternet' (Malcolmson 2015) is real, and from an influence perspective, deeply troubling. A clearly defined group, whether on- or off-line, is much easier to target, as its ideologies and prejudices are easy to determine, and to exploit. A recent study (Vosoughi, Roy, and Aral, 2018) suggests that false information spreads more quickly than truth; the degree to which fake news is deliberately written to chime with the target audience's world-view and push their 'hot buttons' is surely an important driver in this.

‘Ha Ha Only Serious’<sup>1</sup>: the weaponization of irony. If we examine much of the mis/disinformation spread online, and in particular that produced by the so-called ‘alt-right’, one of the hallmarks of this particular form of ‘fake news’ is a particular rhetorical stance: knowing, sardonic, and couched in a would-be humorous manner, where the apparent tone can be seen to defuse or lessen the effect of material which if presented straight would be seen as overtly sexist, racist, or Islamophobic/antisemitic. The use of cartoons, manipulated images, and elements drawn from popular culture further presents what is being said as ‘just a joke’, thereby presenting those who seek to combat these messages as killjoys. This is not merely adolescent iconoclasm or gallows humour; it is a deliberate strategy, part of an attempt to win over new recruits by masking hate speech as simply something done ‘for the lulz’ (Moussa 2017). In 2017, the *Huffington Post* (Feinberg 2017) obtained a copy of the style guide used by the white supremacist website *The Daily Stormer*, where the site’s owner, Andrew Anglin, states explicitly that the aim of his approach is to propagandize through humour:

*The goal is to continually repeat the same points, over and over and over again. The reader is at first drawn in by curiosity or the naughty humour, and is slowly awakened to reality by repeatedly reading the same points [...] the undoctrinated should not be able to tell if we are joking or not. There should also be a conscious awareness of mocking stereotypes of hateful racists. I usually think of this as self-deprecating humour - I am a racist making fun of stereotyped racists, because I don't take myself super seriously.*

*This is obviously a ploy and I actually do want to gas kikes. But that's neither here nor there.*  
[Feinberg 2017; Moosa 2017]

The approach taken by *The Daily Stormer*, it must be emphasised, is not unique to that website; it is a central element in much of today’s ‘fake news’; while a central element of alt-right propaganda (Bokhari and Yiannopolous 2016), far from unique to this end of the political spectrum. As Zannetou et al (2018) show, the use of memes as a tool of influence is widespread, and the ability to identify and counter their use (ideally through automated detection and removal before posting online) would be hugely useful as a tool of counter-influence. This will be a major challenge for the future, requiring the ability, not simply to recognise key words or phrases (an elementary task) but understanding what they mean in a particular context. There have been recent advances in this field, most notably in the ability to detect through automated analysis (Horne and Adali 2017; Joshi, Bhattacharyya, and Carman 2017; Hazarika et al 2018), but there is much more to be done. A technical approach cannot be the sole solution; what is needed is a multidisciplinary methodology, employing IT and AI skills, a knowledge of linguistics (and in particular rhetoric), and expert knowledge from across the domain of influence studies, drawing on military, political, and commercial expertise – it should be noted that one of the best introductions to online influence campaigns is a work written by the former Director of Marketing for American Apparel (Holiday 2013).

## **2. Background**

Mourao & Roberston (2019) argue that no single definition exists for FN and that the definition is fluid. Pomerantsev & Weiss (2014) noted that modern day disinformation campaigns were founded on the premise that anything can be said, and reality can be created to support the statement. The task become easier online where data verification is weak as well as time consuming, and sensing can be manipulated.

The “fuzzy” nature of FN creates a greater countering challenge for binary-based technologies. Younger (2018) discussed the blurred of boundaries defined by ambiguity of real and virtual. Some examples of blurring in news are partial truths and facts taken out of context. These “fuzzy” lines make difficult the challenge of determining the contents required for constructing a resilient foundation. Furthermore, “fuzzy” lines and logic are generally problematic for automated solutions. Automated solutions remain binary in nature and varying degrees of accuracy can be found in news stories that may be factually accurate but biased or factually inaccurate as well as biased.

At its heart, however, we can say that FN, like propaganda, is essentially a persuasive form, a mode of communication, which seeks to influence its target to act or believe in a way planned by the creator(s) of the information to which they are exposed. It is, in short, a contemporary example of rhetoric, defined by Aristotle as “the faculty of observing in any given case the available means of persuasion,” (Aristotle, *Rhetoric*, Book I,

---

<sup>1</sup> A phrase that aptly captures the flavour of much hacker discourse. Applied especially to parodies, absurdities, and ironic jokes that are both intended and perceived to contain a possibly disquieting amount of truth, or truths that are constructed on in-joke and self-parody. (From: <http://catb.org/~esr/jargon/html/H/ha-ha-only-serious.html>)

Part II, Paragraph I). It relies on the effective crafting of language, and rests on a number of key underlying principles.

Three common rhetorical groupings are ethos, pathos and logos, are examined. Many different factors influence human belief and decision-making systems; these three because of their universal appeal provide a starting point. A brief discussion of the groupings follows.

Ethos defines the character of either the target or the messenger (Cockcroft & Cockcroft 2005, pp.28-54). A messenger desiring to appear credible to the audience, and will rely on audio-visual aids, such as wearing a suit or uniform. Similarly, the target will be linked to questionable character traits through unflattering pictures, crude or vulgar phrases. These modern versions of the Classic techniques of laus and vituperatio are well-known (Leonard 2019) but not easily evaluated in the digital environment. Images arrive as a series of bits that may or may not be lacking context, but cannot be immediately verified. However, descriptors may provide a potential marker to measure the distance between event facts, and emergent narratives that are customized to the audience's values (*Ibid*). Invoking pathos leads the reader to choose a position based on feelings, not logic.

Biases and beliefs provide shortcuts to decision-making (Croskerry, 2013; Haselton et al. 2013) these same biases are exploited in the deployment of pathos ([psywarrior.com](http://psywarrior.com)). Pathos based words constitute a rather large search criteria; however, adverbs and punctuation can act as filters (Sample et al. 2018). For example “!” is specifically associated with strong emotions with urgency ([Merriam-webster.com](http://Merriam-webster.com)). One obvious set of tools to draw on in the analysis of pathos-driven FN is Sentiment Analysis, and while there are clear limitations to this, the application of software such as VADER, LIWC, and other automated analysis hold promise for future research (Gonçalves et al 2013; Hutto and Gilbert 2014; Ribeiro et al 2016; Tausczik and Pennebaker 2010)

Logos appeals to logic (Cockcroft and Cockcroft 2005, pp.81-134). Logical constructs should be deployed such as “if-then”, or other derivations such as “when-then”. Another appeal to logic might involve questioning, and the use of “?” may be an early signal. As Varpio (2018) observes, the use of ‘signposting’ words (e.g., *first*, *next*, *specifically*, *alternatively*, *also*, *consequently*, etc.) and phrases (e.g., *as a result*, *and yet*, *for example*, *in conclusion*, etc.) are also clear markers of a logos-based rhetorical approach. However the appearance of logical signposts and textual markers does not of necessity indicate a truly logical argument; a deeper level of fine-grained analysis is necessary to determine the validity of the arguments presented. Work done by Song, et al. (2017) offers some interesting pointers towards future work on automated analysis of argument.

Usage of these three rhetorical methods (ethos, pathos and logos) in propaganda can be observed through several tools. Some common propaganda tools are; name calling, using glittering generalities, transfer, testimonial, plain folks, bandwagon, fear, bad logic and unwarranted extrapolation ([www.pbs.org](http://www.pbs.org) 2019). These tools can interact with the target audiences' basic human emotional needs for safety and acceptance.

All successful rhetoric relies on the use of the Aristoteian triad (with a heavy bias towards ethos and particularly pathos). A fourth element to be considered is *kairos*, or the ability to say the right thing at the right time. These underlying concepts are embodied and reinforced through the use of carefully chosen rhetorical figures, creating a text/speech, which is crafted, however artless it may appear (the 'just plain folks' approach). A full list of all rhetorical figures is beyond the scope of this paper (see Cockcroft and Cockcroft 2005 for a detailed discussion); what is important is that they can be identified and classified, and can be used to construct stylistic analyses of language which can act as a tool for attribution, and indeed for constructing counter-texts which use the same techniques. Automated identification of rhetorical figures is an important element of any attempt to develop a means of identifying FN, and there is a growing number of scholars developing computer-aided rhetorical tagging of these features (Dubremetz, M., and Nivre, J. (2017); Gawryjolek, J., C. Di Marco, C.D., and Harris, R.A. (2009); Harris, R. and Di Marco, C.D. (2009); Kelly, A.R., Abbott, N.A., and Harris, R.A. (2010); Mladenovic, M. (2016)).

Technically speaking, the challenges to accurately identifying fake news are considerable. Linguistic tools offer the ability to rapidly quantify content, however, these tools were not designed to detect the various forms of deceptive data. Some of the features of tools that make possible insights with large data are in conflict with deceptive data processing. When text is initially read and processed punctuation is removed, words are stemmed; stop words are removed before the sentences are placed into vectors. Thus, brief list of relevant items follows with discussion.

- The removal of punctuation such as "?" interferes with the ability to identify *anthypophora*, the use of a question for dramatic effect. Similarly, the presence of "!" in text, is placed for dramatic effect. Thus, we can say that pre-processing will need to add punctuation metrics into rules.
- The creation of the "us vs them" narrative uses pronouns including "us", "them", "they", "we" all pronouns that are eliminated in the initial language processing where sentences are read into vectors and stop words are stripped out. Pre-processing to strip out the words removes the pronouns, and in some cases this is necessary for other operations using computational linguistics as the frequently used stop words create misleading weights that would skew other operations.
- The word stemming process is used to reduce words into the root thereby accounting for tense use (e.g. the verbs study, studied, studying are all reduced to study). Active voice is more forceful than is passive voice.
- An additional problem deals with the initial processing to put word in all lower case. This eliminates the same word being grouped differently because of case sensitivity. In many cases, the use of all uppercase letters in a word, sometimes referred to as shouting, is of high importance and can be stripped away before processing begins.

This short list is certainly not comprehensive, but does illustrate some of the challenges being faced by those who wish to use computational linguistics to detect deceptive data. This list also underscores the value of combining the knowledge of linguists with traditional cybersecurity research where traditional cybersecurity professionals rely on software that may work but was not designed for the task. The merging of disciplines as will likely yield more insights that can be used to improve the fidelity and efficacy of solutions to the fake news problem.

### **3. Proposed solution**

The proposed solution acknowledges the need for interdisciplinary contributions by focusing on characteristics of fact-based narratives. The importance of determining the training data characteristics cannot be overstated. Certain linguistic features and human behaviours could be used to detect the presence of propaganda in support of FN (Sample et al. 2018), or an influence operation. These features and behaviours are readily observable in the digital environment. Algorithms can perform countering actions but they must learn with correct data. The detection training data is an anticipated output for this effort, and will rely on combining behavioural science, linguistics and mathematics. A brief description (*Ibid*) proposed detection method (Sample et al., 2018) of these tools is shown in Table 1.

**Table 1:** Propaganda technique to detection mapping

Technique	Description	Detection method
Name calling	Link person or idea to negative attribute.	Pattern spread
Glittering generalities	Link person or idea to positive attribute.	Pattern spread
Transfer	Link authority or respect to idea.	Pattern spread
Testimonial	Celebrity endorsement	Pattern spread
Plain Folks	Ideas or person is normal person	Computational linguistics
Bandwagon	Everyone agrees, join the group	Computational linguistics
Fear	Warns of negative results if idea not followed	Computational linguistics
Bad logic	Faulty logic to promote a cause	Computational Linguistics and Metadata analysis
Unwarranted extrapolation	Extrapolate big predictions from small data	Computational linguistics

Underpinning the identification of these key propaganda techniques will be a series of automated tools which will apply Corpus Linguistics (CL) methodology to enable a finer-grained analysis. A database of identified FN texts will act as a baseline reference corpus; this will be balanced by a reference corpus of 'genuine' news texts (a series of corpora will be required for each specific social media platform), to allow iterative and progressively refined analysis. These corpora will allow the analysis of any possible FN text through the identification of repeated lexical features (words and phrases), keywords (words and phrases which appear more frequently in the text under examination than in a baseline text, or which are seen to be general hallmarks of FN), and collocations (significant combinations of words). This analysis is the first step towards more detailed stylometric analysis, and a more precise identification of FN at the point of uploading to a social media platform, rather than after it has entered the public domain. Two further forms of analysis are also envisaged:

1. Veracity analysis: while there is as yet no completely accurate tool for automated identification of deception, there is a significant body of research which suggests that there are definite linguistics markers of deceptive communication, and that these markers can be identified automatically (Arciuli, Mallard, and Villar (2010); Burgoon, Blair, Qin, and Nunamaker (2003); Burgoon(2018); Ghosh, Fabbri and Muresan, 2017; Newman, Pennebaker, Berry, and Richards (2003); Zhoud, Burgoon, Nunamaker, and Twitchell (2004).)
2. Tonal analysis: as stated above, much FN is couched in a deliberately ironic, knowing manner, which requires qualitative, rather than quantitative analysis. This is an area for future development, but the work cited above on sarcasm detection offers a potentially fruitful path for study.
3. Network and geolocation analysis: the use of an automated tool such as FireAnt (<http://www.laurenceanthony.net/software/fireant/>) allows the tracking of the spread of FN across a social media platform, enabling the identification of key transmitters by both time and location; this is an invaluable aid to determining attribution.

#### **4. Conclusion**

Human intelligence and AI must work together, in the complex world of hybrid warfare (Younger 2018). Cambridge Analytica illustrated the power of the pairing (Berghel 2018) behavioural science and data science in support of hybrid warfare. The techniques used were controversial, yet highly effective in 2016. Bots and trolls manipulated sentiment analysis (Rosenblatt 2018), thereby calling into question the efficacy of relying on this metric (Hu et al. 2012). This subject demands a robust countering solution that minimally draws inputs from disciplines listed in this paper.

Before applying automated techniques to counter the fake news problem the data needs to be understood in depth to a deeper level of understanding than current automated solutions provide. The ability to identify fake news based on existing patterns recreates the signature-based paradigm that plagues many of the cyber security products. Understanding the data, both live and training data, requires the researchers to move beyond the bits that comprise the object to understanding the context in which the data was created and resides using paradigms from disciplines such as linguistics, psychology, anthropology, economics, political science and even military science and likely other disciplines not listed here. When addressing fake news cybersecurity should remember that while propaganda is not new, the current version is a highly effective weapon in the hybrid warfare arsenal.

#### **References**

- Arciuli, J., Mallard, D., and Villar, G. (2010) "“Um, I can tell you're lying”: Linguistic markers of deception versus truth-telling in speech", *Applied Psycholinguistics*, 31, pp. 397 - 411.
- Bakshy, E., Messing, S., and Adamic, L.A. (2015), "Exposure to ideologically diverse news and opinion on Facebook", Vol. 348, Issue 6239, pp. 1130-1132.
- Berghel, H., 2018. "Malice Domestic: The Cambridge Analytica Dystopia", *Computer*, vol. 5, pp.84-89.
- Bernays, E. (1928). Propaganda. Routledge, London.
- Bokhari, A., and Yiannopoulos (2016) "An Establishment Conservative's Guide To The Alt-Right" [online], *Breitbart.com*, 29 March, <https://www.breitbart.com/tech/2016/03/29/an-establishment-conservatives-guide-to-the-alt-right/#>
- Booth, R., Weaver, M., Hearn, A., Walker, S. and Walker, S. 2017, November 14, "Russia used hundreds of fake accounts to tweet about Brexit, data shows", *The Guardian*. Available: <https://www.theguardian.com/world/2017/nov/14/how-400-russia-run-fake-accounts-posted-bogus-brexit-tweets>.
- Burgoon, J.K., Blair, J.P., Qin, Tiantian, and Nunamaker, J.F. (2003) "Detecting Deception through Linguistic Analysis", International Conference on Intelligence and Security Informatics ISI 2003: Intelligence and Security Informatics, pp. 91-101
- Burgoon, J.K. (2018) "Predicting Veracity From Linguistic Indicators", *Journal of Language and Social Psychology*, Vol 37, Issue 6.
- Butler, D.A., 2017. Who Owns the Moon, Mars, and Other Celestial Bodies: Lunar Jurisprudence in Corpus Juris Spatialis. *J. Air L. & Com.*, 82, p.505.
- Cockcroft, R., and Cockcroft, S. 2005, *Persuading People: An Introduction to Rhetoric*. London, Palgrave.
- Croskerry, P. 2013, "From mindless to mindful practice—cognitive bias and clinical decision making", *New England Journal of Medicine*, vol. 368, no. 26, pp.2445-2448.
- Dubremetz, M., and Nivre, J. (2017) "Machine Learning for Rhetorical Figure Detection: More Chiasmus with Less Annotation", Proceedings of the 21st Nordic Conference of Computational Linguistics, pp. 37–45.
- Ellul, J. (1962) Propagandes, A. Colin, Paris.

- Feinberg, A. (2017) "This Is The Daily Stormer's Playbook" [online], *Huffington Post* 14 December, [https://www.huffingtonpost.co.uk/entry/daily-stormer-nazi-style-guide\\_n\\_5a2ece19e4b0ce3b344492f2?guccounter=1](https://www.huffingtonpost.co.uk/entry/daily-stormer-nazi-style-guide_n_5a2ece19e4b0ce3b344492f2?guccounter=1)
- Gawryjolek, J., C. Di Marco, C.D., and Harris, R.A. (2009) "An annotation tool for automatically detecting rhetorical figures – System demonstration", Proceedings of the IJCAI-09 Workshop on Computational Models of Natural Argument.
- Gentzkow, M., and Shapiro, J.M. (2010) Ideological Segregation Online and Offline (NBER Working Paper No. 15916), NBER, Cambridge, Mass.
- Ghosh, D., Fabbri, A.R., and Muresan, S. (2017) "The Role of Conversation Context for Sarcasm Detection Online Interactions" [online], <https://arxiv.org/pdf/1707.06226.pdf>
- Gonçalves, P., et al. 2013, "Comparing and Combining Sentiment Analysis Methods", *Proceedings of the first ACM conference on Online social networks*, pp. 27-38.
- Harris, R. and Di Marco, C.D. (2009) "Constructing a rhetorical figuration ontology", *Persuasive Technology and Digital Behaviour Intervention Symposium*, pp. 47–52.
- Hazarika, D., Poria, S., Gorantla, S., Cambria, E., Zimmermann, R., and Mihalcea, R. (2018) "CASCADE: Contextual Sarcasm Detection in Online Discussion Forums", Proceedings of the 27th International Conference on Computational Linguistics, pp. 1837–1848.
- Herman, E.S. (1988) Manufacturing Consent: The Political Economy of the Mass Media, Pantheon Books, New York.
- Holliday, R. (2013) Trust Me I'm Lying: Confessions of a Media Manipulator, Penguin, London.
- Isle, B., and Smith, T. 2018, "Real world examples suggest a path to automated mitigation of disinformation, ", *IEEE International Conference on Big Data*, pp.4395 – 4399.
- Haselton, M.G., Nettle, D. and Murray, D.R., 2015. The evolution of cognitive bias. *The handbook of evolutionary psychology*, pp.1-20.
- Hamouda, H. 2018, "Trustworthiness of citizen journalists videos from the perspective of archival science", *IEEE International Conference on Big Data*, pp. 4390 – 4394.
- Horne, B. & Adali, S. 2017) "This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News", *Association for the Advancement of Artificial Intelligence*.
- Hu, N., Bose, I., Koh, N.S. and Liu, L., 2012. Manipulation of online reviews: An analysis of ratings, readability, and sentiments. *Decision Support Systems*, 52(3), pp.674-684.
- Hutto, C.I. and Gilbert, E. 2014, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text", *ICWSM*, The AAAI Press, pp. 216-25.
- Joshi, A., Bhattachryya, and Carman, M.J. (2017) "Automatic Sarcasm Detection: A Survey", ACM Computing Surveys (CSUR), Volume 50 Issue 5, Article No. 73.
- Kelly, A.R., Abbott, N.A., and Harris, R.A. (2010) "Toward an ontology of rhetorical figures", Proceedings of the 28th ACM International Conference on Design of Communication - SIGDOC '10, pp. 123-130.
- Lee C.E. and Kent, JL. 2017 October 30, "Facebook says Russian-backed election content reached 126 million Americans", NBC Nightly News. Available: <https://www.nbcnews.com/news/us-news/russian-backed-election-content-reached-126-million-americans-facebook-says-n815791>.
- Marwick, A. and Lewis, R. (2017) *Media Manipulation and Disinformation Online*, Data & Society, New York.
- Meza, S. 2017 November 2, "'Fake news' named word of the year", Newsweek. Available: <http://www.newsweek.com/fake-news-word-year-collins-dictionary-699740>.
- Mladenovic, M. (2016) "Ontology-based Recognition of Rhetorical Figures", Infotheca, Vol. 16, No. 1–2, pp. 24–47.
- Moosa, T. (2017) "Neo-Nazis are trying to spread hatred through comedy. This isn't funny" [online], *The Guardian* 19 August, <https://www.theguardian.com/commentisfree/2017/dec/19/neo-nazis-hatred-comedy-racist-daily-stormer>
- Mourao, R.R. and Robertson, C.T. 2019, "Fake news as discursive integration: An analysis of sites that publish false, misleading, hyperpartisan and sensational information, *Journalism Studies*. Available: <https://doi.org/10.1080/1461670X.2019.1566871>.
- Newman, J., Pennebaker, J.W., Berry, D.S., and Richards, J. M. (2003) "Lying Words: Predicting Deception From Linguistic Styles", *Personality And Social Psychology Bulletin*, Vol. 29, No. 5, pp. 665-75.
- Pomerantsev, P., and Weiss, M. 2014 "The menace of unreality: How the Kremlin weaponizes information, culture and money". *The Interpreter*, vol. 22.
- Public Broadcasting System 2019, "Propaganda techniques", Classroom materials. Available: <https://www-tc.pbs.org/weta/reportingamericaatwar/teachers/pdf/propaganda.pdf>
- Psywarrior website: Available <http://www.psywarrior.com/>
- Ribeiro et al. 2016, "SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods", EPJ Data Science (2016) 5:23 DOI10.1140/epjds/s13688-016-0085-
- Rosenblatt, S. 2018 March 26, "Exacerbating our fake news problem: Chatbots". Available: <https://www.the-parallax.com/2018/03/26/fake-news-chatbots/>.
- Sample, C., Justice, C and Darraj, E. 2018, "A Model for Evaluating Fake News", *Proceedings from NATO CyCon US Conference*, Washington, DC. Available: <https://www.hSDL.org/?view&did=818918>.
- Song, Yi, Deane, P. and Beigman Klebanov, B. 2017. "Toward the Automated Scoring of Written Arguments: Developing an Innovative Approach for Annotation". *ETS Research Report Series*, doi: 10.1002/ets2.12138.
- Tausczik, Y. R. and Pennebaker, J. W. (2010) 'The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods', *Journal of Language and Social Psychology*, 29(1), pp. 24–54. doi: [10.1177/0261927X09351676](https://doi.org/10.1177/0261927X09351676).

***Char Sample, Keith Scott and Emily Darraj***

- Taylor, P.M. (2003) *Munitions of the mind*, Manchester University Press, Manchester.
- Varpio, L. 2018, "Using rhetorical appeals to credibility, logic, and emotions to increase your persuasiveness", *Perspective Medical Education* 7: pp.207–210, doi: 10.1007/s40037-018-0420-2
- Verrall, N. & Mason, D. 2019, "The Taming of the Shrewd", *The RUSI Journal*, doi: 10.1080/03071847.2018.1445169.
- Vosoughi, S., Roy, D., Aral, S. (2018) "The spread of true and false news online", *Science*, Vol. 359, Issue 6380, pp. 1146-1151.
- Webster dictionary: Available: <http://Meriam-webster.com>.
- Younger, A. 03 December 2018. Available: <https://www.gov.uk/government/speeches/mi6-c-speech-on-fourth-generation-espionage>
- Zannetou, S., Caulfield, T., Blackburn, J., De Cristofaro, E., Sirivianos, M., Stringhini, G., and Suarez-Tangil, G. (2018) "On the Origins of Memes by Means of Fringe Web Communities", IMC '18 Proceedings of the Internet Measurement Conference 2018, pp.188-202.
- Zhoud, L., Burgoon, J.K., Nunamaker, J.F., and Twitchell, D. (2004) "Automating Linguistics-Based Cues for Detecting Deception in Text-based Asynchronous Computer-Mediated Communication", *Group Decision and Negotiation* 13, pp. 81–106.
- Zuiderveen Borgesius, F. J. & Trilling, D. & Möller, J. & Bodó, B. & de Vreese, C. H. & Helberger, N. (2016). Should we worry about filter bubbles?. *Internet Policy Review*, 5(1), pp.1-16.

# A Novel Intrusion Detection System Architecture for Internet of Things Networks

Leonel Santos<sup>1,2,3</sup>, Ramiro Gonçalves<sup>3,4</sup> and Carlos Rabadão<sup>1,2</sup>

<sup>1</sup>School of Technology and Management, Polytechnic Institute of Leiria, Portugal

<sup>2</sup>Computer Science and Communication Research Centre, Leiria, Portugal

<sup>3</sup>University of trás-os-montes e alto douro, Vila Real, Portugal

<sup>4</sup>INESC TEC (formerly INESC Porto), Porto, Portugal

[leonel.santos@ipleiria.pt](mailto:leonel.santos@ipleiria.pt)

[ramiro@utad.pt](mailto:ramiro@utad.pt)

[carlos.rabadao@ipleiria.pt](mailto:carlos.rabadao@ipleiria.pt)

**Abstract:** The Internet of Things (IoT) is rapidly becoming ubiquitous and applied in different domains such as human health, building automation, industrial control and environmental monitoring, introducing new security and privacy challenges. Thus, the security of data, devices and communications of IoT networks are a concern due to the sensitivity of the data used, legal and privacy issues, and the diversity of devices and protocols used. In addition, traditional security mechanisms cannot always be feasible and adequate because of the number, heterogeneity, and resource limitations of IoT devices. In this work, we are concerned with the design of an Intrusion Detection System (IDS) to protect IoT networks from external and internal threats in real time. To do this, after studying the various traditional IDS solutions, as well as new IDS proposals designed specifically for IoT networks, we conclude that there are still several improvements to be made to this type of 2nd line defense mechanism. The design proposed will consider the specific architecture of an IoT network, the scalability and heterogeneity of this type of environment, the minimization of the use of resources, and the maximization of the efficiency in the detection of intrusions. To do so, we consider the various detection methods available and the various types of attacks to which this type of network is exposed. The proposed IDS is network-based and relies on a hybrid architecture (centralized / distributed). As methods of detection, the signature / anomaly-based methods will be used simultaneously. Finally, it is emphasized that this proposal does not require the modification of the IoT software, nor does it influence the performance of the applications in the IoT devices.

**Keywords:** internet of things, intrusion detection, cybersecurity, network security, network attacks

---

## 1. Introduction

Internet of Things (IoT) is a new paradigm that enables many novel applications in different domains such as home automation, industrial process, human health and environmental monitoring and the security and privacy preserving of data collected is a constant concern.

With the rapid growth of the number of devices being connected to the internet, network administration responsibilities like devices management, traffic monitoring and security is a concern that must be on everyone's mind (Santos et al, 2019). According to various reports, by 2020 there will be fifty billion devices connected to the internet, and each person will own around seven devices. Also, with the rapid growth of the IoT that is entering the mainstream, more and more data is being collected and transmitted over networks and internet, making them more exposable to attacks, increasing the risk of cyber security attacks (Atzori et al, 2010).

Because IoT landscape is heterogenous, fragmented and not supportive of interoperability it is very hard to design specific security mechanism. Some solutions for enhancing IoT security have been developed and include methods for providing data confidentiality and authentication, access control within the IoT network, and trust and privacy among users and things. However, even with those mechanisms, IoT networks still vulnerable to attacks. Then, the development of more security tools specific to IoT are required and systems like Intrusion Detection System (IDS) could be used to address that necessity (Santos et al, 2018).

Despite the maturity of IDS technology for traditional networks, current solutions are inadequate for IoT because they will not be flexible enough against the complex and heterogenous IoT ecosystem (Alaba et al, 2017). Characteristics of IoT devices, such as low-cost design, resource constraints, large scale, heterogeneity, preference of functions over security, higher privacy requirements, and harder trust management, make very difficult to apply many traditional security solutions. The challenges in the design of security of IoT systems, such as the network architecture, scalability, heterogeneous devices and communications, integration with the

physical world, resource constraints, privacy, the large scale, trust management, and less preparation for security, explain and enforce the need for development of IDS for IoT (Sha et al, 2018).

Considering that the development of IDS for IoT systems is a new important challenge for the researches in this field, the research team decided to undergo an extensive analysis on the existing literature related to the development of IDS solutions specifically developed for IoT systems and just a few surveys related with this topic were found (Zarpelão et al, 2017; Santos et al, 2018). Despite the existence of some IDS proposals developed specifically for IoT systems, Zarpelão et al. (2017) and Santos et al. (2018) conclude that is necessary to improve them to address more features such as: cross-layer detection, given that the majority are only able to detect attacks in one layer of IoT architecture; ability to defend against a more wide range of attacks, because most of the revised works address only internal routing attacks; providing variety in detection technics used, as most apply signature-based detection methods; address more IoT technologies, since the majority of available works address 6LoWPAN networks; and to ensure the security of IDS alert and management messages, because most doesn't protect IDS internal communications. Considering the mentioned, our goal with this work is design an IDS solution that could be applied on IoT systems, in order to address the challenges and weakness identified.

In this paper, we propose a new IDS architecture specifically designed for IoT networks. This architecture follows a hybrid placement strategy for IDS modules. That is, it involves both centralized and distributed components. Regarding detection method, our proposed IDS system follows a hybrid detection method approach. That is, it applies both signature and anomaly-based techniques. The collection of network data is made using probes in all IoT layers. Collected data is forward to the IDS modules. The two IDS modules, local and remote, apply different detection methods due to their different computational capabilities. IDS local module runs on the border router and use signature-based method to analyse captured data, and the IDS remote module runs on a cloud-based system and use anomaly-based method. One of the main advantages of this approach is that no software modification of IoT devices and application is required. Furthermore, all the IDS modules are connected via secure communication channels in order to avoid security and privacy issues.

The rest of this paper is organized as follows. Section 2 introduces some relevant terms regarding IoT and IoT security. Section 3 introduces relevant terms regarding IDS and present the most important IDS solutions for IoT. In section 4, we describe our proposed design, including the architecture and its main components. Finally, in section 5, we present a brief set of conclusions complemented with future work considerations.

## **2. Internet of things**

The IoT has won attention recently because of the expansion of appliances connected to the Internet (Atzori et al, 2010). IoT simply means the interconnection of vast heterogeneous network frameworks and systems in different patterns of communication, such as human-to-human, human-to-thing, or thing-to-thing (Al-Fuqaha et al, 2015). Moreover, the IoT is a realm where physical items are consistently integrated to form an information network with the specific end goal of providing advanced and smart services to users (Botta et al, 2016).

Typically, the architecture of IoT is divided into three basic layers (Al-Fuqaha et al, 2015):

- Perception layer: is implemented as the bottom layer in IoT architecture. Its main objectives are to connect things into IoT network, and to measure, collect, and process the state information associated with these things via deployed smart devices.
- Network layer: is implemented as the middle layer in IoT architecture. The network layer is used to receive the processed information provided by perception layer and determine the routes to transmit the data and information to the IoT hub, devices, and applications via integrated networks.
- Application layer: is implemented as the top layer in IoT architecture. The application layer receives the data transmitted from network layer and uses the data to provide required services or operations.

Zegzhda et al (2015) propose three useful topologies: point to point, star and mesh. The latter is decentralized, and preferable for IoT systems.

Different alliances, consortiums, special interest groups, and standard development organizations have proposed a considerable amount of communication technologies for IoT, what may carry a big challenge for end-to-end security in IoT applications (Meddeb et al, 2016).

Most popular technologies for IoT include infrastructure protocols like IEEE 802.15.4, BLE, WirelessHART, Z-Wave, LoRaWAN, 6LoWPAN, DTLS and RPL, and application protocols like CoAP and MQTT.

In cyber security, the Confidentiality – Integrity – Availability (CIA) triad is well known. Just a few papers however relate CIA back to IoT. Besides CIA, Lin et al (2017) adds more features to be addressed like Identification and Authentication, Privacy and Trust. Alaba et al (2017) outline some security challenges in each layer of IoT architecture presenting common vulnerabilities and attacks.

Perception layer: the security challenges in this layer focus on forging collected data and destroying perception devices by the following attacks: node capture; malicious code injection; false data injection; replay or freshness; cryptoanalysis and side channel; eavesdropping and interference; and sleep deprivation.

Network layer: the security challenges focus in the impact of the availability of network resources through the next attacks: denial of service (DoS); spoofing; sinkhole; wormhole; man-in-the-middle (MITM); routing information; sybil; and unauthorized access.

Application layer: challenges in this layer focus on software attacks like phishing attack and malicious virus/worm and malicious scripts.

### **3. Intrusion detection in IoT**

#### **3.1 Intrusion detection system**

According to Halme et al (1995), an Intrusion Detection System is an anti-intrusion approach that aims to discriminate intrusion attempts and intrusion preparation from normal system usage.

A typical IDS is composed of sensors, an analysis engine, and a reporting system. Sensors are positioned at different network places or hosts and their main task is to collect data. The data collected are sent to the analysis engine, which is responsible to examine the collected data and detect intrusions. If an intrusion is detected by analysis engine, the reporting system generates an alert to network administrator.

IDSs can be classified as Host-based IDS (HIDS) and Network-based IDS (NIDS). HIDS is attached to a device/host and monitors malicious activities occurring within the system. NIDS connects to one or more network segments and monitors network traffic for malicious activities (Santos et al, 2018).

IDS placement strategy approaches can be classified as distributed, centralized, and hybrid (Zarpelão et al, 2017).

In centralised approach the entire IDS is placed in a central, either remote or host-based location. In the distributed strategy, the IDS nodes are places among multiple or all nodes within the network and responsibility is divided amongst them. The hybrid placement strategy combines any strategy of the above. Often found in tandem with multiple detection types.

IDS detection methods approaches can be classified as signature-based, anomaly-based, specification based, and hybrid (Zarpelão et al, 2017).

In signature-based approaches, IDSs detect attacks when system or network behaviour matches an attack signature stored in the IDS internal databases. If any system or network activity matches with stored patterns/signatures, then an alert will be triggered. This approach is accurate and very effective at detecting known threats, and their mechanism is easy to understand. However, this approach is ineffective to detect new attacks and variants of known attacks, because a matching signature for these attacks is still unknown.

Anomaly-based IDSs compare the activities of a system at an instant against a normal behaviour profile and generates the alert whenever a deviation from normal behaviour exceeds a threshold. This approach is efficient to detect new attacks, however, anything that does not match to a normal behaviour is considered an intrusion and learning the entire scope of the normal behaviour is not a simple task.

Specification is a set of rules and thresholds that define the expected behaviour for network components such as nodes, protocols, and routing tables. Specification-based approaches detect intrusions when network behaviour deviates from specification definitions. Therefore, specification-based detection has the same purpose of anomaly-based detection: identifying deviations from normal behaviour. However, there is one important difference between these methods: in specification-based approaches, a human expert should manually define the rules of each specification. Manually defined specifications usually provide lower false positive rates in comparison with the anomaly-based detection.

Hybrid approaches will involve any combination of the above, whereby issues related to the efficacy of one technique is mitigated by the strengths of another.

### **3.2 IDS solutions for IoT**

IDS as security measures have been considered by researchers for protecting networks with heterogeneous IoT devices. However, IDS in traditional networks have different requirements than IDS for IoT. Therefore, adapting traditional IDS approaches in IoT environments is not an easy and straightforward task. Features such as limited computation power of smart devices, different network structures, and various developed protocols of IoT devices introduce new challenges that should be addressed by an IDS for IoT (Sha et al, 2018). Over the recent years, several review articles have been published on IDSs for technologies related to IoT such as mobile ad hoc networks, wireless sensor networks, cloud computing, and cyber-physical systems. Although these articles primarily focus on the design of IDSs for several IoT related elements, only one presented by Zarpelão et al (2017) and Santos et al (2018) provide a study of IDS techniques specific for the IoT paradigm. Below we briefly describe the most important recent IDS solutions for IoT.

Kalis is one of the first developed IDS that aims at protecting IoT devices irrespective of the IoT protocol or application used (Midi et al, 2017). Kalis is a network-based, hybrid signature/anomaly-based, centralized, online IDS. The selected detection strategy depends on specific network characteristics. Furthermore, Kalis obtains knowledge from modules installed in the network, and attempts to prevent DoS attacks based on the current network topology, traffic analysis, and mobility information. Kalis can support new protocol standards and allows knowledge sharing between the nodes for better detection. Experimental results show that Kalis has better detection performance than traditional IDS.

Another remarkable work in the field is the SVELTE (Raza et al, 2013). This is a signature- and anomaly-based IDS, developed to protect IoT devices from routing attacks based on the IPv6 Routing Protocol for Low Power and Lossy Networks (RPL). Some of the considered attacks include altering information, sinkhole forwarding, and selective forwarding. SVELTE follows a hybrid module placement approach in which a centralised module, called 6LoWPAN Border Router (6BR), performs heavy calculations and several resource constrained modules are responsible for monitoring tasks. The 6BR has three components. The first one is the 6LoWPAN Mapper which recreates the network based on the information obtained from IoT nodes. The second component is the IDS one which analyses information and detects intrusion. The last one is a mini firewall which stops malicious traffic from entering the network. The first and third components are embedded into the IoT nodes.

Despite good progress in developing IDS for IoT, current solutions have several limitations. Kalis, for example, requires installation of specialised detection modules for detecting each type of attack. This could create a complex network and could lead in poor detection performance. Moreover, it uses IEEE 802.11 as communication technology. This means that interference between the smart sensors and Kalis nodes is possible if they are in close proximity. SVELTE has also some limitations as it requires the modification of sensors' software. This, however, would be very inconvenient for networks with large numbers of sensors, which is a typical case in many IoT application domains. All in all, a new technologically improved solution is needed to protect IoT networks from a wide range of possible attacks. The aforementioned limitations have been taken into account when designing our proposed IDS solution.

## **4. Proposed architecture**

### **4.1 System and security requirements**

Since applications for the IoT areas as diverse as smart cities, human health, surveillance, and smart energy will require fundamental assurances from the infrastructure in terms of security, and this certainly includes the

ability of detecting attacks and intrusions against the availability of IoT devices and of the application itself, in a timely fashion, we can note that currently there is a lack of systems and mechanisms designed and capable to detect attacks and intrusions in IoT networks and, with this goal in mind, we target the following design requirements for an IoT IDS:

- Cross-layer detection: The IDS should be able to detect attacks at all layers of IoT architecture.
- External and internal intrusion detection: ability to detect intrusions originated at external hosts and at internal devices.
- Near real time detection: ability to detect intrusions within a reasonable time frame.
- Scalability: capability to accommodate a growing amount of analysis that result of the expansion of IoT network in terms of size or traffic.
- Interoperability and extensibility: support different communication mediums and protocols, intrusion detection mechanisms, among others. Must be extensible to new standards, technologies, and intrusion types as they emerge.
- Reconfigurability: support different intrusion policies, during the lifetime of the IoT system.
- No software changes: minimization of the IDS footprint in the use of resources and software modification in IoT devices.
- No performance overhead: the IDS should not impact the performance of the IoT devices' applications.
- Protection of IDS communications: ensure security of communications between probes and IDS components.

The intrusion architecture we describe next is designed to materialize the previous requirements. We start by analysing the system architecture for intrusion detection in IoT networks and its components, and next we discuss how processing and intrusion detection is implemented.

## **4.2 Architecture and components**

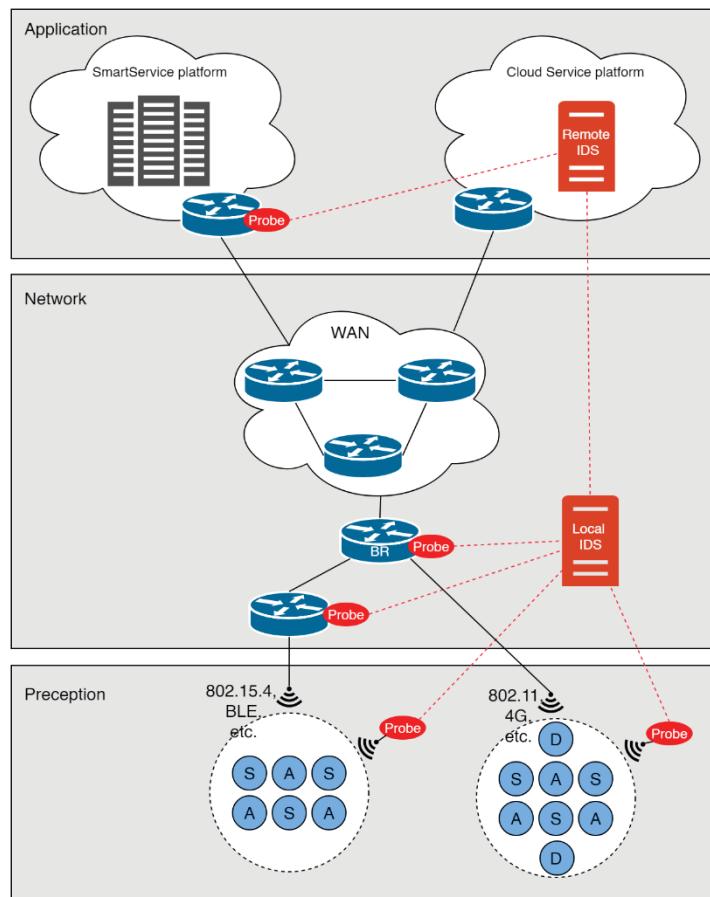
The design of intrusion detection mechanisms for IoT networks must consider a careful usage of the resources available in constrained sensing and actuating platforms while, on the other hand, it may adapt and benefit from the availability of more resources in other class of devices, in particular in border routers or base stations or in cloud-based systems.

Considering the system requirements presented, the only solution that guarantee the careful usage of resources and the non-modification of software of IoT devices is the use of a hybrid approach to intrusion detection architecture. We suggest that a combination of centralized intrusion analysis and a distributed data collection is the best solution for this type of environments. Figure 1 illustrates the system architecture proposed for intrusion detection in IoT networks.

In order to minimize problems related to the wider detection capabilities of the various types of threats, both internal and external, and to possibilities the coverage of more IoT technologies, the monitoring and capture of IoT communications will be done in the 3 layers of the IoT architecture (perception, network, and application), as illustrated in Figure 1, ensuring the detection of internal or external intrusions that may occur in any layer of an IoT application. The monitoring and capturing of these communications will be done using probes that will be placed in all layers, in order to provide a holistic view in the detection of intrusions in IoT networks.

In the proposed solution, the probes are responsible for monitoring and capturing the communications of the segment of the IoT network where they are installed. In the perception layer, the probes will act as dedicated devices and will be monitoring and capturing the internal communications of IoT networks. These IoT communications are made by various IoT devices such as sensors, actuators, among other types of devices, and can use various technologies and protocols (802.15.4, BLE, 802.11, etc.). The probes used in the perception layer support several technologies and protocols, which guarantees total coverage to all technology types and protocols through minimal hardware adaptions. At the network layer, probes are monitoring and capturing incoming and outgoing communications passing through the interfaces of routing devices, such as routers or border routers. In these cases, probes functions may be supported and made available by the operating systems of the routing devices, so it is not necessary to install any probe through additional software. The probe that will

be monitoring and capturing the access communications to the platform of IoT smart service in the application layer, will be installed in the physical or virtual router of access to the platform of IoT smart service and aims to capture the connections made to this platform.



**Figure 1:** IDS System architecture for IoT networks

Once captured, all IoT communications will be forwarded to the IDS devices. Probes must be able to communicate almost in real time with the local and remote IDS devices. Captures made by existing probes in the perception and network layers will be forwarded to the local IDS device, and the captures made by the probe located in the application layer will forward captures to the remote IDS device.

The local IDS device will be responsible for aggregating and analysing the communications captured by the probes that are in the perception and network layer of the IoT network. The data aggregated by this device will be forwarded to the remote IDS device.

The primary role of the remote IDS device is to receive and store the aggregate communications forwarded by the local IDS device, as well as receive the communications captured by the probe located in the application layer. All stored communications in remote IDS will be scanned on this device.

In terms of communications exchanged between the components of the proposed IDS system, they must be preferably made using hard-wired links, and be transmitted using an encrypted channel, in order to guarantee the security and privacy of messages. In addition, communications between perception and network layers probes and local IDS device must use a dedicated VLAN for this type of communications, adding another layer of security to internal communications of the proposed IDS system.

### 4.3 Detection methodology

Our proposal to analysis of IoT communications with the purpose of detecting intrusions, is a solution that will use a hybrid detection method. Since signature-based and anomaly-based techniques have different features and usages, both will be used simultaneously to accumulate the advantages of both. The analysis of the

communications will be done in a centralized way and using devices external to the IoT application. These external devices have the computational resources that these tasks and analysis techniques require, releasing IoT devices from these tasks, assuring scalability to the proposed solution.

Therefore, in order to detect intrusions, it is proposed that the analysis of IoT communications should be done in a local device and in a remote device installed in a cloud-based system. Regarding the analysis of captured IoT communications, both local IDS and remote IDS devices have responsibilities in analysing IoT communications considering their expected computational capabilities.

Because signature-based detection method is not so computationally demanding, it will be used in the local IDS device. As mentioned, in cases where the border router has computational capacity, it can host the local IDS process as an integrated module. If the border router does not have those capabilities, the local IDS process may be installed on a dedicated device for that purpose. To perform the analysis of the communications monitored and captured through the probes, the local IDS device will have a knowledge database with signatures of known attacks and threats, as well as traffic considered normal for the IoT application in operation. This database can be updated whenever there are changes to the IoT application in operation or whenever new signatures of attacks and threats appear. If an attack or threat is detected, an IDS alert message will be generated and sent to remote IDS device in order to be registered in the alert IDS database.

The anomaly-based detection method will be performed in the remote IDS device. This IDS device must be installed in a cloud-based system, providing the execution of more computationally demanding analysis tasks and providing scalable resources. To perform the analysis of the monitored and captured communications through all the probes, the remote IDS device relies on the use of more complex and demanding intrusion detection techniques such as techniques based on statistical methods, machine learning, or other methods that may be used or even new techniques that may arise in this area. These techniques can access to a knowledge database with information that will allow them to identify traffic considered normal and abnormal for the IoT application in operation, among others. This database can be updated whenever there are changes to the IoT application in operation or whenever new attacks and threats arise. If an attack or threat is detected, an IDS alert message will be generated and logged in the alert IDS database located in this device.

#### **4.4 System analysis**

The ability to monitor and capture IoT communications passively and securely export them for remote analysis on IDS devices offers great advantages. In the previous section, it was described that analysis of the data as it is sent by existing probes in the 3 layers greatly increases the intrusion detection capacity in the perception and above layers. Forwarding captured IoT communications in raw format allows access to all the information necessary for the process of analysis done on IDS devices.

Many types of attacks and threats can be identified with a higher success rate, while minimizing any additional load or complexity in resource utilization in the IoT network itself. Nonetheless, this creates a more secure IoT network, although with a higher cost in implementation due to the introduction of additional hardware and the need for secure links between probes and IDS devices. However, as the number of IoT application deployments increases across a variety of scenarios such as health, industry, etc., this additional cost can counterbalance the impact of a potential security and privacy breach. A final issue that can occur with such a passive system is, of course, the subversion of probes. In order to avoid issues of subversion of probes, this possible problem can be mitigated by the existence of several probes monitoring the same location, allowing a comparison of the captured data.

### **5. Conclusion**

As attention in the IoT raises, its application will include more data sensitive projects. As such, guaranteeing its security is a priority. With preventive procedures hard to be implemented due to inherent architectural constraints, solutions must turn to second line methods of defence. We studied IDS as one such defence and concluded that despite the diversity of existing solutions available; none are able to defend against all types of attacks (from the perception layer up) due to their architectural implementation.

In this paper, we proposed a new approach for an IDS for IoT networks. We presented a list of system and security requirements, the design of a high-level IDS architecture, and described its main components and

functions. In terms of placement strategy, the proposed approach had a hybrid architecture and involves both centralized intrusion detection and distributed data collection modules for detecting real time intrusions originating from external networks as well as from internal compromised nodes. In order to take advantages from various detection methods, our proposed system use a hybrid detection method with both signature and anomaly-based. This involves the use of various types of probes to collect data and securely transport it to the IDS devices, located locally and remotely in a cloud-based system. This solution ensures that there aren't software changes and performance overhead on IoT devices and application, as well as, scalability and extensibility due to the use of cloud-based systems to analyse the collected data and to maintain information about known or new types of intrusions.

In our future work we plan to implement and test the proposed system approach. The system will be tested on a variety of IoT applications to examine the effect of monitoring multiple different protocols in varied environments, upon the data collection and analysis process.

The approach presented in this work could be considered as a relatively simple one, although more development and research will be needed to confirm it is optimal in a wide diversity of IoT applications.

## Acknowledgements

This work was supported by Portuguese national funds through the FCT - Foundation for Science and Technology, I.P., under the project UID/CEC/04524/2019.

## References

- Alaba, F. A., Othman, M., Hashem, I. A. T., & Alotaibi, F. (2017). Internet of Things security: A survey. *Journal of Network and Computer Applications*, 88, 10-28.
- Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., & Ayyash, M. (2015). Internet of things: A survey on enabling technologies, protocols, and applications. *IEEE communications surveys & tutorials*, 17(4), 2347-2376.
- Atzori, L., Iera, A., & Morabito, G. (2010). The internet of things: A survey. *Computer networks*, 54(15), 2787-2805.
- Botta, A., De Donato, W., Persico, V., & Pescapé, A. (2016). Integration of cloud computing and internet of things: a survey. *Future generation computer systems*, 56, 684-700.
- Halme, L. R., Bauer R. K. (1995), AINT misbehaving - A taxonomy of anti-intrusion techniques, Proc. of 18th NIST-NCSC National Information Systems Security Conference (pp. 163-172).
- Lin, J., Yu, W., Zhang, N., Yang, X., Zhang, H., & Zhao, W. (2017). A survey on internet of things: Architecture, enabling technologies, security and privacy, and applications. *IEEE Internet of Things Journal*, 4(5), 1125-1142.
- Meddeb, A. (2016). Internet of things standards: who stands out from the crowd?. *IEEE Communications Magazine*, 54(7), 40-47.
- Midi, D., Rullo, A., Muderikar, A., & Bertino, E. (2017, June). Kalis—A system for knowledge-driven adaptable intrusion detection for the Internet of Things. In 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS) (pp. 656-666). IEEE.
- Raza, S., Wallgren, L., & Voigt, T. (2013). SVELTE: Real-time intrusion detection in the Internet of Things. *Ad hoc networks*, 11(8), 2661-2674.
- Santos, L., Rabadão, C., & Gonçalves, R. (2018, June). Intrusion detection systems in Internet of Things: A literature review. In 2018 13th Iberian Conference on Information Systems and Technologies (CISTI) (pp. 1-7). IEEE.
- Santos, L., Rabadão, C., & Gonçalves, R. (2019, April). Flow Monitoring System for IoT Networks. In 7th World Conference on Information Systems and Technologies. Springer.
- Sha, K., Wei, W., Yang, T. A., Wang, Z., & Shi, W. (2018). On security challenges and open issues in Internet of Things. *Future Generation Computer Systems*, 83, 326-337.
- Zarpelao, B. B., Miani, R. S., Kawakani, C. T., & de Alvarenga, S. C. (2017). A survey of intrusion detection in Internet of Things. *Journal of Network and Computer Applications*, 84, 25-37.
- Zeghdha, D., & Stepanova, T. (2015, July). Achieving Internet of Things security via providing topological sustainability. In 2015 Science and Information Conference (SAI) (pp. 269-276). IEEE.

# An Analysis of Security Features on Web Browsers

Siddhartha Sengupta and Joon Park

School of Information Studies, Syracuse University, New York, USA

[jspark@syr.edu](mailto:jspark@syr.edu)

**Abstract:** Today, while Web browsers are widely used to access the Internet resources, there are constant attempts to find exploits and vulnerabilities that can compromise the user's security and privacy. In this paper, we analyze the security features provided by the most popular browsers, including Chrome, Internet Explorer, Firefox, Safari, and Opera, with our hands-on experiments and discuss how these features can enhance the security and privacy for users. We identify and analyze the security features in the browsers, considering the concept of sandbox isolation, plug-ins, password storage/synchronization, Google Safe Browsing, private browsing, cookie management, and other security/privacy related features. We believe our work will help users to choose a right browser for their purposes as well as protect their sensitive data and personal information on the Web. Furthermore, our findings and comparison will help the browser developers to improve the quality and security of their future products and services.

**Keywords:** internet security, browser vulnerability, privacy protection, web browser security

---

## 1. Introduction

A Web browser is an application which helps users to browse through the Internet. Web browsers provide a graphical user interface (GUI) that enables the user to read and navigate through Web pages. The user's browser communicates with the Web servers over the Internet and access information specified on the webpages, using the URLs to visit.

The reliance of today's society on the Internet cannot be understated. We should come rely on it for answers to our questions and curiosity. Access to all that information is done through Web browsers. With the ever-growing popularity of the Internet, from a potential hacker's standpoint, it means there are more potential victims (Ye et. al 2005; Herzberg and Jbara 2008; Antonatos et. al 2008). Typically, browsers are not configured to provide much security in their default state. This can cause many kinds of malware to be installed in the host machines. For instance, password storage via Web browsers is a serious concern as of now. The entire list of passwords can be accessible very easily by knowing the victim's login credential on the device or having administrator access to the victim's machine.

Therefore, we analyze the security features provided by the most popular browsers with our hands-on experiments and discuss how these features can enhance the security and privacy for users. We identify the security features in the browsers, analyse the hardening measures that the developers have undertaken to protect the potential victims from multiple kinds of attack vectors, and compare them with their peers.

## 2. Related work

Awang et. al (2014) discuss the list vulnerabilities reported primarily for Firefox and IE. They brought forward reasons why the vulnerabilities would exist in a browser and then they listed out the specific vulnerabilities for Firefox and Internet Explorer. They tested how secure a browser is using an online tool such as Qualys Browser Check (Qualys BrowserCheck 2016). The online tool just checks if the versions of the installed plug-ins are updated or not and the more comprehensive security check is processed through installation of their plug-ins. The test was conducted on a sample of 10 PC's and they used the online tool on the installed Web browsers to find how many are insecure. As a result, they reported that the Web browsers were insecure mostly due to the plug-ins not being updated. A list of browser-specific (Firefox and IE only) countermeasures was provided, which included cookie management and clearing private data from browsers. They lacked research into the other popular browsers (e.g., Safari and Google Chrome) and did not compare those browsers' security features to Firefox and IE. Also, in terms of tools used, the only metric they used to check how secure a browser is and whether they had updated plug-ins or not, they failed to consider the architecture of a browser and the how the underlying operating system is used to make it more secure. Their countermeasures that were provided explained only what to do and not what would those actions actually achieve in making a browser more secure.

Drake et. al (2011) proposed to compare the browsers based on anti-exploitation techniques. They believe that the anti-exploitation software can eliminate the severity of a single vulnerability or an entire class of exploits.

The browser with the best anti-exploitation technique would be the most resistant to an attack hence more secure compared to its peers. The comparison was done between Google Chrome, IE, and Firefox with the base host a Windows 7 (32bit) machine. They show an extensive analysis of how a chosen anti-exploitation technique is implemented in the Web browsers. This analysis was highly quantitative in nature, but their comparison did not take into consideration the types of users that use browsers. Since their work published several years ago, browser security has come a long way and most of the features discussed are common these days. Differentiating between browsers can be done in a more refined way by including the kinds of users that would use browsers. We have considered browsers usage statistics to correlate between them and the number of vulnerabilities reported, hence defining that a vulnerability in a browser with more usage would be more critical in nature, not just consider the severity of the vulnerability as chosen by a data source.

### **3. Common security vulnerabilities in existing web browsers**

While Web browsers being widely used to access the Internet resources, there are constant attempts to find exploits and vulnerabilities that compromise the user's security and privacy. Typically, browsers are not configured to provide much security in their default state. This can cause many kinds of malware to be installed in the host machines. For instance, when an unsuspecting host clicks on malicious links on the browser, it can trigger a malware on the victim's machine. This malware may have the capability of changing the DNS server in the network settings of the machine, leading to users being redirected to infected websites, alter user searches, replace ads, block anti-virus software, and promote fake security products. Other possible means of attacks could happen through the buffer overflow and code execution. These usually occur because of poor input-validation implemented in webpages and can be quite harmful. These exploits are usually more concerned with Web application developers, as they need to ensure that there is proper input-validation conducted in the fields of the webpage that takes in input.

Web pages these days load multiple types of resources, including texts, images, videos, cookies, Flash, and JavaScript. Loading webpages by using advanced technologies such as JavaScript is unavoidable these days due to its popularity but on the other hand it does pose a risk for end users by enabling malicious actors to deliver scripts over the Web and run them on client computers. Since its release, however, there have been several JavaScript security issues that have gained widespread attention (Sun and Ryu 2017). There are two measures that can be taken to contain this JavaScript security risk. The first measure is sandboxing Web contents so that they can only access certain resources and perform specific tasks. The second measure is implementing the same origin policy, which prevents scripts from one site from accessing data that is used by scripts from other sites.

A cross-site scripting is another kind of attack in which attackers try to inject malicious scripts to perform malicious actions through websites. In cross-site scripting, malicious code executes in the browser and affects users by either installing malware or stealing session cookies/tokens in order to gain privileges on the user's machine. The rate at which vulnerabilities in browsers are found is increasing very fast and developers are constantly releasing security patches to fix those vulnerabilities. Therefore, it becomes extremely crucial to keep the browser timely updated with its latest version. However, this is not always possible, especially when people keep running the same session for a long time (e.g., say, weeks).

### **4. Analysis on security features in existing web browsers**

In this section, we identify the security features we analyzed in popular Web browsers and compare the results based on our experiments. Table 1 summarizes our results with the comparison.

#### **Security Protection with Sandbox:**

Sandboxing in Web browsers is a feature pioneered by Google and was first implemented in Chrome (Laperdrix et. al 2016). It places a restriction on the rendering engine and prevents one instance from interacting with the local disk and the operating system. This stops any arbitrary code or remote access attempts from writing on the local machine. Therefore, it can limit the remote access to the local machine if the Web page has any malicious code. Sandboxing is implemented into browsers by separating the browser processes from the local machine, placing restrictions on its read/write permissions. These restrictions create a logical sandbox for the browser process. In this way even if the browser is compromised, the user's local machine still has an added layer of security to protect any read/write attempts on the local memory. The browser processes can be divided into two parts: the rendering engine and the browser kernel. The rendering engine is responsible primarily for

parsing data and decoding images, and the browser kernel is how the rendering engine instances interact with the local machine's operating system. The rendering engine is restricted and kept in a sandbox while the browser kernel is outside the sandbox.

All the major browsers that we have evaluated employ sandboxing in one way or the other. Google Chrome employs sandboxing both for Web contents and plug-ins. This creates a separate process for each tab created in Google Chrome, which prevents tabs from affecting another opened tab and access to the local machine's hard drive. Firefox has not implemented sandboxing in that way yet, but it does have sandboxing enabled for NPAPI plug-ins (Chrome 2018) like Adobe and Flash. IE11 calls its version of sandboxing as Enhanced Protection mode (EPM) that prevents any malicious code contained in the website from modifying your system files. Safari also employs sandboxing to the extent that Web content that is referred in Web pages from different sources are sandboxed. Additionally, Google Chrome uses the underlying Windows token system to stop processes from interacting with the local disk (Shinder 2010).

**Table 1:** Comparison of security features in popular web browsers

Security Features	Chrome	IE11	Firefox	Safari	Opera
Sandboxed	Yes	Yes	Yes	Yes	Yes
Default plugins	Yes	Yes	No	Yes	Yes
Flash	Yes	Yes	Yes	Yes	Yes
NPAPI (less secure)	No	No	Yes	Yes	Yes
PPAPI (more secure)	Yes	No	No	No	Yes
ActiveX controls	No	Yes	No	No	No
Checks if up-to-date	Yes	Yes	Yes	Yes	Yes
Password accessible through browser	Yes	No	Yes	Yes	Yes
Password saved/location	User Profile Folder/Chrome	Saved in the Windows Registry	User Profile Folder/Firefox	Keychain	User Profile Folder/Opera
Master password	None	None	Yes	None	None
Password syncing	Google Account	Microsoft Account	Setup Firefox Account	iCloud	Opera account
Google Safe browsing	Used	Not used	Used	Used	Not used

#### **Security Concerns with Plug-ins:**

Plug-ins are provided by 3rd parties, which provide additional software features of varying types that users can install into a browser. They enable users to have extra features that the browser inherently does not have. For example, a Flash plug-in enables users to view Flash video in the browser, and without it a webpage would prompt the user to install Flash in order to view the contents. Plug-ins would be the weakest point in terms of security in a browser. Most of the vulnerabilities in a browser are due to interactions between the browser and the plug-ins. Especially, out-of-date plug-ins are a serious security risk, as most of the plug-in developers release

critical updates after vulnerabilities are found. Therefore, it's of utmost importance that plug-ins should be kept up-to date.

Traditionally, most of the plug-ins had to be installed by the users themselves, but these days browsers come with certain default plug-ins installed during the browser installation. This greatly mitigates the risk of an attack through plug-ins because most of these plug-ins can be trusted to be safe. On the other hand, 3rd party plug-ins cannot be simply trusted because an untrustworthy plug-in can be a major risk for the user to use in the browser (Binu et. al 2016). Default plug-ins, being validated plug-ins chosen by the developers which are deemed safe and vital enough to be installed along with the browser, help in increasing user accessibility while also greatly reducing the attack surface of the browser. In addition to the default plug-ins, users still need to install additional plug-ins in many cases. This increases the risk of a user downloading a malicious plug-in. Browsers now provide a way where the user can enable the browser to check the status of all the installed plug-ins and keep them updated on its own. This reduces the hassle of checking for outdated plug-ins by a user. Safari and Firefox have this feature enabled in the browser's default state where they do not allow out-of-date plug-ins to run and force the user to update it. Chrome on the other hand uses its strong suite of default plug-ins which are updated automatically with each version update of Chrome.

#### **Security Features with Plug-in APIs:**

Netscape Plug-in Application Programming Interface (NPAPI) is an API that allows plug-ins to be developed by browsers (Chrome 2018). NPAPI plug-ins work by declaring the type of contents it can handle, when the browser encounters such a content type, it loads the appropriate plug-in. This architecture was first used in by Netscape browsers in 1995 and then adopted by other browsers thereafter. These plug-ins can be of two types: in-process and out-of-process (Plugin Architecture 2019). In-process plug-ins run within a rendering process and are faster when compared to out-of-process plug-ins. Out-of-process plug-ins require access to the local filesystem and network in order to work, hence they pose more of a risk as they cannot be sandboxed. Google developed a new plug-in API called the Pepper plug-in API (PPAPI) from the scratch in order to address performance issues with NPAPI (Chromium 2019). Apart from that, PPAPI also offered easier portability over different platforms. PPAPI type plug-ins are now deployed in Chrome and Opera. Firefox and Safari still use NPAPI plug-ins. Mozilla do not plan to implement PPAPI at all in Firefox according to their website.

Microsoft Internet Explorer dropped NPAPI in support of its own technology developed to mirror functions and enhancements, provided by plug-ins and called it ActiveX. These controls work only for Microsoft applications and are used by default in Internet Explorer. ActiveX controls are plagued with vulnerabilities because of lesser restrictions placed on them. IE11 does have the ability to prevent older version of ActiveX controls from running call the ActiveX filtering, but the technology is still not secure enough and with better substitutes available, avoiding ActiveX controls reduces the attack surface of the user substantially. In general, PPAPI would be considered more secure than NPAPI type plug-ins. For instance, Google Chrome and Opera enabling the development of PPAPI plug-ins help in providing more security features to users, although Opera still supports NPAPI versions in certain plug-ins.

#### **Password Storage:**

Browsers provide options to store passwords of users so that users need not enter passwords every time they open a new browsing session. Passwords are usually stored in the local user's folder (e.g., C:\Users\username). This file is an SQLite database where all the saved passwords for a particular browser. It contains the login details such as usernames, URLs, and passwords. Each browser installed in the user's machine has such a file. Passwords located in those folders are encrypted and a decryption tool is required to access them. Google Chrome, Firefox, and Opera utilize this method to store passwords on the local machine. Another way to access passwords is through the browsers directly so that the Web browsers show the saved passwords in their settings menu. These passwords can be accessed either directly or by putting in the user's device password. They are usually encrypted using the device login password. In order to view these passwords through the browser, the attacker either needs access to the victim's device login password. Alternatively, the attacker can reset the victim's login credentials so that the stored passwords can be accessed using the new credentials. In Windows machines, password storage can be managed by using the Windows registry, which offers a more secure way of storing passwords than using the user's folder. Internet Explorer employs this method of storing passwords.

The usage of master passwords can provide an extra layer of authentication to retrieve or save a password in the Web browser. Master passwords are used to encrypt the saved passwords stored by the browser on top of the protection provided by encrypting stored passwords, using separate credentials. For instance, Firefox asks the user to set an independent password which isn't tied to any account (e.g., Google or Microsoft). This reduces the risk of the master password getting hacked as it would be tougher to get access to this password rather than a password which is connected to an account. Master passwords, through these added layers of encryption, have been known to be more secure (Blocki 2016; Gaw 2006) as long as the master password is securely managed. However, once the master password is compromised, the entire list of the user's passwords would be accessible by the attacker.

**Password Synchronization:**

Passwords and other stored data can be synchronized via an existing account such as Google or Microsoft. This service offers more flexibility to the user as he or she can log in to the account by using different machines and browsers with all the bookmarks and saved passwords loaded from its base repository. However, having a centralized location for all your passwords is not advisable and poses a risk of losing it all in one go. If one of those synchronized password lists has been compromised, the entire password list of the user can be accessible by the attacker. IE does not provide the option to sync stored passwords over different devices. This reduces the possibility of passwords being hacked, but this reduces the usability as users who are accustomed to using saved passwords would need to save all their passwords again when using a new machine. Google Chrome offers the possibility to sync all the accounts and passwords over multiple devices through the Google account, which creates a possible risk as all the passwords being tied to one single Google account. Firefox offers the option allowing syncing by creating a Firefox account, but it also offers an added layer of security by allowing users the option to use a master password, which makes it more secure and usable by a user than IE.

**Google Safe Browsing:**

Google Safe Browsing is an industry standard used to prevent malware and phishing websites from causing harm. Chrome, Firefox, and Safari are the browsers that employ Google Safe Browsing service, while both IE11 and Opera do not. Google Safe Browsing enables applications to check URLs against Google's constantly updated lists of suspected phishing, malware, and unwanted software pages. Google Safe Browsing provides the following features: 1) Warn users before they click on links in your site that may lead to malware-infected pages, 2) Prevent users from posting links to known phishing pages from your site, and 3) Check a list of pages against Google's lists of suspected phishing, malware, and unwanted software page. 32-bit hash prefixes of the URL are used to compare with a list of known phishing websites, using the Safe Browsing Lookup API. This would be an answer for privacy concerns over user's URL's being recorded by the service. However, Google Safe Browsing, despite being the most marketed anti-phishing tool by Google and the other browsers, is still vulnerable and can be worked around (Nirmal et. al 2015; Danchev 2013). Hackers are known to use contents cloaking to work around Google Safe Browsing to detect a malicious website. Therefore, simply relying on just on Google Safe Browsing isn't enough and should not be considered a one-stop solution. There are more extensions available in all the browsers which provide added protection for users apart from Google Safe Browsing.

**Private Browsing:**

This is a feature provided by most browsers in slightly different ways with the basic rationale behind it being able to surf the Web without storing the user's surfing activity data (e.g., cookies, browsing history, download history, passwords, etc.) on the local machine, such as Chrome's Incognito, IE's InPrivate, and Safari's Private browsing modes.

When a private browsing mode starts, the browsers open a new window with none of the previous tabs opening again, thus disallowing the chance of the tabs in the different modes to interact with each other. Incognito mode offered by Chrome does not send pre-existing cookies when an incognito window is opened which allows it to trump over other privacy modes. None of the browsers allow new cookies, download and browsing history, and passwords to persist after a private browsing session is closed. However, all the downloaded files are not removed from the local disk. There are still some risks involved while private browsing because of how certain browsers allow plug-ins to be enabled while in that mode (Liou et. al 2016; Gao et. al 2014). Cookies do get deleted after a session ends in all browsers. In our experiments, IE 11 was the only browser to store passwords

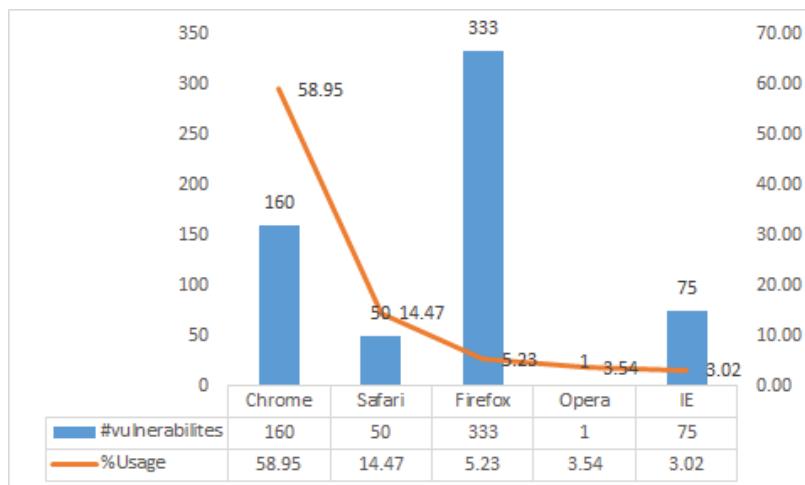
during the session, all other browsers did not allow the option to save passwords. Form data was not saved in any browser during private browsing. It is important to understand the difference between Saved and Recorded. “Not Recorded” means that the data is not recorded by the browser at all and “Not Saved” means that the data recorded but gets discarded after the session ends.

#### **Cookie Management:**

Cookies are information that helps Web servers identify users, customize Web pages based on the user’s previous visiting history, and/or continue the previous session without repeating the same procedures. After each session, session cookies are deleted. However, persistent cookies can be stored in the user’s machine as a text file until they expire. Later, when the user is visiting the same or other Web servers associated, the corresponding cookies are automatically sent to the Web servers by the browser. Then, the Web servers can retrieve the information about the user from the cookies and use it for their purposes. By configuring the Web server, technically, cookies can collect any information about the user’s browsing activities such as name, email address, home address, phone number, session ID, and any other information. Previously, some Web servers stored even the user’s credentials or other sensitive information in their cookies, but now days they realize the security concerns with cookies and stop storing sensitive information in the cookies. However, there are still many cases that cookies collect users’ personal information and cause privacy issues (Park and Sandhu 2000). For instance, cookies can be used to track the user’s Internet usage. Therefore, it is important to understand how the browsers handle cookies properly, considering the security and privacy concerns. In our experiments, while comparing the default settings of how cookies are set for the different browsers, we found that only in Chrome and Opera, first party cookies (i.e., cookies generated by the visiting Web server) were set and no 3rd party cookies (i.e., cookies generated by the non-visiting Web servers) were set. Firefox and Internet Explorer allowed all cookies to be set. Safari in its default settings did not allow 3rd party cookies to be set.

## **5. Discussion**

Figure 1 shows the number of vulnerabilities reported in each browser and its average usage (CVE 2018; StatCounter 2018). As we can see in the figure, in general, except for Firefox, the number of vulnerabilities is tied to the browser usage percentage because more vulnerabilities would be found and reported by more users. Chrome being the most popular browser has predictably higher number of vulnerabilities. Internet Explorer which is the least popular browser still seems to have quite a number of vulnerabilities even though its usage has reduced a lot over the years. The most anomalous behaviour is shown by Firefox’s stats, which has a low percentage usage but has a high number of vulnerabilities, possibly due to flexibility that Firefox offers for using extensions and plug-ins which are mostly deprecated in other browsers and over time naturally.



**Figure 1:** Number of vulnerabilities vs. usage for each browser

Chrome security features include anti-phishing and anti-malware that’s built-in with the browser, which helps prevent users from being exposed to fraudulent websites. Its built-in plug-in system removes the necessity to install 3rd party plug-ins, reducing the risks of an attack. The default plug-ins and Sandboxing are the most influential security features on our list as they prevent a large array of attack vectors without burdening the user in any way. Chrome still hasn’t added the concept of master password and all the saved and synced data is tied

to the Google account which is a concern as its acts as a single point of entry for accessing all that sensitive information.

Firefox offers users to install various extensions and plug-ins that can be used to enhance the security through the browser. Firefox inherently employs considerable features such as Google Safe Browsing and its move to completely shift to HTTPS browsing in order provide more secure connection between the client machine and the Web server is seen as a move towards the right direction. The original philosophy behind the development of Firefox was to enhance user experience through a community platform and this shows with the amount of security enhancements available on the market, but we have to keep in mind there is always the risk of fraudulent plug-ins. Once a user is able to distinguish between the real and fake ones, one can fully utilize the options available and reduce the risks of an attack by using security plug-ins as the same time. Otherwise, Firefox's idea of being plug-in dependent may increase the chance of someone being vulnerable to an attack.

Opera offers several security features, but it hasn't been tested enough by the users trying to find vulnerabilities in it due to its low popularity. Opera is still allowing NPAPI plug-ins to run, which is cause of concern as those types of plug-ins are less secure. The lack of public visibility for Opera browser makes it lesser of a target for attackers and researchers alike and this makes it less tested and scrutinized compared to Chrome and IE.

Safari's plan to stick with NPAPI plug-ins and not migrate to PPAPI makes it less secure, although its password storing method and plug-in control make up for that. This browser also lacks any master password functionality, while allowing passwords to accessed directly through the browser. Its popularity is low primarily because of Chrome being available in Mac machines and Safari does not support a Windows version anymore.

Internet Explorer method of saving passwords locally on the machine using Windows registry is by far the safest method compared to how Firefox and Chrome save passwords locally. Furthermore, the fact that Internet Explorer does not provide a list of saved passwords from the browser provides one less method through which a password can be hacked. However, IE11 still uses ActiveX controls, which is a concern and more secure substitutes available in the market. Overall, IE11 lags behind other browsers in terms of security features.

## **6. Conclusion**

With the ever-growing popularity of the Internet, from a potential hacker's standpoint, it means there are more potential victims. Therefore, in this paper, we have analyzed the security features provided by the most popular browsers, including Chrome, Internet Explorer, Firefox, Safari, and Opera, with our hands-on experiments and discussed how these features can enhance the security and privacy for users. We have identified and analyzed the security features in the browsers, considering the concept of sandbox isolation, plug-ins, password storage/synchronization, Google Safe Browsing, private browsing, cookie management, and other security/privacy related features. We believe our work will help users to choose a right browser for their purposes as well as protect their sensitive data and personal information on the Web. Furthermore, our findings and comparison will help the browser developers to improve the quality and security of their future products and services.

## **References**

- Antonatos, S., Akritidis, P., Lam, V., and Anagnostakis, K. (2008). *Puppetnets: Misusing Web Browsers as a Distributed Attack Infrastructure*. ACM Transactions on Information and System Security (TISSEC), Volume 12 (2), Article 12, 38 pages.
- Awang, N., Ahmad, A., and Ahmad, S. (2014). *Preventing Web Browser from Cyber Attack*. GSTF Journal on Computing (JoC), Volume 2 (1).
- Binu, P., Sreekutty, H., and Sreekutty, V. (2016). *Security Plugin for Mozilla Which Integrates Cryptography and Steganography Features*. IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Chennai, 2016, pp. 1-6.
- Blocki, J. and Sridhar, A. (2016). *Client-CASH: Protecting Master Passwords against Offline Attacks*. In Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security (ASIA CCS '16). ACM, New York, NY, USA, pp.165-176.
- Chrome (2018). *NPAPI Plugins*. [online]. Available at: <https://developer.chrome.com/extensions/npapi>
- Chromium (2019). *Getting Started: Background and Basics - The Chromium Projects*. [online]. Available at: <http://www.chromium.org/nativeclient/getting-started/getting-started-background-and-basics#TOC-Pepper-Plugin-API-PPAPI>
- CVE Details (2018). [online]. Available at: <https://www.cvedetails.com/>

- Danchev, D. (2013). *Comparative Review: Opera Leads in Browser Anti-phishing Protection*. [online]. Available at: <http://www.zdnet.com/article/comparative-review-opera-leads-in-browser-anti-phishing-protection/>
- Drake, J., Mehta, P., Miller, C., Moyer, S., Smith, R., and Valasek, C. (2011). *Browser Security Comparison: A Quantitative Approach*. Technical Report, Accuvant Lab.
- Gao, X., Yang, Y., Fu, H., Lindqvist, J., and Wang, Y. (2014). *Private Browsing: An Inquiry on Usability and Privacy Protection*. In Proceedings of the 13th Workshop on Privacy in the Electronic Society (WPES '14). ACM, New York, NY, USA, pp.97-106.
- Gaw, S. and Felten, E. (2006). *Password Management Strategies for Online Accounts*. In Proceedings of the second symposium on Usable privacy and security (SOUPS '06). ACM, New York, NY, USA, pp.44-55.
- Herzberg, A., and Jbara, A. (2008). *Security and Identification Indicators for Browsers against Spoofing and Phishing Attacks*. ACM Transaction on Internet Technology (TOIT), Volume 8 (4), Article 16, 36 pages.
- Laperdrix, P., Rudametkin, W., and Baudry, B. (2016). *Beauty and the Beast: Diverting Modern Web Browsers to Build Unique Browser Fingerprints*. IEEE Symposium on Security and Privacy, San Jose, CA, 2016, pp. 878-894.
- Liou, J., Logapriyan, M., Lai, T., Pareja, D., and Sewell, S. (2016). *A Study of the Internet Privacy in Private Browsing Mode*. In Proceedings of the 3rd Multidisciplinary International Social Networks Conference on Social Informatics, ACM, New York, NY, USA, Article 3, 7 pages.
- Nirmal, K., Janet, B., and Kumar, R. (2015). *Phishing - the Threat That Still Exists*. International Conference on Computing and Communications Technologies (ICCCT).
- Park, J. and Sandhu, R. (2000). *Secure Cookies on the Web*. IEEE Internet Computing, Volume 4(4), pp.36-44.
- Plugin Architecture - The Chromium Projects (2019). [online]. Available at: <http://www.chromium.org/developers/design-documents/plugin-architecture>
- StatCounter Global Stats - Browser, OS, Search Engine Including Mobile Usage Share (2018). [online]. Available at: <http://gs.statcounter.com/>
- Sun, K. and Ryu, K. (2017). *Analysis of JavaScript Programs: Challenges and Research Trends*. ACM Computing Surveys, Volume 50 (4) Article 59, 34 pages.
- Qualys BrowserCheck (2016). *Qualys BrowserCheck*. [online]. Availale at: [https://browsercheck.qualys.com/?scan\\_type=js](https://browsercheck.qualys.com/?scan_type=js)
- Shinder, D. (2010). *Better Security Through Sandboxing*. [online]. Available at: [http://www.windowsecurity.com/articles-tutorials/windows\\_os\\_security/Better-Security-through-Sandboxing.html](http://www.windowsecurity.com/articles-tutorials/windows_os_security/Better-Security-through-Sandboxing.html)
- Ye, Z., Smith, S., and Anthony, D. (2005). *Trusted Paths for Browsers*. ACM Transactions on Information and System Security (TISSEC), Volume 8 (2), pp.153-186.

# Alert Correlation Using Diamond Model for Cyber Threat Intelligence

Youngsup Shin<sup>1</sup>, Changwan Lim<sup>1</sup>, Mookyu Park<sup>2</sup>, Sungyoung Cho<sup>1</sup>, Insung Han<sup>1</sup>, Haengrok Oh<sup>1</sup> and Kyungho Lee<sup>2</sup>

<sup>1</sup>Agency for Defense Development, Seoul, Republic of Korea

<sup>2</sup>Korea University, Seoul, Republic of Korea

[shinyoungsup@add.re.kr](mailto:shinyoungsup@add.re.kr)

[goat1009@naver.com](mailto:goat1009@naver.com)

[ctupmk@korea.ac.kr](mailto:ctupmk@korea.ac.kr)

[sycho@add.re.kr](mailto:sycho@add.re.kr)

[insung.han@add.re.kr](mailto:insung.han@add.re.kr)

[haengrok@add.re.kr](mailto:haengrok@add.re.kr)

[kevinlee@korea.ac.kr](mailto:kevinlee@korea.ac.kr)

**Abstract:** Information security has gathered great attention leading to a variety of network sensors and Intrusion Detection Systems (IDS), generating numerous threat events. Large number of threat events are difficult to be managed by passive countermeasures of security manpower, lacking in prompt situation recognition and preemptive responses. Therefore, automated cyber threat analysis techniques based on big data or machine learning are required for effective security control and threat analysis. Also, correlation analysis with Cyber Threat Intelligence (CTI) that occurred in the past enables high level analysis of intrusion intent as well as preemptive response. Therefore, approach to autonomous alert correlation methods using machine learning algorithm such as Bayesian network, Hidden Markov Model (HMM), Support Vector Machine (SVM) and neural network are studied for threat analysis recently. In this paper, we propose analysis method for alerts generated by Security Information and Event Management system (SIEM) in two parts. In the first part, we apply Bayesian network to generate attack scenario and infer intent of the intrusion. We define the causality of alerts generated by SIEMs through alert correlation algorithm based on Bayesian network. This facilitates identification of the invasion pathway as well as prediction of the next attack. In the second part, we employed Diamond model to the generated attack scenarios for threat analysis using CTI. Rather than merely plotting an attack graph, it applies the Diamond model to the attack graph based on the cyber kill chain, allowing the analyst to identify more information at a glance. In order to apply Diamond model, we expanded features of each attack such as asset information of the system or vulnerabilities. Accordingly, each attack scenario could be reconstructed as CTI format and compared with threats occurred in the past. Therefore, we demonstrated the feasibility of successful identification and rapid response of the overall attack situation.

---

**Keywords:** cyber threat intelligence, cyber threat analysis, alert correlation, diamond model

## 1. Introduction

The number and influence of cyber attacks has increased rapidly, and the sophistication of attack methods has also greatly improved. As a result, system administrators or security officers are responding to cyber attacks through an intrusion detection system (IDS). In fact, as network traffic increases, IDS generate a large number of alerts. This hinders system administrator or security officers from responding to cyber attacks.

In this paper, we propose a model for generating a chain of past alerts through an alert correlation algorithm and predicting anticipated attacks that expressing them in cyber threat intelligence. System administrators or security officers have great difficulty in analysing large amounts of threat alerts and need to automate them. In order to understand the entire attack scenario, it is necessary to analyse not only specific alerts but also other related alerts. The alert correlation algorithm is used to link alerts associated with specific alerts to form a chain. This allows to determine the attack path so far. Also, in order to predict further attacks, it infers the alerts that are likely to come after a certain alert. It is possible to find the correlation analysis data between accumulated alerts through machine learning. These past alerts and anticipated attacks are organized into chains, also coupled with models that can aid in analysis. The information in the alert is presented in an easy-to-analyse format that allows system administrators or security officers to figure out how to prevent attacks.

## **2. Trends of cyber threat analysis**

### **2.1 Cyber kill chain**

The concept of Cyber kill chain was first introduced by Lockheed Martin (2014), USA. Recent attacks attempt very sophisticated attacks through a planned attack campaign called APT. In order to protect important assets, the attacker's intentions and strategies must be clearly understood. Therefore, we introduced the concept of cyber kill chain in order to identify each stage of an attack campaign and establish a good response strategy. This model identifies what the attacker must accomplish in order to achieve the goal. It distinguishes the necessary steps to make the attack successful, and if one of these steps is not achieved, the attack will fail.

It consists of 7 steps, reconnaissance, weaponization, delivery, exploitation, installation, command & control and actions on objectives. Reconnaissance is a planning stage in which an attacker conducts research to understand the goal, and it is very difficult to detect it from the defender's point of view, but the attacker's intentions can be seen. Weaponization combines malware and exploits into payloads, usually through automated tools. Defenders cannot detect weaponization, but they can infer them by analysing malware artifacts. Delivery is the start of operations by delivering malware to the target, and it is the most important step to prevent operation against the defender. Exploitation is the step in which an attacker evaluates an access privilege using vulnerability and requires custom capability to prevent zero-day exploits. During the installation phase, an attacker typically installs a backdoor in the target environment to maintain access for a long period of time. The defender must detect installation activity logs through endpoint analysis. Command & Control is a step in which malware opens a command channel to allow an attacker to manipulate the target remotely and is the opportunity for the defender to stop the operation last. In the Actions on objectives phase, the attacker produces the desired result. The defender must identify the damage as soon as possible.

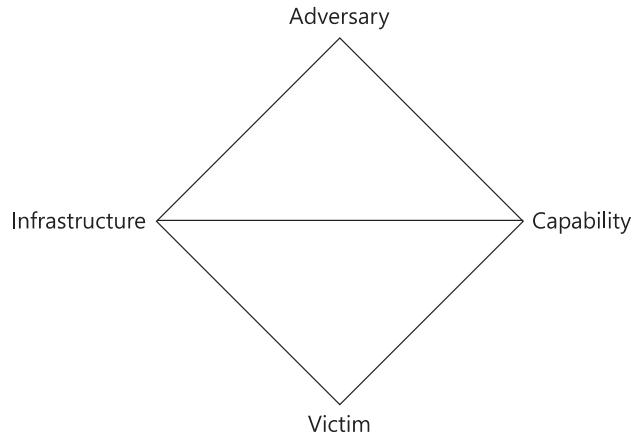
### **2.2 Diamond model**

In Caltagirone, Pendergast & Betz (2013)'s paper, it presents a diamond model which is a comprehensive model including activity documentation, synthesis, correlation analysis, etc. This model ensures the repeatability and accuracy of the analysis results and enables automatic correlation analysis between various events. Confidence can be used to classify events for organize attack campaigns, and It can be used for attack forecasting to construct and execute mitigation strategies. It is possible to integrate correlated intelligence in real time to have network defence capability. It integrates external context and cyber threat intelligence that applies to real-time intrusion detection, represents each attack event effectively.

Express the core features and meta-features of the attack event as a diamond shape. It sorts the steps performed by the attacker in order to achieve the goal and expresses the flow of the attacker's behaviour. In order to fully understand malicious behaviour, it shows not only a single event but also a progressive progression. Effectively presents a list of features to be represented in the event. Use the information that is currently acquired to document and identify what features are empty. This identifies the knowledge gaps and identifies additional tasks.

The core features of the diamond model are adversary, capability, infrastructure, and victim. An adversary is an individual or organization that attacks a computer system or network by intention or necessity. Most of the events are blank at the time of discovery and there are adversary operators that perform actual intrusion activities and adversary customers who actually benefit from attacks. Capability is the tool or skill that the adversary used in the event. It is important information to make an attack mitigation decision or to predict the potential response of adversary. Capability capacity is a vulnerability that can be exploited by an attacking ability, and adversary arsenal is a complete set of adversary's attack capabilities. Infrastructure refers to physical and logical communication structures such as IP address, domain name, email address, and USB device. The first type of infrastructure is that adversary controls or possesses physical proximity. Second, it is a controlled infrastructure by intermediary that can be considered as an adverse by the victim to hide the source or property of the activity, such as zombie host, malicious code load server, or hacked email account. The last core feature, victim, can be an individual, organization, e-mail address, IP address, etc., where the attacking abilities of the adversary are used. Victim persona is the person or organization that owns the assets under attack and non-technical analysis is required. A victim asset is an attack surface that can be a network, system, email address,

IP address, etc. and requires technical analysis such as vulnerability analysis. This is shown in Figure 1(Caltagirone et al. 2013)



**Figure 1:** The core features of diamond model of intrusion analysis

The diamond model has six meta-features: timestamp, phase, result, direction, methodology, and resource. The timestamp is the date and time when the event occurred, and functions as important role in grouping attack patterns, pattern analysis and malicious behaviour. Since malicious activity consists of two or more events, it represents step of each event in phase. Result is used to determine the probability of success of an adversary's performance against a particular attacking ability or some victim. A wide range of views on attack intentions can be drawn, the success of attacks or the degree of infringement in terms of Confidentiality, integrity and Availability are examples. Direction is used for attack mitigation options, detection location selection. The methodology represents a general classification of attack activity. Finally, resource refers to one or more external resources required to successfully perform an attack. Other features include data sources, detection methods, and detection signatures, which can be used as important information for future use.

Using the cyber kill chain model with the attack graph, the attack path of the attacker is divided into stages and the attack route is shown. This paper presents an analytical methodology focusing on specific features and presents new perspective that can be applied to other analytical models (Caltagirone et al. 2013).

### 2.3 Cyber threat intelligence

In Gartner's document, cyber threat intelligence is evidence-based knowledge, including context, mechanisms, indicators, implications and actionable advice, about an existing or emerging menace or hazard to assets that can be used to inform decisions regarding the subject's response to that menace or hazard (McMillan 2013). Information about threat agents and how they are used, and how to prevent or detect attacks, embodies policies and actions. Information such as attacker, victim, infrastructure, motivation, and attack method for a specific attack scenario can help the victim identify the situation of the attack more easily. By using cyber threat intelligence correctly, it helps defenders to detect attacks by providing behavioural indicators at every stage of the attack (Shackelford 2015).

### 3. Alert correlation methods

IDS and other security sensors have been popularized, and administrators have been handling many alerts generated from various sensors. It is difficult to identify the sequence of attacks because of large number of false alerts. Also, since it provides only piecemeal information, it is difficult to understand contextually. To solve these problems, researches on Alert correlation have been actively started. Alert correlation algorithms receive large amounts of alerts from heterogeneous systems, consolidate information and eliminate false alerts. It also detects attack patterns and provides a high level of intelligence. Alert correlation algorithms have been studied in various forms. Algorithms can be broadly divided into three categories: Similarity-based, Knowledge-based, and Statistical-based (Mirheidari et al. 2013). A detailed description of each type follows.

### 3.1 Similarity-based algorithms

Similarity-based algorithm is a method of comparing the similarity between alert and alert or alert and meta-alert (alert cluster). If the alerts indicate similarity, they are merged (Dwivedi & Tripathi 2015). If there is no similarity, a new type of meta-alert is generated. It also compares the similarity of graphs generated through alerts (Wu et al. 2010). The advantage of a similarity-based algorithm is that it can be clustered without having to define the attack type correctly. This algorithm can define correlation through simple rules, hierarchical rule, and machine learning.

### 3.2 Knowledge-based algorithms

This algorithm defines attacks based on prior knowledge. Prior knowledge refers to the network and connection information within the system (Kang & Na 2012), or a general scenario of an attack (Yang et al. 2009). Knowledge-based algorithms can associate a series of alerts generated by entering network and asset information or attack scenarios in advance. However, a knowledge-based algorithm has a disadvantage that it cannot deal with new types of attacks.

### 3.3 Statistical-based algorithms

Just as it is called, the basic concept of a statistical-based algorithm is to classify attacks through statistical analysis. Statistics based algorithms do not require prior knowledge of attack scenarios because statistical analysis and learning of attacks are used to derive attack steps. However, it is not possible to analyse the dependent and abnormal alerts on the learned domain. Statistical-based algorithms are used to detect repeated alert patterns or to predict causal relationships between alerts. In particular, research has been conducted to analyse causal relationships using Bayesian network (Ren et al. 2010, Ahmadian Ramaki & Rasoolzadegan 2016), Neural Network (Zhu & Ghorbani 2006), and Hidden Markov Model (Holgado et al. 2017, Sendi et al. 2012).

In this paper, we correlate alerts using Bayesian network-based statistical-based algorithms.

## 4. Proposed method

The proposed method can be expressed in two steps: alert correlation and applying diamond model.

### 4.1 Alert correlation for attack scenario reconstruction

The Alert Correlation step follows the strategy of statistically analysing alerts generated from security devices and extract the causal relationship between alerts. The basic idea of alert correlation in this study follows method discussed by Ren et al. (2010). We calculated the probabilistic relationship between alert and its components based on the Bayesian network. The applied alert correlation method performs statistical analysis on the alerts collected in the background to extract the main features that affect the causality of the alert type pair. Also, create correlation table and a relevance table are that define the causality and its constraint through the analysed information. The generated tables are used to reconstruct the path of the alert that occurs in real time and to predict the attack afterwards. A description of each procedure is provided below.

#### 4.1.1 Correlation probability calculation and feature selection

The basic concept of this part is to understand causal relationships between alert types that occur within the same time window. Defining the relationship between the intrusion alert means defining the similarity of the alert feature. Since all features of an alert do not have the same effect on causality, choosing a critical feature is an important step. To select a feature that has a large effect on causality, we need to calculate the correlation probability between the alerts based on the features. When the alert types A and B both have the feature  $f_j$  of the same domain, the equation for calculating the correlation probability of A and B is as follows.

$$P(B|A[f_j] = \text{dom}(B, f_j)) = P(B \wedge A[f_j] = \text{dom}(B, f_j)) / P(A[f_j] = \text{dom}(B, f_j))$$

Next, we compare the calculated correlation probability with the probability of occurrence alert type B  $P(B)$ , and preset threshold  $\theta$  to identify important features. Alert features can be divided into the following four categories.

- $P(B|A[f_j] = \text{dom}(B, f_j)) = P(B)$ :  $f_j$  does not affect the probability of occurrence of alert type B.
- $P(B|A[f_j] = \text{dom}(B, f_j)) < P(B)$ : alert type A reduces the probability of occurrence of alert type b if feature a has feature  $f_j$ . Which means,  $f_j$  is a feature with a negative effect.
- $P(B) < P(B|A[f_j] = \text{dom}(B, f_j)) < \theta$ : alert Type increases the probability of occurrence of alert type B if a has feature  $f_j$ . Which means,  $f_j$  is a feature with a positive effect.
- $P(B|A[f_j] = \text{dom}(B, f_j)) > \theta$ :  $f_j$  is a feature that has a critical effect. If alert type a has feature  $f_j$ , it increases probability of occurrence of alert type B critically.

Therefore, we can obtain the maximum set of  $f_j$  satisfying  $P(B|A[f_j] = \text{dom}(B, f_j)) > \theta$  using greedy algorithm. This set is the best set of features that have a critical impact on the correlation probability of alert type pairs. This feature set is used as a constraint on causality.

#### 4.1.2 Creating correlation table and relevance table

Through the method above, we can calculate the critical feature set between the alert type pair and the correlation probability. The next step is to construct two tables through the calculated values. First, the correlation table shows the causal information extracted from the alert types. The role of the correlation table is to represent all pairs of alerts that can be part of an attack scenario. This table contains the correlation set between two alert types and the feature set which used as the relevance constraint. This table is used as a component when extracting scenarios of attacks that occurred.

Next, the relevance table stores information about the alert type. This table contains the probability of occurrence of each alert type and the relevant alert types. This table is used to determine causality for two consecutive alerts.

#### 4.1.3 Attack scenario reconstruction

Attack scenario reconstruction is performed using correlation table and relevance table. If a malicious attack occurs in real time, refer to the relevance table. Search whether the alert type corresponding to the attack and the relevant alert type occurred in the past. If there was a relevant alert in the past, check the constraints by referring to the correlation table and correlate the two alerts. Repeat this method to reconstruct the attack path of the attack. It can also predict future threats from current alerts. The attack scenario can be extracted by combining the reconstructed past attack path and the predicted future threats. Reconstructed attack scenario is shown in Figure2 (Ren et al. 2010).



**Figure 2:** An example of reconstructed attack scenario from LLDS 1.0 data set. Each circle represents attack state and arrow represents causal relationship and its correlation probability

## 4.2 Applying diamond model

Alerts generated from IDS or SIEM do not include all features required by the Diamond model. This is called knowledge gap in the diamond model. The first step is to reduce the knowledge gap between the actual alert and the Diamond model. For this, we use the cyber threat taxonomy. There are some researches under way to classify threats using cyber threat taxonomy (Simmons et al. 2014). Our method uses taxonomy to reclassify alerts and provide additional information. The information about each node of attack chain generated through alert correlation is classified into core features and meta-features of diamond model. The attack chain, represented simply by the name of the node, evolves into a node of the diamond model, which allows the analyst to see more information at a time. The information processed by the diamond model through the alert correlation is used as cyber threat intelligence and provides much more information than simply viewing the alert.

#### 4.2.1 Threat reclassify

First, we reclassify threats (generated alerts) according to the cyber threat taxonomy. Each of alerts can be distinguished by the kill chain phases, tactics, techniques and procedures (TTPs). The taxonomy is constructed with refer to MITRE CAPEC and ATT&CK. The example of cyber threat taxonomy is shown in Table 1.

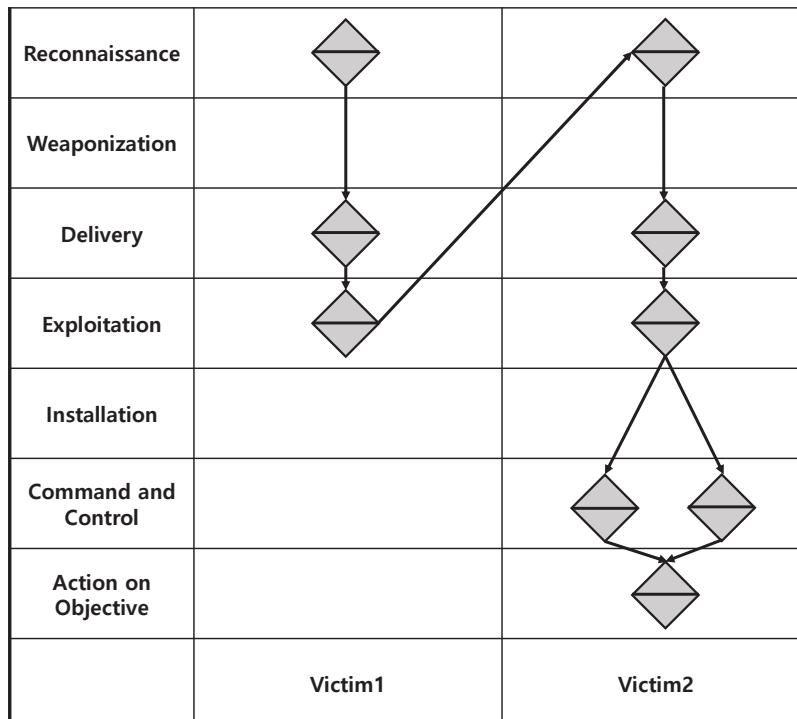
**Table 1:** The cyber threat taxonomy consisting of kill chain phase, tactics, techniques and procedures

Kill chain phase	Tactics	Techniques	Procedures
Delivery	Login attempt	Abnormal login attempt	Abnormal login attempt
		brute force attack	brute force attack
	Web hacking	Command injection	SQL injection XML injection
		Code injection	XSS
		File inclusion	Local File Inclusion Remote File Inclusion

From that, each states of the attack scenarios are mapped to its kill chain phases and reclassified into corresponding technique. This process expresses attack in normalized form and enables a contextual understanding of occurred attack.

#### 4.2.2 Feature expansion

Next, we supplemented additional feature information for each attack state. Information about attackers, assets and infrastructures can be added by using past threats, open cyber threat intelligence and expert knowledge of internal systems. For example, ip address of a threat can be mapped to specific attack group and destination ip can be mapped to asset of the system. Each of information is stored in database and automatically mapped to the feature when the attack occurs. This additional information can be used to apply the regenerated attack scenario to the diamond model. The information processed with the diamond model can be used as cyber threat intelligence and the information expressed through the diamond model is more helpful for decision making than viewing the initial alert. The attack scenario using the diamond model is shown in the Figure 3.



**Figure 3:** The example of diamond model applied to attack scenario (Caltagirone et al. 2013). Each diamond represents attack state and it contains its features. Arrows represents causal relationship and correlation probability. This provides overall information of the attack and enables a contextual understanding of the attack situation

## 5. Conclusion

We integrate alert correlation and diamond model for cyber threat intelligence. Machine learning is performed on threat alerts generated through various heterogeneous detection sensors, and alert correlation between threats can be performed to define the causal relationship between threats. Based on the derived correlation model, it is possible to automatically reconstruct the past threat scenarios, predict future developable threats, and synthesize the reconstructed and predicted threat chains. This threat chain is represented by cyber threat intelligence through the diamond model to provide information about entire attack scenario. This effectively supports the analyst's determination.

Following these suggestions, automated techniques can help you get quick, accurate results from big data-based analytics out of reliance on expert analysis. It is possible to prevent a part that can be missed or a case where it takes a long time. In addition, there is an opportunity to prevent attack before the attacker reaches the final goal of cyber attack by escaping the method that responded after the cyber attack. An immediate solution with preemptive and proactive response opportunities can minimize damage and prevent further attacks.

In the future, further studies will be conducted to verify the model by inputting data and to generate more information than the observed data through the cyber threat intelligence database built with big data.

## Acknowledgements

This work was supported by Defense Acquisition Program Administration and Agency for Defense Development under the contract. (UD160066BD)

## References

- Ahmadian Ramaki, A. & Rasoolzadegan, A. (2016), 'Causal knowledge analysis for detecting and modeling multi-step attacks', *Security and Communication Networks* 9(18), 6042–6065.
- Caltagirone, S., Pendegast, A. & Betz, C. (2013), The diamond model of intrusion analysis, Technical report, Center For Cyber Intelligence Analysis and Threat Research Hanover Md.
- Dwivedi, N. & Tripathi, A. (2015), Event correlation for intrusion detection systems, in '2015 IEEE International Conference on Computational Intelligence & Communication Technology', IEEE, pp. 133–139.
- Holgado, P., VILLAGRA, V. A. & Vazquez, L. (2017), 'Real-time multistep attack prediction based on hidden markov models', *IEEE Transactions on Dependable and Secure Computing*.
- Kang, D. & Na, J. (2012), 'A rule based event correlation approach for physical and logical security convergence', *IJCSNS Intern. Journal of Computer Science and Network Security* 12(1), 28–32.
- Martin, L. (2014), 'Cyber kill chain®'.  
<http://cyber.lockheedmartin.com/content/dam/lockheedmartin/rms/documents/cyber/GainingtheAdvantageCyberKillChain.pdf>
- McMillan, R. (2013), 'Open threat intelligence'. <https://www.gartner.com/doc/2487216/definition-threat-intelligence>
- Mirheidari, S. A., Arshad, S. & Jalili, R. (2013), Alert correlation algorithms: A survey and taxonomy, in 'Cyberspace Safety and Security', Springer, pp. 183– 197.
- Ren, H., Stakhanova, N. & Ghorbani, A. A. (2010), An online adaptive approach to alert correlation, in 'International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment', Springer, pp. 153–172.
- Sendi, A. S., Dagenais, M., Jabbarifar, M. & Couture, M. (2012), 'Real time intrusion prediction based on optimized alerts with hidden markov model', *Journal of networks* 7(2), 311.
- Shackleford, D. (2015), 'Who's using cyberthreat intelligence and how?', SANS Institute. Retrieved January 24, 2018.
- Simmons, C., Ellis, C., Shiva, S., Dasgupta, D. & Wu, Q. (2014), Avoidit: A cyber attack taxonomy, in '9th Annual Symposium on Information Assurance (ASIA'14)', pp. 2–12.
- Wu, Q., Gu, Y., Cui, X., Moka, P. & Lin, Y. (2010), A graph similarity-based approach to security event analysis using correlation techniques, in '2010 IEEE Global Telecommunications Conference GLOBECOM 2010', IEEE, pp. 1–5.
- Yang, S. J., Stotz, A., Holzapfel, J., Sudit, M. & Kuhl, M. (2009), 'High level information fusion for tracking and projection of multistage cyber attacks', *Information Fusion* 10(1), 107–121.
- Zhu, B. & Ghorbani, A. A. (2006), 'Alert correlation for extracting attack strategies', *IJ Network Security* 3(3), 244–258.

# Digital Dematerialisation in the Portuguese tax System

Ana Paula Silva<sup>1, 2</sup> and Susana Aldeia<sup>3, 4</sup>

<sup>1</sup>IPVC- Instituto Politécnico de Viana do Castelo, Valença, Portugal

<sup>2</sup>Research on Economics, Management and Information Technologies - REMIT, Oporto, Portugal

<sup>3</sup>Portucalense University, Research on Economics, Management and Information Technologies (REMIT), Portucalense Institute for Legal Research (IJP) - Porto, Portugal  
<sup>4</sup>Polytechnic Institute of Cávado and Ave (IPCA), Research Centre on Accounting and Taxation (CICF) - Barcelos, Portugal

[anahrmsilva@gmail.com](mailto:anahrmsilva@gmail.com)

[susanaldeia@sapo.pt](mailto:susanaldeia@sapo.pt)

**Abstract:** The importance of the digital dimension of the world has been transforming society, in many domains, increasingly so. The tax system felt the need to follow this tendency, and over the last decades, Portuguese companies were obliged to adapt to this new reality as their routines underwent several changes thereof. The Portuguese case study is of particular relevance since it is widely acknowledged it has pioneered cutting edge dematerialization, and in this regard, it has also been taken as a model by other countries which are now replicating some long-taken Portuguese initiatives towards such dematerialization. That process represents an important step in the relationship between taxpayers and the Tax Authority because it increases the tax information's security. This paper provides a critical review of the main dimensions of digital dematerialization in the Portuguese jurisdiction. Starting from the deductive method of literature review, which covers the doctrine, the legislation, and the jurisprudence, this paper includes practical exercises whereby we demonstrate the tightness of the tax calendar companies struggle with, somehow counterbalanced by the easiness at which declarations may be submitted. The control by Tax Administration has been increasingly tighter, to the extent that a particular terror climate has been laid down, in an attempt to act as an effective deterrent of companies' tax evasion.

**Keywords:** dematerialization, tax dematerialization, tax compliance, Portuguese tax system

## 1. A review on Portuguese dematerialized compliance under a legal framework

Nowadays it is impossible to ignore the importance of the digital dimension of the world because the electronic revolution changed society, completely. The tax system felt the need to embrace the tendency of a nearly paperless future, and over the last decades, Portuguese companies were obliged to adapt to this new reality as their routines underwent several changes thereof. The Portuguese case study is of particular relevance since it is widely acknowledged it has pioneered cutting edge dematerialization (Oliveira and Machado, 2018; Reis and de Lima, 2009), and in this regard, it has also been taken as a model by other countries which are now replicating some long-taken Portuguese initiatives towards such dematerialization. Several independent best practice awards have been received by the Portuguese tax administration (Reis and de Lima, 2009). In fact, in the field of taxation in Portugal, there is total dematerialization of procedures and files based on a single portal- the "Finance Portal" (*Portal das Finanças*). Dimensions of dematerialization worth highlighting include: i) electronically issued documents, such as invoices, with legal value assigned; ii) online access to the integrated tax situation of individuals and corporate entities; iii) online availability of all types of official forms; iv) the possibility of taxpayers and accounting professionals to contact the Portuguese Tax Administration by e-mail, using the application form of the service (*e-balcão*) to raise questions and see their doubts clarified in writing; and, most importantly, v) an exhaustive online filing system, which allows self-assessment, filing and payment of taxes with simple internet access (Oliveira and Machado, 2018; Reis and de Lima, 2009). Dematerialised compliance covers, amongst many others, VAT returns, corporate and personal income tax statements, withholding tax statements, inventories'annual communication for Tax Administration, issuance of invoices and their monthly communication for Tax Administration, goods transportation documents, social security compliance, and year-end accountability through the so-called 'Simplified Corporate Information return (IES - *Informação Empresarial Simplificada*) towards Tax Administration, Bank of Portugal, Statistics Portugal, Institute of Registries and Notaries. This important step of tax compliance dematerialization implies to taxpayers of the Portuguese jurisdiction more security in tax information. The similar process that was previously in force required the submission of tax statements manually and the data's introduction of the declarations presented manually, in the system of the tax authority, was also done manually. This process represented significant data entry errors because it required the intervention of human resources in the processing of tax data.

Accounting professionals may secure diversified portfolios of clients that include both individuals and corporate entities, thereby accruing all sorts of dematerialized compliance obligations, of various frequencies, ranging from monthly to annually. Table 1 provides a reasonably comprehensive review on monthly, fully dematerialized, compliance, while Table 2 compiles annual such compliance. For parsimony, quarterly dematerialized compliance as well as dematerialized compliance of other regularity is left uncovered (e.g., communication of transport documents, statement of admission of new employees, statement of tax planning schemes proposed/accompanied by promoters, tax Planning communication by users). As data from Table 1 and Table 2 were considered to provide sufficient grounds for the scope of this paper: while Portugal has pioneered cutting edge's tax compliance dematerialization, a significant problem of the Portuguese tax system resides in its voluminous tax compliance activities, where day-to-day compliance costs are high (OECD Economic Surveys: Portugal, 2010).

**Table 1:** Monthly dematerialized compliance impending over accounting professionals

Monthly deadline	Dematerialized routine	Applicable legislation and administrative instructions
10 <sup>th</sup>	Statement of income paid and tax withheld (employment income), for Tax Administration, regarding previous month ( <i>DMR- Declaração Mensal de Remunerações</i> ) – filing	Art. 119, paragraph 1c) of the Personal Income Tax Code (CIRS- <i>Código do Imposto Sobre o Rendimento das Pessoas Singulares</i> ) Ordinance No. 6/2013, of January 10
	Statement of income paid and social security contributions, for Social Security, regarding previous month - filing	Art. 40, paragraph 2 of the Social Security Contribution Regimes Law (Law No.110/2009 as further amended)
	VAT return and annexes for the taxpayers covered by the monthly VAT regime (filing and payment)	Art. 27; Art. 29, paragraph 1; and Art. 41, paragraph 1a) of the VAT Code (CIVA- <i>Código do Imposto sobre o Valor Acrescentado</i> ) Ordinance No. 375/2003, of May 10
20 <sup>th</sup>	Statement of corporate and personal income tax withheld and stamp duty, regarding previous month(filing and payment)	Art. 98, paragraph 3 of the Personal Income Tax Code (CIRS) Art. 94, paragraph 6 of the Corporate Income Tax Code (CIRC- <i>Código do Imposto sobre o Rendimento das Pessoas Coletivas</i> ) Ordinance No. 523/2003, of July 4
	Recapitulative statement (intracommunity supplies of goods and services)	Art. 23 and art. 30 of the VAT Regime in Intracommunity Transactions (RITI - <i>Regime do Iva nas Transações Intracomunitárias</i> ). Art. 29, paragraph 1l) of the VAT Code (CIVA)
	Reporting of invoices issued, regarding previous month	Art. 3 of the Decree-Law No. 198/2012, of August 24 (as further amended) Ordinance No. 426-A/2012, of December 28
	Statement of social security contributions (filing and payment)	Art. 43 of the Social Security Contribution Regimes Law (Law No.110/2009 as further amended)
	Employment Compensation Fund (FCT- <i>Fundo de Compensação do Trabalho</i> ) and Employment Compensation Guarantee Fund (FGCT- <i>Fundo de Garantia de Compensação do Trabalho</i> ) (filing and payment)	Law No. 70/2013, of August 30
30 <sup>th</sup> / 31 <sup>st</sup> <sup>(1)</sup>	Annual vehicle tax (IUC- <i>Imposto Único de Circulação</i> ) (filing and payment)	Art. 4, art. 16, and art. 17 of the Single Transit Tax Code (CIUC- <i>Código do Imposto Único de Circulação</i> )
Last day of each month	Statement of income paid or placed at the disposal of non-resident entities - Form 30	Art. 119, paragraph 7 of the Personal Income Tax Code (CIRS)

While dematerialization made filing and payment secure, the preparation of tax returns often remains burdensome, accounting for the bulk of total compliance time. Furthermore, and despite progress operated, there is still scope for increased coordination between tax administration and social security. For example, Table

<sup>(1)</sup>Last day of the month in which the license plate was issued.

1 shows that corporate entities still have to file separate monthly returns for social security contributions and personal income tax withheld from employees' pay.

**Table 2:** Annually dematerialized compliance impending over accounting professionals

Annually deadline		Dematerialized routine	Applicable legislation and administrative instructions
Month	Day		
January	31 <sup>st</sup>	Statement of income paid, tax withheld, tax deductions, social security and health contributions, and subscriptions (except employment income), regarding previous year – Form10	Art. 1, paragraph 1c)ii) of the Personal Income Tax Code (CIRS) Ordinance No. 383/2015, of October 26 Circular letter No. 20181/2016, of January 4
		Inventory communication, regarding December 31 of the previous year	Ordinance No. 2/2015, of January 6
February	28 <sup>th</sup>	Option to adopt the simplified tax system by small business entities	Art. 86-A of the Corporate Income Tax Code (CIRC)
		Statement of income paid and tax withheld as final withholding, regarding previous year –Form39	Ordinance No. 371/2015, of October 20
March	31 <sup>st</sup>	First special payment on account (PEC- <i>Pagamento Especial por Conta</i> ): self-assessment, filing, and payment	Art. 106 of the Corporate Income Tax Code (CIRC)
April	15 <sup>th</sup>	Single Report concerning the social activity of the company, to be sent to the Office of Strategy and Studies of the Ministry of Economy (RU- <i>Relatório Único</i> )	Ordinance No. 55/2010, of January 21 (as further amended)
May	31 <sup>st</sup>	Corporate income tax, municipal surtax and state surtax (self-assessment, filing, and payment) – Form 22	Art. 120, paragraph 1 and art. 104, paragraph 1b) of the Corporate Income Tax Code (CIRC)
June	30 <sup>th</sup>	Reporting of creation or contributions to stock option/ subscription/ award/ other schemes on behalf of employees and/or board members, regarding previous year- Form 19	Art. 119, paragraph 7 of the Personal Income Tax Code (CIRS)
July	15 <sup>th</sup>	Filing of 'Simplified Corporate Information' return (IES- <i>Informação Empresarial Simplificada</i> ), regarding the previous year	
		Law No. 8/2007, of January 12	
		Ordinance No. 64-A/2011, of February 3	
		Art. 113 of the Personal Income Tax Code (CIRS)	
		Art. 117 and 121 of the Corporate Income Tax Code (CIRC)	
		Art. 29, paragraph 1d)e)f)h) of the VAT Code (CIVA)	
		Art. 52 of the Stamp Duty Code (CIS- <i>Código do Imposto do Selo</i> ) Decree-Law No. 292/2009, of October 13	
	31 <sup>st</sup>	First payment on account (self-assessment, filing, and payment) concerning the ongoing year	Art. 104 of the Corporate Income Tax Code (CIRC) Ordinance No. 514/2003, of July 2
		First additional payment on account (self-assessment, filing, and payment) concerning the ongoing year	Art. 105 - A of the Corporate Income Tax Code (CIRC)
		Statement of issuance or circulation of securities, regarding previous year- Form 34	Art. 120 of the Personal Income Tax Code (CIRS)
		Statement of exempt income, income subject to reduced tax rates or regarding which no withholding tax is required, regarding previous year – Form 31	Art. 119, paragraph 2 of the Personal Income Tax Code (CIRS)
September	30 <sup>th</sup>	Second payment on account (self-assessment, filing, and payment) concerning the ongoing year	Art. 104 of the Corporate Income Tax Code (CIRC) Ordinance No. 514/2003, of July 2

Annually deadline		Dematerialized routine				Applicable legislation and administrative instructions			
Month	Day								
		Second additional payment on account (self-assessment, filing, and payment) concerning the ongoing year				Art. 105 - A of the Corporate Income Tax Code (CIRC)			
October	31 <sup>st</sup>	Second special payment on account (PEC- <i>Pagamento Especial por Conta</i> ): self-assessment, filing, and payment				Art. 106 of the Corporate Income Tax Code (CIRC)			
December	15 <sup>th</sup>	Third payment on account (self-assessment, filing, and payment) concerning the ongoing year				Art. 104 of the Corporate Income Tax Code (CIRC)			
		Third additional payment on account (self-assessment, filing, and payment) concerning the ongoing year				Ordinance No. 514/2003, of July 2			
						Art. 105 - A of the Corporate Income Tax Code (CIRC)			

## 2. Real-life examples

Portuguese accounting legislation allows for simplified accounts in the case of small/medium-sized or micro-enterprises, more evidently after the Decree-Law 98/2015 of June 2 became effective from 1 January 2016 (to comply with Directive 2013/34/EU of the European Parliament and of the Council, of June 26). Based upon two case studies, we show such accounting relief on the smaller is somewhat impaired with their tax compliance (dematerialized) workload.

### 2.1 The case of a micro-entity

Table 3 compiles monthly, quarterly, and annual dematerialized compliance obligations of one given micro-entity: a private label clothing manufacturer, employing as few as four full-time workers, and whose figure for annual sales turnover is around 60.000€. This business has been generating profit, so three payments on account are due; nonetheless, the additional payment on the account does not apply due to the small size.

**Table 3:** Dematerialised tax compliance impending over one real-world micro-entity

Compliance obligation	Deadline											
	Jan	Feb	March	April	May	June	July	Au.	Sep	Oct	Nov	Dec
Statement of income paid and tax withheld (employment income), for Tax Administration, regarding previous month– filing	10	10	10	10	10	10	10	10	10	10	10	10
Statement of income paid and social security contributions, for Social Security, regarding previous month - filing	10	10	10	10	10	10	10	10	10	10	10	10
Statement of corporate and personal income tax withheld and stamp duty, regarding previous month (filing and payment)	20	20	20	20	20	20	20	20	20	20	20	20
Statement of corporate and personal income	20	20	20	20	20	20	20	20	20	20	20	20

Compliance obligation	Deadline											
	Jan	Feb	March	April	May	June	July	Au.	Sep	Oct	Nov	Dec
tax withheld and stamp duty, regarding previous month (filing and payment)												
Reporting of invoices issued, regarding previous month	20	20	20	20	20	20	20	20	20	20	20	20
Statement of social security contributions (filing and payment)	20	20	20	20	20	20	20	20	20	20	20	20
Employment Compensation Fund and Employment Compensation Guarantee Fund (filing and payment)	20	20	20	20	20	20	20	20	20	20	20	20
VAT return and annexes - quarterly VAT regime (filing and payment)		15			15			15			15	
Statement of income paid, tax withheld, tax deductions, social security and health contributions, and subscriptions (except employment income)– Form10		31										
Inventory communication, regarding December 31 of the previous year		31										
First special payment on account: self-assessment, filing, and payment			31									
Single Report concerning the social activity of the company				15								
Corporate income tax, municipal surtax and state surtax (self-assessment, filing, and					31							

Compliance obligation	Deadline											
	Jan	Feb	March	April	May	June	July	Au.	Sep	Oct	Nov	Dec
payment) – Form 22												
Filing of 'Simplified Corporate Information' return							15					
First payment on account (self-assessment, filing, and payment)							31					
Second payment on account (self-assessment, filing, and payment)								30				
Second special payment on account: self-assessment, filing, and payment									31			
Third payment on account (self-assessment, filing, and payment)												15

## 2.2 The case of a small entity

Table 4 exhibits further compliance obligations of a small entity, as compared to those of the micro-entity presented in Table 3. This small entity is a supermarket with a total annual turnover of around 8.500.000 euros and 43 employees.

**Table 4:** Further dematerialized tax compliance impending over one small real-world entity as compared to a micro-entity

Compliance obligation	Deadline											
	Jan	Feb	March	April	May	June	July	Au.	Sep	Oct	Nov	Dec
Communication of transport documents	Daily											
Recapitulative statement (intracommunity supplies of goods and services)	20	20	20	20	20	20	20	20	20	20	20	20
VAT return and annexes - monthly VAT regime (filing and payment)	10	10	10	10	10	10	10	10	10	10	10	10

The bottom line of this section is clear: the resource-constrained micro-entities are disproportionately overwhelmed with (dematerialized) tax compliance. Our second real-life example refers to a company which is much larger than the former; yet, not much further compliance is required from it.

## 3. Discussion

Tax dematerialization has brought several benefits, both to Portuguese Tax Administration and Portuguese citizens. Firstly, the inherent centralization of information on individuals and corporate entities facilitates

detection of tax evasion and thereby not only increases tax revenue but also incites a generalized feeling of greater tax justice. Faulty taxpayers are easily, or even automatically, spotted and notified to pay their tax liabilities and bear due penalties. Benefits extend to concerned professionals and taxpayers as they may conveniently accomplish their tax obligations and see their doubts clarified in writing from their offices/homes, which results in substantial savings of time and money(Oliveira and Machado, 2018; Reis and de Lima, 2009). Consistently, Oliveira and Machado (2018) acknowledge tax dematerialization in Portugal has promoted more excellent proximity between taxpayers and concerned professionals, on the one hand, and Tax Administration, on the other hand, as time constraints and bureaucratic requirements are, to a great extent, overcome, resulting in improved effectiveness and efficiency. It is also worth highlighting that dematerialization has allowed tax compliance to be simplified in many instances as one same statement may respond to the information requirements of different stakeholders. For example, the annual return and correspondent annexes (IES) provides information to Tax Administration, the Bank of Portugal, Statistics Portugal, and the Institute of Registries and Notaries.

Nonetheless, could it be that the advantages reaped from the speed of the dematerialized Portuguese tax system are counterweighed by the disadvantages thereof? For example, it is widely acknowledged that a substantial compliance burden, worsened by a fast pace changing tax landscape which makes it challenging to keep ahead and be fully compliant, acts as a significant deterrent of foreign direct investment (e.g., Leonidou, 2017; OECD Economic Surveys: Portugal, 2010).

In 2010, unlisted Portuguese companies in the non-finance sector compulsorily adopted a new Accounting Standards System (*Sistema de Normalização Contabilística* - SNC), a principles-based accounting standards system, to align Portuguese accounting standards with the International Accounting Standards Board (IASB) model, which is Anglo-Saxon based. Before the adoption of this IFRS-based accounting system, Portugal's accounting regime was based on the Continental European model of rules-based regulation, primarily geared towards government and tax needs, rather than to meet investors' needs (Ferreira, Lara, & Gonçalves, 2007; Guerreiro, Rodrigues & Craig, 2012). While Portugal's international accounting convergence called for a substantial detachment from its previous accounting regime, arguably, the high level of Portuguese tax digitalization has facilitated an exponential growth of tax compliance obligations, resulting in accounting professionals overlooking the clearer disengagement between accounting and taxation that characterizes the IFRS model. It can be inferred from Table 1, taken together with Table 2(section 1), that accounting professionals must devote a substantial deal of their time to filing dematerialized statements on a nearly daily basis. Therefore, despite obvious benefits arising from tax digitalization, it has also highlighted context costs with companies, most notably micro-entities, being overwhelmed with tax compliance obligations. It raises the risk of penalties many resource-constrained SMEs cannot afford to bear, overdraws attention towards compliance, as much as lowers commitment to International Financial Reporting Standards' (IFRS) *de facto* implementation, arguably contributing to perpetuate the long-standing practice of tax-oriented accounting. Despite filing being much quicker than bureaucratic paper procedures requiring face-to-face interaction, this benefit may be to a certain extent written off by an enormous tax compliance workload in Portugal. Consistently, Dâmaso and Martins (2015), based on a large sample of Portuguese chartered accountants, found the majority of them did not recommend the simplified tax regime, and the main motive for those who recommended it was *not* a decrease of compliance costs. Also, increased tax enforcement, which has been taking place in Portugal, increases the tax burden for companies, for example by demanding more time gearing up for questions from Tax authorities (Leonidou, 2017). Arguably, it may also act as a deterrent of compliance, encouraging the shadow economy as a means to achieve tax savings, to escape non-compliance penalties, and to avoid the context costs from timely compliance. One of the motivations for deliberately concealing market activities from public authorities is, according to Schneider (2007), the avoidance to fulfill certain administrative procedures, such as completion of statistical questionnaires and any forms. Barbosa, Pereira, and Brandão (2013) concluded that despite the share of the shadow economy in Portugal decreased sharply between 1997 and 2000 (from about an estimate of 52% to 13.4% of the GDP), since 2001 it has been rising again, most evidently after 2007, having reached a peak of 24.2% of the GDP in 2011. It was precisely in 2006 that the Portuguese government developed a programme for administrative simplification and modernization, called the "SIMPLEX" Programme, which entailed dematerialization of procedures and files. Lastly, under the pretext of tax simplification and aid in tax compliance, the Portuguese Tax Administration ends up upholding taxpayers' confidential information.

#### **4. References**

- Barbosa, Eduardo, Pereira, Samuel and Brandão, Elísio (2013) "TheShadowEconomyinPortugal: An Analysis Using the MIMIC Model," Working Papers (FEP) -- Universidade do Porto, No. 514, pp 1-48.
- Circular letter No. 20181/2016, of January 4.
- Dâmaso, Maria and Martins, António (2015) "The New Portuguese Simplified Tax Regime for Small Business," Journal of Accounting & Finance, Vol 15, No. 5, pp 76-84.
- Decree-Law 98/2015 of June 2.
- Decree-Law No. 198/2012, of August 24 (as further amended).
- Decree-Law No. 292/2009, of October 13.
- Decree-Law No. 102/2008, of June 20 (as further amended) - VAT Regime in Intracommunity Transactions (RITI - Regime do Iva nas Transações. Intracomunitárias).
- Decree-Law No. 287/2003, of November 12 (as further amended) - Stamp Duty Code (CIS- Código do Imposto do Selo).
- Decree-Law No. 442-B/88, of November 30 (as further amended) - Corporate Income Tax Code (CIRC- Código do Imposto sobre o Rendimento das Pessoas Coletivas).
- Decree-Law No. 442-A/88, of November 30 (as further amended) - Personal Income Tax Code (CIRS- Código do Imposto Sobre o Rendimento das Pessoas Singulares).
- Decree-Law No. 394-B/84, of December 26 (as further amended) - VAT Code (CIVA- Código do Imposto sobre o Valor Acrescentado).
- Directive 2013/34/EU of the European Parliament and of the Council, of June 26.
- Ferreira, Leonor, Lara, Juan and Gonçalves, Tiago (2007)"Accounting conservatism in Portugal: Similarities and differences facing Germany and the United Kingdom," Revista de Administração Contemporânea, Vol 11, pp 163-188.
- Guerreiro, Marta, Rodrigues, Lúcia and Craig, Russel (2012) "Factors influencing the preparedness of large unlisted companies to implement adapted International financial reporting standards in Portugal," Journal of International Accounting, Auditing, and Taxation, Vol 21, No. 2, pp 169-184.
- Law No. 70/2013, of August 30.
- Law No.110/2009, of September 16 (as further amended) - Social Security Contribution Regimes Law.
- Law No. 22-A/2007, of June 29 (as further amended) - Single Transit Tax Code (CIUC - Código do Imposto Único de Circulação).
- Law No. 8/2007, of January 12.
- Leonidou, Natalie (2017) "Compliance burdens rise despite lower tax rates and more incentives", International Tax Review, June, p. 10-10.
- OECD Economic Surveys: Portugal (2010) "Towards a less distortive and more efficient tax system", Vol 2010, No. 16, pp 57-91.
- Oliveira, Fernanda, and Machado, Carla (2018) "Papers, my friend, are blowing in the wind: towards a paperless Administration," Perspectives of Law and Public Administration, Vol 7, No. 1, pp 1-22.
- Ordinance No. 2/2015, of January 6.
- Ordinance No. 383/2015, of October 26.
- Ordinance No. 371/2015, of October 20.
- Ordinance No. 6/2013, of January 10.
- Ordinance No. 426-A/2012, of December 28.
- Ordinance No. 64-A/2011, of February 3.
- Ordinance No. 55/2010, of January 21 (as further amended).
- Ordinance No. 523/2003, of July 4.
- Ordinance No. 514/2003, of July 2.
- Ordinance No. 375/2003, of May 10.
- Reis, Miguel, and de Lima, João Velez (2009) "The Portuguese tax system - a first balance of the adopted simplified measures," International Tax Review, Vol 20, No. 1, pp. p. 53-53.
- Schneider, Friedrich (2007) "Shadow Economies and Corruption all over the World: New Estimates for 145 Countries", Economics, No. 2007-9, pp 1-47.

# Information Security is More Than Just Policy; It is in Your Personality

Petteri Simola<sup>1</sup>, Toni Virtanen<sup>1</sup> and Miika Sartonen<sup>2</sup>

<sup>1</sup>Finnish Defence Research Agency, Human Performance Division, Finland

<sup>2</sup>Finnish Defence Research Agency, Concepts and Doctrine Division, Finland

[petteri.simola@mil.fi](mailto:petteri.simola@mil.fi)

**Abstract:** It has been estimated that human factors (HF) account for 27% of data breaches on the global scale (Ponemon institute 2018). Even with clear and often strict policies in place, with clear sanctions, employees still are considered to be the weakest link in the field of information security (IS). This paper seeks to find one explanation to this phenomenon in military context by exploring military cadets' attitudes towards IS, as well as their reasons and justifications for using neutralisation techniques in order to transgress from organisational IS regulations. Neutralisation techniques offer a way of rendering existing norms inoperative by justifying behaviour that violates those norms (Rogers & Buffalo 1974; Sykes & Matza 1957). These techniques are as follows: Condemnation of the condemners, The Metaphor of the ledger, Denial of injury, Denial of responsibility, Appeal to higher loyalties and Defence of necessity. 144 military cadets completed a survey assessing their use of neutralisation techniques (Siponen & Vance 2010) in addition to assessing their personality by the Five Factor (Konstabel, et. al. 2012) and the Dark Triad (Jones & Paulhus, 2014) models of personality. The Dark Triad model supplements the Five Factor model with more sinister aspects of personality: Machiavellianism, Narcissism and Psychopathy, which are still considered to be sub-clinical. Even though the tendency to use neutralisation techniques was relatively low, there still was a significant correlation between personality traits and the use of neutralisation techniques. Those high in Machiavellianism ( $r. 0.19 - 0.4$ ) and Neuroticism ( $r. 0.23 - 0.4$ ) were more likely to use these techniques whereas high scores on Conscientiousness ( $r. -0.18 - -0.27$ ) and Extraversion ( $r. -0.27 - -0.42$ ) decreased this likelihood. The results suggest that a more individualised approach in IS education could be useful. Understanding how one's personality can sensitise oneself to certain kinds of neutralisation techniques can help an individual to acknowledge his or her strengths and vulnerabilities in IS behaviour.

**Keywords:** Information security, personality, neutralisation, military cadets, big five, dark triad

---

## 1. Introduction

Information security has always been a top priority in any military organisation, and thus significant effort is typically invested in protecting critical information assets. The same goes for any commercial company that wishes to protect its business. Despite the efforts, both intentional and unintentional failures to follow security regulations, as well as security breaches, occur to both military and commercial operators (Furnell & Clarke 2012; Siponen, Pahnila & Mahmood, 2006). As a recent military example, a U.S military base in Afghanistan was revealed after the fitness application Strava published its GPS data (Hern, 2018). To remedy the situation, the information security organisations very often focus on more sophisticated technical solutions and protocols, in order to improve information security. It has been acknowledged, however, that these technical tools are at risk of not keeping up with fast development of security threats (Furnell & Clarke 2012), at least if the human component, which is often the weakest link (Acuña 2016), is not accounted for. According to a recent Ponemon Institute study (2018), 48 % of all data breaches are caused by malicious or criminal attacks, 27 % happen due to human error and 25 % are caused by both IT and business process failures and system glitches. Even though malicious acts are the most crucial, human error has significant impact. Previously it has been estimated that over half of the information security breaches are indirectly or directly caused by employees' violations of information security regulations and rules (Dhillon and Moores 2001; Stanton et al. 2005).

In general, it is reasonable to assume that people typically follow regulations, and disobeying them is thus an exception. The reasons for an individual to decide not to follow information security regulations and directions may be diverse and differ between situations. To break the rules, individuals often need to have a reason, or some kind of justification to do so. There are obvious reasons for malevolent behaviour, such as revenge or financial benefit. In many cases, however, the reasons for violating regulations are more mundane, such as getting a task done easier or faster, without any conscious intention to harm the organisation. Even in these small everyday violations, people often find it necessary to justify one's actions.

Siponen and Vance (2010) proposed that one of the ways employees justify their actions may be related to a phenomenon called neutralisation techniques, which has its roots in criminology research. Neutralisation techniques offer a way for the individual to render existing norms inoperative by justifying norm-violating

behaviour. These techniques are as follows: Condemnation of the condemners, The Metaphor of the ledger, Denial of injury, Denial of responsibility, Appeal to higher loyalties and Defence of necessity. (Rogers & Buffalo 1974; Sykes & Matza 1957). These techniques and their relevance to information security are described briefly on table 1.

**Table 1:** Neutralisation techniques

Technique	Explanation in information security context
<b>Condemnation of the condemners</b>	One neutralises his or her actions by blaming those who point out norm-violating behaviour. In the information security context, an employee could say that it is not wrong to violate information security policies that are unreasonable.
<b>The Metaphor of the ledger</b>	Is based on the idea of compensating occasional bad acts with good acts. In the information security context employees could argue that their general adherence to security policies compensates their occasional violation of these policies.
<b>Denial of injury</b>	Involves justifying an action by minimising the harm it causes. In the information security context, an employee might argue that it is ok to violate information security policies, if no harm is done to the company.
<b>Denial of responsibility</b>	Is employed when a person committing a deviant act defines himself as lacking responsibility for his or her actions. In the information security context, it has been reported that employees denied their responsibility to comply with confidential emails encryption policy because they rationalised that the policy was unclear.
<b>Appeal to higher loyalties</b>	Is employed by those who feel they have a dilemma that must be resolved at the cost of violating a law or policy. In the information security context, an employee could argue that he or she must violate a policy in order to get his or her work done.
<b>Defence of necessity</b>	Is based on the justification of the rule-breaking being viewed as necessary. In the information security context, an employee could claim that they did not have time to comply with the policies due to tight deadlines.

Table adapted from Puhakainen, 2006; Siponen & Vance 2010

Siponen and Vance (2010) demonstrated that neutralisation techniques have an effect on employees' intentions to violate information security policies, and thus their research adds to the understanding of the reasons behind security violations. People vary, however, in their likelihood of deviating from expected behaviour, and not everyone uses neutralisation techniques. One fundamental factor that mediates behaviour and the way we perceive the world is personality. According to American Psychological Association (APA) personality is defined as follows: personality refers to individual differences in characteristic patterns of thinking, feeling and behaving (APA 2019). Even though personality research is a widely studied field in psychology, its relation to information security has gained only limited attention. We argue that personality mediates our compliance to follow information security regulations and thus our information security related risk assessment.

The Big Five personality model, also known as the Five Factor or the OCEAN personality model, is probably the best known and the most studied personality model (Konstabel, Lönnqvist, Walkowitz, Konstabel & Verkasalo 2012). It is widely used in both civilian and military recruitment processes and aptitude testing. The Big Five model divides personality traits into five factors: Openness to experience, Conscientiousness, Extraversion, Agreeableness and Neuroticism. Openness to experience measures how open and curious an individual is for new ideas and experiences, versus how cautious and wary of change he or she is. Conscientiousness describes how organized and self-disciplined, versus disorganized and spontaneous he or she can be. High score in Extraversion typically indicates sociability and the tendency to seek stimulation from the company of others, whereas people low in extraversion can be more reflective and reserved with other people. Agreeableness depicts the tendency to be more cooperative and compassionate towards others, versus being more suspicious and antagonistic. A person with very high agreeableness is also considered to be a bit naïve, while very low Agreeableness tends to indicate competitiveness and argumentativeness. Neuroticism refers to the emotional stability of the person and the tendency to have negative feelings, such as anxiety, more easily. People with very low Neuroticism can be seen as emotionally very stable, but sometimes they appear uninspiring or even unconcerned. Those with very high Neuroticism score are viewed as dynamic, but sometimes insecure or unstable.

The Dark Triad personality model appends the Five Factor model with three personality traits, which can be considered to be more malicious, selfish and manipulative in nature. These are Narcissism, Machiavellianism, and Psychopathy (Jones & Paulhus 2014). Narcissism is comprised of egotism, sense of grandiosity and lack of

empathy towards others. Machiavellianism depicts a manipulative and exploitative character, while Psychopathy has a more antisocial and impulsive personality, and the inability to be remorseful for their actions.

## **2. The survey**

### **2.1 The aim of the survey**

The aim of the survey was twofold: first, to assess military cadet's tendencies to justify behaviour that violates information security, and secondly, to further assess these tendencies' association to personality.

### **2.2 Participants and procedure**

144 first year military cadets (2 female, 142 males; mean age 22.6, sd 1.9) from the Finnish Defence University (FDU) participated in the survey. Cadets were asked to fill in questionnaires after their last indoor lecture of the day. Participation to this study was voluntary; of the 200 first year cadets, 56 refused to participate or were unable to fill all questionnaires, thus participation rate was 72 % from the total population.

### **2.3 Measures**

A paper and pencil questionnaire with three separate subcategories, concerning personality (Big Five and Dark Triad) and tendencies to explain possible security violations with external reasons (neutralisation techniques) were administered to students. Subcategories are described below.

#### **The Big-5 personality**

Short-five questionnaire is a 60-item scale assessing the five basic factors of personality, namely: Extraversion, Neuroticism, Openness, Conscientiousness and Agreeableness (Konstabel et al. 2012). Each five factors are assessed with 12 items. For each item, participants were asked on a 7-point scale, ranging from 1 (strongly disagree) to 7 (strongly agree) to rate the extent to which statement describes themselves. Scores for each factor range from -36 to 36.

#### **Dark Triad traits**

Short Dark Triad questionnaire is 27-item scale assessing subclinical Narcissism, Machiavellianism and Psychopathy (Jones & Paulhus 2014). Each three factors are assessed with 9 items. For each item, participants were asked to rate the extent to which statement describes themselves, on a 5-point scale, ranging from 1 (strongly disagree) to 5 (strongly agree). Scores range from 9 to 45.

#### **Neutralisation techniques**

Neutralisation techniques questionnaire developed by Siponen and Vance is an 18-item scale assessing different neutralisation techniques, namely Condemnation of the condemners, Appeal to higher loyalties, Defence of necessity, Metaphor of the ledger, Denial of injury and Denial of responsibility. Each technique is assessed with 3 items. For each item, participants were asked to rate the extent to which statement describes themselves, on a 5-point scale, ranging from 1 (strongly disagree) to 5 (strongly agree). Individual items are described in article Siponen & Vance (2010). Scores range from 3 to 15.

## **2.4 Statistical analyses**

All statistical analyses were conducted with IBM SPSS v.22. Differences were considered statistically significant at a p-value < 0.05. Bivariate correlations were conducted using Spearman's rank-order correlation as normality assumptions were not met in all variables. The eight predictor factors from Big Five and Dark Triad were found to have multivariate normal distribution. Normality tests for individual variables were made using Komogorov-Smirnov test. Due to central limit theorem multivariate normal distribution is sufficient even if individual variables would not be completely normally distributed. Multicollinearity of the predictor factors was assessed using bivariate correlation. The factors correlated with each other somewhat, but there was no extremely strong correlation between factors that would indicate that they are measuring the same phenomenon. Multivariate outliers for the eight predictor factors from Big Five and Dark Triad were screened using Mahalanobis distance metric. Mahalanobis calculates the distance for each individual from the centroid value of the Big Five and Dark

Triad factors. This distance was then evaluated using chi<sup>2</sup> table ( $df=8$ ,  $\alpha=0,01$ ). No multivariate outliers were found. Separate linear regression analysis was conducted for each neutralisation technique in order to test their connection to Big Five and Dark Triad variables. The regression model had two steps, first including all the Big Five traits as explanatory variables and then appended with the Dark Triad traits in order to see how that could increase the explanatory power of the regression model.

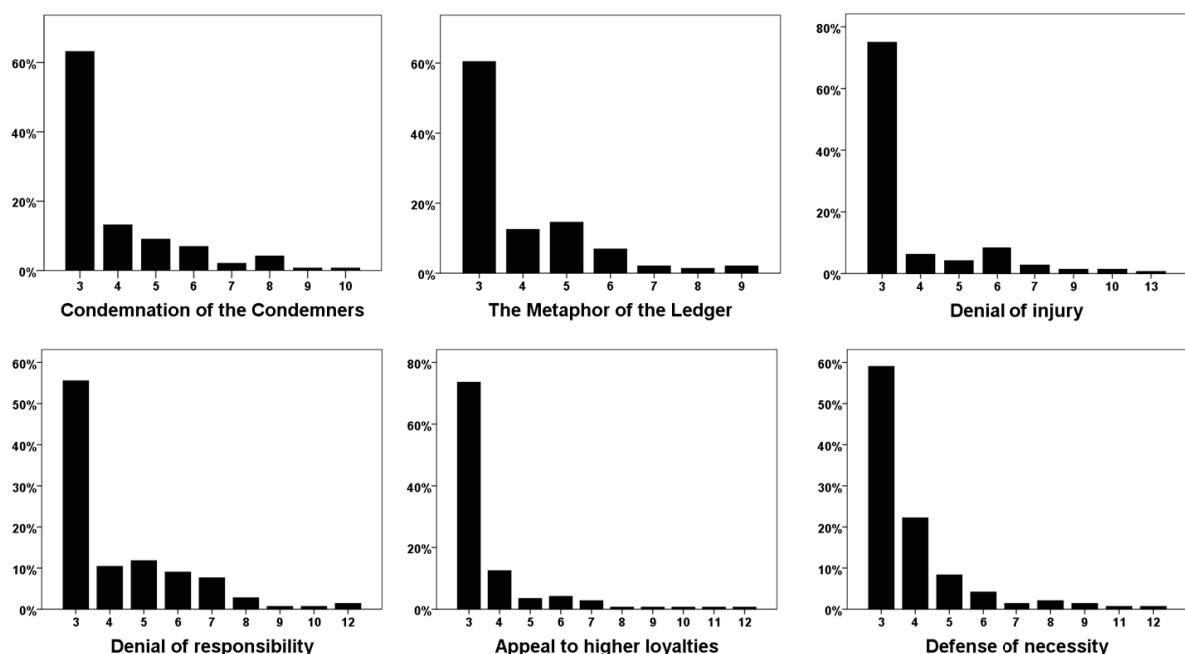
### 3. Results

Descriptive statistics (mean and standard deviations) as well as simple correlations between personality traits (Big Five and Dark Triad) and neutralisation techniques are presented in table 2. Distributions of attitudes toward using neutralisation techniques are illustrated in figure 1.

**Table 2:** Descriptive statistics and correlations between personality traits and neutralisation techniques

	Condemnation of the condemners	Metaphor of the ledger	Denial of injury	Denial of responsibility	Appeal to higher loyalties	Defence of necessity	Mean (SD)
Mean (SD)	3.9 (1.5)	3.9 (1.4)	3.8 (1.7)	4.3 (1.9)	3.7 (1.6)	3.9 (1.6)	
Neuroticism	.252**	.397**	.361**	.278**	.286**	.228**	-21.8 (7.6)
Extraversion	-.180*	-.244**	-.265**	-.233**	-.199*	-.261**	13.9 (8.6)
Openness	-.141	-.370**	-.045	-.226**	-.267**	-.253**	7.2 (9.3)
Agreeableness	-.191*	-.155	-.114	-.042	-.208*	-.134	13.8 (7.3)
Conscientiousness	-.276**	-.379**	-.419**	-.237**	-.341**	-.265**	20.6 (7.4)
Machiavellianism	.401**	.188*	.312**	.237**	.364**	.330**	20.6 (5.3)
Narcissism	-.092	-.105	-.009	-.042	-.011	-.009	28.0 (5.1)
Psychopathy	.172*	.119	.221**	.078	.146	.072	18.7 (4.5)

\*p<0.05, \*\*p<0.01



**Figure 1:** Distributions of attitudes toward using neutralisation techniques

### 3.1 Linear regression

The multiple linear regression models were employed with the neutralisation techniques as a dependent variable. The explanatory variables were Big Five and Dark Triad variables. In the first model (model 1) only Big Five variables were included in the model, whereas in the second model (model 2) Dark Triad variables were added to the model (table 3). Adding the Dark Triad variables increased the predictive power of the regression model for every neutralisation technique except Metaphor of the ledger. Different neutralisation techniques were related differently to Big Five and Dark Triad variables.

**Table 3:** Regression models

			R <sup>2</sup>	Adjusted R <sup>2</sup>	Std. Error of the Estimate	Change Statistics	
						R <sup>2</sup> Change	Sig. F Change
<b>Condemnation of the condemners</b>	Model 1	Big-five	.088	.055	1.472	.088	.025
	Model 2	Big -five and Dark triad	.215	.168	1.382	.127	.000
<b>Metaphor of the ledger</b>	Model 1	Big-five	.245	.217	1.242	.245	.000
	Model 2	Big -five and Dark triad	.272	.229	1.233	.028	.172
<b>Denial of injury</b>	Model 1	Big-five	.165	.135	1.546	.165	.000
	Model 2	Big -five and Dark triad	.220	.174	1.510	.055	.026
<b>Denial of responsibility</b>	Model 1	Big-five	.118	.085	1.774	.118	.004
	Model 2	Big -five and Dark triad	.170	.120	1.740	.052	.042
<b>Appeal to higher loyalties</b>	Model 1	Big-five	.113	.080	1.495	.113	.005
	Model 2	Big -five and Dark triad	.187	.139	1.447	.074	.008
<b>Defence of necessity</b>	Model 1	Big-five	.118	.086	1.496	.118	.004
	Model 2	Big -five and Dark triad	.219	.173	1.423	.101	.001

Further detailed analysis was made using the model 2 regression for each individual personality variable, with the exception of Denial of responsibility, where only model one was significant (table 4). The strongest effect for explaining the use of the neutralisation technique Condemnation of the condemners was Machiavellianism; Metaphor of the ledger technique was best explained with high Neuroticism and low Openness. Machiavellianism was also the main factor explaining the use of Denial of injury technique. Low Openness, high Agreeableness and Machiavellianism were related to the use of Denial of responsibility. Machiavellianism was also highest in the Appeal to higher loyalties. Low Openness and high Machiavellianism were also the strongest predictors of the use of the Defence of necessity neutralisation technique.

**Table 4:** Standardised beta coefficients

	Condemnation of the condemners	Metaphor of the ledger	Denial of injury	Denial of responsibility	Appeal to higher loyalties	Defence of necessity
<b>Neuroticism</b>	.128	<b>.294**</b>	.126	.200t	.192t	.153
<b>Extraversion</b>	.028	.021	-.098	-.064	-.007	-.053
<b>Openness</b>	-.094	<b>-.264**</b>	-.051	<b>-.223*</b>	-.181t	<b>-.226*</b>
<b>Agreeableness</b>	.121	.059	.183	<b>.223*</b>	.085	.199
<b>Conscientiousness</b>	-.043	-.132	-.202t	-.069	-.015	-.050
<b>Machiavellianism</b>	<b>.446***</b>	-	<b>.268**</b>	<b>.245*</b>	<b>.309**</b>	<b>.376***</b>
<b>Narcissism</b>	-.116	-	-.028	.106	.051	.053

	Condemnation of the condemners	Metaphor of the ledger	Denial of injury	Denial of responsibility	Appeal to higher loyalties	Defence of necessity
Psychopathy	.011	-	.058	-.071	.006	-.033

<sup>t</sup>p<0.1, \*p<0.05, \*\*p<0.01, \*\*\*p<0.001

#### 4. Discussion

Our aim was to assess military cadet's attitude toward violation of security norms, and to assess the association between the tendency to violate these norms using neutralisation techniques, and personality (Big Five and Dark Triad).

Military cadets' attitude toward violations of security norms was rather strict. The majority of cadets (86.4 - 94.4 %) disagreed with justifying the violation of information security (disagreed or strongly disagreed). A small portion of the cadets were relatively neutral towards this kind of attitude, and only a very few considered it acceptable to justify information security violations if necessary. Even though the results are encouraging, as the majority considered following information security rules as important, there still exists a small portion of cadets (and future military officers) that either had neutral or somewhat positive attitude toward justifying information security violations. When it comes to security breaches, only one rule ignorer may be enough to jeopardize organizations' information security. Thus, it can be concluded that even if it may be impossible to change everyone's behaviour and attitude towards information security, it should still be our goal.

There are multiple underlying variables that may explain why some individuals are more prone than others to accept information security norm violations. Among these, one significant influencing factor that either predisposes or prevents maladaptive behaviour and affects attitudes is personality. In this survey, of all the Big Five traits, low Openness score and high Neuroticism score were best in explaining the use of neutralisation techniques. Previous research has found the relationship between Openness and security violations to be complex. For example, McBride and colleagues (2012) found impact of Openness to vary from protective to compromising, depending on individual self-efficacy or how they perceive level of threat and punishment. Persons with high Openness are often curious individuals who desire new experiences. It can be assumed that this type of personalities are not prone to justify violations with external reasons, even though there are evidence that they are more likely to ignore information security protocols in comparison to those with low Openness (Carter, Warkentin & McBride 2012). The effect of high Neuroticism in our survey is in line with previous research. Individuals with high Agreeableness, high Conscientiousness and low Neuroticism have been shown to be less likely to violate security norms (Johnston, Warkentin, McBride, & Carter. 2016). Persons with high Neuroticism have a tendency to have negative feelings, such as anxiety, more easily. These negative emotions and increased anxiety may result as a need to justify actions with external reasons. Of the Dark Triad, Machiavellianism was strongly related to the use of neutralisation techniques. The tendency to be manipulative and deceptive has common features with most of the neutralisation techniques, as they all are ways of diverting the blame away from oneself. Even though Psychopathy has in previous studies been linked to counterproductive work behaviour (Deshong et al. 2015), its link to security violations is unclear. It can be assumed that individuals with this tendency may lack the feelings of guilt and the need for social acceptance that would prompt the use of neutralisation techniques. Of individuals with high Narcissism score, which includes egotism, sense of grandiosity and lack of empathy towards others, it can be assumed that these persons expect themselves to have rules different from others, and thus there are fewer reasons to explain one's behaviour.

##### 4.1 Limitations

There are several limitations to be accounted for when interpreting the results of this study. Firstly, we only measured attitudes related to information security, not real actions, which is a common problem in this field of research. It is widely known that an intention to do something does not always lead to action; therefore, an intention to behave in a certain way may not predict true behaviour. Secondly, questions related to information security behaviour are prone to elicit socially accepted answers, which can limit the predictive power of the study. The third limitation is that this study is a cross-sectional study and causalities therefore cannot be accounted for.

## **4.2 Conclusion**

Most military cadets' attitudes towards information security were positive. There were only a few individuals prone to accept violation of security norms. However, it should be pointed out that also the individuals with neutral attitudes may pose a possible risk, as neutral attitudes may shift when benefits from violations are high and the likelihood of getting caught is low. These conditions are less likely to affect those with positive attitude towards information security.

We found a link between personality and the use of neutralisation techniques. Further research, however, is still needed. Nevertheless, the link between these two highlights the requirement for a more individualised information security education, which should acknowledge the various components that drive individual behaviour. We believe that personalised information security education is more beneficial than mass education. Understanding how employee's own personality can affect his/her information security behaviour should increase employee's awareness and thus drive towards a more secure behaviour.

## **References**

- American Psychological Association, APA (2019). Available from: <https://www.apa.org/topics/personality/> [Accessed 14<sup>th</sup> Jan 2019]
- Acuña, D.C. (2016) Effects of a comprehensive computer security policy on computer security culture. In: MWAIS 2016 Proceedings, Paper 10
- Furnell, S., & Clarke, N. (2012). Power to the people? the evolving recognition of human aspects of security. Computers & Security, 31(8), 983-988.
- Carter, L., Warkentin, M., McBride, M. (2012). *Exploring the role of individual employee characteristics and personality on employee compliance with cybersecurity policies*. RTI - Institute for Homeland Security Solutions.
- Deshong, H. L., Grant, D. M. & Mullins-Sweatt, S. N. (2015). Comparing models of counterproductive workplace behaviors: The Five-Factor Model and the Dark Triad. Personality and Individual Differences, 74, 55–60.
- Dhillon, G., and Moores, S. (2001) "Computer Crimes: Theorizing About the Enemy Within," Computers and Security (20:8), pp. 715-723.
- Hern, A. (2018) Fitness tracking app Strava gives away location of secret US army bases. The Guardian, 28th Jan 2018, Available from: <https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases> [Accessed 30th Dec 2018].
- Johnston, A. C., Warkentin, M., McBride, M., & Carter, L. (2016) Dispositional and situational factors: influences on information security policy violations. European Journal of Information Systems, 25(3), 231-251.
- Konstabel, K. , Lönnqvist, J. , Walkowitz, G. , Konstabel, K. and Verkasalo, M. (2012), The 'Short Five' (S5): Measuring personality traits using comprehensive single items. Eur. J. Pers., 26: 13-29.
- McBride, M., Carter, L., and Warkinten, M. (2012) Exploring the Role of Individual Employee Characteristics and Personality on Employee Compliance with Cyber Security Policies. (Prepared by RTI International – Institute for Homeland Security Solutions under contract 3-312-0212782.)
- Puhakainen, P. (2006) A Design Theory for Information Security Awareness, Oulu, Finland: University of Oulu.
- Siponen, M., Pahnila, S., & Mahmood, A. (2006) Factors influencing protection motivation and IS security policy compliance. Innovations in Information Technology, 2006, 1-5.
- Siponen M. & Vance A. (2010) Neutralization: New Insights into the Problem of Employee Information Systems Security Policy Violations. MIS Quarterly 34(3): 487–502
- Sykes G. & Matza D. (1957) Techniques of Neutralization: A Theory of Delinquency. American Sociological Review 22(6): 664–670.
- Stanton, J., Stam, K., Mastrangelo, P., and Jolton, J. (2005) "Analysis of End User Security Behaviors." Computers and Security (24:2), pp. 124-133.
- Rogers JW & Buffalo MD (1974) Neutralization techniques: toward a simplified measurement scale. Pacific Sociological Review 17(3): 313–331.

# Cyber Security Risk Modelling and Assessment: A Quantitative Approach

**Abderrahmane Sokri**

**DRDC CORA, Government of Canada, Canada**

[Abderrahmane.Sokri@drdc-rddc.gc.ca](mailto:Abderrahmane.Sokri@drdc-rddc.gc.ca)

**Abstract:** The extensive use of information systems has become both a crucial enabler and a critical vulnerability in *all spheres* of public and private activities. A cyber-security breach may result in interruption, modification, degradation, fabrication, interception, and unauthorized use of an information asset. The resulting damage can be immediate causing direct financial losses or prospective gradually harming the national safety and reputation. In the context of cyber security, risk is present where a threat intersects with a corresponding vulnerability which allows it to manifest. It can be formally expressed as a function of three elements: Probability that a threat may become harmful, Probability that a vulnerability may be exploited, and Resulting impact. While these three elements can be expressed either qualitatively or quantitatively, they are generally described in qualitative terms in the context of cyber security. This paper presents common cyber security risk assessment methods and shows how a risk analysis can be conducted in cyberspace. It proposes a new cyber risk formulation combining statistical and Monte Carlo simulation techniques. Risk analysis is defined here as a process that aims to identify, analyze, and reduce or transfer risk. A case study using the most common threats, vulnerabilities, and impacts is presented to illustrate the approach. In this study, a Program Evaluation and Review Technique (PERT) distribution is used to represent the inherent risk curve. A correlation analysis using stochastic simulation is conducted to show how sensitive the overall risk is to the different threats. Risk drivers are therefore assessed and displayed graphically using a pairwise association. The paper results and insights can assist civilian and military decision-makers in identifying critical risk drivers and the need for contingency plans. Statistical techniques and Monte Carlo simulation objectively derive the most likely cyber risk profile. The Loss Exceedance Curve shows for each loss the likelihood of exceeding it. The what-if analysis determines which risk mitigation strategies would have the most impact.

**Keywords:** cyber security, risk analysis, threat, vulnerability, impact

---

## 1. Introduction

A growing number of organizations rely on the Information and Communication Technologies (ICT). The extensive use of ITC, the growing interconnectivity, and the standardisation of communications have made the protection of the confidentiality, integrity, and availability of their connected systems and data a vital issue (Khan and Hussain, 2010; Branagan, 2012; Bernier et al., 2012; Aslanoglu and Tekir, 2012; Cherdantseva et al., 2016; Sokri, 2018; Yeo et al., 2019).

A loss of confidentiality refers to the improper disclosure of information. It may result, for example, in stealing corporate secrets and intellectual property. A loss of integrity is the unauthorized modification or destruction of information. It may result, for example, in misrouting logistics and improper financial transactions. A loss of availability is the disruption of access to or use of information or an information system. It may result, for example, in a loss of productivity (Hubbard and Seiersen, 2016; Aslanoglu and Tekir, 2012; Majumder, Mathur, Javaid, 2019).

Malicious intruders and skilled attackers may use a broad range of cyber means to have a combination of these effects. They may ultimately target specific vulnerabilities and cause significant damage to critical infrastructures (Bier et al., 2009). Critical infrastructures include (but are not limited to) (1) networked infrastructures (e.g., telecommunications networks, water supply), (2) information infrastructures (e.g., flight control, emergency broadcasting), and (3) physical infrastructures (e.g., hospitals, sports stadiums).

Damage inflicted by a successful attack on critical infrastructures could be immense (Mirkovic and Reiher, 2004). It can not only cause immediate financial losses but can also have long term effects (Melo and Gondim, 2012). It can pose serious threats to a country's safety (Acquaviva, 2017) and reputation (Aslanoglu and Tekir, 2012). Examples of such incidents include the infamous Stuxnet worm (Ghofouri et al., 2016, Yao et al., 2019). This unprecedented sophisticated cyber-attack attempted to destroy Iran's nuclear enrichment centrifuges by operating the motors at destructive speeds. Iran's centrifuges were potentially damaged and over 60,000 computers from different countries were infected by this Advanced Persistent Threat (Aslanoglu and Tekir, 2012).

Any form of possible destruction, degradation, manipulation, or exploitation against an information asset can be viewed as risk. In the context of cyber security, risk is present where a threat intersects with a corresponding vulnerability which allows it to manifest (Bowen et al., 2006). It can be formally expressed as a function of three elements (Schneidewind, 2009; Branagan, 2012; Abdel-Basset et al., 2019):

- Probability that a threat may become harmful,
- Probability that a vulnerability may be exploited, and
- Resulting impact.

In this context, a threat can be defined as a potential cause of an unwanted event which may result in harm to a system or organisation (Standards Australia, 2004; Branagan, 2012; Blyth and Kovacich, 2006, Abdel-Basset et al., 2019). The cyber threats include upgraded viruses, Trojan horses, worms, and advanced persistent threat (APTs). These threats are usually included as links or attachments in email messages (Sokri, 2018). A threat does not spontaneously occur. It needs an agent or a threat-source to manifest (Branagan, 2012). For example, a worm (the threat) occurs when an attacker (threat-source) can spread it from a computer to another and gain access to computers remotely.

A vulnerability is a weakness in an information system that may be exploited by a threat to make the system collapse (Zhang, 2012). A vulnerability may be caused by technology (e.g. no firewall), human behaviour (e.g., no password protection), and system settings (e.g., no proper privileges) (Martins et al., 2012).

Impact or consequence is the outcome of an event. It expresses the magnitude of its consequence in terms of loss, harm, injury and disadvantage (Standards Australia, 2004). A cyber-security breach may result in interruption, modification, degradation, fabrication, interception, and unauthorized use of an information asset. The information asset can be physically (e.g., hardware) or logically (e.g., software) based.

Interruption occurs when an information asset becomes unusable, unavailable or lost (e.g., denial of legitimate access to the asset). Modification occurs when an opponent tampers with the asset (e.g., unauthorised manipulation of information). Degradation means a slowdown in the rate of information delivery (e.g., decrease in the quality or precision of an asset). Fabrication is the counterfeiting of an asset (e.g., insertion of unauthentic transactions into a computer network). Interception is an unauthorised access to an asset (e.g., e.g., monitoring a computer network). Unauthorized use occurs when unauthorised party uses an asset for his own purpose (e.g., unauthorised use of software) (Blyth and Kovacich, 2006; Bernier et al., 2012).

While these three elements can be expressed either qualitatively or quantitatively, they are generally described in qualitative terms in the context of cyber security. A qualitative assessment uses a non-numerical classification as in Table 1 to provide a high level subjective judgment of the result of an event. Table 1 is adapted from Kim and Cha (2012), Ghanmi and Wong (2014), and Veerasamy et al. (2012). It represents the likelihood of threat, the likelihood of vulnerability, and the corresponding consequence using ordinal scaling techniques. A quantitative assessment assigns numerical values to these variables (Gallotti, 2019).

**Table 1:** Likelihood and impact assessment scales

Likelihood	Definition	Impact	Definition
5. Almost certain	Expected to occur in most circumstances	5. Severe	Failure to operate resulting in massive financial/technical damage
4. Likely	Will probably occur in most circumstances	4. Major	Shutdown of a critical business unit resulting in extensive financial/technical damage or loss of reputation
3. Possible	Could occur at some time	3. Moderate	Short interruption of a critical process or system resulting in a limited financial loss
2. Unlikely	Not expected to occur	2. Minor	Interruption with no financial loss.
1. Rare	Occurs in exceptional circumstances only	1. Insignificant	No significant effect on target

The aim of this paper is to show how a risk analysis can be conducted in cyberspace. Risk analysis is defined here as a process that aims to identify, analyze, reduce, or transfer risk (Melo and Gondim, 2012). A case study using the most common threats, vulnerabilities, and impacts is presented to illustrate the approach.

The paper is organized into five sections. Following the introduction, section 2 offers a succinct presentation of cyber security risk assessment methods. Section 3 presents a new risk analysis approach for cyber security. In section 4, a case study is presented to illustrate the suggested approach. Concluding remarks as well as a future research question are indicated in section 5.

## **2. Cyber security risk assessment methods**

There is a wide range of risk assessment methods applied in various areas including health, environmental and cyber domains (Cherdantseva et al., 2016; Rocchetto, Ferrari, and Senni, 2019). Methods for conducting cyber risk analysis vary from qualitative, through semi-quantitative, to quantitative methods, depending on data availability (Van Asselt, 2018). While qualitative methods use subjective reasoning to provide a high level judgment of cyber risk (Chew et al., 2008), quantitative methods strive to measure it numerically. Semi-quantitative methods use different combinations of qualitative and quantitative modeling.

### **2.1 Qualitative risk analysis methods**

Researchers primarily use qualitative methods to assess cyber risk (Pamula et al., 2006; Tang and Shen, 2009). Qualitative methods are a quick and cost-effective way of prioritizing risks. Risk is generally expressed in terms of descriptive variables and ordinal metrics such as Likert scale (Suh and Han, 2003). Qualitative risk analysis is an expert judgment-based approach that strongly depends on expert input. It elicits risk rankings from a group of experts when there are data gaps (Van Asselt, 2018). This approach can be separated along methodological lines into three main techniques: Delphi techniques, scenario analysis, and decision trees (Suh and Han, 2003; Van Asselt, 2018; White, 2019).

Delphi is a technique used for the elicitation of opinions of experts on a complex issue (Brown, 1968). The objective is to achieve the most reliable consensus of opinion of the group of experts (Rowe and Wright, 1999). Delphi can be used for nearly any complex problem involving forecasting, estimation, or decision making (Green et al., 2007). It consists of a series of intensive questionnaires (at least in two rounds) where a group of experts are anonymously asked to express and justify their opinions. They are also asked to revise their earlier answers in light of the feedback from other fellow respondents (Brown, 1968). The Delphi process is stopped when a consensus emerges. Four key features characterize the Delphi technique: anonymity, iteration, controlled feedback, and the aggregation of group response (Rowe and Wright, 1999; Dosumu, 2018).

The scenario analysis examines potential future events using alternative possible scenarios. A scenario is defined as a set of sequences of actions that may have an observable impact. Each scenario is defined by its optimistic, pessimistic, most likely values. This technique should normally provide appropriate strategies and countermeasures to prevent the expected cyberattacks (Kim and Cha, 2012). In the context of cyber security, the scenario analysis should take into consideration the threat scenarios, their targets, their impacts, and any actors that may reduce or exacerbate these impacts.

Decision tree is a straightforward method based on a set of clearly defined questions to obtain qualitative information about the risks associated with a given threat. The analyst generally asks experts to classify the risk as a low, medium, or high (Van Asselt, 2018). In cyber security, experts may be asked to identify threats and vulnerabilities, estimate their probability of occurrence, and recognize their impacts.

### **2.2 Semi-quantitative risk analysis methods**

Semi-quantitative approach uses different combinations of qualitative and quantitative methods. It mainly includes two methods: Multi criteria decision analysis (MCDA) and risk matrix.

MCDA methods solve problems involving multiple conflicting weighted criteria. They can be applied to a wide range of decision making problems to determine the impact of various criteria on an overall ranking. In the context of security risk ranking, the set of relevant criteria may include not only technical information (physical and logical Indicators) but also subjective elements such as expert judgment and acceptability to society. The results of this approach are strongly contingent on the weights of criteria. By varying the weights, the ranking and priorities change (Triantaphyllou, 2000).

Risk matrix is generally a two-dimensional graphical representation of risk. It visualizes both the likelihood of occurrence and the impact of the risk using different classes. Classes that could be used for the impacts and likelihood of occurrence are given in Table 1. Events with combinations of high likelihood and high impact are placed in the high risk corner and are likely to require further analysis. Events in the low risk corner would require less emphasis. They may just be included in a watch list for monitoring. In the context of cyber security, it may have three dimensions: Threat, vulnerability, and impact. The risk matrix method may be respectively quantitative, qualitative, or semi-quantitative, if the classification is numerical, non-numerical, or both (Sokri and Ghanmi, 2016; Hubbard and Seiersen, 2016; Van Asselt, 2018; Gonzalez-Granadillo et al., 2018).

### **2.3 Quantitative risk analysis methods**

Quantitative methods in risk assessment are mostly probabilistic. They typically use graph-based methods supported by simulation and mathematical models (Cherdantseva et al., 2016). For example, tree-based risk assessment methods, such as vulnerability tree and event tree, are used to determine the probability of a top undesirable event. A top event such as vulnerability existence represents a pivotal event for a particular failure scenario (Patel et al., 2008).

These methods may be split into two main categories: (1) inductive or forward search techniques and (2) deductive or backward search techniques. Inductive methods such as event tree trace from possible causes to undesired events whereas deductive methods such as attack trees trace from undesired events to possible causes (Taylor et al., 2002; Ralston et al., 2007). Quantitative risk assessment methods are powerful and long-established, but the unavailability of objective data limits their applicability, impedes their validation, and reduces the trustworthiness of their results (Cherdantseva et al., 2016). For more information on this topic, the reader is referred to Edgar and Manz (2017).

Many ways for how to quantify the overall cyber risk were suggested in the literature. The majority of the existing methods use the traditional step-by-step analysis to handle cyber risk. This bottom-up description places uncertainty on each threat and each vulnerability in the information asset. A build-up estimate is constructed, for each single vulnerability, for each threat that may exploit it, and for the corresponding consequence if the threat does occur, at the lowest level of detail. While this method provides detailed insights into risk contributors, it can be at times too time-consuming and expensive to implement (Schneidewind, 2009; Cherdantseva et al., 2016).

This paper suggests a threat driver method to derive the overall cyber risk for an information asset. This top-down approach of analyzing risk utilizes known prioritized risks from the risk register. Instead of placing uncertainty on each vulnerability and each threat, this approach applies the list of historical risks to the entire asset. A Monte Carlo simulation is conducted to generate a probability distribution that shows how a range of losses is possible during a given period of time. Technical details about this approach are highlighted and discussed in section 3.

## **3. Cyber risk assessment measures**

This section suggests a stochastic approach to portray the overall shape of cyber risk. The approach applies statistical techniques and Monte Carlo simulation to known risks to objectively derive the most likely cyber risk profile.

### **3.1 The overall risk**

The overall risk against an information asset  $a$  is generally calculated as an average loss (Cherdantseva et al., 2016). It can be expressed as

$$R_a = \sum_{t=1}^n p_t l_t, \quad (1)$$

where

- $p_t$  is the probability of a successful attack compromising the asset  $a$ ,
- $l_t$  is a loss resulting from the successful attack, and
- $n$  is the number of threats on the list of the identified risks (risk register).

The probability of a successful attack is derived as the ratio between the number of successful attacks and the total number of attacks against the asset.

### 3.2 Stochastic simulation

A threat driver-based stochastic simulation is conducted to derive the cyber risk profile. Instead of assigning fixed numerical values (single point estimates) to both the probability and impact of a given threat, as is standard in the existence literature, this approach uses a range of values within which the true value of the considered variable may lie. The technique involves simulating the losses of all possible threats that might occur. It also provides a means for conducting sensitivity analysis to determine key threat drivers for further investigation and risk mitigation.

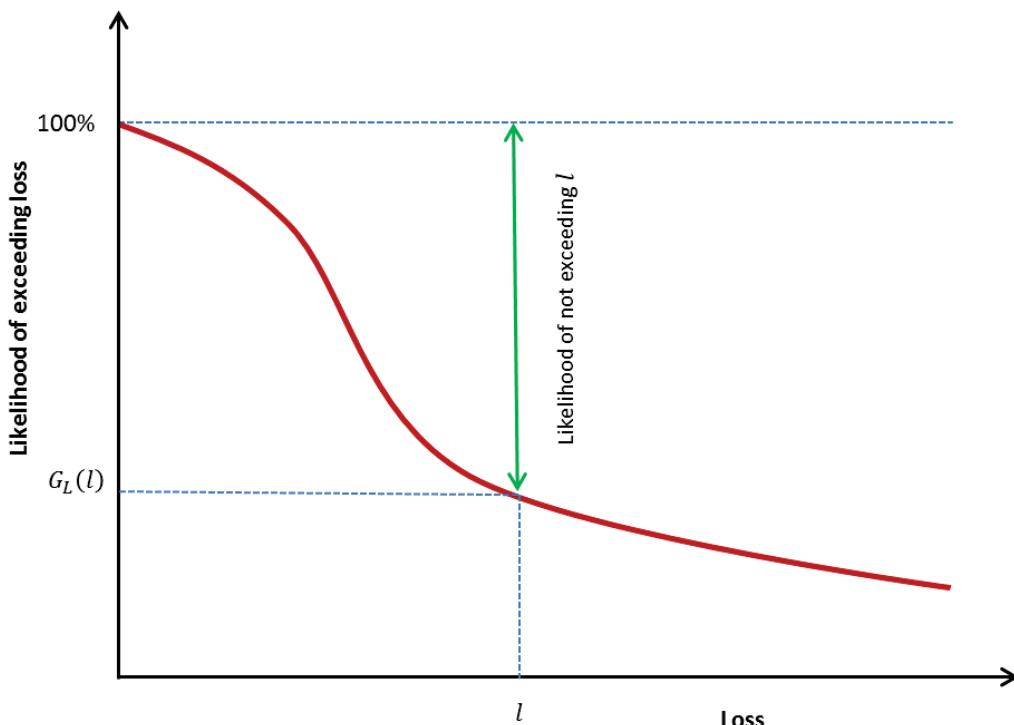
### 3.3 Risk profile

The cyber risk profile can be derived from the output of the stochastic simulation and presented using the Complementary Cumulative Distribution Function (CCDF) of the cyber risk. More commonly known as a Loss Exceedance Curve (LEC) or a probability of exceedance, this mathematical curve shows for each loss the likelihood of exceeding it. The concept of LEC is extensively used in financial portfolio risk assessment, actuarial science, and nuclear power (Hubbard and Seiersen, 2016). It is also known as the survival function in survival analysis and the reliability function in engineering. It allows decision-makers to determine a loss contingency, and identify key threats through sensitivity analysis.

To understand the concept of LEC and how it can be employed in the cyber context, let the loss  $L$  be a real-valued continuous random variable,  $F$  its Cumulative Distribution Function (CDF), and  $G$  its CCDF. As shown in equation 2, for each potential loss  $l$ ,  $G(l)$  is the probability of obtaining a value greater than  $l$ .

$$G_L(l) = 1 - F_L(l) = P(L > l). \quad (2)$$

An example of LEC adapted from Hubbard and Seiersen (2016) is provided in Figure 1. As shown in this graph, the lower the loss value, the higher is the likelihood of exceeding it (United States Air Force, 2007; Sokri and Solomon, 2014; Sokri and Ghanmi, 2017; Dreyer et al., 2018).



**Figure 1:** Prototype of a loss exceedance curve

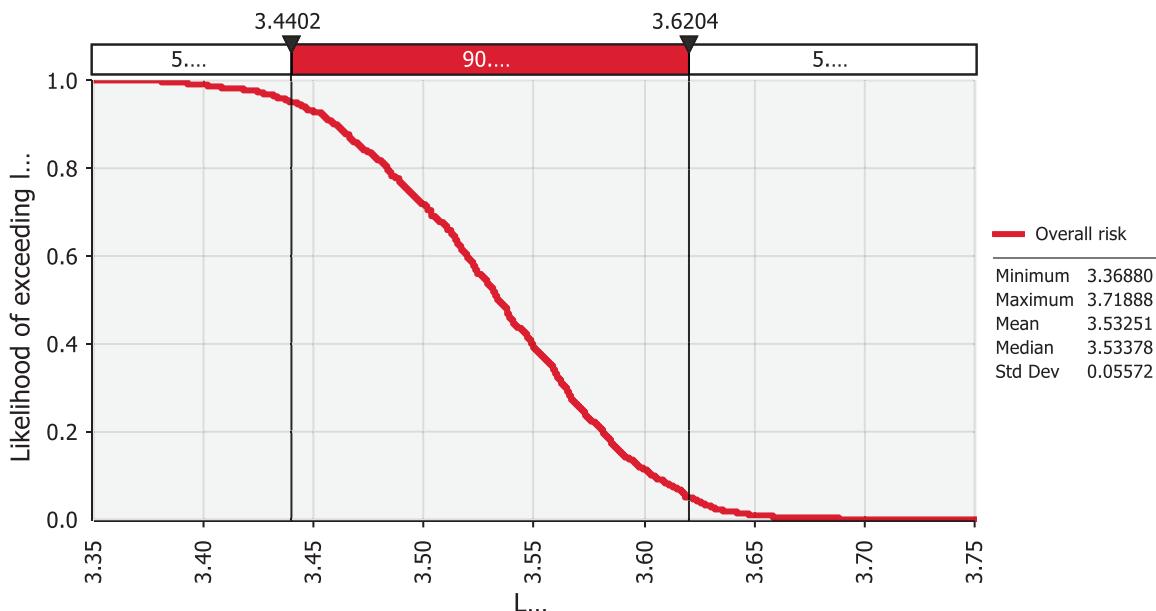
#### 4. Illustrative example

To illustrate the suggested approach, consider the set of threats against an information asset shown in Table 2. In this table, losses are given in millions of dollars. Notation and symbols are the same as in equation 1. This illustrative example outlines the potential impact of eleven categories of attacks on the asset. It shows, for example, that the probability of a successful Denial of Service (DoS) compromising the asset is approximately 10%, the damage inflicted by this attack is \$11 million, and the corresponding expected loss is around \$1.1 million. To avoid issues with sensitive information, we used illustrative data adapted from the existing cyber security literature (Schneidewind, 2009; Khan and Hussain, 2010).

**Table 2:** Cyber risk quantification

Threat	$p_t$	$l_t$	$p_t \cdot l_t$
Account Compromise	1.31%	7	0.0920
Corruption of Database	0.13%	1	0.0013
DoS	10.03%	11	1.1034
Packet Sniffer	2.93%	6	0.1761
Probe	12.41%	9	1.1172
Root Compromise	0.13%	5	0.0063
Scan	7.41%	8	0.5925
Spyware	1.59%	2	0.0319
Trojan Horse	1.92%	4	0.0770
Virus	3.10%	10	0.3100
Worm	0.81%	3	0.0243

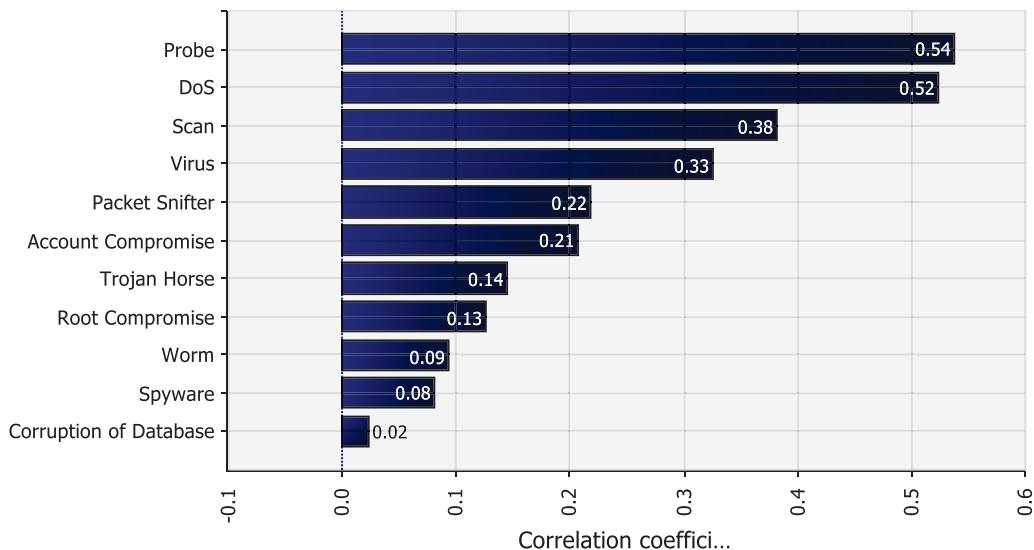
A Program Evaluation and Review Technique (PERT) distribution was used to represent the inherent risk curve. The inherent risk is the risk observed before any mitigation. The LEC in Figure 2 depicts the likelihood of exceeding a given amount of loss during a year. It shows for example that the probability of losing \$3.55 or more would be approximately 40%. It also indicates that the 90% confidence interval of the inherent risk is [\$3.44 million, \$3.62 million]. The median is the value separating the lower half of the probability distribution from the higher half. As the LEC is continuous, this value corresponds to the 50<sup>th</sup> percentile and amounts to approximately \$3.53.



**Figure 2:** The inherent risk curve

A correlation analysis using stochastic simulation is conducted to show how sensitive the overall risk is to the different threats. This what-if analysis can determine which risk mitigation strategies would have the most impact (Sokri and Ghanmi, 2017). In this technique, the pairwise association between the overall risk and each threat is estimated. Critical risk drivers are therefore identified and displayed graphically using the correlation coefficients. The correlation coefficient, for example, determines the strength and direction of this association.

The higher the correlation coefficient, the more significant the threat is in influencing the overall risk [3]. As shown in Figure 3, the most critical threat is Probe. This risk factor has the highest impact on the overall loss. The correlation coefficient between the overall project and this threat is 0.54. A strategy to mitigate this specific risk should be defined first.



**Figure 3:** Pairwise association between the overall risk and each threat

## 5. Conclusion

The extensive reliance on ITC has changed the face of the battlefield and the definition of security. In the context of cyber security, risk is present where a threat meets an exploitable vulnerability. The resulting impact can be immediate causing direct financial losses or prospective gradually harming the national safety and reputation.

The common methods for cyber risk analysis are presented and discussed in this paper. A new cyber risk formulation combining the probability of exceedance and simulation is proposed. A case study using the most common threats, vulnerabilities, and impacts is also presented and discussed to illustrate the approach. This approach can produce a more complete and accurate risk analysis particularly when including or excluding security controls.

Further efforts will be undertaken to use this approach to improve the Security Assessment and Authorization (SA&A) program. These efforts include, for example, the determination of the number of security controls to be used in the program by quantitatively comparing the inherent risk, the tolerance risk, and the residual risk.

## References

- Abdel-Basset M., Gu M., Mohamed M., and Chilamkurti, N. (2019) "A framework for risk assessment, management and evaluation: Economic tool for quantifying risks in supply chain", Future Generation Computer Systems, Vol. 90, pp 489-502.
- Acquaviva, J.R. (2017) "Optimal Cyber-Defence Strategies for Advanced Persistent Threats: A Game Theoretical Analysis", Master Thesis, the Pennsylvania State University, 2017.
- Aslanoglu, R. and Tekir, S. (2012) "Recent Cyberwar Spectrum and its Analysis", Proceedings of the 11th European Conference on Information Warfare and Security, Laval, France. pp 45-52.
- Bernier, M., LeBlanc S. and Morton, B. (2012) "Metrics Framework of Cyber Operations on Command and Control", Proceedings of the 11th European Conference on Information Warfare and Security, Laval, France, pp 53-62.
- Bier, V.M., Cox, L.A. and Azaiez, M.N. (2009) "Why Both Game Theory And Reliability Theory Are Important in Defending Infrastructure Against Intelligent Attacks" (Chapter 1). In: Game Theoretic Risk Analysis of Security Threats, Bier, V.M. and Azaiez, M.N. (eds) Springer: New York, pp 1–11.
- Blyth, A. and Kovacich, G. (2006) "Information Assurance", Springer-Verlag Ltd, London
- Bowen, P. Hash, J. and Wilson, M. (2006) "Information Security Handbook: A Guide for Managers". National Institute of Standards and Technology (NIST) Special Publication 800-100.
- Branagan, M. (2012) A risk simulation framework for information infrastructure protection. Ph.D. Dissertation, Queensland University of Technology, Australia.

### **Abderrahmane Sokri**

- Brown, B. (1968) "Delphi process: A methodology used for the elicitation of opinions of experts". Santa Monica: The RAND Corporation, Santa Monica, California.
- Cherdantseva, Y., Burnap, P., Blyth, A., Eden, P., Jones, K., Soulsby, H. and Stoddart, K. (2016) "A review of cyber security risk assessment methods for SCADA systems", *Computers & security*, Vol. 56, pp 1–27.
- Chew, E., Swanson, M., Stine, K., Bartol, N., Brown, A. and Robinson, W. (2008) "Performance measurement guide for information security", National Institute of Standards and Technology, NIST Special Publication 800-55.
- Dosumu O. (2018) "Assessment of the Likelihood of Risk Occurrence on Tendering and Procurement of Construction Projects". *Journal of Construction Business and Management*, Vol. 2, Issue 1, pp 20-32.
- Dreyer P, Jones T, Klima K, Oberholtzer J, Strong A, Welburn JW, Winkelman Z. (2018) "Estimating the Global Cost of Cyber Risk". RAND Corporation, Santa Monica, California.
- Edgar, T.W. and Manz, D.O. (2017) "Research Methods for Cyber Security". Syngress.
- Gallotti, C. (2019) "Information security: Risk assessment, management systems, the ISO/IEC 27001 standard". Ed. Lulu Press.
- Ghafoori, A., Abbas, W., Laszka, A., Vorobeychik, Y. and Koutsoukos, X. (2016) "Optimal Thresholds for Anomaly-Based Intrusion Detection in Dynamical Environments". Proceedings of the Decision and Game Theory for Security: 7th International Conference, New York.
- Ghanmi, A. and Wong, A. (2014) "Risk Management for Defence Major Delivery Projects: Canadian Perspective". TTCP TP4 Risk Management Workshop, 22-23 May 2014, Arlington, VA.
- Gonzalez-Granadillo G., Dubus S., Motzek A., Garcia-Alfaro J., Alvarez E., Merialdo M., Papillon S. and Debar H. (2018). "Dynamic risk management response system to handle cyber threats", *Future Generation Computer Systems*, Vol. 83, pp 535-552.
- Green, K.C., Armstrong, J.S. and Graefe, A. (2007) "Methods to Elicit Forecasts from Groups: Delphi and Prediction Markets Compared", *Foresight*, Issue 8.
- Hubbard, D.W. and Seiersen, R. (2016) How to Measure Anything in Cybersecurity Risk, Wiley, New York.
- Khan, M.A. and Hussain, M. (2010) "Cyber security quantification model", *Bahria University Journal of Information & Communication Technology*, Vol. 3, Issue 1.
- Kim, Y. and Cha, S. (2012) "Threat scenario-based security risk analysis using use case modeling in information systems", *Security and Communication Networks*, Special Issue Paper, Vol. 5, p. 293–300.
- Majumder S., Mathur A., Javaid A.Y. (2019) "Cyber-Physical System Security Controls: A Review". In: Guo S., Zeng D. (eds) *Cyber-Physical Systems: Architecture, Security and Application*. EAI/Springer Innovations in Communication and Computing. Springer, Cham, pp 187-240.
- Martins, J., Santos, H., Nunes, P. and Silva, R. (2012), "Information Security Model to Military Organizations in Environment of Information Warfare". Proceedings of the 11th European Conference on Information Warfare and Security, Laval, France, pp 172-179.
- Melo, L.P. and Gondim, P.R. (2012) "Risk Assessment and Real Time Vulnerability Identification in IT Environments" (Chapter 9). In: *Information Assurance and Security Technologies for Risk Assessment and Threat Management: Advances*, Chou, T. (eds) Idea Group Inc (IGI), pp 229-253.
- Mirkovic, J. and Reiher, P. (2004) "A taxonomy of DDoS attack and DDoS defense mechanisms", *SIGCOMM Computer Communication Review*, Vol. 34 (2), pp 39–53.
- Pamula, J., Ammann, P., Jajodia, S. and Swarup, V. (2006) "A Weakest-Adversary Security Metric for Network Configuration Security Analysis", 2nd Workshop on Quality of Protection, Alexandria, VA.
- Patel, S., Graham, J. and Ralston, P. (2008) "Quantitatively assessing the vulnerability of critical information systems: A new method for evaluating security enhancements", *International Journal of Information Management*, Vol. 28, p. 483–491.
- Ralston, P.A.S., Graham, J.H. and Hieb J.L. (2007) "Cyber security risk assessment for SCADA and DCS networks", *ISA Transactions*, Vol. 46, p. 583–594.
- Rocchetto M., Ferrari A., Senni V. (2019) "Challenges and Opportunities for Model-Based Security Risk Assessment of Cyber-Physical Systems". In: Flammini F. (eds) *Resilience of Cyber-Physical Systems. Advanced Sciences and Technologies for Security Applications*. Springer, Cham.
- Rowe, G. and Wright, G. (1999) "The Delphi technique as a forecasting tool: issues and analysis", *International Journal of Forecasting*, Vol. 15 (4).
- Schneidewind, N. (2009) "Systems and Software Engineering with Applications", Wiley-IEEE Press.
- Sokri, A. (2018) "Optimal Resource Allocation in Cyber-Security: A Game Theoretic Approach", *Procedia Computer Science*, Vol. 134 pp 283–288.
- Sokri, A. and Ghanmi, A. (2016) "Schedule Risk Analysis for Defense Acquisition Projects", Book Chapter in *Risk Management: Perspectives and Open Issues – A Multidisciplinary Approach*, edited by McGraw-Hill Education.
- Sokri, A. and Ghanmi, A. (2017) "Risk Analysis of Defence Acquisition Projects: Methods and Applications", DRDC Scientific Report DRDC-RDDC-2017-R124.
- Sokri, A. and Solomon, B. (2014) "Cost Risk Analysis and Contingency for the NGFC", DRDC CORA TM 2013-224.
- Suh, B. and Han, B. (2003) "The IS risk analysis based on a business model", *Information & Management*, Vol. 41(2), pp 149-158.
- Tang, X. and Shen, B. (2009) "Extending Model Driven Architecture with Software Security Assessment", Third IEEE International Conference on Secure Software Integration and Reliability Improvement (SSIRI), pp 436-441.

***Abderrahmane Sokri***

- Taylor, C., Krings, A. and Alves-Foss, J. (2002) "Risk Analysis and Probabilistic Survivability Assessment (RAPSA): An Assessment Approach for Power Substation Hardening". Proceedings of ACM Workshop on Scientific Aspects of Cyber Terrorism, Washington DC.
- Triantaphyllou, E. (2000) "Multi-criteria Decision Making Methods: A Comparative Study". Boston, MA, USA: Kluwer Academic Publishers.
- United States Air Force (2014) "Cost Risk and Uncertainty Analysis Handbook", Maryland, United States, 2007.
- Van Asselt, E.D., Van der Fels-Klerx, H.J., Raley, M., Poulsen, M., Korsgaard, H., Bredsdorff, L., et al. (2018). Critical review of methods for risk ranking of food-related hazards, based on risks for human health. *Critical Reviews in Food Science and Nutrition*, Vol. 58 (2), pp 178–193.
- Veerasamy, N., Grobler, M. and Von Solms, B. (2012) "Building an Ontology for Cyberterrorism", Proceedings of the 11th European Conference on Information Warfare and Security, Laval, France, pp 286-295.
- White R. (2019) "Risk Analysis for Critical Infrastructure Protection". In: Gritzalis D., Theocharidou M., Stergiopoulos G. (eds) *Critical Infrastructure Security and Resilience. Advanced Sciences and Technologies for Security Applications*. Springer, Cham, pp 35-54.
- Yao Y., Sheng C., Fu Q., Liu H., and Wang D. (2019) "A propagation model with defensive measures for PLC-PC worms in industrial networks", *Applied Mathematical Modelling*, In Press.
- Yeo, S., Sue Birch, A. and Jörgen Bengtsson, H.I. (2019) "The Role of State Actors in Cybersecurity: Can State Actors Find Their Role in Cyberspace?" (Chapter 2). In: *National Security: Breakthroughs in Research and Practice*, ed. Information Resources Management Association, pp 16-43.
- Zhang, J. (2012) "Information Security Risk Management Framework: China Aerospace Systems Engineering Corporation", Master Thesis, University of South Australia.

# Taxonomies of Cybercrime: An Overview and Proposal to be Used in Mapping Cyber Criminal Journeys

**Tiiia Somer**

**Tallinn University of Technology, Department of Software Sciences, Estonia**

[Tiiia.Somer@taltech.ee](mailto:Tiiia.Somer@taltech.ee)

**Abstract:** Is cybercrime different from „traditional crime“? Is there such a thing as cybercrime? There are several proposals for definition of cybercrime, and several analyses and proposals for a taxonomy. This paper presents an overview of currently available taxonomies and considers how useful these could be in mapping cyber criminal journeys. In order for individuals, enterprises and states to be informed about the motivations, methods and modi operandi of cyber criminals and thereby begin to protect themselves, it is useful to understand how a cyber crime takes place from its initial stages of motivation and intent, until exit. This paper uses the methodology of journey mapping as a basis for proposing a four-level taxonomy of cyber crime. This is significant because a common framework and taxonomy will help to more effectively analyse, as well as find targeted preventive and awareness-building measures in the fight against cybercrime.

**Keywords:** taxonomy, cybercrime, journey mapping

---

## 1. Introduction

In order to better understand cybercrime, its processes, and tactics, techniques and procedures used by cyber criminals, we need a robust framework. This paper reviews a number of well-known taxonomies and approaches used currently, and analyses if these can be used to understand cyber criminal journeys. Cyber crime is a vast and growing problem internationally. Cybercriminal revenues have reached 1,5 trillion USD annually in 2018 (McGuire 2018). The main goal of cyber criminals is to gain some kind of personal benefits: often times it is financial, but not only. Other goals might be curiosity, thrill, distribution of illegal material, harassment, influencing populations, espionage, stealing of intellectual property, etc.

## 2. Existing taxonomies

Currently there does not exist a universally accepted definition of cyber crime, and this is reflected in the taxonomies that have been developed. Some of the definitions and classifications follow the logic used for “traditional crimes”, others use approaches based on technology, adversaries or threats, for yet others law is the basis. Common points do exist among these taxonomies, but there are differences in structure, definitions and content. There is a clear distinction between the approaches using technical nature of cyber crime as basis on one hand and those using impacts of crime as basis on the other hand. Researchers have studied aspects of crime, such as technical (e.g. malware types and prevention of these) aspects of execution of crimes, or human aspects (human error, victimisation, etc.), but understanding of cyber crime as a process and system remains less developed. The current approaches to studying cyber crime are based on law, impacts, victims, methods and attack vectors, criminals/ attackers, motivations of perpetrators and criminal gain.

Current taxonomies and approaches to building a taxonomy are based on one of the following:

- Approaches based on criminology
- Approaches based on technologies, adversaries and threats
- Two dimension taxonomies
- Three dimension taxonomies
- Proposals by international bodies

### 2.1 Approaches based on criminology

Approaches based on criminology are built on traditions of criminal justice and they see cyber crime as regular crime, where computers are a tool which are enablers for “traditional” crimes. Brenner (2006) states that the definition of cybercrime as “a crime committed on a computer network” should reflect existing legal frameworks, both national and international. Brenner also discusses differences between cybercrime and cyberterrorism, and notes that crimes are in general committed for personal gain or personal revenge. Terrorism

has different motives, but may use the same methods, and the same applies to cyber warfare where the techniques used may be identical to cybercrime.

**Table 1:** Taxonomy based on criminology

Wall (2007)	Three different types of crime	- Traditional crimes, committed through use of ICT - Partially new crimes, modified crimes - New crimes, enabled by ICTs
-------------	--------------------------------	--

## 2.2 Approaches based on technological aspects, attackers, attack vectors or threats

A different approach has technological aspects, attackers, attack vectors or threats as a basis, or the aspects of computer and network infrastructure as targets. These proposals use the nature of computer security flaws as a basis, look at existing data on security incidents, look at adversaries and attacks, or impacts.

**Table 2:** Taxonomy based on technological aspects, attackers, attack vectors or threats

Landwehr et al (1994)	The nature of computer security flaws as a basis	- Flaws by genesis (how the flaw arises) - Flaws by time of introduction - Flaws by location (hardware and software)
Howard (1997); Howard and Longstaff (1998)	Based on existing data on security incidents	- Attackers (hackers, criminals, terrorists, vandals) - Tools (scripts, toolkits, user commands) - Access (implementation or design vulnerabilities, access permissions) - Results (corruption, deletion or disclosure of data, theft of resources, denial of service) - Objectives (intellectual challenge, peer status, financial gain, damage)
Hansman and Hunt (2005)	Four-category model	- Attack vectors (the means by which the target is reached) - Targets (hardware, software, network, data) - Specific vulnerabilities and exploits (security flaws) - Payload (the outcome and effects)
Kjaerland (2005, 2006)	Built on earlier work and added a quantitative component	- Source sectors (top level domains) - Method of operation (resource theft, social engineering, malware, denial of service) - Impact (disruption, distortion, destruction, disclosure) - Target services (commercial or governmental)
Meyers et al. (2009)	Based on attack vectors	- Viruses; - Worms; - Trojans; - Buffer overflows; - Denial of service; - Network attacks; - Physical attacks; - Password attacks/user compromise; and - Information gathering
Simmons et al. (2014)	AVOIDIT taxonomy	- Attack Vector - Operational Impact - Defence - Information Impact - Target
Rogers (1999, 2001, 2006)	Adversaries	- Script kiddies, newbies, novices; - Hacktivists, political activists; - Cyberpunks, crashers, thugs; - Insiders, user malcontents; - Coders, writers; - White hat hackers, old guard, sneakers; - Black hat hackers, professionals, elite; - Cyberterrorists

### 2.3 Two-dimension and three-dimension taxonomies

Some authors have proposed two dimension taxonomies to classify cybercrime into two categories, and others have proposed three-dimensional taxonomies of cybercrime.

**Table 3:** Two-dimension and three-dimension taxonomies

Furnell (2001), Koenig (2002), the Australian High Tech Crime Centre (2003), Lewis (2004), and Wilson (2008), Foreign Affairs and International Trade of Canada (2004)	Two-dimensional classification of crimes	- crimes committed using computers and networks (hacking, viruses); - traditional crimes that are facilitated by the use of computers (illegal pornography, online fraud)
Alkaabi et al. (2010)	Two-dimensional classification of crimes	- Type I crimes, where the computer, computer network, or electronic device is the target of the criminal activity - Type II crimes, where the computer, computer network, or electronic device is the tool used to commit or facilitate the crime
Wall (2007)	Three- dimensional classification of crimes	- Computer integrity crimes (hacking, cracking and denial of service attacks, activities that prevent legitimate access to systems or modify, corrupt or delete software and data). - Computer-assisted crimes (virtual robberies, scams and thefts). - Computer content crimes (digital storage and communication of pornography, violence and offensive materials)
Goodman (1997)	Three types of cybercrime	- crimes in which the computer is the end target - crimes where the computer is the tool - crimes where there is an incidental presence of computer equipment
Ghernaouti (2013)	Distinguished cybercrime from cyberconflicts, wars and terrorism	- Cybercrimes against people (including activities affecting their dignity and integrity, frauds, identity crimes and privacy related offences); - Cybercrimes against assets (including the theft of data, the theft of services and resources, counterfeiting, software piracy, surveillance and espionage, manipulation of information, fraudulent acquisition of intellectual property); and - Cybercrimes against states (including destabilization, information warfare, and attacks on critical infrastructures)

### 2.4 Taxonomies proposed by international bodies

Another set of taxonomies are those proposed by international bodies. These are significant because of their visibility and influence in shaping policies and providing a framework for legislation. The best known among these are the Council of Europe Convention on Cybercrime, The UN Manual on the prevention and control of computer related crime (1999) and a three domain taxonomy proposed by INTERPOL.

**Table 4:** Taxonomies proposed by international bodies

Council of Europe Convention on Cybercrime		- Offences against the confidentiality, integrity and availability of computer systems and data - Computer related offences (forgery, fraud) - Content related offences - Offences related to infringements of copyright and related rights - Dissemination of racist and xenophobic material, and threats and insults motivated by racism or xenophobia, through computer systems
--	--	--

The UN Manual on the prevention and control of computer related crime (1999)	Proposed to address the problems of international cooperation in computer crimes and criminal law	<ul style="list-style-type: none"> <li>- Fraud by computer manipulation</li> <li>- Computer forgery</li> <li>- Damage to or modification of computer data or programs</li> <li>- Unauthorised access to computer systems and services</li> <li>- Unauthorized reproduction of legally protected computer programs</li> </ul>
INTERPOL	Three domain taxonomy	<ul style="list-style-type: none"> <li>- Attacks against computer hardware and software (botnets, malware and network intrusion)</li> <li>- Financial crimes (online fraud, penetration of online financial services and phishing)</li> <li>- Abuse (especially of young people, in the form of grooming or 'sexploitation')</li> </ul>

The review of taxonomies and approaches shows that while there is agreement on the importance and growing significance of cybercrime, there is no consensus on a common definition or taxonomy.

Some proposals for the definitions and classifications follow the logic used for “traditional crimes”, others use approaches based on technology, adversaries or threats, yet others use law to base their proposals on. The main difference in these taxonomies arises from the basic definition: what is cyber crime? Some authors consider any crime, involving computers at any stage, a cyber crime (pure cyber crimes, cyber-enabled and cyber-dependent crimes), others consider only “pure cyber crimes” or “cyber-dependent” crimes as cyber crimes. There are common points in all of these taxonomies, but the structure, definitions and content differ. There is a clear distinction between the approaches which use technical nature of cyber crime as basis on one hand and those using impacts of crime as basis on the other hand. Researchers have studied aspects of crime, such as technical aspects (e.g. malware types and prevention of these) of execution of crimes, or human aspects (human error, victimisation, etc.), but understanding of cyber crime as a process and system remains less developed.

The cybercrime ecosystem is dynamic, and constantly evolving (McGuire 2018), therefor it is not easy to apply any of the existing taxonomies to thoroughly understand cybercrime.

### **3. Cyber criminal journey mapping**

Cyber crimes can be seen as a process where resources are required and decisions are made, which together constitute the modus operandi of a crime. In order to understand cyber criminal modi operandi, we have in our previous work undertaken an exercise to map cyber criminal journeys – how a crime is conducted, from the criminal view. Journey mapping is a method which utilises methods from various disciplines – criminology, information security, military science, economics – with the aim to understand how a crime takes place. In our work, we have explained a crime in four principal stages: intent/ motivation; preparation; execution; and monetization and exit. This work is significant as it allows us to understand cyber crime as a process, and can provide different stakeholders involved in investigation, fight against, and prosecution of cyber crimes a common platform for understanding cyber crimes.

In our effort to model cyber criminal processes we have used phase-based approach, customer journey mapping and crime scripting. Phase-based approach has been used in the armed forces (U.S. Department of Defence, 2007) to find capability shortfalls and make decisions on the development of new capabilities (Tirpak, 2000). Mapping is also used in business to map customer satisfaction, and crime scripting is used in criminology to map criminal journeys throughout a crime cycle. Given the complex nature of cyber crime, it is valuable to gain deep understanding of the mechanics of it. We analysed several crimes and took them apart, thereafter made generalisations for similar types of crime (DDOS, Malware development, DRM cracking, VoIP attacks, extortion, espionage, IP theft, crypto cracking). In essence, our mapping used attack vectors, providing a sequence of actions making up a crime event, generalised from a number of known crimes.

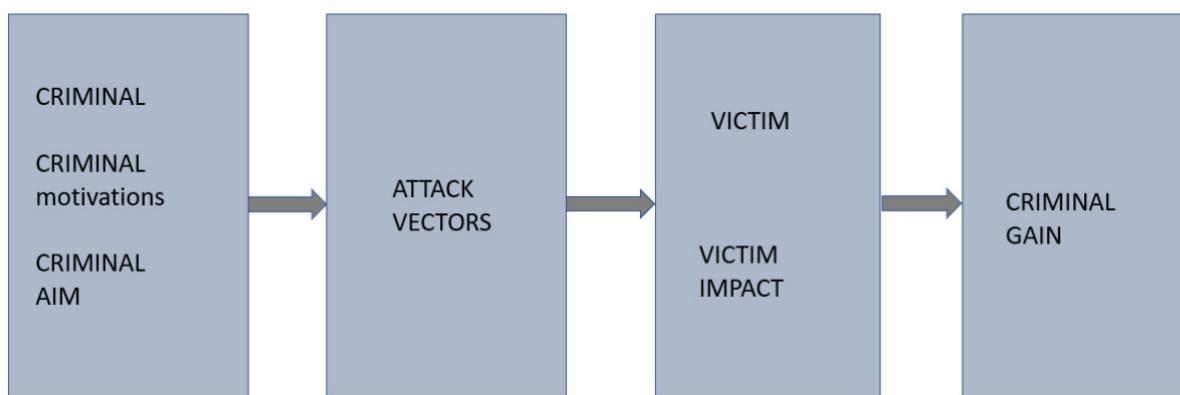
In analysing different types of cyber crime, it has been difficult to come up with a usable taxonomy that would let us best break the crime down to small parts. The schematic representation of sequence of actions provides us with a cognitive representation of how we believe a sequence of events will occur (Abelson 1981, Borron 2013), e.g. the steps a criminal makes in an attempt to commit cybercrime. Such scripting can be used to present different crimes, but are believed to be of particular use for new or complex crimes (Brayley et al 2011). Levi and Maguire (2004) have suggested, in their review of organised crime reduction strategies, that such scripts could be used as an innovative way to get a more detailed understanding complex new crimes. Our initial analysis of

cybercrime was based on attack vectors. Having modelled these, it became apparent that the steps criminals pass in different phases for the different attack vectors do not differ between themselves too much. In our modelling it became evident that there are four main aspects that need to be considered: criminal, attack vectors, victim, and criminal gain. Also it became clear that there is a need for a taxonomy.

#### 4. Proposed taxonomy for mapping cyber criminal journeys

The main difficulty in creating a taxonomy of cyber crime is the definition of crime, which is a legal issue. Crimes exist when they are identified as such in legislation, and this is done in specific national contexts. Cyber crime is an international phenomenon, and this makes its study even more difficult and creates additional problems of terminology and taxonomy.

In our approach, we have not considered cyber-enabled crime as cybercrime, rather such crimes are enablers of cybercrime. We have defined cybercrime as “pure” and “cyber-dependent” cyber crimes, i.e. such crimes, that would not happen without the use of ICTs. The aim of mapping cyber criminal journeys is to understand how a crime takes place from the criminal’s point of view. Important aspects in this process are the criminal, his/ her motivations, victim, impacts on him/her, attack vectors and methods, and criminal gain, or exit from crime journey. Schematically this can be represented as follows:



**Figure 1:** Actors and aspects in a cybercriminal cycle/ journey

A robust taxonomy is seen as an essential starting point to addressing these questions. The need for a taxonomy to respond to cyber crime is a practical measure: without an understanding of cyber crime, meaningful responsive measures cannot be developed (Moitra 2005). Any taxonomy created can not be a static or complete document. This is so due to the connected digital world that we live in, the ever-evolving nature of cyber crime, ever-evolving cyber criminal ecosystem, and the varied legal frameworks.

The principal objective in conducting cyber criminal journey mapping is to understand how a cyber crime takes place. Therefore we propose to use an appropriately generic taxonomy, that can be further sub-divided as the processes, actors, tactics, techniques and systems change. After researching various crime types, the cyber criminal ecosystem, victimology and cyber criminal business models, we propose a four-dimensional taxonomy:

- Perpetrator, including their motivation and aim, business models, ecosystem and preparation to conduct the crime;
- Attack vector, including enabling capabilities to conduct the crime;
- Victim, including the impact of crime on victim;
- Exit, including monetization of crime.

The basis for the proposed taxonomy is the underlying crime cycle, or crime journey, i.e. the stakeholders (perpetrator and victim), capabilities (attack vectors, but also preparation for crime), and enablers (monetization of crimes, but also the cyber criminal ecosystem). This classification covers all aspects in a crime and each can be further sub-divided as depicted on Figure 2 and explained below.

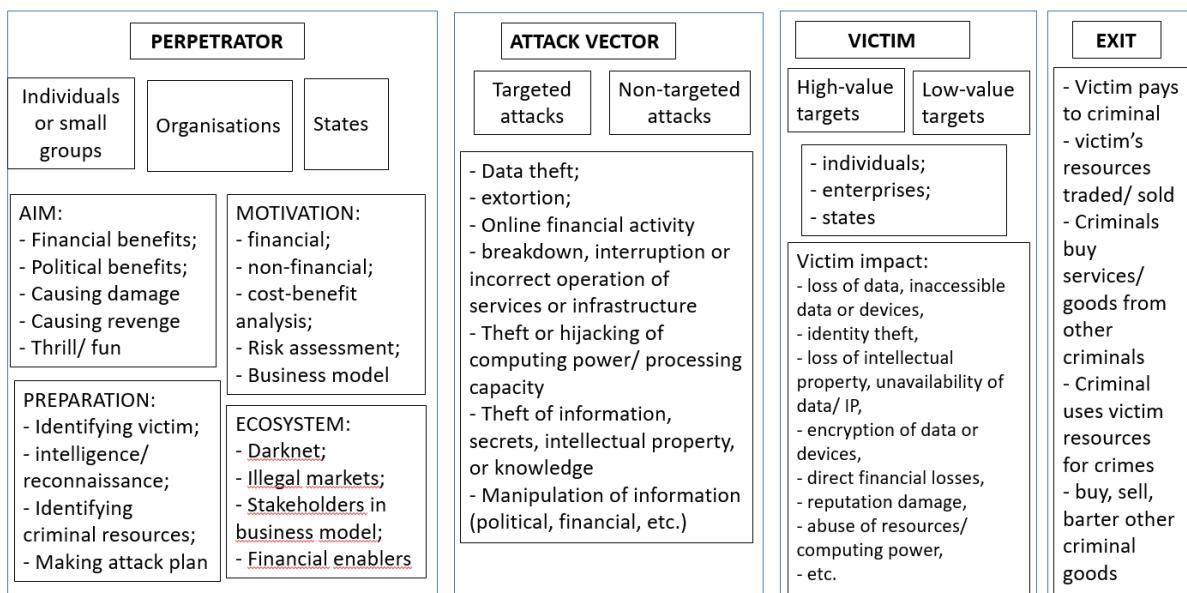


Figure 2: Four-dimensional taxonomy proposed for cyber criminal journey mapping

#### 4.1 Perpetrator

In quite short time cyber crime has moved from low volume crime performed by individuals to being high volume crime, which is organised and industry-like (UN 2013). Published evidence on cyber crime is based on a small number of case studies or interviews, and mostly focus on the methods of the crime (McGuire, Dowling 2013). Research from five or more years ago suggested that the cyber criminal world was not highly organised, however research published within the last few years suggests otherwise. Now there are estimates that more than 80% of cybercrime acts use some kind of organised activity (UN 2013, Broadhurst et al 2014). The UN cyber crime study states that cyber crime often requires a “high degree of organisation to implement, and may lend itself to small criminal groups, loose ad hoc networks, or organised crime on a larger scale”, and that the typology of cyber criminal groups reflect that of the conventional world. It also notes that many cyber crime acts require high levels of organisation and considers it likely that conventional organised crime groups are also active in cyber crime. (UN 2013).

Bearing this in mind, cyber criminals can be broadly classified into four general types of players and/or groups, often interacting amongst each other:

- 1. Individuals and small criminal groups (opportunistic or planned attacks),
- 2. International criminal organisations,
- 3. States (i.e. foreign intelligence agencies),
- 4. Legitimate organisations.

##### 4.1.1 Perpetrator motivations

Criminal motivation, or intent, is where the potential criminal makes a conscious decision to engage in a criminal act. In spite of it being a conscious decision, it doesn't have to be well-reasoned, rational, or completely thought through; and timeframes for decision-making are very short. Having said that, there is an element of rationality, with cost-benefit analysis, risk assessment, etc. interplaying with the decisions to commit a crime. In general terms, perpetrator motivations can be:

- 1. Satisfaction, peer-approval, publicity, status, revenge, etc.,
- 2. Challenges, having fun,
- 3. Organisational/ community obligations (hacktivism, working on behalf of a government organisation or individuals acting on a sense of pride or obligation),
- 4. Direct or indirect monetary gain.

Based on the above, we propose the following classification for perpetrators:

- Actor (individual, small groups, international (criminal) organisations, states),
- Aim (get financial benefits, get political benefits, cause damage, cause revenge, thrill/ fun,
- Motivation (financial or non/financial, cost-benefit analysis, risk assessment),
- Preparation for conduct of crime (identifying victim, intelligence/ reconnaissance, identifying criminal resources, making attack plan)

#### **4.2 Victim**

People and organisations/ enterprises usually fall victim to cyber crime through either their actions or non-actions. One can fall victim during the course of their use of information technology, i.e. using e-mail, browsing the web, using removable media, or by not taking appropriate action to keep their systems or use of systems secure, i.e. bad password management; not updating or poor management of systems, software or hardware. Once falling victim to cyber crime, the victim will face damages.

The impacts of, or damages from, cybercrime can be loss of data, hijacked accounts, loss of intellectual property, unavailability of data/ IP, inaccessible data or devices, encryption of data or devices, identity theft, direct financial losses, reputation damage, abuse of resources, abuse of computing power, etc.

Victims can be either high or low value targets. Adopted from Ghernaouti (2013), victims of cybercrime can broadly be classified into three general categories:

- 1. individuals (identity crimes, privacy related offences, direct financial crimes, extortion, etc);
- 2. enterprises (theft of data or intellectual property, theft of services and resources, counterfeiting, economic espionage, etc;
- 3. states (destabilization, information or cyber warfare, attacks on critical infrastructures).

#### **4.3 Attack vectors**

Typical attacker modus operandi is gaining or blocking access to a victim's data, device, or functionality (European Police Force, 2014). The overview of current taxonomies and approaches to cybercrime presented earlier in this paper, show two approaches in presenting attack vectors: 1) concrete technical means of conducting attacks (viruses, worms, Trojans, denial of service, network attacks, user compromise, etc.); 2) general attack methods (unauthorised access, malicious codes, interruptions of services, theft or misuse of data/ devices).

For discussing attack vectors in the journey mapping context we propose to use a general approach of attacks (such as theft of data, extortion, etc.), which can be further sub-divided into enablers (such as malware, botnets, ransomware, etc). Attacks can further be subdivided to be either targeted or non-targeted attacks. The attack vectors proposed are the following:

- Data theft
- Extortion
- Online financial activity
- Breakdown, interruption or incorrect operation of services or infrastructures
- Theft or hijacking of computing power/ processing capacity
- Theft of information, secrets, intellectual property, or knowledge
- Manipulation of information (political, financial, etc)

#### **4.4 Exit strategies**

In general, the gain from cybercrime can be divided into two: financial gain or non-financial gain. An exit strategy is a "means of leaving one's current situation, either after a predetermined objective has been achieved, justifying premises or decision makers for any given operational planning changed substantially, or as a strategy

to mitigate imminent or possible failure". In the case of financially motivated crime, monetization as a separate phase will come into the picture. In cases of hacktivism, or state-sponsored actions, the monetization phase is (often, but not always) excluded.

A financially-motivated attacker must decide who and what to attack, attack successfully and then monetize access. In general, there are five options to generate income:

- Victim pays to criminal directly (e.g. extortion);
- Victim's resources are turned to tangible assets (i.e. victim's resources will be sold and traded);
- Criminal pays for goods and/or services to another criminal (real money, cryptocurrency, re-sellable money equivalents, or goods and services);
- Criminal gets access to victim resources and uses these for other (criminal) actions;
- Buying, selling or bartering other (sometimes illegal) goods and services.

In terms of cybercrime, an exit strategy will let the criminal decide when to leave the crime scene, either for the risks or costs outweighing the benefits, for not being able to make enough profit, or simply for risks of getting caught by law enforcement authorities. The exit strategy will also cover hiding one's tracks and analysing law enforcement agencies' abilities to infiltrate crime rings or their capabilities of investigating crimes.

## **5. Future work**

Future work to validate this taxonomy will be undertaken in the form of case studies with real-life cyber crime cases in the second half of 2019, in cooperation with police forces from Estonia, Germany and the U.K.

## **6. Conclusion**

Cybercrime as being a vast and growing problem is acknowledged widely. Even so, the understanding of cybercrime as a complete system or process is not studied in detail, and to a big extent this is influenced by the lack of appropriate taxonomy. Any taxonomy to be used in mapping cyber criminal journeys should be appropriately generic and based on stakeholders and actions interacting within a cyber crime cycle. Therefor we propose a four-dimension taxonomy:

- Perpetrator, including their motivation and aim, business models, ecosystem and preparation to conduct the crime;
- Attack vector, including enabling capabilities to conduct the crime;
- Victim, including the impact of crime on victim;
- Exit, including monetization of crime.

The basis for the proposed taxonomy is the underlying crime cycle, or crime journey, i.e. the stakeholders (perpetrator and victim), capabilities (attack vectors, but also preparation for crime), and enablers (monetization of crimes, but also the cyber criminal ecosystem). This classification covers all aspects in a crime and each is further sub-divided.

We believe that this generic taxonomy will allow us to analyse and develop a better understanding of cybercrime as a process and system, where criminals and victims interconnect with each other and where attack vectors, enablers and exit strategies from crime are analysed in a systematic context. This is significant because it would help develop an understanding of how cybercrime business processes, but also tactics, techniques and procedures utilised by criminals function. This could potentially lead to identifying pinch points in the cybercrime processes, find better countermeasures, and develop novel policy or technical approaches in the fight against cybercrime.

## **References**

- Alkaabi A, Mohay G, McCullagh A, Chantler A. (2010). Dealing with the problem of cybercrime. Conference Proceedings of 2nd International ICST Conference on Digital Forensics & Cyber Crime. Abu Dhabi
- Australian High Tech Crime Centre (AHTCC). (2003). Fighting the Invisible. Platypus Mag J Aust Fed Police. 2003;80:4–6
- Brenner SW. (2006). Cybercrime, cyberterrorism and cyberwarfare. Rev Int Droit Penal. 77(2006/3):453–71
- Borrión, H. (2013). "Quality assurance in crime scripting", Crime Science 2013, 2:6.  
<http://www.crimesciencejournal.com/content/2/1/6>

- Brayley, H, Cockbain, E, Laycock, G. "The value of crime scripting: Deconstructing Internal Child Sex Trafficking", *Policing*, Volume 5, Number 2, pp. 132–143
- Broadhurst, R., Grabosky, P., Alazab, M. and Chon, S. (2014). Organizations and Cyber crime: An Analysis of the Nature of Groups engaged in Cyber Crime
- Clarke, Ronald R. (ed.). (1997). *Situational Crime Prevention: Successful Case Studies*. Second Edition. New York: Harrow and Heston
- Council of Europe Convention on Cybercrime (2001).
- Foreign Affairs and International Trade Canada. (2004). Available from: <http://www.dfaid-maeci.gc.ca/internationalcrime/cybercrime-en.asp>
- Furnell S. (2001). The Problem of Categorising Cybercrime and Cybercriminals. 2nd Australian Information Warfare and Security Conference. Perth, Australia; 2001. p. 29–36
- Ghernaouti S. (2013). *Cyberpower: Crime, Conflict and Security in Cyberspace*. EPFL Press
- Goodman M. (1997). Why the Police Don't Care about Computer Crime. *Harv J Law Technol.* 1997;10(3):465–94
- Hansman S, Hunt R. (2005). A taxonomy of network and computer attacks. *Comput Secur.* 2005;(21):31–43
- Howard J. (1997). An analysis of security incidents on the internet, 1989–1995. Carnegie Mellon University; Available from: <http://www.cert.org/archive/pdf/JHThesis.pdf>
- Howard J, Longstaff T. (1998). A common language for computer security incidents. Sandia National Laboratories; 1998. Report No.: Technical Report SAND98- 8667. Available from: [http://www.cert.org/research/taxonomy\\_988667.pdf](http://www.cert.org/research/taxonomy_988667.pdf)
- Interpol (2014). Available from: <http://www.interpol.int/Crime-areas/Cybercrime/Cybercrime>
- Kshetri N. (2006). The Simple Economics of Cybercrimes. *IEEE Secur Priv.* 2006;4(1):33–9
- Landwehr C, Bull A, McDermott J, Choi W. (1994). A taxonomy of computer program security flaws, with examples. *ACM Comput Surv.* 1994;26(3):211–54
- Kjaerland M. (2005). A classification of computer security incidents based on reported attack data. *J Investig Psychol Offender Profiling.* 2005;(2):105–20.
- Kjaerland M. (2006). A taxonomy and comparison of computer security incidents from the commercial and government sectors. *Comput Secur.* 2006;(25):522–38
- Koenig D. (2002). Investigation of Cybercrime and Technology-related Crime.
- Levi, M. and Maguire, M. (2004). "Reducing and Preventing Organised Crime: An Evidence-Based Critique", *Crime, Law & Social Change*
- Lewis B. (2004). Preventing of Computer Crime Amidst International Anarchy [Internet]. 2004. Available from: [http://goliath.ecnext.com/coms2/summary\\_0199-3456285\\_ITM](http://goliath.ecnext.com/coms2/summary_0199-3456285_ITM)
- McGuire, M. (2018). Into the Web of Profit. Understanding the Growth of Cybercrime Economy. Bromium
- Meyers C, Powers S, Faissol D. (2009). Taxonomies of Cyber Adversaries and Attacks: A Survey of Incidents and Approaches. Lawrence Livermore National Laboratory; 2009 Apr. Report No.: LLNL-TR-419041
- Moitra S. (2005). Developing Policies for Cybercrime. *Eur J Crime Crim Law Crim Justice.* 2005;13(3):435–64.
- Rogers M. (1999). A new hacker taxonomy. University of Manitoba.
- Rogers M. (2001). A social learning theory and moral disengagement analysis of criminal computer behavior: an exploratory study. University of Manitoba.
- Rogers M. (2006). A two-dimensional circumplex approach to the development of a hacker taxonomy. *Digit Investig.* 2006;(3):97–102
- Simmons C, Shiva S, Bedi H, Dasgupta D. (2014). AVOIDIT: A Cyber Attack Taxonomy. Proceedings of the 9th Annual Symposium on Information Assurance (ASIA '14). Albany, NY, USA; 2014.
- Tirpak, J. (2000). Find, Fix, Track, Target, Engage, Assess. Air Force Magazine, 83:24–29, 2000. URL <http://www.airforce-magazine.com/MagazineArchive/Pages/2000/July%202000/0700find.aspx>
- United Nations manual on the prevention and control of computer-related crime (1999). United Nations
- United Nations Office on Drugs and Crime (2013). Comprehensive Study on Cybercrime
- U.S. Department of Defense (2006). Joint Publication 3-13 Information Operations, February 2006. URL [http://www.dtic.mil/doctrine/new\\_pubs/jp3\\_13.pdf](http://www.dtic.mil/doctrine/new_pubs/jp3_13.pdf)
- Wall D. (2007). *Cybercrime*. Cambridge: Polity Press
- Williams L. (2008). Catch Me If You Can: A Taxonomically Structured Approach to Cybercrime. Forum on Public Policy
- Wilson C. (2008). Botnets, Cybercrime, and Cyberterrorism: Vulnerabilities and policy issues for congress.

# On Neutrality and Cyber Defence

Marcel Stolz

Centre for Technology and Global Affairs (CTGA) / Cyber Security Analytics Group, University of Oxford, UK

[marcel.stolz@oriel.ox.ac.uk](mailto:marcel.stolz@oriel.ox.ac.uk)

**Abstract:** Neutrality is a concept of international law, which the Swiss Confederation adopted in 1815. Its current idea goes beyond the legal requirements set by The Hague Convention of 1907. The convention comprises mainly of territorial definitions relevant for conventional warfare. Switzerland, on the other hand, defines the idea of neutrality more broadly. Considerations on its cyber neutrality policy could also be applied to other countries that claim to be neutral or impartial. We take Switzerland as an example for a case study, as it has the longest continuous tradition of neutrality. Due to the non-territorial character of cyber conflict, the conventional practices of Switzerland's neutrality and foreign policy are confronted with challenges: Major challenges are posed by the clash between national interest and the self-restrictions of neutrality policy. The national interest demands a strong cyber security capacity and an effective defence capability. However, this might only be achieved by means of international collaboration and knowledge exchange. This collides with the principle of impartiality and non-intervention in international conflict, a core-concept of neutrality. A new concept for neutrality in cyberspace has to be developed, which builds on the foundation of Switzerland's tradition as a neutral country in the international community. This paper outlines the inherent problem of neutrality and cyber defence. We describe Switzerland's neutral tradition, how it has developed since 1815 and its current characteristics, described as *active neutrality* policy. Furthermore, we illustrate Switzerland's involvement in cyber activities and outline where these involvements reach their limits. Finally, an outlook on future implementations of neutrality policy is made.

**Keywords:** neutrality, cyber defence, Switzerland, cyberspace dilemma in kinetic warfare

---

## 1. Introduction

Neutrality of a state or nation is a concept of *conventional* or *kinetic* warfare. Its initial implementations have developed on the basis of informal practices of non-aligned states. A legal definition of neutrality has been adopted in international law by The Hague Convention (1907). While the convention codifies the rights and duties of neutral nations in war on land and sea, it does not cover all aspects of neutrality. In particular, there is no generally accepted legal codification of the rights and duties of neutral nations in cyber conflict. With the emergence of conflicts in cyberspace, it is of vital interest for neutral nations to find a common understanding of the implementation of neutrality in cyberspace. Neutral nations need to know what is expected from them in this new domain of conflict—and what they may expect from the status of neutrality. Hence the motivation of this paper to cover neutrality with a focus on cyber defence.

When seeking solutions for a common definition of neutrality in cyberspace there is more to take into account than a legal definition of neutrality in conventional warfare. The history of implementation and practice of the foreign and security policy of neutral nations has to be considered, since neutrality existed already before its legal definition. This implementation has changed over the centuries and provides material for case studies of neutrality, which can lead to a more thorough understanding of what state neutrality means. We refer to this history of the implementation as *neutral tradition*.

### 1.1 Paper outline

In this paper we investigate which steps are required in order to raise the idea of state neutrality to cyberspace. We do this by means of a theoretical analysis of neutrality policy on the one hand and an analysis of current approaches to cyber defence theories in international relations on the other hand. Since the most prominent and longest continuous example of a nation's successful implementation of neutrality is found in Swiss history, we mostly rely on examples and case studies related to Switzerland. As a framework of thought in cyber studies we mostly rely on the work of Lucas Kello (2017), as it provides a thorough analysis of the contradictions between classical defence thinking and the nature of cyber conflict and extends international relations theory to the domain of cyberspace. It is briefly introduced in Section 2. In Section 3, we analyse work on the neutral tradition and explore the case of Switzerland's implementation of neutrality. This will include frameworks developed for considerations on neutrality as well as challenges that neutrality policy had to face and prevail in the past. Thereafter, Section 4 discusses some straightforward examples of Switzerland's reaction on cyber incidents in recent years. Building on the findings from Sections 2,3, and 4 we deduce and explain challenges that current neutrality

policy has to face with regard to cyberspace in Section 5. Finally, Section 6 concludes and outlines possible future work on the topic of neutrality policy and its implications on cyber defence.

## **2. Cyber studies in political science**

Since the field of international relations, which commonly describes interstate conflict and cooperation, does not reflect adequately the nature of cyberspace—in particular its pervasiveness—Kello expands international relations theory in order to incorporate the *virtual weapon*, i.e. the effect of cyber operations on the International System. He claims that the virtual weapon fundamentally disrupts the *Conventional Model* of states as the dominant units of international relations theory. While the Conventional Model relies on a Clausewitzian approach to international conflict, Kello declares that he expands theory to post-Clausewitzian considerations. We believe that his analysis of technological revolution and its influence on global conflict provides a thorough basis in order to explain the challenges to conventional neutrality policy. The conventional theories of international relations neglect the novel character of the virtual weapon.

Kello defines third-order, second-order and first-order technological revolutions and analyses their effect on the Conventional Model. We shortly outline the basic concepts of these revolutions:

Third-order technological revolution, also known as systemic disruption, refers to “important adjustments within the sharp limits of the states system”. While the rational order of the International System is affected, the moral order is not. An example of systemic disruption is “the replacement of one or a group of dominant states in the system by another”. The basic mechanisms of the states’ contest for power are not changed.

Second-order revolution, also known as systemic revision affects the moral order of the International System. It maintains the Conventional Model: A state or group of states repudiates the “shared basic purposes of the units and rejects the accepted methods of achieving them, in particular restraints on the objectives and means of war”. Nevertheless, “it leaves the system’s organising principle, or state supremacy, intact”. Examples are the emergence of Soviet communism and its idea of world revolution or the strive for permanent peace as a founding principle of the European Union.

First-order revolution, or systems change, finally, overthrows the mechanisms of the Conventional Model in that it challenges the supremacy of states. Examples are alien (i.e. non-state) actors who try to dominate the actions of conventional states by means of new mechanisms of influence and power. It yields an altogether new type of International System beyond the concept of the supremacy of states.

We argue that this classification of types of revolutions is essential to describe the challenges of cyberspace to neutrality policy. A consequence of this systems change is the state of *unpeace*: While nations are not formally at war with one another, their hostile activities in cyberspace do not correspond to the conventional understanding of peaceful co-existence either. A further aspect of this systems change is the *diffusion of power*. In a system where states no longer hold the supremacy of power, other actors and groups have increased means of power to influence the actions of states. As states no longer hold the monopoly of power, defence is also fragmented: private (and other) actors, which used to be subunits of states, develop their own defence mechanisms in order to secure their interests. This further erodes the supremacy of states, as large technology firms gain influence not only in the defence of their own systems but also in the balance of cyber power.

A particular shift in defence thinking with effect on neutrality policy is the irrelevance of territoriality and the pervasion of dimensions of warfare. This aspect is partially introduced by Joseph S. Nye (2011) when he defines the nature of *cyberpower*. This novel aspect of warfare has yet to be fully understood and applied in military strategy. In cyberspace, most countries have a virtual border with most other countries and are, therefore not only vulnerable by means of a violation of their territorial borders but directly through cyberspace. These factors have an impact on conventional neutrality policy.

## **3. Switzerland's tradition of neutrality policy**

In order to understand more clearly the traditional implementation of neutrality policy, we outline the theory of neutrality policy in the upcoming subsections. We do this on the example of Switzerland, as its history provides the majority of interesting cases and Swiss neutrality is the longest example of a continuous and successful implementation of neutrality policy. It has lasted for over 200 years now. We first briefly outline the history of

Switzerland's neutrality. Thereafter, we discuss thought frameworks and terms of neutrality policy. Finally, we illustrate Switzerland's current approach to neutrality policy within the Conventional Model. The content is mostly based on Suter (1998), Riklin (1991) and Holenstein (2014).

### **3.1 A very brief history of Swiss neutrality**

Some historians have claimed that Swiss neutrality policy goes back as far as 1515 (Bonjour, 1946). It is arguably the year of 1515—following the event of the *Marignano Defeat*—which marks a turning point with regard to the involvement of the Swiss states in conflicts outside their territory. The Swiss states continuously refrain from fighting in foreign conflicts during the following centuries and take measures to avoid being drawn into foreign conflicts. However, the actual starting point of Swiss neutrality is the year 1815, following the Vienna Congress and the Treaty of Paris, as is generally accepted by contemporary historians (Suter, 1998). Switzerland's neutrality was defined as *permanent*. This neutrality was challenged several times. Suter (1998) provides an essay highlighting some precedential situations. Neutrality was not only chosen by the Swiss but rather imposed by the great powers of Europe during the process of reorganisation after the Napoleonic wars. Switzerland's neutrality was based in its geostrategic location at the centre of Europe and the inability of the superpowers (Prussia, Russia, England, France and Austria) to decide who this geographically difficult terrain should be assigned to. This highlights an important precondition for neutrality: It cannot merely be chosen by a country. Neutrality has to lie in the interest of a majority of influential powers.

The underlying principle of neutrality is to *prevent bias towards any belligerent power even before a conflict is recognisable*. This goes beyond the legal definition of neutrality, which only covers the declaration of neutrality in an already ongoing conflict.

Suter explains the different nuances of neutrality over time. During the 19<sup>th</sup> century, neutrality was a means for sovereignty of a lone and yet very young liberal and direct democracy surrounded by mostly monarchic states. It was challenged several times when Switzerland had to mobilise its forces to deter a Prussian / German invasion, e.g. in the *causa Wohlgemuth*: Switzerland functioned as a safe hideaway for political dissidents, for example the German social democrats after Bismarck's anti-socialist laws were passed. From this time arises the principle that neutrality should not only be permanent but also *armed* in order to prevent or deter attacks. A neutral nation cannot rely on any foreign power to protect its interests. Its army has to be of substantial size and self-dependant. It is noteworthy that even in 1998 the only armies of larger size on the European continent were those of Russia, Ukraine, Turkey and France (Suter, 1998).

Armed neutrality became particularly important during the world wars. During the First World War, Switzerland experienced a strong ethnic divide and had to remain neutral in order to prevent internal instability. After the First World War, Switzerland chose to join the League of Nations, marking the start of what is known as *differential* neutrality: In respect of Switzerland's neutral status, Switzerland to participate in economic sanctions only, military sanctions of the League were ignored. Switzerland returned to its *integral* neutrality in 1938, leaving the League when it became obvious that it had failed its goal of collective security.

Swiss neutrality in the Second World War was far from perfect. Secret consultations between the Swiss head of armed forces, General Guisan, and the French general staff covered the case of Germany circumventing France's Maginot line through Switzerland. A German invasion would be a breach of Swiss neutrality, yet the consultations themselves were as well. Suter justifies them with the absence of alternatives and the presence of an imminent and existential threat. Minor breaches with regard to trade politics followed: Switzerland was forced to collaborate economically with the axis powers. Suter concludes that neutrality was one of several factors for relative peace, the other being: the usefulness of the functioning Swiss economy for the axis powers, the mobilised army, the ideological resistance (*geistige Landesverteidigung*), the humanitarian and diplomatic services for all belligerent powers and the luck of not being at the geostrategic centre of the war.

On this background, Switzerland pursued its strategy of integral neutrality throughout the Cold War. It would not join the UN, it would not participate in any economic or military sanctions and it would remain highly armed and provide diplomatic and humanitarian services for the international community. The Cold War shaped the still applied principle of *courant normal* in the case of economic sanctions: In order to not upset the sanctioned nation, Switzerland would not participate in sanctions. Yet, in order to avoid upsetting the sanctioning nations, it would also not allow any circumvention of sanctions by freezing its economic interactions to the *regular scope*

of economic trade with the sanctioned country. Only after 1990 Switzerland started to integrate more into supranational institutions, joining the UN in 2002 and participating in UN economic sanctions. Early in the 21<sup>st</sup> century, Swiss Foreign Minister Micheline Calmy-Rey coined the term of *active neutrality*. It emphasises the responsibility of a trusted third party in peacekeeping processes. Micheline Calmy-Rey decided that Switzerland should become more active in foreign politics by means of expressing discontent in cases of clear breaches of international law.

### **3.2 Neutrality concepts**

The two main works on the concept of neutrality policy emerge from Suter (1998) and Riklin (1991). From the previous subsection, we recall the different definitions of neutrality, based on Suter (1998) and Widmer and Kreis (2007):

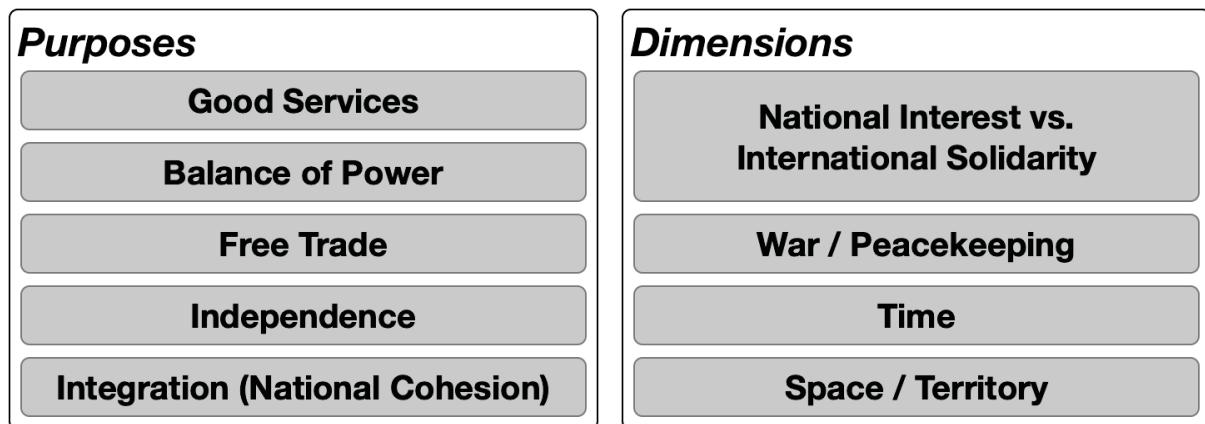
- Military non-alliance: no obligation or contract exists to support any belligerent power, yet the country may choose to support any side;
- Conflict-specific neutrality: a state declares its neutrality for a specific conflict. The nation should not be biased towards any of the belligerent powers and provide equivalent trade options for both sides. International treaties on the rights and duties of neutral nations are in place;
- Permanent neutrality: the declaration that a nation will not be biased towards any belligerent power, even before any conflict is foreseeable;
- Armed neutrality: the declaration of a nation that it is willing to defend its neutrality by its own means;
- Integral neutrality: a very strict and absolute interpretation of neutrality with isolationist aspects. No supranational organisations (e.g. EU, UN) should be joined, no economic sanctions should be followed;
- Differential neutrality: international organisations may be joined and economic sanctions may be followed if they are widely accepted by the international community. No military sanctions are followed. The aim to maintain unbiased towards any power is upheld;
- Active neutrality: a neutral nation's recognition of its responsibility for international peace and collaboration as a trustworthy third party. The neutral nation expresses its discontent in cases of clear breaches of international law.

Suter also defines three preconditions for the formation of Swiss neutrality: first, the absence of a consensus among the major powers due to their diverging geopolitical interests. Second, the absence of consensus within Switzerland on its foreign ties due the diverging ethnical, cultural and cantonal interests. Third, Switzerland's status as a small state with little military threat potential. These preconditions can be broken down to the *balance of power*, *internal peacekeeping* and an *absence of threat* to the major powers.

We also note Suter's emphasis of the *requirement of self-defence*: Switzerland's borders were threatened only decades after 1815 by powers that had promised to guarantee them.

Finally, we note the precedence of *national security over neutrality policy*: neutrality policy is an important principle supporting Switzerland's security strategy: it aims at not being drawn into any conflict of foreign powers. When the country's national security is existentially threatened, a historic precedent for prioritising national security over neutrality has been set. Suter further underlines this claim by stating that the authors of the first constitution of the modern federal state did explicitly abstain from adding neutrality as one of the founding principles of the Swiss Confederation in 1848.

These principles exfiltrated from Suter's essay are consistent with Riklin's theory of *purposes* of neutrality. Riklin describes five purposes of Swiss neutrality, as shown in Figure 1. The purposes explain what major national or international interests require neutrality. The dimensions are the domains in which the relevance of these purposes are measured. The importance of a purpose might change over time, or is more or less relevant from a territorial (geopolitical) perspective, etc.



**Figure 1:** The functions and dimensions of state neutrality, according to Riklin (1991)

### 3.3 Current Implementation of neutrality policy

The current implementation of neutrality policy is based on a new implementation of differential neutrality with the addition of active neutrality. While it emphasises the aspect of international collaboration and the importance of providing good services to the international community, the other purposes remain relevant for Switzerland. Former diplomats underline the importance of a credible and trustworthy implementation of neutrality policy (Widmer et al, 2007) in order to maintain the option of active neutrality. Furthermore, they emphasise that, even though the geopolitical environment for Switzerland might seem peaceful in post-Cold War Europe, neutrality should still be upheld as basic principle of foreign policy in order to maintain independence and avoid foreign conflicts. Finally, neutrality enjoys very strong support in the Swiss population. Its acceptance rate is over 80%, sometimes even over 90% (Widmer et al, 2007). In the Swiss democratic system this means that an official breach of neutrality policy is politically infeasible.

## 4. Switzerland's activities in cyberspace

In order to understand the potential and limits of neutrality policy in cyberspace, we shortly outline the current state of governmental cyber activities in Switzerland. This section also covers examples of cyber incidents in government institutions.

### 4.1 Switzerland's cyber institutions and its cyber strategy

The main governmental organisation that analyses cyber incidents on a daily basis is *MELANI*. It monitors current cyber risks and major bugs and functions as an information hub between the private sector, critical national infrastructure providers, the intelligence service and also runs the government's CERT (*Computer Emergency Response Team*). Its competencies are limited to analyses, monitoring and distribution of information. A small unit connected with MELANI sits in the Swiss intelligence service in the Department of Defence. It is assumed that this unit might be capable of carrying out offensive actions.

Apart from the cybercrime prosecution unit of the federal police, the other main governmental cyber institution are the *Swiss Armed Forces*. Even though there has been a high amount of publicity on the introduction of a cyber unit, it is questionable what capability this unit has so far. The main recruiting scheme has only started in 2018, with approximately 18 recruits per semester (Swiss Armed Forces, 2018).

Switzerland has also released a National Cyber Strategy, meanwhile in its second iteration (Informatiksteuerorgan des Bundes ISB, 2018). The document outlines the responsibilities of MELANI and further government institutions with regard to incident management and protection of critical infrastructures. It also has a specific section on Cyber Defence. It does not go into depth and only covers main aspects. Nevertheless, active protection measures are a focus area for development of capabilities. Furthermore, a particular focus lies in the provisioning of intelligence information and the attribution of cyberattacks.

Noteworthy is also the section on international collaboration in cyber matters, where Switzerland aims at bringing together nations to generate a basic understanding of problems in cyberspace and potential resolution mechanisms.

## **4.2 Cyber defence case examples**

In recent years, several attacks on Swiss government institutions have been registered and publicly reported. Notable attacks are the series of data breaches at the Department of Foreign Affairs (Schweizerische Depeschenagentur, 2012), the RUAG case in 2016 (GovCERT.ch, 2016) and further attacks on the Department of Defence in 2017 (Der Bundesrat, 2017). The RUAG case is particularly interesting, as a detailed technical report is provided. It was not a direct attack on the Swiss Department of Defence. Rather, it attacked a governmentally owned defence contractor, RUAG, in order to gain access to confidential data shared with the contractor. The attack in 2017 is a follow-up on the 2016 attack, indicating that the initial attacker did presumably not get all the data intended from the attack in 2016.

The release of the technical report is particularly noteworthy, as it provides insights into how Switzerland responded to a significant foreign attack on one of its government institutions. A response is challenging. Conventionally, a military response to a military attack would be permitted for a neutral country. Self-defence, in particular a direct response to an attack, is always permitted. However, in cyberspace the authorship of an attack is easily blurred. States can plausibly deny that they authored an attack and so the attribution to an attack remains a matter of more or less substantial speculation. Hence, a direct response to a cyberattack by means of a counterattack is not a recommendable solution. The technical report released on the RUAG case provides an interesting solution to this problem. The report was released publicly and provides a detailed technical analysis of the attack. Furthermore, it attributes the attack to a hacker group commonly associated with government activities of a specific country. Hence, the report publicly but indirectly blames a country for the attack. Furthermore, it provides a platform for Switzerland to present its technical capabilities with regards to attack analysis. The very detailed report bears an indirect message to any potential attackers: "beware, we know who is attacking us and we have the means to discover your attack!" Furthermore, it leaves room for speculation on the potential capabilities of Switzerland in the offensive domain, which might follow a cyberattack on Switzerland.

## **5. Future of Switzerland's neutrality policy**

We have discussed the current state of cyber studies, the history of neutrality policy and the current state of Switzerland's cyber activities. The aim of the current section is to merge the findings of each previous section and to define the challenges they pose when brought together. We then outline potential solutions to the challenges.

### **5.1 Challenges**

In Section 2, we discussed the problem of the scholarship gap regarding cyber studies in political science and the theory of revolutions disrupting the International System. Cyberspace has the potential of launching a first-order revolution, or *systems change*. Kello (2017) argues that cyberspace has already caused the International System to change, by significantly empowering non-state actors. This notion of a systems change is presumably the main challenge for the conventional implementation of neutrality policy. As we introduced in Section 3, the history of neutrality policy has evolved in an international system of sovereign states, i.e. the Conventional Model. While third- and second-order revolutions maintain this system—and examples of such revolutions have occurred over the last 200 years of history—, first-order revolutions impacts the way the International System works. Hence, some of the considerations on neutrality policy from Section 3 might not correspond to the current state of international relations anymore. Independence might have become a naïve wish in an interconnected world. Similar considerations are true for free trade in times of publicly fought trade wars. On the other hand, neutrality purposes, such as good services or the balance of power receive a new meaning: the emergence of non-state actors requires to also balance their power and to provide diplomatic or humanitarian services for them. It is questionable to which extent this might be accepted by states in the future.

Moreover, we have seen that the justification for neutrality has been made on the basis of geostrategic, i.e. territorial, considerations. In cyberspace, However, Switzerland has an indirect virtual border with any country, even if only indirectly. Hence, the basic environment for neutrality has become more complex.

A further aspect is the precedence of national security over neutrality. In the cyber age, where every component of a company or government can potentially be attacked, it becomes questionable whether neutrality is still maintainable. Is the cyber threat enough to cause an existential threat to national security? Or does neutrality,

extended to cyberspace, offer additional safety and security from attacks? This matter requires careful evaluation and balancing of the potential of neutrality in cyberspace, its advantages, and the disadvantage in terms of cyber capability, cyber capacity and knowledge exchange, as this might have an impact on bias and trustworthiness.

Finally, we have outlined in Section 3 that an important precondition for neutrality is not only the choice of a country to be neutral but also a common interest of the international community—or at least a significant part of it—in favour of the neutrality of this country. It is unclear whether there is a general international interest for the case of a neutral country in cyberspace. An interest would require a benefit for the international community, which has yet to be elaborated.

We conclude that the main challenges are:

- the effect of the unprecedented systems change on neutrality policy;
- the impact of the previous point on the purposes of neutrality;
- the rupture of the geostrategic / territorial nature of a country's security;
- the precedence of national security over neutrality policy in the cyber era;
- the benefit of neutrality in cyberspace for the global community.

## **5.2 Provisional outline of possibilities**

Following the definition of challenges for neutrality in cyberspace, we would like to outline some potential solutions to these challenges relying on current developments in cyberspace.

First of all, the effects of the systems change have yet to manifest themselves in the International System. While there is an increased number of actors beyond state control in cyberspace, there is also an increased initiative of states to maintain the supremacy of states (i.e. the Conventional System). A prominent example is the UN Group of Governmental Experts (UN GGE), where Switzerland pushes strongly for a consensus-finding approach to answer challenges of cybersecurity (BAKOM, 2018). This approach is also described in the National Cyber Strategy of Switzerland (Informatiksteuerorgan des Bundes ISB, 2018). In case this approach fails, other actors apart from states will have true relevance, a process which has already started in the recent years. A typical example are the increased influence of large tech companies. Further examples are the emergence of hacker groups and their influence on global affairs. This increase of actors might increase the importance of neutrality policy: in terms of national security, neutrality might help avoid being targeted by the new global actors. Furthermore, the presence of a generally trustworthy mediator is beneficial for all parties.

The purposes of neutrality should remain unchanged by the systems change. However, their significance will be impacted; some purposes might increase in importance while others decline. As outlined previously, the significance of good services and balance of power could increase.

The rupture of the geostrategic and territorial nature of a country's security definitely has an impact on Switzerland's security. There is an increased interest in remaining neutral, as more virtual borders equal more potential attack origins and neutrality helps avoid potential conflict. Also, more virtual borders provide more potential partners for direct free trade. Hence, this challenge will presumably function as an accelerator for neutrality in the future.

The precedence of national security over neutrality policy is the most problematic challenge. The significance of the virtual weapon for national security is currently being discussed. An important aspect neutrality is the defence capability. Would a neutral country be able to accumulate sufficient knowledge and capabilities? Or is increased collaboration with foreign powers an indispensable asset for the defence strategy of small countries? We cannot answer these questions sufficiently for the time being.

When it comes to the benefit of neutrality policy for the international community, there are straightforward possibilities: A neutral country can ensure communication and diplomatic relations between parties who do not

have any direct and official means of communication. Furthermore, a neutral country can provide analysis services of cyberattacks, as has been suggested by Mäder (2019). However, this can result in the country being targeted. On the other hand, the neutral country can gain important insights from the analysis work.

We conclude that a future of neutrality in cyberspace is possible. However, its detailed shape yet has to be defined. We can imagine a neutrality policy which takes into account the different nature of cyberspace by means of defining a *dual* neutrality policy: the conventional, stricter implementation for conventional conflicts, while applying a more flexible approach to neutrality in cyberspace, in order to allow more collaboration with other states for cyber capacity building. Furthermore, with the emergence of new actors in cyberspace, a very strict interpretation of neutrality policy in cyberspace is possible, in order to avoid being drawn into any conflicts. A particular interest also lies on the topic of the role of neutral countries for the global community.

## **6. Concluding remarks and future work**

We outlined challenges and potential solutions to neutrality policy in cyberspace. Further considerations are required in order to outline the borders of neutrality policy in the cyber era. In particular, the balance between national security and neutrality policy requires a more thorough analysis. As preliminary work, the neutrality purposes have to be freshly evaluated. The potential of a dual neutrality approach needs yet to be analysed. Finally, a study on the benefit of neutrality in cyberspace for the global community is necessary and a definition of the role of a cyber neutral country in an era of *global* conflict is required.

We would like to conclude by stating that the potential of neutral nations in cyberspace is high with regard to the system changing character of the virtual weapon, if a common ground for the benefit and requirement of cyber neutral nations can be found.

## **References**

- BAKOM (2018) „Engagement des Eidgenössischen Departements für auswärtige Angelegenheiten EDA im Bereich Cyber“, [online], BAKOM,  
<https://www.bakom.admin.ch/dam/bakom/de/dokumente/informationsgesellschaft/strategie2018/faktenblaetter/FB%20EDA%20Cybersecurity.pdf.download.pdf>  
Bonjour, E. (1946) *Geschichte der schweizerischen Neutralität. Drei Jahrhunderte eidgenössischer Aussenpolitik*, Helbing & Lichtenhahn, Basel.
- Der Bundesrat (2017) „Cyberangriff auf die Bundesverwaltung entdeckt und Massnahmen ergriffen“, [online], Swiss Confederation, <https://www.admin.ch/de/start/dokumentation/medienmitteilungen.msg-id-68135.html>
- GovCERT.ch (2016) *APT Case RUAG Technical Report*, Swiss Confederation, Bern.
- Holenstein, A (2014) *Mitten in Europa: Verflechtung und Abgrenzung in der Schweizer Geschichte*, Hier und Jetzt, Baden.
- Informatiksteuerorgan des Bundes ISB (2018) *Nationale Strategie zum Schutz der Schweiz vor Cyber-Risiken (NCS) 2018-2022*, Swiss Confederation, Bern.
- Kello, L. (2017) *The Virtual Weapon and International Order*, Yale University Press, New Haven.
- Mäder, L. (2019) „Wenn der feindliche Zugang zum Computer gleich mitgeliefert wird“. [online], NZZ,  
<https://www.nzz.ch/schweiz/wenn-der-feindliche-zugang-zum-computer-gleich-mitgeliefert-wird-ld.1467220?mktcid=nled&mktcval=107&kid= 2019-3-18>
- Nye, Jr., Joseph S. (2011) *The Future of Power*, PublicAffairs, New York.
- Riklin, A. (1991) „Funktionen der schweizerischen Neutralität“, *Passé pluriel. En hommage au professeur Roland Ruffieux*, Editions universitaires, Freiburg, pp 361–394.
- Schweizerische Depeschenagentur (2012) „Bundesanwaltschaft untersucht erneuten Hacker-Angriff auf EDA“, [online],  
Tageswoche, <https://tageswoche.ch/allgemein/bundesanwaltschaft-untersucht-erneuten-hacker-angriff-auf-eda/>
- Suter, A. (1998) *Neutralität. Praxis, Prinzip und Geschichtsbewusstsein*, Eine kleine Geschichte der Schweiz, Suhrkamp, Berlin.
- Swiss Armed Forces (2018) „Erste Erfahrungen mit dem Cyber-Lehrgang der Armee“, [online], Swiss Confederation,  
<https://www.vtg.admin.ch/de/armee/detail.news.html/vtg-internet/verwaltung/2018/18-09/erste-erfahrungen-mit-dem-cyber-lehrgang-der-armee.html>
- The Hague Convention (1907) *Rights and Duties of Neutral Powers and Persons in War on Land*, The Hague,  
<https://www.loc.gov/law/help/us-treaties/bevans/m-ust00001-0654.pdf>
- The Hague Convention (1907) Rights and Duties of Neutral Powers in Naval War, The Hague,  
<https://www.loc.gov/law/help/us-treaties/bevans/m-ust00001-0723.pdf>
- Widmer, P. and Kreis, G. (2007) *Die Schweizer Neutralität: beibehalten, umgestalten oder doch abschaffen?*, Werd-Verlag, Zürich.

# **Integrative Approach to Understand Vulnerabilities and Enhance the Security of Cyber-Bio-Cognitive-Physical Systems**

**Todor Tagarev<sup>1</sup>, Nikolai Stoianov<sup>2</sup> and George Sharkov<sup>3</sup>**

**<sup>1</sup>Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Sofia, Bulgaria**

**<sup>2</sup>Bulgarian Defence Institute, Sofia, Bulgaria**

**<sup>3</sup>ESI Centre Eastern Europe, Sofia, Bulgaria**

[tagarev@bas.bg](mailto:tagarev@bas.bg)

[n.stoianov@di.mod.bg](mailto:n.stoianov@di.mod.bg)

[gesha@esicenter.bg](mailto:gesha@esicenter.bg)

**Abstract:** Rapid technological advances provide numerous benefits to our ways of work and leisure, banking and transportation, delivery of products and health assistance. The increased interconnectedness among devices, people, networks, and systems, however, introduces a level of complexity surpassing the experience accumulated so far. While the security of communications, network and information systems can be considered a well-established discipline, the study of security of cyber-physical systems is fairly recent. Furthermore, the dependencies of live organisms, including humans, with integrated sensors and electronics, of perceptions and cognition, and variety of drones on influences from cyberspace have been subject of only few, mostly incidental studies. The interdependencies among cyber, physical, biological systems, and humans in situation assessment and decision-making roles create new potential vectors of attack by malicious actors. If exploited, they will lead to cross impact among domains that are usually studied separately. Authors from three Bulgarian institutions, combining research and policy-making experience, embarked on the task to elaborate a comprehensive cybersecurity research agenda. This paper presents their concept for an integrative approach to the exploration of ‘systems of systems.’ The study is structured along five domains: communications and information systems and networks; cyber-physical system; bio-integrated systems; cognitive processes, i.e. the processes of shaping perceptions, assessing a certain situation and options and making decisions; and drones, remotely controlled or autonomous, the latter case being particularly reliant on advances in artificial intelligence. This paper outlines the problem of vulnerability of each of the five domains to influences from cyber space. Then it presents some advances in cross-domain understanding of vulnerabilities, supported by examples of cybersecurity studies, and provides the outlines of a corresponding, interdisciplinary research agenda, built around the concept of systems of systems. The authors conclude by predicting that the field of cybersecurity will be subject to considerable growth in coming years, requiring multi- and inter-disciplinary competencies and scientific support.

**Keywords:** cyber security, cyber-physical systems, bio-integrated systems, hybrid threats, system of systems, decision-making

---

## **1. Introduction**

Advances in sensors, communications and information technologies, biotechnologies (Lentzos, 2016), nanotechnologies (Ionescu, 2016), microelectromechanical systems (MEMS), storage and generation of energy (Jha, 2010) allow for rapid development and implementation of new products and services. In parallel, interconnectedness increases exponentially. In this development, sometimes designated as the “Internet of Things” (IoT), researchers anticipate a merger of the physical and virtual worlds, ripe of benefits, but also of vulnerabilities (Costigan and Lindström, 2016).

Sensors are increasingly used to collect data and information on the functioning of industrial systems, pipelines, vehicles and their environment, physiological processes and psychological status (Paradiso, Faetti and Werner, 2011). Sensor data is transferred via networks and integrated with information from diverse databases. The definition of control inputs, or decision making more generally, is automated to increase speed and efficiency, while in a growing number of applications the human in the loop is replaced by artificial intelligence. Then various types of actuators implement such decisions, with the consequent impact on transport, industrial or physiological processes.

Thus, technological advances provide benefits to our ways of work and leisure, banking and transportation, delivery of products and health assistance. However, in the race to gain a competitive advantage and exploit commercially the opportunities provided by technology, developers do not always pay sufficient attention to safety and security. Products often have intrinsic vulnerabilities that may lead to malfunctioning of equipment

or interruption of services in ‘normal’ conditions, under the influence of natural hazards, or unintended human activity. More importantly, if behaviour in the past can be used to predict the future, individuals or group actors with malicious intent will continue to search for ways to exploit vulnerabilities.

The increased interconnectedness among devices, people, networks, and systems introduces a level of complexity surpassing the experience accumulated so far. While the security of communications, network and information systems can be considered a well-established discipline, the study of security of cyber-physical systems is fairly recent (Rehman et al, 2018). Furthermore, the dependencies of live organisms, including humans, with integrated sensors and electronics, of perceptions and cognition, and the increasing variety of drones on influences from cyberspace have been subject of only few, mostly incidental studies.

The interdependencies among cyber, physical, biological systems, and humans in situation assessment and decision-making roles create new potential vectors of attack by malicious actors. If—or, rather, when—exploited, they will lead to cross impact among domains that are usually studied separately.

This paper calls for an integrative approach to the study of such ‘systems of systems.’ The examination is structured along five domains:

- communications and information systems (CIS) and networks;
- cyber-physical system;
- bio-integrated systems;
- cognitive processes, i.e. the processes of shaping perceptions, assessing a certain situation and options and making decisions; and
- drones, remotely controlled or autonomous, the latter case being particularly reliant on advances in artificial intelligence (AI).

This remaining text is structured as follows. Section 2 outlines the problem of vulnerability of each of these five domains to influences from cyber space. Section 3 presents some advances in cross-domain understanding of vulnerabilities. The fourth section provides the outlines of a corresponding interdisciplinary research agenda, built around the concept of systems of systems. We conclude with the prediction that the field of cybersecurity will be subject to considerable growth in coming years, requiring the reflection of multiple perspectives and variety of competencies.

## **2. Increasing domain vulnerability**

Equipment, services and systems in each of the five domains have been subject of cyberattacks. Without going into detail, this section provides examples of vulnerabilities that have been exploited with malicious intent or the possibility for such exploitation has been demonstrated in trials.

### **2.1 Communications and information systems and networks**

Computers, even not connected to other devices, were found vulnerable in the early 1970s and attacked by viruses, spread via floppy disks, already in the early 1980s. Connections between computers made the spread of viruses easier, and the ways and means of attack rapidly proliferated with the emergence of Internet and wireless communications. By now, with the understanding of the importance of the communications and information infrastructure for almost any facet of the functioning of modern societies, network and information systems security is an established discipline with an evolving body of strategies, policies and legislation (e.g. the EU Directive 2016/1148), standards (e.g. ISO/IEC 27000 series), organizations such as (national/NATO) computer incident response centres/capabilities (N/CIRC) and computer emergency response teams (CERTs), etc.

Nevertheless, cybersecurity remains a challenge, as witnessed for example by the call of the U.S. Defense Advanced Research Projects Agency for developing hardware design tools to provide built-in cyber security against computer hardware vulnerabilities in military and commercial electronic systems (Keller, 2017).

## **2.2 Cyber-Physical Systems**

Broadly examined, Cyber-Physical Systems (CPS) are “systems that integrate computing systems with the physical components” and include, *inter alia*, aerospace, manufacturing, process control, robotics, and power grid systems (Suh et al, 2014; Lezzi, Lazoi and Corallo, 2018) and other critical infrastructures.

There are numerous examples of successful cyberattacks on physical systems, e.g. the attacks on regional power distribution networks in Ukraine at the end of 2015 (Lee, Assante and Conway, 2016). Diverse efforts are already underway to examine systematically the problem, find and implement relevant solutions. The U.S. National Institute of Standards and Technology (2017) has published a cyber framework, enabling “organizations – regardless of size, degree of cybersecurity risk, or cybersecurity sophistication – to apply the principles and best practices of risk management to improving security and resilience.”

Likewise, the IEEE Systems Council (2017) established a Technical Committee on Cyber-Physical Systems to promote “interdisciplinary research and education in the field of cyber-physical systems” and address “the close interactions and feedback loop between the cyber components such as sensing systems and the physical components such as varying environment and energy systems”.

## **2.3 Cybersecurity of bio-integrated systems**

There is a clear trend towards increasing use of wearable electronic sensor systems for long-term monitoring of physiological signals (Yeo et al, 2013). The use of electronics, sensors and actuators—or ‘Implantable Medical Devices’ (IMDs)—for both diagnostic and therapeutic purposes is also proliferating (Rogers, 2012). However, legacy IMDs in particular suffer from cybersecurity vulnerabilities that, if exploited, may have dramatic consequences for the patients (Tabasum et al, 2018).

Particularly widespread is the use of actuators—insulin pumps in this case—for treating patients with diabetes. While this increases benefits to patients, the complexity of the resulting system is growing, making breaches of security possible (Paul, Kohno and Klonoff, 2011; Klonoff, 2015). Years ago, it has been demonstrated that by a cyber hack an attacker is able to deliver deadly dosage of insulin (Goodin, 2011). Other programmable, implantable and external biomedical devices, such as pacemakers, defibrillators, pain management pumps, vagus nerve stimulators, etc., may also be vulnerable to unauthorized access, commonly referred to as ‘hacking’. This intrusion may lead to compromise of confidential patient data or loss of control of the device itself, which may be deadly (Frenger, 2013; Storm, 2015).

The dependence of human and other living organisms on advanced information technologies will only increase with the implementation of movement enhancing mechanisms, e.g. exoskeletons (Veneva, 2018), sensors and actuators for farm animals (Andersson, 20016), and for the purposes of farming (Penido, 2019).

These and other similar cases raised the concerns and growing understanding that it is “imperative that time and funding is invested in maintaining and ensuring the protection of healthcare technology and the confidentiality of patient information from unauthorized access” (Kruse et al, 2017). Yet, compared to cyber-physical systems, research frameworks are less developed for this domain.

## **2.4 Cyber aspects of cognitive processes**

Two trends shape the future of cognition in networked environments saturated by sensors, databases, advanced visualization, information processing and decision support technologies. First, electronic media and social media are increasingly used to influence decision making by providing manipulated contextual information, spreading fake news, propaganda, using bots and cyborgs, changing search engine results, etc., and thus influencing perceptions, the assessment of the situation, generation of options, and the selection of one or another course of action. These developments are often analysed as a component of *hybrid warfare* (Pocheptsov, 2018), but more specific analysis frameworks are also emerging (Wardle and Derakhshan, 2017).

The second trend relates to the power of augmented and virtual realities, and will be examined in the next section.

## **2.5 Proliferation of drones and advances in artificial intelligence**

The fifth and the final domain in our examination covers the ever wider use of drones and advances in artificial intelligence. Unmanned aerial, ground/surface or underwater vehicles, collectively known as UxVs, attract significant investments and already find civilian (Al Amir and Al Marar, 2019), military (Scharre, 2018), and wider security (Koslowski and Schulzke, 2018) applications. Most of the drones currently in use are remotely controlled, while the new generations rely increasingly on the application of artificial intelligence performing their tasks either individually or in swarms (Pastore, Galdorisi and Jones, 2017; Scharre, 2018).

Cybersecurity was not a key concern in the design of the early generations of drones, and they are often vulnerable to cyberattacks (O’Malley, 2017). This statement applies not only to drones designed for civilian applications, but also for military drones, as exemplified by the case with the Iranian hijacking of the U.S. stealth reconnaissance RQ-170 Sentinel drone in 2011.

The enhanced understanding of vulnerabilities has already triggered rigorous studies of the problem (Costello, 2017; Bunse and Plotz, 2018), and development of methods and procedures aiming to increase the security and resilience of drones and drone fleets (Loukas, 2019; Lukina, 2018).

## **3. Advances in cross-domain understanding of vulnerabilities**

Efforts to understand and create strategies to countering threats from cyberspace to communications and information systems and, to a lesser extent, to cyber-physical systems, are well established. This section presents selected examples to demonstrate cross impact between two or more domains.

### **3.1 Cyber-bio-physical systems**

Although Suh et al (2014) include biomedical systems in the examination of CPS, in our view they form a distinct domain, the analysis in which requires advanced knowledge of biology and physiological processes. A number of examples—real and simulated—have proven that scientific rigor is needed as the best defence to promote cybersecurity in the best interest of patients. All medical devices therefore need better cybersecurity (Ransford et al, 2017).

Recently, the Berkeley Engineering School announced that is entering a new domain of biological systems and the human body. The stated aim is to “develop technology that restores sensory, motor and cognitive functions in people disabled by injury or neurological conditions”, thus adding to ongoing efforts to build intelligent exoskeletons that enhance strength, reduce fatigue or restore mobility. In the announcement it is underlined that work at the intersection of the cyber, the physical and the biological builds on expertise in systems design, achieving security, privacy, usability and resilience in complex networks (Sastry, 2015).

Implantable devices are used not only for hospital patients and post-hospital monitoring and treatment, but also by people who continue their common way of life, including work in sensitive positions, e.g. operators of critical infrastructure. A plausible scenario has already been suggested, in which hacking of an implantable defibrillator by terrorists results in a severe national security threat to the United States (Frenger, 2013). Hence, the wider use of implantable devices will further increase cyber risks to safety and security.

### **3.2 Cyber-Cognitive-Physical systems**

One of the key challenges of the biomedical cyber-physical system is to combine cognitive neuroscience with the integration of physical systems. An example of such integration is the application of novel tools and the assistance to people with disabilities (Chai et al, 2017).

There is a growing recognition that the cognitive aspects “need to be addressed in a holistic and multidisciplinary way in order to reach the full potential and obtain the benefits of cyber-physical systems” (Delicato, 2019). That applies to the development of disaster management architectures and networks (Jahir, 2019), the use of IoT-based infrastructure for smart cities (Marques, 2019), etc.

Studies on the security of cyber-cognitive-physical systems are still scarce (Gonzalez, 2017) and do not capture holistically vulnerabilities due to the interlinkage between cognitive and cyber-physical systems, potential cascading effects, respective methods of risk analysis and protective measures.

### 3.3 Artificial intelligence and cognitive processes

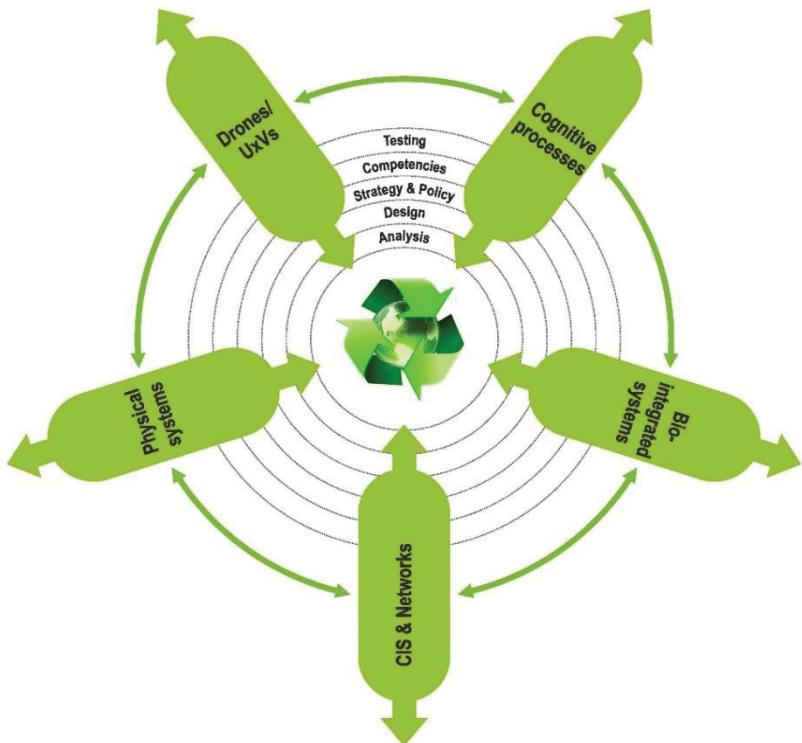
The seemingly most challenging cross domain is formed by the interplay of cognitive processes and the application of artificial intelligence to speed up decision making, enhance individual perceptions and influence group behaviour, economy, and conflict (Payne, 2018). Artificial intelligence already has the capacity to support propaganda by producing high-quality, and yet fake, stories, audio and video which, with properly selected timing, may influence the decisions of state leaders and introduce or increase existing tensions even among allies (Valášek, 2017).

Scientists and policy-makers are already addressing the social and ethical dilemmas of artificial intelligence (e.g. Science and Technology Committee, 2016), while the security aspects are still to be confronted comprehensively. At the same time, artificial intelligence methods and tools, like Machine learning and Deep learning, have been used heavily over the past two years in advanced ‘automatic’ models of classical cyberattacks. For example, according to the Internet Security Threat Report, spear-phishing is still the most used and successful cyberattack vector, and its increased efficiency is mainly due to automating this process (Symantec, 2018). By using a combination of natural language processing, ontologies and machine learning techniques for processing publicly available data, adversaries are able to generate more legitimate-looking malicious messages with efficient reduction in effort, time, and resources. Likewise, AI-based methods are needed to detect and signal such sophisticated massive attacks by profiling the behaviour of users or applications (Wilkins, 2018).

The demonstrated by the “Cambridge Analytica” case, the power of big data analysis and machine learning techniques, as well as developments in cognitive research, allowed the development of advanced defence methods and tools against adversarial attacks by implementing “next generation cognitive security operations centers” (Demertzis et al, 2019).

## 4. Towards a comprehensive research architecture

Understanding all possible interactions in cyber-bio-cognitive-physical systems, consequent effects, and the benefits of potential measures for protection poses considerable challenges. It requires developing research and expert capacity in the following fields (see Figure 1):



**Figure 1:** Integrative approach to the study of cyber-bio-cognitive-physical systems' security

- *Analysis*, i.e.
- *capacity to analyse the impact of vulnerabilities and consequences of unintentional actions or malicious attacks within and across domains, including understanding of possible cascading effects and overall impact on the security of modern society and its wellbeing;*
- *identification of possible vectors, means and tactics of cyberattacks;*
- *benchmarking and identification of best practices of cybersecurity within each domain and of systems of systems;*
- *identification of requirements for standardization and elaboration of standards;*
- *Design of products, services and systems with full integration of cybersecurity concepts and standards;*
- *Testing and certification of components, products, software components and applications, systems and system interactions;*
- *Competencies*, i.e. tracking, determining and foreseeing competence requirements, elaborating competence frameworks, training and certification requirements, education curricula and training courses; certification of personnel;
- *Strategy & Policy*, i.e. defining missions, roles, and responsibilities; organizational arrangements; procedures; capabilities; cooperation, coordination, and collaboration; risk-informed decision making on financing, budgeting and management of investments, etc.

Examples of analyses within individual domains were provided in section 2 of this paper. Among possible cross-domain interactions, research is most advanced across the cyber and physical systems. It has covered Internet of Things, Internet of Services and cyber-physical systems. In a recent report, from a multi-disciplinary perspective Dumitrache and co-authors reviewed principles and paradigms, models and architectures of CPS and suggested a conceptual framework and generic system architecture for the study of cyber-physical systems (Dumitrache, 2017).

The Technical Committee on CPS of IEEE Systems Council (2017) has defined research areas of primary interest “to develop innovative interdisciplinary techniques that can address unique CPS challenges such as the fast increase of system scale and complexity, the close interactions with dynamic physical environment, and the significant uncertainties in sensor readings” and “to bridge CPS research to a variety of emerging domains including big data analytics, smart energy systems, smart health, smart cities, etc.” Uncertainty has also been addressed with a view of testing such systems (Zhang et al, 2017).

It has been recognized that the design of cyber-physical systems requires multi-disciplinary competences. Even when CPS are designed to be autonomous, associated uncertainty requires interaction with humans for engineering, monitoring, controlling, performing operational maintenance, etc. And while the human-in-the-loop (HITL) concept is applied widely, it evolves from a reductionist point of view and exhibits certain limitations. Therefore, a model of Bio-CPS, grounded on theoretical biology, physics and computer sciences and based on the key concept of human systems integration, has been proposed to analyse such systems (Fass and Gechter, 2015).

Apparently, research frameworks in cross-domains are less developed. The authors have developed initial research architecture with a more narrow purpose in mind – to support the planning and the organization of a national research programme on “Cybersecurity and Resilience of Systems of Systems”.

It is safe to predict, that in a multi-disciplinary case like this one, relevant research paradigms, frameworks, and architectures will be developed in an evolutionary manner, combining bottom up exploration of interfaces, processes and phenomena of particular interest with integrative, comprehensive top-down studies to identify and fill in gaps. In the process, multidisciplinary fora will allow to transfer ideas and exchange of experience and knowledge from one domain to another, thus enriching the understanding of cyber-bio-cognitive-physical systems.

## **5. Conclusion**

The field of cybersecurity will be subject to considerable growth in coming years, requiring the involvement of multiple perspectives, variety of research disciplines and organization of respective competencies. This paper

outlined five domains of key interest—CIS and networks, cyber-physical systems, bio-integrated systems, cognitive processes, drones and artificial intelligence—and reviewed recent examples of vulnerabilities, approaches and solutions within each domain, as well as emerging frameworks for the study of cross-domain impact of malicious cyber activities. This understanding is currently used by the authors in the ECHO project (2019) to design a multi-sector assessment framework and, with the comprehensive picture in view, to design and select scenarios and use cases for in-depth study of cybersecurity vulnerabilities.

Approaches from natural sciences and engineering have focused mostly on the exploitation of technical and software vulnerabilities within CIS and CPS and ways to protect against cyberattacks. Studies of systems of systems have already suggested more general frameworks. The future, nevertheless, will require expansion of the study frameworks to the fields of biology and physiology, cognitive sciences (e.g. psychology, neuroscience, linguistics, philosophy of mind, anthropology, and sociology), artificial intelligence, studies of complexity and self-organization, etc.

Of particular interest will be interaction with the fields of security and defence studies which, in their foundation, go beyond protection and look into ways to prevent and even preempt the actions of an intelligent, purposeful opponent. Hence, researchers and practitioners go beyond ways of protecting and defending key assets, and consider attack as potentially more effective approach (Emmott, 2017), which may turn into the biggest overall defence policy shift in decades (Ali, 2017).

Possibly, the studies and the practice of cybersecurity will become as broad, maybe even broader, than defence with its knowledge base and practice in law, international relations, history, political science, public administration, ethics, psychology, operations, doctrine, management, technologies, industrial base, etc. – a truly multidisciplinary field of study.

## **Acknowledgements**

This paper reflects results in the project ECHO that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 830943.

## **References**

- Al Amir, N. and Al Marar, M.S. (2019) "Eye in the Sky: How the Rise of Drones Will Transform the Oil & Gas Industry", paper at Abu Dhabi International Petroleum Exhibition and Conference ADIPEC 2018 (2019).
- Ali, R. (2017) "NATO's Little Noticed but Important New Aggressive Stance on Cyber Weapons," [online] *Foreign Policy*, 7 December, <http://foreignpolicy.com/2017/12/07/natos-little-noticed-but-important-new-aggressive-stance-on-cyber-weapons>.
- Andersson, L.M., Okada, H., Miura, R., Zhang, Y., Yoshioka, K., Aso, H. and Itoh, T. (2016) "Wearable Wireless Estrus Detection Sensor for Cows", *Computers and Electronics in Agriculture*, Vol 127, No. 1, September, pp 101-108.
- Bunse, C. and Plotz, S. (2018) "Security Analysis of Drone Communication Protocols", *Lecture Notes in Computer Science* 10953, pp 96-107.
- Chai, R., Naik, G.R., Ling, S.H. and Nguyen, H.T. (2017) "Hybrid Brain–Computer Interface for Biomedical Cyber-Physical System Application Using Wireless Embedded EEG Systems", *BioMedical Engineering OnLine*, Vol 16, No. 5, <https://doi.org/10.1186/s12938-016-0303-x>.
- Costello, P.J. (2017) "Identifying and Exploiting Vulnerabilities in Civilian Unmanned Aerial Vehicle Systems and Evaluating and Countering Potential Threats against the United States Airspace", paper at Proceedings of the Conference on Integrating Technology into Computer Science Education ITiCSE, pp. 761-762.
- Costigan, S.S. and Lindström, G. (2016) "Policy and the Internet of Things", *Connections: The Quarterly Journal*, Vol 15, No. 2, pp 9-18.
- Dan Goodin, D. (2011) "Insulin Pump Hack Delivers Fatal Dosage over the Air", *The Register*, 27 October.
- Delicato, F.C., Zhou, X., Wang, K. I-K. and Guo, S. (2019) "Special Issue: Advances and Trends on Cognitive Cyber-Physical Systems", *Ad Hoc Networks*, Vol 88, pp 1-4.
- Demertzis, K., Tziritas, N., Kikiras, P., Sanchez, S.L. and Iliadis, L. (2019) "The Next Generation Cognitive Security Operations Center: Adaptive Analytic Lambda Architecture for Efficient Defense against Adversarial Attacks", *Big Data Cognitive Computing*, Vol 3, No. 1, 6, <https://doi.org/10.3390/bdcc3010006>.
- Dumitache, I., Sacala, I.S., Moisescu, M.A. and Caramihai, S.I. (2017) "A Conceptual Framework for Modeling and Design of Cyber-Physical Systems", *Studies in Informatics and Control* Vol 26, No.3, September, pp 325-334.
- ECHO (2019) European network of Cybersecurity centres and competence Hub for innovation and Operations, <https://www.echonetwerk.eu/>.

- Emmott, R. (2017) "NATO Mulls 'Offensive Defense' with Cyber Warfare Rules," [online] Reuters, 30 November, <https://www.reuters.com/article/us-nato-cyber/nato-mulls-offensive-defense-with-cyber-warfare-rules-idUSKBN1DU1G4>.
- EU Directive 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union, *OJL* 194, 19 July 2016, pp 1-30.
- Fass, D. and Gechter, F. (2015) "Towards a Theory for Bio-Cyber Physical Systems Modelling", In: Duffy, V., ed., *Digital Human Modeling. Applications in Health, Safety, Ergonomics and Risk Management: Human Modeling. DHM 2015*, Springer, Cham.
- Frenger, P. (2013) "Hacking Medical Devices a Review - Biomed 2013", *Biomedical Sciences Instrumentation*, Vol 49, pp 40-47.
- Gonzalez, C., Ben-Asher, N. and Morrison, D. (2017) "Dynamics of Decision Making in Cyber Defense: Using Multi-agent Cognitive Modeling to Understand Cyberwar," in *Lecture Notes in Computer Science*, Vol. 10030, pp. 113-127.
- IEEE Systems Council (2017) "Cyber Physical Systems Technical Committee", November, <http://www.ieeesystems council.org/pages/cyber-physical-systems-technical-committee>.
- Ionescu, A.M. (2016) "Nanotechnology and Global Security", *Connections: The Quarterly Journal*, Vol 15, No. 2, pp 31-47.
- Jahir, Y., Atiquzzaman, M., Refai, H., Paranjothi, A. and LoPresti, P. (2019) "Routing Protocols and Architecture for Disaster Area Network: A Survey", *Ad Hoc Networks*, Vol 88.
- Jha, A.R. (2010) *Next-Generation Batteries and Fuel Cells for Commercial, Military, and Space Applications*, CRC Press, Boca Raton, FL.
- Keller, J. (2017) "DARPA asks industry to develop built-in cyber security against computer hardware vulnerabilities", *Military & Aerospace Electronics*, April 2017, <http://www.militaryaerospace.com/articles/2017/04/cyber-security-computer-hardware.html>.
- Klonoff, D.C. (2015) "Cybersecurity for Connected Diabetes Devices", *Journal of Diabetes Science and Technology*, Vol 9, No. 5, 16 April, pp 1143-1147.
- Koslowski, R. and Schulzke, M. (2018) "Drones Along Borders: Border Security UAVs in the United States and the European Union", *International Studies Perspectives*, Vol 19, No. 4, November, pp 305-324.
- Kruse, C.S., Frederick, B., Jacobson, T. and Monticone, D.K. (2017) "Cybersecurity in Healthcare: A Systematic Review of Modern Threats and Trends", *Technology and Health Care*, Vol 25, No. 1, pp 1-10.
- Lee, R.M., Assante, M.J. and Conway, T. (2016) "Analysis of the Cyber Attack on the Ukrainian Power Grid: Defense Use Case", [online] Electricity Information Sharing and Analysis Center (E-ISAC), Washington, D.C. [https://ics.sans.org/media/E-ISAC\\_SANS\\_Ukraine\\_DUC\\_5.pdf](https://ics.sans.org/media/E-ISAC_SANS_Ukraine_DUC_5.pdf).
- Lentzos, F. (2016) "Biology's Misuse Potential", *Connections: The Quarterly Journal*, Vol 15, No. 2, pp 48-64.
- Lezzi, M., Lazoi, M. and Corallo, A. (2018) "Cybersecurity for Industry 4.0 in the current literature: A reference framework", *Computers in Industry*, Vol 103, December, pp 97-110.
- Loukas, G., Karapistoli, E., Panaousis, E., Sarigiannidis, P., Bezemskij, A. and Vuong, T. (2019) "A Taxonomy and Survey of Cyber-Physical Intrusion Detection Approaches for Vehicles", *Ad Hoc Networks*, Vol 84, pp 124-147.
- Lukina, A., Tiwari, A., Smolka, S.A., Esterle, L., Yang, J. and Grosu, R. (2018) "Resilient Control and Safety for Cyber-Physical Systems", paper at 3rd Workshop on Monitoring and Testing of Cyber-Physical Systems MT-CPS 2018, 8429482, pp 16-17.
- Marques, P., Manfroi, D., Deitos, E., Cegoni, J., Castilhos, R., Pignaton, E. and Kunst, R. (2019) "An IoT-based Smart Cities Infrastructure Management Architecture", *Ad Hoc Networks*, Vol 88.
- National Institute of Standards and Technology (2017) *Framework for Improving Critical Infrastructure Cybersecurity*, Version 1.1, December.
- O'Malley, J. (2017) "Drones Wide Open to Hijack Threats", *Engineering & Technology*, Vol 12, no. 3, <https://eandt.theiet.org/content/articles/2017/03/drones-wide-open-to-hijack-threats/>.
- Paradiso, R., Faetti T. and Werner, S. (2011) "Wearable monitoring systems for psychological and physiological state assessment in a naturalistic environment", paper at IEEE Conference on Engineering in Medicine and Biology Society, Boston, MA, August-September, 2250-3.
- Pastore, T., Galdorisi, G. and Jones, A. (2017) "Command and Control (C2) to Enable Multi-domain Teaming of Unmanned Vehicles (UxVs)", OCEANS, pp 1-7.
- Paul, N., Kohno, T. and Klonoff, D.C. (2011) "A review of the security of insulin pump infusion systems", *Journal of Diabetes Science and Technology*, Vol 5, No. 6, 1 November, pp 1557-1562.
- Payne, K. (2018) *Strategy, Evolution, and War: From Apes to artificial intelligence*, Georgetown University Press, Washington, D.C.
- Penido, É.C.C., Teixeira, M.M., Fernandes, H.C., Monteiro, P.B. and Cecon, P.R. (2019) "Development and Evaluation of a Remotely Controlled and Monitored Self-propelled Sprayer in Tomato Crops", *Revista Ciencia Agronomica*, Vol 50, No. 1, pp 8-17.
- Pocheptsov, G. (2018) "Cognitive Attacks in Russian Hybrid Warfare", *Information & Security: An International Journal*, Vol 41, pp 37-43.
- Ransford, B., Kramer, D.B., Kune, D.F., De Medeiros, J.A., Yan, Ch., Xu, W., Crawford, T. and Fu, K. (2017) "Cybersecurity and Medical Devices: A Practical Guide for Cardiac Electrophysiologists", *Pacing and Clinical Electrophysiology*, Vol. 40, No. 8, August, pp 913-917.

- Rehman S., Gruhn V., Shafiq S., Inayat I. (2018) "A Systematic Mapping Study on Security Requirements Engineering Frameworks for Cyber-Physical Systems", In: Wang G., Chen J., Yang L. (eds) Security, Privacy, and Anonymity in Computation, Communication, and Storage. SpaCCS 2018. Lecture Notes in Computer Science, vol 11342. Springer, Cham.
- Rogers, J.A. (2012) "Bio-integrated electronics", paper at 2012 IEEE International Electron Devices Meeting, San Francisco, CA, December.
- Sang C. Suh, S.C., Tanik, U.J., Carbone, J.N. and Eroglu, A. eds. (2014) *Applied Cyber-Physical Systems*, Springer, New York.
- Sastry, S.S. (2015) "The Cyber-Bio-Physical Research Frontier", [online], *Berkeley Engineering*, 16 April, <http://engineering.berkeley.edu/2015/04/cyber-bio-physical-research-frontier>.
- Scharre, P. (2018) "How Swarming Will Change Warfare", *Bulletin of the Atomic Scientists*, Vol 74, No. 6, November, pp 385-389.
- Science and Technology Committee (2016), *Robotics and Artificial Intelligence*, House of Commons, London.
- Storm, D. (2015) "Researchers Hack a Pacemaker, Kill a Man(nequin)", *Computerworld*, 8 September.
- Symantec (2018) Internet Security Threat Report, Vol 23, March [online], <https://www.symantec.com/content/dam/symantec/docs/reports/istr-23-2018-en.pdf>.
- Tabasum, A., Safi, Z., Alkhater, W. and Shikfa, A. (2018) "Cybersecurity Issues in Implanted Medical Devices", 2018 International Conference on Computer and Applications, ICCA 2018, Beirut, August, 8460454, 110-115.
- Valášek, T. (2017) "How Artificial Intelligence Could Disrupt Alliances," [online], *Carnegie Europe*, 31 August, <http://carnegieeurope.eu/strategiceurope/72966>.
- Veneva, I.P., Chakarov, D., Tsveov, M., Trifonov, D.S., Zlatanov, E. and Venev, P. (2018) "Active Assistive Orthotic System: (Exoskeleton) Enhancing Movement", in Habib, M., ed., *Handbook of Research on Biomimetics and Biomedical Robotics*, IGI Global, Hershey, PA, pp. 48-75.
- Wardle, C. and Derakhshan, H. (2017) *Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking*, Council of Europe report DGI(2017)09, Strasbourg, 31 October 2017.
- Wilkins, J. (2018) "Is Artificial Intelligence a Help or Hindrance?", *Network Security*, Vol 2018, No. 5, May, pp 18-19.
- Yeo, W.-H., Webb, R.C., Lee, W., Jung, S. and Rogers, J.A. (2013) "Bio-integrated Electronics and Sensor Systems", Proceedings Volume 8725, *Micro- and Nanotechnology Sensors, Systems, and Applications V*; 87251; SPIE Defense, Security, and Sensing, Baltimore, Maryland, United States.
- Zhang, M., Ali, S., Yue, T., Norgren, R. and Okariz, O. (2017) *Uncertainty-Wise Cyber-Physical System Test Modeling*, Technical report 2016-02, Simula Research Laboratory.

# Cyber Resilience Strategy and Attribution in the Context of International law

Pardis Moslemzadeh Tehrani

Faculty of Law, University of Malaya, Malaysia

[pardismoslemzadeh@um.edu.my](mailto:pardismoslemzadeh@um.edu.my)

**Abstract:** Recent cyber-attacks around the globe, which are largely due to the emergence of the internet of things (IOT), indicate that what was until recently a futuristic threat has now become a reality. The 'fifth dimension war' portrays 'cyber warfare' as a new battlefield which impels states to protect their critical infrastructure against attacks. Large-scale disruption of critical infrastructure caused by cyber-attacks undermines confidence in the state, emphasizing outdated security policies. It is difficult to accurately attribute the source, purpose or motives of cyber-attacks. The anonymity conferred by encrypting and hiding the sources of attack facilitates the implementation of cyber-attacks by terrorist organizations. It is impossible to prevent all attacks on critical infrastructure, even if protection measures are implemented appropriately. This implies that technology and updated infrastructure alone are not sufficient as a response. Additional measures along with specific procedures to facilitate faster recovery of assets that have suffered cyber-attacks are essential in responding to damage to critical infrastructure. Cyber resilience's key concepts such as prevention, preparedness and response are the new policies of developed regions. This article aims to further illuminate the underlying principles of national critical infrastructure defence in cyber warfare with the objective of providing a cyber resilience strategy. It also attempts to examine the issue of attribution and the applicability of international rules to state and non-state actors for malicious cyber activities in the context of attribution. The examination is conducted via a multilevel legal analysis.

**Keywords:** cyber resilience, attribution, cyber terrorism, critical infrastructure, international law

---

## 1. Introduction

The nature of cyberspace, which is constantly changing and provides good cover and anonymity, makes the origin of cyber-attacks very difficult to pinpoint. Critical assets and information structures must be better understood, protected and maintained against cyber threats. Technical skills to counter the growing risk of cyber threats are limited. Often the weaknesses found in organizations in post-event analysis include porous controls around intrusion detection, monitoring and incident response. To counter this, organizations must subject their code and their systems to both internal and external examination to ensure robust cyber security configurations able to withstand attack. Critical capabilities include not just prevention and protection strategies but those enabling them to recover quickly following attacks or absorb and frustrate such attacks (Masys, 2015).

Cyber resilience helps rebalance the cost-benefit calculation in favour of defenders. All nations, particularly those at most risk, should reorient their policies based on this principle. Due to rapid technological advances, it is critical for nations to develop, implement and adhere to recovery plans that enable them to re-establish operational capacity after cyber-attacks. (Peter, 2017). Cyber capability is considered an integral part of the strategic diplomacy of all countries. However, evidence shows that states also have an interest in employing the cyber capabilities of non-state actors, and that they use either espionage or sabotage to execute attacks that are below the threshold of armed attack to in order to protect themselves from potential punishment (Robinson, 2016).

Statistical estimates imply that significant cyber vulnerabilities and hazards may leave the critical infrastructure of countries open to attack and even incapacitation. That is, cyber criminals now have the ability to 'crash' critical systems by exploiting cyber weaknesses due to the increasing penetration of the Internet. It is because of this that despite the existence of a broad legal regime to combat cybercrime, it is problematic to legally classify a cyber-attack against infrastructure as a use of force or an armed attack. There are many reasons why it is difficult to apply the rules of international humanitarian law to cyber-attacks. Chief among these is the difficulty of estimating the impact of a given cyber-attack in a proper manner and determining the identity and political motivation of the perpetrator/s. International groups of experts have attempted to address these issues by the creation of the Manual on the International Law Applicable to Cyber Warfare (the so-called "Tallinn Manual") which is the most comprehensive existing guide to the applicability of international law rules to cyber conflict, cyber warfare and cyber-attacks (Pipyros, Mitrou, Gritzalis, & Apostolopoulos, 2016).

Despite such progress, there is still confusion regarding the applicability of international law to cyber warfare. State and non-state actors may use cyber-attacks in the context of military operations. In recent events, cyber-attacks have been used as part of hybrid warfare. Cyber hacking by states is not expressly regulated or controlled in the current international regime. There is no specific principle of international law applicable, and specific treaties that expressly regulate the use of cyber-attacks have not yet been developed. While scholars point to the potential application of the law of armed conflict, this law has not been invoked thus far to respond to cyber-attacks (Hathaway et al., 2012).

This article explores the term 'national critical infrastructure' and considers ways in which such infrastructure can be targeted in cyber-attacks. It then discusses the notion of cyber resilience as an alternative to deterring cyber-attacks. The discussion of cyber resilience is followed by a discussion of the issue of attribution as an element of a resilience strategy, since a possible response to cyber-attack requires identifying the offender and responding accordingly. Finally, the article demonstrates how existing international law can be applied to cyber warfare, how it is deficient and what may be done to improve it.

## **2. National critical infrastructure**

Critical infrastructure refers to assets which are fundamental to the functioning of a society, that is, infrastructure which provides essential services which underpin the functioning of a society. It constitutes the 'backbone' of a nation's economy, security, and health ("What Is Critical Infrastructure?" 2017). Critical infrastructure also includes certain functions, sites and organizations which although not critical to the maintenance of essential services need protection due to the potential danger to the public their malfunction may cause, such as nuclear and chemical sites ("Critical National Infrastructure," 2018). A good example of disruption of critical infrastructure, and possibly the most instructive example to date, is the Stuxnet attack of 2010 which caused substantial damage to Iran's nuclear program. Critical infrastructure is defined and viewed differently in different countries due to divergent perspectives on issues such as national interest, security policy and defence development. It is commonly analysed by sectors; the United States has 16 sectors, while the UK has 13. (Lopez, Setola, & Wolthusen, 2012).

The term 'critical national infrastructure' is defined by the Centre for the Protection of National Infrastructure (CPNI) in the UK as: "those facilities, systems and sites necessary for the delivery of the essential services upon which daily life ... depends and which ensure the country continues to function socially and economically" (Current and emerging trends in cyber operations: Policy, strategy and practice, 2015). Following an attempted spear-phishing attack, the CPNI released new guidance to educate the public which explained spear-phishing as "a targeted form of email deception that results in exploitation or compromise of individual devices and organizational networks". It also explained more complex issues such as geo-spatial and multidimensional boundaries and hierarchical control systems such as SCADA systems across the UK. The potential threats posed to national interests have motivated all stakeholders to ensure cyber initiatives are coordinated in line with national risk assessments which were previously established (Lemieux, 2015).

The situation in the United States is somewhat different from that in the UK. Attempts to protect critical infrastructure commenced under President George W. Bush following the 2007 distributed denial of service (DDOS) attack on Estonia. His government established the Comprehensive National Cyber Security Initiative (CNCI) in 2008 which was declassified in March 2010 when President Barak Obama released public information on the CNCI and its main recommendations. The National Infrastructure Protection Plan (NIPP), which evolved from concepts introduced in 2008, outlines how government and private sector participants in the critical infrastructure community can work together to manage risks and achieve security and resilience outcomes by 2013. It is updated regularly based on risk, policy, and strategic environments. It attempts to provide a foundation for an integrated approach to achieve "[a] nation in which physical and cyber critical infrastructure remains secure and resilient, with vulnerabilities reduced, consequences minimized, threats identified and disrupted, and response and recovery hastened."

The existence of national critical infrastructure enables perpetrators to attack strategic targets without physical intervention for the first time in history. It is widely perceived that the target of cyber-terrorists is critical infrastructure. The recent examples of cyber-attacks against critical infrastructure indicate the ongoing vulnerabilities and illustrate the urgent need to replace protection with resilience. A proper level of cyber security cannot be achieved without the necessary effort to improve the resilience of critical infrastructure.

### **3. Cyber resilience strategy**

Resilience in terms of cyber infrastructure is commonly understood as the capability of the critical infrastructure or service to either rapidly “bounce back” following an attack or absorb and frustrate the potential of attacks. Such strategies are complementary to existing prevention and protection policies which are the pillars of current critical infrastructure protection programs. Heath-Kelly defines resilience as “attempts to avoid security failure through retrospective, anticipatory and “decapitalized” operations, rather than actually addressing disaster recovery” (Heath-Kelly, 2015).

There was no definition before the 1990s for the term ‘critical infrastructure’. (Lewis, 2014). The term appears to have been introduced in recognition of a world which deals with imminent threats which lead to the development in advance of resilience strategies able to confront emergent disasters which must be thought of, prepared for, and dealt with continuously. Scholars have divided resilience into different categories with various typologies. Theoretical resilience scholarship focuses on how resilience is suited to liberal notions of security, while empirical resilience scholarship attempts to put the policy-driven incorporation of resilience into a growing number of social domains. Governmental resilience philosophy depicts resilience as an optimal recovery from an adverse event through security processes (Dunn Cavelti et al., 2015). Technologies also need to be prepared for the actual process of ‘coping’ which makes the resilience discourse “a discourse of futurity” (Schott, 2013). This discussion has brought scholars to the point that, although there are some contradictions in the existing literature on what resilience is and how different types of resilience work, resilience needs to be studied in many different forms and contexts. It is no longer the only paradigm with security relevance.

### **4. Deterrent strategy and attribution**

The structure and the anonymity of the internet and the difficulty in identifying the person behind the attack rather than the source machine of an attack and are highly salient to the issue of attribution to individuals or states (“The Law of Attribution: Rules for Attributing the Source of a Cyber-Attack,” 2017). This concern implies the need to create a legal solution to the problems posed by attribution. The combination of all these pieces of the puzzle makes it difficult to create actual legal or political frameworks for attribution (Rid & Buchanan, 2015). Making a successful attribution is a nuanced and multi-layered process requiring careful management, training, and leadership. Attribution is a political stakes game and governments must invest in identifying the perpetrators based on the damage level and importance of the target. Individual governments must decide how to approach attribution and when attribution has been established well enough for action. Attribution is important for governments since any response to a specific offence requires first identifying the offender. In many cases, the challenges of establishing attribution allow offenders to stay online and hide their identity behind the attribution problem.

The difficulty in detecting malicious threats that affect networks causes many intrusions to go unpunished. Policy makers attempt to utilize deterrence, which has been used throughout history as a useful tool to discourage unwanted behaviour. States use this policy as rational actors who try to make decisions via cost-benefit analyses. Such deterrence can be done by changing the decision-making calculus, punishment and denial (Nevill & Hawkins, 2016). Deterrence by punishment features prominently in the national security strategy of various countries. During the Cold War it was widely used and played a fundamental role. The most industrialised and connected countries are the most vulnerable countries, while less advanced and thus less vulnerable countries have an advantage in this respect. States need to identify the ‘red line’ to distinguish between accepted and unaccepted behaviour for punishment. Threats must also be supported by the ability, resources and intent to follow through by inflicting punishment, and punishment needs to be appropriate and informed by contemporary norms and international law. The deterrent by punishment approach is the most popular defensive approach, partly due to its high profile during the Cold War; other approaches are not as popular due to the difficult nature of their application. The US, the UK and Australia have attempted to achieve deterrence through a combination of punishment and denial in order to pursue their cyber stability by covering all bases and establishing a level of complexity in applying deterrence concepts to cyberspace. It must be borne in mind that deterring cyber actions through threat is not as easy as deterring physical attacks. Establishing a deterrent framework requires a variety of steps which are quite difficult and complicated to carry out in cyber space. Obstacles to effective deterrence include establishing a threshold, communicating threats, detecting intrusions and attributing responsibility.

As mentioned above, distinguishing between acceptable and unacceptable behaviour is significant since it draws the ‘red line’. It is necessary for the deterrence policy to define clearly the punishable behaviour. According to current definitions, a cyber-attack must cause physical destruction or loss of life. Therefore, other malicious acts such as cybercrime and cyber espionage do not attain the threshold. It is important that this grey zone be defined since it needs to be addressed in national security policy. Appropriate punishments for malicious behaviour must be defined not only to comply with international law but also to draw a direct connection between the behaviour and the punishment to reinforce the deterrence policy (Jensen, 2012). According to the principle of proportionality, the retaliation must reflect the scale and scope of the original act. It is necessary to agree on what constitutes an appropriate response to cyber acts and how to distinguish between them to establish international cyber norms. A framework has been proposed by Tobias Feakin, Australia’s inaugural Ambassador for Cyber Affairs, to identify proportionate real-world responses to different types of cyber incidents; however, it the framework has not been officially adopted and integrated into international normative behaviour at this time. (Feakin, 2015)

Another issue regarding deterrence strategies is that in most cases malicious activities remain unpunished as the threats do not cross the chosen threshold. The difficulty in detecting all malicious incidents also affects deterrence strategy and many intrusions go unpunished. Basically, all threats that are lower than the threshold remain unpunished, which undermines deterrent threats. The approach of online actors is to launch numerous low-intensity attacks to remain unnoticed; in case any single act is detected, that act is likely to fall below an adversary’s threshold for retaliation. The assumption is that these actions will probably either not be noticed or will not trigger a response. The ability to avoid retaliation by use of the strategy outlined above causes the credibility of retaliation threats to be undermined.

The immediacy of a punishment is also very important as it correlates with its effectiveness in establishing a deterrent. The time lapse between the crossing of a red line and the associated retribution make it difficult to identify the connection between the two. Thus, the challenges of detecting the offender affect a defender’s ability to establish deterrence by punishment in cyber space. Such inconsistency confuses public perception of thresholds and reduces the credibility of the deterrence posture. This disparity between the broad range of potential actions in cyber space, weak detection and unclear response frameworks is yet to be effectively addressed (Goodman, 2010). Notwithstanding the fact that effective attribution offers multiple benefits such as strengthening the existing deterrence framework and improving collective security, the difficulty of accurately attributing responsibility makes it hard to establish a credible deterrent threat (Rid & Buchanan, 2015). The attribution process in cyberspace is challenging, and even if the perpetrator has been identified by the cyber forensic specialist, the process of imposing punishment is often slow, unreliable or impossible. The source of an attack may be masked by a variety of techniques including botnet, proxy servers and onion routing. Identifying an IP address as the source of an attack does not offer conclusive evidence about the identity and intention of the perpetrator (Council, 2010). The inherent difficulty of attribution was demonstrated by the hacking of French television channel, TV5 Monde, in April 2015. The sophisticated incident was initially identified as an act of the Islamic State, but was later found more likely to be the handiwork of Russian-based hackers (Nevill & Hawkins, 2016). Several types of sanctions are available that might be justified by a legal chain of attribution. Available sanctions include economic punishment, denial of participation in future international treaties and agreements, hack-back countermeasures, unilateral and multilateral action and military responses.

## **5. International rules relating to the responsibilities of states and non-state actors**

The importance of attribution becomes clear in the discussion of state responsibility since it is essential for states to sanction responsible malicious actors. In addition, attribution assists a state in claiming legitimacy for sanctions or countermeasures. The purpose of the law is to determine the outcome of a conflict while legitimizing that outcome determination to others. Such sanctions are applied in the context of international law and international relations, since in this context states lack an overarching authority to compel compliance via force. Therefore, they need to operate under norms established and legitimized by customary international law (Goldsmith & Posner, 1999).

Cyber capabilities can be used in a range of ways. There are two main types of cyber-attack which are most often applied by non-state actors employed by states. The first is cyber-enabled espionage – at the strategic or operational level – which compromises the confidentiality of information and information systems and has the potential to distribute secrets and sensitive information to adversaries. The second type is cyber-enabled

sabotage, which can cause severe physical ramifications, especially when targeting infrastructure such as energy or transportation networks or when used to confuse the target and undermine command and control decision-making by manipulation (Robinson, 2016).

As mentioned previously, attribution is necessary to validate subsequent legal action. In order to do this, states need to identify the source of an attack as well as legitimize its attribution to the satisfaction of other state actors in order to justify whatever recourse or countermeasure follows ("The Law of Attribution: Rules for Attributing the Source of a Cyber-Attack," 2017). In the context of attribution, the same adjustment of law can account for differences in punishment after the attribution of an attack to a state. Different states may interpret *mens rea* in different ways since states are composed of a multitude of individuals with a variety of mindsets. However, it is possible for states to adjust their standards of scrutiny of the law of attribution based on the burden of proof it requires. The standard of proof as a core element of procedure can be notched higher or lower based on the severity of the chosen remedy. The *mens rea* requirement is only one means of fine tuning the burden of proof, while the evidentiary standard of proof presents another holistic way to incorporate the seriousness of a penalty into the generalized requirements of a procedural framework.

A recent Cyber Conflict Project carried out by Yale University's Centre for Global Legal Challenges contributes significantly to the discussion of this matter by proposing the application of an adversarial legal framework, i.e. one primarily characterized by impartial decision makers issuing judgments on disputes based on evidence and arguments presented by parties and legal representatives ("The Law of Attribution: Rules for Attributing the Source of a Cyber-Attack," 2017). The adversarial model is well suited to the context of attribution since attribution frequently relies on technical evidence and concerned parties are in the best position to acquire and present such evidence. The adversarial system for an attribution framework points the way towards a general rule of what a widely accepted law of attribution in respect of cyber infringements might look like. A significant advantage of the adversarial system is that it has an impartial adjudicator and will be driven by the concerned parties in terms of both legal argumentation and the production of facts and evidence. The Centre for Global Legal Challenges argued that the current consensus on a preponderance of the evidence standard fits the goals of attribution as it provides the optimal balance of deterrence and information production since it lowers the barriers to the establishment of attribution (and hence, increases the potential for countermeasures) while still requiring a requisite level of persuasion that would incentivize the production of relevant intelligence and information regarding the cyber-attack. When it is ruled that a state is responsible for launching a cyber-attack as a result of the above-mentioned steps being carried out, the accused state might attempt to defend itself from attribution by placing the blame on a non-state actor. Therefore, the law of attribution comes into play by following the trail to an individual hacker and then attempting to connect that person to a state in terms of legal responsibility. The doctrine of state responsibility has long existed beyond the realm of cyber-attacks and has been addressed frequently in other contexts. International law already possesses a state responsibility doctrine for attributing the malicious behaviour of non-state actors to states.

International Law Commission Draft Articles (2001) sets out ways for international states to be held responsible for the acts of non-state actors. Articles 4 and 8 of the Draft have been recognized as customary international law by the ICJ and are held to set the standard interpretation of the state responsibility doctrine. Both the first and second edition of the Tallinn Manual draw on the ILC's Draft Articles to formulate their conception of the state responsibility doctrine in the context of cyber-attacks. According to the Draft, if a non-state actor is acting as an organ of the state or is acting under the instructions, directions, or control of the state, the wrongful behaviour of the non-state actor is attributable to that state. It also holds actions attributable to a state in cases where the action has been conducted by individuals who may be recognized as organs of the state. It also extends to individuals who may be considered de facto organs of the state (Scharf & Day, 2012). It has been understood that the conditions described in Article 4 and 8 set tests for the notion of control. Such control tests have been employed in rulings by courts such as the ICJ.

There is currently a significant debate about the limitations of the existing international law as regards state responsibility. The current framework was criticized by Oona Hathaway, who argued that states can still escape responsibility by handing illegal tasks to non-state actors. She also believes that the control test may actually disincentivize state efforts to control rogue or malicious behaviour, since that test might make it impossible to hold the state responsible for wrongdoing. The control doctrine is also criticized by Peter Margulies, who notes that it requires specific, comprehensive control to be established, and claims that such a standard would exclude very significant examples of states directing non-state actors to conduct cyber-attacks. These comments led to

the proposition of a solution, the “virtual test”. This test, suggested by Margulies, shifts the burden to states to demonstrate non-responsibility in cases where states funds and equip private entities or parties who engages in cyber-attacks. According to Hathaway, this approach carries risks as the potential attachment of liability to any existing relationship between the government and a non-state actor who has carried out wrongful acts might incentivize governments to relinquish control over non-state actors within its reach. The argument might be made that the “funding and equipping” requirement is an indicator of state contribution and must be followed by the control test, will be applied only in cases that states materially supports such entities. Further, the concept of funding and/or equipping a non-state entity is not clearly defined by Margulies. Hathaway further notes that states should have access to an affirmative defence if they can prove that they took reasonable steps to prevent violations of international law.

The combined weight of the above arguments could support the proposition that the law of attribution, properly re-interpreted and applied to cyber situations, may potentially be useful in addressing the novel challenges of cyber-attack attribution. The virtual control test along with an affirmative defence of “reasonable care” would allow the doctrine of attribution to attribute individuals’ cyber-attacks to states, while allowing states sufficient means of protecting themselves from liability by taking good faith measures to prevent wrongdoing (“The Law of Attribution: Rules for Attributing the Source of a Cyber-Attack,” 2017).

## **6. Conclusion**

The Stuxnet cyber-attack revealed how vulnerable states are to cyber-attack and how important it is for states to protect their national critical infrastructure. The attack indicated that while the threat of cyber-attacks has grown rapidly, response capabilities have not kept pace. The difficulty and uncertainty in tracing cyber-attack and attributing them correctly must lead states to prioritize cyber resilience. Cyber resilience interweaves different temporal strands which reinforce long-term security for protection of critical infrastructure. Cyber resilience abilities can affect opponents’ decision-making and deter the deployment of attacks. As such, cyber resilience constitutes a form of cyber defence and enables states to protect national infrastructure which may otherwise be targeted by terrorist organizations. Damage to critical infrastructure can bring states to their knees. Having a comprehensive cyber resilience policy implies the ability to verify where cyber-attacks originate, to establish advance cyber policy and to having the ability to deploy traceback and related forensic tools which can go some way towards establishing attribution. Establishing an attribution policy is the first step in developing an advanced cyber resilience policy.

This discussion has shown how important it is for states to establish effective cyber resilience policies to face new and growing threats. This is particularly important for highly developed and connected countries such as the United States, the United Kingdom and others, and such states are already taking steps towards such policies. The discussion has also shown how the concept of attribution and its applicability to state actors behind non-state wrongdoers may be used by states to justify counter measures. The doctrine of state responsibility serves as a means to establish responsibility of states for the acts of non-state actors in specific conditions. Although concerns have been raised by scholars such as Hathaway and Margulies, the law of attribution is currently poised to develop in new and fruitful ways in order to address novel challenges. The doctrine of attribution may be used to protect states from liability in appropriate cases which also allowing for the attribution of individuals’ cyber-attacks to states.

## **References**

- Council, N. R. (2010). Proceedings of a Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options for US Policy: National Academies Press.
- Critical National Infrastructure. (2018).
- Current and emerging trends in cyber operations: Policy, strategy and practice. (2015). (F. Lemieux Ed.). USA: Palgrave Macmillan.
- Dunn Cavelti, M., Kaufmann, M., & Søby Kristensen, K. (2015). Resilience and (in) security: Practices, subjects, temporalities. *Security Dialogue*, 46(1), 3-14.
- Economics of Information Security and Privacy. (2010). (T. M. D. J. P. C. Ioannidis Ed.): Springer.
- Feakin, T. (2015). Developing a Proportionate Response to a Cyber Incident. *Council on Foreign Relations*, Aug. <http://www.cfr.org/cybersecurity/developing-proportionate-response-cyber-incident/p36927>.
- Goldsmith, J. L., & Posner, E. A. (1999). A theory of customary international law. *The University of Chicago Law Review*, 1113-1177.
- Goodman, W. (2010). Cyber deterrence: Tougher in theory than in practice? Retrieved from

- Hathaway, O. A., Crootof, R., Levitz, P., Nix, H., Nowlan, A., Perdue, W., & Spiegel, J. (2012). The law of cyber-attack. *California Law Review*, 817-885.
- Heath-Kelly, C. (2015). Securing through the failure to secure? The ambiguity of resilience at the bombsite. *Security Dialogue*, 46(1), 69-85.  
[<integamate.52.3.fm.pdf>](#).
- Jensen, E. T. (2012). Cyber Deterrence. *Emory Int'l L. Rev.*, 26, 773.
- The Law of Attribution: Rules for Attributing the Source of a Cyber-Attack. (2017).
- Lemieux, F. (2015). Current and emerging trends in cyber operations: Policy, strategy and practice: Springer.
- Lewis, T. G. (2014). Critical infrastructure protection in homeland security: defending a networked nation: John Wiley & Sons.
- Lopez, J., Setola, R., & Wolthusen, S. (2012). Critical Infrastructure Protection: Advances in Critical Infrastructure Protection: Information Infrastructure Models, Analysis, and Defense (Vol. 7130): Springer.
- Masys, A. J. (2015). The Cyber-Ecosystem Enabling Resilience Through the Comprehensive Approach Disaster Management: Enabling Resilience (pp. 143-154): Springer.
- Nevill, L., & Hawkins, Z. (2016). Deterrence in cyberspace.
- Peter, A. S. (2017). Cyber resilience preparedness of Africa's top-12 emerging economies. *International Journal of Critical Infrastructure Protection*, 17, 49-59.
- Pipyros, K., Mitrou, L., Gritzalis, D., & Apostolopoulos, T. (2016). Cyberoperations and International Humanitarian Law: A review of obstacles in applying International Law rules in Cyber Warfare. *Information & Computer Security*, 24(1), 38-52.
- Rid, T., & Buchanan, B. (2015). Attributing cyber attacks. *Journal of Strategic Studies*, 38(1-2), 4-37.
- Robinson, N. (2016). NATO: changing gear on cyber defence. NATO, available at: [www.nato.int/docu/review/2016/Also-in-2016/cyber-defense-nato-security-role/EN/](http://www.nato.int/docu/review/2016/Also-in-2016/cyber-defense-nato-security-role/EN/) [accessed 15 June 2016].[Google Scholar].
- Scharf, M. P., & Day, M. (2012). The International Court of Justice's Treatment of Circumstantial Evidence and Adverse Inferences. *Chi. J. Int'l L.*, 13, 123.
- Schott, R. M. (2013). Resilience, normativity and vulnerability. *Resilience*, 1(3), 210-218.
- What Is Critical Infrastructure? (2017).

# A Biological Framework for Characterizing Mimicry in Cyber-Deception

Steven Templeton<sup>1</sup>, Matt Bishop<sup>1</sup>, Karl Levitt<sup>1</sup> and Mark Heckman<sup>2</sup>

<sup>1</sup>Department of Computer Science, University of California, USA

<sup>2</sup>Department of Computer Science, University of San Diego, USA

[templets@cs.ucdavis.edu](mailto:templets@cs.ucdavis.edu)

[bishop@cs.ucdavis.edu](mailto:bishop@cs.ucdavis.edu)

[levitt@cs.ucdavis.edu](mailto:levitt@cs.ucdavis.edu)

[mrheckman@usd.edu](mailto:mrheckman@usd.edu)

**Abstract:** Deception, both offensive and defensive, is a fundamental tactic in warfare and a well-studied topic in biology. Living organisms use a variety deception tools, including mimicry, camouflage, and nocturnality. Evolutionary biologists have published a variety of formal models for deception in nature. Deception in these models is fundamentally based on misclassification of signals between the entities of the system, represented as a tripartite relation between two signal senders, the “model” and the “mimic”, and a signal receiver, called the “dupe”. Examples of relations between entities include attraction, repulsion and expected advantage gained or lost from the interaction. Using this representation, a multitude of deception systems can be described. Some deception systems in cybersecurity are well-known. Consider, for example, all of the many different varieties of “honey-things” used to ensnare attackers. The study of deception in cybersecurity is limited compared to the richness found in biology. While multiple ontologies of deception in cyber-environments exist, these are primarily lists of terms without a greater organizing structure. This is both a lost opportunity and potentially quite dangerous: a lost opportunity because defenders may be missing useful defensive deception strategies; dangerous because defenders may be oblivious to ongoing attacks using previously unidentified types of offensive deception. In this paper, we extend deception models from biology to present a framework for identifying relations in the cyber-realm analogous to those found in nature. We show how modifications of these relations can create, enhance or on the contrary prevent deception. From these relations, we develop a framework of cyber-deception types, with examples, and a general model for cyber-deception. The signals used in cyber-systems, which are not directly tied to the “Natural” world, differ significantly from those utilized in biologic mimicry systems. However, similar concepts supporting identity exist and are discussed in brief.

**Keywords:** cyber-deception, mimicry, bio-inspired computing, computer security

---

## 1. Introduction

Deception has been a tactic of survival in Nature for millions of years, and has likely been a tactic in military operations since shortly after the first opposing forces met. As warfighting has changed, so have the tactics of deception. The reliance upon information systems in modern warfare requires cultivating new deception tactics. Whereas humans have studied deception for several millennia, it has been a part of Nature for hundreds of millions of years (Garrouste, et al., 2016). Given the richness of deception techniques found in the natural world, it can provide insight into how we may propose new forms to support our offensive and defensive needs.

Deception in the forms of mimicry and crypsis (hiding) have been documented by biologists since at least the time of Aristotle (Peck, 1965). The scientific study of how organisms can, through deception, gain advantage from an adversary includes subjective taxonomies of the features of mimicry systems, as well as mathematical models of interactions between the organisms involved in the deception. Biologists have identified dozens of different types of mimicry, each detailing a class of interactions. This lends itself to collections of attributes and labels but does not look at the interactions in a general scheme not directly tied to the classified examples. Wickler (1965), Vane-Wright (1976), Pasteur (1982), Starret (1982) and others developed and expanded conceptual frameworks for mimicry based on the abstract nature of the interaction. Using these frameworks examples of deception, both mimicry and crypsis, can be analysed. This can be extended to examples not just from biology, but examples from military tactics, and for deception in cyber-environments.

Deception in cyber-environments has been studied in the past yielding taxonomies (Rowe & Rothstein, 2004) and methods to discover mimicry attacks in restricted settings (Giffin, et al., 2006). While academically interesting, the utility of narrow approaches or of cataloguing a multitude of aspects characterizing specific examples of deception is limited. Without a unifying science of the relations between entities in a deception system, the ability to develop or identify unknown forms is doubtful. The goal of this work is to understand

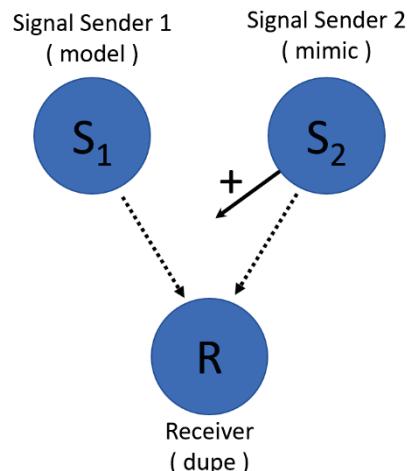
cyber-deception using a scientific framework allowing us to discover new forms of deception and the ability to detect them.

Historically, the majority of examples of deception documented from nature fall into a few categories of interaction; however, rare, unusual, and clever examples in other categories exist. This is true with cyber-deception as well; most known examples will be variants of the same types of interaction. Accordingly, efforts to detect deception have focused on these common classes. By looking at under-represented (or unrepresented) classes of deception, new methods of deception may be designed, and accordingly, allow for the creation of methods to detect previously unconsidered possibilities for deception.

Mimicry and crypsis are intentional; there is intent to deceive. Whether offensive or defensive, deception is active. This is an important idea for the future and must be expanded greatly beyond "honey-things", first implemented in a significant way beginning with the Deception Toolkit (Cohen & others, 1998). Deception will be a significant defensive component and will play a prominent role in systems employing offensive countermeasures. Both of these are well represented in Nature.

Although the words deception, crypsis, camouflage, masquerade, and mimicry may have specific definitions, they often overlap, and are commonly used loosely. Many authors have independently defined new terms to define both specific and general concepts. For example, the term *adaptive resemblance* has been used emphasizing the evolutionary aspect (Starrett, 1993). Rather than adopt any one in particular, unless the distinction is important to note, we use the term *mimicry* as a generalization for all, and crypsis where the primary function is concealment.

Mimicry may be defined formally as a system in which a sender ( $S_2$ , *the mimic*) presents effectively identical signals as that of a different organism ( $S_1$ , *the model*), where both have in common at least one receiver ( $R$ , *the dupe*) that reacts to both signals in the same way because it is advantageous to the receiver to react in that way to the signals of the model, although it may be disadvantageous to react in such a way to the counterfeit signals (Figure 1). The mimic always gains advantage from the dupe's inability to distinguish between signals of the model and the mimic. In this context, the term signal is a generalization for information transmitted from the senders to the receiver, regardless of how the information is communicated. We adopt this generalized concept of model, sender, signals and receiver from that originally described by Wickler as it avoids characterizing participants as enemy, predator, etc.

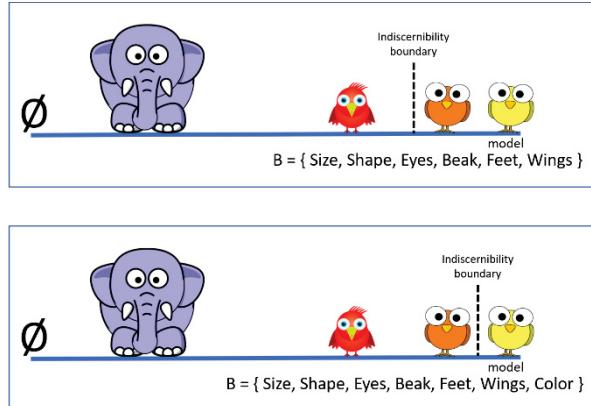


**Figure 1:** Illustration of basic tripartite mimicry system. Both model and mimic send signals to the dupe resulting in the mimic gaining advantage from the system

The distinction between crypsis and mimicry is not always clear, and a particular method may have elements of both. However, in general, in mimicry the sender's signals are intended to be perceived by the receiver, whereas with crypsis the sender's goal is to avoid detection by imitating a background that is neutral to the receiver.

Studying mimicry involves identifying key communicative signals and their effect upon the relevant receiver and the resultant effect upon the sender. Additionally, understanding the cost/benefit of these signals to the mimic is important in evaluating the value of mimicking a particular set of signals.

If we consider a receiver's perception of a spectrum of types of signals, ranging from having no similarity to the model to actually being the model, we can place sender  $S_2$  along this spectrum based on, how discernible  $S_2$  is from the actual model. The shorter the distance between  $S_2$  and the model, the more effective  $S_2$  will be as a mimic. Once sender  $S_2$  has passed the indiscernibility boundary, mimicry will, with respect to the receiver's capabilities, be effective. By including those features that, relative to the capabilities of the receiver, provide the highest level of mimicry with the least cost and least likelihood of failure, the more valuable will be the method of mimicry. Conversely, for the receiver, perceiving cost-effective signals not available to the mimic that distinguish the mimic from the model, can reveal the attempted deception.



**Figure 2:** Discernability between model and mimic depends on signals employed. Based on the signal set of size, shape, eyes, beak, and feet, the elephant and left bird are distinguishable from the model, however the mimic is not. This results in effective mimicry

As an example, consider the task of mimicking a particular type of bird. Features (signals) can include the model bird's size, shape, eyes, beak, feet, wings and colour (Figure 2). Other signals could include heart-rate, temperature, blood count, or DNA. If the dupe (Receiver) only perceived size, providing any of the other signals would be, relative to the receives capabilities, unnecessary and wasteful. Similarly, UDP packets authenticated only by the source IP address are subject to spoofing (Bishop & Heberlein, 1996). However, by verifying that the Time-to-Live (TTL) field is consistent with the true source the deception may be detectable (Templeton & Levitt, 2003). As a mimic, the goal is to move the indiscernibility boundary beyond that of the dupe, but not wastefully further. These additional signals provide no mimicry value and unnecessarily increase cost to the sender. As a defender, we wish to avoid deception by using a set of features sufficient to make it infeasible for the mimic to be successful at deception, such that their cost does not exceed expected loss from being deceived.

When the signals in an environment change, such as when human activity occurs, the new signals can confuse receivers. For example, the female Australian Jewel Beetle is appearing as large brown oblong. The much smaller males are preferentially attracted to the largest females. Interestingly, discarded beer bottles, which are brown and oblong, are misinterpreted by the males as enormous females which with the males aggressively try to mate (Gwynne & Rentz, 1983). The beetle employed a minimal set of signals, based on the original environment, to make decisions. This behaviour holds in cyber-environments as well. For example, in the context of adversarial machine learning, a trained traffic sign recognition system can, with correct placement of a few small black and white rectangles, be duped to interpreting a stop sign as a speed limit sign (Eykholt, et al., 2018). Just like the beetle, the trained recognition system used a minimal set of signals to make decisions. By understanding the signals used, and being able to produce those unexpected, an adversary can affect mimicry.

In nature signals are derived from physics and chemistry and are perceived relative to the capabilities of the receiver. Example signals include odour, emitted light, reflected light, moves and posture, body outlines, sound, temperature, and location. These are received chemically, optically, acoustically or tactically. However, "The 'universe' of cyber-security is an artificially constructed environment that is only weakly tied to the physical universe" (JASON, 2010). In cyber-environments signals derive from the unique constructs of the system. Actions that change the state of the physical world, e.g. changing the state of indicator lights, or the reported

temperature of a processor or hard drive, could be part of a deception; most cyber-signals have no direct analogues to those in nature. The location of a file, or its size are perhaps analogues of location or body outline. Attributes such as file name, file contents, hostname, MAC address, IP address, port number/protocol, TTL field, or user name, are just some of the available signals. With the exception of cyber-physical signals, signals are received as input to a monitoring process.

## **2. Related work**

### **2.1 Mimicry in biology**

One of the earliest formal works on mimicry systems (Poulton, 1898) describes relations between the signals presented by the mimic and those of the model, but focused only on the purpose of the mimicry and not a formal model of the relation between the entities involved. The formal concept of the tripartite relation between signal sender, receiver and dupe was first described in 1965 in (Wickler, 1965) and expanded in (Vane-Wright, 1972). These the basis for most models of mimicry systems used. In 1982, Pasteur published a seminal review of mimicry systems (Pasteur, 1982). This paper presented a multitude of specific named types of mimicry, but more importantly, it described models of mimicry including those of the previously mentioned authors.

A comparison of abstract models of mimicry is presented in (Starrett, 1993). These were classified based on the abstractness of the model in terms of semantic and cryptic resemblances. Starret used three types of resemblances to compare systems: *distinctive* (clearly appearing as the model), *attributive* (having some qualities that could appear like the model – e.g. having a similar silhouette), and *implied* (not looking like a model, but appear in some way that the dupe reacts – e.g. xenophobic reaction to unusual appearance or sounds).

In response to inconsistency of mimicry models, (Dalziell, 2016) presents a conceptual framework based on whether or not the receiver perceives the similarity between model and mimic, and in so doing benefits the mimic. The authors attempt to address the multiplicity of classification systems and terminology.

In Maran's book, *Mimicry and Meaning: Structure and Semiotics of Biological Mimicry* (Maran, 2017), the author revisits biological mimicry, providing a history of the mimicry concept, models of the structure of mimicry, and includes analysis and criticism of existing mimicry models. Typically, these models focus on one aspect of the mimicry system: formal structure, perceptual correspondence between participants, characteristics of resemblance, aspects of communication, or parallels to human cultural processes.

(Kloock & Getty, 2018) present a mathematical model of aggressive mimicry. Extending a model of foraging under predations, they argue that previous models of aggressive mimicry are more aligned with protective forms of mimicry. In particular, that the level of predation affects the behaviour of the prey. Their model considers pressure for changes in the quality of mimicry (e.g. resource use) based on changing conditions.

### **2.2 Bio-inspired computing**

Nature and biology have been used as the inspiration for many concepts in computer science. The progression from neural networks, genetic algorithms, ant colony optimization (Dorigo, 1992), swarm computing (Kennedy, 1995), and artificial immune systems (Forrest, et al., 1997), to more recently cuckoo search (Gandomi, et al., 2013) and artificial plant optimization (Cui & Cai, 2013). Typically, these are used as optimization algorithms as nature is known to optimize resource use (Johnson, 2013) and have been applied to areas such as network routing (Ducatelle, et al., 2010), clock synchronization (Leidenfrost & Elmenreich, 2009), content distribution (Sun, 2013), and security (Fink & Oehmen, 2012), (Forrest, et al., 1997), (Rauf, 2018) (Hassanien, et al., 2014), (Mazurczyk, et al., 2016). This work differs in that it does not define an algorithm, but the basis for a model for characterizing deception using biological models of mimicry.

### **2.3 Modelling cyber-deception**

Deception has been and will continue to be a significant tactic in military operations, both kinetic and cyber. "Cyber deception is a deliberate and controlled act to conceal our networks, create uncertainty and confusion against the adversary's efforts to establish situational awareness, and to influence and misdirect adversary perceptions and decision processes", (Climek, et al., 2016). Works such as (Whaley, 1982) look at deception in

general, taking a psychological view related to stage magic. An early project, The Deception Toolkit (Cohen & others, 1998) provided a collection of tools to create the appearance that the protected system had multiple known vulnerabilities. Steganography and Trojan Horse programs are examples of deception where information is hidden within another.

Formal studies of cyber-deception have resulted in taxonomies which classify attributes of examples of deception, but do not create models of deception. One taxonomy (Rowe & Rothstein, 2004) lists 8 general principles derived from military literature, 6 of which are effectively mimicry. Rowe & Rothstein presents a list of semantic cases that can be altered from the expected to create deception. These are similar to the signals we present for mimicry in cyber systems. (Rowe, 2004) presents a ranked collection of tactics that could be used to support offensive or defensive cyber-deception along with “excuses” to help convince the adversary that the system is truthful. The author notes that deception is generally thought of in only a few categories and that subtle possibilities exist, a sentiment we embrace. Concepts of what makes deception effective are presented in (Underbrink, 2016), while (Cybenko, et al., 2016) discusses method of assessing the covertness of deception operations. (Briskin, et al., 2016) discusses design considerations for building cyber-deception systems. None of these looks to an overall framework for modelling deception and the relations between entities in the system.

### **3. Classification of mimicry systems**

#### **3.1 Basic concepts – the tripartite relation**

Deception in the form of mimicry can be understood as a tripartite relation between the *signal senders* (model and mimic) and the receiver (dupe), whereby the mimic gains advantage from the dupe. In this system, the model, as Signal Sender 1, is designated  $S_1$ ; the mimic, as Signal Sender 2, is designated  $S_2$ ; and the dupe is designated by R (**Figure 1**). When the set of signals perceived by the Receiver are indiscernible from that of the Model, mimicry occurs. The perceived set includes both those emitted or blocked. This implies that mimicry occurs when the receiver experiences a perceived difference between the signals of model and the mimic. Then,

$$[P_R(\overline{\mathcal{S}}_{S_1})] \cong [P_R(\overline{\mathcal{S}}_{S_2})] \rightarrow \text{Mimicry},$$

where  $\overline{\mathcal{S}}_{S_1}, \overline{\mathcal{S}}_{S_2}$  are sets of signals sent by  $S_1$  and  $S_2$  respectively, with  $P_R()$  being the perception function of R.

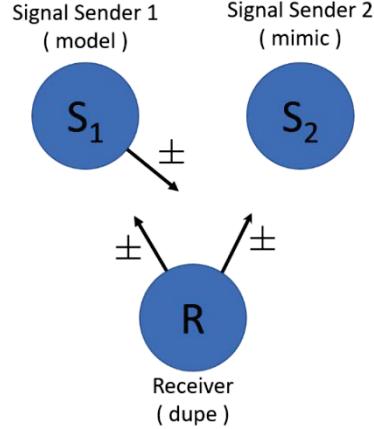
Vane-Wright (1976) expanded this basic model to consider the advantage or disadvantage to the signal-receiver in response to the model and mimic, and the advantage or disadvantage to the model as a result of the receiver's response (**Figure 3**). The normal response of the signal-receiver to the mimic is assumed to always be advantageous, otherwise, evolutionarily, the mimicry would not be preserved.

Noting that both the reactions between Model, Mimic, and Receiver, as well as the effect of actions of the three, characterize the mimicry system, we expand the Wickler/Vane-Wright model to explicitly include these relations. These are listed in Table 1.

*Reactions* characterize how one entity naturally responds to another; are they attracted, repulsed or neutral to the other. Small fish are attracted to worms which an anglerfish mimics, but would be repulsed by the anglerfish itself. Insectivores are attracted to *Kallima* “leaf-wing” butterflies, but are neutral to the leaves they mimic.

*Effects* characterize how one entity affects another, generally as a benefit, harm or net neutral result. Predation is harmful to the prey; warning coloration is protective to the mimic. Modalities such as camouflage are beneficial to the mimic in either protecting the mimic from predation or supporting its predation on others. Unlike relations, the act of mimicry can affect the other entities in the mimicry system apart from how the mimic (in the absence of mimicry) would. In the anglerfish example, the mimic causes harm to the dupe, and by eating the model's predator, benefits the model. More complex are the probabilistic effects of repeated interactions. For example, the more often a bird eats a noxious insect, the less likely the bird will take its mimic. Conversely, the more often the bird harmlessly eats a mimic, the more likely it will be to target the noxious species. Much the same, if a harmful creature were to mimic a benign model, repeated interaction would benefit the model through avoidance. These and other higher-order effects are similarly important in cyber-security. Banners declaring that the user is being monitored and will be punished for policy violations, yet never result in any action against the user will become ignored. Frequent encounters with phishing email may result in users

avoiding legitimate communications; consider how users and automated email scanning systems have changed with respect to legitimate email in the presence of phishing attacks.



**Figure 3:** From Vane-Wright (1976), members of the mimicry system gain or lose advantage from the other members. The mimic is assumed to always take advantage from the dupe. We extend this to include the negative effects the mimic causes the model

### 3.2 Relation vectors

Each instance of a relation will have a value. For simplicity we are using  $\{+, -, \odot\}$  representing, for Reactions, attractive, repulsive, and neutral respectively. With Effect relations, these indicate that the entity is benefited, harmed, or neutral (i.e., no net effect). We use  $| |$  to indicated the value of a relation. The equation  $|(R : S_1)| \rightarrow +$  indicates that the dupe is attracted to the Model.  $|(R \circ S_1)| \rightarrow \odot$  indicates that the model is neutral to the dupe; for example, a background substrate.

**Table 1:** List of entity relations in mimicry system

r#	Relation	Description	
1	$(R : S_1)$	Reactions	Reaction of Dupe to Model
2	$(S_1 : R)$		Reaction of Model to Dupe
3	$(R : S_2)$		Reaction of Dupe to Mimic (without mimicry)
4	$(S_2 : R)$		Reaction of Mimic to Dupe
5	$(S_1 : S_2)$		Reaction of Model to Mimic
6	$(S_2 : S_1)$		Reaction of Mimic to Model
7	$(R \circ S_1)$	Effects	Effect of actions of Dupe on Model
8	$(S_1 \circ R)$		Effect of actions of Model on Dupe
9	$(R \circ S_2)$		Effect of actions of Dupe on Mimic
10	$(S_2 \circ R)$		Effect of actions of Mimic on Dupe
11	$(S_1 \circ S_2)$		Effect of actions of Model on Mimic
12	$(S_2 \circ S_1)$		Effect of actions of Mimic on Model

Six additional relations characterizing how interactions within the mimicry system affect the system can also be considered (Table 2). Here we use  $S_2'$  to indicate the entity actively exhibiting mimicry.

The goal of mimicry is to alter the signals sent by the mimic changing the response function such that the  $(R : S_2')$  attraction (or avoidance) is enhanced. In our simplified model this equates to  $|(R : S_2')| > |(R : S_2)|$  where  $|(R : S_2')| \rightarrow \{+, \odot\}$  for attraction, and to  $|(R : S_2')| < |(R : S_2)|$  where  $|(R : S_2')| \rightarrow \{\odot, -\}$  for repulsion.

**Table 2:** Higher order effects

r#	Relation	Description	
13	$(R \circ S_2')$	Higher-order effects	How behaviour of Dupe affects behaviour of Mimic
14	$(S_2' \circ R)$		How mimicry affects behaviour of Dupe
15	$(S_1 \circ S_2')$		How behaviour of Model affects efficacy of Mimic
16	$(S_2' \circ S_1)$		How mimicry affects Model
17	$(S_2' \circ S_2)$		How mimicry affects efficacy of Mimic
18	$(S_2' \circ S_2)$		How mimicry affects Mimic (always positive, either for individual or population)

Excluding the higher-order effects, we can represent a mimicry system as a 12-element vector. Given 3 possible values for each of the 12 relations yields over  $\frac{1}{2}$  million potential mimicry systems. Many of these vectors map to well-known instances of mimicry in nature and cyber-deception; others may not be feasible, with the remainder suggesting possible new forms of mimicry. Correlation of vector elements will also occur.

Table 3 shows the relation vectors for 4 examples of mimicry from Nature; table 4 shows examples from security. Those from Nature show variation in the relation values, while those from security are nearly identical. While not exhaustive, this illustrates that commonly cyber-deception are variants of “honey-things”. Considering other regions of the relation vector space, we can generate unusual deception scenarios.

Some examples of cyber-deception with unusual relation vectors (Table 5) include:

**False Honeypots:** A real server behaves as if it were a honeypot causing attackers to avoid it. This would be a reversal of plausible “excuses” (Rowe, 2004) to prevent an attacker from believing the system was deceptive.

**Counterfeit DDoS Threat:** Received email demands payment or DDoS of organization will occur. The email claims to be from an organization known for large scale DDoS attacks. Here the model is a DDoS email threat from known menace. The mimic cannot cause a DDoS. The dupe is the organization who a DDoS would cause harm. This is an example of where the  $(S_1:S_2)$  relation increases awareness of a real threats, but repeated instances of the mimic without harm decrease believability.

**Table 3:** Example mimicry system vectors – biological examples

r#	Relation	Antennarius Anglerfish $S_1$ : worm $S_2$ : anglerfish R: fish that eats worm		Haplochromis Cichlids $S_1$ : eggs $S_2$ : male cichlid R: female cichlid		Batesian Mimicry $S_1$ : noxious butterfly $S_2$ : Harmless Butterfly R: insectivore		Camouflage $S_1$ : plant leaf $S_2$ : “leafwing” butterfly R: insectivore	
1	$(R : S_1)$	Reactions	Attracted to	+	Attracted to	+	Avoids	-	Ignores
2	$(S_1 : R)$		Oblivious to	-	Inanimate	⊕	Avoids	-	Inanimate
3	$(R : S_2)$		Avoids	-	Ignores	⊕	Attracted to	+	Attracted to
4	$(S_2 : R)$		Attracted to	+	Attracted to	+	Avoids	-	Avoids
5	$(S_1 : S_2)$		Neutral	⊕	Inanimate	⊕	Neutral	⊕	Inanimate
6	$(S_2 : S_1)$		Neutral	⊕	Neutral to	+	Neutral	⊕	Attracted/Protects
7	$(R \circ S_1)$	Effects	Preys on	-	Hatch	+	Eats/Kills	-	Neutral
8	$(S_1 \circ R)$		Neutral	⊕	Inanimate	⊕	Noxious	-	Neutral
9	$(R \circ S_2)$		Neutral	⊕	Improve Hatching	+	Eats/Kills	-	Eats/Kills
10	$(S_2 \circ R)$		Preys on	-	Improves	+	Food	+	Food
11	$(S_1 : S_2)$		Neutral	⊕	Inanimate	⊕	Dec. Interest	+	Inanimate
12	$(S_2 : S_1)$		Reduces Predation	+	Improves	+	Inc. Interest	-	Neutral

**Tricked into Updating:** Deception can be used to benefit the dupe. For example, rather than the frequent pop-ups nagging users to update their software or threatening email from management, users can be tricked into applying patches. Instead a staff member contacts the user, saying they can't update and wondered if patching works on the user's system.

**False Compromised Server:** A computer has a known vulnerability actively being exploited, but cannot be patched. Instead, the computer is altered to appear already compromised to the malware.

**Table 4:** Example mimicry system vectors – cyber examples

r#	Relation	Phishing $S_1$ : legitimate email $S_2$ : malicious email R: office worker		DoS Attack $S_1$ : legitimate packets $S_2$ : spoofed UDP packets R: server		Malicious Link $S_1$ : safe link $S_2$ : homographic link R: user		DTK $S_1$ : true vulnerabilities $S_2$ : false vulnerabilities R: attacker	
1	$(R : S_1)$	Reactions	Accepting	+	Accepting	+	Attracted to	+	Attracted to
2	$(S_1 : R)$		Targeting	+	Targeting	+	Neutral	⊕	Neutral
3	$(R : S_2)$		Avoiding/Rejecting	-	Avoiding/Rejecting	-	Avoid	-	Avoid

r#	Relation	Phishing <i>S<sub>1</sub>: legitimate email S<sub>2</sub>: malicious email R: office worker</i>		DoS Attack <i>S<sub>1</sub>: legitimate packets S<sub>2</sub>: spoofed UDP packets R: server</i>		Malicious Link <i>S<sub>1</sub>: safe link S<sub>2</sub>: homographic link R: user</i>		DTK <i>S<sub>1</sub>: true vulnerabilities S<sub>2</sub>: false vulnerabilities R: attacker</i>		
4	(S <sub>2</sub> : R)	Effects	Enticing	+	Enticing	+	Neutral	○	Neutral	○
5	(S <sub>1</sub> :		Inanimate	○	Inanimate	○	Inanimate	○	Inanimate	○
6	(S <sub>2</sub> :		Inanimate	○	Inanimate	○	Inanimate	○	Inanimate	○
7	(R o S <sub>1</sub> )		Reads/Clicks	+	Neutral	○	Exploits	-	Exploits	-
8	(S <sub>1</sub> o R)		Provides Info	+	Provides Data	+	Supports	+	Supports	+
9	(R o S <sub>2</sub> )		Activates	+	Accepts	+	Activates	○	Inanimate	○
10	(S <sub>2</sub> o R)		Exploits	-	Exploits	-	Exploits	-	Ensnakes	-
11	(S <sub>1</sub> :		Neutral	○	Neutral	○	Neutral	○	Neutral	○
12	(S <sub>2</sub> :		Reduces	-	Blocks	-	Neutral	○	Neutral	○

### 3.3 Signals in cyber-environments

In Nature signals have distinctive characteristics that convey information. Audible signals have characteristics such as pitch, volume, duration. Similarly, visible signals have hue, intensity and duration. As tactics in deception it may be possible to affect physical properties of equipment, however, cyber systems will typically use signals not based on physics and chemistry, but on the nature of information systems.

**Table 5:** Relation vectors of unusual instances of cyber-deception

r#	Relation	False Honeypot <i>S<sub>1</sub>: honey-server S<sub>2</sub>: real server R: attacker</i>		DDoS Threat <i>S<sub>1</sub>: DDoS attacker S<sub>2</sub>: incapable attacker R: business</i>		Update Trick <i>S<sub>1</sub>: needy staff S<sub>2</sub>: update service R: user</i>		Compromised Server <i>S<sub>1</sub>: compromised server S<sub>2</sub>: vulnerable server R: attacker</i>		
1	(R : S <sub>1</sub> )	Reactions	Avoids	-	Avoids	-	Positive	+	Avoids	-
2	(S <sub>1</sub> : R)		Neutral	○	Harms	-	Positive	+	Neutral	○
3	(R : S <sub>2</sub> )		Attracted to	+	Ignores	○	Ignores to avoids	-	Attracted to	+
4	(S <sub>2</sub> : R)		Avoids	-	Targets	+	Neutral	○	Avoids	-
5	(S <sub>1</sub> : S <sub>2</sub> )		Neutral	○	Neutral to hostile	-	Supportive	+	Neutral	○
6	(S <sub>2</sub> : S <sub>1</sub> )		Neutral	○	Neutral to fearful	-	Supportive	+	Neutral	○
7	(R o S <sub>1</sub> )		None	○	Pays Ransom	+	Helps	+	Avoids	-
8	(S <sub>1</sub> o R)		Wastes time/Tracks	-	Harms	-	Helps	+	Neutral	-
9	(R o S <sub>2</sub> )		Compromise	-	Neutral	○	Stifles	-	Compromises	-
10	(S <sub>2</sub> o R)	Effects	Gives information	+	Neutral	○	Patches	+	Is compromised	-
11	(S <sub>1</sub> o S <sub>2</sub> )		Neutral	○	Neutral	○	Supports	+	Neutral	○
12	(S <sub>2</sub> o S <sub>1</sub> )		Neutral	○	Neutral	○	Neutral	○	Neutral	○

The signals arising from the model and mimic may be generated by different aspects of the organisms. The method of creating the signals can be fundamentally different from the signals perceived by the dupe. For example, when the mimic produces analogous chemical signals, the dupe senses these chemically. However, when the mimic takes on a morphological form which is perceived by the dupe by reflected sound waves, the source is of a different nature than that received. In this way the dupe has an information processing function which translates the signals from the model and mimic, which may be created by different means, into those perceived by the dupe.

Perception of signals can elicit different types of information such as the size of an object. As an example, the distance, size and direction of motion of an insect are perceived from the acoustic echo of a bat. The equivalent

in cyber-environments occurs whenever data received is turned in to information. For example, when geo-location information is added to an IP address, or when assigned privileges for a username are obtained.

In cyber-environments signals will be messages from either internal or external sources. The characteristics of the particular message will distinguish it from others. For an IP packet, this would include source address, protocol, destination port, size, etc. The source (e.g. keyboard, USB drive, network) are other examples.

#### **4. Concluding remarks**

This work, inspired by long established ideas in evolutionary biology, lays the groundwork for a methodology for studying mimicry in cyber-deception. By examining the relations between the entities in a mimicry system we can group similar forms of deception even though they may not outwardly appear related. By looking at classified examples from Nature we can postulate analogous instances of cyber-deception, instances perhaps not yet realized. Additionally, examining instances of the mimicry relation space not tied to any examples, we may discover forms of deception not found in Nature, but could exist in cyber-environments.

Mimicry occurs when the mimic's signals, as perceived by the dupe, are indiscernible from those of the model; therefore, understanding the nature of cyber-signals is important to understanding deception. Given the richness of deception as mimicry found in the natural world, it can be a useful guide in developing and detecting new deception tactics, for offense or defence.

The field of adversarial machine learning may provide many options for studying cyber-mimicry. Because a trained system has learned to interpret a set of signals, by determining what signals are being used in the decision process, an adversary can cause one thing to mimic another. In the sign example, this is presumably to cause harm. However, it could just as easily be used to hide objects (e.g. armoured tanks), make objects seem uninteresting or attractive. This could be done in physical-space by creating situations that, while innocuous, would appear worthy of focussing attention and directing resources away from sites in truth more threatening.

Many other aspects of mimicry were not discussed in this work, but are worth continuing to develop. In addition to expanding work on types of signals, other relations between entities in the model need be considered: extrinsic vs. intrinsic mimicry; the types of "species" involved (all the same, all different, a mix); automimicry, where the mimic mimics part of itself; how signals compose; and persistence of signals. The cost of sending or receiving (perceiving) particular signals and how acts of deception affect the deception system.

#### **References**

- Bishop, M. & Heberlein, L. T., 1996. Attack class: Address spoofing. s.l., s.n., pp. 371-377.
- Briskin, G. et al., 2016. Design considerations for building cyber deception systems. In: Cyber Deception. s.l.:Springer, pp. 69-95.
- Climek, D., Macera, A. & Tirenin, W., 2016. Cyber Deception. Journal of Cyber Security and Information Systems, 3.VOLUME 4.
- Cohen, F. & others, 1998. The deception toolkit. Risks Digest, Volume 19.
- Cui, Z. & Cai, X., 2013. Artificial plant optimization algorithm. In: Swarm Intelligence and Bio-Inspired Computation. s.l.:Elsevier, pp. 351-365.
- Cybenko, G., Stocco, G. & Sweeney, P., 2016. Quantifying Covertness in Deceptive Cyber Operations. In: Cyber Deception. s.l.:Springer, pp. 51-67.
- Dalziell, A. a. W. J., 2016. Mimicry for all modalities. Ecology letters, 16(6), pp. 609-619.
- Dorigo, M., 1992. Optimization, learning and natural algorithms, s.l.: s.n.
- Ducatelle, F., Di Caro, G. A. & Gambardella, L. M., 2010. Principles and applications of swarm intelligence for adaptive routing in telecommunications networks. Swarm Intelligence, Volume 4, pp. 173-198.
- Eykholz, K. et al., 2018. Robust physical-world attacks on deep learning visual classification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1625-1634.
- Fink, G. A. & Oehmen, C. S., 2012. Final Report for Bio-Inspired Approaches to Moving-Target Defense Strategies.
- Forrest, S., Hofmeyr, S. A. & Somayaji, A., 1997. Computer immunology. Communications of the ACM, Volume 40, pp. 88-96.
- Gandomi, A. H., Yang, X.-S. & Alavi, A. H., 2013. Cuckoo search algorithm: a metaheuristic approach to solve structural optimization problems. Engineering with computers, Volume 29, pp. 17-35.
- Garrouste, R. et al., 2016. Insect mimicry of plants dates back to the Permian. Nature communications, Volume 7, p. 13735.
- Giffin, J. T., Jha, S. & Miller, B. P., 2006. Automated discovery of mimicry attacks. s.l., s.n., pp. 41-60.
- Gwynne, D. & Rentz, D., 1983. Beetles on the bottle: male buprestids mistake stubbies for females (Coleoptera). Australian Journal of Entomology, 22(1), pp. 79-80.

- Hassanien, A. E., Kim, T.-H., Kacprzyk, J. & Awad, A. I., 2014. Bio-inspiring Cyber Security and Cloud Services: Trends and Innovations. s.l.:Springer.
- JASON, M. I. T. R. E., 2010. Science of cyber-security, s.l.: MITRE Corp..
- Johnson, A. T., 2013. Teaching the principle of biological optimization. *Journal of biological engineering*, Volume 7, pp. 1-7.
- Kennedy, R., 1995. J. and Eberhart, Particle swarm optimization. s.l., s.n.
- Kloock, C. & Getty, T., 2018. A mathematical model of aggressive mimicry. *Behavioral Ecology*.
- Leidenfrost, R. & Elmenreich, W., 2009. Firefly clock synchronization in an 802.15. 4 wireless network. *EURASIP Journal on Embedded Systems*, Volume 2009, p. 7.
- Maran, T., 2017. Mimicry and meaning: Structure and semiotics of biological mimicry (Vol. 16). Berlin: Springer.
- Mazurczyk, W. et al., 2016. Bio-inspired cyber security for communications and networking. *IEEE Communications Magazine*, 6, Volume 54, pp. 58-59.
- Pasteur, G., 1982. A classificatory review of mimicry systems. *Annual Review of Ecology and Systematics*, Volume 13, pp. 169-199.
- Peck, A. L., 1965. Aristotle's Historia Animalium IX. s.l.:Harvard University Press, Cambridge, Mass..
- Rauf, U., 2018. A Taxonomy of Bio-Inspired Cyber Security Approaches: Existing Techniques and Future Directions. *Arabian Journal for Science and Engineering*, pp. 1-16.
- Rowe, N. C., 2004. A model of deception during cyber-attacks on information systems.
- Rowe, N. C. & Rothstein, H. S., 2004. Two taxonomies of deception for attacks on information systems.
- Starrett, A., 1993. Adaptive resemblance: a unifying concept for mimicry and crypsis. *Biological Journal of the Linnean Society*, Volume 48, pp. 299-317.
- Sun, L., 2013. Epidemic Content Distribution in Mobile Networks, s.l.: s.n.
- Templeton, S. J. & Levitt, K. E., 2003. Detecting Spoofed Packets. s.l., IEEE, pp. 164-175.
- Underbrink, A. J., 2016. Effective cyber deception. In: *Cyber Deception*. s.l.:Springer, pp. 115-147.
- Vane-Wright, R. I., 1976. A unified classification of mimetic resemblances. *Biological Journal of the Linnean Society*, Volume 8, pp. 25-56.
- Whaley, B., 1982. Toward a general theory of deception. *The Journal of Strategic Studies*, Volume 5, pp. 178-192.
- Wickler, W., 1965. Mimicry and the evolution of animal communication. *Nature*, Volume 208, p. 519.

# Agile Technology Development to Improve Scenario-Based Learning Exercises

**Benjamin Turnbull, David Ormrod, Nour Moustafa and Nicholas Micallef**

**University of New South Wales, Australia**

[benjamin.turnbull@unsw.edu.au](mailto:benjamin.turnbull@unsw.edu.au)

[drdave@linux.com](mailto:drdave@linux.com)

[Nour.Moustafa@unsw.edu.au](mailto:Nour.Moustafa@unsw.edu.au)

[Nicholas.Micallef@unsw.edu.au](mailto:Nicholas.Micallef@unsw.edu.au)

**Abstract:** Technology is increasingly playing a part in defence and military decision-making, and this must be reflected in how defence, government and military train with such systems. Even beyond the traditional mechanisms that are considered cyber, the interconnectivity we take for granted has utility in real conflict, and therefore must be accounted for in training scenarios. Scenario-based learning is an effective teaching mechanism, allowing for both in-depth knowledge transfer and experimentation. Scenarios allow a level of immersion not present in other forms of learning and can expand on what is currently possible to provide greater insight into theoretical situations. However, designing and integrating technology into scenario-based exercises is difficult and often results in failure, from the perspective of the relevance of the learning outcomes to real life. This work has two distinct aims; to develop a framework that adapts and expands upon known software development project management paradigms to enhance both the quality and success rate of technology integration into scenario-based training, and to implement this for a scenario. This paper outlines the processes, challenges and mechanisms specifically faced when developing and integrating new technology for scenario-based learning opportunities, then discusses a specific implementation and the lessons learned.

---

**Keywords:** wargame, social-media, online social network, cyber-security, software-engineering, scenario-based-learning

---

## 1. Introduction

Technology is increasingly playing a part in defence and military decision-making, and this must be reflected in how defence, government and military train with such systems. Even beyond the traditional mechanisms that are considered cyber, the interconnectivity we take for granted has utility in real conflict, and therefore must be accounted for in training scenarios. Access to new information sources has the advantage of providing additional intelligence, can provide data that would otherwise be unavailable, and is near real-time. It also has disadvantages; it is noisy and unrepresentative. Balancing the rapid pace of the twenty-four-hour news cycle and changes in the operating environment with immediate decision-making process provides a challenging aspect for exercises at all scales and timeframes.

Implementing technology into scenario-driven environments is still in its infancy; there is no shortage of technologies used in the real world that would be of potential use if implemented for exercises. There are also additional considerations; data sources, legal requirements, connectivity, the balance between realism and complexity, and specific programming requirements particular to an individual exercise or series.

This work outlines a framework to assist technology developers work with scenario developers to add digital technology into scenario-based learning and experiment activities. It focuses on maintaining flexibility and minimising risk throughout this process. The framework has also been tested, and this work also outlines the test process. This work specifically outlines the rapid development of two simulated technologies for use in military and defence simulations and exercises; an intelligent bot-driven social network, and multiple Internet news platforms. A tailored approach has multiple benefits; it allows for organisers to approach the exercise from its potential learning outcomes, provides greater context specific to each training opportunity, and also adds the potential for immersion.

## 2. Outlining the need

Scenario-based learning is an effective learning mechanism that draws upon experiential learning and learning-by-doing theories (Dewey, 1938; Kolb, 2014). There are several concepts that underpin these; individuals retain more information if they are active participants rather than passive observers, make decisions and see results, and makes for more motivated learners. Scenario-based learning and training exercises have been effectively used across many fields, as a means of training and education reinforcement; from nursing (Cant & Cooper, 2010; Weller, 2004) to military training (Gritzalis & Papageorgiou, 2016; Macedonia, 2002) to incident response

(Perry, 2004) through to astronaut training (Loftin et al., 1994; Voas, 1961). Fields that have been successful for simulation-based learning all have several things in common; they apply to practical, hands-on areas, they all refer to their own programmes as either training or reinforcement learning, and games and simulation are primarily used in situations where it is not practical to use a real environment. There are several reasons for this; technological limitations, policy reasons, and in many cases, to practice and understand before using equipment for the first time. Scenarios are used to develop or refine procedure or policy, understand new workflow or systems or to learn and refine participant skill. Scenario-based learning is also used as a form of experiential learning, to facilitate greater understanding and development. Within defence and the military, the term ‘exercise’ has a specific meaning, but the fundamental concepts are able to be shared.

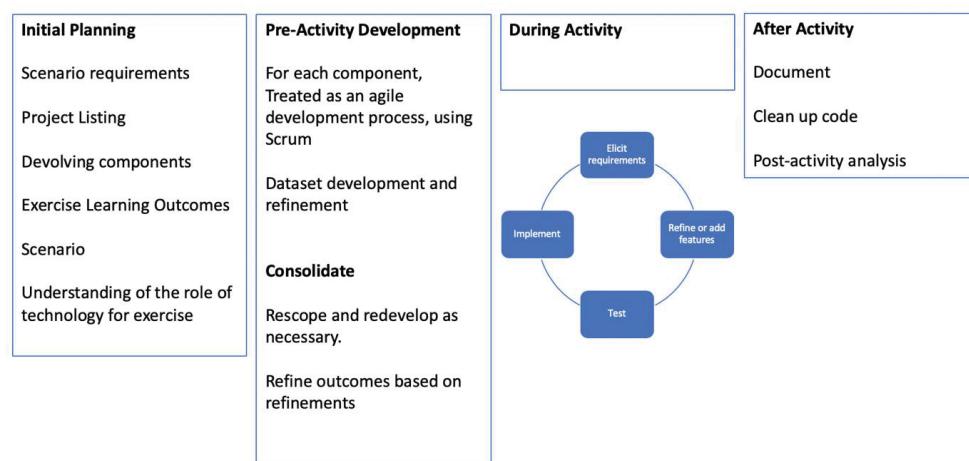
There is an increasing movement to integrate additional technology into exercises; to add realism, immersion and to assist in automating physical processes and components. The lessons learnt and skills obtained by exercise participants should be representative and helpful in the real world, otherwise the training value of the activity is diminished. Therefore, technologies used in the real world to generate outcomes should be available or at least represented in some way in exercises, so that the lessons learnt and skills practiced are representative. However, there is still a dearth of software suited to these tasks outside of specific weapon systems. More generalised digital technologies available outside of exercises are often not provided because of security concerns or difficulties in tailoring them to the wargame itself. Wargame exercises and scenario-based learning environments are different from the majority of software systems created for industry, and even from real military systems. Technology integrated into scenario-driven exercises are fundamentally different; they are highly experimental, they are to be operated within a fixed development time with no possibility of extension, and they are often changing throughout to fit scenario needs. Specifically, military exercises are organised and planned in advance, but the details and scenario are increasingly refined up until live execution. Any and all software developed for, or run, will need to therefore be operational both within this timeframe, but with the understanding of changes throughout this process. This type of work rules out the use of many software engineering methodologies, including waterfall and iterative approaches. The dates of exercises are set and unable to be changed, so development is heavily compressed. This approach does not work well with many agile approaches to software creation, including many agile approaches.

Therefore, with both the expected increased use of technology development in military exercises and a lack of suitable approaches designed specifically to develop software in this space, a different approach is necessary. This work seeks to augment existing, established processes to account for the uniqueness of military exercises.

### 3. Approach

#### 3.1 Framework outline

This work used a tabletop exercise and brainstorming session to develop a framework to assist in the development of software for an exercise. It then implemented this framework, developing multiple projects for an exercise. The results allowed a process of refinement. The completed Agile Scenario Technology framework is presented in Figure 1.



**Figure 1:** The agile scenario technology framework

The framework itself consists of four separate components; each with separate aims and outcomes. These phases are the *initial pre-activity*, *pre-activity development*, *during activity* and *post-activity*. Each will be discussed separately.

### **3.2 Initial pre-activity phase**

By far the largest and most complex component of the framework occurs in the **initial pre-activity phase**. There are two aims to this phase; to work with the scenario developers to understand the needs actors, flow and pacing of the scenario itself, and to devolve this into a series of requirements for one or more IT-based projects to develop or source.

The first aim, to work with the scenario developers, seeks to develop a comprehensive understanding of the needs, the role of technology within the scenario and the scenario itself. What is the learning objective or experimental aim of this scenario? What is the scenario itself? Who are the actors, and who are participants? Has the scenario been planned before? What role is technology envisioned to have in this work? How do these link to the scenario objectives?

Whilst understanding the objectives and learning outcomes of a scenario may not seem immediately vital, working with changing environments, timeframes and details can sometimes mean that the scenario organisers may not be immediately available. In some situations, critical stakeholders may not arrive until immediately prior to, or shortly after, an exercise or wargame has commenced. As a result decisions may need to be made by the developers. Understanding the aims will allow for more comprehensive and better decisions to be made.

The second aim of the initial pre-activity phase is to take the high-level needs from the scenario organisers and to decompose these into manageable requirements. The scenario needs may involve one or more of the following:

- The development of one or more independent systems,
- The development of one or more interconnected systems,
- Analysis of the best existing system to use to achieve an end,
- Configuration of existing system or platform,
- Customisation of existing systems to suit scenario needs,
- Integration system to connect disparate software or systems, or
- An outline of the data requirements appropriate to the exercise or scenario.

One of the central considerations of this framework is risk minimisation. When considering projects, redundancy is a potentially important aspect. However, staffing constraints are real, and appropriately apportioning teams is difficult. The options are to run similar, or the same project across multiple teams, or to build a single team that has a greater chance of completing the system.

When partitioning requirements into separate projects, there is the potential that common interfaces will be required to provide interactivity. It is the authors assertion that the best, most comprehensive and efficient way of producing interfaces is to trust in the different project groups and let them develop organically. Developing use-cases, mandating interface design before development can be overly prescriptive and does not allow for changes that are likely to occur during development.

Choice of programming language is another important aspect that must be considered before the project can commence. It is not required, nor even expected that all projects are written in the same programming language. Common interfaces and APIs using standards negate this need. Language choice is context dependent and may be determined from the need to integrate existing systems or libraries.

### **3.3 Pre-Activity development phase**

The second phase of the Agile Scenario Technology framework is **Pre-activity Development**. There are two primary aims of the pre-activity development phase; to develop and configure the software projects themselves, and to periodically refine the requirements.

For each individual project, the determinations made in the previous phase dictate the development process. Depending on the number of projects, this may be partitioned into multiple teams, each working on a smaller number of projects. The authors recommend a modified scrum approach, given the flexibility and adaptability the process requires. Scrum also has the benefit of focusing on working product over documentation, agility in aims and greater flexibility than other software development paradigms. Scrum is designed for small teams and rapid development. The acknowledged disadvantages for scrum, specifically, that is does enforce the development of maintainable, sustainable, robust code, and can lead to issues in documentation over long periods, is less applicable in a fast-moving, experimental setting.

The development, generation or sanitisation of datasets is also within this phase of the framework. It is both related to, and independent of software development, and relates to both testing for correctness, but more importantly for ensuring the resulting platform is fit for purpose. Data is vital to the success of any simulated scenario or exercise. Robust data is a necessary consideration. As with software, data may be acquired from existing streams, taken from public or paid datasets or programmatically generated. Each has advantages and disadvantages. Existing datasets and live streams from real systems are often easy to collect and can be used for comparison purposes. However, depending on the context for use, they may only be partially suitable. Depending on the data source, method of collection, use-case and jurisdiction, there may be additional legal or licence issues with using existing datasets. These are context dependent but must be considered. Programmatically generating datasets allows for a greater level of control than using existing data but comes at a higher complexity to create and validate. Data generation is an art in of itself and much of the requirements are beyond the scope of this work. There are multiple methods, but the creation of interesting, varied, believable and contextually appropriate datasets is a challenging research area.

The refinement process occurs at predefined stages leading up to the scenario or exercise. At multiple points, decisions may be needed to refine or reduce development scope. Software engineering is difficult to estimate (Leung & Fan, 2002), and there may be unexpected difficulties in the development, installation or configuration that do not become apparent until later in the development process. This true even for project implementing or configuring existing software systems; despite due diligence, sometimes the selected system is not fit for the purpose for which it was intended.

The fixed deadline of an upcoming scenario or exercise necessitates changes from accepted software engineering approaches; the delivery timetable is non-negotiable and therefore the only remaining variable is what can be delivered. The periodic synchronisation between different development teams and scenario organisers provides feedback on which aspects of development are behind, on, and ahead of schedule. This gives organisers opportunities to adjust expectations, systems and requirements approaching the scenario.

### **3.4 During activity phase**

During the **activity phase** of the Agile Scenario Technology framework, the main aims are twofold; to adapt software, data and platforms as necessary, and to solicit feedback and participant opinions. The latter serves as a basis for the post-activity phase. Depending on the design of both the scenario and the software, there also may be periodic tasks and events to be triggered. This may involve alterations to mechanics, the release of new information or other aspects in response to participant gameplay. The purpose of this phase is not to evaluate the activity itself, but the software and how it assists. This consists of multiple aspects; how fit-for-purpose the software is, how it integrates into the exercise, its ease of use, and any other issues that present.

As with the nature of experimental scenarios, exercises or events driven by participants, the initial deployment of software is unlikely to be complete within itself. Fixes, alterations and adjustments may be necessary to keep the scenario moving, fix bugs or to improve the learning outcomes. Often complexities arise relating to scale or differences in deployment environment may need to be fixed while the scenario is operational.

Changes may therefore be necessary during or between phases of the scenario. These might be pre-planned (changes to the environment that are not automated, based on time or in response to participant intervention) or in response to necessary changes in the scenario itself. The latter may be in response to unexpected events, or to manage difficulty. For example, if participants are finding the scenario too difficult, additional resources may be opened up, removed, or altered to balance scenario. This is particularly relevant for ‘Capture the Flag’ or team vs team competitive scenarios, where unbalanced teams reduce the learning opportunities. Depending

on the nature of the scenario, the software developed, the organisers and the participants, the pace during the activity phase may be either frenetic or slow. Some aspects may be known in advance, but the context is important.

There are multiple methods of evaluating the effectiveness of software, how fit-for-purpose it is, and the value it provides (Karat, 1997). This includes passive processes such as observation, and active measures including interviews and surveys. Another method is to instrument the software itself, collecting attributes of how the software is used. These metrics include active engagement time, mouse clicks, and even video eye-tracking (Goldberg & Kotval, 1999). In the latter of these, there is an additional overhead in both instrument design and in analysis, but the data is detailed and provides definitive outcomes.

### **3.5 Post-Activity phase**

The final phase of the Agile Scenario Technology framework is the post-activity phase. This phase has two major aims, to understand the positive and negative feedback to capture a 360-degree understanding of the systems developed and how they related to the learning outcomes, and to package and refine code for delivery or publication.

Working with the scenario developers, participants and from their own observations, a post-activity analysis is created. This provides feedback for future improvement, in addition to discovering and acknowledging the successful aspects of the technologies deployed.

It is expected that code developed in a compressed timeframe, especially if incorporating developmental changes made during a live exercise, will be imperfect. Providing opportunity to ensure high code quality and documentation is invaluable. Much of this can be done after the exercise has completed. The result is a software environment or code that can be deployed at future exercises, maintained by external groups or otherwise distributed.

## **4. Framework implementation and exercise**

As previously outlined, this project was developed specifically for a military scenario-based exercise.

The aims and objectives of this exercise are beyond the scope of this work, but there was a strong interest in incorporating technology into this work. The scenario was experimental in nature, lasted for four days, with approximately fifty participants. The addition of technology to this work was one of the fundamental principles underlying the aims and processes, and therefore, the scenario was developed to incorporate both an interpersonal and virtual component. The following section follows the Agile Scenario Technology framework.

### **4.1 Initial pre-activity**

Three components were outlined, in addition to a common component. These are; an extensible news-system to replicate news websites, a social network simulation platform utilising intelligent agents, and a common datastore backend. Separate, but integrated to this, was a data generation utility, designed to populate the graph datastore.

Both components, the news and the agents, were developed in the Python 3.6 programming language (Python.org, 2019). There were multiple reasons for this; it was the programming language all authors had in common, it has a number of existing libraries to other external components and has an active community. The disadvantages of Python, specifically its performance speed, were considered but ultimately deemed acceptable.

### **4.2 Pre-activity development**

The social network team was comprised of three people and was developed in the style outlined in the previous section of this work. It was self-described as a modified Scrum approach. The team utilised Git (Spinellis, 2012), Slack (Slack Inc, 2019), Atlassian Jira and Confluence (Zalan, Muzychenco, & Burshtein, 2009). This worked well for the most part, although the enterprise-grade software was often overkill for the small number of people within the project.

There were multiple technologies used to create the social networks. The first note of interest is that it was a prerequisite that it would be necessary to run this work entirely separated from the Internet, and therefore self-hosted social networks were required. As such, multiple open-source self-hosted platforms were used, and Elgg (Sharma, 2008) and Mastodon (Göndör & Küpper, 2017) were selected.

Mastodon is an open-source Twitter-like social network, designed to facilitate friends, connections, follows, likes, and short messages. Mastodon is well-built, tested and used in multiple real-world scenarios. Mastodon also has an API for developers to access some functions within the system. This had significant advantages and also disadvantages; the simple API had access to several features and could be used for the majority of use-cases. The disadvantages are within the limited features of the API and the need for per-user connectivity.

The Python Mastodon API exposes aspects of Mastodon that a standard chatbot would be expected to use; authentication, messaging, and following. However, this work required more advanced methods such as adding and modifying users. The Python Mastodon API also requires one connection and authentication per user. Again, this is well within the expected use-case for this API. The response was two-fold; to utilize the existing API where possible, and to write directly to the database for operations not considered by the standard API. Testing indicated that approximately 30,000 concurrent user connections was possible with the Python API.

Elgg is an open source social media platform, often used on Intranets. It has multiple options, skins and forms of customization, allowing it to replicate existing platforms such as Facebook, short messaging platforms akin to Twitter, and blogs. However, it has no native Python API, and the community is not as active as it has been previously, increasing the difficulty to build one. The development response was the same as Elgg, to reverse-engineer the Elgg database schema and to develop modules that directly write to the database as necessary. For Elgg, this proved non-trivial, given the modular nature of the platform.

The existing social networks became the front-end to the developed system. The backend was written as a series of Python intelligent agents, initialised from the datastore. Each agent represented a single user, and queried the datastore for current ‘state’, including information such as the contact information for friends, family, colleagues and neighbours. This provided a baseline for communications, friends, follows and interactivity.

One of the primary purposes of this was to create lifelike social networks, with simulated users having conversations. However, in conversation, it was noted by the scenario organisers, that data would need to be anonymised or generated, and that live data was not permissible without filtering and alteration. The alternative to utilising existing datasets verbatim was to generate new ones. This is not without its own issues; development of datasets in bulk is a time-consuming and difficult process. It also risks not appearing realistic, as it has been generated by one or more people, rather than by the large groups that organically generate the real content. It was therefore decided to generate the data for conversations from existing datasets as a training mechanism for a machine learning approach. Twitter datasets were used, and a supervised machine learning algorithm was then able to take a smaller training sample of conversations and use these as the basis to algorithmically generate new conversations based on the examples provided.

This work specifically used Recurrent Neural Networks with Long Short-Term Memory as the machine learning algorithm to generate conversations between agents (Zaremba, Sutskever, & Vinyals, 2014). Recurrent Neural Networks are an emerging machine learning technique with excellent results in natural language processing (Cheng, Dong, & Lapata, 2016).

The development of the social network simulation platform had several minor issues, mostly related to the number of interconnected components. The use of existing social media platforms necessitated additional software infrastructure in the form of virtual machines and systems, the graph datastore had network issues when operated at scale, and the data generation was a prerequisite for other system components. However, for the development of a small team, it was an overall successfully deployed product.

#### News Platforms

The news platform was a project operated by two team members. They used the same methodology as the social media network development team. The project itself was written in a Python web framework, Django.

The system comprised of two components; a user front-end that was the ‘online news site’ and a back-end administrative area. The latter was used by the scenario media team to add new news stories as appropriate.

The major issue faced with the news platform was not discovered during the development phase and was a product of the change in deployment environment. It is discussed in the next phase. This project developed a first iteration, which lacked automated and scheduling abilities. This was functional, although there are several areas for future improvement.

### **4.3 During activity**

As outlined, the aims of the activity phase are to refine, develop and to elicit feedback. In this case, there were several areas of development that were able to be refined during the scenario itself. This allowed the systems to progressively improve.

Due to time limitations, this work was based on observation. Future iterations of this work will involve semi-structured interviews or other forms of assessment. Participant interaction, in the form of interviews or surveys, require human ethics approval. The specifics of this vary by participant type, local policy and objectives. Future work will be to organise these aspects in advance, providing a more comprehensive approach.

One issue that was only discovered immediately when implemented on the target, disconnected network was the embedded use of Javascript frameworks. Upon installation to a network disconnected from the Internet, there were issues with both systems in rendering fonts and high load times. This was traced back to a common issue, the use of Content Delivery Networks (CDNs) to deliver fonts. These required external internet access, and browsers would time out waiting for responses before continuing to load pages. This was solved at a network level, by routing the [www.googleapis.com](http://www.googleapis.com) domain to a system that failed quickly, allowing rendering to continue. Wherever possible, this was then coded to use local fonts and the dependency removed. Not all third-party systems allowed this however. The ability to work in an offline environment would be a consideration when choosing future frameworks.

To connect with the scenario being operated during the learning exercise, the news system was replicated five times. This allowed different views of the same events, and different participants to view content relevant to them. The news systems were controlled by a small number of scenario organisers dubbed ‘the media team’. The media team were responsible for adding and curating content in time with events within the scenario.

It was decided, with the news systems active, that news media would become one of the primary mechanisms for conveying information, and therefore were used at a much higher rate than expected. This did not pose an issue, although it was noted in system analysis that the systems were working at capacity. With the exception of the network issue outlined above, the news systems only received minor tweaks and cosmetic changes after installation to allow for visual differentiation for participants. Branding, colours and layout designs were modified to improve readability and distinguishability.

The Social media platform had two major functions during the exercise; to provide background discussion as a mechanism for enhancing participant immersion, and to highlight specific users and topics. The first component of this was automated and based on the machine-learning algorithms outlined in the previous section. The latter of these was a combination of manually controlled users, based on the media team requirements, and code developed during the exercise, to provide the impression of large followers, message propagation and interest. This allowed a small number of active ‘users’ to appear active in real-time.

Future iterations would connect these systems, to allow greater automation. The scripts generated during the week were in response to discussion on how the system could be extended. This would have increased productivity immediately but was not possible in the timeframe outlined.

### **4.4 Post-Activity**

The aim of the post-activity phase was to clean up the software for future use, to document it correctly, and to elicit feedback from participants, developers, users and the exercise organisers on both the positive and negative aspects of the developments.

Controversially, one of the major outcomes discussed amongst the social media platform teams was that the next version would be a partial rewrite of the existing codebase. Although the initial architecture and design choices were appropriate at the time, and worked for the duration of the exercise, there were several areas that would be unable to be expanded upon with the current system. To pay this technical debt, it has been decided to rewrite significant portions of the work to increase flexibility and scalability. The reasoning for this was that the first version of this work provided an understanding of the needs and potential performance bottlenecks that could then be architected more cohesively in future versions.

Some of the scripts and software artefacts produced during the exercise week were rushed and contained examples of poor and insecure coding technique. Hard-coded passwords, large sections of copied and pasted code, and no documentation or commenting. At the conclusion of the activity, these aspects of the code were updated and augmented.

## **5. Conclusion**

This work has outlined and implemented the Agile Scenario Technology framework, designed specifically to assist software development for exercises and scenario-driven learning environments. Exercises are used heavily in the military and defence, but not exclusively, and there is opportunity to expand the role of technology in this space. Although current software engineering methodologies are well-designed, the context needed in this space necessitates an adapted approach.

As outlined, one of the highest risks in developing technology to implement in scenario-driven learning environments is the risk of incomplete or inadequate components. The Agile Scenario Technology framework seeks to minimise this risk to the extent possible. First, by using COTS, MOTS and open source systems wherever possible. This can also be seen as a pragmatic means of avoiding ‘not invented here’ syndrome (NIH) (Garlan & Perry, 1994). NIH is prevalent in research organisations, but has the capacity to slow development, at least initially. However, this does increase the focused need on integration from multiple vendors or developers, many of whom will not be present. The use of existing systems, platforms, technologies and libraries cannot be underestimated; the wealth of technology available allows experienced integrators to achieve amazing things, fast.

Scrum is the most popular of agile software development methodologies, but not the only one. This work adapted scrum as it was the most applicable to this work, but there is no single correct approach to developing software. The flexibility provided by this approach allowed significant gains in compressed timeframes to achieve an end. There are improvements that can be made, and these are outlined. An agile approach for developing and integrating technology into a wargame exercise allowed us to pivot quickly, and to learn and test additions as the exercise was run. However, the unique limitations of an exercise meant that the alterations made were crucial. Context is important in software development, and the context for developing a system to be used once on a short time-frame is different from what most software methodologies are aiming for; a commercial, public launch. This work was a success, and there is now a roadmap to refine and continue work in this area.

## **References**

- Cant, R. P., & Cooper, S. J. (2010). Simulation-based learning in nurse education: systematic review. *Journal of advanced nursing*, 66(1), 3-15.
- Cheng, J., Dong, L., & Lapata, M. (2016). Long short-term memory-networks for machine reading. *arXiv preprint arXiv:1607.0733*.
- Dewey, J. (1938). Experiential education. New York: Collier.
- Garlan, D., & Perry, D. (1994). *Software architecture: practice, potential, and pitfalls*. Paper presented at the Proceedings of the 16th international conference on Software engineering.
- Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics*, 24(6), 631-645.
- Göndör, S., & Küpper, A. (2017). *The Current State of Interoperability in Decentralized Online Social Networking Services*. Paper presented at the 2017 International Conference on Computational Science and Computational Intelligence (CSCI).
- Gritzalis, D., & Papageorgiou, S. (2016). PANOPTES: The Greek National Cyber Defence Exercise. Retrieved from <https://www.infosec.aueb.gr/Publications/CEER-ENISA-2016%20Gritzalis%20Papageorgiou.pdf>
- Karat, J. (1997). User-centered software evaluation methodologies. In *Handbook of human-computer interaction* (pp. 689-704): Elsevier.
- Kolb, D. A. (2014). *Experiential learning: Experience as the source of learning and development*: FT press.

- Leung, H., & Fan, Z. (2002). Software cost estimation. In *Handbook of Software Engineering and Knowledge Engineering: Volume II: Emerging Technologies* (pp. 307-324): World Scientific.
- Loftin, R. B., Kenney, P. J., Benedetti, R., Culbert, C., Engelberg, M., Jones, R., . . . Nguyen, L. (1994). *Virtual environments in training: NASA's Hubble space telescope mission*. Paper presented at the Interservice/Industry Training Systems & Education Conference.
- Macedonia, M. (2002). *Games, simulation, and the military education dilemma*. Paper presented at the Internet and the University: 2001 Forum.
- Perry, R. W. (2004). Disaster exercise outcomes for professional emergency personnel and citizen volunteers. *Journal of Contingencies and Crisis Management*, 12(2), 64-75.
- Python.org. (2019). Python.org. Retrieved from [www.python.org](http://www.python.org)
- Sharma, M. (2008). *Elgg social networking*: Packt Publishing Ltd.
- Slack Inc. (2019). Slack, About Us. Retrieved from <https://slack.com/about>
- Spinellis, D. (2012). Git. *IEEE software*(3), 100-101.
- Voas, R. B. (1961). Project Mercury Astronaut Training Program. *Supra ref*, 4, 96.
- Weller, J. M. (2004). Simulation in undergraduate medical education: bridging the gap between theory and practice. *Medical education*, 38(1), 32-38.
- Zalan, T., Muzychko, O., & Burshtein, S. (2009). *Atlassian: Supporting the world with legendary service*: NeilsonJournals Publishing.
- Zaremba, W., Sutskever, I., & Vinyals, O. (2014). Recurrent neural network regularization. *arXiv preprint arXiv:06733*.

# The Possibilities of Cyber Methods as Part of Maritime Warfare: Baltic Sea

Maija Turunen

Finnish National Defence University, Helsinki, Finland

[maija.turunen@ftia.fi](mailto:maija.turunen@ftia.fi)

**Abstract:** This paper describes how the cyber warfare methods together with the electronic warfare methods can influence the traditional naval missions and change the war character. The level of control of sea is the main element, which affects the formation of the maritime war character. Maritime actions may also be part of the strategic communication of the states which are based on the state's military strategies. The states of the Baltic Sea constitute an interesting case scenario because they are militarily allied in a different way, highly digitalized and dependent on the maritime transport. The harassment of maritime transport and the interruption of the states' western maritime connections would affect the coastal states' society of the Baltic Sea as a whole. All the Baltic Sea states actors increase their maritime capabilities by developing and testing for example cyber and electronic warfare methods and armaments. In the digital world, cyber and electronic defence capabilities play a key role as part of the credible defence of the states. The digitalization of services and activities, as well as the increasingly growing interfaces of the Defence Forces with private sector service providers, add the vulnerabilities and threats. It is questionable whether we can talk about the cyber warfare and whether the cyber-based methods are comparable to the traditional maritime warfare methods. This paper seeks to illustrate how it might be possible to combine cyber and electronic warfare methods to the maritime warfare and how the objectives of the Baltic Sea states' respond to the threats posed by these methods. These possibilities, threats and the activities already carried out are assessed in the light of the international law and the constructive theory of the war character.

**Keywords:** cyber methods, maritime warfare, electronic warfare, Baltic Sea, military strategies

---

## 1. Introduction

The Baltic Sea is relatively low and in some places a cramped water area. The international open sea area is very narrow in some places. This contributes to the increasing need of states to seek to maintain and secure their strategic interest in using the sea. Strategic areas for control of sea as well as sea lanes in the Baltic Sea are the Danish Straits, the Gotland Island, the Gogland, the narrows of the Gulf of Finland and the Aland Islands.

Everyone in the Baltic Sea coastal states has the common to dependence on the use of high technology and services in all sectors of society. This also increases the attractiveness of the use of cyber and electronic warfare methods. Cyber and electronic warfare methods have a significantly increasing position as a weapon of maritime warfare and an important part of other types of weapons. The development trend seems to be partially placing conventional maritime warfare systems on smart unmanned ships (UMS) or aircrafts (Boothby 2018, 24).

Taking care of the merchant shipping and thus the security of supply, is also one of the key tasks in addition to military and diplomatic tasks of the navy and the growing number of tasks. Naval co-operation and guidance for shipping (NCAGS) will become more important. In this task, the navy must cooperate in future with civilian actors. The most important civilian partner in surveillance and controlling merchant shipping is Vessel Traffic Service (VTS). Other important players include fairway maintenance, sea chart creators, hydrographic surveys, ports, pilots and in the Baltic Sea also icebreakers. All of these actors have important information and sensor systems that also provide information to the navy to create awareness maritime situation. This contributes to the increase of cyber vulnerabilities.

In principle, international war rules prohibit attacks on civilian targets. In the western countries, however, many of the support functions important for national defence have been privatized or implemented based on business principles. There is a risk that the measures, which are important for the protection of cyber attacks are compromised by their high cost. Commercial vessels or other civilian maritime operators are more vulnerability for cyber-attacks. By attacking to civilian systems, that produces situation awareness or other support services for the military, can in worst case also interfere with the operation of military systems and command, control, communications, computers, combat, intelligence, surveillance and reconnaissance (C5ISR).

The warships and their communications systems, sensors and weapons are often protective against cyber and electronic attacks. Their defensive systems can at the best be so powerful that they affect the operation of

attacker systems. For example, the United States Navy have placed self-defence close-in weapon systems (CIWS) aboard its warships. These systems can detect the signatures of incoming missile and rocket attacks and automatically respond with lethal force to defeat them (Thurnher 2018, 104). The connections of the warships' communication systems are dependent on the satellites and/or the link connections provided by the air forces or other maritime or land vehicles. They are increasingly reliant on space-based systems for navigation, communication, intelligence, surveillance and reconnaissance and targeting. Using and denying the use of cyberspace is already critical (Speller 2018, 193-194).

NATO's maritime actions has taken on the U.S. strategic thinking, so called All-Domain Access Doctrine (U.S. Navy 2015, 19-21) is emphasized as well as cooperation between state and non-state actors and partner countries, including the dimensions of the cyber defence and electronic warfare. However, there is not possible to dominate all domains at all times (Speller 2018, 202-203). Instead, it is possible try to create time windows and situations for actions taken domain superiority to penetrate the adversary's defences and maintain them as required to accomplish the mission changing character of war.

The control of sea is an essential factor in enabling the navy to fulfil its essential duties, such as protecting merchant shipping and information and service connections, projection power and preventing hostile marine military actions. The control of sea can be complete, temporal and regional or controversial. The sovereign state should have complete sea control in its own territorial waters. However, the degree and effectiveness of control may vary from one domain to another. Physical power and performance controlled by the sea are tied to the cyber and electronic defence capability and capabilities of the new generation weaponry systems to allow control in modern warfare.

The cyber space military dimension emphasizes status awareness, cyber deterrence, cyber physical systems and affecting the cognitive level (Lehto 2014b, 158-171). The advancers' command processes and systems can be hit from three different dimensions: 1) dimension of information; 2) dimension of decision making; and 3) dimension of communication connections (Lehto 2014a, 68). Cyber methods can be used to prevent or interfere with the warships' communication and the availability of information systems, to track vessel communications or to change available information, such as GPS-information.

In addition to the cyber methods of warfare, the use of the electronic warfare methods can be even more effective in the maritime domain than in the pursuit of the goal. Such military actions that use electromagnetic or directed energy against the enemy or electromagnetic spectrum control are considered as electronic warfare (EW). Electronic warfare includes three major subdivisions: electronic attack, electronic protection and electronic warfare support (U.S. Joint Chiefs of Staff 2007, 1-1 - 1-4). The objects of electronic warfare at sea can be, for example, radio links, radars, precision weapons, positioning and navigation systems, computers and satellite connections. Electronic warfare methods can, for example, interfere with or prevent enemy communication and other electronic activities, such as intelligence, as well as misleading, paralyzing or electronically destroying adversary's hardware and arming.

The Commander of the Russian Electronic Warfare Force, General Latochkin, has drawn attention to the role and importance of electronic warfare increased especially at the operational level. In his view, one of the focal areas of electronic warfare is the capability to fight together in the information and communication space, where it is important to protect the resources and target them to the enemy C5ISR. Lastochkin also raised up the unclear differences in the information environment between strategic, operational and tactical levels. In these circumstances, asymmetric actions that make it possible to level out the enemy's information superiority acquire special significance. According to Lastochkin, the EW's growing role has changed nature of modern warfare (McDermott 2018).

### **1.1 The theory of war character**

This research relies on the theory of the character of war. The character of war considered in the constructive frame of reference. Critical theories of constructivism also apply to the transformation processes of military doctrines through strategic communication in the formulation of the war character. The war character can be defined as follows: "The character of the war means the common perceptions in the international system of the nature, needs and possibilities of the use of armed forces, as well as the effective principles and operating models of the armed forces." In the theory of the war character, war is viewed as a pragmatic and changing

phenomenon. The war character is seen in the international system and the security environment, as well as in these preconceived operational logic, strategic communication, rules and as a construct associated with identities of the actors (Raitasalo – Sipilä 2008, 9). As the theoretical framework, the character of the war creates the normative criteria for the use of armed forces, when legitimate war can be invoked and interpretation through strategic communication. The norms of international law define the framework for what the war is considered and how to warfare. The problematic aspect of this normative framework in the war is, in particular, the obsolescence of the norms of international law, slow pace of change and poor applicability to cyberwarfare or cyber operations. The methods of cyber warfare can be used to create “instability zones” between levels of society and between elements of warfare without the goal being to strive for conventional warfare or to defeat an adversary.

## **1.2 Methods and materials**

The conceptual and technical background is a literature survey explaining the ideas and measures behind maritime cyber and electronic warfare. A context analysis of the doctrines and the strategies are used for justifying and supporting the model of war character. The primary research material consists of the official documents by the Baltic Sea states. Secondary, the sources include theoretical literature and academic papers in the field of military activities on the Baltic Sea, cyber and electronic warfare.

## **2. The context of international law**

U.S. Joint Forces Command (2016, ii) anticipates that the future security environment will be defined by twin overarching challenges: “Contested norms will feature adversaries that credibly challenge the rules and agreements that define the international order. Persistent disorder will involve certain adversaries exploiting the inability of societies to provide the government functioning in a stable and legitimate way. Confrontations involving contested norms and persistent disorder are likely to be violent, but also include a degree of competition with a military dimension short of traditional armed conflict.”

The Charter of United Nations has been considered the highest authority of international law. For example, North Atlantic Treaty recognizes its priority and key principles. This position can be questionable due to the obsolescence and weak applicability of the Charter to modern technology-weighted warfare. However, Article 51 (right of self-defence) of the Charter continues to serve as a justification for the state defence.

The United Nations Convention on the Law of the Sea (UNCLOS) is the most important provision for the international maritime navigation. UNCLOS defines the basic principles of the shipping and the rights and obligations of the states for use of the sea. The basic principle is the freedom of navigation. The passage of a foreign ship shall be considered prejudicial to the peace, good order or security of the coastal state if in the territorial sea it engages for example any exercise or practice with weapons of any kind or act aimed at interfering with any systems of communication or any other facilities or installations of the coastal state. The coastal state may take the necessary steps in its territorial sea to prevent passage, which is not innocent (art. 25).

The protocol additional to the Geneva Conventions of 12 August 1949, and relating to the protection of victims of international armed conflicts (PROTOCOL I of 8 June 1977), also includes some rules that are relevant for cyber and electronic warfare methods. The basic rules are, that the methods or means of warfare are limited and employ weapons, projectiles and material and methods of warfare of a nature not to cause superfluous injury or unnecessary suffering (art.35). The development, acquisition or adoption of a new weapon, means or method of warfare, the states are under an obligation to determine whether its employment would in some or all circumstances, be prohibited by the rules of international law applicable (art. 36). This is supported by Article 51, which discriminates prohibited attacks. Unlimited attacks which are of a nature to strike military objectives and civilians or civilian objects without distinction, are prohibited.

Armed forces in different states also increasingly use the same commercial solutions, information infrastructures and service providers as civilians. As a result of which, for example, limiting the cyber or electronic operation and its effects to the cyber domain used by the military is very challenging. This not even aim to avoid this because the wounding of the so-called civil-based information infrastructure can often be easier to cause uncertainty and chaos in the attacked society and indirectly weaken the ability of the armed forces to control their own cyber domain. On the other hand, the cyber methods can be used well targeted so that their impact

on civilian objects can be limited and controlled. Consequently, the use of cyber methods does not appear to be implicitly prohibited under Article 51.

### **3. Unaligned countries Finland and Sweden – the credible silent deterrence**

Finland and Sweden are Non-NATO states but cooperate closely with NATO countries through partnership agreements and European Union (EU) security co-operation. The European Council has adopted the Maritime Security Strategy (2014). The strategy identified risks and threats for example intentional unlawful acts at sea and in ports against ships, cargo, crew and passengers and port facilities and critical maritime and energy infrastructure, including cyber attacks. The purpose of the implementation is to strengthen the EU response and one of the main areas of strategy covers development maritime awareness, surveillance and information sharing.

The cooperation of Finland and Sweden with the NATO includes for example exchanging information on hybrid warfare, coordinating training and exercises, developing better joint situational awareness to address common threats and developing joint actions. Finland and Sweden participate in the enhanced NATO Response Force (NRF) in a supplementary role and subject to national decisions. They have signed a memorandum of understanding on Host Nation Support which, also following a national decision, allows for logistical support to Allied forces located on, or in transit through, their territory during exercises or in a crisis.

Due to the differing geopolitical position and the different political attitudes towards the development of the Defence Forces and the various external pressures on the defence of the sovereignty, there are some differences between Finland and the Sweden. Russian aggression against Ukraine and in particular the annexation of the Crimea are an awakening of the both states, especially in Sweden. Sweden has pushed down its army's capabilities and resources for decades. Now the situation is changed.

The Sweden's Defence Policy from 2016 to 2020 particularly emphasizes openness and transparency in the cooperation of the relations and defence with the U.S. and Finland. Sweden has introduced the New Total Defence Concept. In the new focus, the regional focus is a priority, emphasis the national defence and planning for wartime scenarios. Sweden is re-establishment of permanently based regular army units on the island of Gotland, increasing the reinforcement of an anti-submarine warfare capability and active cyber capabilities. Sweden will take new mobile sensors into use and keep operational mine-systems to ensure the sensor functionality of anti-submarine capabilities. According Sweden's Defence Policy that means for example an ability to carry out active operations in the cyber domain and legitimacy for the Armed Forces to protect the Swedish sovereign rights and national interests outside the Sweden's territory.

Finland has also started to accelerate the development of its naval and air forces. The ability to protect against electromagnetic harassment and utilize artificial intelligence has been emphasized in procurement. Unlike the strategy papers of the neighbouring countries, the Finnish strategy papers are very defensive and do not directly allow the use of military means, such as cyber methods, for such purposes that could be interpreted as being offensive or affecting a potential adversary. Finland's Security Strategy for Society (2017) defines that the primary objective of the development and maintenance of Finland's defence capability is to establish a deterrence but Finland is preparing to meet more multifaceted threats, which combine military and non-military means.

However, Finland does not hide its awareness of the challenges posed by its geopolitical position. The focus areas of maintaining and using the defence system during the span of the Finland's Defence Report (2017) are for example an intelligence, cyber-defence and long-range strike capability. Also Finland's Defence Forces will continue developing a cyber defence capability in accordance with the national Cyber Security Strategy (2013).

Based on the aforementioned surveys of the Finnish and Swedish maritime and cyber warfare strategic alignments, it can be concluded that the outwardly directed strategic communication of both states tacitly emphasizes the existence and decisive development of the military power and capabilities rather than threatening or displaying the military power and capabilities. It is evident that both countries have a quiet doctrine that Russia respects the power, so these non-aligned states must have their own credible defence but also a close relationship with the NATO.

#### **4. NATO focuses on cyber defence and airspace management**

NATO has increased its ability to respond to hybrid threats. In the Baltic Sea, NATO's focus is to increase airspace control alongside cyber capabilities. According to Brussel Summit Declaration (2018), NATO will reinforcing a maritime posture and has taken concrete steps to improve their overall maritime situational awareness. They will prepare strategic assessments on the Baltic Sea and reinvigorate collective maritime warfighting skills in key areas, including an anti-submarine warfare, amphibious operations and a protection of the sea lines of communications. Vananga has estimated that the deterrence became the new mind-set of the alliance but there is a lack of the clarity in the command and control (C2) structure and the rules of an engagement in case of conflict (Vananga 2018, p.42; Howard 2018, 87). Given that, NATO's challenge is not only to develop the common C2 system but also to integrate the nationals C2 systems with that (Howard 2018, 88). Alongside the extensions associated with C2, Vego (2009, 5) raises the challenges of combining the logistical support and sustainment of multiservice and especially multinational forces, differences in doctrine and weapons and equipment.

The cyber defence is mentioned as part of NATO's core task of the collective defence. NATO will continue to implement cyberspace as a domain of operations. Reaffirming NATO's defensive mandate, the leaders of NATO members determined to employ the full range of capabilities, including cyber, to deter, defend against and to counter the full spectrum of cyber threats, including those conducted as part of a hybrid campaign (NATO 2018). That was a strong strategic message from NATO leaders that the tolerance to accept targeted hybrid attacks has fallen.

#### **5. And how about Russia?**

The Military Doctrine of Russia (2014) highlights that one of the main tasks of Russia is to deter and to prevent military conflicts is to create conditions to reduce the risk of using information and communications technologies for the military-political purposes. As a way to reduce these risks, Russia has started building its own internet network (RUNET) which they could, if necessary, detach from the global internet. In addition to strengthening its own cyber security and digital sovereignty, RUNET represents Russia's ambition to challenge the Western dominance in the information space and protect the Russian culture and values (Ristolainen 2017, 13-15).

That task can also be interpreted as a direct command to use cyber operations to create political and military conditions appropriate to Russia. Combining that with the justification of Article 31 (the Armed Forces may be employed outside the country to protect the interests of the Russia and its citizens), we can speak of a strong justification for the military and other security authorities (especially the intelligence agencies) to take even offensive measures everywhere in the world where Russian citizens live or otherwise have activities. Of course, provided that the purpose is also the maintenance of international peace and security.

According the Maritime Doctrine of Russia (2015) addition to combating aggression against Russia, the Navy maintains military and political stability on the world's seas, including through introduce of the Russian flag and military power. Russia has been busily presenting its flag in the Baltic Sea, especially during military exercises of other countries. The focus of Russian Maritime Doctrine has shifted from the oceans to the management of the Russian coast and neighbouring areas and to foster Russian interests in these areas (Russian Ministry of Defence 2017).

As emphasized in the Russian military doctrine, also in the Maritime Doctrine highlights development and implementation the rules of international law, the provisions of the international treaties and safeguarding the national interest of the Russia recurring challenges and threats to the national security of the Russia from the ocean and the sea. All of the Russian strategy papers / doctrines have highlighted compliance with international treaties and obligations. Russia has also actively demanded the updating of international agreements or the conclusion of new agreements, cyber-technology, including artificial intelligence, exploitation in military use, and denial of the space armed. At the same time, Russia strongly invests in the development of these capabilities. Such activity can be seen as quite rational. By achieving a strong position and thus a strong deterrent effect, for instance in space equipment, the Russian negotiating position at the contract table will improve. Achieved capabilities can be bargained while at the same time ensuring that adversaries cannot develop their own capabilities at the same level. Achieved superiority also allows trading in other areas of regulation and cross-subsidization.

The Maritime Doctrine confirms narratives maintained by Russia's leadership in Russia as a fortress surrounded by adversaries. Article 52 states, that: "The decisive factor in relations with NATO remains the tactics of the Alliance to advance its military infrastructure to Russia's borders and the attempt to globalize the efforts, which is unacceptable to the Russian Federation." As an example of maintaining and strengthening such a narrative, the Russian Ministry of Defence, General Soigu has expressed his concern about the convergence of the Baltic non-allied countries, Finland and Sweden with NATO (Russian Ministry of Defence, 2018). Unlike General Soigu claimed and as explained in Section 3 above, Finland and Sweden do not have full access to NATO exercises and command structures and also NATO has no unrestricted access to Finnish or Swedish territorial waters and airspace.

Concerning the Baltic Sea, the Maritime Doctrine highlights the importance to ensure transport accessibility to the Kaliningrad region, to the conduct of potentially hazardous underwater objects, the condition of undersea pipelines and to the development of military capabilities, as well as the system of military installation of the Baltic Fleet. Kaliningrad where the Russian Baltic Sea Naval Staff is located is a pain point for Russia and for the West. In the absence of a land connection, Kaliningrad is fully dependent on the sea and air defence in crises. This also hampers the management of the cyber environment of the area, exposing it to various threats.

## **6. Conclusions and discussion**

The methods of cyber and electronic warfare alone and combined with other methods of the warfare, have changed the nature of the maritime warfare. They have a strong influence to the character of the war, to the formation, change, and demands of the war that this reaction requires. The sea control and the control of airspace beyond are no longer sufficient but the defender must also control the cyber domains and air space as far as its marine combat systems require.

The maritime warfare will become increasingly automated as unmanned combat ships and artificial intelligence weapon systems develop and become more common. At the same time, the number of players in the battlefield are increasing. In addition to governmental actors, non-governmental service providers and possibly non-governmental organizations with their own political goals will also be involved. There is the war of all against all, *bellum omnium contra omnes*. The threats and struggle of the domains develop and become more complicated.

The Baltic Sea is a strategically important area for Russia, but also for NATO because of its credibility. NATO's eastern partners, as well as Finland's access to the Baltic Sea and to the Atlantic Ocean is a lifeline for the maritime transport. The Baltic Sea is a sensitive area due to the dependence of coastal states' activities and population on high technology and maritime transport. Russia's position in the Baltic Sea is not easy because it must be able to protect the maritime transport in the port of Primorsk, Ust-Luga and Sankt Petersburg but also in Kaliningrad, which lies between the NATO countries. This position also creates the basis for a narrative to Russian people in Russia, as a state surrounded by enemies. In relation to other Baltic Sea states, Russia seeks to correct this position, for example through a strategic communication. Displaying the flag and potential power in the Baltic Sea as well as a regular communication with non-allied countries with the need to defend themselves as demanded as power relations change, working alongside other non-public methods, creating fear and the balance of power.

The occupation of the Crimean Peninsula serves as a good example of a change in the character of the war. An individual event can quickly change the perception of military threats, military actors and when a war equated the conflict is born and how it can be responded to. At the same time, new issues arise, for the example legitimacy of the methods and aims of the warfare and the factors that are important to increase a credible military power and which military means are effective in achieving the goal of the war. Despite the condemnation of the Crimea occupation, Russia can be considered to have made a great service to the Baltic Sea coastal states by inspiring them to focus heavily on the development of the capabilities of their defence forces and hybrid warfare capabilities.

## **References**

- Boothby, William. (2018) Dehumanization: Is there a Legal Problem Under Article 36? In "Dehumanization of Warfare. Legal Implications of New Weapon Technologies", Von Heinegg, WH. and Frau, R. and Singer, T. (eds.), Switzerland, pp.21-52
- Finland's Ministry of Defence: Finland's Cyber Security Strategy (2013) [online]  
[https://www.defmin.fi/files/2378/Finland\\_s\\_Cyber\\_Security\\_Strategy.pdf](https://www.defmin.fi/files/2378/Finland_s_Cyber_Security_Strategy.pdf)

- Finland's Government's Defence Report (2017) [online]  
[https://www.defmin.fi/files/3688/J07\\_2017\\_Governments\\_Defence\\_Report\\_Eng\\_PLM\\_160217.pdf](https://www.defmin.fi/files/3688/J07_2017_Governments_Defence_Report_Eng_PLM_160217.pdf)
- European Union Maritime Security Strategy (2014) [online] [https://ec.europa.eu/maritimeaffairs/policy/maritime-security\\_en](https://ec.europa.eu/maritimeaffairs/policy/maritime-security_en)
- Howard, Glen E. (2018) Enabling Deterrence: U.S. Security Policy Toward the Baltic. In "Security of the Baltic Sea Region Revisited amid the Baltic Centenary." Sprūds, A. and Andžāns, M. (eds.). [online] The Rīga Conference Papers, pp.83-98, <http://www.liia.lv/en/publications/security-of-the-baltic-sea-region-revisited-amid-the-baltic-centenary-the-riga-conference-papers-2018-741>
- Joint Declaration on EU-NATO Cooperation by the president of the European Council, the president of the European Commission, and the secretary general of the North Atlantic Treaty Organization 10.7.2018. [online]  
[https://www.consilium.europa.eu/media/36096/nato\\_eu\\_final\\_eng.pdf](https://www.consilium.europa.eu/media/36096/nato_eu_final_eng.pdf)
- Lehto, Martti. (2014a) Kyberaistelun toimintaympäristön teoreettinen tarkastelu. In "Kyberaistelu 2020" Kuusisto, T. (ed.), Maanpuolustuskorkeakoulun taktiikan laitos. Julkaisu-sarja 2: Asiatietoa, No. 1/2014, pp. 67-89
- Lehto, Martti. (2014b) Kyberaistelu ilmavoimaympäristössä. In "Kyberaistelu 2020" Kuusisto, T. (edit.), Maanpuolustuskorkeakoulun Taktiikan laitos. Julkaisusarja 2: Asiatietoa, No. 1/2014, pp. 157-178
- McDermott, Roger N. (2018) Russia's Evolving Electronic Warfare Capability: Unlocking Asymmetric Potential. [online]<https://www.icds.ee/blog/article/russias-evolving-electronic-warfare-capability-unlocking-asymmetric-potential/>
- NATO Public Diplomacy division: BRUSSELS SUMMIT DECLARATION Issued by the Heads of State and Government participating in the meeting of the North Atlantic Council in Brussels 11-12 July 2018. PRESS RELEASE 2018 PR/CP(2018)074 (11 July) [online] [https://www.nato.int/cps/en/natohq/official\\_texts\\_156624.htm](https://www.nato.int/cps/en/natohq/official_texts_156624.htm)
- Protocol additional to the Geneva Conventions of 12 august 1949, and relating to the protection of victims of international armed conflicts (PROTOCOL I of 8 June 1977. [online]  
[https://www.icrc.org/en/doc/assets/files/other/icrc\\_002\\_0321.pdf](https://www.icrc.org/en/doc/assets/files/other/icrc_002_0321.pdf)
- Ristolainen, Mari. (2017) Should 'RuNet 2020' be taken seriously? Contradictory views about cybersecurity between Russia and the West. In "GAME CHANGER - Structural transformation of cyberspace. Kukkola, J and Ristolainen, M and Nikkarila, J-P (eds.) Finnish Defence Research Agency Publications 10. Riihimäki 2017, pp.7-26
- Russian Ministry of Defence: Military Doctrine of the Russian Federation 2014. [online] www.mil.ru
- Russian Ministry of Defence: Maritime Doctrine of the Russian Federation 2015. [online] www.mil.ru
- Russian Ministry of Defence (2017) Fundamentals of the state policy of the Russian Federation in the field of naval activities for the period up to 2030, I /pp. 8-17. [online] www.mil.ru
- Russian Ministry of Defence (2018) Moscow hosted the meeting of Defence Ministry Board Session. [online]  
[http://eng.mil.ru/en/news\\_page/country/more.htm?id=12187378@egNews](http://eng.mil.ru/en/news_page/country/more.htm?id=12187378@egNews)
- Speller, Ian (2014) Understanding Naval Warfare, Routledge, Oxon
- Speller, Ian (2018) Understanding Naval Warfare. Second edition. New York
- Swedish Ministry of Defence (2016) The Sweden's Defence Policy. [online]  
[https://www.government.se/globalassets/government/dokument/forsvarsdepartementet/sweden\\_defence\\_policy\\_2016\\_to\\_2020](https://www.government.se/globalassets/government/dokument/forsvarsdepartementet/sweden_defence_policy_2016_to_2020)
- Raitasalo, Jyri and Sipilä, Joonas (2008) Näkökulmia sotaan. In "Sota – teoria ja todellisuus. Näkökulmia sodan muutokseen" Raitasalo, J. and Sipilä, J. (eds.). Helsinki, 2008, pp.1-10
- Thurnher, Jeffrey S. (2018) Feasible Precautions in Attack and Autonomous weapons. In "Dehumanization of Warfare. Legal Implications of New Weapon Technologies" Von Heinegg, WH. and Frau, R. and Singer, T. (eds.), Switzerland, pp.99-117
- United Nations Convention on the Law of the Sea (UNCLOS). [online]  
[http://www.un.org/Depts/los/convention\\_agreements/texts/unclos/unclos\\_e.pdf](http://www.un.org/Depts/los/convention_agreements/texts/unclos/unclos_e.pdf)
- U.S. Joint Chiefs of Staff: Electronic Warfare. [online] Joint Publication 3-13.1, 25 JAN 2007, Washington, DC  
<https://apps.dtic.mil/dtic/tr/fulltext/u2/a464647.pdf>
- U.S. Navy (2015) A Cooperative Strategy for 21st Century Seapower. [online]  
<https://www.navy.mil/local/maritime/150227-CS21R-Final.pdf>
- Vanaga, Nora (2018) NATO's Conventional Deterrence Posture in the Baltics: Strengths and Weaknesses. In "Security of the Baltic Sea Region Revisited amid the Baltic Centenary" Sprūds, A. and Andžāns, M. (eds.) [online] The Rīga Conference Papers 2018. pp. 31-42. <http://www.liia.lv/en/publications/security-of-the-baltic-sea-region-revisited-amid-the-baltic-centenary-the-riga-conference-papers-2018-741>
- Vego, Milan (2019) Operational Warfare at Sea. Oxon, USA

# Grasping Cybersecurity: A set of Essential Mental Models

Jan van den Berg

Delft University of Technology, Delft, The Netherlands

[j.vandenberg@tudelft.nl](mailto:j.vandenberg@tudelft.nl)

**Abstract:** For most people, cybersecurity is a hard to grasp notion. Traditionally, cybersecurity has been considered as a technical challenge and still many specialists view it equivalent with information security, with the notions of confidentiality, integrity and availability as starting points of thinking. And although others searched for a broader perspective, the complexity and ambiguity of the notion still thwarts a common understanding. While developing and executing a MSc cybersecurity program for professionals, the lack of a common understanding of what cybersecurity entails was again observed. Stimulated by this, we started to look for and define a new, transdisciplinary conceptualization of cybersecurity that everyone can agree upon. It resulted in two scientific papers published. This paper describes the outcomes of the continuation of our research journey. It turned out that the earlier introduced description of two key notions, namely that of *cyberspace* and that of *cybersecurity*, can still be considered as adequate starting points. Here, we describe a set of additional mental models that elaborates them and provides more detail to the meaning of the two key notions. In practice, it turned out that the additional mental models strongly support the description and analysis of existing and upcoming cybersecurity challenges and helps to understand how everybody, in his or her various roles, can or should contribute to reducing the related cyber risks to adequate levels. We further discovered that for certain cybersecurity challenges, especially those related to efficient cyber risk mitigation, we could not yet identify an adequate sub-set of mental models. This defines the agenda for near future cybersecurity research.

**Keywords:** cyberspace, cyber activities, cybersecurity, cyber risk management, mental models, holistic view, cyber situational awareness, cyber risk assessment, cyber risk mitigation

---

## 1. Introduction

For most people, including those responsible for it, cybersecurity is a hard to grasp notion. Traditionally, cybersecurity has been considered as a technical challenge and still many specialists view it equivalent with Information or IT Security, with the notions of Confidentiality, Integrity and Availability (CIA) as starting points of thinking. Also in information security standards like the (famous) ISO/IEC 27000-series (ISO/IEC JTC 1, 2018), (ISO/IEC JTC 1, 2005), the key asset chosen is ‘information’ and the ‘preservation of the confidentiality, integrity and availability’ of information is defined as the key information security challenge.

This conceptualization maybe rather clear to IT specialists (like those working in hardware and software R&D), for policy makers, strategic managers and end-users, among others, this cybersecurity framing is difficult to grasp and does not invite for proper actions from their side. As a consequence, many actors in cyberspace have difficulties in defining what their role can and should be in securing the digital environment. It leads to the situation that in many cyber sub-domains, a coherent cybersecurity approach is missing. Based on this, we argue that there is need for a re-conceptualization of what the cybersecurity challenge entails. More precisely we claim that there is a need for a broad view on cybersecurity that (a) everybody can grasp, and (b) enables that everybody understands how he/she can contribute to securing cyberspace, in each of his/her cyber activity roles.

While developing and executing an executive MSc Program Cybersecurity for professionals (Cyber Security Academy, 2019), we worked on the creation of a holistic view on cybersecurity and discovered that mental models turn out to be very useful to create a common conceptualization, understanding, and language about what cybersecurity essentiality is. Our work resulted in two papers (Van den Berg et al., 2014), (Van den Berg, 2018), in which we brought forward (the) two key elements of cybersecurity being (*i*) a clear *conceptualization of cyberspace*, and, (*ii*) a basic *definition* of what cybersecurity, i.e. *securing cyberspace*, encompasses.

During further cybersecurity research as well as continued execution of the MSc program, we elaborated these ideas by collecting all kinds of additional models and best practices in attempts to deepen the new conceptualization. This paper describes how far we have reached now by sketching the set of mental models that are thought to be most essential. In addition, we describe in which part of the cybersecurity challenge we are still missing some basic mental models. In a way to validate the proposed set of essential mental models, we also describe some examples of cybersecurity research in which these models have been applied.

The remainder of this paper is structured as follows. In section 2, we describe the basic model of *cyberspace* (consisting of three layers) and three supportive mental models, one for each layer. This creates the basics for describing in section 3 what the *cybersecurity challenge* essentially is using again one basic model (related to a cyber risk management cycle), supplemented by a series of supportive mental models. We here also identify some gaps in our body of cybersecurity knowledge (inviting for 3 key topics of cyber research to be performed in the near future). In section 4, we provide, in an attempt to validate the proposed set of mental models, the results from recent research, in which these models have been applied. Finally, in section 5, we draw conclusions and summarize future research topics.

## **2. Cyberspace and its security concerns**

### **2.1 Three-layer model of cyberspace**

The ISO/IEC standard 27032 (ISO/IEC JTC 1, 2012) - having the aim ‘to clarify the terms and demonstrate global leadership with this and the other cybersecurity standards projects now in progress’ - defines cyberspace as ‘the complex environment resulting from the interaction of people, software and services on the Internet, supported by worldwide distributed physical information and communications technology (ICT) devices and connected networks.’

This framing is somewhat related to Enterprise Resource Planning (Jacobs, F.R., Weston Junior, F.C.T., 2007) type of thinking where, in a layered approach, business processes, as executed by people and machines, are enabled by supportive IT services. The framing also relates to the purport of the well-known ‘People, Process, Technology’ triangle, where – in the context of software applications – people are split into end-users (of the applications) and application creators (i.e., IT specialists), where (business) processes relate (and should be aligned) to the strategic business goals of the organization (to be fixed by the strategic management), and where the supportive software applications enable better business-decisions by relevant decision makers (Halo Business Intelligence, 2009).

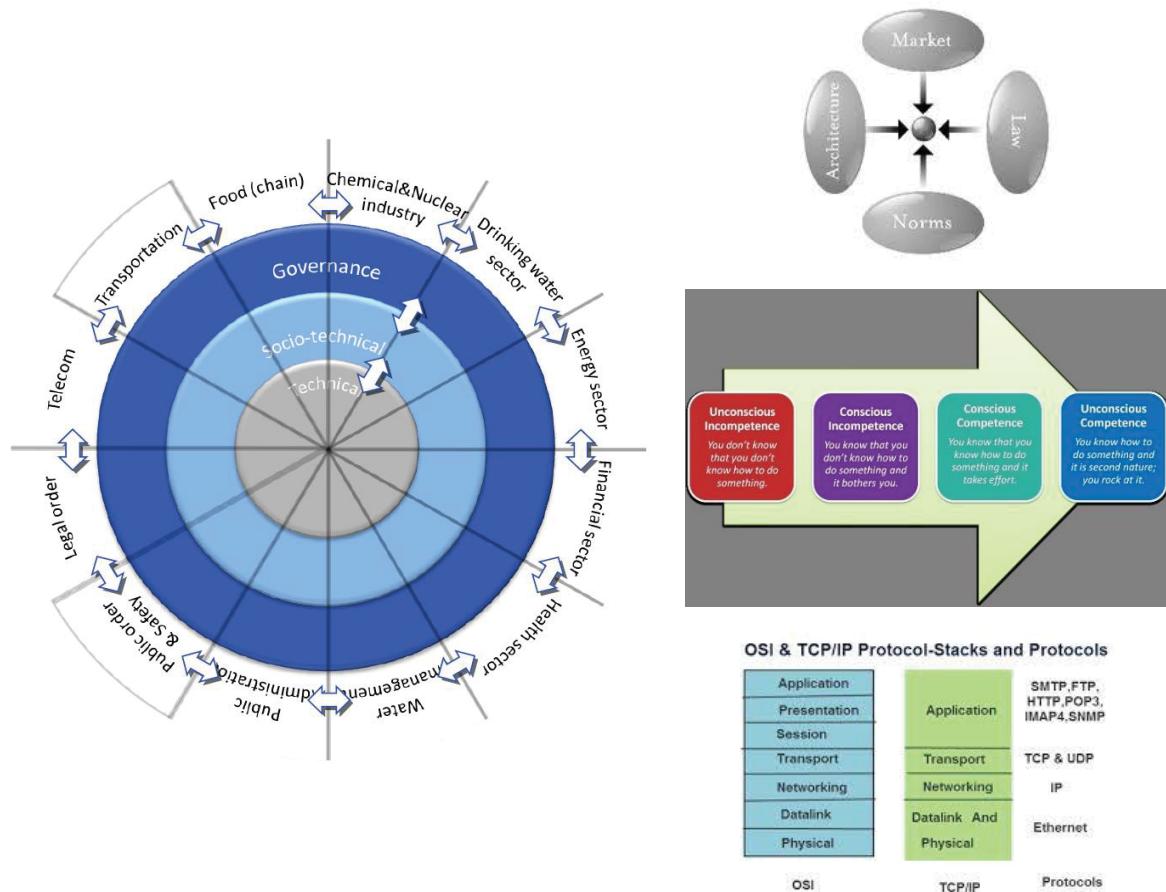
Based on an analysis of existing information security approaches and inspired by the above-given frameworks of thinking, we have provided a new conceptualization of cyberspace (Van den Berg, 2018) consisting of a three layer model, shown in figure 1, left side. It concerns the basic cyberspace model. The *middle layer* concerns the *socio-technical layer of cyber activities* as being executed by people and smart IT, in attempts to reaching their personal, business, or societal goals. Examples of cyber activity behavior include searching on the world wide web, execution of financial transactions, manufacturing goods and products, controlling critical infrastructures, law enforcement pursuits (around, for example, privacy breaches and selling illegal products in the dark web), up till criminal cyber activities of all kind, and cyber warfare operations. The *inner layer* of the cyberspace model concerns all IT that enables the cyber activities. So, in other words we can say that *cyber activities* are, basically, *IT-enabled activities*. The *outer layer* of the model concerns the *governance layer of rules and regulations* that should be put in place to properly organize the two underlying layers, including their security. The three layer model visualization further shows a sub-division of cyberspace in example *cyber sub-domains* to emphasize that cyber activities in different domains have often different characteristics and, as a consequence of that, different security requirements.

Due to the continuous process of strong digitization in almost all domains of society, the amount and variety of cyber activities we currently execute at home, when traveling, at work, and beyond, is enormous, and is still growing. For many of us, it is quite challenging to adequately cope with the rapid digitization developments and to stay competent as ‘*homo digitalis*’. The basic challenge for *adequate cyber behavior*, which include secure cyber behavior, maybe therefore be formulated as becoming *unconscious cyber competent*. This is visualized in the middle at the right side of figure 1: the basic challenge is that every cyberspace actor, regarding all his/her cyber activities, takes the path from the state of being ‘*unconscious incompetent*’, via ‘*conscious incompetent*’ and ‘*conscious incompetent*’, to the final state of being ‘*unconscious competent*’.

To clarify the key issues of the governance layer, we have chosen as mental model the picture shown at the top of the right side of figure 1. It concerns the *four modalities of regulation* in cyberspace as proposed by Lawrence Lessig (Lessig, 1999) being *laws* (next to rules, policies, and regulations), *norms* (informal societal rules), *markets* (to create the right incentives for stakeholders), and *architecture* (which concern physical or technical constraints on cyber activities). It should be clear that this framing of the four modalities of regulation is precisely

in line with the three layer model of cyberspace: the modalities laws, norms and markets steer cyber activities (in layer 2) from a governance perspective (in layer 3), while the modality architecture (in layer 1) put constraints on cyber activities using a technical approach.

For the *technical layer*, the key issues relate to the two protocol stacks that are in use to describe computer networks, namely, the OSI and TCP/IP protocol stacks. These stacks are in the core of what every cyberspace actor should understand to a certain extent.



**Figure 1:** The 3-layer model of cyberspace (left), and three models describing the key cyberspace issues per layer (right)

## 2.2 Security concerns of cyberspace layers

Having conceptualized cyberspace in three layers, we can determine the *security concerns per layer*. To do so, let us start considering the key assets in cyberspace being the cyber activities. We may say that *cyber security* basically concerns the *security of cyber activities*, which actually is about the *security of cyber behavior!* It is easy to see that the security requirements of a cyber activity strongly depend of the type of the activity and its context. For example, the requirements related to the execution of a financial transaction in a public environment relate to secure payment behavior: careful use of debit/credit card, checking the amount before paying, shielding the keyboard of the payment equipment while typing your pin number, and inspection of the correctness of the receipt. When considering the automatic control of a critical infrastructure like water supply, the cyber security requirements are very different and are basically about guaranteeing continuous automatic supply of clean water to recipients, monitoring of this process, and committing necessary interventions ‘through SCADA systems attached to distributed control systems (DCS), programmable logic controllers (PLCs), remote terminal units (RTUs) and field devices’ (ISACA, 2016). More in general, we can say that issues of secure cyber activity behavior (also) relate to minimization or no use of USB-sticks, to always choosing strong passwords, to never clicking on a URL in an email, and to very limited (or no) downloading of files from the Internet, among others.

Looking at the technical layer, we can use the classical requirements of information security being *confidentiality, integrity, and availability* (CIA) (ISO/IEC JTC 1, 2018). For the financial payment example just mentioned, all the three CIA requirements are relevant, while more refined technical requirements might be added here like secure identification, authentication and access (IAA) control, and non-repudiation. Note that, as a consequence of separating cyberspace in layers, we explicitly *discriminate between cyber security* (being the security of cyber activities in layer 2) and *information security* (being the security of IT of layer 1), a distinction that is quite different from the current practice.

Continuing our way of thinking, we further make the observation that incidents in the technical layer (often termed information security breaches) are actually (cyber) threats for the cyber activities executed in the socio-technical layer. If such cyber threats, emerging as information security incidents in the technical layer, result into incidents in the socio-technical layer, we can term these incidents cyber security incidents or cyber security breaches, which again shows an important difference in meaning of ‘cyber’ and ‘information’. In short, within our conceptualization of cyberspace and cyber security, information security is truly something else than cyber security.

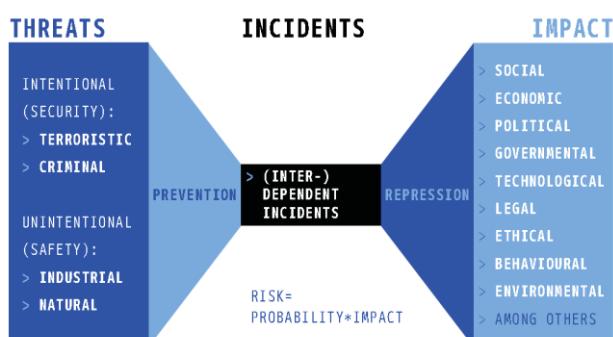
Finally, the security concern of the governance layer encompasses the fixation of rules and regulations (using the governance modalities mentioned above and in accordance with the chosen ‘risk appetite’: see below) for both the socio-technical and the technical layer. So, these governance rules and regulations should be related to both secure cyber activity behavior (in layer 2) and secure IT (in layer 1).

### 3. Modeling the cybersecurity challenge

Before diving into the cyber security challenge (which, as we will argue, basically concerns a risk management challenge), it is relevant to put forward the modern notion of risk. According modern standards, risk is the potential of gaining or losing something of value, or, according (ISO/TC 262, 2018), the (positive or negative) ‘effect of uncertainty on objectives’. In the financial world, this phenomenon is well-known since investments can result into an actual return on an investment that is higher (opportunity) or lower (potential loss) than the expected return. In cyberspace we observe similar symptoms since digitization usually offers expected opportunities like efficiency, cost reduction, convenience, et cetera, but at the same it enhances the ‘cyberattack surface’ creating higher cyber risks. While for decision makers on cyber security it is always wise to take this two-sided view on cyber risk into account, the focus of our discussion in the remainder of this paper is mostly focused on the negative part of it.

#### 3.1 The bowtie and the cyber risk management cycle

A well-known basic model used in safety and security science is the bowtie model. Basically, it reasons from (intentional and unintentional) threats to incidents next to the impact of the latter. Incidents occur with a certain probability or likelihood, and risk of a threat is defined as the expected impact of this threat, i.e., Risk = Likelihood times Impact. In cyberspace, the bowtie model can be used to model cyber threats, cyber incidents and their impact. To prevent cyber incidents from happening, preventive measures can be taken to reduce the probability of their occurrence. To reduce impacts of a given incident, repressive measures can be taken like measures related to detection and recovery. For more details on (the use of) the bowtie model, we refer to (Zipp, 2015). In figure 2, at the left side, a visualization of the bowtie is provided.



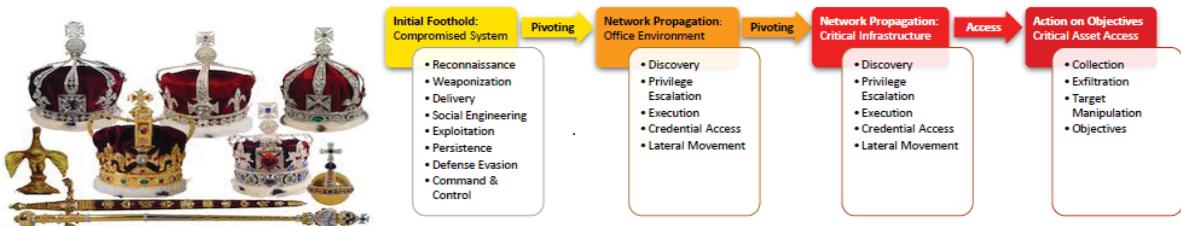
**Figure 2:** The bowtie model (left), and the basic cyber risk management cycle (right)

Since cyber activities are threatened, related cyber incidents may occur, sometimes with high impact. This clarifies why cyber security is actually a risk management challenge. Here, again, standards can help like the ISO standard (ISO/TC 262, 2018) mentioned above. It mentions that for proper risk management a risk management process should be executed. Here we provide this risk management cycle on the right in Figure 2, in the context of securing cyberspace. Within our framing, this concerns risk management of cyber activities.

- Repeat ‘forever’  
 (in all ‘relevant’ cyber sub-domains)
1. Identify the *critical cyber activities*  
 (sometimes termed the ‘crown jewels’)
  2. Identify & assess their *cyber risks*  
 (potential gains & losses)
  3. Define *acceptable* cyber risk levels
  4. Decide way(s) of *dealing with the risks*
  5. Design & Implement *cyber risk measures*
  6. *Monitor effectiveness of measures taken.*

### 3.2 Additional mental models

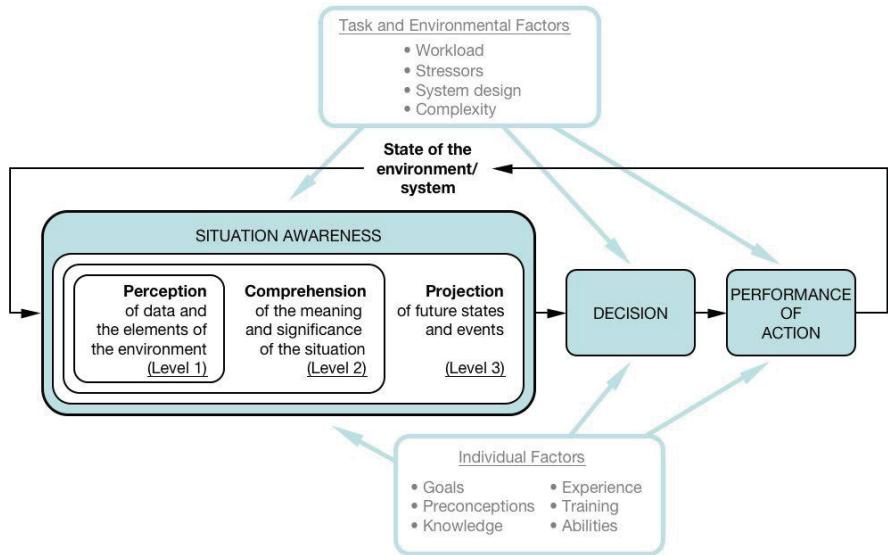
Having defined and visualized the basic mental cyber security model in Figure 2, we can now sketch a set of additional models that provide background details. This is done by considering each of the six steps of the basic cyber risk management cycle. The first step concerns the identification of the critical cyber activities as executed by a person, organization or society. The critical cyber activities are the IT-enabled activities we mostly depend on and are, in case of being disrupted, expected to create the highest impact. In society, critical cyber activities are related to critical infrastructures like transport (of goods and people), supply of water & energy (electricity, oil, gas), as well as the financial, healthcare, and first aid services. In a digitalized corporate environment, data are often considered as critical and sometimes termed the ‘crown jewels’ (Fredriksen, 2018). Within our framing, we consider not its data but the critical cyber activities of a corporation as its crown jewels (which usually relate to its critical business processes) and they need to be cyber secured with the highest priority. For a visualization of the crown jewels, we refer to Figure 3 (left). But before being able to think about their security, they should be first identified and defined, which is still not common practice in many organizations as we have often observed.



**Figure 3:** Two additional mental models related to the first two steps of the basic cyber risk management cycle: crown jewels (left) and the unified kill chain (right)

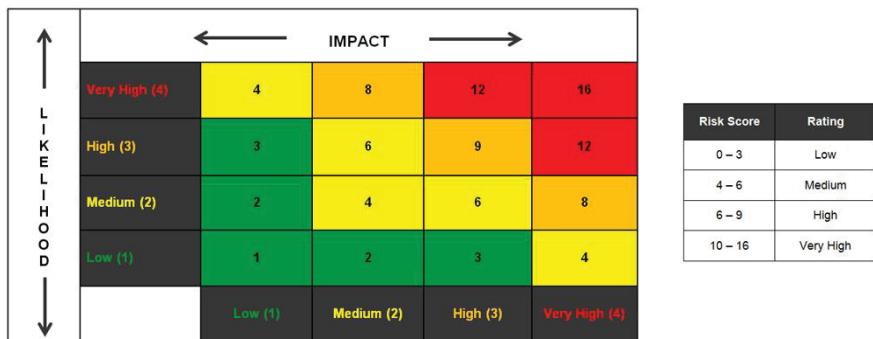
Once a person, organization or society has identified his/her/its critical cyber activities, the second risk management cycle step of identifying and assessing their cyber risks can be taken up. In case of considering intentional attacks, the ‘unified kill chain’ model can be applied for analyzing and defending against possible attack behavior: this model (visualized in Figure 3, right) describes in detail all possible steps an attacker can choose in attempts to disrupt (y)our critical cyber activities (Pols, 2018).

Such an analysis is of course only possible if we carefully monitor the cyber activities taking place on (y)our IT systems connected to the Internet or, in other words, we need to create sufficient ‘cyber situation awareness’. The general notion of situation awareness was introduced in 1995 by Micah Endsley (Endsley and Jones, 2016) and is here applied in the context of securing cyberspace (Figure 4).



**Figure 4:** Another additional mental model related to the second step of the basic cyber risk management cycle: (cyber) situational awareness (Endsley and Jones, 2016)

Having organized adequate cyber situational awareness, one can try to assess the risk related to possible cyber activity incidents in terms of likelihood times impact. There are numerous methods available to make such assessments (for an overview we refer to (ISO/IEC TC 262, 2009)), but a picture showing the principle ideas is shown in Figure 5. The risk values found are shown with a color and range from low risk (green) to very high (red). This picture completes the set of four additional mental models related to the first two steps of the basic cyber risk management cycle.



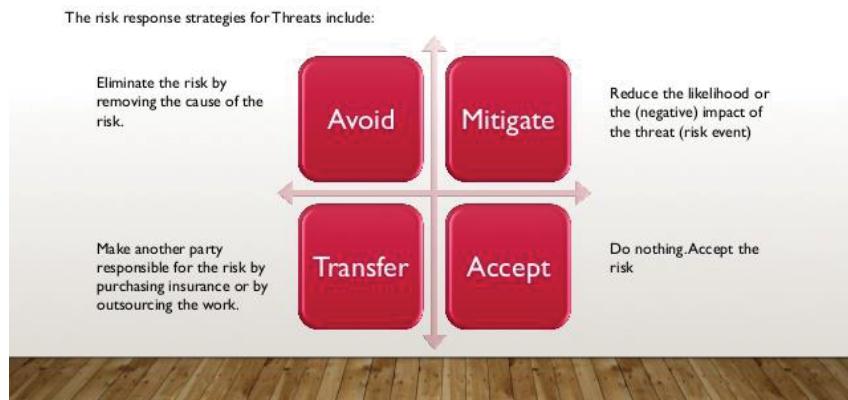
**Figure 5:** Yet another mental model related to the second step of the basic cyber risk management cycle: cyber risks assessment

The next step of the basic cyber risk management cycle concerns the fixation of acceptable cyber risk levels for each of the critical cyber activities. This relates the so-termed risk appetite of an individual, organization or society. Defining the risk appetite falls for the larger part outside the scope of science since it concerns mostly a choice and is often based on personal judgements. However, some remarks are of relevance here. For critical infrastructures, governments often determine the required risk levels as is common practice in the worlds of, for example, finance, energy supply, flood defense and (also) IT systems. Being not compliant with the related rules and regulations can result in severe penalties, which of course influences the risk appetite of an organization. Similarly, shareholders do have their ideas about managing (cyber) risk and the related risk appetite, so their voice is often decisive in the risk appetite choice of an organization.

Having assessed the relevant cyber risks and having chosen the acceptable cyber risk levels, the fourth step concerns the decisions how to deal with the assessed risks. A well-known principle from safety & security science tells us that there exist basically four response strategies to negative risk (Dorfman, 2007). The first one is *avoidance* by stopping the risky cyber activity at stake. The second one is *transfer* by making another party responsible for the risk through insurance or outsourcing. The third option is simply *accepting* the risk in case it is within your risk appetite. And the final response strategy is *mitigating* the risk to the defined acceptable risk level.

level by reducing the probability and/or impact of the cyber threat. The four possible risk response strategies have been visualized in Figure 6.

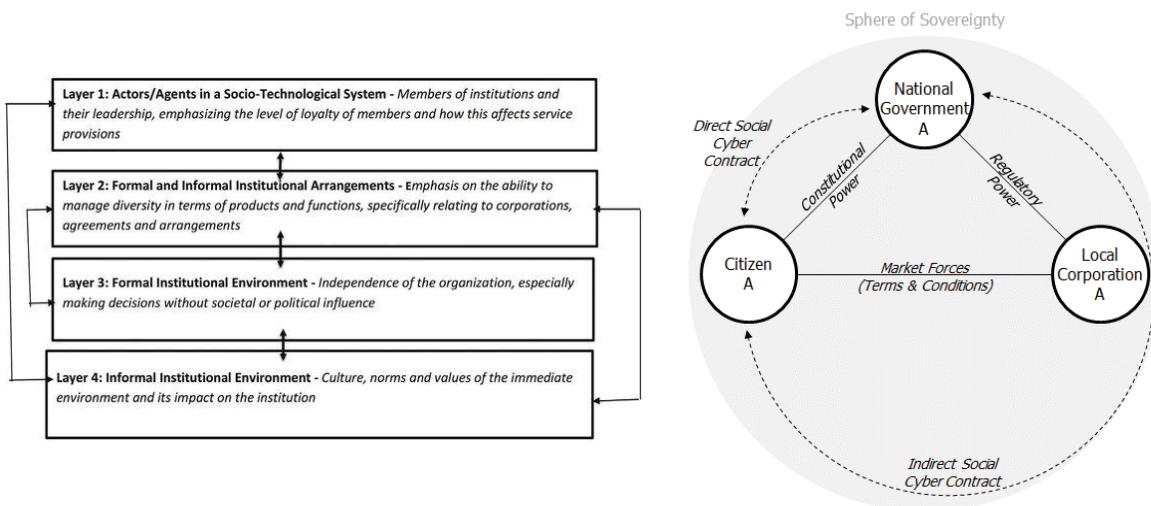
## Risk Response Strategies - Threats



**Figure 6:** A mental model related to the fourth step of the basic cyber risk management cycle: risk response strategies

Step 5 of the basic cyber risk management cycle concerns the design and implementation of cyber risk measures, which is relevant in case you have adopted the strategy of risk mitigation in the previous step. This concerns a complex challenge since an abundance of preventive and repressive mitigation measures exist. Within our conceptualization of cyberspace, the challenge boils down to designing and implementing a ‘balanced set of cyber risk mitigation measures’ in the three layers. A simple example may illustrate the basic idea. Consider the case of using USB sticks, which is often a risky cyber activity since malware can be easily and fast transferred if sticks are used in different IT environments. Measures to mitigate such a malware infection risk at the socio-technical layer concern measures related to cyber activity behavior: someone, who feels herself a potential target for an infection attack via a USB-stick, might decide not to use such a device nor allow anyone else to use it on her PC or laptop. The identified USB infection threat might also be mitigated at the technical layer by disabling all USB ports in the IT environment at stake or, in a less restrictive approach, by monitoring USB stick injections and scanning on infections before allowing data retrieval from such sticks. Finally, at the governance layer, rules might be made official that USB sticks are not allowed inside a certain IT environment and, in case of a cyber incident occurrence due to a violation of this rule, the (financial or other) consequences are for the person who violated the rule.

In practice, cyber risk mitigation is usually a much more complex challenge than the simple example shows. Considering for example again critical infrastructures in our digitized society, we immediately observe that usually many stakeholders (often a combination of public and private actors) are involved, each one with specific responsibilities for the risk mitigation challenge. This thwarts the design and implementation of balanced risk mitigation approaches that are both effective and efficient. Actually, we identify here a huge research topic for the near future, since, up to our knowledge, so far little attention has been paid to the cyber risk mitigation challenge of designing and implementing balanced sets of cyber risk mitigation measures. However, we are not completely empty-handed with respect to filling in this knowledge gap. For example, to implement the often-heard adage of public-private partnership (PPP) in cyberspace, models from institutional economics are of relevance. The first one, a visualization of which is given in Figure 7 (left), relates to the ‘institutional design for complex technological systems’ in such a way that ‘socially desired objectives are realized’ (Koppenjan and Groenewegen, 2005). It discusses ‘arrangements between actors that regulate their relations: tasks, responsibilities, allocation of costs, benefits and risks’, so we argue that this theory can be very helpful for institutional design around securing cyberspace. A second model refers to social contract theory (Bierens et al., 2017). In this paper it is argued that also for cyberspace (next to existing societal domains), an appropriate social contract has to be fixed where it discriminates between a direct social contract (between citizens and the government) and indirect social contract (between citizens and the government but via private organizations), a visualization of which is given in Figure 7 (right).



**Figure 7:** Two mental models related to the fifth step of the basic cyber risk management cycle: the four-layer diagram of institutionalization (left) and the direct and indirect social contract model (right)

Finally we consider step six of the basic cyber risk management cycle, where we have to take care of monitoring the effectiveness of the measures taken. We argue here that this strongly relates to the creation of cyber situation awareness discussed above. Measuring the effectiveness of measures taken is certainly not an easy task but through smart monitoring of cyber activities in the socio-technical layer and of IT-processes in the technical layer, we may create the necessary insights. And we are aware of certain examples. Currently, it has become commonplace to monitor and analyze real-time all kinds of financial transactions in attempts to prevent fraudulent ones, among other issues. In addition, national intelligence services are busy in monitoring cyber activities of foreign state (or state-sponsored) actors related to, for example, espionage or the distribution of fake news, and the police monitors the dark web on all kinds of illegal activities. And finally, researchers are monitoring and analyzing all kinds of Internet traffic to, for example, discover new threats based on new upcoming attack tools. Since such monitoring activities are strongly related to the creation of cyber situation awareness discussed above, we see no need to introduce an additional mental model here.

### 3.3 Three key challenges for the near future

Reconsidering the analysis results provided so far, we finalize this section by mentioning three key challenges for the near future in an attempt to arrive at adequate cyber security levels:

- *1. Creation of Cyber Situation Awareness:* although some progress has been made, the state of the art of understanding what happens in cyberspace is still insufficient. It is often heard that states, organizations and individuals have limited insights on the (in)correctness of relevant cyber activities and, as a consequence, the related cyber risks. Cyber Situation Awareness is crucial for understanding cyber risks, for measuring the effectiveness of cyber risk mitigation measures, as well as for dealing with fundamental dilemma's like between cyber opportunities and negative cyber risks (discussed above), and between privacy and cyber security. To illustrate the latter: if cyber risks are very high (e.g., societal disruptive), people tend to be more willing to accept privacy limitations to help law enforcement agencies to catch the perpetrators.
- *2. Methodologies for a Arriving at Balanced Sets of Cyber Risk Mitigation Measures:* being aware of the possibility to take all kinds of cyber risk mitigation measures in both the technical, socio-technical and governance layer of cyberspace, methodologies (or even a set of best practices) for selecting an balanced set of those measures that is both efficient and effective do not exist. This is considered to be an important research challenge for the near future.
- *3. Implementing Public-Private Partnerships (PPPs) for Securing Cyberspace:* although we observe a growing amount of emerging initiatives of increased cooperation in all kinds of cyber sub-domains, one might say that – compared to PPP implementations in physical world (land, water, air and space) domains – the development of those for securing cyberspace is still in its infancy. This may be caused by the enormous complexity of cyberspace with almost 4 billion people connected via the world-wide Internet on the one hand, and a few big players on the other (the famous 'Big Five' tech companies). However, governments

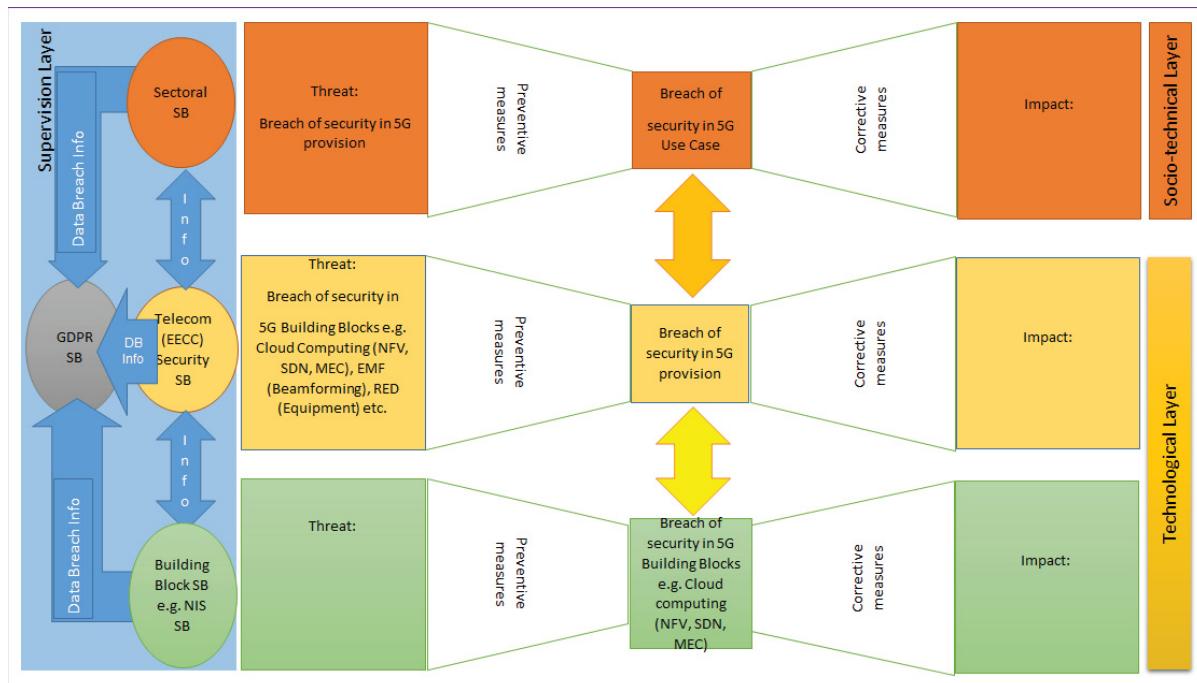
can not escape from their responsibility, as part of their social contract with their citizens, to take together the lead in creating public-private partnerships in the benefit of a more secure (fifth domain of) cyberspace. Researchers can support this development by suggesting suitable looking relation arrangements between relevant cyberspace stakeholders.

#### 4. Use of mental models

As already mentioned above, the set of mental models introduced in this paper has been collected based on discussions and research collaboration with cyber security professionals while following an executive master's program (Cyber Security Academy, 2014). In attempt to validate the choice of models presented, we here briefly review the use of a few mental models during the execution of the research of some students when composing their final thesis. The cyberspace model of three layers has been often applied, for example, to structure the results of an analysis of the security of eHealth services of Dutch General Practitioners (Willems, 2017), or to structure a large set of requirements for designing a multi-stakeholder roadmap for implementing consumer vulnerability management (Bastiaanse, 2018).

In another thesis, the model of Lessig on cyber regulations modalities was applied to analyze policy strategies to make smartphone Virtual Private Networks (VPNs) available for consumers (Ghaoui, 2017).

As a final example, we mention here a thesis written on the design of a model for cyber security supervision of 5G in the Netherlands (Wazir, 2019). In this thesis, both the bowtie model and the three-layer model of cyberspace have been applied and have been integrated in one conceptual supervision model, which is shown in Figure 8. For more details, we refer to the related Master's theses, most of which are already available online.



**Figure 8:** The Triple Bow Tie Cyber Security Supervision model with the 5G use case layer, the 5G provisioning layer and the 5G building blocks layer (Wazir, 2019).

#### 5. Conclusions and future research

In this paper, a set of essential mental models have been presented that together cover the key elements of cyberspace and those of securing cyberspace, i.e., cyber security. They provide an overview on what cyberspace entails and what the related cyber security challenges encompass. In day-to-day discussions between cyber security professionals and during the execution of cyber security research, these models and the related frameworks of thinking have turned out to be useful and effective. The choice of making an explicit distinction between IT (the enabling technology of cyberspace in layer one) and the use of it (in terms of cyber activities in layer two) is crucial for understanding the difference between classical information (or IT) security and cyber

security. Cyber security is the new notion and concerns the challenge of sufficiently securing our cyber activities, the things we *do* using modern IT. The consequence of this framing is that everybody can now understand what cyber security is about since cyber activities are human-based and avoid the one-sided focus on (for many difficult to grasp underlying) IT. The second model with underling mental models emphasizes that cyber security is essentially a cyber risk management challenge. Here, risk is being considered as a two-side notion of being both an opportunity (positive risk) and a potential loss (negative risk). This invites for discussions on assessing at the same time potential advantages and disadvantages of (further) digitization. In this way, cyber security is not just only a cost factor but also a factor of potential profit, which, for sure, facilitates discussions on cyber security at board level. Finally, it is argued that, in one way of another, the design and implementation of the basic cyber risk management cycle is crucial for trying to create a sufficiently secure cyber environment in the cyber sub-domain at stake.

During our research journey we encountered three key challenges related to the implementation of the cyber risk management cycle that need soon further investigation, namely, (i) the creation of well-established cyber situation awareness in all kinds of cyber sub-domains, (ii) the design of methodologies for arriving at a balanced set of cyber risk mitigation measures in all cyber sub-domains, and (iii) the implementation of public-private partnerships in cyberspace.

## **References**

- Bastiaanse, H. (2018), "Multi-stakeholder roadmap for implementing consumer vulnerability management", *Master's thesis*, Cyber Security Academy, The Hague.
- Bierens, R., Klievink, B., and Van den Berg, J. (2017), "A social cyber contract theory model for understanding national cyber strategies". In Jansen, M. et al., editor, Proceedings of IFIP EGOV-EPART 2017 Conference (EGOVEPART2017), volume 10428 of *Lecture Notes in Computer Science*, pp 166 - 176, St Petersburg.
- Cyber Security Academy (2014), "About the CSA", <https://www.csacademy.nl/en/about-csa> (last access: 2019-02-03).
- Dorfman, M.S. (2007), *Introduction to Risk Management and Insurance* (9<sup>th</sup> ed.), Prentice Hall, Englewood Cliffs.
- Endsley, M. and Jones, D. (2016), *Designing for Situation Awareness* (Second ed.), CRC Press.
- Fredriksen, G. (2018), "Protecting the Crown Jewels", Forbes, <https://www.forbes.com/sites/forbestechcouncil/2018/08/13/protecting-the-crown-jewels/#3eb2ba30a5a9> (last access: 2019-02-17)
- Ghaoui, N. (2017), "Policy strategies for VPN for consumers in the Netherlands", *Master's thesis*, Cyber Security Academy, The Hague.
- Halo Business Intelligence (2009), "People Process Technology, The Golden Triangle Explained", *white paper*, <https://halobi.com/blog/people-process-technology-the-golden-triangle-explained/> (last access: 2019-02-03)
- ISACA (2016), "The Merging of Cyber Security and Operational Technology", *white paper*, ISACA.
- ISO/TC 262 (2018), ISO 31000:2018, *Risk Management, Guidelines*, ISO, Geneva, Switzerland.
- ISO/IEC JTC 1 (2018), ISO/IEC 27000:2018, *Information Technology - Security Techniques - Information Security Management Systems - Overview and Vocabulary*, ISO, Geneva, Switzerland.
- ISO/IEC JTC 1 (2005), ISO/IEC 27001:2005, *Information Technology - Security Techniques - Information Security Management Systems - Requirements*, ISO, Geneva, Switzerland.
- ISO/IEC JTC 1 (2012), ISO/IEC 27032:2012, *Information Technology - Security Techniques - Guidelines for Cybersecurity*, ISO, Geneva, Switzerland.
- ISO/IEC TC 262 (2009), ISO/IEC 31010:2009, *Risk Management - Risk Assessment Techniques*, ISO, Geneva, Switzerland.
- Jacobs, F.R. and Weston Junior, F.C.T. (2007), "Enterprise resource planning (ERP) - A brief history", *Journal of Operations Management*, vol. 25, No. 7, pp 357 - 363.
- Koppenjan, J. and Groenewegen, J. (2005), "Institutional design for complex technological systems", *Technology, Policy and Management*, vol. 5, No. 3, pp 240 - 257.
- Lessig, L. (1999), *Code and Other Laws of Cyberspace*, Basic Books, New York.
- Pols, P. (2018), "The Unified Kill Chain, Designing a Unified Kill Chain for analyzing, comparing and defending against cyber attacks", *Master's thesis*, Cyber Security Academy, The Hague.
- Van den Berg, J. et al., (2014), "On (the Emergence of) Cyber Security Science and its Challenges for Cyber Security Education", Proceedings of the NATO IST-122 Cyber Security Science and Engineering Symposium, Tallinn, Estonia, October 13–14. (Winner of the Best Paper Award).
- Van den Berg, J. (2018), "Cyber Security for Everyone". In Stefanie Frey and Michael Bartsch, editors, *Cyber Security Best Practices*, pp 571 – 583, Springer Vieweg, Wiesbaden.
- Wazir, F. (2019), "Can NL trust 5G? A conceptual model for cyber security supervision of 5G in the Netherlands", *Master's thesis*, Cyber Security Academy, The Hague.
- Willems, D. (2017), "Caring for security: an analysis of the security of eHealth services of Dutch General Practitioners", *Master's thesis*, Cyber Security Academy, The Hague.
- Zipp, A. (2015), "BowTieXP, Bowtie Methodology Manual", *white paper*, preliminary version, IP Bank.

# A Legal Perspective of the Cyber Security Dilemma

Brett van Niekerk and Trishana Ramluckan  
University of KwaZulu-Natal, Durban, South Africa  
[vanniekerkb@ukzn.ac.za](mailto:vanniekerkb@ukzn.ac.za)  
[ramluckant@ukzn.ac.za](mailto:ramluckant@ukzn.ac.za)

**Abstract:** The cyber security dilemma is an evolution of the international relations theory of the security dilemma (also known as the spiral model). In the traditional model, a nation attempting to increase its security inadvertently decreases the security of one or more nations. Even though the nations act defensively and do not seek conflict, mistrust results in the situation escalating towards conflict. Cyber operations has altered the status quo of national security, and has resulted in the investigation of the security dilemma applied to cyber-space. Unlike the traditional security dilemma, escalation in cyber-space and any potential online conflict may remain invisible to many outside of the cyber security and national security communities. Most studies focus on the feasibility or existence of the cyber security dilemma and related mechanisms in cyber conflict, but there has been no discussion on the legal perspectives surrounding these aspects. The studies on the cyber security dilemma are limited to the case of conflict only in cyber-space, and do not take into account a dilemma including both physical and cyber elements, nor a clear inclusion of cyber-influence operations. The aim of the paper is to fill the gap in knowledge relating the legal perspectives of the cyber security dilemma by discussing the application of international law to concepts encompassed by the cyber security dilemma and its relation to cyber-influence. A qualitative methodology will be followed, using document analysis of various discussions on the applicability of international law to cyber-conflict. In particular, the Tallinn Manuals, which are studies on the intersection of conflict in cyber-space and international law, will be used to form the basis of the analysis and legal discussion. In addition to a cyber-only conflict, considerations for scenarios where there is escalation to economic sanction or physical conflict need to be discussed.

**Keywords:** cyber conflict, cyber law, cyber operations, cyber security dilemma, international relations

---

## 1. Introduction

Cyber-space is the new domain that nations attempt to project their power and influence: “coercive cyber capabilities are becoming a new instrument of state power, as countries seek to strengthen national security and exercise political influence. Military capabilities are being upgraded to monitor the constantly changing cyber domain and to launch, and to defend against, cyber attacks” (International Institute for Strategic Studies (IISS), 2014: 19). Whilst there have been numerous high-profile incidents involving possible state-backed cyber-operations, traditional models of national power, international relations, conflict, and international law have yet been adequately adapted to this new domain of operations.

The security dilemma is a model whereby nations seeking to avoid conflict by increasing their national security inadvertently cause an escalation and ultimately conflict (Herz 1950; Jervis 1978). Rueter (2011), Libicki (2016), and Buchanan (2017) consider a variation of this model for cyber-space, known as the cyber security dilemma. These deliberations focus primarily on the possibility of a cyber security dilemma occurring, and there is limited discussion on the perspectives of international law.

Currently international law is struggling to translate existing laws and conventions into cyber-space to tackle the changes in modern conflict; even the concept of national sovereignty in cyber-space is being challenged. This has implications for the cyber security dilemma as there may be disproportionate responses to perceived threat online. This paper aims to fill this gap by providing a legal discussion on aspects on the cyber security dilemma. The research presented here is conducted via deskwork, where the prevalent commentaries on international law and cyber security are applied to key concepts and challenges emanating from the cyber security dilemma model.

Section 2 presents a background to international law and cyber security, introducing some of the major legal texts that will be used to discuss the cyber security dilemma. Section 3 provides an overview of the Cyber Security Dilemma, which is followed by the legal analysis in Section 4. The paper is concluded in Section 5.

## 2. International law and cyber security

Darnton (2005: 1) indicates that “there is a requirement in international law for signatories to the Geneva Conventions to assess new forms of warfare in terms of lawfulness.” However, over ten years later there are still challenges evident in translating international humanitarian law to cyber-space. Darton’s (2005) study focuses

on a broader concept on information warfare; however, the challenges such as proportionality, discrimination, neutrality, and the definition of armed conflict when dealing with a non-physical situation are relevant. Rowe (2007) considers the specific case of potential war crimes arising from the use of cyber-attacks, in particular the used of deception and secrecy and the issues of collateral damage. Rowe's suggested responses include boycotts and reparations against the offending nation, and the possibility of blockades against their networks; however, he notes that network forensics in a case of a possible war-crime will need to be more thorough than for cyber-crime.

Henry, Stange, and Trias (2010) raise the issue of defining an act of war in cyber-space; they propose that a state-sponsored operation that results in damage to critical infrastructure and the significant loss (or threat of loss) of human life be considered as acts of war in cyber-space. Nitu (2011) also raises issues in the application of international law, and suggests that the law needs to change in order to address the rapid advancements in technology.

Melzer (2011) and Schmitt (2013; 2017) provide discussion on the application of international law to cyber-security and cyber-operations. A necessary focus of these documents is on the United Nations Charter, in particular article 2(4), and the concepts of *jus ad bellum* (the body of law that governs the right to go to war or use force in international relations) and *jus in bello* (governing the way in which war is conducted, i.e. 'just' war). These documents are not legally binding; however, through their discussion they provide a guide on the relevance of international laws and how to apply them.

The Budapest Convention on Cybercrime (Council of Europe, 2001) is a binding treaty that allows for mutual assistance in investigating cyber-crimes, and mechanisms to obtain required information from participating nations. The document, however, does not specifically consider state-sponsored cyber-attacks in any detail. The document however is old, and whilst it is broad in its scope, may find challenges in the applicability of modern cyber-security. As this is a treaty, it is only binding to nations are signatories, limiting the impact of the document.

The Singapore Norm Package is a set of six norms created in an attempt to preserve the trust and stability of the Internet (Global Commission on the Stability of Cyberspace (GCSC), 2018):

- Norm to Avoid Tampering,
- Norm Against Commandeering of ICT Devices into Botnets,
- Norm for States to Create a Vulnerability Equities Process,
- Norm to Reduce and Mitigate Significant Vulnerabilities, Norm on Basic Cyber Hygiene as Foundational Defence, and
- Norm Against Offensive Cyber Operations by Non-State Actors.

Whilst this is not a binding legal document, it provides useful guidelines of expected behaviour at a national and international level to aid in decision-making.

The Paris Call for Trust and Security in Cyberspace (2018) is voluntary agreement by state and non-state actors to mitigate malicious activity on the Internet. It specifically mentions and supports the United Nations Charter and the Budapest Convention on Cybercrime.

Shackelford (2017) raises a concern that many attempts considering international law to address cyber security is that the international humanitarian law is being applied "below the threshold" (p. 3). This however does not consider the challenge that the threshold may very well be changing, and the traditional concepts of physical force may no longer be sufficient. Darnton (2005) indicates there could be a case of propaganda that does not directly cause material damage; however, the impacts to a country's population could be severe. These challenges support Nitu's (2011) proposal that the laws themselves need to evolve. Rowe, Garfinkel, Beverly, and Yannakogeorgos (2011) raise challenges with monitoring cyber-weapons in terms of arms control. This will be relevant to all treaties and international legal requirements relating to cyber security, as these need to be monitored for them to be effective. Without the ability to effectively monitor compliance to international cyber security norms, these will fail and result in mistrust.

### 3. The cyber security dilemma

#### 3.1 The security dilemma

The security dilemma, also known as the spiral model, is an international relations theory that considers the scenario of nations-states taking measures to increase their security and avoid conflict, but inadvertently escalate tensions to a point of conflict (Herz 1950; Jervis 1978). Some theorists only consider the dilemma to exist should both nations have benign intentions, and that the dilemma will not exist if there is aggressive intent by one or both nations (Rueter, 2011).

The inadvertent escalation is considered to occur due to mutual fear and mistrust between the two nations, therefore they make worst-case assumptions with regards to their adversary's intentions (Wendt, 1992). However, some theorists consider the escalation and outbreak of conflict a result of anarchy in international relations (Herz, 1950). An illustration of the mistrust is the English views of the German naval build-up in 1908, which the German ambassador reported was viewed as either dangerous or as an intent to attack the UK (Kydd, 1997).

The theory of defensive realism indicates that clear signalling and/or communication of intent could mitigate the dilemma; however, offensive realism takes the view of nations attempting to increase their national power and thus resulting in conflict (Rueter, 2011). The implementation of the 'hotline' between Moscow and Washington, D.C. during the cold war, and clear communication of US and Soviet troop movements in the Middle East to avoid accidental attacks are examples, amongst others, of how communication can prevent the dilemma (Buchanan, 2017).

A measure of adversary capabilities is known as the offence-defence balance, which is defined by Glaser and Kaufmann (1998: 46) as the "cost of the forces that the attacker requires to take territory to the cost of the defender's forces." The offence-defence balance is related to the security dilemma by Jervis (1978) as shown in Table 1. The advantage in the offence-defence balance can lie with either aggressive actions or defensive actions, and the ability by the adversaries to distinguish between offensive and defensive actions results in four possible states. The security dilemma is defined in state 2, where the defence has the advantage, but the posture is indistinguishable from offensive postures. In this state, countries are likely to expand their defences, but this is likely to be misunderstood by the adversary as a sign of coming aggression. State 4 is likely to be peaceful, where countries can build their defences and this is clearly distinguishable; State 1 is the opposite, where aggression is likely and countries cannot distinguish between aggressive and defensive actions, increasing the likelihood of conflict. In state 3, the offense has the advantage implying aggressive intent and easily distinguishable actions; therefore, the dilemma cannot exist.

**Table 1:** Four worlds (Jervis, 1978)

		Advantage	
		Offence	Defence
Offense-defence posture	Indistinguishable	1. Doubly dangerous	2. Security dilemma, but security requirements may be compatible
	Distinguishable	3. No security dilemma, but aggression possible. Status-quo states can follow different policy than aggressors. Warning given.	4. Doubly stable

Often a status quo is developed in international relations, however the introduction of new technologies can affect this balance, resulting in a scenario where a security dilemma occurs. An example of this is the introduction of the aircraft carrier which then made the Washington treaty obsolete as it limited the number of size of battleships and the number the countries could possess (Buchanan, 2017). Nuclear weapons also altered the power-balance and the strategy of deterrence and mutually aided destruction, followed by nuclear weapons treaties, providing some balance. In contemporary times, cyber-operations are the main disruptive technology, which will be discussed in terms of the cyber security dilemma in the next section.

### 3.2 The cyber security dilemma

Three authors have considered the application of the security dilemma to cyber-space: Buchanan (2017), Libicki (2016) and Rueter (2011). An in-depth discussion of the cyber security dilemma is provided by Buchanan (2017), who proposes three key considerations:

- A platform is required to enable future cyber operations, therefore nations need to prepare in advance should they wish to have the option of conducting cyber operations.
- There are legitimate defensive reasons for nations to conduct network intrusion into the networks of other nations.
- Penetrations into critical, sensitive, or strategic networks will be considered threatening by the victim.

A view of the cyber security dilemma is the attempts by a state to improve its digital infrastructure security, thereby decreasing the cyber security of another nation (Rueter, 2011). This is a direct translation of the security dilemma into the cyber security dilemma. Due to the similarities of techniques used in cyber operations, Rueter (2011) contends that it is difficult to differentiate between aggressive and benign operations. There is therefore a high likelihood of misinterpretation of intent; this is an indicator of the dilemma occurring as is illustrated in Table 1. Contrary to Jervis (1978), Rueter considers the offensive advantage of cyber-attacks as an indicator of the dilemma, whereas in kinetic warfare the dilemma occurs when the defence has the advantage.

Libicki (2016) contends that the cyber security dilemma may not be as prevalent as one would think. Unlike Buchanan (2017) and Rueter (2011), Libicki (2016) considers there to be a clear distinction between offensive cyber operations and defensive cyber operations, except for the case of active defence. From an economics perspective, the vendors are usually the ones that need to improve products in order to mitigate cyber-attacks, however a nation improving its cyber defence is likely to benefit other nations (Libikci, 2016).

The challenge of attribution provides a counter-argument to the cyber security dilemma, where it cannot exist between two nations if the source of the intrusion cannot be determined. The victim will however mostly likely suspect possible perpetrators, and the general fear will still result in the destabilising situation (Buchanan, 2017). The possibility of deception by a third country to intentionally increase tensions between states is therefore possible. The offense-defence balance is discussed for cyber-space by Slayton (2017), where the main criticism is that the traditional costs does not work for cyber-security. A stronger cyber-power in theory would not be concerned about retaliation by a weaker state and should be able to prevent attribution on its cyber-operations. The stronger cyber-powers should also be more able to determine attribution for aggressors against them. However, stronger cyber-powers are likely to be more reliant on technologies than the weaker nations, and are therefore more susceptible to disruption, with any in-kind response being less effective against weaker and less reliant states (Buchanan, 2017). This asymmetric nature may allow weaker nations to create disproportionate impacts against stronger countries, who may have to resort to other means in response.

## 4. Legal analysis of the cyber security dilemma

Concepts in the above discussion of the cyber security dilemma raise a number of questions or points of contention:

- States have legitimate defensive reasons for network intrusions
- How certain does attribution need to be prior to response?
- What is an appropriate or proportionate response?
- At what point is it appropriate to escalate?
- Legality and ethics of false flag operations.

The *Singapore Norm Package* allows for the fact that nations may need to conduct network intrusions for law enforcement or intelligence purposes (GCSC, 2018); this concurs with Buchanan's (2017) view. According to Melzer (2011: 28):

*Cyber operations causing neither death, injury or destruction, nor military harm, on the other hand, such as those conducted for the purposes of general intelligence gathering (no direct causation of harm), for purely criminal purposes or otherwise unrelated to the hostilities (no belligerent nexus), would fall short of the concept of "hostilities" and, thus, would not be governed by [international*

*humanitarian law] IHL on the conduct of hostilities and, if conducted by civilians, would not entail loss of protection against direct attacks.*

This again explicitly indicates that intelligence collection is not governed by international humanitarian law. According to Schmitt (2017), there is a need to assess if the network intrusion can be considered as a breach of sovereignty of the victim; this may be considered if a governmental function or national infrastructure is affected. It is also noted that unless a cyber-crime can be attributed to a nation, then it does not violate the sovereignty of the victim nation. This could present challenges as in cyber-operations a proxy could be hired (e.g. a cyber-criminal on the Dark web) to conduct the operations and provide plausible deniability to the nation, creating further difficulties in attribution. With reference to point 1 above, if there are legitimate defensive reasons such as intelligence collection, then network intrusions are at some level permissible, provided they do not affect the sovereignty of another state.

In an international context, the use or threat of force against a country needs to be legally attributable to the perpetrator country, this includes when non-state actors act in an authorised capacity on behalf of the country (Melzer, 2011). Due to the common use of proxies in cyber-operations, the operation itself and the aggressive intent needs to be attributable to the suspected perpetrating or sponsoring nation (Melzer, 2011). This aspect of intent is discussed in Section 3.2. Whilst it is indicated that attribution and evidence of intent is required, in Melzer (2011) it is not specified how strong the attribution needs to be, and the difficulty due to common cyber-operations techniques is acknowledged. Schmitt (2017) concurs that there needs to be evidence of attribution as well as the intention of the nation, particularly when a non-state actor is used as a proxy, however provides more details as to when a third-party can be considered as acting on a nation's behalf. The challenge of providing attribution again highlighted by Schmitt (2017: 91):

*Accordingly, the mere fact that a cyber operation has been launched or otherwise originates from governmental cyber infrastructure, or that malware used against hacked cyber infrastructure is designed to 'report back' to another State's governmental cyber infrastructure, is usually insufficient evidence for attributing the operation to that State. That said, such usage can serve as an indication that the State in question may be associated with the operation.*

This difficulty in attribution may need to be addressed on a case-by-case basis (Schmitt, 2017); however, it is possible that the victim is able to use their own cyber capability to obtain evidence of attribution. As mentioned above, the need to assess intent may become more important should there be an inadvertent impact on critical infrastructure that has a significant economic or social impact or results in loss of life. An accidental impact or the fear of interference on an electoral process will also require the determination of intent, as the *Paris Call for Trust and Security in Cyberspace* (2018) is intended to "strengthen our capacity to prevent malign interference by foreign actors aimed at undermining electoral processes through malicious cyber activities." Should the intent of the operation be benign, then it is not prohibited, however should the operation's presence be detected, there could be suspicion of electoral interference. This follows the mistrust between nations in such scenarios that can result in a destabilising situation.

Points 3 and 4 can be considered together as they both rely on the discussion on whether the cyber-operation can be considered as an act of force, and there being clear evidence that the cyber-operation resulted in the impact being claimed as an act of force. Should the exact nature of the impact be unclear and/or limited, there is likely little that can be done other than accusations, complaints, and sanctions. Should there be significant damage hindering the functioning of the nation and its society, or loss of life, then the use of force in response may be permissible (Melzer, 2011). The concept of proportionality in self-defence by the victim state is important when considering a response to a network intrusion and its impacts (Schmitt, 2017). This proportionality needs to be dealt with on a case-by-case basis. There is also a difference amongst the concepts of proportionality when considering countermeasures, a response to an act of force under *jus ad bellum*, and the conducting of cyber-operations under *jus in bello* (Schmitt, 2017). Ultimately, "a use of force involving cyber operations undertaken by a State in the exercise of its right of self-defence must be necessary and proportionate" (Schmitt, 2017: 348). In this regard, the responses by a nation suffering from a network intrusion is limited, as is the right to intentionally escalate.

A challenge may arise when a cyber-operation accidentally impacts on critical systems; industrial control systems are sensitive and with cyber-operations there is a risk of unintended consequences or digital collateral damage. In this instance, Rowe's (2007) proposal of reparations may be a suitable route. The responsible nation may not

wish to publicly admit to the operation or possessing specific cyber-capabilities, but may be willing to pay reparations to the victim nation to prevent the possibility of public attribution.

According to Melzer (2011), deception is prohibited when it feigns the intent to negotiate, injury, protected status or civilian status, and results in the killing, injury or capture of an adversary. Therefore tactical deception in cyber-operations in order to achieve the disruption of communications or infrastructure is not explicitly prohibited. This indicates that the use of civilian or cyber-criminal elements to conduct cyber-operations may not be explicitly prohibited should this be limited to disruption in cyber-space and no real-world casualties occur. The use of civilian actors for conducting cyber-operations on behalf of the state indicates that they therefore become agents of the state (Melzer, 2011). With specific reference to point 5 above, the use of adversary uniforms is prohibited, which then equates to false flag operations with the intent of making it appear the cyber-attack came from a third-party (Melzer, 2011). The challenge lies in proving that there is deception in the first place, and also in providing sufficient attribution to the actual perpetrator. This will therefore necessitate intelligence-gathering operations, which then refers back to point 1. The policing of such deception and false flags is therefore problematic, which may give nations the confidence of attempting such operations.

The texts considered indicate an allowance to conduct intelligence operations that may include intrusions into another nation's networks. Should the 'victim' nation feel threatened or consider the intrusion as a breach of their sovereignty, then they would need to prove attribution of the suspected perpetrators and the aggressive intent. Intelligence would be required to provide attribution and information on intent, where the victim country may employ network intrusions of its own to investigate suspected perpetrators. Suspected but innocent states may then detect and need to investigate these intrusions, resulting in a cascading and destabilising situation. Whilst there is a requirement for necessity and proportionality in responses and restrictions on escalation, it is possible that a nation will consider a preventative cyber-attack to stop what it considers continuous cyber-operations into its networks. With the uncertain applications of international humanitarian law, and the need for attribution but difficulty in obtaining it that is not adequately addressed by the existing laws, it can be considered that the current international humanitarian law may contribute to the occurrence of a cyber-security dilemma. Nitu (2011) suggests that the laws may need to be changed; as there is still uncertainty regarding the applicability and/or suitability of the current international law, there may need to be a change. However, having countries agree to these laws and signing the treaties is still proving to be difficult.

Generic responsibilities for nations to provide cyber security are outlined in the *Singapore Norms Package* (GCSC, 2018) and a more detailed view in Schmitt (2017). Following Libicki's (2016) view that countries improving their national cyber security posture will be beneficial to all other countries, if there is a concerted effort to follow such norms it will be more difficult in general to conduct cyber-operations, acting as a deterrent to unnecessary network intrusions and thus mitigating the cyber security dilemma to some extent.

## 5. Conclusion

With the increasing prevalence of cyber-operations as a projection of national power, there is a move towards applying traditional international relations models of conflict to cyber-space. There has been increasing attention to the possibility of a cyber security dilemma, where cyber-operations with defensive intent are misinterpreted and result in destabilising situations. The majority of discussions on this topic focussed on the feasibility of the dilemma, however there has been limited considerations from an international law perspective. A series of commentaries on international law and cyber security are considered to provide a legal perspective to the cyber security dilemma. Network intrusions that are for defensive purposes are permissible; however there can be limitations to their scope. The response to such intrusions requires attribution and a determination of intent, which may prove to be difficult. Any response needs to be proportionate to the original cyber-operation. The use of non-state actors to conduct cyber-operations on behalf of the state is not prohibited, however deception in intentionally making the attack appear to be attributable to a third nation is prohibited. The challenges become in policing such deception, as this is commonly used in cyber-operations. Due to the uncertainty in the application of the laws, and the particular need for attribution and assessing intent resulting in additional intelligence gathering operations and network intrusions, there can be a situation of increasing instability. Therefore the current standing of international law with regards to cyber-operations could actually facilitate the creation of a cyber security dilemma.

## **Acknowledgements**

The first author received funding by the South African National Research Foundation, Grant no. 115059.

## **References**

- Buchanan, B. (2017) *The Cyber Security Dilemma: Hacking, Trust and Fear between Nations*. Oxford University Press: Oxford.
- Council of Europe. (2001) Convention on Cybercrime, Budapest, European Treaty Series - No. 185.
- Darnton, G. (2005) "Information Warfare, Revolutions in Military Affairs, and International Law," *Journal of Information Warfare* 4(1), pp. 1-20.
- Glaser, C.L., and Kaufmann, C. (1998) "What is the Offense-Defense Balance and Can We Measure it?" *International Security*, 22(4), Spring, pp. 44-82.
- Global Commission on the Stability of Cyberspace. (2018) Norm Package Singapore, November, [online], accessed 30 January 2019, <https://cyberstability.org/wp-content/uploads/2018/11/GCSC-Singapore-Norm-Package-3MB.pdf>.
- Henry, W.C., Stange, J.M., and Trias, E.D. (2010) "Pearl Harbor 2.0: When Cyber-Acts Lead to the Battlefield," *Journal of Information Warfare* 9(2), pp. 45-55.
- Herz, J.H. (1950) "Idealist Internationalism and the Security Dilemma", *World Politics*, 2(2), pp. 157-180.
- International Institute for Strategic Studies. (2014) "Chapter One: Conflict analysis and conflict trends," *The Military Balance*, 114(1), pp. 19-22.
- Jervis, R. (1978) "Cooperation under the Security Dilemma", *World Politics*, 30(2), pp. 167-214.
- Kydd, A. (1997) "Game Theory and the Spiral Model", *World Politics*, 49(3), pp. 371-400.
- Libicki, M.C. (2016) "Is There a Cybersecurity Dilemma?" *The Cyber Defense Review*, Spring, pp. 219-140.
- Melzer, N. (2011) *Cyberwarfare and International Law*, UNIDIR, [online], accessed 22 January 2019, <http://unidir.org/files/publications/pdfs/cyberwarfare-and-international-law-382.pdf>.
- Nitu, A. (2011) "International Legal Issues and Approaches Regarding Information Warfare," *Journal of Information Warfare* 10(2), pp. 48-57.
- Paris Call for Trust and Security in Cyberspace (2018), 12 November, [online], accessed 18 January 2019, [https://www.diplomatie.gouv.fr/IMG/pdf/paris\\_call\\_cyber\\_cle443433.pdf](https://www.diplomatie.gouv.fr/IMG/pdf/paris_call_cyber_cle443433.pdf)
- Rowe, N. (2007) "War Crimes from Cyber-weapons," *Journal of Information Warfare* 6(3), pp. 15-25.
- Rowe, N., Garfinkel, S.L., Beverly, R., and Yannakogeorgos, P. (2011) "Challenges in Monitoring Cyberarms Compliance," *International Journal of Cyber Warfare and Terrorism* 1(2), pp. 35-48.
- Rueter, N.C. (2011) *The Cybersecurity Dilemma*, Master's dissertation, Duke University.
- Schmitt, M.N. (2013) *Tallinn Manual on the International Law Applicable to Cyber Warfare*, Cambridge: Cambridge University Press.
- Schmitt, M.N. (2017) *Tallinn Manual 2.0: On The International Law Applicable to Cyber Operations*, Cambridge: Cambridge University Press.
- Shackelford, S.J. (2017) "The Law of Cyber Peace," *Chicago Journal of International Law* 18(1), pp. 1-48, [online], accessed 18 February 2019, <http://chicagounbound.uchicago.edu/cjil/vol18/iss1/1>.
- Slayton, R. (2017) "What Is the Cyber Offense-Defense Balance? Conceptions, Causes, and Assessment," *International Security*, Winter, 41(3), pp. 72-109.
- Wendt, A. (1992) "Anarchy is what States Make of it: The Social Construction of Power Politics", *International Organization*, 46(2), pp. 391-425.

# An Analysis of Selected Cyber Intelligence Texts

Brett van Niekerk<sup>1</sup>, Trishana Ramluckan<sup>1</sup> and Petrus Duvenage<sup>2</sup>

<sup>1</sup>University of KwaZulu-Natal, Durban, South Africa

<sup>2</sup>University of Johannesburg, Johannesburg, South Africa

[vanniekerkb@ukzn.ac.za](mailto:vanniekerkb@ukzn.ac.za)

[ramluckant@ukzn.ac.za](mailto:ramluckant@ukzn.ac.za)

[duvenage@live.co.za](mailto:duvenage@live.co.za)

**Abstract:** Cyber intelligence is a growing discipline related to cyber security, resulting in a number of whitepapers, advisories and services from cyber security companies and professional bodies. Global spending on cyber threat intelligence services has demonstrated rapid increases, yet a corresponding mitigation of advanced threats is yet to be seen. Despite this growth, there is still variation on the definition and the focus of cyber intelligence, which could account for the limited gains. Cyber intelligence is important for understanding both the threat environment and identifying what is relevant to the user's context based on a number of internal factors to the organisation or nation. This paper provides a document analysis of a number of publicly available cyber intelligence and cyber threat intelligence documents (in the form of advisories and whitepapers) using text mining, content analysis, and thematic analysis. These techniques are relevant to both qualitative research and intelligence analysis and are employed here for assessing similarities amongst the documents, identifying common themes and categories, assessing the emphasis thereof, and identifying gaps common to these documents. Gaps in the coverage of the documents are identified in that they do not consider cyber counterintelligence or the legal considerations for cyber intelligence, and cyber intelligence collection operations are only briefly mentioned. Cyber counterintelligence and cyber intelligence collection operations are sub-disciplines of cyber intelligence, alongside cyber threat intelligence. The analysis also indicates that documents on cyber intelligence and cyber threat intelligence do not align. Such occurrences and some implications thereof are discussed.

**Keywords:** qualitative analysis, cyber intelligence, cyber counterintelligence, cyber law

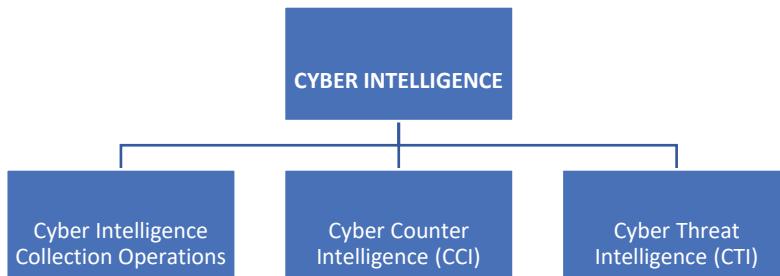
---

## 1. Introduction

The 2019 US National Intelligence Strategy is placing emphasis on cyber security (Office of the Director of National Intelligence, 2019); by extension this implies increasing focus on cyber intelligence. Cyber threat intelligence services have seen a great deal of attention and an increase in subscriptions, with a growth of 129% in four years (Duvenage & van Niekerk, 2016), however this has not yet translated into widespread effective mitigation of cyber threats. Despite the increase in the prevalence of cyber threat intelligence, there is still uncertainty surrounding the definitions and the focus of cyber intelligence. This uncertainty can negatively affect the understanding of the topic and ultimately the effectiveness of implementations, which could account for the limited gains.

It is worthwhile to provide the definition of intelligence, as this useful to distinguish amongst various concepts covered in this paper. Intelligence is a product of the collection, processing, analysis, and interpretation of information about nations, actors, threats, and operational areas (Joint Chiefs of Staff, 2013); it needs to be in context and actionable. There is a difference between cyber intelligence and cyber espionage: cyber intelligence seeks to gain information and insights about relevant cyber threats, whereas cyber espionage is the use of cyber techniques to steal secrets (Duvenage & van Niekerk, 2016).

Lee (2014a; 2014b) considers cyber intelligence to be comprised of the following sub-disciplines: cyber threat intelligence (CTI) (Lee, 2014e), cyber intelligence collection operations (Lee, 2014c) and cyber counterintelligence (CCI) (Lee, 2014d). CTI can be seen as subscription services providing information input for intelligence analysis, whereas collection operations can be considered as obtaining specific external and internal information for analysis. Counterintelligence is both provocative and defensive intelligence operations: mitigating adversary collection operations whilst seeking to learn more regarding their objectives and techniques (Lee, 2014b; 2014d). These components are depicted in Figure 1.



**Figure 1:** Components of cyber intelligence

Both intelligence and cyber security are multi-disciplinary topics. Cyber security traditionally will contain the technical subjects; however, the strategic political and legal aspects are coming to the fore. Similarly, intelligence studies have a traditional focus on politics and history, but there is a growing situation where other disciplines can contribute to intelligence studies (Marrin & Madison, 2017); particularly with the growth of computing and cyber security. In the same manner, academic research methodology can be useful in intelligence analysis for cyber security.

The Introduction continues below in Section 1.1. where an overview of related research and publications is provided. Section 2 presents the methodology, introducing the documents and the techniques used to assess them. Section 3 provides the results of the analysis of the texts, which is followed by a discussion of the results in Section 4. A discussion on the techniques used in this study and their multi-disciplinary use for cyber intelligence analysis is provided in Section 5. The paper is concluded in Section 6.

### 1.1 Related research and publications

A number of previous research outputs have considered cyber intelligence from different perspectives. One of the first works considering cyber intelligence was by Bodmer, Kilger, Carpenter, and Jones (2012), which focussed on counter intelligence principles for network security. Yucel and Koltukuz (2014) provide a list of articles for topics such as cyber espionage, open source intelligence, social media intelligence, threat and intrusion detection, and cyber weapons. Falk (2016) proposes an initial ontology describing aspects related to CTI. Sample, Cowley, Watson, and Maple (2016) investigate the possibility of cultural influences in CTI, and found possible links between national culture and activities of the threats, which could conceivably be used for attribution or deception. Duvenage, Jaquire, and von Solms (2016) consider the relationship between CCI and CTI, and categorise offensive and defensive CCI. Bellaby (2016) motivates for cyber intelligence whilst warning against misuses, and considers some legal aspects to guide conducting cyber intelligence operations. One of the gaps that has not been assessed relates to the cyber intelligence documents produced by professional bodies to guide organisations and cyber security professionals. This paper analyses eight selected texts from three professional bodies. The next section describes the texts and analysis techniques in more detail.

## 2. Methodology

Qualitative analysis of cyber intelligence whitepapers and best practices was conducted, in particular content analysis and thematic analysis of the documents. In total eight documents were selected from professional bodies based on their relationship with the topic. These texts include: ISACA Tech Brief on Threat Intelligence (2017), Centre for the Protection of National Infrastructure Threat Intelligence whitepapers by Context (2015) and MWR InfoSec (2015), and Intelligence and National Security Alliance (INSA) documents as listed below:

- Cyber Intelligence – Setting the Landscape for an Emerging Discipline (INSA, 2011);
- Operational Levels of Cyber Intelligence (INSA, 2013);
- Strategic Cyber Intelligence (INSA, 2014a);
- Operational Cyber Intelligence (INSA, 2014b);
- Tactical Cyber Intelligence (INSA, 2015).

The software NVivo was used to conduct the analysis – this provides content analysis in terms of the frequency of major words in the texts. Thematic analysis provides an indication of the prevalent themes in the texts. Document correlation and cluster analysis is also performed. This indicates relationships amongst the various themes and documents.

### **3. Analysis of cyber intelligence white papers**

Figure 2 provides a cluster analysis of the documents based on word similarity, determined using Pearson's correlation. There are two main clusters: the documents focussing on threat intelligence, and the INSA documents focussing on the operational levels of cyber intelligence. From the cluster, it can be seen that the document on tactical cyber intelligence is closest to the threat intelligence documents, followed by operational and then strategic. This indicates that the focus on the threat intelligence is tactical in nature, aligning more to the cyber-security analysts rather than aiding organisational or national decision makers in defining cyber-security strategies.



**Figure 2:** Sources clustered based on word similarity

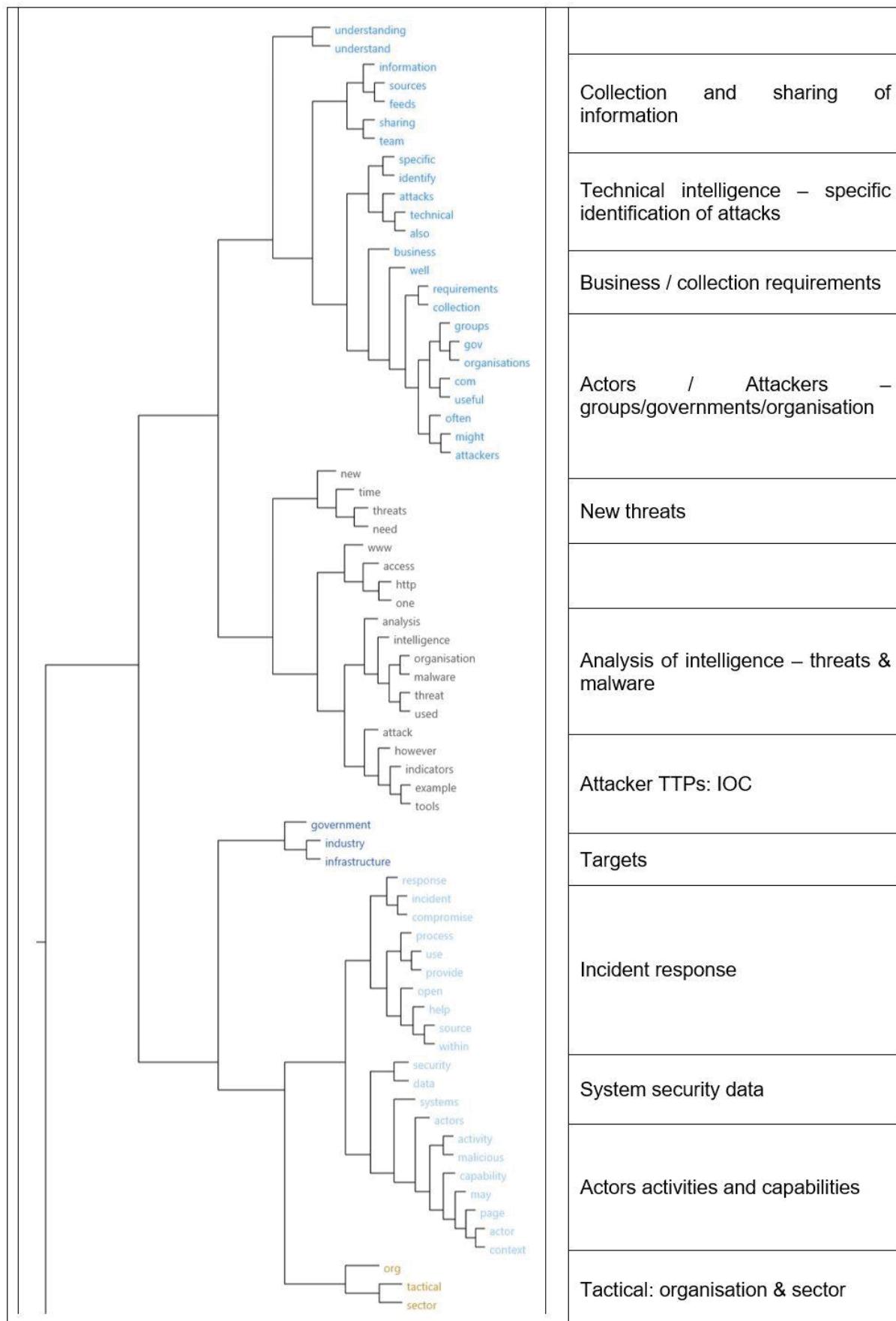
The overall word frequency is visualised in Figure 3. The larger the word, the more common it was. Despite there being more documents in cyber intelligence operational levels (five of the eight), the word ‘cyber’ was not as prevalent as the word ‘threat’. The words following this in prevalence were: ‘information’, ‘security’, ‘organisation’ and ‘network’.

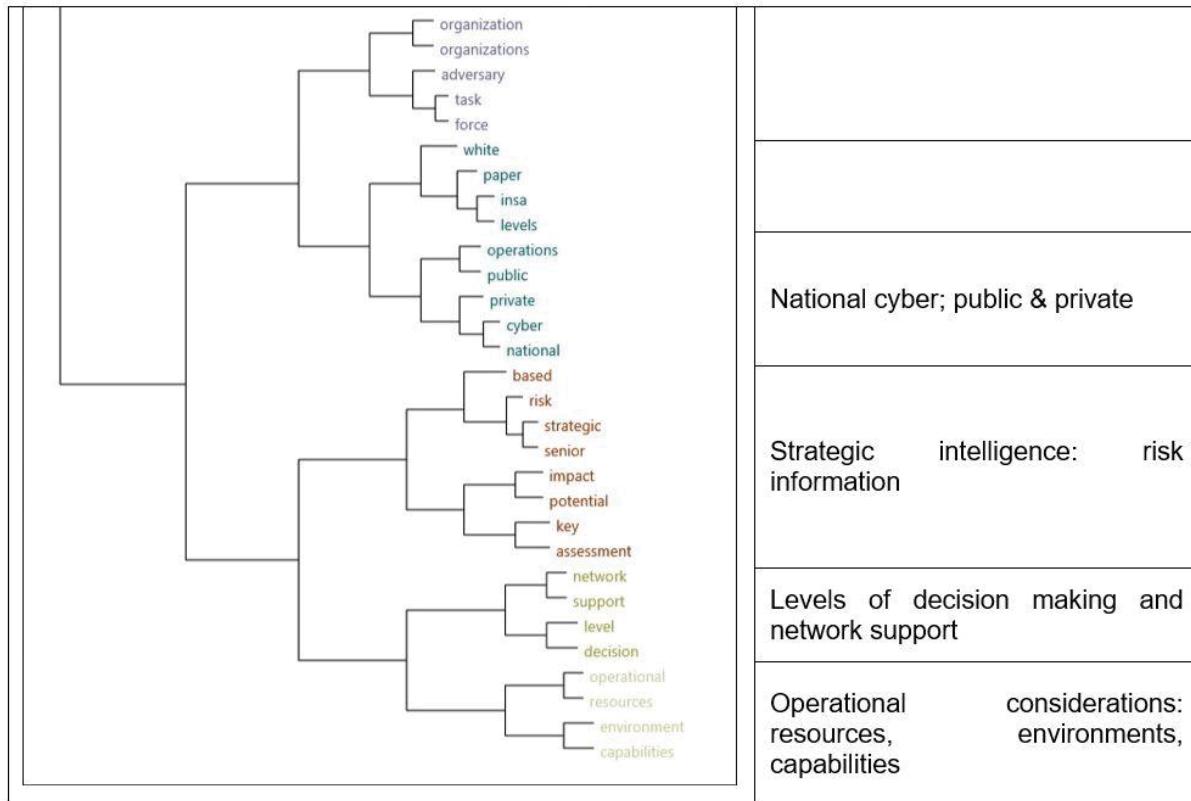


**Figure 3:** Word cloud from all texts

The content analysis determined the frequency of words. Figure 4 clusters these words to determine key themes across the texts (indicated on the right). Key themes that emerge are: collection and sharing of information; technical intelligence; business requirements; actors; new threats; intelligence analysis; attacker tactics, techniques and procedures (TTPs); targets; incident response; system security data; Actors activities and capabilities; tactical aspects in organisations sectors; national cyber consideration e.g. public vs private; strategic intelligence and risk; decision making levels; and operational considerations.

From the clustering of the words, it is implied that threat intelligence is tactical and focussed at the organisational level, whereas strategic cyber intelligence is more related to a national focus. Aspects that are noticeably absent from the word cloud and word cluster analysis are the concept of cyber intelligence collection operations, cyber counterintelligence, and any legal or governance aspects related to cyber intelligence. The linkages of analysis techniques to the usage of intelligence is also not covered in detail.





**Figure 4:** Cluster analysis of key words

#### 4. Discussion of results

The discussion is divided into two section: the first (Section 4.1) provides a discussion on the explicit content analysis, and the second (Section 4.2) provides a discussion on the gaps of content coverage that are indicated above.

##### 4.1 Discussion of document analysis

The two distinct clusters in the document correlation implied that threat intelligence and cyber intelligence are not the same; and that there is some misalignment between those considering cyber intelligence holistically and those considering threat intelligence. The tactical cyber intelligence appears closest to the threat intelligence, indicating that the focus of the threat intelligence is towards detection of possible indicators at an analyst level, and there is limited threat intelligence to aid strategic decision makers. This view concurs with that of Lee (2014e).

There appears to be a focus on CTI. As Lee(2014b) and Duvenage, Jaquire, and von Solms (2016) indicated, the focus on CTI is more due to a catch-phrase in marketing services than real understanding of the concepts. CTI as is provided cannot be considered true intelligence as it is a stream of data; it is not in context and it is limited in how actionable it is. Whilst network security devices and cyber security analysts can use CTI as an aid in detecting possible threats, the real intelligence is the analysis and interpretation of any indicators that have found matches in the network. Therefore, CTI on its own is of limited use; it needs to correlated with internal data (processing) to provide possible threat detections, which can then be analysed and interpreted to form an improved view of the threat environment.

##### 4.2 Gaps in document coverage

As is indicated in Section 3, there are gaps in the coverage of the documents. This includes the concept of cyber intelligence collection operations and cyber counterintelligence (Section 4.2.1), legal and governance considerations (Section 4.2.2), and the linkage of analysis techniques and usage (Section 4.2.3).

#### **4.2.1 Cyber intelligence collection operations and cyber counterintelligence (CCI)**

None of the documents consider cyber counterintelligence, and cyber intelligence collection is only briefly considered at a high level. As is evident from Figure 3, the term "collection" does not occur frequently. These are key constructs in cyber intelligence, and the limited coverage illustrates the gaps in professional literature.

Intelligence collection is considered at a high level, however specific techniques for collection are not covered in detail. Certain sources may be best suited for a specific level of intelligence (tactical/technical, operational, or strategic), and the collection methods for those sources may differ. The INSA (2014b) document provides the most detail for the collection processes and considerations. As CTI is essentially a data stream with a very broad context, collection operations are important to gain insights into the internal environment against which the CTI can be correlated. This is known as passive collection (Lee, 2014c). Active collection is the process of obtaining intelligence from an adversary's network (Lee, 2014c). This activity is considered unethical or illegal for non-state actors to conduct, as is outlined by the Singapore Norm Package by the Global Commission on the Stability of Cyberspace (GCSS, 2018), which will be discussed in more detail in Section 4.2.2.

It is also important to note that the network environment can influence the choice of collection techniques. For instance, industrial control networks are particularly sensitive, therefore collection operations from critical infrastructures may have unintended adverse impacts resulting in downtime of the control systems. The increase of connected 'smart' devices (Internet of Things) through the Fourth Industrial Revolution can provide challenges. These devices can increase the attack surface, and due to the quantity of the devices, there may be an overload of information, which may result in additional pre-processing to filter out. As many devices can connect in an ad-hoc or uncontrolled manner, they may be invisible to collection operations if they are not connected during the period of the collection. Therefore, insecure devices could be missed, giving an inaccurate picture of the threat exposure.

Cyber counterintelligence is an important concept as this can be seen as defending against adversary network intrusions and espionage or intelligence collection operations, whilst obtaining critical information regarding their tactics, techniques, procedures and objectives. Offensive CCI interacts with the adversary (Lee, 2014d) in order to gain an understanding of their objectives and preferred methods. This involves the use of deception such as fake online profiles to engage with the adversary, or the use of honeypots and honeynets to monitor their activity (Bodmer, Kilger, Carpenter, & Jones, 2012). The use of deception has been debated from legal and ethical viewpoints and will be discussed further in Section 4.2.2. Defensive CCI can use red teams and threat analysis techniques to identify specific areas for cyber security improvement specifically based on an organisation's threat and vulnerability landscape (Lee, 2014d). During the past four years it was especially cyber intelligence's component of cyber counterintelligence that has been gaining traction in main stream business (The Economist, 2015; Panda Security, 2018). The growing commercial prominence of cyber intelligence and its subsets raises the academic imperative to clearly delineate concepts within this emerging (academic) field.

#### **4.2.2 Legal and governance considerations**

With reference to Bodmer, Kilger, Carpenter, & Jones (2012,108), every country or organisation has the capabilities in terms of rules or governance to restrict counterintelligence activities, preventing it from invading the rights of its private citizens. However, in some countries, absolute power belongs to the State, under the pretence of representing freedom and liberty for all its citizens. Therefore, prudence is warranted as counterintelligence professionals conduct their operations. Cyberspace poses a particular challenge as online activities, spread at rapid pace from country to country, with the ability to gather information.

Bodmer, Kilger, Carpenter, & Jones (2012,410), further state that since this may be the case, a number of national and international cyber laws have been developed and are applicable in various countries, the implications of which need to be understood, regardless of the country an individual may reside in or who is hosting the IP address. While some countries may choose to remain unaffected or indifferent, countries like the US intend prosecuting those individuals (Bodmer, Kilger, Carpenter, & Jones, 2012: 410).

The South African King IV Report Section 13d provisions for the "proactive monitoring of intelligence to identify and respond to incidents," which includes any adverse social media activities as well as cyber-attacks (Institute of Directors Southern Africa (IoDSA), 2016)

The Global Commission on the Stability of Cyberspace 2018's Singapore Norm Package states that "Non-state actors should not engage in offensive cyber operations and state actors should prevent and respond to such activities if they occur." With reference to Nijman (2010), non-state actors do not have a legal personality, which results a challenge concerning the legality of intelligence collection from adversary networks by non-state actors. This creates a legal challenge when a government entity attempts to utilise the services of non-state actors as part of their intelligence operations. Since non-state actors have no legal personality within the context of International Law, the question of attribution of liability may arise. With reference to the United Nations (UN) Convention on the Law of the Sea (LOSC), companies may enter into an international contract regarding various services, but in so doing they also incur obligations and responsibility under international law. While the Convention may apply to some shared responsibility between states and international organisations, it can be argued that non-state actors, on the basis of the "contract", would be responsible for wrongful acts and thereby be in breach of the contract. This, however, still falls outside the ambit of International Law, and may just be attributable in terms of a "contractual obligation" (d'Aspremont, Nollkaemper, Plakokefalos and Ryngaert, 2015).

The next consideration is the legality of using "honeypots" for the purpose of counter intelligence. According to Spitzner (2003) there are three main issues that are commonly maintained regarding honeypots which are entrapment, privacy, and liability. By definition entrapment is "a law-enforcement officers or government agent's inducement of a person to commit a crime, by means of fraud or undue persuasion, in an attempt to later bring a criminal prosecution against that person" (Black's Law Dictionary, 7th Ed). From the definition, honeypots do not qualify as a form of entrapment, as the "attacker", usually breaks into a honeypot on their own initiative and coercion is not involved. Regarding privacy, the use of honeypots does affect privacy. The exemption under Service Provider Protection means that security technologies can gather information on people including attackers, as long as that particular technology is being used for the protection of the environment, which results in these technologies are exempt from privacy restrictions (Spitzner, 2003). Spitzner (2003) states that liability implies a person can be sued if their honeypot is used to harm others, however, it is a civil matter and not a criminal one.

There remains the need for cyber intelligence governance as cyber intelligence needs to cooperate with legal counsel pertaining to the issues (Bodmer, Kilger, Carpenter, & Jones, 2012: 169).

#### *4.2.3 Analysis techniques and usage*

Depending on the sector or organisation type, the use of cyber intelligence, and therefore the analysis techniques, may differ. The intelligence required for a private organisation will be different to that of a nation state, where the military, intelligence community, and law enforcement are producers and consumers of cyber intelligence. Non-state organisations are more likely to be intelligence consumers, where the processing may be limited to distilling the intelligence further to make it more relevant to the organisation's context. The various organs of state mentioned above will be required to provide collection, processing, analysis and interpretation within their mandates to achieve their objectives; they are therefore both producers and consumers of cyber intelligence.

As with the cyber intelligence collection operations, the INSA (2014b) document has the most detail of potential analysis techniques for cyber intelligence. The analysis techniques need to be carefully selected, as they need to be relevant to the data types and the objectives of the intelligence product. As an example, Heuer and Pherson (2015) describe seven structured analytical techniques to achieve twelve objectives in analysis. Specific cases of employing these techniques are illustrated in Beebe and Pherson (2015). As an extension to this, Section 5 discusses the research methodology used in this paper as a relevant intelligence analysis technique.

### **5. Research methodologies and intelligence analysis techniques**

This section reflects on research methodology used in this paper and its relevance to cyber intelligence analysis. The research process is very similar to that of intelligence analysis: a problem or question is identified, data is collected, process, and analysed to provide interpretations that are then compiled into an output (e.g. a dissertation, paper, or intelligence report).

The qualitative analysis using tools such as Nvivo can quickly provide insights into large bodies of text which would be very time consuming for human analysts to process and assess. Data sources could be official

documents (as is used in this paper), or forum discussions on social media or the dark web, or transcripts from speeches. A specific use case could be the assessment of foreign cyber security strategies. By clustering the documents, an analyst could gain an insight into possible influence and/or collaboration amongst nations. This could also simplify analysis of a national cyber security posture as similar nations can be identified and experience from previous analyses can be leveraged.

On its own, this analysis will not be sufficient. There still needs to be interpretation of the results, where bias in the analysis can still occur. As Heuer and Pherson (2015) point out, critical analysis has led to major intelligence mistakes; therefore, this technique needs to be followed by other analytics techniques to validate the interpretation. This concern is common both the academic research and intelligence analysis. Therefore, the structured analytic techniques advocated by intelligence analysts may prove to be very beneficial in academic research.

## **6. Conclusion**

Cyber intelligence is a rapidly growing field; however, it still suffers from being poorly defined. This paper analysed selected texts from professional bodies. The results indicate that there is a possible misalignment between the drive for cyber threat intelligence and the growth of cyber intelligence as a discipline. Aspects that were found to be omitted are cyber counterintelligence and the legal and governance aspects, with more focus on cyber intelligence collection operations and analysis techniques required. For cyber intelligence to mature as a discipline, the inclusion of these areas, and aligning all areas is imperative. Due to the increasing prevalence in industry, there is a need to ensure that academic studies in these areas grow. The qualitative research methods employed in this paper may prove to be useful in cyber intelligence analysis, and intelligence analysis techniques can be used to improve academic research.

## **Acknowledgements**

The first author received funding by the South African National Research Foundation, Grant no. 115059.

## **References**

- Beebe, S.M., and Pherson, R.H. (2015) Cases in Intelligence Analysis: Structured Analytic Techniques in Action, 2<sup>nd</sup> ed., Los Angeles: Sage.
- Bellaby, R.W. (2016) "Justifying Cyber-intelligence?" *Journal of Military Ethics*, 15(4), pp. 299-319.
- Bodmer, S., Kilger, M., Carpenter, G., and Jones, J. (2012) Reverse Deception: Organised Cyber Threat Counter-Exploitation, New York: McGraw-Hill.
- Context. (2015) Integrating Threat Intelligence: Defining an Intelligence Driven Cyber Security Strategy, Centre for the Protection of National Infrastructure, [online], accessed 12 April 2016,  
[https://www.ncsc.gov.uk/content/files/protected\\_files/guidance\\_files/CPNI\\_CONTEXT\\_CERT-Threat\\_Intelligence.pdf](https://www.ncsc.gov.uk/content/files/protected_files/guidance_files/CPNI_CONTEXT_CERT-Threat_Intelligence.pdf).
- d'Aspremont, J., Nollkaemper, A., Plakokefalos, I and Ryngaert, C. (2015). Sharing Responsibility Between Non-State Actors and States in International Law: Introduction. *Netherlands International Law Review*, Volume 62, Issue 1, pp 49–67.
- Duvenage, P. Jaquire, V. & von Solms, S. (2016) "Conceptualising Cyber Counterintelligence: Two Tentative Building Blocks," Proceedings of the 15<sup>th</sup> European Conference on Cyber Warfare and Security, Munich, 7-8 July, pp. 93-103.
- Duvenage, P. and van Niekerk, B. (2016). "Cyber Intelligence and Counterintelligence," ISACA South Africa Conference 2016, Johannesburg, 29-30 August.
- The Economist. (2015) Manage like a spymaster: Counter-intelligence Techniques may Help Firms Protect Themselves Against Cyber-attacks, 27 August, [online], accessed 24 May 2016,  
<https://www.economist.com/business/2015/08/27/manage-like-a-spymaster>.
- Falk, C. (2016) "An Ontology for Threat Intelligence," Proceedings of the 15<sup>th</sup> European Conference on Cyber Warfare and Security, Munich, 7-8 July, pp. 111-116.
- ISACA. (2017) Tech Brief: Threat Intelligence, [online], accessed 16 November 2017, <http://www.isaca.org/Knowledge-Center/Research/ResearchDeliverables/Pages/Threat-Intelligence.aspx>.
- Garner, BA. (1999). Black's Law Dictionary. 7<sup>th</sup> Edition. United States of America.
- Global Commission on the Stability of Cyberspace. (2018) Norm Package Singapore, November, [online], accessed 30 January 2019, <https://cyberstability.org/wp-content/uploads/2018/11/GCSC-Singapore-Norm-Package-3MB.pdf>.
- Heuer, R.J., and Pherson, R.H. (2015) Structured Analytic Techniques for Intelligence Analysis, Los Angeles: Sage.
- Institute of Directors Southern Africa. (2016) King IV: Report on Corporate Governance for South Africa 2016, [online], accessed 28 January 2019, <https://www.iodsa.co.za/page/KingIV>.
- Intelligence and National Security Alliance. (2011) Cyber Intelligence – Setting the Landscape for an Emerging Discipline, September, [online], accessed 7 March 2016, <https://www.insaonline.org/cyber-intelligence-setting-the-landscape-for-an-emerging-discipline/>.

- Intelligence and National Security Alliance. (2013) Operational Levels of Cyber Intelligence, September, [online], accessed 7 March 2016, <https://www.insaonline.org/operational-levels-of-cyber-intelligence/>.
- Intelligence and National Security Alliance. (2014a) Strategic Cyber Intelligence, March, [online], accessed 7 March 2016, <https://www.insaonline.org/strategic-cyber-intelligence/>.
- Intelligence and National Security Alliance. (2014b) Operational Cyber Intelligence, October, [online], accessed 7 March 2016, <https://www.insaonline.org/operational-cyber-intelligence/>.
- Intelligence and National Security Alliance. (2015) Tactical Cyber Intelligence, December, [online], accessed 7 March 2016, <https://www.insaonline.org/tactical-cyber-intelligence/>.
- Joint Chiefs of Staff. (2013). Joint Intelligence, Joint Publication 2-0,
- Lee, R.M. (2014a) "An Introduction to Cyber Intelligence," Tripwire Blog, 16 January, [online], accessed 15 March 2018, <https://www.tripwire.com/state-of-security/security-data-protection/introduction-cyber-intelligence/>.
- Lee, R.M. (2014b) "Developing Your Cyber Intelligence Analyst Skills," Tripwire Blog, 27 January, [online], accessed 15 March 2018, <https://www.tripwire.com/state-of-security/security-data-protection/developing-cyber-intelligence-analyst-skills/>.
- Lee, R.M. (2014c) "Cyber Intelligence Collection Operations," Tripwire Blog, 25 February, [online], accessed 15 March 2018, <https://www.tripwire.com/state-of-security/security-data-protection/cyber-intelligence-collection-operations/>.
- Lee, R.M. (2014d) "Cyber Counterintelligence: From Theory to Practice," Tripwire Blog, 5 May, [online], accessed 15 March 2018, <https://www.tripwire.com/state-of-security/security-data-protection/cyber-counterintelligence-from-theory-to-practice/>.
- Lee, R.M. (2014e) "Cyber Threat Intelligence," Tripwire Blog, 2 October, [online], accessed 15 March 2018, <https://www.tripwire.com/state-of-security/security-data-protection/cyber-threat-intelligence/>.
- Marrin, S., and Madison, J. (2017) Intelligence Studies, Intelligence Analysis, and Multidisciplinary Learning, The National Academies of Sciences, Engineering, Medicine, [online], ccessed 28 January 2019, [http://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse\\_179893.pdf](http://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_179893.pdf).
- MWR InfoSecurity. (2015) Threat Intelligence: Collecting, Analysing, Evaluating, Centre for the Protection of National Infrastructure, [online], accessed 12 April 2016, [https://www.ncsc.gov.uk/content/.../MWR\\_Threat\\_Intelligence\\_whitepaper-2015.pdf](https://www.ncsc.gov.uk/content/.../MWR_Threat_Intelligence_whitepaper-2015.pdf).
- Nijman, J. E. (2010). Non-State Actors and the International Rule of Law: Revisiting the “Realist Theory” of International Legal Personality, Non-State Actors in International Law, Politics and Governance Series. 5.
- Office of the Director of National Intelligence. (2019). National Intelligence Strategy of the United States of America 2019, [online], accessed 28 January 2019, [https://www.dni.gov/files/ODNI/documents/National\\_Intelligence\\_Strategy\\_2019.pdf](https://www.dni.gov/files/ODNI/documents/National_Intelligence_Strategy_2019.pdf).
- Panda Security (2018) The hunter becomes the hunted: How cyber counterintelligence works, 6 July, [online], accessed on 06 November 2018, <https://www.pandasecurity.com/mediacenter/panda-security/cyber-counterintelligence/>.
- Sample, C., Cowley, J., Watson, T., and Maple, C. (2016) "Re-thinking Threat Intelligence," International Conference on Cyber Conflict (CyCon U.S.), Washington, DC, USA, pp. 1-9.
- Spitzner, L. (2003). Honeypots: Are They Illegal? Symantec [online], accessed 8 February 2019, <https://www.symantec.com/connect/articles/honeypots-are-they-illegal>
- Yucel, C., and Koltuksuz, A. (2014) "An Annotated Bibliographical Survey on Cyber Intelligence for Cyber Intelligence Officers," Proceedings of the 13<sup>th</sup> European Conference on Cyber Warfare and Security, Piraeus, Greece, 3-4 July, pp. 213-220.

# A Legal Understanding of State-Linked Cyberattacks and Malicious Cyber Activities

Murdoch Watney

University of Johannesburg, Gauteng, South Africa

[mwatney@uj.ac.za](mailto:mwatney@uj.ac.za)

**Abstract:** Many countries have fallen victim to state-linked cyberattacks and malicious activities. State-linked cyber operations pose serious legal threats and challenges to the stability and security of cyberspace. The discussion aims to establish a legal understanding pertaining to the differences and similarities between state-linked cyber operations such as cyberattacks and malicious cyber activities. In many instances, the terms are used interchangeably. Conduct that may be considered as malicious activities are referred to as cyberattacks. However, cyberattacks and malicious activities are not the same and the consequences and motive for the state-linked cross-border cyber operation may differ. State-linked cyberattacks is defined as a cyber operation that is reasonably expected to cause injury or death to persons or damage or destruction to objects such as DDoS attacks or ransomware attacks. State-linked malicious cyber activities consist of theft of information (espionage), disinformation and false websites. Malicious activities do not cause physical harm to persons or objects. However, the harm in the instance of espionage may consist of financial loss and/or undermining trust in the ability of the government to protect sensitive information or sowing political and social discord such as interference in another country's elections or referendums. Cyber operations evoke various debatable questions such as how should a victim state respond to state-linked cyber operations and when does state behaviour in a foreign cyberspace constitute an act of cyber war or information war? Drawing a clear distinction is relevant when it comes to a victim state's response to a foreign state's cyber operation in their cyberspace on national, international and global level. Stability and security in cyberspace may be achieved by means of international norms governing state behaviour specifically cross-border cyber operations. Although a country should not abuse the cyberspace of another country, the discussion debates the negotiation and enforcement of cyber norms governing state behaviour and how countries should respond to unlawful cyber operations on national and global level.

**Keywords:** cyberattack, malicious cyber activities, cross-border state cyber operations, law, state cyber operations and international law, norms for state behaviour in cyberspace

---

## 1. Introduction

In 2018 the Prime Minister of Australia, Scott Morrison, called on all countries to refrain from various types of malicious activities in cyberspace (Faulconbridge, Deutsch and Lambert, 2018). The legal question that requires an investigation relates to: what is understood with state-linked malicious activities and cyberattacks? Does a state-linked cyberattack include a malicious activity or do these cyber operations differ and if they differ, why is drawing a distinction relevant from a legal perspective?

The reality today is that the near-total digitalisation of business models makes the global economy and security of a country more vulnerable to cross-border cyberattacks and malicious activities, not only from states but also from criminal organisations and other nonstate actors.

The Black Hat Europe Attendee Survey 2018 report referred to as "Europe's Cybersecurity Challenges" shows that nearly two-thirds of European security professionals (65%) believe a major attack on critical infrastructure spanning multiple European countries will occur in the next two years. This figure indicates that concerns over such an attack have not ebbed since the 2017 survey (Black Hat Europe Attendee Survey report, 2018).

Although the threat of cyberattacks and malicious activities have been around for some time, it now poses serious legal threats and challenges as a result of the almost complete digitalisation of governments and businesses. The legal risks and challenges are escalated and aggravated where a foreign state is involved in cross-border cyber operations. Unlike traditional military attacks, a cyber operation can be launched instantaneously from any distance with little obvious evidence of any build-up. In many instances such cyber operation is extremely hard to trace back with any certainty to its perpetrators making a response to the cyber operation harder (Ranger, 2018). Cross-border state cyber operations have resulted in governments and/or organisations debating on how such cyber operations should be dealt with on a national, global and international level.

It is against the above background that the main focus of the discussion centres on establishing a legal understanding of what constitutes state-linked cyber-attacks and malicious activities on national, international and global level. It is a complex topic which explores various inter-related issues.

## **2. Defining a state-linked cyberattack and malicious activities**

There is no universal definition for a cyberattack or malicious cyber activity. However, it is important to conceptualise the terms, cyberattack and malicious activities. In many instances the terms are used interchangeably without drawing a clear distinction between it. However, it will be shown hereafter that when it comes to addressing cross-border state operations, the distinction is very important as it has an impact on how a victim state may address such a cyber operation on a national level and how states on an international and global level may approach it.

The Tallinn Manual 1.0 which was published in 2013 was followed by the Tallinn Manual 2.0 in 2017. The manuals were written at the invitation of NATO and consist of an academic, non-binding study on how international law applies to cyber conflicts and cyber warfare. The 2017 Tallinn Manual 2.0 defines a "cyber operation" as "the employment of cyber capabilities to achieve objectives in or through cyberspace" (Tallinn Manual 2.0 at 564). A "cyber-attack" is defined in Tallinn Manual 2.0 as "a cyber operation, whether offensive or defensive, that is reasonably expected to cause injury or death to persons or damage or destruction to objects" (Tallinn Manual 2.0 at 415; Rule 92).

The examples hereafter will show that not all cyberoperations focus on attacking the systems directly. Many, especially those from Russia, aim to disrupt other nations for political gain through disinformation campaigns (O'Flaherty, 2018). If a cyberattack is typified by physical harm, then espionage (theft of information) or false websites or spread of false information would not constitute an attack but constitute malicious activities. Such a distinction would necessitate an investigation into whether the victim state's response to a cyberattack would differ from that of a malicious activity.

Many of the cyberattacks and malicious activities are cross-border and linked to a foreign state. In September 2018 FireEye Chief Executive Officer Kevin Mandia, whose U.S. security company was hired by Google to defend against state-sponsored cyberattacks, said in an interview that the "dead reality" is that every major cyberattack is somehow state-condoned (Gurdus, 2018). A 2018 report by the Swedish Security and Defence Industry Association (SOFF) indicates that 90% of a researcher's time is spent looking for targeted attacks many of which are nation-state backed and aimed at either stealing secrets or at sabotage (SOFF, 2018).

## **3. Discussion of examples of state-linked cyberattacks and malicious activities**

The motive for state-linked cyberattacks and malicious activities and the response thereto are best illustrated by means of examples. Efrony and Shany (2018) discusses in their article, "A rule book on the Shelf? Tallinn Manual 2.0 on Cyberoperations and Subsequent State Practice" various examples of state-linked cyber operations with reference to the facts, attribution and response of the victim state.

Example 1: The 2016 U.S. presidential campaign/Democratic National Committee (DNC) hack is an example of a state-linked malicious cyber activity.

The 2016 U.S. elections were a game changer pertaining to foreign interference in elections and referendums. It gave other countries facing elections a warning of possible foreign interference and how to mitigate such interference.

The facts are as follows: In May 2016, six months before the U.S. election day, the Democratic National Committee (DNC) invited a cyber security firm (CrowdStrike) to investigate a suspected breach of its network. The investigation team identified intrusions by two well-known hacking actors in cyberspace, namely "Cosy Bear" and "Fancy Bear". Both "Bears" are associated with the Main Intelligence Directorate (GRU), a Russian military intelligence organisation. (Efrony and Shany, 2018; Pollard, Segal and Devost, 2018).

The political motives underlying the operation was designed to sow discord, heighten societal divisions and undermine confidence in democracy. The DNC hack had the essential components of an "influence cyberoperation" intended to modify attitudes and shape opinions through the dissemination of information and

### ***Murdoch Watney***

messages. No significant harm was caused to any computer system but it may have resulted in instability and mistrust in the democratic election process.

The malicious activity was publicly attributed to Russia. Russia denied attribution. It is interesting to note how the U.S. as victim state responded to the malicious activity. The U.S. was not sure how to react to the foreign interference. Efrony and Shany (2018) indicates that the U.S. made use of criminal indictments and diplomatic relations.

After the Brexit referendum and President Trump's election in 2016, other countries took note of foreign interference in an election process. EUROACTIVE (Bulckaert, 2018) report that the 2017 French presidential campaign which was also affected by Russian interference managed to maintain its democratic integrity. In October 2018 Western countries such as Australia, Canada, New Zealand and the UK accused Russia of cyber operations aimed at undermining Western democracies (Faulconbridge, Deutsch and Lambert, 2018).

Example 2: Two large-scale cyberoperations, namely WannaCry and Petya/Notpetya were carried out using leaked U.S. National Security Agency (NSA) hacking tools that exploit vulnerabilities in existing computer networks. The tools had allegedly been stolen and leaked online by a group called "Shadow Brokers," which is reportedly affiliated with Russia (Efrony and Shany, 2018; Ranger, 2018). The cyberattacks had geopolitical consequences.

On 12 May 2017, the WannaCry malware spread like wildfire, affecting hundreds of thousands of computers of companies, government agencies and individuals in more than 150 countries, including Russia, the U.S., China, Germany, Ukraine, and the UK. The affected computers had all data stored in them encrypted and displayed a message demanding payment of a ransom in Bitcoins within three days in order to unlock the computer. Otherwise, access to the data would be permanently lost. In August 2017 it was reported that 52.2 Bitcoins (at the time, around \$143,000) were withdrawn from online wallets for depositing the ransom payments. It turned out, however, that paying the ransom did not ensure unlocking the computer.

Efrony and Shany (2018) indicates that the motivation of those behind the WannaCry operation remains unclear. Ostensibly, it looked like a large-scale criminal ransom operation intended to raise cash. The global manner in which it was executed and the fact that computers remained locked even after the ransom was paid has raised doubts about the true nature of the perpetrators' motives. One possibility is that the operation was designed to draw attention to the NSA's stockpile of cyberweapons which exploit undisclosed vulnerabilities in popular software. Alternatively, it might just have been meant to cause havoc and destruction.

Cybersecurity firms succeeded, within a few days, in identifying the technical footprints of the hackers' group ("Lazarus"), which has been linked to North Korea. The UK government was the first government to officially announce that North Korea was behind the WannaCry attack. Other governments such as Australia, the U.S., Canada, New Zealand and Japan have also attributed the attack to North Korea.

Notwithstanding the attribution of responsibility to North Korea by multiple states, no act of retaliation has been reported (Efrony and Shany, 2018). The UK Security Minister called on Western nations to develop a "doctrine of deterrent" to prevent further cyberattacks. Microsoft's president further called on governments to come together as they did in Geneva in 1949 and adopt a new Digital Geneva Convention that makes clear that these cyberattacks against civilians, especially in times of peace, are off-limits and a violation of international law.

"Petya," a ransomware program targeting Microsoft Windows operated computers, was first revealed in March 2016. On 27 June 2017, a new variant of Petya malware—dubbed "NotPetya" (to distinguish it from the first variant)—was launched against computer systems in Ukraine, using the leaked NSA tool that had already been exploited in the WannaCry operation. The malware spread promptly and globally and affected many companies, institutions and facilities in more than sixty countries including Russia, Poland, the U.S., Germany, UK, Israel, Italy, Netherlands, and others. Almost 80 percent of affected companies were, however, Ukrainian. The malware encrypted all data on the infected systems and demanded a ransom of \$300 in bitcoin for decrypting the files.

NotPetya turned out to be a wiper malware like the Stuxnet. Thus, the damage it caused was irreversible as it did not have the capacity to decrypt the files it had encrypted. NotPetya caused the victim companies unprecedented losses. The tactics, techniques, and procedures (TTPs) of NotPetya operators, as analysed by

cybersecurity experts, lead with high confidence to the conclusion that Russia was behind this operation and that it was related to the ongoing armed conflict in Ukraine (Greenberg, 2018). Ukraine was the first state to point the finger at Russia for orchestrating the NotPetya operation.

Initially, no states affected by NotPetya (except Ukraine) came out with a clear, direct and public attribution claim but in February 2018, the U.S. published a direct and official attribution, claiming that the Russian military launched NotPetya to destabilize Ukraine which turned out to be the most destructive and costly cyberattack in history. Australia joined the U.S. and the UK and published a similar statement attributing responsibility to Russia for the operation (Ranger, 2018).

#### **4. Challenges in addressing state-linked cyberattacks and malicious activities**

The challenges arising from above-given examples are briefly as follows:

- The above discussed examples illustrate that although attribution is not easy to establish, once ascertained, there are victim states who are prepared to publicly attribute an unlawful cyber operation to a specific state. In April 2018 the UK and U.S. made a joint statement blaming Russia for cyberattacks on businesses and consumers (O'Flaherty, 2018). The announcement – which is the first time two nations have come together to show solidarity in this area – saw the National Cyber Security Centre (NCSC), U.S. Department of Homeland Security and the FBI warn businesses and citizens that Russia is exploiting network infrastructure devices such as routers around the world. The aim is to allegedly lay the groundwork for future attacks on critical infrastructure such as power stations and energy grids. It is widely agreed that Russia is one of the most – if not the most – accomplished nations in the world in its ability to perform state sponsored attacks, disinformation and espionage. Countries such as the U.S., Israel, China, North Korea and Iran have the technical knowledge to launch cyberattacks (Perlroth and Krauss, 2018).
- In some instances, there can only be speculation pertaining to which foreign state was responsible for the attack. For example, in August 2017 a petrochemical company with a plant in Saudi Arabia was hit by a cyberattack (Perlroth and Krauss, 2018). The attack was not designed to simply destroy data or shut down the plant, but it was meant to sabotage the firm's operations and trigger an explosion. Initially it was thought that Iran was responsible as there have been tension between Iran and Saudi-Arabia (Perlroth and Krauss, 2018). In October 2018 new research indicated that if it was Iran, it had a lot of Russian assistance, and that when the malware needed to be fine-tuned, the Russian institute provided the expertise (Sanger, 2018). However, not all states are prepared to attribute a cyberattack to a specific country. For example, Singapore suffered its biggest ever cyberattack in 2018. Singapore indicated that state actors were likely behind the cyberattack citing the scale and sophistication of the hack which hit the medical data of about a quarter of the population, but Singapore did not attribute the attack to a specific nation-state (Kwang, 2018).
- Wheeler (2018) indicates that the first step to improving cyber defense is to determine what constitutes a cyberattack by a foreign power. For example, in the run-up to the 2017 German parliamentary elections, a string of cyberattacks led to fears of Russian meddling, but according to the Charter of the United Nations (UN), unless armed force has been brought to bear within the borders of a country, no internationally recognized act of aggression has occurred (Wheeler, 2018). Similarly, cyberattacks in the Netherlands in 2017 and 2018 resulted in the denial of government funding and vital services to citizens, but because conventional battlefield weapons were not used, the UN Charter's provisions were not violated.
- It is clear from the examples that neither governments nor businesses know how to respond to cyberattacks and malicious cyber activities. The alleged Russian hacking during the 2016 U.S. presidential campaign was not a violation of the UN Charter's prohibition on the use of force. It amounted to information warfare but not cyber warfare. Cyber warfare consists of an attack between states of such significant scale and severity that it could be seen as the equivalent of a physical attack. Attacks by individual hackers or even groups of hackers would not be considered cyber warfare unless they are being aided and directed by a state. Ranger (2018) indicates that a common trend is states providing support to hackers in order to create plausible deniability of their own actions.
- Ranger (2018) indicates that while Western strategists tend to regard cyber warfare and information warfare as separate entities, Chinese and Russian military theorists regard the two as closely linked.
- State behaviour such as cyberattacks and malicious activities in the cyberspace of another country result in insecurity, instability and mistrust between nations. It is important to understand how state behaviour in cyberspace on an international level may be addressed.

## **5. Establishing international binding norms governing state behaviour in cyberspace**

There are several ideas being discussed at international, global and national levels. Some of them are focused on the UN or UN agencies, while others try to draw on the lessons of the past, actively avoiding the challenges those structures bring by considering radical ideas outside of the UN framework (Ruiz, 2018). What all the approaches agree on is that the so-called “inadvertent escalation” – accidental war – through state actions in cyberspace is a real and increasing threat and one that urgently needs to be addressed (Ruiz, 2018).

The following proposals to govern cyberspace have been made:

### **5.1 Tallinn Manual 1.0 and 2.0**

The purpose of the Tallinn Manuals is to explain how international law apply to cyber warfare, but as indicated, it is not binding. Efrony and Shany (2018) indicates that much of the criticism against the Tallinn Manual has been focused on how it represents a NATO and specifically a Western outlook on what cyber warfare should be.

### **5.2 Digital Geneva Convention**

As indicated, Microsoft's president has proposed a new Digital Geneva Convention. Wheeler (2018) also proposes a digital Geneva Convention that produces a few deep and well-enforced rules surrounding the conduct of war in cyberspace. However, if a group of experts actually did convene to create a binding digital Geneva Convention, it is unclear from what source it would derive its authority. It could be the Tallinn Manual but it is non-binding and it is also not an official North Atlantic Treaty Organisation (NATO) publication. Moreover, as Wheeler (2018) indicates, the alliance itself is currently on shaky ground and there is no guarantee that the U.S. would abide by any agreement.

### **5.3 International norms for state behaviour in cyberspace under the auspices of the U.N. Group of Governmental Experts (GGE)**

The United Nations Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security (GGE) is an UN-mandated working group and is currently composed of 25 government representatives.

In 2017 the UN GGE failed to reach an agreement. It appears that the main issue in dispute was the application of international law to cyberspace and the right to use self-defence in the face of a cyberattack. After the failure of the GGE to reach a consensus report in 2017, governments took some “cooling-off time” to evaluate the present situation and to consider the way forward.

At the end of 2018 both the U.S. and Russia submitted proposals to the UN Committee on Disarmament and International Security for implementing a process commencing in 2019 to develop international cyber norms (Johnson, 2018). There are sections in the Russian proposal that may be contentious. One section reaffirms “the right and duty of states to combat, within their constitutional prerogatives, the dissemination of false or distorted news, which can be interpreted as interference in the internal affairs of other States.” As discussed, the U.S. intelligence agencies concluded that Russia interfered in the 2016 U.S. presidential elections. The proposal also states that nations have a duty to “abstain from any defamatory campaign, vilification or hostile propaganda for the purpose of intervening or interfering in the internal affairs of other States” and avoid embedding malware or other cybersecurity vulnerabilities in critical infrastructure technology or global supply chain (Johnson, 2018). In the latter regard, note should be taken of the 2010 Stuxnet which targeted Iran’s nuclear centrifuges (Clarke and Knake, 2012; Ranger, 2018). The U.S. is widely believed to have worked with Israel to develop and deploy the Stuxnet malware that set back Iran’s nuclear program in 2009-10 (Johnson, 2018).

On 5 December 2018 the UN General assembly adopted a resolution that established an open-ended working group, acting on the basis of consensus, for the further development of norms and principles of responsible behaviour of states in cyberspace and ways of their implementation. The first organizational meeting is scheduled in June 2019.

The GGE's recommendations to the UN secretary general about norms nations should honour in cyberspace will contribute greatly towards restoring stability, security and trust in cyberspace, but unfortunately the group's work is often hindered by larger squabbles between the U.S, Russia, China and other nations. It is apparent that there exists a great deal of mistrust between countries. The U.S. has viewed Russia's calls for norms as a cynical way to attempt to limit American cyberactivity, while sending out surrogates to conduct operations on Russia's behalf (Sanger, 2018).

#### **5.4 Global Commission on the Stability of Cyberspace (GCSC)**

In 2017 the Global Commission on the Stability of Cyberspace (GCSC), which is based in the Netherlands, was established by the Hague Centre for Strategic Studies and the East West Institute and supported by numerous organizations, including the Internet Society and Microsoft. The GCSC is a multi-stakeholder commission of experts which sets out to bring perspective by developing proposals for norms and policies to enhance international security and stability and guide responsible state and non-state behaviour in cyberspace.

In November 2018 the GCSC announced the release of its new norm package featuring six new global norms to help promote the peaceful use of cyberspace. The norms were developed with the express purpose of being adopted by public and private sectors towards an architecture to improve international security and stability in cyberspace. These norms are non-binding and would voluntarily be applicable on a global level.

### **6. Consequence of absence of international treaty on state behaviour in cyberspace**

The above discussion illustrates that no treaty pertaining to state behaviour has been reached on an international level. Countries have different approaches to the governance of the Internet. A distinction in this regard is drawn between a multi-stakeholder model and a multi-national model. Countries are not able to agree on an international level on the Internet governance model. The United States is aligned with a group of countries that insists that existing international law fully applies to cyberspace, whereas Russia is aligned with another group that wants a new treaty pertaining to state behaviour tailored specifically to this domain.

Wheeler (2018) indicates that it appears that many countries are beginning to coalesce around the idea that some forms of active countermeasures are justified in self-defense, if not in actual reciprocation, under international law. In 2014 NATO took the important step of confirming that a cyberattack on one of its members would be enough to allow them to invoke Article 5, the collective defence mechanism at the heart of the alliance (Ranger, 2018). In 2016 it defined cyberspace as an operational domain, an area in which conflict may occur (Ranger, 2018).

Nations are building cyber defences and offence capabilities. There is a risk that some countries may be at the early stages of a cyberwar arms race as they realise that having a cyber warfare strategy is a necessity. This means that more states may be stockpiling cyber weapons such as zero-day attacks which may be exploited and result in attacks escalating quicker (Ranger, 2018). Cyber weapons may be analysed and even potentially be repurposed and re-used by the country it was used against. Ranger (2018) refers to Stuxnet which was the first genuine cyber weapon in that it was designed to cause physical damage.

In September 2018 the U.S. introduced a new Cyber Deterrence initiative. The cyber strategy of the U.S. Department of Defence (DoD) calls upon the military to "defend forward" and "to prepare for war". The term "defend forward" is understood as the DoD penetrating other systems to either understand their adversaries better or disrupt and halt state-linked cyber operations before it happens.

The European Union (EU) is increasingly cooperating in cyber defence with a view to strengthen its capacities. On 18 October 2018 the European Council called for measures to build strong cybersecurity which include the ability to respond to and deter cyber-attacks.

Ranger (2018) proposes that countries must deter foreign countries from unlawful cyber operations by making the cost of such cyber operations too high. Government and businesses need to work together to address the threat of state cyber operations especially as U.S. businesses are a favourite target of nation-state attacks (Lemos, 2018). Unfortunately, businesses are not always in a position to repulse these cyber operations as state cyber operations – even from smaller states – are usually too persistent to repulse on a regular basis. Lemos

(2018) suggests that a government should assist businesses in various ways aimed at dissuading foreign states from unlawful cyber operations. A government can impose sanctions on a foreign state perpetrator. Unfortunately, it should be borne in mind that criminal indictments, sanctions and diplomatic interactions do not always dissuade states as many states still continue to “poke and pry at the computer systems of their rivals” (Ranger, 2018). Lemos (2018) proposes that the U.S. government may consider a more “painful” countermeasure. In this regard, the U.S. by introducing the Cyber Deterrence initiative is looking at a different approach to responding to attacks especially with regard to the critical infrastructure that may be vulnerable. It is therefore important that a government should reconsider what is a critical infrastructure and protect such structure.

## **7. Conclusion**

It is important that state-linked cyber operations are deterred but establishing solutions in addressing such operations are proving challenging. The discussion highlighted that the existing international law framework fails to provide legal remedies to address state-linked malicious operations.

Wheeler (2018) indicates that the challenge of today is the rapid speed at which cyberspace morphs and evolves. It is changing faster than international summits can be convened, making obsolete any deal that takes longer than a week or two to negotiate. Even if one country can come to an internal agreement on what constitutes a cyberattack from one private party to another, there is no guarantee that two countries would do the same.

It is clear from the above discussion that governments are still a long way off from being able to draft new legally-binding conventions that would be both monitorable and enforceable, or even understandable in terms of having agreed on definitions of key components. It may be that for now, legally non-binding norms may be the best way forward.

Ranger (2018) correctly opines that countries need to talk more, to understand where the boundaries lie and which kind of behaviour is acceptable. Unfortunately, ideological division among states on approaches toward the use of information and communications technologies now appear to be deeper than ever before. It means that in the absence of an international agreement, like-minded countries may create a coalition to work on the attribution of, response to and consequences for conducting cyber operations, giving the term ‘deterrence’ the meaning of consequences in the broader cyber policy area.

## **References**

- Black Hat Europe Attendee Survey Report (2018) “Europe’s Cybersecurity Challenges”, [online],  
<https://globenewswire.com/news-release/2018/11/14/1651150/0/en/2018-Black-Hat-Europe-Research-Low-Confidence-in-GDPR-Lack-of-Privacy-via-Social-Media-Impending-Nation-Wide-Critical-Infrastructure-Attacks-More.html>.
- Bulckaert, N. (2018) “How France successfully countered Russian interference during the presidential election”, [online],  
<https://www.euractiv.com/section/elections/news/how-france-successfully-countered-russian-interference-during-the-presidential-election/>.
- Clarke, R. and Knake, R. (2012) *Cyber War* HarperCollins Publishers New York, U.S.
- Efrony, D. and Shany, Y. (2018) “A rule book on the shelf? Tallinn Manual 2.0 on Cyberoperations and Subsequent State practice” [online], <https://www.cambridge.org/core/journals/american-journal-of-international-law/article/rule-book-on-the-shelf-tallinn-manual-20-on-cyberoperations-and-subsequent-state-practice/54FBA2B30081B53353B5D2F06F778C14>.
- Faulconbridge, G., Deutsch, A. and Lambert, L. (2018) “West accuses ‘pariah state’ Russia of global hacking campaign”, [online], <https://www.reuters.com/article/us-britain-russia-cyber/west-accuses-pariah-state-russia-of-global-hacking-campaign-idUSKCN1ME1HE>.
- Fruhlinger, J. (2018) “What is a cyber attack? Recent examples show disturbing trends”, [online],  
<https://www.csionline.com/article/3237324/cyber-attacks-espionage/what-is-a-cyber-attack-recent-examples-show-disturbing-trends.html>.
- Greenberg, A. (2017) Petya ransomware epidemic may be spill-over from cyberwar”, [online],  
<https://www.wired.com/story/petya-ransomware-ukraine/>.
- Gurdus, E. (2018) “Every cyber attack is related to geopolitical conditions says CEO of cybersecurity company hired by Google”, [online], <https://www.cnbc.com/2018/09/25/fireeye-ceo-every-cyberattack-is-related-to-geopolitical-conditions.html>
- Johnson, D.B. (2018) “Russia jockey to shape new global cyber norms”, [online], <https://fcw.com/articles/2018/11/09/us-russia-cyber-norms.aspx>.

### **Murdoch Watney**

- Kwang, K. (2018) "Singapore health system hit by 'most serious breach of personal data' in cyberattack; PM Lee's data targeted", [online], <https://www.channelnewsasia.com/news/singapore/singhealth-health-system-hit-serious-cyberattack-pm-lee-target-10548318>.
- Lemos, R. (2018) "Five Ways Government can help businesses fight nation-state attacks", [online], <https://www.eweek.com/security/five-ways-government-can-help-businesses-fight-nation-state-attacks>.
- O'Flaherty, K. (2018) "Cyberwarfare: The threat from Nation States", [online], <https://www.forbes.com/sites/kateoflahertyuk/2018/05/03/cyber-warfare-the-threat-from-nation-states/#258321cd1c78>.
- Perlroth, N. and Krauss, C. (2018) "Cyber attack in Saudi-Arabia had a deadly goal. Experts fear another try", [online], <https://www.nytimes.com/2018/03/15/technology/saudi-arabia-hacks-cyberattacks.html>.
- Ranger, S. (2018) "Cyberwarfare could turn every gadget you own into a weapon on a virtual battlefield. And the damage will be felt in the real world", [online], <https://www.zdnet.com/article/cyberwar-a-guide-to-the-frightening-future-of-online-conflict/>.
- Ruiz, M.M. (2018) "Q&A with Marina Kaljurand on the future of cyberspace", [online], <https://hewlett.org/qa-with-marina-kaljurand-on-the-future-of-cyberspace/>.
- Sanger, D. E. (2018) "Hack of Saudi Petrochemical Plant was Coordinated from Russian Institute", [online] <https://www.nytimes.com/2018/10/23/us/politics/russian-hackers-saudi-chemical-plant.html>.
- Swedish Security and Defence Industry Association (SOFF) (2018) "State Sponsored Cyber Attacks", [online], [https://soff.se/wp-content/uploads/2018/03/Cybersecurity\\_statsunderstödda-aktörer.pdf](https://soff.se/wp-content/uploads/2018/03/Cybersecurity_statsunderstödda-aktörer.pdf).
- Pollard, N., Segal, A and Devost, M.G. (2018) "Trust war: dangerous trends in cyber conflict", [online], <https://warontherocks.com/2018/01/trust-war-dangerous-trends-cyber-conflict/>.
- Wheeler, T. (2018) "In Cyberwar there is no rules", [online], <https://foreignpolicy.com/2018/09/12/in-cyberwar-there-are-no-rules-cybersecurity-war-defense/>

# A Methodology for the Comparative Analysis of Strategic Culture and Cyber Warfare

Andrew Williams

University of New South Wales, Canberra, Australia

[andrew.williams3@student.adfa.edu.au](mailto:andrew.williams3@student.adfa.edu.au)

**Abstract:** A state's strategic culture influences policy and behaviour in cyberspace. If cultures are misunderstood, work towards the establishment of international norms may be degraded. A nuanced approach is required to support security and strategic studies for the establishment of norms beyond the technical realms of military science and operations in cyberspace. Doctoral research at the University of New South Wales is investigating this cross-discipline field to provide an analysis of Australian strategic culture and its influence on cyber warfare. Strategic culture, born from the analysis of nuclear strategy in the Cold War, has a mixed 40-year history with multiple theories and methodologies. Research on strategic culture and cyber warfare is not highly developed, providing little opportunity for comparative analysis of strategic cultures and their influence on cyber warfare. A workable comparative methodology will advance the field and provide practical outcomes for policy formation and establishment of norms. The paper proposes a mixed methods framework combining qualitative primary document research with key informant interviews and a broader survey. The methodology enables an understanding of the roles of elites in policy formation and provides stratification to demonstrate the influence of the culture on the main agents, namely military and related defence personnel. Furthermore, the methodology supports comparative static and longitudinal studies of nation states in cyberspace to enable policy makers to coalesce understanding of strategic cultures for the establishment of international norms in cyberspace.

**Keywords:** methodology, strategic culture, norms, cyber warfare

---

## 1. Introduction/background

Strategic culture influences policy and behaviour though is largely unexplored in cyber warfare where technology dominates, and social sciences are less influential. The academic body of knowledge in the cross-discipline field of Strategic Culture and cyber warfare is scant and lacking an agreed methodology. Indeed, strategic culture itself is a divided sub-field. An agreed methodology is essential to progress global comparative research capable of supporting the establishment of norms. As stated by Johnston (1995, pp. 63-4), 'Done well, the careful analysis of strategic culture could help policymakers establish more accurate and empathetic understandings of how different actors perceive the game being played, reducing uncertainty and other information problems in strategic choice'.

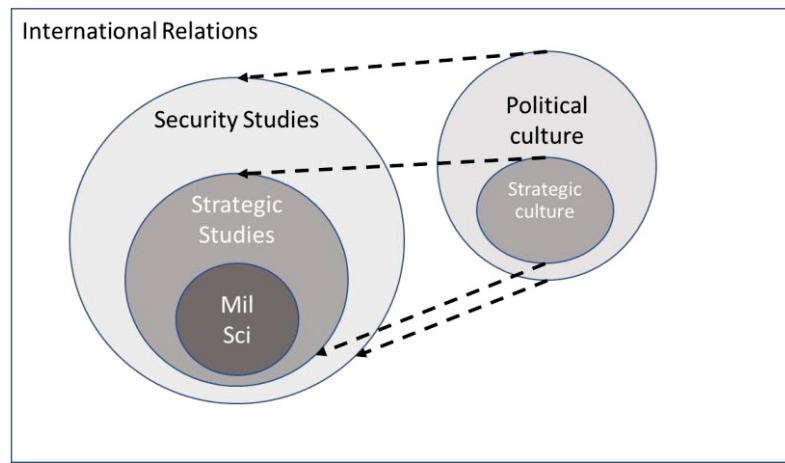
Establishment of an agreed methodology will enable comparative studies to inform national policy and influence international norms. Initially such work will be static in nature, while enabling consideration of longitudinal studies as the body of knowledge deepens.

This paper provides a methodology to enable international comparative research on the influence of strategic culture on cyber warfare. The paper is aligned with thinking on strategic studies and limits strategic culture as a subset of political culture. The paper considers the state of the field for strategic culture, including previous comparative research methods, before analysis of current research into the cross-discipline. The paper then provides the comparative methodology.

## 2. Literature review

### 2.1 Situating the field

Betts (1997, p. 9) provided a clear definition of strategic studies, placing it between the subfields of security studies and military science, such that it forms 'how political ends and military means interact under social, economic, and other constraints'. Further, Macmillan et al. (1999, p. 11) define strategic culture as a subset of political culture. As such we may consider that strategic culture is embedded within strategic studies and is a subset of political culture as shown in Figure 1.



**Figure 1:** Strategic studies and strategic culture within international relations

## 2.2 The origins and evolution of the idea of strategic culture

The strategic culture body of knowledge is just over forty years old, and lacking in consistency in philosophy, methodology and output. The origin of the idea of strategic culture is attributed to Snyder (1977, p. 8) and his analysis of Soviet Cold War nuclear strategy, where he identified strategic culture as being the combination of ideas, responses and habitual behaviours acquired and shared by a national strategic community.

The notion of generations of strategic culture theory is contested. Johnston (1995) divided the research into three generations with criticisms for each; the first focussing on multiple input factors for US/Soviet nuclear strategy ('over-determined and under-determined'), the second linking strategic culture and behaviour ('ambiguous instrumentality') and the third targeting the realist edifice ('organisational culture as an intervening variable'). Various academics have subsequently proposed or discussed a fourth generation (Bloomfield 2012; Haglund 2014; Howlett 2006; Libel 2016; Pirani 2016). The fourth generation employs constructivist/interpretivist worldviews, places culture as an intervening variable, adopts the notion of competing strategic sub-cultures, and incorporates both ideas and behaviour. This methodology, and associated research, is within this concept of the fourth generation of strategic culture theory.

A balance must be struck between methodological rigour and the substance of strategic culture. In response to Johnston (1995), in particular his demand for scientific rigour and falsifiability, Gray (1999) defended the work of the first generation and provided an update to the theory. He later reaffirmed his dissatisfaction with Johnston's theory, drawing on Hedley Bull to argue the risk that 'the demands of rigor and precision in theory construction are allowed to triumph over the substance of the subject, [concluding that] while a rigorous method is admirable, it ought not to take precedence over an inconvenient reality' (Gray 2009, p. 223). The challenge is, therefore, to provide a sufficiently robust and academically rigorous methodology to enable comparative studies without permitting theory to triumph over substance.

## 2.3 Comparative strategic culture

Limited comparative research is available to benchmark methodology. Two comparative edited works are considered, namely Booth and Trood (1999), and Johnson et al. (2009).

Macmillan and Booth (1999) provide a 'framework for analysis' in Booth and Trood (1999). This research provides comparative analysis of national strategic cultures rather than influence against a capability such as cyber. The authors emphasise they are 'seeking to identify those lasting features of *thought* in particular societies, and how they manifest themselves in *behaviour*' (Macmillan & Booth 1999, p. 363). The fundamental assumption is 'that decision-makers always have some element of choice, and that something that can be called "strategic culture" plays a part in shaping the decisions they reach' (Macmillan & Booth 1999, p. 364). The framework lists three sources of strategic culture as the 'most significant' while acknowledging these do not 'exhaust sources of explanation'; Geography and Resources, History and experience, and Political structure and defence organisation (Macmillan & Booth 1999, pp. 365-6). The authors note a sensitivity to criticism that they are not exhaustive. The framework subsequently provides a short list of elements to consider, namely: Political

culture and strategic culture, incorporating the history of strategic thought, strategic doctrine and a profile of traditional strategic culture; Contemporary strategic policy incorporating nuclear strategy, conventional strategy, disarmament and arms control, unconventional strategy, independence/interdependence, security, defence decision-making, the new strategic environment, and strategic culture, society and identity (Macmillan & Booth 1999, pp. 366-9). The framework provides a useful list of research questions, though does not provide a methodology.

The second comparative analysis relates to Strategic Culture and Weapons of Mass Destruction. Johnson (2009) provides a different framework for analysis compared with Booth and Trood (1999). Johnson (2009, pp. 245-54) discusses four variables ‘that are most likely to have an effect on security policy, or are sufficient to understanding the pivotal components of a foreign society’s cost/benefit analysis’, including ‘identity, values, norms and perceptive lens’. Johnson (2009, p. 254) notes ‘inputs such as geography, history, access to technology, political experience, religious traditions, education, demographics, common texts, and so on *create* identity, values, norms, and a group’s perceptive lens’. Therefore, to arrive at the four variables, perhaps better termed ‘themes’, one should analyse the inputs, then aggregate into themes.

## **2.4 Strategic culture and cyber**

The first identified work on strategic culture and cyber is the analysis of integration of Russian cyber power into grand strategy by Wirtz (2015). He sought ‘national styles’ in cyber strategy that might produce idiosyncratic behaviour to enable attribution. He acknowledges the technical nature of cyber warfare and the lack of ‘connection between technical exploitation and grand political strategy’ (Wirtz 2015, p. 29). Wirtz (2015, p. 30) rightly notes that ‘Russia, more than any other nascent actor on the cyber stage, seems to have devised a way to integrate cyber warfare into a grand strategy capable of achieving political objectives’. His brief analysis of Russian strategic culture and technology adds little to understanding of strategic culture, other than Russian fears of Western technological advantage, and the Clausewitzian notion of war and politics. He surmises that Russia has strategically employed non-lethal force to circumvent NATO’s deterrence posture, forcing NATO to escalate and risk being viewed as the aggressor. There is no discussion of methodology, limiting the ability to advance the field.

Tosbotn (2016) offers a constructivist worldview and a methodology for quantitative and qualitative analysis. The quantitative assessment is a frequency count of a number of key terms, including ‘cyber’. Tosbotn (2016, p. 17), notes ‘the boost in frequency over time affirms “cyber” as a prominent and aspiring dimension of the perceived security environment’. The qualitative element of this research is fulfilled through broad statements regarding the quantitative element of the research. While this provides an element of richness to the frequency analysis it does not represent a mixed methods approach. There is brief mention of holding Russia as an “other” (Tosbotn 2016, p. 19) which is broadly aligned with Bloomfield (2011).

Johnson (2018) uses an operationalised definition of strategic culture to define ‘cyber strategic culture’ by specifically directing the amalgamation of culture, history and national interests to cyberspace’ (Johnson 2018, p. 17). The methodology uses ‘a comparative case study analysis of three past cyber-attacks in conjunction with a policy development analysis’ (Johnson 2018, pp. 17-8). The limitation of this methodology, acknowledged by Johnson (2018, p. 19), is ‘specific to the UK and their own national cyber security strategies, so it will not necessarily be useful in applying to other nations as threats and abilities vary state to state’.

Persoglia (2018) also proposes a definition of ‘cyber strategic culture’ with a qualitative case study on the United States. This thesis primarily employs a deductive thematic analysis methodology. The framework centres on ‘offensiveness’ and ‘defensiveness’. Each theme is sub-divided into sub-themes: defensiveness into ‘network security/defence, cyber-resilience, and cooperation’; offensiveness into ‘pre-emption/prevention’, ‘retaliation’ and ‘domination’ (Persoglia 2018, pp. 68-86). The concept of *a priori* deductive themes is beneficial for comparative analysis, though there are issues associated with this particular taxonomy. For example, resilience is necessary whether a state has an offensive or defensive posture, and cooperation may occur in a defensive or offensive posture. Similarly, deterrence is a defensive strategy, though has been deemed offensive. A sub-culture schema, such as that proposed by Bloomfield (2011, p. 80) is advantageous, though requires further consideration.

The aforementioned research lacks a supportive methodology to enable future comparative work. Such research may further cloud the cyber warfare debate. Further, we should be cautious at the development of a distinct 'cyber strategic culture' beyond a national strategic sub-culture schema. Such a schema influences policy and behaviour across all elements of military capability, including cyber warfare. The opposite opinion would see a strategic culture for each domain, i.e. a maritime, land, air, space etc. strategic culture. The methodology therefore proposes the analysis of the influence of a national strategic sub-culture schema on policy and behaviour associated with cyber capability.

### **3. The proposed methodology**

#### **3.1 Underpinnings**

The concept of stability in constructivist research is far from resolved. Adler (2012, p. 122) states 'Constructivists use a large variety of *methods*: positivist, post-positivist, quantitative, qualitative, and a combination of these'. With instability in the concept of strategic culture, and in constructivist methodology, we risk a field that is limited in ability to provide comparative research. A means to ensure richness is to broaden the methodology, employing mixed methods. Further, an element of pragmatism 'open[s] the door to multiple methods, different worldviews, and different assumptions, as well as different forms of data collection and analysis' (Cresswell 2014, p. 11). The proposed methodology draws on constructivist philosophy, taking advantage of pragmatism to provide a rich mixed-methods approach. It similarly heeds the concerns of Bull, as echoed by Gray, while acknowledging the requirement for academic rigour desired by Johnston.

The methodology enables analysis of the influence of strategic culture on cyber warfare, rather than developing a distinct cyber strategic culture. As such, it analyses cyber policy and related behaviour in the context of broadly defined strategic culture and an associated framework for thematic analysis. The 'convergent parallel mixed method' (Cresswell 2014, pp. 219-23) comprises three elements; qualitative policy research, qualitative *Key Informant* interviews and a hybrid survey.

#### **3.2 Definition of strategic culture**

An agreed definition of Strategic Culture is required to baseline comparative research. The broad definition allows for strategic culture to vary over time and not be immutable or monolithic. It is related to strategic studies rather than being further reduced to a singular capability such as cyber. Furthermore, the definition incorporates policy and behaviour. The definition for use in this cross-discipline field is that provided by Bloomfield (2011, p. 288):

*The habits of ideas, attitudes, and norms toward strategic issues, and patterns of strategic behaviour, which are relatively stable over time.*

#### **3.3 Thematic analysis**

As noted above, a deductive framework enables consistency in analysis. Thematic analysis provides sufficient rigour for data analysis as Bull's concerns eliminate the requirement for cumbersome theory. The following framework draws heavily on the literature review, namely the comparative frameworks, and similarly noting the aforementioned concerns of Macmillan and Booth (1999).

##### **3.3.1 Strategic culture**

Research should consider previous studies that pre-date cyber warfare capability to ensure depth of understanding. Macmillan and Booth (1999, p. 367) provide for; the history of strategic thought, strategic doctrine and a profile of traditional strategic culture. Similar consideration may be given to 'history and experience' and 'political structure and defence organisation' (Macmillan & Booth 1999, pp. 365-6). Consideration should then be given to changes in policy and behaviour following the advent of cyber warfare capability.

##### **3.3.2 Contemporary strategic culture and cyber capability**

The following themes and associated questions provide the framework for analysis.

*Strategic sub-culture schema*

A list of strategic sub-cultures is to be provided with the framework, expanding on the concept of a ‘schema set’ (Bloomfield 2011, p. 80), to support identification of like-minded and opposed nations. Sub-cultures to be included, for example in the Australian case are; neutrality, offence, forward defence, continental defence (defensive neo-realism), internationalism, and pacifism. Which sub-cultures are within the schema set for the nation in question, and which is/have been dominant and/or latent following the advent of cyber warfare capability? Is the influence of the dominant sub-culture similar for cyber versus more traditional military capabilities?

*Geography and resources*

While cyber operations occur in the virtual space, there remains a fundamental connection with the physical environment. The notion of layers is well considered in US military doctrine (USDoD 2018). Similarly, the near-instantaneous nature of the passage of data across cyberspace reduces the traditional notion of geography, with physical separation associated with time required to reach an ally or adversary eroded. Notions of traditional adversaries and allies remain, best aligned with the Bloomfield (2011) friend/foe calculus.

Resource questions include indigenous industrial capacity, supply-chain security, and availability of cyber education and workforce.

*Conventional strategy*

Is cyber capability viewed as a conventional capability? How has conventional strategy evolved considering cyber warfare capability?

*Cyber strategy*

What are the characteristics of cyber warfare and cyber security policy? How well aligned is cyber strategy with traditional strategic culture, including traditional classic and/or modern strategists? To what extent does the nation have an offensive or defensive cyber posture – policy, capability and behaviour? How has policy changed over time and responded to global events? Are cyber strategy and behaviours aligned? How has the legal framework evolved with cyber strategy? Is cyber a domain or an element of another domain?

*Independence and Interdependence*

Does the nation have a focus on self-reliance in conventional and cyber capability? Have alliances and partnerships altered through the emergence of cyber warfare? What messaging is required to shape perceptions regarding cyber capability? Does the nation follow an internationalist approach? Does the nation comply with international regimes and norms? Does thinking on international law shape cyber capability?

*Security*

Does the nation have a narrow or broad sense of cyber security enabled by Defence capability? Is cyber security strategy and behaviour aligned? Are elements of national power secured by Defence capability?

*Decision-making*

Is the military strategic decision-making process consistent for conventional and cyber capability? Are behaviours aligned with policy? How does the legal framework influence decision-making?

### **3.4 Mixed methods elements**

#### *3.4.1 Qualitative policy and behaviour analysis*

Policy has long formed the primary source for the analysis of strategic culture. Defence white papers and other key policy documents since inception of cyber capability form the primary reference set. Foreign policy should

be considered to the extent they refer to Defence capability, without extending to security studies and political culture. Political, bureaucratic and military elite speeches, and parliamentary debates are also primary sources. Secondary sources are analyses of such policy, statements and debate. Behavioural elements also need to be captured to determine whether cyber policy reflects strategic culture, and whether behaviours/outcomes are aligned with policy.

Consistency in coding is fundamental to ensuring consistency in output. Deductive coding (Miles et al. 2014, p. 81) against the framework provides the start state, without limiting the opportunity for other codes to emerge. The collection of terms in the framework may then be aggregated against the broader themes.

### **3.4.2 Qualitative key informant interviews**

The opportunity to speak with the elite provides opportunity to clarify individual perception of strategic culture, and further ability to analyse the influence of strategic culture on cyber policy formation and behavioural implementation. *Key Informants*, according to Payne and Payne (2004, p. 134),

*are those whose social positions in a research setting give them specialist knowledge about other people, processes or happenings that is more extensive, detailed or privileged than ordinary people, and who are therefore particularly valuable sources of information to a researcher, not least in the early stages of a project.*

*Key Informants* in this context are members of the political, government bureaucratic and military elite who are engaged in cyber warfare capability. Key Informants may also extend beyond these three classes to incorporate private sector individuals such as Defence industry or other business figures with an interest in cyber warfare. The selection of *Key Informants* is *a priori* purposeful (Hood 2007, p. 157), with the researcher carefully selecting those with known specialist knowledge. Interviews should seek to extend the quantity of *Key Informants* through referral during interview.

Interview design should be *semi-structured* to capture the framework, commencing with a *tour* question, and enabling *probes* to delve deeper. The purpose of the *tour* question is to enable the ‘interviewees to act more or less as guides, walking you through their turf while pointing out what they think is important on the way’ (Rubin & Rubin 2005, p. 8). This enables the interviewee to feel relaxed in their setting, and the interviewer may identify areas to focus on in later probes. *Main questions* (Rubin & Rubin 2005, pp. 10-2) form the majority of the interview and are based on the deductive framework. *Probes* (Rubin & Rubin 2005, pp. 12-9) enable the interviewer to clarify responses or to delve more deeply into particular areas of the interview without requiring pre-prepared questions.

The number of *Key Informant* interviews required is dependent on *saturation*, defined by Schutt (2012, p. 528) as ‘the point at which subject selection is ended in intensive interviewing, when new interviews seem to yield little additional information’.

### **3.4.3 Survey**

The survey enables analysis of influence at a lower level. A means to achieve consistency in populations for comparative research comes through military education systems. Most militaries with the ability to generate cyber capability have more advanced military education systems that progress through career levels. This provides for *convenience sampling* (Fink 2017, pp. 99-103), namely a *purposive sample*. The sample selected assumes mid and senior level staff courses. Individuals selected for such courses are likely to have knowledge of strategy and may have knowledge of cyber capability. As noted by Fink (2017, pp. 99-100), those willing to complete the survey may be more concerned than those who don’t, may wish to complain or brag, or may not have time. The selection of mid and senior level staff courses provides stratification to assess the depth of influence of strategic culture.

For consistency across the research elements, the survey instrument should be composed in a similar fashion to the *Key Informant* interviews, providing for ease of side-by-side analysis. The survey provides the opportunity for quantitative and qualitative elements.

### **3.5 Improving validity and understanding bias in comparative studies**

A summary of validity and reliability factors and methods is provided by Cresswell (2014, pp. 201-2). Of particular note Cresswell (2014, p. 202) states that ‘good qualitative research contains comments by the researchers about how their interpretation of the findings is shaped by their background, such as their gender, culture, history, and socioeconomic origin’. Greater consideration of this aspect is required as research expands from singular PhD research to broader international efforts.

Validity is enhanced through triangulation, namely side-by-side analysis, making ‘the comparison within a discussion, presenting first one set of findings and then the other’ (Cresswell 2014, p. 222).

## **4. Application of the methodology**

### **4.1 Employment**

This methodology is currently employed in PhD research at the UNSW Canberra, specifically to analyse the influence of strategic culture on Australian Defence cyber warfare capability. The research covers the period from 2000, with the first mention of cyber in a Defence White Paper (Australia 2000), through to the present. The mixed methods elements have been employed in parallel, with ongoing data collection and analysis.

Several themes are emerging. First and foremost, Australia is at the tipping point of dominant sub-culture from continental defence (defensive neorealism) to forward defence. Similarly, application of internationalism via the Western global rules-based order is evident. While much of cyberspace remains in the virtual world, notions of traditional friend/foe calculus and geography remain important. Australian Defence Force application of cyber warfare is limited to military operations, with planning and approval according with conventional capability. Other elements, such as application of traditional strategic thought and deeper understanding of the legal framework, manifesting as legislative over common law comprehension, are viewed as less important.

### **4.2 Limitations and future direction/research**

Analysis of policy is complicated by the often-classified nature of cyber warfare capability. The lengths to which a nation is willing to openly discuss cyber behaviours will influence validity, potentially distorting comparative research. For example, in the Australian context, knowledge of an offensive capability remained out of public discourse until revealed in unclassified policy (Australia 2016). Analysis of policy and behaviour is limited when examples are excluded from public debate, though the longevity of strategic culture enables generalisations. As such, the development of an Australian offensive cyber capability likely occurred in the period 2000-2016, perhaps representing a marker of a shift in dominant strategic culture from continental defence (defensive neorealism) to forward defence.

The methodology provided herein enables consistent multi-national research, noting further assessment is required, beyond the scope of the PhD, to support multi-national comparison. The methodology will be refined during the PhD research project, largely limited to the Australian context. Primarily, the code list and associated definitions require ongoing revision against the deductive framework. Iterative coding cycles against qualitative elements will strengthen the research. As noted in section 3.3.2 of this paper, the sub-culture schema is important, particularly when expanded to multi-national comparative studies. This research will incorporate and expand on Bloomfield’s concept of the schema for Australia, with future research required to produce a globally supportive schema. The intent of the schema is to enable researchers in this sub-field to test for evidence of sub-cultures and record relative strength and influence over time. Further research may then analyse a combination of nations’ strategic cultures over time to determine alignment and contention. The outcomes of this may provide policy experts with opportunity to explore opportunity through alignment in dominant or lesser sub-cultures.

## **5. Conclusion**

Drawing on a 40-year scholarly effort, we may construct with confidence, a theory of national strategic culture for military strategy in cyberspace. The paper finds that strategic cultural elements rooted in the past continue to strongly influence cyber strategy, including geography, conventional strategy and existing institutions.

This paper provides a consistent fourth generation methodology for comparative analysis of the influence of strategic culture on cyber warfare. Limited research in a cross-discipline field with inconsistencies and disagreements has produced little to establish academic rigour worthy of the attention of political and military elites. The field is unlikely to gain traction until a consistent and coherent methodology is applied, with comparative research outputs enabling discussion within the elite.

The mixed methods framework enables an understanding of the roles of elites in policy formation and provides stratification to demonstrate the influence of the culture on the broader defence population. Furthermore, the methodology supports comparative static and longitudinal studies of nation states in cyberspace to enable policy makers to coalesce understanding of strategic cultures for the establishment of international norms.

## **Acknowledgements**

Acknowledgement and thanks to Prof Greg Austin, UNSW Canberra Cyber, for his guidance on strategic culture and cyber security. Thanks also to Mr Hywel Evans, cyber law expert, and Ms Karine Pontbriand, fellow PhD student, for their contribution to editing.

**Disclaimer:** *This paper does not represent the views of the Australian Government nor the Australian Department of Defence.*

## **References**

- Adler, E 2012, 'Constructivism in International Relations: Sources, Contributions, and Debates', in W Carlsnaes, T Risse & BA Simmons (eds), *Handbook of International Relations*, 2nd edn, Sage, England.
- Australia, *Defence 2000: Our future Defence Force*, 2000, Do Defence, Commonwealth of Australia, Canberra.
- , *Australia's Cyber Security Strategy*, 2016, DPMC, Commonwealth of Australia, Canberra, Australia.
- Betts, RK 1997, 'Should Strategic Studies Survive?', *World Politics*, vol. 50, no. 1, pp. 7-33,
- Bloomfield, A 2011, 'Australia's Strategic Culture: An investigation of the concept of strategic culture and its application to the Australian case', Queen's University.
- 2012, 'Time to Move On: Reconceptualizing the Strategic Culture Debate', *Contemporary Security Policy*, vol. 33, no. 3, pp. 437-61,
- Booth, K & Trood, R (eds) 1999, *Strategic Cultures in the Asia-Pacific Region*, Macmillan Press Ltd, Great Britain.
- Cresswell, JW 2014, *Research Design: Qualitative, Quantitative and Mixed Methods Approaches*, 4 edn, Sage Publications Inc, Los Angeles.
- Fink, A 2017, *How to conduct surveys: a step-by-step guide*, 6 edn, Sage, Los Angeles, USA.
- Gray, CS 1999, 'Strategic culture as context: the first generation of theory strikes back', *Review of International Studies*, vol. 25, no. 1, pp. 49-69, Cambridge University Press, Cambridge Core,
- 2009, 'Out of the wilderness: Prime Time for Strategic Culture', in JL Johnson, KM Kartchner & JA Larsen (eds), *Strategic Culture and Weapons of Mass Destruction: Culturally based insights into Comparative National Security Policymaking*, palgrave macmillan, New York, pp. 221-41.
- Haglund, DG 2014, 'What Can Strategic Culture Contribute to Our Understanding of Security Policies in the Asia-Pacific Region?', *Contemporary Security Policy*, vol. 35, no. 2, pp. 310-28,
- Hood, JC 2007, 'Orthodoxy Vs. Power: The defining traits of Grounded Theory', in A Bryant & K Charmaz (eds), *The SAGE handbook of Grounded Theory*, SAGE publications Ltd, Wiltshire, pp. 151-64.
- Howlett, D 2006, 'The future of strategic culture', <<https://fas.org/irp/agency/dod/dtra/stratcult-future.pdf>>
- Johnson, JL 2009, 'Conclusion: Toward a standard methodological approach', in JL Johnson, KM Kartchner & JA Larsen (eds), *Strategic Culture and Weapons of Mass Destruction: Culturally based insights into comparative national security policymaking*, Palgrave Macmillan, New York, pp. 243-57.
- Johnson, K 2018, 'National Resilience in Cyberspace: The United Kingdom's National Cyber Security Strategy Evolving Response to Dynamic Cyber Security Challenges', University of Glasgow
- Charles University.
- Johnson, LL, Kartchner, KM & Larsen, JA (eds) 2009, *Strategic Culture and Weapons of Mass Destruction*, Palgrave Macmillan, NY, USA.
- Johnston, AI 1995, 'Thinking about strategic culture', *International Security*, vol. 19, no. 4, p. 32, EBSCOhost, tsh, item: 9508175317.
- Libel, T 2016, 'Explaining the security paradigm shift: strategic culture, epistemic communities, and Israel's changing national security policy', *Defence Studies*, vol. 16, no. 2, pp. 137-56,
- Macmillan, A & Booth, K 1999, 'Appendix: Strategic Culture - Framework for analysis', in K Booth & R Trood (eds), *Strategic Cultures in the Asia-Pacific Region*, Macmillan Press Ltd, Great Britain, pp. 363-72.
- Macmillan, A, Booth, K & Trood, R 1999, 'Strategic Culture', in K Booth & R Trood (eds), *Strategic Cultures in the Asia-Pacific Region*, Macmillan Press Ltd, Great Britain, pp. 3-26.

**Andrew Williams**

- Miles, MB, Huberman, AM & Saldaña, J 2014, *Qualitative data analysis : a methods sourcebook*, Edition 3. edn, Thousand Oaks, California : SAGE Publications, Inc., Thousand Oaks, California.
- Payne, G & Payne, J 2004, *Key Concepts in Social Research*, SAGE Publications, London.
- Persoglia, D 2018, 'Between Defence and Offence: An Analysis of the US "Cyber Strategic Culture"', University of Glasgow Charles University.
- Pirani, P 2016, 'Elites in Action: Change and Continuity in Strategic Culture', *Political Studies Review*, vol. 14, no. 4, pp. 512-20,
- Rubin, H & Rubin, I 2005, 'Designing main questions and probes', in *Qualitative Interviewing (2nd ed.): The Art of Hearing Data*, 2nd edn, SAGE Publications, Inc., Thousand Oaks, California, pp. 152-72 viewed 2017/09/11, <<http://methods.sagepub.com/book/qualitative-interviewing>>.
- Schutt, RK 2012, *Investigating the Social World: The Process and Practice of Research*, 7 edn, SAGE Publications Inc., USA.
- Snyder, J 1977, *The Soviet Strategic Culture: Implications for Limited Nuclear Operations*, The RAND Corporation, Santa Monica USA.
- Tosbotn, RA 2016, 'NATO and Cyber Security: Critical junctures as catalysts for change', Universiteit Leiden.
- USDoD, *Joint Publication 3-12 (R), Cyberspace Operations*, 2018, Do Defense, USA.
- Wirtz, JJ 2015, 'Cyber War and Strategic Culture: The Russian Integration of Cyber Power into Grand Strategy', in K Geers (ed.), *Cyber War in Perspective: Russian Aggression against Ukraine*, NATO CCDCOE Publications, Tallinn, Estonia, pp. 29-37 7 May 2018, <[https://ccdcoc.org/sites/default/files/multimedia/pdf/CyberWarinPerspective\\_Wirtz\\_03.pdf](https://ccdcoc.org/sites/default/files/multimedia/pdf/CyberWarinPerspective_Wirtz_03.pdf)>.

# Cyber Warfare, Terrorist Narratives and Counter Terrorist Narratives: An Anticipatory Ethical Analysis

**Richard Wilson**

**Towson University, Department of Philosophy/Computer and Information Sciences,**

**Towson University, USA**

**Hoffberger Center for Professional Ethics, University of Baltimore, USA**

[wilson@towson.edu](mailto:wilson@towson.edu)

**Abstract:** Cyber Jihadists and cyber terrorists, have used the internet and social media platforms in order to engage in an information warfare using a variety of stories, narratives and a master narrative in the attempt to accomplish their goals. Recently white supremacists have used similar methods to encourage terrorist acts. This analysis is aimed at exploring the concepts of cyber warfare, terrorist narratives, a phenomenological theory of intentionality applied to these narratives, counter terrorist narratives and to the ethical problems posed by the construction of counter terrorist narratives. The exploration of these themes will provide a foundation for understanding how white supremacists use similar tactics to attract followers and to inspire terrorist attacks. The study of Cyber Jihad and cyber terrorism can give us insight into how all terrorists including domestic terrorists operate using master narratives and sub narratives. At the outset the effort to describe the conflict between terrorist Islamist extremist narratives and counter terrorists narratives presents us with terminological difficulties. The discussion of cyber warfare presents a number of terminological difficulties when it is applied to events related to terrorists. These difficulties stem from the definition of cyber warfare itself. According to Healey Cyberwarfare is defined as "Action by a nation-state to damage or disrupt another nation's computers or networks, which are heavily damaging and destructive – similar to the effects achieved with traditional military force – and so are considered to be an armed attack. An act of war that is mediated in full or in part through cyberspace." (Healey, James, 2013, p. 15) One difficulty with this definition of cyber warfare is related to claim that actions of warfare have to be perpetrated by nation state actors against other nation state actors. The so called war on terrorism conducted against ISIS and other groups such as Al Qaeda, The Taliban, and ISIS is actually a series of actions perpetuated by state actors against non state actors. ISIS, which is the focus of our analysis of narratives, is an aspiring state engaged in terrorism against all non believers. In the struggle to establish the Caliphate. ISIS engages in kinetic terrorist activities and extends terrorist activities into the cyber domain. The actions of ISIS in the cyber domain employ new media and social networking platforms in what can be called acts of information warfare while encouraging all forms of terrorism. In the tradition of Islamist extremism ISIS uses all of the technologies available to disseminate their message.

**Keywords:** cyber jihad, cyber warfare, cyber terrorism, information warfare, influence operations, ethics, anticipatory ethics

---

## 1. Introduction

To carry out this analysis we will assume the following working definitions of cyber terrorism. Cyberterrorism is defined as "a premeditated, politically motivated criminal act by subnational groups or clandestine agents, against information and computer systems, computer programs, and data, that results in physical violence, where the intended purpose is to create fear in noncombatant targets." (Colarik, 2006) Another more recent definition states that "cyberterrorism is the use of cyber capabilities to conduct enabling, disruptive, and destructive militant operations in cyberspace to create and exploit fear through violence or the threat of violence in the pursuit of political change." (Brickey, Jonalan, 2012) One difficulty with giving a clear account of 'cyber terrorism' is also related to attempting to define 'terrorism'. There is no universally agreed upon definition of terrorism. Cyber terrorism can be defined as being concerned with inspiring fear in an audience which now includes all forms of new media. However, this definition does not go far enough in describing how ISIS uses computer systems to further its agenda.

How is cyber terrorism pursued using the internet and social media platforms? According to Weimann, (Weimann, Gabriel, 2015) terrorists have 3 main reasons for using the internet and social networking platforms. 1<sup>st</sup>, they are the easy way to reach mainstream audiences, 2<sup>nd</sup> social media companies provide a service for free. 3<sup>rd</sup> social media platforms provide a means for social networking which allows terrorists to reach out to target audiences. Social networking platforms allow terrorists to target potential recruits, members and followers, disseminate propaganda and target young people.

According to Jenkins, "Terrorists use the Internet to disseminate their ideology, appeal for support, spread fear and alarm among their foes, radicalize and recruit new members, provide instruction in tactics and weapons,

gather intelligence about potential targets, clandestinely communicate, and support terrorist operations. The Internet enables terrorist organizations to expand their reach, create virtual communities of like-minded extremists, and capture a larger universe of more-diverse talents and skills." (Jenkins, B. M , 2011) The internet has been employed by ISIS for a variety of activities including to announce and broadcast its terrorist agenda, to spread its influence, to recruit new members as well as to generate funding. What is taken as central in what follows is how social media help to organize how terrorists use the internet with the spread of narratives being employed as a way to organize their activities.

## **2. Narratives**

At the center of the cyber activities pursued by terrorist groups is the notion of a 'narrative'. According to Halverson (Halverson et al. 2011), a narrative is a system of stories that relate to one another with a coherent theme where the whole of the stories exceeds the sum of its parts. (p. 1) This set of stories is joined together in what is called a 'master narrative'. A master narrative in turn is "deeply embedded in a culture, provides a pattern for cultural life and social structure and creates a framework for communication about what people are supposed to do in certain situations. Like all narratives, master narratives have components such as story forms and archetypes that can be used to understand their structures." (Halverson et al., 2011 p.7)

A master narrative is an overarching narrative that gathers together a group of narratives while presenting the ideology of the group committed to the narrative. The master narrative gathers together a series of stories and gives them a coherence by organizing them in a master narrative. Master narratives are described in the following way, "a master narrative is a transhistorical historical narrative that is deeply embedded in a particular culture. By 'transhistorical' we do not mean that master narratives are 'born' as such. In fact, 'they grow up' to attain that stature over time through repetition and reverence in a particular culture." (Halverson et al., 2011). A master narrative gains significance when the narrative is repeated over and over again as part of what can be called an indoctrination process. One way to approach the influence of stories and narratives and the ideas they introduce to audiences is through literary theory. Taking an approach from literary theory, a narrative can be defined as a mode of fiction in which the narrator is communicating directly to the reader. A broader definition of a narrative would involve the notion of storytelling. According to Owen Flanagan, "Evidence strongly suggests that humans in all cultures come to cast their own identity in some sort of narrative form. We are inveterate storytellers."

(Flanagan, Owen,1995) Flanagan suggests that stories play an important role in everyday existence, which would mean that narratives can play the same role.

Stories are an important part of every culture and society, but in the context of terrorist narratives, they become an important means for perpetuating terrorist ideology and culture. The stories and narratives of Islamic fundamentalists gain legitimacy through the reverence accorded to the stories and through the repetition of these stories as well as through the authority of the teller of the story. Further understanding of the influence of stories and narrative can be gained by employing distinctions from the phenomenological tradition in philosophy. According to this tradition consciousness is intentional. (Husserl, Ideas) This means story telling is intentional while at the same time the stories and narratives also have an intention.

## **3. Stories, story telling and story told**

With stories there is a basic structure, there is a story teller, there is the telling of the story, there is the story told and there is an audience. These distinctions can be schematized as follows:

Story teller      Telling of the story      Story Told      Audience

Each of these concepts can be explored through phenomenological reflective analysis (Embree, Lester, 2011) and can be examined as they relate to terrorist narratives. Terrorist narratives are stories that are directed towards audiences.

### **a. Story Teller**

The role of the teller of the story is important in the telling of the master narrative and related narratives. The master narrative gains stature because of the authority of the story teller. With terrorist narratives the story

teller is an authority. A story teller's perspective is important when a story is told to an audience. There are important distinctions to be made between first and third person narratives. Generally, a first-person narrator brings greater focus on the feelings, opinions, and perceptions of a particular character in a story, and on how the character views the world and the views of other characters. If the writer's intention is to get inside the mental world of a character, then it is a good choice, although a third-person narrator is an alternative that does not require the writer to reveal all that a first-person character would know. By contrast, a third-person omniscient narrator gives a panoramic view of the world of the story, looking into many characters and into the broader background of a story. The master narrative of extremists is told from 2 perspectives. 1st it is told from a 3<sup>rd</sup> person perspective while, 2nd, it is also told from a 1<sup>st</sup> person perspective by those engaged in waging Jihad and those engaged in terrorist activities, or through the testimony of future suicide bombers who often record tapes told in the 1<sup>st</sup> person the suicide bombing occurs.

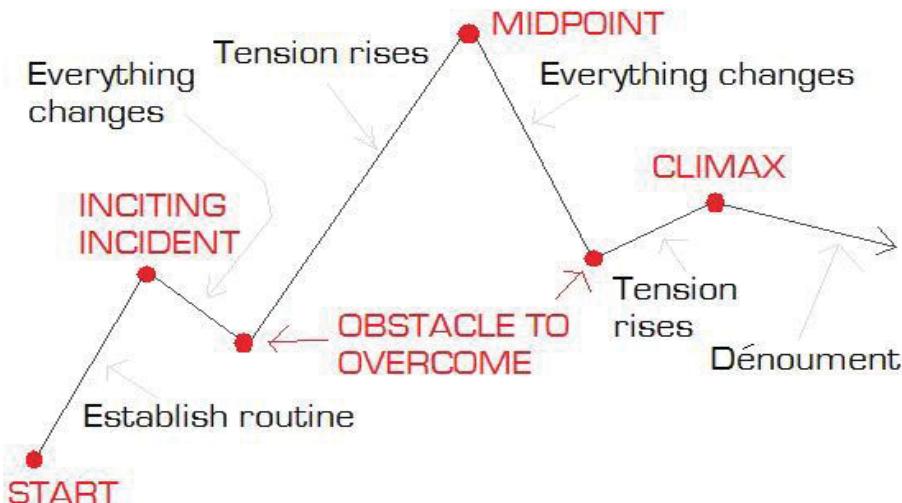
b. Telling of the story

Terrorist narratives are used by story tellers to engage in storytelling to gain the attention of the widest range of audiences. When members of terrorist groups such as ISIS recount stories about a variety of subjects including terrorist attacks and suicide bombings, they have the intention of making the terrorist act the center of the attention of the audience and of the media. The purpose of recounting stories of terrorist kinetic and cyberwarfare actions and activities can be interpreted as having the intention of having the stories that are told, be at the center of an audience's and the media's attention. One way to view terrorism is to interpret it as taking place within the context of stories and narratives told by the proponents of terrorism. Authoritative story tellers employ an overarching master narrative to bring a cohesiveness to the stories and narratives and present this master narrative to a wide range of audiences. The goal of these activities is to gain followers, get financial support and to recruit new members. The master narrative can also combine 4 separate types of narratives, which include a political narrative, a moral narrative, a religious narrative, and social-psychological narrative (Leuprecht et al. 2010),

c. Story Told

A narrative can be defined as a report of events that are presented to an audience arranged in a logical event based sequence. A story is often taken to be a synonym of a narrative. A narrative or story is told by a story teller who may be a direct part of the terrorist experience, in which case he or she often shares the experience as a first-person narrator. As stated above the story teller may also only observe the events as a third-person narrator and gives his or her verdict from an overarching perspective of a master narrative. One approach that can be taken for understanding the structure of stories and narratives can be found in Aristotle's Poetics (Aristotle, Poetics). The terrorist master narrative has a number of features in common with a Greek tragedy. An important feature of the master narrative is the focus on the series of 'tragedies' that have befallen Islamic Culture. Just as the audience of a Greek tragedy is invited to reflect upon the sequence of events occurring in a Greek Tragedy, so audiences at which the master narrative is aimed, are invited to reflect upon the events from a 3<sup>rd</sup> person perspective told by the story teller in the 1<sup>st</sup> person of extremist jihadi narratives. The story tellers of the grand narrative invite an audience to reflect upon the grand narrative of the extremist jihadi stories and master narrative. The elements involved in telling a story in order to create an impact is depicted below. (Malby,F. C. Narrative Arc Shaping Your Story)

Aristotle also describes the elements of successful imitation that relate to those who will imitate the action of a hero of the narrative. The poet or story teller must imitate either things as they are, things as they are thought to be, or things as they ought to be. The master narrative combines these elements. The poet must also imitate in action and language (preferably metaphors or contemporary words). Errors come when the elements of the story are imitated incorrectly - and thus destroys the essence of the story. Imitation is a crucial ingredient when audiences are encouraged to perform actions described in terrorist narratives.



## THE STORY ARC

### 4. The story teller

A narrative can be defined as a report of events that are presented to an audience arranged in a logical event based sequence. A story is often taken to be a synonym of a narrative. A narrative or story is told by a story teller who may be a direct part of the terrorist experience in which case he or she often shares the experience as a first-person narrator. The story teller may also only observe the events as a third-person narrator and gives his or her verdict from an overarching perspective of a master narrative. Another approach that can be taken for understanding the structure of stories and narratives can be found in Aristotle's Poetics. The terrorist master narrative has a number of features in common with a Greek tragedy. An important feature of the master narrative is the focus on the series of 'tragedies' that have befallen Islamic Culture. Just as the audience of a Greek tragedy is invited to reflect upon the sequence of events occurring in a Greek Tragedy, so audiences at which the master narrative is aimed, are invited to reflect upon the events from a 3<sup>rd</sup> person perspective told by the story teller in the 1<sup>st</sup> person of extremist jihadi narratives. The story tellers of the grand narrative invite an audience to reflect upon the grand narrative of the extremist jihadi narrative.

Story teller	Telling of the story	Story Told
--------------	----------------------	------------

Although there are many story tellers who tell the story of the terrorist master narrative there are a number of ways in which it is told. Bin Laden had a specific way of telling stories and of recounting the master narrative. His talks were focused on a solitary figure sitting in one position. Anwar Al-Jawlaki was a proficient public speaker who spoke English which allowed him to attract non Arabic speakers. He made public lectures and produced tapes and videos that were distributed across the internet, that extended Jihadist ideas beyond the limitations of bin Laden. ISIS uses a multimedia approach that goes beyond either of these earlier examples, "With the use of smartphone apps, social media sites, and free file hosting, the organization exploited most of the commonly available online communication tools. (Nance, Malcolm and Christopher Sampson, Hacking ISIS, 2017)

### 5. Telling of the story

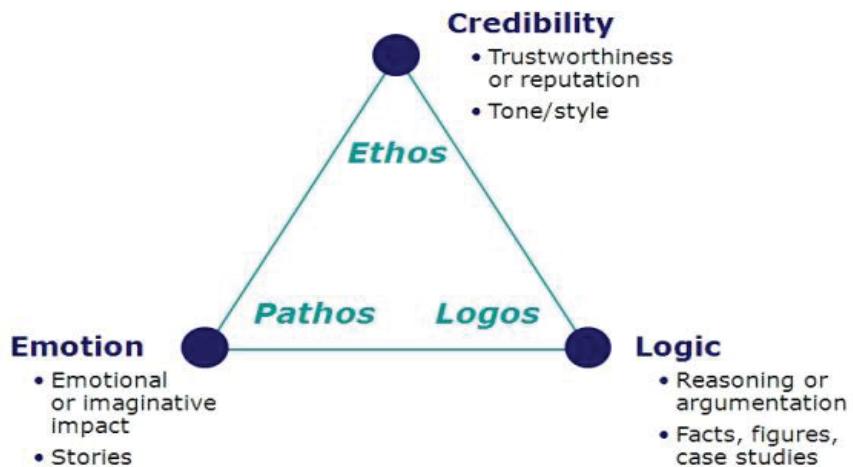
There are a number of elements that play a role in the influence created by the Jihadist master narrative but these narratives rely for their success upon elements that can also be related to what was identified by Aristotle in his Rhetoric. (Aristotle, Rhetoric, 2018) The Rhetoric is concerned with clarifying the elements involved in persuading an audience. The narratives told by extremists employ the same elements that Aristotle employs in the Rhetoric. These elements are ethos, logos, and pathos. Narratives, like any other stories, in order to be active, rely on rhetoric elements and especially on the three components offered by Aristotle in his Rhetoric; (1)ethos (2) logos, and (3) pathos. How are each of these elements defined?

Ethos: refers to the appeals that the story teller makes to the audience to establish credibility, and to ensure that the audience can trust him or her.

Logos: appeals to facts such as events, traumas, conflicts, depravation, or inequalities.

Pathos: refers to the appeals that the story teller makes to the audience's emotions, including anger, hate , pride, and frustration.

The following diagram (Aristotle's rhetorical triangle) captures the elements central to persuading an audience in Aristotle's Rhetoric, techniques which are also being used by story tellers telling extremist stories who employ these same elements to convince audiences when telling the master narrative of the legitimacy of that narrative.



## 6. The story told and the master narrative

The story that is told is actually a combination of stories that are gathered together and organized by an overarching master narrative. The themes discussed in the master narrative are deeply embedded in Muslim culture. What are the elements of the master narrative according to Islamist extremists that are deeply embedded?

- a. The Master Narrative (Adapted from Weimann, Gabriel. Terrorism in Cyberspace)
  - 1. an account of 'The golden age of Islam' which occurred during the time of the Prophet.
  - 2. There is an account of cultural pollution that led to the end of the golden age.
  - 3. There has been a repression and ethnic cleansing of Muslims since the loss of the Ottoman Caliphate, with repression due to the Gulf War, Afghanistan and Iraq.
  - 4. There is a Zionist/Crusader Axis with Islam as the main enemy.
  - 5. The experience of defeat has had an affect upon the global (ummah) community and to Westernization of Muslim life.
  - 6. Islam is under attack and this attack is led by the US and other western crusaders while jihadists and terrorists are defending against this attack.
  - 7. Martyrdom and The End of Days

## 7. Audience

Extremist narratives are directed towards a number of audiences and each narrative is aimed at convincing each of the audiences of the truthfulness of the master narrative. The same elements as those identified in Aristotle's Poetics and Rhetoric have been skillfully employed to convince a wide range of audiences to sympathize and emphasize with the themes that are part of the master narrative. There are actually a number of audiences and

stakeholders at which terrorist narratives and the terrorist master narrative are aimed. Terrorist groups are concerned with trying to gain the attention of the following groups of audiences. These audiences include the following groups:

Among Muslims

- Non sympathetic Muslims in Muslim nations
- Neutral Muslims in Muslim nations
- Sympathetic Muslims in Muslim nations
- Sympathetic Muslims in non Muslim nations
- Supporters of Terrorism
- Those sympathetic to terrorism
- Vulnerable audiences
- Potential converts
- The World audience

One of the main goals of the narratives and master narratives of terrorists is to Inspire fear in all audiences but they are also aimed at recruiting new members. Phenomenological reflection can be employed to see what cajols audiences in the direction of radicalization, the creation of empathy and sympathy for the story tellers, and empathy with the themes discussed in the master narrative. The purpose of the master narrative is to direct audiences away from any sense of empathy with non believers in the Master Narrative and recruit them into sympathizing with terrorist groups and the goals of terrorism.

## **8. Radicalization**

Radicalization and self radicalization are major problems in both global terrorism and domestic terrorism. To understand the structure of radicalization. The U.S. Senate Committee on Homeland Security and Governmental Affairs lists the following stages in the radicalization process.( U.S. Congress, Homeland Security and Governmental Affairs Committee, 2008)

- 1. Pre-Radicalization: ... [T]he point of origin for individuals before they begin the radicalization process. It is their life situation before they were exposed to and adopted jihadi-Salafi [ideology] ... as their own ideology.
- 2. Self-Identification: ... [T]he phase where individuals, influenced by both internal and external factors, begin to explore Salafi Islam, gradually gravitate away from their old identity, and begin to associate themselves with like-minded individuals and adopt this ideology as their own.
- 3. Indoctrination: ... [T]he phase in which an individual progressively intensifies his beliefs, wholly adopts jihadi-Salafi ideology and concludes, without question, that the conditions and circumstances exist where action is required to support and further the cause... While the initial self identification process may be an individual act, ... association with likeminded people is an important factor as the process deepens.
- 4. Jihadization: ...[T]he phase in which members of the cluster accept their individual duty to participate in [terrorist activities] and self-designate themselves as holy warriors or mujahedeen. Ultimately, the group will begin operational planning for the ... terrorist attack. These “acts in furtherance” will include planning, preparation and execution.

Radicalization is a process through which individuals, or groups come to adopt increasingly extreme political, social, or religious ideals. The master narrative of the terrorist group attempts to recruit new members are in conflict with and have aspirations that reject or undermine the status quo or contemporary ideas and expressions of targeted nations.

## **9. The master narrative: Story told**

Through ideas, values and images related to narratives terrorist organizations project propaganda and promote the development of ideas that are aimed at underwriting processes of radicalization which are deployed through online platforms. Narratives and the master narrative give their audience a sense of where they came from and towards where they are going. As we have seen the Master Narrative of jihadi extremists employs a number of

elements already found to be at work in ancient tragedies and which are described in Aristotle's Poetics. These elements play an important role in and contribute to the persistence of terrorist propaganda, seduction and recruiting. The grand narrative of jihadi extremism also employs elements from Aristotle's rhetoric to persuade audiences.

A number of authors present slightly different views of what is contained within the master narrative projected by Islamic extremists. According to Weimann (Weimann 2015) the core elements in radical narratives include:

- 1. An account of 'The golden age of Islam' which occurred during the time of the Prophet. "This age, which was characterized by harmony, peace and justice forms the millennial vision to which violent radicals desire to return." (Weimann 2015)
- 2. There is an account of cultural pollution that led to the end of the golden age. "Cultural pollution accumulated cultural practices or innovation, moral decay, and Muslim "sinfulness" did to end the golden age." (Weimann 2015)
- 3. There has been a repression and ethnic cleansing of Muslims since the loss of the Ottoman Caliphate, with repression due to the Gulf War, Afghanistan and Iraq. Defeats included, "the loss of statehood (the Ottoman caliphate) in the early 20<sup>th</sup> century; Palestine displacement and the loss of Jerusalem in the mid 20<sup>th</sup> century; repression in areas such as North Africa, Egypt, and Chechnya; ethnic cleansing in the Balkans; infidel occupation of holy places in Saudi Arabia since the 1991 Gulf war; and the occupation of Afghanistan in 2001 and Iraq in 2003." (Weimann 2015)
- 4. There is a Zionist/Crusader Axis with Islam as the main enemy. "The United States, Israel, and the United Kingdom/European Union are seen as being united in a 'Zionist'/'Crusader' axis, with Islam as the main enemy. The populations of these countries are seen as 'collectively guilty' and therefore are not subject to the norm of civilian protection. This narrative has moved from engagement with the 'near' enemy (e.g., expelling the West from the Middle East). (Weimann 2015)
- 5. The experience of defeat has had an affect upon the global (ummah) community and to Westernization of Muslim life.
- 6. Islam is under attack and this attack is led by the US and other western crusaders while jihadists and terrorists are defending against this attack.
- 7. Martyrdom and The End of Days

The terrorist master narrative gathers together a series of stories and narratives that are never presented by jihadists, Islamist extremism and the story tellers as terrorist narratives, there is a normalizing of the terrorist agenda. The master narrative of Jihadists normalizes their extreme views. As argued above the master narrative has many elements that are reminiscent of Greek tragedy. However, the master narrative instead gives the appearance of describing a sequence of events, just as stated in Aristotle's poetics, where there is a unifying theme that is centered on the idea that an unjust global war is being waged upon Islam. The unjust war is described from the perspective of 4 narratives, political, moral, religious and social-psychological. What gives the master narrative legitimacy is a discussion by authority figures of universal themes that attract and then reinforce ideas within sympathetic audiences.

## **10. Ethical issues**

According to James Rachels (Rachels, James. *The Elements of Moral Philosophy*), "Morality is, at the very least, the effort to guide one's conduct by reason – that is, to do what there are the best reasons for doing – while giving equal weight to the interests of each individual affected by one's action." This conception of morality gives a fundamental picture of what it means to be a conscientious moral agent. A moral agent "is someone who is concerned impartially with the interests of everyone affected by what he or she does; who carefully sifts facts and examines their implications; who accepts principles of conduct only after scrutinizing them to make sure they are justified; who is willing to 'listen to reason' even when it means revising prior convictions; and who, finally, is willing to act on these deliberations."

There are a number of issues that arise with understanding terrorist narratives but also with the idea of constructing counter terrorist narratives.

- 1. Extremists are concerned with drawing the attention of a wide range of audiences. To contest a terrorist narrative by disputing facts and by asking an audience to reason critically about the terrorist narrative requires a sophistication of the part of the audience.
- 2. Can the 5 strategies suggested by Halverson et al. (2011) succeed?
- 3. Can the depoliticizing suggested by Rane succeed?
- 4. To recast the conflict of narratives as a conflict between stories is to see some of the difficulties related to constructing counter terrorist narratives. Stories are often fictive accounts of events. In the context of the terrorist master narrative the stories are at the levels of myth for the believers. The problem of constructing a counter terrorist narrative must confront the issue of the counter narrative also being a story. Counter terrorist strategy needs to incorporate all of these factors into any new method suggested for countering terrorism.
- 5. To recast the conflict of narratives as a conflict between stories is to see some of the difficulties related to constructing counter terrorist narratives.
- A. *Stories are often fictive accounts of events. In the context of the terrorist master narrative the stories are at the levels of myth for the believers in the master narrative.*
- B. *The problem of constructing a counter terrorist narrative must confront the issue of the counter narrative also being a story.*
- C. *Counter terrorist strategy needs to incorporate all of these factors into any new method suggested for countering terrorism.*

Another approach would be to construct counter terrorist narratives using the techniques employed above to analyze extremist narratives. The ideas of Aristotle and Phenomenological reflection can be deployed to understand terrorist narratives but also to construct counter terrorist narratives. This would include the idea that counter terrorist narratives must focus on the story teller, storytelling, story audience interconnection. Authorities must tell this counter terrorist story and engage in identifying how the stages of radicalization need to be reversed through literary techniques.

## **11. Anticipated ethical issues**

Technologies and Artefacts alter social existence. The Internet has greatly altered the trajectory of extremism and terrorism. While ISIS engages in kinetic activities involving extreme aggression there is also a sophisticated effort to work simultaneously on the Internet in Cyberspace. Hybrid conflict employs and combines kinetic and cyber means to accomplish their goals. How is the Internet used? To disseminate propaganda and ideas, to recruit, to stockpile training manuals and videos. The use of the internet and social networking platforms has led to the “Rise of the Lone Wolf”. The emerging threat in terrorism is perhaps better described as the rise of the “individual actor terrorist” as opposed to the Lone Wolf terrorist. The internet and social media allow for self radicalization and connections with other actors through social media postings, and now through video capture and that is occurring as the event is taking place. We can now anticipate ethical issues related to the rise of individual actor terrorists.

- (1) Individual actor terrorists and lone wolf terrorists are remnants of successful combatted and disrupted group actor terrorism.
- (2) The internet is a vehicle for the dissemination of extreme ideologies and has influenced individual actor terrorists.
- (3) Individual actor terrorists and LW empirical data is lacking. Data has to be gathered on these ongoing threats in order to better understand how they self radicalize online.
- (4) 1st hand accounts and reports related to individual actor terrorists are missing.
- (5) Individual actor terrorists and LWS combine broad structures of prevalent extreme ideology with personal grievances.
- (6) Motivational patterns of individual actor terrorists and LWS shift over times – they are not static ideal types.
- (7) The espousal of a particular ideology alone does not guarantee that radicalization occurs towards terrorist violence.

- (8) There is no single personality type of the individual actor terrorists and LWT.
- (9) Mental Health Issues – Psychological disorder does not cause individual actor terrorists and LWT to become cognitively disordered.
- (10) Individual actor terrorists and LWT's exhibit varying degrees of social ineffectiveness and are often socially isolated.
- (11) Responses to individual actor terrorists and LWTs should be based on democratic principles and respect for human rights and adequate safeguards need to be built into responses, to prevent erosion of civil liberties.

## **12. Recommendations**

One item we can take for granted is that technology will continue to rapidly develop. The study of how Individual actor terrorists and lone wolf terrorists self-radicalize using the latest technology must continue to be studied. We need to study historical cases as well as present cases in order to anticipate future developments. This is the best avenue for developing ideas related to anticipatory analysis where from anticipating ethical issues we develop strategies for developing policy. This analysis provides perhaps the important avenue for developing emerging norms that will help us understand individual actor terrorists and lone wolf terrorists.

Are there new norms are there new ethical issues that are technology and artefact driven?

Examples:

- 1. ISIS use of Internet and Social Media that is employed to recruit, disseminate and stockpile video's and manuals.
- 2. An emerging type of cyber aggression is state sponsored hacktivism.
- 3. Individual actors can be state sponsored, can be group sponsored and can self-radicalize with little or no contact with others
- 4. Counter terrorist narratives need to be developed according to the same structure that is employed by those constructing and deploying the terrorist master narrative. Literary and philosophical texts such as the work of Aristotle can be used to gain insight into how to counter extremist narratives.

The strongest conclusion that emerges is rethinking the idea of lone wolf terrorist and focusing on individual actor terrorists. Deeper analysis must be conducted in order understand how individual actors use the internet and the latest technology to self radicalize and eventually perform terrorist actions. Most important is perhaps the idea that domestic terrorism needs to be more deeply studied, more clearly defined and identified as a current threat.

## **References**

- Allam, Hannah. Is imam a terror recruiter or just an incendiary preacher? MCCLATCHY Newspapers (November 22, 2009). <https://www.mcclatchydc.com/news/nation-world/world/article24564601.html>
- Al Raffie, Dina. 2012."Whose Hearts and Minds? Narratives and Counter-Narratives of Salafi Jihadism." Journal of Terrorism Research 3 (2):13-21.
- Aly, Anne, Stuart McDonald, Lee Jarvis and Thomas Chen, Violent Extremism Online: New perspectives on terrorism and the Internet, Routledge, New York, 2016.
- Archetti, Cristina. Understanding Terrorism in the Age of Global Media. Palgrave, Mcmillan, NY,2012.
- Betz, David, 2008. "The Virtual Dimension of Contemporary Insurgency and Counterinsurgency." Small Wars & Insurgencies 19 (4): 510-40.
- Aristotle. Poetics. Penguin Classics, 1997.
- Aristotle, Rhetoric, Hackett Publishing Company, 2018
- Aristotle's rhetorical triangle, <http://bbwbettiepumpkin.com/pathos-ethos-logos/> Jan. 13,2019.
- Brickey, Jonalan, Defining Cyberterrorism: Capturing a Broad Range of Activities in Cyberspace, Combating Terrorism Center at West Point, August 23, 2012 <https://www.ctc.usma.edu/posts/defining-cyberterrorism-capturing-a-broad-range-of-activities-in-cyberspace>.
- Casebeer, William A., and James Russell, 2005. "Storytelling and Terrorism: towards a Comprehensive 'Counter-Narrative Strategy.'" Strategic Insights 6 (3), <http://calhoun.nps.edu/bitstream/handle/10945/11132/casebeerMar05.pdf>
- Change Institute. 2008. Studies in Violent Radicalization: The Beliefs. Ideologies and Narratives. February. European Commission-directorate-General for Justice, Freedom and Security. [http://www.changeinstitute.co.uk/images/publications/changeinstitute\\_beliefsideologiesnarratives.pdf](http://www.changeinstitute.co.uk/images/publications/changeinstitute_beliefsideologiesnarratives.pdf).

**Richard Wilson**

- Colarik, Andrew M. *Cyber Terrorism*, Idea Group Publishing, USA, 2006, p.47.
- Dal Cin, Sonia, Mark P. Zanna, and Geoffrey T. Fong. 2004 "Narrative Persuasion and Overcoming Resistance." In *Resistance and Persuasion*, edited by Eric S. Knowles and Jay A. Linn, 175-92. Mahwah,NJ: Erlbaum.
- Embree, Lester, *Reflective Analysis*, Zeta books, Bucharest, 2011.
- Fink, Naureen Chowdhury, and Jack Barclay. 2013. Mastering the Narrative Counterterrorism Strategic Communication and the United Nations. Washington DC: Center on Global Counterterrorism Cooperation.  
[http://www.globalcenter.org/wp-content/uploads/2013/03/Feb2013\\_CT\\_StratComm.pdf](http://www.globalcenter.org/wp-content/uploads/2013/03/Feb2013_CT_StratComm.pdf).
- Flanagan, Owen. *Consciousness Reconsidered*, MIT Press, 1995.
- Friedman, Thomas.2009. "America vs. The Narrative." New York Times. November 28.  
<http://www.nytimes.com/2009/11/29/opinion/29friedman.html>.
- Gartenstein-Ross, David, and Laura Grossman. 2009. Homegrown Terrorists in the U.S. and U. K.Washington, DC: FDD.
- Halverson, Jeffrey, Steven Corman, and H. L. Goodall.2011. *Master Narratives of Islamic Extremism*. New York: Palgrave Macmillan.
- Holtmann, Philipp. (2013) "Countering Al-Qaeda's Single Narrative." *Perspectives on Terrorism* 7 (2): 141-146.
- Husserl, Edmund. *Ideas for a Pure Phenomenology and Phenomenological Philosophy: First Book: General Introduction to Pure Phenomenology*. Hackett Publishing Company. 2014.
- Jacobson, Moichael, *Terrorist Dropouts: Learning from those who have Left*, The Washington Institute for Near East Policy, *Policy Focus #101*, January 2010
- Jenkins, B. M. "Is Al Quaeda's Internet Strategy Working? Santa Monica, CA: RAND Corporation. 2011.
- Lentini, Peter, (2013) *Neojihadism: Towards a New Understanding of Terrorism and Extremism?* Cheltenham, UK: Edward Elgar, publishing.
- Lia, Brynjar. 2008. "Al-Qaida's Appeal: Understanding Its Unique Selling Points." *Perspectives on Terrorism* (2) 8 3-10.  
<http://www.terrorismanalysts.com/pt/index/phppot/article/view/44/html>.
- Leuprecht et al. 2010 Containing the Narrative: Strategy and Tactics in countering the Storyline of Global Jihad." *Journal of Policing, Intelligence*
- Malby,F. C. *Narrative Arc Shaping Your Story*, <https://fcmalby.com/2014/05/14/narrative-arc-shaping-your-story/>.
- Mazzoco, Philip J., and Melanie C. Green . 2011. "Narrative Persuasion in Legal Settings: What's the Story?" *The Jury Expert* 23 (3): 27-38. <http://www.thejuryexpert.com/2011/05/narrative-persuasion/>.
- Office for Security and Counterterrorism. 2009. "The United Kingdom's Strategy for Countering International Terrorism." Home Office of the United Kingdom. <http://www.official-documents.gov.uk/document/cm75/7547/7547.asp>.
- Presidential Task Force, 2009. *Rewriting the Narrative: An Integrated Strategy for Counterradicalization*, March.
- Washington, D. C.: Washington Institute for Near East Policy.
- Nance, Malcolm and Christopher Sampson. *Hacking ISIS: How to Destroy the Cyber Jihad*, 2017.
- Rabasa, A and Benard,C (2015) *Eurojihad: Patterns of Islamist Radicalization and Terrorism in Europe*. Cambridge: Cambridge University Press.
- Rachels, James and Stuart Rachels. *The Elements of Moral Philosophy*, McGraw Hill Education, 8<sup>th</sup> ed. 2015.
- Rane, Halim, "Narratives and Counter-narratives of Islamist extremism" in Aly, Anne, Stuart McDonald, Lee Jarvis and Thomas Chen, *Violent Extremism Online: New perspectives on terrorism and the Internet*, Routledge, New York, 2016.
- Schmid, Alex 2010. "The Importance of Countering Al-Qaeda's 'Single Narrative.'".
- Slater, Michael, 2002. "Entertainment Education and the Persuasive Impact of Narratives." In *Narrative Impact: Social and Cognitive Foundations*, edited by Melanie C. Green, Jeffrey J. Strange, and Timothy C. Brock, 157-81. Mahwah, NJ: Erlbaum.
- U.S. Congress, Homeland Security and Governmental Affairs Committee, *Violent Islamist Extremism, the Internet, and the Homegrown Terrorist Threat: Majority and Minority Staff Report; Joseph Lieberman, chairman, Susan Collins, ranking minority member*, 2008, 4.
- Weimann, Gabriel. *Terrorism in Cyberspace*, Woodrow Wilson Center Press / Columbia University Press 2015.
- World Public Opinion. 2009. *Public Opinion in the Islamic World on Terrorism,,Al Qaeda and US Policies..* February 25.  
WorldPublicOpinion.org.

# **Cambridge Analytica, Facebook, and Influence Operations: A Case Study and Anticipatory Ethical Analysis**

**Richard Wilson**

**Towson University, Department of Philosophy/Computer Science and Information Sciences, Towson University, USA**

**Hoffberger Center for Professional Ethics, University of Baltimore, USA**

[wilson@towson.edu](mailto:wilson@towson.edu)

**Abstract:** This discussion is aimed at analyzing the influence operations carried out by Cambridge Analytica in Facebook focusing on the activities performed during 2016. After the 2016 worldwide elections, it is important to note the role that Influence operations will potentially play in influencing national elections as well as international geopolitical relationships. We now need to be aware of how influence campaigns and influence operations will play an increasingly important role in competition between nation states. Anticipatory ethics and the identification of the ethical issues related to what is now referred to as the Facebook – Cambridge Analytica scandal can provide a fundamental basis for determining policy about social media in order to mitigate influence operations as part of information and cyber warfare operations. This analysis interprets influence operations from the perspectives of a variety of stakeholders in order to explore what this emerging type of social media warfare, may look like in the future. The object of this research is to identify the ethical issues with the types of influence operations carried out in the Facebook – Cambridge Analytica scandal and to use the insights gained in order to formulate policy about social media. Recommendations will be made from this research about what should be of concern for policy makers about influence operations in the future.

**Keywords:** Facebook, Cambridge Analytica, cyber warfare, information warfare, influence operations, ethics, anticipatory ethics

---

## **1. Introduction**

This analysis is a case study and anticipatory ethical analysis of Facebook, Cambridge Analytica (CA), and influence operations. (For a general background account of the (CA) scandal see: Bloomberg April 10, 2018.) The goal of this case study and research is to provide an example of how non-state actors including corporations and private companies can influence political events through influence operations. The mechanisms involved in this influence operation are now capable of being used in competition between nation-state actors and need to be studied as part of cyber warfare strategies if the complete range of cyber strategy is to be understood. This analysis is aimed at examining the role of the influence operations of Cambridge Analytica in connection with the 2016 United States political campaign and presidential election, by studying the effect of and ethical issues arising from the Cambridge Analytic – Facebook scandal.

The Cambridge Analytica scandal gives an example of how Cyber warfare can function simultaneously as information warfare, which is focused on influence campaigns and operations. This is exhibited in how Cambridge Analytica employed Facebook to attempt to influence individuals through influence operations. This involved constructing false narratives that revolved around presenting false statements as facts. The presentation of false statements as true is a fundamental strategy of Information warfare. (see:Underwood, Kimberly. A New Front in Information Warfare) This was done as part of a broader cyber warfare campaign aimed at getting political power. According to Clausewitz, " War is not an independent phenomenon, but the continuation of politics by different means." (Clausewitz) Influence campaigns can function within information warfare as a means for gaining political power.

What has most recently been seen in competition between adversaries at the geopolitical level, is the use of influence operations and campaigns to attempt to have a psychological impact upon civilian populations. Psychological influence which is the focus of psychological operations, are directed at attempting to influence thoughts, ideas, and behaviors, in targeted populations. Influence operations were deployed in the United States 2016 presidential election which involved influence campaigns where Facebook was used by Cambridge Analytica in the effort to influence voters and create dissension within the civilian population. These influence campaigns were aimed at an effort to cause the civilian population to change its ideas or to endorse already believed in ideas, in the effort to sway voter behavior and the outcome of the U. S. presidential election. While this may not seem to be a subject that needs to be addressed at the same level as cyber warfare, it might be more accurate to recognize that the sophisticated operations of information warfare, influence campaigns, and

influence operations should now be considered to part of the norms that will be used in the future cyber warfare campaigns and “competition” between nation states. The use of influence operations can be identified by examining how influence operations were employed by Russia in the influence operations carried out during the 2016 presidential election. (Wirtz, James J. Cyber War and Strategic Culture: The Russian Integration of Cyber Power into Grand Strategy). With the collapse of the Soviet Union and the decline of the kinetic military might of the USSR, the Russian government undertook a change in political and military strategy. This Grand strategy was developed as an alternative way to have geopolitical influence at the global level. (Suciu, Peter. Why cyber warfare is so attractive to small nations). For this to be accomplished by Russia in today's geopolitical climate, cyber influence campaigns need to be used in a way that can potentially impact and undermine the political structure of adversaries. The best strategy for gaining political advantage is to employ influence operations and campaigns in order to gain and political advantage in the future.

Influence operations present a better alternative for carrying out military and political activities that potentially have a political outcome favoring a specific political ideology. Campaigns of information warfare and influence operations are also related to what have been called psychological operations (Goldstein, Frank, ed. Psychological Operations: Principles and Case Studies - Fundamental Guide to Philosophy, Concepts, National Policy, Strategic, Tactical, Operational PSYOP). The amount of resources needed to carry out an influence campaign in the effort to sway the population of a nation, to influence ideas in audiences, while causing a population to potentially vote for a candidate, is minuscule compared to the expenditure of large amounts of money required to engage in kinetic warfare. Identifying the ethical issues related to information warfare and influence operations is needed before conducting an anticipatory ethical analysis and this also requires identifying the stakeholders involved in the influence of operations carried out during 2016 presidential election. The reason for conducting an anticipatory ethical analysis is to attempt to provide a foundation for establishing the most important principles that will be necessary for guaranteeing legitimate elections, democratic values and discussions.

## **2. Ethical analysis**

To understand the ethical issues involved with Facebook and Cambridge Analytica requires an understanding of ethics. Ethics relates to agents who perform actions. Dwight Furrow identifies the focus of ethical analysis as involving series of factors. As Furrow states, ethics is related to evaluating actions and actions are performed by those capable of being moral agents. As Furrow says, “When we evaluate an action, we can focus on various dimensions of the action. We can evaluate the person who is acting, the intention or motive of the person acting, the nature of the act itself, or the consequences.” (Furrow, Dwight. Key Concepts in Philosophy.)

Two points made in this passage can be applied to the Facebook – Cambridge Analytica scandal. First, ethical issues related to the Facebook – Cambridge Analytica case are based upon the idea that what those who perpetuate influence operations do, are actions, and second, these actions are an extension of a person’s intentions. The actions of influence operations are capable of being evaluated based upon the intentions and actions of the person engaged in the activity. If the distinctions are applied to influence operations, there are three possible levels of ethical evaluation. We can evaluate the actions of a person controlling the actions of an influence operation, the intentions of the person controlling and directing the actions involved in influence operations, and the consequences of the actions intended by the person controlling the actions of an influence operation.

The actions of agents deploying influence operations, which may involve propaganda, misinformation, and disinformation, are subject to ethical evaluation based upon the actions of the person disseminating the disinformation, the intentions of that person and the consequences produced by propaganda, misinformation, and disinformation. Ultimately it is the person or persons, who are controlling the disinformation, who are subject to moral evaluation. (For more on this see: Gunkel, David, The Machine Question). If we want to identify the ethical issues with disseminating disinformation, we need to ask, what actions are performed when the disinformation is disseminated, what is the character of the person controlling the disinformation that is disseminated, what are the intentions of the disinformation being disseminated, and what are the consequences of the disinformation being disseminated?

### **3. Facebook and Cambridge Analytica: Technical issues**

What are the technical issues related to the influence operations? It has been argued that there has always been always skepticism about Facebook data sharing but now, since information about Cambridge Analytica has come out, (Chang, Alvin. "The Facebook and Cambridge Analytica Scandal, Explained with Simple Diagram.") it has been proven that this skepticism was not unwarranted. One technical issue with social media companies like Facebook, is related to the issue of whether having an advertising-supported Surveillance Capitalism economic model is capable of supporting user interests. What occurs from the perspective of business, is that users enter their personal data into a social media database and algorithms are used to target users with ads for products based upon the data the users have entered into the platform. To analyze the CA scandal, a variety of stakeholders perspectives will be considered, some of the stakeholder perspectives that need to be taken into consider will include Facebook management, Cambridge Analytica, the Trump campaign, users who took the Facebook App test, data brokers, Robert Mercer, Christopher Wylie and the United States government. (Zuboff, Shoshana. *Big Other: Surveillance Capitalism and the Prospects of an Information Civilization* and "Why the business model of social media giants like Facebook is incompatible with human rights.") If we examine how Facebook makes money, we find that it is through advertisements. Users of Facebook, import their data into Facebook, which, is then sold to third-party clients, as was the case of Cambridge Analytica. This includes Facebook employing data mining techniques to identify products in order to market these products to individual Facebook users. (Scherr, Ian. *Facebook, Cambridge Analytica and data mining: What you need to know*). Social media platforms are aimed at making money and advertisements are the best option for accomplishing this task. However, the ads that are shown to the user are based upon the information that the user previously entered into the platform, which eventually also includes their search history. (Hetherington, James. *Facebook Whistleblower Reveals Cambridge Analytica Is Tip of Data-Mining Iceberg.*) Facebook simply employs the user's information as a way for them to make money. As Mark Zuckerberg argues, having an advertising-supported model, is the only rational model that can support building this service to reach millions of people. (Mark Zuckerberg defends Facebook's advertising-supported business model.)

### **4. Psychological profiling**

According to a documentary (on *Channel 4 News: Cambridge Analytica Uncovered: Secret filming reveals election tricks* (Data, Democracy and Dirty Tricks, Channel 4 News, 19 Mar 2018. and Porter, John.), psychological profiling is one of the oldest methods in any marketing manual and in advertising campaigns. Psychological profiling is at the center of the Facebook business model. How did Cambridge Analytica use Facebook for psychological profiling? In 2014 a personality quiz called "this is your digital life", was created by Aleksander Kogan. (Wagner, Kurt. Here's how Facebook allowed Cambridge Analytica to get data for 50 million users and "Another Facebook quiz could have stolen data under the guise of research,"). This quiz harvested some of the personal data from users of Facebook who took the quiz. The quiz targeted users and focused on personal interests which raised questions about personal privacy. Data Related to 50 million Facebook users was collected for supposedly academic purposes. This data included information about the user's demographics: their age, sex, and education. Information was also collected about the user's psychographics: their interests, opinions, and values. The quiz also constructed a list of personality profile traits about users, based on their answers to the quiz questions. This was problematic in two senses, first, data mining in Facebook was supposed to be restricted to academic projects. Second, it was also potentially a breach of personal privacy when it gathered data for non-academic purposes. As a result of what occurred, Facebook CEO Mark Zuckerberg has had to defend the situation involving Cambridge Analytica. For users of Facebook, the issue of data collection and privacy were troublesome and it created tensions between Facebook and users, as well as tensions between Facebook and United States government. (Statt, Nick. *The Justice Department and FBI are reportedly investigating Cambridge Analytica over Facebook scandal.*)

There are a variety of issues related to the intentions of the various stakeholders involved in the Facebook – Cambridge Analytica case and the issue of who is morally responsible for what occurred in the scandal. Using the distinctions from Furrow we can identify ethical issues related to intentions. For example, gathering data related to individual persons, which then became available to third parties, such as Cambridge Analytica, without users knowing it, seems to be an unacceptable breach of privacy and a violation of personal security. The intentions of users is to engage in social activities related to social existence. This is in potential conflict with the intention of Cambridge Analytica, which was aimed at gathering user data from Facebook that could be employed to engage in political influence operations. There is also the intention of Facebook to make money. Unfortunately, the majority of users may thoughtlessly opt into the convenience of using Facebook, which offers

free membership, without worrying about privacy issues. Since Facebook is a complex platform, adding specific regulations to limit the security risks for users is difficult to implement. It seems that Facebook needs to protect user's privacy at the same time that users have to the need for strong self-advocacy in order to maintain personal privacy (Waterfield, Phee. Cambridge Analytica Under Fire for Data Harvesting).

## **5. Influence campaigns and operations**

In addition to ethical issues connected directly to individual privacy seen in the Facebook - Cambridge Analytica scandal, there were also the potential for risks that emerged related to democratic governments and for the concept of democracy in general. For example, once Cambridge Analytica had obtained the data about users from the Facebook quiz, it appears that Cambridge Analytica used their technical abilities to micro target individuals, the groups they belonged to, and entire states without anyone at first knowing it. (Babu, Oana. Advertising, Microtargeting and Social Media). The extent of influence operations and campaigns during the 2016 election has now been documented. (Howard, Phillip, Bharath Ganesh, and Dimitra Liosiou. The IRA, Social Media and Political Polarization in the United States, 2012-2018). This type of 'targeting', presents a challenge to the legitimacy of elections, since the intention of the targeting is to attempt to shape the ideas of potential voters. There is also the intention to create disruption among voters by introducing divisive stories and disinformation, which may in turn influence ideas about the legitimacy of elections and the democratic process. These Intentions, are a major concern since this data gathering is then capable of being used to intentionally perform an action related to micro targeting personal accounts and even private messaging. Influence operations were aimed at influencing voter behavior. As those familiar with psyops would claim, human beliefs and desires, and hopes and fears, are the most powerful triggers in Information Warfare and Influence Campaigns. (Goldstein, Frank ed. Psychological Operations: Principles and Case Studies - Fundamental Guide to Philosophy, Concepts, National Policy, Strategic, Tactical, Operational PSYOP.) Companies that employ data analytic methods such as those employed by Cambridge Analytica, will gather data from millions of users via the Facebook application and then, through information warfare and influence operations, engage in the effort to reshape the ideas of members of society. This can be accomplished by using slogans, memes, and advertisements. Cambridge Analytica conducted early examinations prior to Trump's candidacy (Sheth, Sonam,.Cambridge Analytica began testing out pro-Trump slogans the same year Russia launched its influence operation targeting the 2016 election), testing the potential for the influence of memes and slogans such as posting "Build the Wall" and "Drain the Swamp" messages on Facebook. The goal of these postings was to attempt to view how the sharing of memes could be introduced to and then shared among individuals and groups. The movement of slogans through Facebook could be tracked through 'sharing' from 'friend to friend', and from 'friends of friends, to friends of friends'. (Lapowsky, Issie. Cambridge Analytica Could Have Also Accessed Private Facebook Messages.)

In addition, the company gained additional data about potential voters after purchasing data that was mined from users who downloaded the Facebook app "this is your digital life".(Kalvapalle, Rahul. Facebook app 'This Is Your Digital Life' collected users' direct messages:) Eventually, this method allowed the Trump campaign to micro target individuals, as well as individual cities, with memes, slogans, advertisements, and narratives, where they would potentially have the most impact in attempting to influence targeted audience attitudes. This data was sold to Cambridge Analytica against Facebook's terms of service. (Facebook suspends Cambridge Analytica, which worked for Trump campaign. by NBC News). This is an example of how the information warfare, influence campaigns, and influence operations work, while Cambridge Analytica can easily be described as propaganda and disinformation organization. This type of propaganda campaign, carried out in social media, collects user data in the effort to employ influence operations, as part of pursuing their own or their client's political agenda. What is significant for the future is to recognize that even if the mechanisms used by Cambridge Analytica are now prevented, the future development of manipulation with potentially more sophisticated strategies for conducting influence operations stand to be more highly developed by different organizations.

## **6. Propaganda, misinformation, and disinformation**

Important ethical issues emerge that are related to an audience and the ability of audiences and users to be able to differentiate between accurate or factual information, propaganda, misinformation, and disinformation. With propaganda, according to Christopher Wylie, even before the Trump campaign, Cambridge Analytica and Steve Bannon were testing slogans like images of "Build the wall" and "Drain the Swamp" to attempt to identify and determine potential targets for influence campaigns and influence operations. (How Steve Bannon used Cambridge Analytica to further his alt-right vision for America). Campaign slogans often target human emotions

which are at the heart of both tribal politics and post truth politics. The results of data harvesting can be used to enact psychological and influence operations which aim at influencing attitudes based upon the exploitation of the mental vulnerabilities of the targeted audiences. Influence campaigns and operations can be used to attempt to influence and manipulate attitudes through fear and to try to influence and control the behaviors of targeted audiences. These are the same techniques are used in advertising campaigns. In political campaigns there is an explicit effort to attempt to reinforce beliefs about what audiences want to hear as well as what audiences fear.

According to a CNN report Steve Bannon (How Steve Bannon used Cambridge Analytica to further his alt-right vision for America), wanted to use the type of aggressive messaging tactics usually reserved for geopolitical conflicts, including military strategies, to conduct cultural warfare through social media. The goal of these activities was to move the US electorate further to the political right. According to Christopher Wylie, Cambridge Analytica helped built a psychological warfare tool for Steve Bannon in order to attempt to reshape the perceptions of political issues by voters for a political agenda. Influence campaigns, psychological operations and influence operations have long been part of the political campaigns. (Cyberspace & Information Operations Study Center influence operations and Hitler, Adolf. *Mein Kampf*). On the other hand, attempting to reshape the political focus of society and through this, influence the thoughts and behaviors of individuals and of the general public, involves attempting to undermine the autonomy of individuals. The App created and used in Facebook by professor Kogan, was transformed by Cambridge Analytica into an influence engine directed at the long range goal of influencing the mental attitudes of potential voters. This influence campaign and the influence operations conducted within the campaign potentially had a negative effect upon a society. This is now thought to have included attempting to undermine the confidence in the electoral process. (Howard, Phillip, Bharath Ganesh, and Dimitra Liosiou. The IRA, Social Media and Political Polarization in the United States, 2012-2018). The potential negative influence of disinformation and falsehoods presented as facts was carried out at the political level by a private organization. At the same time Cambridge Analytica had a business partner from Russia. (O'Sullivan, Donie, Drew Griffin and Patricia DiCarlo. Cambridge Analytica's Facebook data was accessed from Russia, MP says.). It is interesting to note that, Cambridge Analytica's motto was "Cambridge Analytica - data-driven behavior change".

## **7. Privacy and data sharing**

It is important to recognize that users of social media need to be better educated about how the social media platforms they use work, as well as the policies that social media companies have about privacy rules and data sharing. From the perspective of critics, the managers of the social media platforms should be concerned about these issues. Aleksandr Kogan developed the personality testing Facebook App called "*this is your digital life.*" <https://pix11.com/2018/04/10/what-is-this-is-your-digital-life-the-facebook-app-you-may-be-alerted-about/> ). Kogan violated Facebook's terms of agreement, when he sold that information, to Cambridge Analytica, Facebook has some degree of responsibility since they allowed the App to mine data from its users. Users may see the situation differently, and from their perspective, this is a violation of ethical standards. For Utilitarian's, there is a focus on outcomes of actions that bring about the greatest amount of happiness for the greatest number. Data mining that is unknown to users seems to do the reverse. Even though Facebook officials and Cambridge Analytica CEO Alexander Nix denied any wrongdoing, users should definitely be concerned with the intention of the types of data harvesting that social media companies such as Facebook are practicing and the uses to which the harvested data is being put.

The kinds of practices engaged in by Facebook and Cambridge Analytica help bring into focus a larger set of concerns, that are related to what large corporations know about users. These concerns are related to data brokers, brokers who collect all sorts of information – such as names, addresses, income. The data collected by data brokers is then sold to 3<sup>rd</sup> party companies, such as Cambridge Analytica. One problem is that while individual users of social media do often share too much of their personal information on a daily basis, it is just a matter of time until companies would figure out how to use this data for profit. In addition it is also a matter of time until they figure out how to use the data in conjunction with other manipulation tools. From the perspective of data brokers, since they are also companies which are in business, it seems as if the consideration and avoidance of ethical issues with data harvesting and big data analytics isn't a concern, because they are simply the 'messenger' for political influence operations, based upon advertisements and marketing strategies aimed at making money.

According to Keith D. Foote (<http://thebernreport.com/robert-mercier-john-bolten-cambridge-analytica/>), Cambridge Analytica was a company that was partly owned by the family of Robert Mercer. Robert Mercer is a computer scientist and co-CEO of Renaissance Technologies. Mercer also played a key role in the Brexit campaign. Robert Mercer invested ten million dollars in Breitbart – run by Steve Bannon, a pro-Trump news outlet. (<https://www.breitbart.com/>). It can be argued that this establishes a strong connection between Cambridge Analytica and the 2016 Trump presidential campaign. It is also difficult to ignore the fact that Cambridge Analytica was also doing business with a large Russian oil company, ([Cadwalladr](#), Carole and Emma Graham-Harrison. Cambridge Analytica: links to Moscow oil firm and St Petersburg university), which indicates that in the past there seems to have been a connection between the Trump campaign and Russia. Multi-pronged influence campaigns employed in the 2016 United States presidential election involving Facebook, Cambridge Analytica, and Russia were used to create dissension among civilian populations in the attempt to influence the outcome of the election due to the strategies of each of these actors,

According to ‘whistle blower’ Christopher Wylie – former research director of Cambridge Analytica, the data mining and influence operations are an example of how modern colonialism works. At the geopolitical level this involves nation states influencing other nation states at the global level. (Crabtree, Justina. Cambridge Analytica is an ‘example of what modern day colonialism looks like,’ whistleblower says). When an organization engages in activities that undermine social well - being questions arise about the consequences of these activities. It is important to have a general public that is able to differentiate between fact based information, propaganda, misinformation, and disinformation, and that has media literacy capabilities, as well as critical reasoning skills. Wylie ask the question, why did Facebook allow such a data mining project to occur without warning users that such an activity was taking place? As is known, US midterm elections and other global elections were scheduled to take place, so it would be beneficial for the public to realize that in many cases propaganda, misinformation and disinformation are distributed through social media. This might also be a large part of how political campaigns in the future will be conducted. Members of the public need to recognize that nations and state-sponsored actors can use a variety of means to attempt to influence political structures, elections, while trying to influence political outcomes. An issue that needs to be recognized is how, Facebook, with specially designed data mining Apps, were employed to disseminate and perpetuate propaganda and fake news. (Howard, Phillip, Bharath Ganesh, and Dimitra Liosiou. The IRA, Social Media and Political Polarization in the United States, 2012-2018).

## **8. Technical conclusions**

Influence and operations campaigns were employed to profile individuals based upon data gathering and data mining techniques and the effort was made to use data gathered from Facebook to influence the ideas and behaviors of a large segment of the voters in the United States. This was an important goal for Cambridge Analytica using the data gathered by Aleksandr Kogan. To be more specific the Cambridge Analytica scandal was more than just a minor “breach”, as Facebook executives have defined it. It was contrary to ethical principles at a number of levels. The intentions, actions, and outcomes aimed at by Cambridge Analytica were all ethically problematic. It is important to recognize that after taking a personality quiz, the data of users that was collected from the quiz included private messages, likes, and also user’s locations.( Danner, Chas, Trump’s Data Firm Exploited Private Information From Millions of Facebook Users). A basic lesson learned from the Facebook/Cambridge Analytica scandal is that Facebook’s privacy policy is not a guarantee of data protection. Mark Zuckerberg has admitted that the data of individuals was not protected as much as it should have been. (Associated Press, 03/21/18 Mark Zuckerberg admits mistakes, outlines steps to protect user data.). The ethical issues with user privacy during the 2016 U. S. election and Brexit show the level of risk that is potentially at issue and the potentiality for more sophisticated strategies to be employed in future elections. While users of social media platforms should know that privacy is not guaranteed and personal data is not safe on any social media accounts, it is important for individuals to practice responsible use of social media and to practice mindful digital interaction within social media. Social media platforms such as Facebook treating users and user’s data as a “product,” and as a commodity, which means, that they are simply a means to make money. One technical issue that needs to be recognized is that ads are directed toward users based upon the information that users provide to social media platforms. Unfortunately many of the users of any social media platforms, will ignore or simply forget about the security issues related to personal data and keep on posting personal information online. As technology continues to become more advanced, voice recognition and face recognition capabilities will be more developed and added to social media and the outcome is likely to be, privacy will be capable of being violated in even more sophisticated fashions.

## **9. Ethical analysis**

With the constant advancements in technology and the trends related to data mining and big data analytics, it is important to attempt to identify potential problems and to attempt to anticipate ethical problems that may emerge, as the basis for policy development. It is important to attempt to develop and apply ethical rules as the basis for policy development, to attempt to prevent and foresee issues with situations that are likely to occur in the future. A set of 5 rules have been developed by a community of scholars to help guide thoughts about computing artifacts. (*Moral Responsibility for Computing Artifacts: Five Rules, Version 24*). These 5 rules were developed to help guide how artifacts including Facebook should be used. In order to apply the 5 rules we assume that social media platforms, algorithms, and data mining techniques such as those employed by Cambridge Analytica, are all sophisticated technological artifacts. The first rule states, “The people who design, develop, or deploy a computing artifact are morally responsible for that artifact, and for the foreseeable effects of that artifact. This responsibility is shared with other people who design, develop, deploy or knowingly use the artifact as part of a sociotechnical system.” The application of this rule would be that the creators of the applications deployed in Facebook need to ask users to consent to their data being harvested and they are responsible for using this data appropriately and describing to users what will be done with the data.

The second rule of the 5 states, “The shared responsibility of computing artifacts is not a zero-sum game. The responsibility of an individual is not reduced simply because more people become involved in designing, developing, deploying or using the artifact. Instead, a person’s responsibility includes being answerable for the behaviors of the artifact and for the artifact’s effects after deployment, to the degree to which these effects are reasonably foreseeable by that person.” This second rule would state that Facebook should be responsible for what the application and creation of a data mining technique (by Cambridge Analytica) did, as well as what they did with the data harvested.

The fifth rule of the 5 rules is, “People who design, develop, deploy, promote, or evaluate a computing artifact should not explicitly or implicitly deceive users about the artifact or its foreseeable effects, or about the sociotechnical systems in which the artifact is embedded.” One application of this rule would be that Facebook should make it transparent to Facebook users what applications by third party users are harvesting as well as how they are using their data. It needs to be made clear to users what their data is being used for, and they should provide a simplistic way to opt out of processes of data sharing.

## **10. Anticipatory ethical recommendations/policy**

Recommendations can now be made in conjunction with conclusions of the preceding application of the 5 rules as the basis of an anticipatory ethical analysis and policy recommendations. What can be anticipated is that it seems inevitable that globally societies will accept free products services from companies such as Facebook, in exchange for agreeing to being targets of advertisements. As a result there should be some set of core best practices that should be followed social media companies. For example, when a social media platform like Facebook, is used by a business such as Cambridge Analytica, the request to access information should be obviously stated and should include an explanation of where the data goes and where the data will go. Applications (Apps) should also not be able to mine data of a user’s friend without the friend clearly agreeing to this ‘sharing’ of this data. Also, current and potential representatives of the government should not have access to and be able to manipulate the use of a user’s personal information, for a political agenda. In addition, while influence operations aimed at influencing audiences there is also ‘evidence’ that the U. S. voting system was breached and although not reprogrammed, in the future without proper safeguards, it could be. The ethical question any stakeholder of social media has to take into consideration is how this will be prevented in future elections. An election system can be influenced in such a way as to potentially reconstruct democracy in a new form of modern colonialism. Interestingly, with Cambridge Analytica, although it has now closed its doors, Facebook is attempting to change policies. In addition all other social media platforms are implementing privacy policies related to user data, which is one of the effects that are results of the Facebook- Cambridge Analytica scandal. All of these points need to be taken into consideration by those who will make policy about social media in the future.

A final point needs to be made. How psychological operations and influence operations are performed, must be seen to be an important part of cyber warfare and needs to be part of cyber warfare education. It needs to be recognized that the study of information warfare, influence campaigns, and influence operations will play an important role in nation-state to nation state competitions and these additions to warfare will continue in

conflicts into the foreseeable future. The result is that the study of cases such as the Facebook – Cambridge Analytica scandal need to be included in Cyber Warfare and Cyber Security education in the future.

## **References**

- "Another Facebook quiz could have stolen data under the guise of research", Available at:  
<https://www.digitaltrends.com/social-media/cubeyou-suspended-accused-facebook-data-misuse/>
- Associated Press, 03/21/18 Mark Zuckerberg admits mistakes, outlines steps to protect user data. Available at:  
<https://www.10tv.com/article/mark-zuckerberg-admits-mistakes-outlines-steps-protect-user-data>
- Babu, Oana. Advertising, Microtargeting and Social Media. *Procedia - Social and Behavioral Sciences*, Volume 163, 19 December 2014, Pages 44-49.
- Bloomberg April 10, 2018. Available at:<http://fortune.com/2018/04/10/facebook-cambridge-analytica-what-happened/>
- Cadwalladr, Carole and Emma Graham-Harrison. Cambridge Analytica: links to Moscow oil firm and St Petersburg university, The Guardian, 17/March/2018. Available at:  
<https://www.theguardian.com/news/2018/mar/17/cambridge-academic-trawling-facebook-had-links-to-russian-university>
- Cambridge Analytica whistleblower Christopher Wylie warns that Facebook threatens Free Speech. Jul 20, 2018.  
Available at: <https://www.washingtonpost.com/news/the-switch/wp/2018/06/19/>
- Chang, Alvin. "The Facebook and Cambridge Analytica Scandal, Explained with a Simple Diagram." Vox, Vox, 23 Mar. 2018, available at: [www.vox.com/policy-and-politics/2018/3/23/17151916/facebook-cambridge-analytica-trump-diagram](http://www.vox.com/policy-and-politics/2018/3/23/17151916/facebook-cambridge-analytica-trump-diagram).
- Chua, Amy. Political Tribes: Group Instinct and The Fate of Nations. Penguin Press, New York, N. Y. 2018.
- Clausewitz quote available at: [https://www.brainyquote.com/quotes/carl\\_von\\_clausewitz\\_174934](https://www.brainyquote.com/quotes/carl_von_clausewitz_174934)
- Crabtree, Justina. Cambridge Analytica is an 'example of what modern day colonialism looks like,' whistleblower says, CBNC, 27 March 2018. Available at: <https://www.cnbc.com/2018/03/27/cambridge-analytica-an-example-of-modern-day-colonialism-whistleblower.html>.
- Danner, Chas, Trump's Data Firm Exploited Private Information From Millions of Facebook Users. March 17, 2018, New York Magazine. Available at: <http://nymag.com/daily/intelligencer/2018/03/trumps-data-firm-exploited-facebook-data-from-millions.html>
- Data, Democracy and Dirty Tricks, Channel 4 News, 19 Mar 2018, Available at: [data-democracy-and-dirty-tricks-cambridge-analytica-uncovered-investigation-expose](http://data-democracy-and-dirty-tricks-cambridge-analytica-uncovered-investigation-expose).
- Facebook suspends Cambridge Analytica, which worked for Trump campaign. by NBC News / Mar.17.2018 / 1:31 AM ET / Updated Mar.17.2018 / 11:20 AM ET / Source: Reuters. Available: <https://www.nbcnews.com/tech/social-media/facebook-suspends-cambridge-which-worked-trump-campaign-n857516>.
- Furrow, Dwight. Ethics: Key Concepts in Philosophy, Continuum, New York, NY. 2005. p. 44.
- Goldstein, Frank, ed. Psychological Operations: Principles and Case Studies - Fundamental Guide to Philosophy, Concepts, National Policy, Strategic, Tactical, Operational PSYOP. U. S. Military. Dept of Defense, April 9. 2017.
- Gunkel, David, The Machine Question, The MIT Press, Cambridge Mass., 2012, pp. 65-74.
- Hetherington, James. Facebook Whistle-blower Reveals Cambridge Analytica Is Tip of Data-Mining Iceberg. 3/20/18 Available at: <https://www.newsweek.com/cambridge-analytica-just-one-hundreds-mining-data-facebook-852744>
- Hitler, Adolf. Mein Kampf, Houghton Mifflin Company, 1971, Chapter 6.
- Howard, Phillip, Bharath Ganesh, and Dimitra Liosiou. The IRA, Social Media and Political Polarization In the United States, 2012-2018, Computational Propaganda Research Project, University of Oxford,2018 available at:  
<https://comprop.oiil.ox.ac.uk/wp-content/uploads/sites/93/2018/12/IRA-Report.pdf>
- How Steve Bannon used Cambridge Analytica to further his alt-right vision for America, 2018-7-30. Available at:  
[http://m.cnn.com/en/article/h\\_de1e28304a404967f4477409a3db3a40](http://m.cnn.com/en/article/h_de1e28304a404967f4477409a3db3a40)  
<https://www.theatlantic.com/technology/archive/2018/03/facebook-cambridge-analytica/555866/>  
<http://thebernreport.com/robert-mercier-john-bolten-cambridge-analytica/https://pix11.com/2018/04/10/what-is-this-is-your-digital-life-the-facebook-app-you-may-be-alerted-about/https://www.breitbart.com/>
- Kalvapalle, Rahul. Facebook app 'This Is Your Digital Life' collected users' direct messages: report, April 13, 2018. Available at: <https://globalnews.ca/news/4143810/aleksandr-kogan-this-is-your-digital-life-messages/>
- Lapowsky, Issie. Cambridge Analytica Could Have Also Accessed Private Facebook Messages. Wired, 04. 10.18. Available at: <https://www.wired.com/story/cambridge-analytica-private-facebook-messages/>
- Mark Zuckerberg defends Facebook's advertising-supported business model. AP, April 03, 2018 Available at:  
<https://brandequity.economictimes.indiatimes.com/news/advertising/mark-zuckerberg-defends-facebooks-advertising-supported-business-model/6359088>
- McIntyre, Lee. Post Truth. The MIT Press, Cambridge, Massachusetts, 2018.
- Moral Responsibility for Computing Artifacts: Five Rules, Version 24. Available at:  
<https://edocs.uis.edu/kmill2/www/TheRules/moralResponsibilityForComputerArtifactsV24.pdf>.
- O'Sullivan, Donie, Drew Griffin and Patricia DiCarlo. Cambridge Analytica's Facebook data was accessed from Russia, MP says. July 17, 2018. Available at: <https://money.cnn.com/2018/07/17/technology/cambridge-analytica-data-facebook-russia/index.html>

**Richard Wilson**

- Porter, John. Watch Channel 4's explosive Cambridge Analytica Uncovered right here Read more at  
<https://www.trustedreviews.com/news/cambridge-analytica-uncovered-channel-4-watch-online-3428542#WtM1yuhUfdhFvdi4.99>. March. 20, 2018.
- Porter, John. Watch Channel 4's explosive Cambridge Analytica Uncovered right here  
Read more at <https://www.trustedreviews.com/news/cambridge-analytica-uncovered-channel-4-watch-online-3428542#WtM1yuhUfdhFvdi4.99>. March. 20, 2018.
- Underwood, Kimberly. A New Front in Information Warfare, SIGNAL Magazine. March 20, 2018. Available at:  
<https://www.afcea.org/content/new-front-information-warfare>
- Scherr, Ian. Facebook, Cambridge Analytica and data mining: What you need to know, April 18, 2018 Available at:  
<HTTPS://WWW.CNET.COM/NEWS/FACEBOOK-CAMBRIDGE-ANALYTICA-DATA-MINING-AND-TRUMP-WHAT-YOU-NEED-TO-KNOW/>.
- Sheth, Sonam,.Cambridge Analytica began testing out pro-Trump slogans the same year Russia launched its influence operation targeting the 2016 election. March 20, 2018. Available at: <https://www.businessinsider.com/cambridge-analytica-trump-russia-ties-2018-3>.
- Statt, Nick. The Justice Department and FBI are reportedly investigating Cambridge Analytica over Facebook scandal, available at: <https://www.theverge.com/2018/5/15/17358802/facebook-cambridge-analytica-justice-department-fbi-investigation>.
- Suciuc, Peter. Why cyber warfare is so attractive to small nations, Fortune, December 21, 2014. Available at:  
<http://fortune.com/2014/12/21/why-cyber-warfare-is-so-attractive-to-small-nations/>.
- Thompson, Nick. Christopher Wylie, Cambridge Analytica whistleblower, speaks out on Facebook controversy, March 19, 2018. Available at: <https://www.cbsnews.com/news/christopher-wylie-cambridge-analytica-whistleblower-speaks-today-facebook-controversy-2018-03-19/>.
- Underwood, Kimberly. A New Front in Information Warfare, SIGNAL Magazine. March 20, 2018. Available at:  
<https://www.afcea.org/content/new-front-information-warfare>
- Wagner, Kurt. Here's how Facebook allowed Cambridge Analytica to get data for 50 million users, Mar 17, 2018. Available at: <https://www.recode.net/2018/3/17/17134072/facebook-cambridge-analytica-trump-explained-user-data>
- Waterfield, Phee. Cambridge Analytica Under Fire for Data Harvesting, 19, March, 2018. Available at:  
<https://www.infosecurity-magazine.com/news/cambridge-analytica-data-harvesting/>.
- "Why the business model of social media giants like Facebook is incompatible with human rights." April 2, 2018.Available at: <http://theconversation.com/why-the-business-model-of-social-media-giants-like-facebook-is-incompatible-with-human-rights-94016>.
- Wirtz, James J. Cyber War and Strategic Culture: The Russian Integration of Cyber Power Into Grand Strategy. CCDCOE, Tallinn, Estonia. 2015. Available at:  
[https://ccdcoe.org/sites/default/files/multimedia/pdf/CyberWarinPerspective\\_Wirtz\\_03.pdf](https://ccdcoe.org/sites/default/files/multimedia/pdf/CyberWarinPerspective_Wirtz_03.pdf)
- Zuboff, Shoshana. Big Other: Surveillance Capitalism and the Prospects of an Information Civilization, 17 Apr 2015. Journal of Information Technology (2015) 30, 75–89. doi:10.1057/jit.2015.5

# Information Warfare: Fabrication, Distortion and Disinformation: A Case Study and Anticipatory Ethical Analysis

Richard Wilson

Towson University, Department of Philosophy/Computer Science and Information Sciences, Towson University, USA

Hoffberger Center for Professional Ethics, University of Baltimore, USA

[wilson@towson.edu](mailto:wilson@towson.edu)

**Abstract:** Information warfare occupies an important place in cyber warfare and is aimed at influencing societies and states in 2 levels. This particularly true of Russian information warfare, which has 2 fundamental tasks, 1<sup>st</sup>, to accomplish political objectives without engaging in kinetic warfare, and 2<sup>nd</sup>, to influence, in a positive way, the international response to the deployment of Russian military and Russian allies. (Information Warfare: Russian Activities). At the individual level having accurate information is important because it is at the foundation of our decision making processes. When we are engaged in projects in the surrounding world, in order to accomplish our goals we need accurate information in order to determine the best ways to accomplish these goals. The same distinctions apply to nation states. They also need to use information in order to develop their policies and to achieve their political and geopolitical goals. Cyber warfare now includes information warfare where information is capable of being fabricated and also distorted, in order to disinform audiences, which has an impact upon individuals and nation states. At this point in time the advance of emerging technologies such as photo editing, Face Apps on smart phones, project voco, voice mimicry and voice imitating with emotion software, are technological developments that create the possibility for altering reports about what occurs in the world around us. Emerging technologies mediate between our experience and the objective world. What this means is that information about the surrounding world is capable of being significantly altered in a number of ways. This is due to how these emerging technologies can be employed to alter how and what information is presented to audiences. There are a number of ways in which technologies can alter and mediate between ourselves and the external world which in the process alters our experience of the surrounding world. This is already occurring in photo editing, Face Apps on smart phones, and will continue to occur as new technologies such as project voco, voice mimicry and voice imitating with emotion software developments, continue to evolve. This analysis is aimed at identifying and examining some of the technologies that can be employed to fabricate and distort information about what seems to be occurring in the external world. This fabrication and distortion has a potentially strong impact upon how audiences experience what has supposedly occurred in the world, which then serves to disinform these audiences. This creates further difficulties related to identifying facts about the world as well as influencing an audiences understanding of what amounts to information. This analysis explores how the use of these technologies will be conjoined to influence warfare to engage in information warfare in the future. An ethical analysis will show these technologies are misused now and how they will be misused now and into the future. Understanding how fabrication, distortion and disinformation are occurring will provide a foundation for developing policy about how to address disinformation warfare. It is important to attempt to identify if these technologies will lead to the fundamental nature of information warfare through the influence of ever more sophisticated influence operations.

**Keywords:** information warfare, influence operations, information, misinformation, disinformation, ethics, anticipatory ethics

---

## 1. Introduction

Information can be contrasted with propaganda, misinformation and disinformation. Information needs to be thought of as related to data and facts. Raw data needs to be interpreted in order to present facts about the external world that can be verified to be either true or false. What the advance of technology has done, in the areas of images and voice recognition technologies, is to create conditions for the possibility of creating completely inaccurate accounts of what has occurred and what is occurring in the external world. This now extends with voice alteration software into the area of what people are actually saying. Photographs can be altered using a number of techniques such as Photoshop while more recently voice recognition software, from a small sample of a speaker's voice, has been able to completely alter what a speaker has said. The combination of these two techniques allows for the hybrid creation of video and audio that allows someone wanting to manipulate images and audio to have a speaker say what the manipulator wants the speaker to say. These technological developments allow this manipulation to be so precise that how someone moves is in complete sync with words that the person never said. This analysis is concerned with identifying the ethical issues that are related to the dissemination of disinformation and the combination of visual and spoken distortion into main stream media and social media, and their potential impact upon social existence in the context of information and social media warfare. These technological developments now allow information warfare and influence

campaigns to be carried out at the levels of fabrication and distortion of information. Tactics used to accomplish the political objectives of information warfare while falling short of kinetic warfare include, damaging information systems, subverting political, economic and social systems, attempting to influence audiences through psychological manipulation, which could lead audiences to make decisions that could destabilize a state and society. (Hodnicki. Joe, 2017). Emerging technological developments also create the possibility for the widespread deployment of disinformation opening up new avenues for information warfare.

## **2. Fabrication, distortion and disinformation**

The concepts of information, propaganda, misinformation, and disinformation show how important it is to be able to detect fabricated and distorted information and fake news, within traditional media, but also within social media. The idea that nation states, and nation state sponsored information warfare, influence campaigns and hacktivism can be effective within social media and have an effect upon audience attitudes and elections, is one area that needs to be the focus of professionals interested in information warfare and influence campaigns. The study of cyber warfare now should include the study of information warfare and information and influence campaigns carried out through social media. Social media warfare and the role that soft warfare (see Gross, Michael, Soft War) can play in influencing audiences, includes undermining elections and the trust that the citizens have of various sources of information within the political arena. Loss of trust in national leaders, traditional sources of information, and in the entire democratic political system, puts in jeopardy national security in the sense that audiences that are vulnerable to manipulation through fabrication, distortion and disinformation, can have an impact upon the attitudes of audiences and upon who wins in national elections. This manipulation can even lead to the loss of audiences trust in the democratic process.

## **3. Information/propaganda/misinformation/disinformation**

As stated above information, political messages or news that are projected towards an audience have to be accurate if it is going to be informative for that audience. For members of an audience to be able to use information as the basis of their critical reasoning in order to make the best decisions in the effort to accomplish their goals, the information has to be accurate. In the context of news in order for a reporter/journalist to be able to inform an audience, they must be able to identify and report facts relevant to the reporting on the event. To differentiate between accurate news and fake news, we also need to be able to differentiate between information, propaganda, misinformation and disinformation. A clearly formulated set of basic definitions of these concepts has been developed by librarians. The differences between Information, propaganda, misinformation and disinformation are described in the following way. To begin, "Information, at its most basic, is data set that has been situated in an appropriate context for relevance. In other words, information tells us something that is understandable and has the potential to become knowledge for us when we view it critically and add it to what we already know."(JHU Library, Evaluating Information). In contrast with information is propaganda, which is defined in the following way. Propaganda is the "systematic propagation of information or ideas by an interested party, esp. in a tendentious way in order to encourage or instill a particular attitude or response. Also, the ideas, doctrines, etc., disseminated thus; the vehicle of such propagation." (from Oxford English Dictionary, 2nd ed., 1989) It is important to note that Propaganda is ideologically driven. Political campaign speeches and party political statements are often, in reality, a form of propaganda. They fit this definition when they present the opposing point of view in an unfavorable light. All political organizations do this on a variety of issues. (JHU Library, Evaluating Information] Propaganda is used in a disingenuous way when it uses an ideological framework to attack those who have opposing political views to one's own. In distinction from information and propaganda is misinformation, "Misinformation is defined as the action of misinforming or condition of being misinformed; or erroneous or incorrect information. Misinformation differs from propaganda in that it always refers to something which is not true. It differs from disinformation in that it is "intention neutral": it isn't deliberate, it's just wrong or mistaken." [ JHU Library, Evaluating Information] A key feature of misinformation is that it is presented by someone which is intention neutral and therefore the conveyor of misinformation can be taken to be legitimately confused about the truth or falsity of the misinformation that they are presenting. The final category is disinformation. According to the JHU librarians, "disinformation refers to disseminating deliberately false information, especially when supplied by a government or its agent to a foreign power or on the media with the intention of influencing policies of those who receive it." This seems to be a straightforward account of disinformation where the disinformers intentions are to present what has the appearance of information to an audience, and which is true, when it is known by the disinformers to be false. There is an intention by the disinformers to present something as true to an audience when in actuality the supposed truthful information is known by the disinformers to be false. For librarians this

may require examining documents. According to the JHU website, "One good starting point in determining whether or not a document may constitute disinformation is to find out who owns the document or domain and then find out what that individual or group's mission or beliefs are. Ask yourself what the document owner has to gain by circulating the document. Always validate or confirm information on individuals, institutions or groups, and countries that you find on the Internet. If you don't know who wrote what you read or why they wrote it, you don't know if it's trustworthy." (JHU Library, Evaluating Information). In the case of the emerging technologies under discussion here video footage and voice recording will be considered to be documents.

#### **4. Introduction to ethics**

After defining the crucial distinctions above we can turn towards ethical issues. In order to understand the ethical issues involved with fabrication, distortion and disinformation of 'documents' we need to have an understanding of ethics. Ethics in a basic definition relates to agents who perform actions and how these actions affect other agents. Dwight Furrow in the following passage identifies the focus of ethical analysis as involving series of factors. As Furrow states, ethics is related to evaluating actions and actions are performed by those capable of being moral agents. As Furrow says, "When we evaluate an action, we can focus on various dimensions of the action. We can evaluate the person who is acting, the intention or motive of the person acting, the nature of the act itself, or the consequences." (Furrow, Dwight. Ethics: Key Concepts in Philosophy)

Two important distinctions are introduced in this passage. 1<sup>st</sup>, ethical issues related to fabrication, distortion and disinformation are based upon the idea that what someone who presents what is fabricated, distorted and disinformation as if it is informative is an action, but this action is an extension of what a person does. 2<sup>nd</sup>, the actions related to the presentation of disinformation as information are only capable of being evaluated based upon the actions of the person controlling them. If this is true and if we endorse Furrow's distinctions identified in the preceding passage and apply them to the use of fabrications, distortions and disinformation, there are three possible levels of ethical evaluation. We can evaluate the actions of a person controlling the acts of fabrication, distortion and disinformation. We can evaluate the intentions of the person controlling and directing the actions of fabrication, distortion and disinformation, and finally, we can evaluate the consequences of the actions intended by the person controlling the actions of fabrications, distortions and disinformation.

We assume that the actions of fabrication, distortion and disinformation are subject to ethical evaluation based upon the actions of the person controlling the fabrication, distortion and disinformation, the intentions of that person and the consequences produced by the use of fabrication, distortion and disinformation. Ultimately it is the person or persons, who are controlling the fabrications, distortions and disinformation, that is subject to moral evaluation. ([Furrow, Dwight. Ethics: Key Concepts in Philosophy]) If we want to identify the ethical issues with fabrication, distortion and disinformation, we need to ask, what actions are performed when with fabrication, distortion and disinformation are used to manipulate audiences, what is the character of the person controlling the use of fabrication, distortion and disinformation, what are the intentions of those using with fabrication, distortion and disinformation, and what are the consequences of the use of fabrication, distortion and disinformation?

The ethical issues with emerging technologies are the result of how these technologies can be used to fabricate, distort, and disinform audiences. They are capable of being employed by a user to influence those who experience fabrications, distortions and disinformation as they are employed by the perpetrator. There is the controller of the fabrications, distortions and disinformation, the fabrications, distortions and disinformation, and who and what is influenced by the activities of fabrication, distortion and disinformation. The intention of the users of fabrication, distortion and disinformation involves instrumental reasoning and establishing a purpose for the use of fabrication, distortion and disinformation, as well as a goal (such as manipulation of the attitudes of an audience). There are also those affected by the purpose of the fabrication, distortion and disinformation which involves the technical issue of being psychologically manipulated. (For more on this see: PSYOP Military Psychological Operations Manual) It is from this interaction between the technical use of the fabrication, distortion and disinformation by the user to influence the attitudes of audiences, and the technical effect of the fabrication, distortion and disinformation as it affects another person, that ethical issues with fabrication, distortion and disinformation arise. A preliminary ethical analysis using standard ethical principles can be developed from how the intentions of the users of with fabrication, distortion and disinformation affects the person or persons affected by the use of with fabrication, distortion and disinformation.

## **5. Methods of fabrication, distortion and disinformation**

As stated above there are a variety of technological developments that are already contributing to the fabrication and distortion of information and to the dissemination and spread of disinformation. These technological developments include software developments in photo editing, Face Apps on smart phones, project voco, voice mimicry and voice imitating with emotion. These developments are all directly related to software developments.

### **A. Photo Editing**

Photo editing software such as Photoshop has altered the way in which photo editing can take place. Photoshop introduced a creative revolution in photo editing by allowing photos to be edited using digital resources. Photoshop is described in the following way, "Adobe Photoshop is a raster graphics editor developed and published by Adobe Inc, It was originally created in 1988 by Thomas and John Knoll. Since then, this software has become the *de facto* industry standard not only in raster graphics editing, but to digital art as a whole; it even went to the point that the software's name itself has become a generic trademark, leading to its usage as a verb (e.g. "to photoshop an image", "photoshopping", and "photoshop contest") although Adobe discourages such use. Photoshop can edit and compose raster images in multiple layers and supports masks, alpha compositing, and several color models including RGB, CMYK, CIELAB, spot color, and duotone.". (See: Adobe Photoshop) Photo editing software in general can be used for a variety of purposes. Someone can alter photographs that they take of another person but they can also alter photographs of other persons that were taken in the past. Edited photos can be used to make alterations that include not just the alteration involving the features of a person but alterations of events. A person can be inserted into an event that never occurred. This would amount to the altering the time, place and space of when events occurred as well as altering the location where an event occurred and in what locations.

### **B. FaceApp**

Advances in the manipulation of photographs has been demonstrated with smartphone technology and what often seems to be a harmless application of smartphone capabilities can be found in FaceApp. This app which was developed by the Russian company Wireless lab employs neural network technology which allows users of the app to generate alterations of faces in photographs. Those who use this app can alter what appears in a photograph. Some of the changes that can occur are that young people can be made to look old, facial expressions can be altered so that a frown can become a smile and it's also possible to improve appearances. Faces can also be distorted. What is involved in the FaceApp software involves a combination of artificial intelligence with neural networking. This combination allows the computational system that is at the center of FaceApp to improve its performance through the transmission of digital signals. This has greatly improved the pixel by pixel transformations of photos by Photoshop. (Vincent, James, "This App Uses Neural Networks to Put a Smile on Anybody's Face,")

### **C. Motion Capture**

Researchers at the Max Planck Institute for Informatics and Stanford University in 2016 demonstrated a motion capture system that showed how facial expressions on real people could be falsified. (Thies, Justus, Michael Zollhofer, Marc Stamminger, Christian Theobalt, et al., "Face2Face:Real-Time Capture and Reenactment of Rgb Videos.") Using film slips of political leaders including Obama, Trump, Putin, and George W. Bush, this research showed that unique expressions could be transposed onto the speaker. The target's facial structure could be preserved with a seamless transposition on the targets face. This ability creates the possibility for completely falsified video. According to Thies, the lead researcher, "We aim to modify the target video in a photo-realistic fashion, such that it is virtually impossible to notice the manipulations.". (Thies, Justus, Michael Zollhofer, Marc Stamminger, Christian Theobalt, et al., "Face2Face:Real-Time Capture and Reenactment of Rgb Videos.").

### **D. Plug and Play networks**

Researchers at the University of Wyoming and computer scientist at the Montreal Institute for learning algorithms and Uber AI Labs have worked collaboratively on the development of an artificial intelligence system that was focused on the generation of unique Images. Completely artificial images were generated of birds, ants,

monasteries and volcanoes that did not exist in the real world. This has been referred to as one type of " plug-and-play generative Network". (Nguyen, Anh, jeff Clune, Yoshua Benguo, Alexey Dosovitskiy, et a.;, "Plug to Play Generative Networks: Conditional Iterative Generation of Images in Latent Space,")

E. Project VoCo

In addition to developments that have occurred with images and video footage, there are also a number of technological developments in the areas of sounds and voice. Project VoCo this is a software application that allows for the rearrangement of people's words from audio clips or files which makes it possible for someone manipulating the audio clip or file to make it seem that the person said something that they didn't say. (Zeyu Jin, Gautham J. Mysore, Stephan Diverdi, Jingwan Lu, and Adam Finkelstein, "VoCo: Text-Based Insertion and Replacement in Audio Narration,") VoCo is a new software that was developed at the Engineering School of Princeton University. It is based on a sophisticated algorithm that utilizes machine learning to recreate the sound of a particular voice and also 'learn' from any mistakes made through human correction. Voco allows the user to edit a transcript of an audio recording of a human voice which involves adding or replacing words. These replacement words are processed through the software and become automatically synthesized into the audio recording in the speaker's voice." (<http://www.digitaljournal.com/tech-and-science/technology/new-software-edits-voices-like-text/article/497243>). Voco has been called Photoshop for the voice.

F. Lyrebird

This is a 2017 development that has a sophistication over VoCo, Lyrebird employs deep learning systems which were first created at the University of Montreal. Lyrebird has developed a voice imitation algorithm that "can mimic a person's voice and have it read any text with the given emotion based on the analysis of just a few seconds of audio recording. Users will be able to generate new and unique voices tailored for their needs."( Dorn, Lori, "Lyrebird, Remarkable Speech Synthesis Technology That Can Learn and Mimic Voices in Seconds.") Lyrebird software can take a sample of the spoken words of a person and the software will then create a voice that says in the same voice anything it is told to say. (see <https://www.wired.com/brandlab/2018/10/lyrebird-uses-ai-find-artificial-voice/>). This presents the possibility for users to be able to generate entire discussions and words spoken by an individual including the ability to generate completely new words and voices from a brief audio recording. This includes not only can the ability to create a sequence of words that the original speaker never uttered but they can also generate from audio sample a completely unique set of words in the target speakers voice.

**6. The social consequences of information that is presented in the cyber world**

The nature of cyber space is that it allows users access to use technologies to create a variety of what can be called 'alternative worlds'. There is also a variety of cyberworlds that can be created through the use of technologies such as those discussed in section 5. These technologies allow for fabrication, distortion and the dissemination of disinformation as well as for the construction of narratives based upon these fabrications and distortions. Modern cyber strategies by individuals, groups and nation state actors are capable of influencing rivals through the use of emerging technological developments related to altering images, sounds and through hybrid recombinations of these technologies. Rival nation states are capable of altering the balance of information, and thereby the balance of power, by using technologies to alter images and sounds with the goal of influencing the ideas of target audiences. In each of these cases there is an effort on the part of the initiator of fabrications and distortions to influence the thoughts and behavior of the target audiences.

A number of issues arise with the development of the various visual and voice technologies that can affect the ideas and behaviors of the targets. There can be a profound impact upon what an audience experiences in the mass media and what that audiences comes to believe. Visual and voice technologies can compromise the believability of what we see and hear. Voices can be altered to imitate a real political actor with emotion added. Those who use the technology will be able to generate entire dialogues with the voice of their choice or design from the ground up, with completely new and unique voices tailored to fit for their needs. Actors can be made to appear other than they appear and can be made to say things they never said. False narratives can be created that are aimed at audiences in order to persuade these audiences. Most significantly these technologies are capable of compromising what an audience sees and hears. Fabricated and distorted images, voices and the narratives built upon them can also influence the believability of what we see and hear. The technologies

analyzed above can be employed to engage in information warfare and influence campaigns that completely disinform audiences. They can be employed to attempt to accomplish the political objectives of information warfare while falling short of kinetic warfare. The goal of using emerging technologies include, damaging information systems, subverting political, economic and social systems, attempting to influence audiences through psychological manipulation, which could lead audiences to make decisions that could destabilize a state and society. What is unique about this analysis is the degree to which technological developments in visual and voice alteration technology are capable of influencing ideas in unsuspecting audiences.

## **7. Misuse possibilities**

The development of these new technologies, related to the alter of visual images and voices, creates possibilities for a great deal of misuse. As these technologies become less expensive and as they become easier to use individuals, groups and nation states can adopt them. The ease of alterations in video and spoken words could include making it sound as if a politician is saying something that they aren't really saying. This could lead to a great deal of confusion and descension by generating controversy as well as inciting violence in crowds that hear what the politician says not being able to recognize that the politician didn't really say what the altered voice pattern says they said. If this kind of technology becomes widely available to average consumers it is not difficult to imagine ordinary users and provocateurs with conspiracy websites being able to fabricate both visual and audio recording making it seem as if politicians and public figures have said things that they have not said. This alteration of audio would also make it possible for politicians to deny that something they said they actually said due to the fact that they can claim that the video recording of what they said has actually been altered by someone to attempt to slander or defame them. How information is presented to audiences and then processed by these audiences provides insight into what is perceived as factual about the real world. Those who present disinformation in the context of a narrative can create, in the minds of their audience, a perceived world or worlds. This is nowhere more obvious than in the construction of the worlds created by fake news. Fabrication, distortion and disinformation can be employed with emerging visual and voice technologies, in order to support disinformation campaigns and influence operations, in order to accomplish to political goals.

## **8. Anticipated ethical issues**

According to James Rachels (James. The Elements of Moral Philosophy), "Morality is, at the very least, the effort to guide one's conduct by reason – that is, to do what there is the best reasons for doing – while giving equal weight to the interests of each individual affected by one's action." This conception of morality gives a fundamental picture of what it means to be a conscientious moral agent. A moral agent "is someone who is concerned impartially with the interests of everyone affected by what he or she does; who carefully sifts facts and examines their implications; who accepts principles of conduct only after scrutinizing them to make sure they are justified; who is willing to 'listen to reason' even when it means revising prior convictions; and who, finally, is willing to act on these deliberations."

As noted above following the distinctions introduced by Furrow, we assume that the actions of fabrication, distortion and disinformation are subject to ethical evaluation based upon the actions of the person controlling the technologies that produce fabrication, distortion and disinformation. We can also access the intentions of persons and the consequences produced by the dissemination of fabricated and distorted information and disinformation. The construction of false narratives is a political and geopolitical reality that can have negative consequences ultimately it is the person or persons, who are controlling the fabrications, distortions and disinformation that are subject to moral evaluation. Furrow, Dwight. Ethics: Key Concepts in Philosophy]

When we apply these distinctions to the concepts of fabrication, distortion, information, propaganda, misinformation and disinformation we see a wide range of misuse possibilities for the developing technologies related to visual images and voices examined above. It seems obvious that in the domains of cyber warfare, information warfare, and influence campaigns developing technologies will be used by agent's in a fashion that stands in stark contrast to how moral agents might use the same technologies. There are conditions that make fabrication, distortion and disinformation particularly dangerous as we look to information warfare carried out in the future. What makes these developments dangerous is that there is a democratization of the technologies. The technologies examined above are and will continue be available to everyone. These technologies and the development of additional more sophisticated video and audio technologies will dramatically blur the lines between visible fact and fiction. Of particular concern are the ways in which these developing technologies produce possibilities for misuse and the construction of false narratives. Some of the ways this technology can

be misused include, making a public figure say something that didn't say, making someone look as if they were somewhere where they weren't, making a public figure say something provocative that they didn't say, making a public figure say something provocative that they didn't say in order to create confusion, generating controversy, inciting violence, inciting violence through the introduction of fabricated emotion in a speakers voice, or simply slandering someone. What is unique about the research is to point out how sophisticated information warfare and influence operations can become using visual and voice alteration technology. Cyber defense must include an understanding of how the technologies examined above will be used and how more sophisticated technologies will be developed in the future.

One final point needs to be stressed, every technology that comes online that can be used to fabricate and distort information will be exploited by individuals, nation states and non-nation state actors to fabricate and distort visual and audio recordings. This means that what is taken to be truthful Information due to the development of technologies, will be capable of being replaced by what has been fabricated, distorted and intentionally presented to audiences to disinform them. The technologies examined above will be further developed and extended into the creation of propaganda, misinformation and disinformation in order to attempt to undermine everything that is contested in individual, group, and nation state conflicts. Due to these developments a clearer distinction must be drawn between misinformation and disinformation. Much of what is currently called misinformation is actually disinformation. Practitioners of cyber offense will continue to claim that fabrications, distortions and disinformation are misinformation As a final point, disinformation, intentionally calling something true that is false, needs to be much more clearly defined and identified as what it actually is, disinformation.

## **References**

- #VoCo.Adobe Max 2016 (Sneak Peeks\_ Adobe Creative Cloud," YouTube video. 7:20, filmed November 2016 in San Deigo, California, posted by "Adobe Creative Cloud," November 4, 2016, <https://www.youtube.com/watch?v=l314XLZ59iw> (accessed Dec 18, 29017).
- Adobe Photoshop, [https://en.wikipedia.org/wiki/Adobe\\_Photoshop](https://en.wikipedia.org/wiki/Adobe_Photoshop)"Adobe Voco 'Photoshop-for-Voice' Causes Concern," BBC News, November 7, 2016, <http://www.bbc.com/news/technology-3789902> (accessed January 18, 2018).
- Chessen, Matt. "Machines Will Soon Program People," @mattlesnake (blog), Medium, May 16, 2017, <https://medium.com/@mattlesnake/machines-will-soon-program-people-73929e84c4c4>.
- Dorn, Lori, "Lyrebird, Remarkable Speech Synthesis Technology That Can Learn and Mimic Voices in Seconds." Laughing Squid (blog), April 24, 2017, <https://laughingsquis.com/lyrebird-remarkable-speech-synthesis-technology-that-can-learn-and-mimic-voices-in-seconds/> (accessed January 20, 2018).
- "Face2Face: Real-Time Face Capture and Reenactment of RGB Videos (CVPR 2016 Oral)," YouTube video, uploaded by Matthias Niessner, March 17, 2016, <https://www.youtube.com/watch?v=ohmajJTpNk>.
- Furrow, Dwight. Ethics: Key Concepts in Philosophy, Continuum, New York, NY. 2005. p. 44.
- Gross, Michael and Tamar Meisels. Soft War: The Ethics of Unarmed Conflict, Cambridge University Press, 2017.
- Helmus, Todd C., Christopher Paul and Russell W. Glenn. Enlisting Madison Avenue: The Marketing Approach to Earning Popular Support in Theaters of Operation, National Defense Research Institute, The Rand Corporation. 2007.
- Hodnicki, Joe, Russia's interference in the 2016 US presidential election and Information warfare, Law Librarian Blog, <https://llb2.com/author/lawlibrarian20/page/56/> Oct. 5, 2017.
- <http://www.digitaljournal.com/tech-and-science/technology/new-software-edits-voices-like-text/article/497243>  
[https://www.tensorflow.org/.mimic\\_a\\_speaker's\\_voice](https://www.tensorflow.org/.mimic_a_speaker's_voice): Bahar Gholipour, "New AI Tech Can Mimic Any Voice," *Scientific American*, May 2, 2017, <https://www.scientificamerican.com/article/new-ai-tech-can-mimic-any-voice/>.  
<https://www.wired.com/brandlab/2018/10/lyrebird-uses-ai-find-artificial-voice/>
- JHU Library, Evaluating Information, Information, <http://guides.library.jhu.edu/c.php?g=202581&p=1334961>
- Information Warfare: Russian Activities, (<https://www.everyreport.com/reports.IN10563.html>). Sept.2, 2016 IN10635.
- Lindner, Isabel and Linda A. Henkel, "Confusing What You Heard with What You Did: False Action-Memories from Auditory Cues," Psychonomic Bulletin & Review 22, no.6 (2015): 1791-97, <http://doi.org/10.3758/s13423-015-0837-0> (accessed December 18, 2017).
- Lomas, Natasha., "Lyrebird Is a Voice Mimic for the Fake News Era," TechCrunch, April 25, 2017, <https://techcrunch.com/2017/04/25/lyrebird-is-a-voice-mimic-for-the-fake-news-era/>.
- Lyrebird,"Copy the Voice of Anyone," SoundCloud, <https://lyrebird.ai/demo> (accessed July 16, 2017)
- Metz, Cade. "Google's Dueling Neural Networks Spar to Get Smarter, No Humans Required," *Wired*, April 11, 2017, <https://www.wired.com/2017/04/google-dueling-neural-networksspar-get-smarter-no-humans-required/>.
- Nguyen, Anh, Jeff Clune, Yoshua Bengio, Alexey Dosovitskiy, et al., "Plug to Play Generative Networks: Conditional Iterative Generation of Images in Latent Space," in Computer Vision and Pattern Recognition (November 30, 2016), <https://arxiv.org/abs/1612.00005> (accessed December 18. 2017)
- PSYOP: Military Psychological Operations Manual. Headquarters, Department of the Army, Mind Control Publishing, 2009.
- Rachels, James and Stuart Rachels. The Elements of Moral Philosophy, McGraw Hill Education, 8<sup>th</sup> ed. 2015.

**Richard Wilson**

- Stewart, Craig.* "Photoshop for audio", "Adobe Prototypes 'Photoshop for Audio,' " Creative Bloq, November 3, 2016, <https://www.creativebloq.com/news/adobe-prototypes-photoshop-for-audio>.
- Thies, Justus , "Face2Face: Real-Time Face Capture and Reenactment of RGB Videos" (unpublished paper, January 2016), <http://niessnerlab.org/papers/2016/1facetoface/thies2016face.pdf>.
- Thies, Justus, Michael Zollhofer, Marc Stamminger, Christian Theobalt, et al., "Face2Face:Real-Time Capture and Reenactment of Rgb Videos," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016): 2387-95, [https://www.cv-foundation.org/openaccess/content\\_cvpr\\_2016/papers/Thies\\_Face2Face\\_Real-Time\\_Face\\_CVPR\\_2016\\_paper.pdf](https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Thies_Face2Face_Real-Time_Face_CVPR_2016_paper.pdf).(accessed December 18, 2017)
- Vincent, James, "This App Uses Neural Networks to Put a Smile on Anybody's Face," Verge, January 27, 2017, <https://www.theverge.com/tldr/2017/1/27/14412814/faceapp-neural-networks-ai-smile-image-manipulation> (accessed 16, 2017)
- "VoCo:text-Based Insertion and Replacement in Audio Narration," YouTube video, 6:02, posted by "Adam Finkelstein," May 11, 2017, <https://www.youtube.com/watch?v=RB7upq8nzIU> (accessed January 18, 2018).
- Vondrick, Carl, Hamed Pirsiavash, and Antonio Torralba, "Generating Videos with Scene Dynamics" (paper presented at the 29th Conference on Neural Information Processing, Barcelona, 2016), <http://carlvondrick.com/tinyvideo/paper.pdf>.
- Warzel,Charlie. "Infocalypse Now," BuzzFeed, February 11, 2018, [https://www.buzzfeed.com/charliewarzel/the-terrifying-future-of-fake-news?utm\\_term=.viEmNOIN3o#.xtPNkBWkwD](https://www.buzzfeed.com/charliewarzel/the-terrifying-future-of-fake-news?utm_term=.viEmNOIN3o#.xtPNkBWkwD).
- Zeyu Jin, Gautham J. Mysore, Stephan Diverdi, Jingwan Lu, and Adam Finkelstein, "VoCo: Text-Based Insertion and Replacement in Audio Narration," ACM Transactions on Graphics 36, no. 4, article 96 (July 2017) <http://doi.org/10.1145/3072959.3073702> (accessed December18,2017).

# ARM Security Alternatives

Raz Ben Yehuda<sup>1</sup>, Roee Leon<sup>1</sup> and Nezer Zaidenberg<sup>2</sup>

<sup>1</sup>University of Jyväskylä, Finland

<sup>2</sup>College of Management Academic Studies, Rishon LeZion, Israel

[rabenyah@student.jyu.fi](mailto:rabenyah@student.jyu.fi)

[roee@trulyprotect.com](mailto:roee@trulyprotect.com)

[scipio@scipio.org](mailto:scipio@scipio.org)

**Abstract:** Many real-world scenarios such as protecting DRM, online payments and usage in NFC payments in embedded devices require a trustworthy “trusted execution environment” (TEE) platform. The TEE should run on the ARM architecture. That is popular in embedded devices. Furthermore, past experience has proved that such TEE platform should be available in source code form. Without the source code 3<sup>rd</sup> parties and user cannot be conducted code review audit. Lack of review put doubt on the system as a trustworthy environment. The popular Android OS supports various TEE implementations. Each TEE OS implementation has its own unique way of deploying trusted applications(trustlets) and its own distinct features. Choosing a proper TEE operating system can be a problem for trust applications developers. When choosing TEE applications developers has many conflicting goals. The developers attempt to ensure that their apps work on as many different Android devices as possible. Furthermore, developers rely on the TEE for certain features and must ensure the suggested TEE provides all the features that they need. We survey multiple ARM TrustZone TEE operating systems that are commonly available and in use today. We wish to provide all the information for IoT vendors and SoC manufacturer to select a suitable TEE.

**Keywords:** virtualization, ARM architecture, TrustZone, trusted computing

---

## 1. Introduction

The proposed solutions for creating TEE on the ARM architecture are all using TrustZone™ feature. TrustZone™ is a unique privilege level on ARM (ARM 2009) whose purpose is to create a Trusted Execution Environment (TEE) (Zaidenberg 2018). Trustzone™ can be found on virtually all modern mobile phones. Additionally, TrustZone™ can be found in other ARM based systems on chip, such as AMD with their "Hiero falcon", AppleMicros X-Gene3, Cavium Thunder X and other systems.

Trust Execution Environment is required on many scenarios. TEE use cases include providing Digital right management (DRM) support. DRM requires a root of trust that can store decryption keys on the end point so that it will not be available to outsiders (Zaidenberg et al 2015b). Using virtualization to create root of Trust was attempted by SONY on PS4 and later shown on PC and MIPS by (Averbuch et al 2013). A complete system for reverse engineering protection by distribute keys for encrypted code execution after attestation (Resh et al 2017) introduced the concept of protecting code and registers by the hypervisor. (Kiperberg et al 2019) proposed creating buffers of protected code in the CPU case using memory addresses that only the hypervisor can address. Virtualization can also provide end point security (Resh et al 2017). Hypervisor can also be used for development aid by catching hypercalls (Khen et al 2011) connecting virtual debug hardware (Khen et al 2013) forensics ( Zaidenberg et al 2015) and Forensic memory dump (Kiperberg et al 2019b). Last hypervisor can be used for end point attestation as shown by (Kiperberg et al 2013, Kiperberg et al 2015) etc.

Using virtualization as a tool for cyber security is also a common practice. Virtualization was initially designed for dynamic provisioning of computing resources. However, virtualization offers higher privileges and execution permissions that can be used to detect threats as well as serve as TEE. ARM didn't offer virtualization or TrustZone™ support until the ARM7a hardware. ARM virtualization and TrustZone™ technologies were proposed as two optional additions. These technologies are available in some 32bit ARM7a models. ARM virtualization and TrustZone™ are now part of the 64bit ARM8a architecture and offered in all ARM8a devices.

ARM vendors offer their own TEE implementations. Some TEE implementations such as Trustonic's TSP and Qualcomm's QSEE are closed source while others are open source or offer providing the TEE source code for a fee. We survey Trusted computing alternatives for implementations. We mainly consider alternatives with available source code. All the surveyed solutions offer a complete solution for the TrustZone™ environment. We also survey some ARM virtualization (not TrustZone™) based alternatives.

## **2. Background**

### **2.1 Trusted Execution Environment**

The ARM architecture design allows both a Trusted Execution Environment (TEE) and a Rich Execution Environment (REE, i.e. normal ARM OS e.g. Android or iOS) to run simultaneously. The Trusted Execution Environment is a secure “area” inside a main processor not effected by the REE operating system. The Trusted Execution Environment runs its own operating system and uses its own set of register and its own memory management unit (MMU). The TrustZone™ operating system is a separate operating system that is running in parallel to the main operating system in an isolated environment. The Trusted Execution Environment guarantees that the code and the data loaded in the TEE are protected with respect to confidentiality and integrity from all application that run on the REE OS.

The Rich Execution Environment is a separate privilege level inside the main processor. The Rich Execution Environment runs its own separate operating system (compared to TEE). The Rich Execution Environment is the standard operating system that the device is running, usually the REE is Google’s Android or Apple’s iOS. The Rich Execution Environment offers significantly more features and applications than the TEE. This is by design and application are supposed to run on the REE and not on the TEE. As a result of offering more features, the attack surface against the REE is much larger and therefore, the REE is more vulnerable to attacks. The Rich Execution Environment receive services such as decryptions and storing decryption keys from the Trusted Execution Environment. According to ARM design the TEE acts as a monitor service for the REE.

The TEE has higher permissions than the REE as well as access to the REE memory, MMU, registers and data structures. The REE should not have access to the TEE memory and data structures. In ARM terminology, the two execution environments are called worlds, the secure world (TEE) and the non-secure world(REE). Context can be switch between secure and insecure worlds through the supervision the “Secure Monitor” running in monitor mode(TrustZone). This switch from secure to insecure world is performed through a special architecture specific assembler instruction called “secure monitor call” or smc. In order to communicate between the secure and non-secure world the user creates a shared memory segment. TrustZone™ splits the SoC device to the secure and non-secure worlds. TrustZone™ controls all the device hardware interrupts. TrustZone™ can route any interrupt to the secure world or to the non-secure world. Like in the memory case, I/O and interrupts routing may change dynamically. TrustZone™ uses its own MMU. Operating systems and Processes that execute in TrustZone™ do not share the same address space with their non-secure world counterparts. Thus, there is no need to have distinct TrustZone™ for each processor. A single TrustZone™ OS can run across multiple ARM processors/cores and manage all the device trusted computing needs. The ARM architecture cryptographic keys are accessible only in TrustZone™, The manufacturer can provide each CPU or platform with device specific keys using e-fuses. These keys are device specific, thus enabling protection in the end unit granularity level.

(For example distributing video that only specific device can decode etc.)

Booting a Trusted Execution Environment must form a chain of trust in which a trust nexus verifies the next component on the boot chain. Each component verifies the next component until the system.

### **2.2 ARM permission model**

The ARMv8 architecture has a unique approach to privilege levels. The ARM platform normally has 4 exception (permission) levels.

ARM also has secure world (TrustZone™) and normal world (non TrustZone™)

ARM Exception levels are described in Table 1 Each of the exception levels provide its own state of special purpose registers and can access these registers of the lower levels but not higher levels. The general-purpose registers are shared.

Thus, moving to a different exception level on the ARM architecture, does not require the expensive context switch that is associated with the x86 architecture.

**Table 1:** Arm exception levels

Exception level	Meaning	Notes
Exception Level 0 (EL0)	Refers to user space code.	Exists in both secure and normal world This is analogous to "ring 3" in x86 platform.
Exception Level 1 (EL1)	Refers to operating system code.	Exists in both secure and normal world This is analogous to "ring 0" in x86 platform.
Exception Level 2 (EL2)	Refers to HYP mode. ARM hypervisor privilege level	Exists in both secure and normal world This is analogous to "ring -1" or "real mode" on the x86 platform.
Exception Level 3 (EL3)	TrustZone™	Refers to TrustZone™ as a special security mode that can monitor the ARM processor and may run a security real time OS. There is no direct analogous modes but related concepts in x86 are Intel's ME or SMM. Naturally it exists only on secure world

ARM7 architecture is similar to ARM8. ARM7 offers virtualization as an extension that is only available to some late ARM7 models. ARM7 also offer TrustZone™ as separate extension. Furthermore, ARM7 is 32bit architecture while ARM8v is 64bit (and 32bit) architecture.

### 3. Virtualization vs. TrustZone™ mode

The first question we must address is how the operating system should be verified. The REE operating system can be verified using HYP mode or TrustZone™. ARM has designed the TrustZone™ mode specifically for attesting and monitoring the Rich operating system. The benefit of using TrustZone™ is that it reserves the HYP mode for real hypervisor without the need to use features such as nested virtualization. Furthermore, only the vendor can install software on the TrustZone™ mode. In some cases, even the vendor (i.e. the manufacturer of the device or phone, not the CPU vendor) has limited access and cannot install software in TrustZone™ mode. However, no such limitation exists on HYP mode. Everybody can install software in HYP mode with no special limitations. This usually makes hypervisor code easier to install. From the device vendor standpoint therefor it is assumed that TrustZone™ is available. (or possibly available) From software vendor standpoint TrustZone is not available but virtualization may be.

The two main drawbacks of using virtualization are that virtualization mode is no longer available for other software that may want to run there. Also, TrustZone™ is monitored on boot by the BSP (Board Support Package) it cannot be modified or replaced as easy as the hypervisor boot loader or driver.

Resh et al (2017) and Seshadri et al (2007) both provide examples of using hypervisor for end point security.

We examine several hypervisor implementations for completion. However it is assumed that a TrustZone™ solution is preferable whenever TrustZone™ is available.

#### 3.1 Virtualization classification

Virtualization is the process of running multiple Operating system on a single hardware or running microkernel to manage single operating system. Hypervisor is the software that provides virtualization. Hypervisors are classified to two modes:

- 1. Full virtualization - The guests operating system is not modified in any way.
- 2. Para-virtualization – The guest operating system is aware it runs as guest. The guest's operating system code is modified. The guest operating system does not attempt to communicate directly with the hardware. Instead the guest operating systems uses hypercalls to communicate with the host hypervisor. When the guest's operating system needs the host for example for I/O access and sometimes in critical sections. The hypercalls trap to the hypervisor to perform a service on behalf of the guest. Using para-virtualization and hypercalls usually yields better performance.

In the taxonomy of virtualization environment, virtualization environments are categorized by their design.

- 1. Complete monolithic - A single software responsible to provide access to the hardware to the guests. For example, VMware ESXi server.

- 2. Partially monolithic - The technology is an extension to the general-purpose operating system, such as KVM in Linux and VMWare Desktop or Microsoft Hyper-V
- 3. MicroKernel These are light weight micro kernels that provide a minimal set of services to the guests, mainly CPU virtualization and hardware access. Xen and seL4 are examples for such micro hypervisors.

Last virtualization pioneers Popek and Goldberg (Popek et al 1974) classified hypervisors to type I and type II

Type I hypervisors (or boot hypervisors) – are hypervisors that start at boot and start various guest operating systems. Examples include VMWare ESXI, Xen and IBM S/390 VM.

Type II Hypervisors (or hosted hypervisors) – are hypervisors that starts under a host operating system that already booted and took control of the machine. Examples include VMWare desktop.

For security and trusted computing purposes only Type I hypervisors are of interest. Type II hypervisors can be disabled by the host OS and thus serve no security purpose.

## **4. Alternatives review**

### **4.1 GlobalPlatform**

GlobalPlatform is a non-profit organization that consists of an alliance of many mobile device manufacturers. GlobalPlatform defines and publishes the standards for mobile devices including a standard for secure digital services for mobile devices. GlobalPlatform (2011) is the current industry standard for TEE platform under ARM.

### **4.2 General Dynamics OKL4**

OKL4 is a microkernel that was originally developed, maintained and distributed by Open Kernel Labs.

The OKL4 operating system was based on the L4 operating system by Liedtke (Liedtke 1996).

The L4 microkernels family in its earlier form was called L3. L3 was a microkernel that was developed by Liedtke in the 1980's on an i386 system and was deployed in few thousands' installations, mainly education institutes. L3 suffered from a high overhead of Inter-process communication (IPC) communication which was over 100us. Liedtke, trying to reduce the IPC overhead problem, had re-implemented L3 completely and reduced significantly the IPC overhead to 5us, on i486. This new design was referred as L4.

L4 had evolved over the years and became a family of L4 microkernels, to name a few, L4-embedded, Codezero, NICTA, sel4 etcetera. NICTA was maintained by OpenLabs which renamed it to OKL4 microkernel and stopped the open source development.

OKL4 (Heiser et al 2010) is deeply discussed in peer reviews press. The OKL4 microvisor supports both paravirtualization and pure virtualization. It is designed for the IoT industry, and supports, ARMv5, ARMv6, ARMv7ve and ARM8va. OKL4 is focused on embedded devices. OKL4 was originally open source software.

On 2012 general dynamics acquired the Open Kernel Labs. After being acquired General Dynamics changed OKL4 source code policy from open to closed source project. The latest available open source OKL4 is from May 2013 and is still available to download from archive.org (and other sources). OKL4 has a sister open source project (supported by GeneralDynamics) called seL4 which is described later.

Installing OKL4 and running it is a challenging task that requires expertise. Open source OKL4 code must also be adapted to modern hardware. Today OKL4 is still under development and support of General Dynamics.

One can also obtain the current OKL4 source code under suitable license and NDAs.

We refer to the latest available open source OKL4 from 2013 and not to current releases (for which source code is not available) thus we are not up to date with current releases (compared to other Trusted Execution Environment alternatives).

### **4.3 Google Trusty TEE**

Trusty is a secure Operating System (OS) that was developed by Google.

Trusty provides a Trusted Execution Environment (TEE) for The Android (only) Operating system.

The Trusty OS doesn't require security specific hardware. Instead, Trusty TEE runs on the same processor as the normal Android OS. However, despite running on the same CPU, Trusty is isolated from the rest of the system. This is done using ARM TrustZone™ features that enable separate MMU for trusty (in TrustZone™) and the normal world OS. TrustZone™ allows Trusty to create an isolated secure execution environment and provide certain services to the non-secure (i.e. Android) OS.

Trusty consists of:

- A small operating system kernel. The trusty Kernel is derived from Little Kernel. Little Kernel is a small operating system that is also used as Android boot loader.
- A Linux kernel driver that acts as mediator between the TEE (Trusty) and REE (Android) environments
- An Android user space library that provide to communication between the REE (Android) and TEE (Trusty) applications using the kernel driver

Trusty is compatible with ARM and Intel processors. On ARM systems, Trusty uses ARM's Trustzone™ to virtualize the main processor and create a secure trusted execution environment. Similar support is also available on Intel x86 platforms using Intel's Virtualization Technology.

### **4.4 Linaro OP-TEE**

Linaro security working group and STMicroelectronics have teamed to create OP-TEE.

OP-TEE follows GlobalPlatform specification (2011) and implements version 1.1 of GlobalPlatform TEE client API and TEE internal API. OP-TEE is an open source project and is widely available under BSD 2-clause license and (Kernel parts) GPLv2 license. According to (Bech 2014) and our testing OP-TEE has a small foot print and minimal effect on the running system. OP-TEE has a large community support. Like Trusty above OP-TEE consists of 3 main components.

- A light weight secure operating system. The OP-TEE operating system consists of several modules such as memory management, interrupt handling, etc. In addition, OP-TEE implements a hardware abstraction layer as it supports various processors and hardware. The OPTEE operating system also provides a capability to run user-space applications (typically referred to as Trustlets) in the secure world. These applications are provided with the GlobalPlatform TEE Internal API which allows them to ask for internal, secure-only, OS services
- A non-secure user-space client that is composed of two components: (1) a user-space/kernel-space mediator and (2) libraries that implement the GlobalPlatform TEE Client API.
- A kernel driver that simply performs the transitions between the secure and non-secure worlds

## **5. Other alternatives**

There are other TEE options that the vendor can use for both EL2 and EL3. These alternatives are not as common or as strong as the alternatives mentioned above and fail in atleast one pre-requisite.

### **5.1 Jailhouse**

Jailhouse was announced by Siemens in November 2013. Baryshnikov (Baryshnikov 2016) has analyzed the Jailhouse system. Jailhouse is a type 2 partitioning microvisor for Linux hosts.

A partitioning microvisor is a microvisor that controls the OS access to resources and isolates the resources from the general-purpose operating system or other guests. Partitioning in microvisor context means the microvisor performs strict allocation of the system resources. The hosting Linux is referred as the Root cell, and the guests are called inmates. Jailhouse itself is not an operating system, it is a resource access controller. Jailhouse is

controlled from the Linux host, and reveals information stored to the Linux host (the root cell), but not the guests (the inmates).

Jailhouse is a bare metal hypervisor, and in most cases, it is pure virtualization hypervisor, and as such can run many types of operating systems, such as FreeRTOS (Barry 2008), Erika3 (Evidence 2019)``, Linux and Zephyr. Jailhouse supports ARMv8, ARMv7a, and x86\_64 architectures. Jailhouse requires the machine to have at least two processors. One processor is used to run the hosting Linux, and the other processors may be assigned to Jailhouse. Jailhouse requires the Linux kernel to provide a contiguous memory at boot time. It requires a memory footprint of few tens of megabytes, usually 50 megabytes.

The Jailhouse configuration is performed through a tool provided by Jailhouse. This tool scan sysfs and procfs, and generates a device tree that describes the hardware as seen by the Linux host. This Jailhouse device tree is referred as the cell configuration file. The user can edit the cell configuration file to create a correct guest configuration. Jailhouse targets the automation, robotics and IoT industries.

## **5.2 QSEE**

QSEE is Qualcomm Secure Execution Environment. In the past it was based on OKL4 until GeneralDynamics and Qualcomm failed to reach agreement regarding licensing. Since 2015-6 Qualcomm has developed QSEE from scratch with no (direct) connection to GeneralDynamics.

QSEE is closed source (and Qualcomm does not provide source code licenses) Therefore QSEE fails our precondition of source availability and is not part of this survey. Prior releases of QSEE suffered from several well documented security problems.

(Beniamini 2016)

## **5.3 seL4**

sel4 like OKL4 is also a based on the L4 microkernel. sel4 (Klein et al 2009) is a microvisor that was implemented by Open Kernels labs (and later GeneralDynamics). seL4 is not as popular as OKL4. One of the strong features of seL4 is the fact that it has been formally verified to be correct. The method of verification is definition of seL4 exact functional specification and proving its operations using rigorous logical means. Later the seL4 codebase was reimplemented in C to make it efficient. Despite the rigorous testing seL4 is not necessarily bug free. The implementation has some assumptions of correctness about the compiler, architecture and C reimplementation.

seL4 compared to OKL4, is an open source kernel. seL4 is the sole kernel that is mathematically proven secured and safe. L4-embedded, or NICTA embedded, was adopted by Qualcomm as a real time operating system for their wireless modem processors firmware.

The basic rules of the L4 kernel design are minimalism. Leidtke (1996) formulated the rule of minimization as follows:

*"A concept is tolerated inside the u-microkernel only of moving it outside the kernel, i.e. permitting competing implementation would prevent the implementation of system required functionality".*

This principle, known also as the no-policy in the kernel is the core of the L4 microkernels design. Though operating systems tends to grow in size over the years, L4 footprint is considerably low. seL4 footprint is 9600 lines of code. As a side effect of the minimization and performance, L4 microkernels, do not strive to hardware abstraction. Half of the seL4 microkernel is agnostic to the underlying hardware.

L4 also demonstrates a new resource management scheme where all memory allocations are user space driven. Another interesting feature of the L4 microkernels, is the fact that interrupts are disabled while executing in kernel mode. This approach simplifies the implementation, increases the performance and eases the kernel verification. Direct process switch, which in general means that sel4 tries to avoid from using the scheduler, is another interesting facet of L4. When a thread reaches a pre-emption point, the kernel switches to the first runnable thread, which in turn, executes on the time slice of the pre-empted thread.

seL4 runs on ARM, supports SMP and Uniprocessor. Like OKL4 seL4 also provides real-time support.

seL4 was recently ported to ARM8 architecture (Heiser 2019)

#### **5.4 TrustTonic**

TrustTonic is known for its TrustZone technology in the mobile world, mainly Android. TrustTonic operating system, Kinibi, is a closed source operating system. Kinibi is wide spread in the Android cellphone world. Kinibi provides data encryption and device authentication. It also gives safe access to peripherals, such as the touch screen, NFC and finger print reader, through its TEE. Since peripheral I/O is provided using the TEE no malware in the REE will affect the I/O. In addition, Kinibi can isolate sensitive code execution and secure data.

Kinibi is verified by the chain of trust, i.e; it is verified by the bootloader each time the device boots.

Furthermore, TEE application has access to the network. This way, a trusted application can access remote services securely.

TrustTonic can also be found in the automotive industry. In this area, TrustTonic approaches data leakage, application overlapping and application re-packaging attacks. Application overlapping attack is an interception technique for stealing sensitive I/O, such as when a user enters its password. A repacking of an application is a method of modifying a program to steal sensitive data. For example, adding a log entry that prints sensitive information.

Trustonic offers an SDK, compliant with GlobalPlatform API standards, to help build Trusted Applications.

Since no open source version of Kinibi exists (not even older version) we left it out of this survey.

#### **5.5 Xen**

Xen was announced in 2003. Xen was developed initially at the university of Cambridge, by Ian Pratt. Xen is a micro kernel hypervisor. Xen provides CPU virtualization using virtual interrupts, MMU and under-guest communication. In Xen, a virtual machine is referred as Domain. Domain0, also known as Dom0, is the first domain. Dom0 must run before any other virtual machine. Dom0 is usually Linux or BSD, and DomU is a virtual machine on top of the other domains. Domain0 requires access to the entire machine's hardware. Domain0 responsibility is management through the Linux kernel. Xen's event channel provides communication between Dom0 and DomU. Whenever DomU issues a virtualized event it uses this event channel. The event channel is used for para-virtualized guests. For a full virtualized guest Xen uses QEMU. Xen's tool stack is the management tool used to control guests. The fact that Xen uses Linux as Dom0 provides Xen with abundant of hardware support and Linux software. Xen boots from the bootloader and is then loads the a para-virtualized host.

Xen's I/O virtualization comes with a performance cost. Virtualized I/O accesses and virtualized interrupts from DomU Xen guests are delegated to Dom0. In addition, if a host interrupt occurs while DomU runs, then this interrupt would be served only when Dom0 is gets the processor. Thus, interrupts and events have a performance overhead.

Xen is available in ARM and x86, runs on SMP and UP. Xen is licensed GPL.

#### **5.6 Xvisor**

Xvisor was announced in Apr 2012. Xvisor (Patel et al 2015), is a monolithic, type 1 hypervisor that is independent of Linux. Xvisor is a monolithic hypervisor that controls the hardware peripherals. Xvisor provides a minimal operating system, thus Xvisor is not a microkernel. Xvisor can emulate devices and provides a path-through access to real devices. As an operating system, Xvisor has a memory management, scheduler, load balancer and threads. Xvisor does not support processes and is not POSIX compliant. Xvisor supports SMP, so that a guest can use two or more processors. There are no restrictions on the number of processors, and Xvisor can also execute on a single processor.

Xvisor provides an IPC between two guests through the use of aliased guest region, which is

a GPA (guest physical address) shared between two guests. In the Xvisor taxonomy, a processor can be Normal VCPU or an Orphan VCPU.

Normal VCPUs serve guests OSes, and Orphan VCPUs belongs to the hypervisor. Xvisor support ARM 32bit and 64bit and x86. Its footprint is less than 10MB, however, since it is a type 1 hypervisor, it is required to change the boot loader. Xvisor is widely targets the infotainment market in the automobile world, and automation in general. It is licensed GPL.

## **6. Discussion**

We surveyed most well known TrustZone and HYP alternatives. If the user requires open source and community review as security requirements then both trust-tee and OP-TEE provide good alternatives.

OP-Tee provides the benefit of following GlobalPlatform standard and is more light weight and provide less features than google Trust-Tee. OP-Tee has been an open source OS but the source is only available to paying customers. However, OP-Tee (and seL4) offer real time support that are required for some systems. We also reviewed seL4 which is also open source. seL4 is an option but is not as widely used as OP-Tee and Trust Tee (or OKL4).

Furthermore, TrustZone™ and other extensions can be used for other means such as real-time processing (Ben Yehuda et al 2018) and control flow analysis (Abera et al 2016). Using the TrustZone™ architecture for other purposes is an interesting future research area

## **7. Conclusion**

We surveyed the popular TEE alternatives available today. Each alternative has its own benefits and drawbacks. SoC vendors and Platform manufacturers can choose the desired implementation based on their requirements and preferences. Out of the popular alternatives we believe that the free alternatives Google TrusTEE and OP-TEE offer sufficient features and fair replacement for OKL4. We believe OKL4 has none security benefits in strict real time environments that are beyond the scope of this review.

## **References**

- Abera, T., Asokan, N., Davi, L., Ekberg, J. E., Nyman, T., Paverd, A., Sadeghi A.R & Tsudik, G. (Abera et al 2016). C-FLAT: control-flow attestation for embedded systems software. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (pp. 743-754). ACM.
- Averbuch, A., Kiperberg, M., & Zaidenberg, N. J. (Averbuch et al 2013). Truly-protect: An efficient vm-based software protection. *IEEE Systems Journal*, 7(3), 455-466.
- ARM, Architecure (ARM 2009). Security technology building a secure system using TrustZone™ technology (white paper). ARM Limited.
- Barry, R. (Barry 2008). FreeRTOS. *Internet, Oct*
- Baryshnikov, M. (Baryshnikov 2016). Jailhouse hypervisor (Bachelor's thesis, České vysoké učení technické v Praze. Vypočetní a informační centrum.).
- Bech, J. (Bech 2014). OP-TEE, open-source security for the mass-market. Core Dump. <https://www.linaro.org/blog/op-tee-open-source-security-mass-market/>
- Ben Yehuda R. and Zaidenberg N. J (Ben Yehuda et al 2018) Hyplets - Multi Exception Level Kernel towards Linux RTOS In the Proceedings of the 11th ACM International Systems and Storage Conference Systor 2018 pp 116-117
- Beniamini G. (Beniamini 2016) "Extracting Qualcomm's KeyMaster Keys - Breaking Android Full Disk Encryption". <https://bits-please.blogspot.com/2016/06/extracting-qualcomms-keymaster-keys.html>
- Evidence Sri (Evidence 2019) "ERIKA Enterprise 3 source code" <https://github.com/evidence/erika3>
- Heiser, G., & Leslie, B. (Heiser et al 2010). The OKL4 Microvisor: Convergence point of microkernels and hypervisors. In Proceedings of the first ACM asia-pacific workshop on Workshop on systems (pp. 19-24). ACM.
- Heiser, G (Heiser 2019) "Whats new in the world of seL4" FOSDEM 2019 <https://www.youtube.com/watch?v=6s5FDX5PkZI>
- Khen, E., Zaidenberg, N. J., & Averbuch, A. (Khen et al 2011). Using virtualization for online kernel profiling, code coverage and instrumentation. In *2011 International Symposium on Performance Evaluation of Computer & Telecommunication Systems* (pp. 104-110). IEEE.
- Khen, E., Zaidenberg, N. J., Averbuch, A., & Fraimovitch, E. (Khen et al 2013). Lgdb 2.0: Using lguest for kernel profiling, code coverage and simulation. In *2013 International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS)* (pp. 78-85). IEEE.
- Kiperberg, M., & Zaidenberg, N. (Kiperberg et al 2013) Efficient Remote Authentication. In *Proceedings of the 12th European Conference on Information Warfare and Security: ECIW 2013*(p. 144). Academic Conferences Limited.
- Kiperberg, M., Resh, A., & Zaidenberg, N. J. (Kiperberg et al 2015). Remote Attestation of Software and Execution-Environment in Modern Machines. In *2015 IEEE 2nd International Conference on Cyber Security and Cloud Computing* (pp. 335-341). IEEE.

- Kiperberg M, Algawi A, Leon R, Resh A. & Zaidenberg N. J. Hypervisor-assisted Atomic Memory Acquisition in Modern Systems (Kiperberg et al 2019b) in Proceedings of 5<sup>th</sup> international conference on information system security and privacy ICISSP 2019
- Kiperberg, M., Leon, R., Resh, A., Algawi, A., & Zaidenberg, N. J. (Kiperberg et al 2019). Hypervisor-based Protection of Code. *IEEE Transactions on Information Forensics and Security*.
- Klein, G., Elphinstone, K., Heiser, G., Andronick, J., Cock, D., Derrin, P., Elkaduwe D., Engelhardt K., Kolanski R. ,Norrish M., Sewell, T., Tuch H. & Winwood S. (Klein et al 2009) . sel4: Formal verification of an OS kernel. In Proceedings of the ACM SIGOPS 22nd symposium on Operating systems principles (pp. 207-220). ACM.
- Liedtke, J. (Liedtke 1996). Toward real microkernels. *Communications of the ACM*, 39(9), 70-77.
- Patel, A., Daftedar, M., Shalan, M., & El-Kharashi, M. W. (Patel 2015). Embedded hypervisor xvisor: A comparative analysis. In Parallel, Distributed and Network-Based Processing (PDP), 2015 23rd Euromicro International Conference on (pp. 682-691). IEEE.
- Popek, G. J., & Goldberg, R. P. (Popek et al 1974). Formal requirements for virtualizable third generation architectures. *Communications of the ACM*, 17(7), 412-421
- Resh, A., Kiperberg, M., Leon, R., & Zaidenberg, N. (Resh et al 2017b). System for Executing Encrypted Native Programs. *International Journal of Digital Content Technology and its Applications*, 11
- Resh, A., Kiperberg, M., Leon, R., & Zaidenberg, N. J. (Resh et al 2017). Preventing Execution of Unauthorized Native-Code Software. *International Journal of Digital Content Technology and its Applications*, 11.
- Zaidenberg, N. J. (Zaidenberg 2018). Hardware Rooted Security in Industry 4.0 Systems. *Cyber Defence in Industry 4.0 Systems and Related Logistics and IT Infrastructures*, 51, (pp 135-151).
- Zaidenberg, N. J., & Khen, E. (Zaidenberg et al 2015). Detecting Kernel Vulnerabilities During the Development Phase. In *2015 IEEE 2nd International Conference on Cyber Security and Cloud Computing* (pp. 224-230). IEEE
- Zaidenberg, N., Neittaanmäki, P., Kiperberg, M., & Resh, A. (Zaidenberg et al 2015b). Trusted Computing and DRM. In *Cyber Security: Analytics, Technology and Automation* (pp. 205-212). Springer, Cham.

# IoT Architectures for Critical Infrastructures Protection

**Alin Zamfiroiu, Bogdan Iancu, Catalin Boja, Tiberiu Georgescu and Cosmin Cartas**

**The Bucharest University of Economic Studies, Romania**

[alin.zamfiroiu@csie.ase.ro](mailto:alin.zamfiroiu@csie.ase.ro)

[bogdan.iancu@ie.ase.ro](mailto:bogdan.iancu@ie.ase.ro)

[catalin.boja@ie.ase.ro](mailto:catalin.boja@ie.ase.ro)

[tiberiugeorgescu@ase.ro](mailto:tiberiugeorgescu@ase.ro)

[cosmin.cartas@csie.ase.ro](mailto:cosmin.cartas@csie.ase.ro)

**Abstract:** Critical Infrastructures, such as telecommunications, power supply and transport, carry out essential functions in modern states. In today's networking world, it is possible that serious damages can extend beyond national borders and affect other states. Infrastructure protection has therefore become a global challenge that needs to be addressed. One of the criteria for assessing the approaches of different countries is whether or not there is a national, convincing Critical Infrastructure Protection strategy. The issue of security has always been an area in which research is being carried out continuously. Internet of Things (IoT) Security is no exception. The number of smart devices is steadily increasing, and the connectivity options are diversifying with the evolution of information technology. Companies and other organizations implement different IoT architectures, but rarely this technology was adopted by the simple customer. Currently, this technology is accessible to individual users. International standards and protocols have been established and products manufactured by different companies have been linked together. Taking these issues into account, people have started to show increased interest in the field of IoT in recent years. In the last period there have been many cyber-attacks targeting IoT technologies all around the world. Based on these cyber-attacks we can learn how to protect our Critical Infrastructure against future attacks and how to improve and adapt our management policies to the new models of attacks. In this paper we present an architecture for behavioural based device identification. In this regard, if the devices will have an unusual behaviour, they can be excluded from system. In conjunction with that, we take into consideration a security model that is using semantic reasoning for critical infrastructure. By using existing ontologies together with specially designed ontologies we can annotate important metadata from the analysed infrastructure. In this manner the semantic reasoner can deduct knowledge that is not explicitly present in the ontology, but can play a key role in assessing the infrastructure's security.

**Keywords:** IoT, architecture, critical infrastructure, security, system, device

---

## 1. Critical Infrastructure protection

Critical Infrastructures (CIs), such as telecommunications, power supply and transport, carry out essential functions in modern states. In today's networking world, it is possible that serious damage extends beyond national borders and affects other states. Infrastructure protection has therefore become a global challenge that needs to be addressed. Today, no state can ignore the need to revise and continuously improve the protection of the infrastructure. However, the definitions of critical infrastructures in different countries are as divergent as the concepts of infrastructure protection that have been developed in these countries: since it is possible to identify some common structural elements between countries, the measures taken so far, achieved by the organizations responsible for infrastructure protection and the level of protection obtained so far differs largely (Choras et al., 2016).

Three main categories for the protection of critical infrastructures around the world can be identified in Critical Infrastructure Protection (2004):

- the first is the Critical Information Infrastructure Protection (CIIP) approach. It serves for software security and protection inside the infrastructure. Protection of physical components is provided separately. Private sector integration is tried at all CIP levels;
- the second approach involves both software protection and the physical protection of critical infrastructures. Physical protection is part of all civil defence models. There is no clear separation between individual components. In addition, the important role of national defence ministries should be emphasized, due to their coordinating role. This approach is called the All Hazards;
- the third and last approach is only found in the Chinese model. Under this approach, there is no cooperation between the public and the private sector.

One of the criteria for assessing the approaches of different countries is whether or not there is a national, convincing Critical Infrastructure Protection (CIP) strategy.

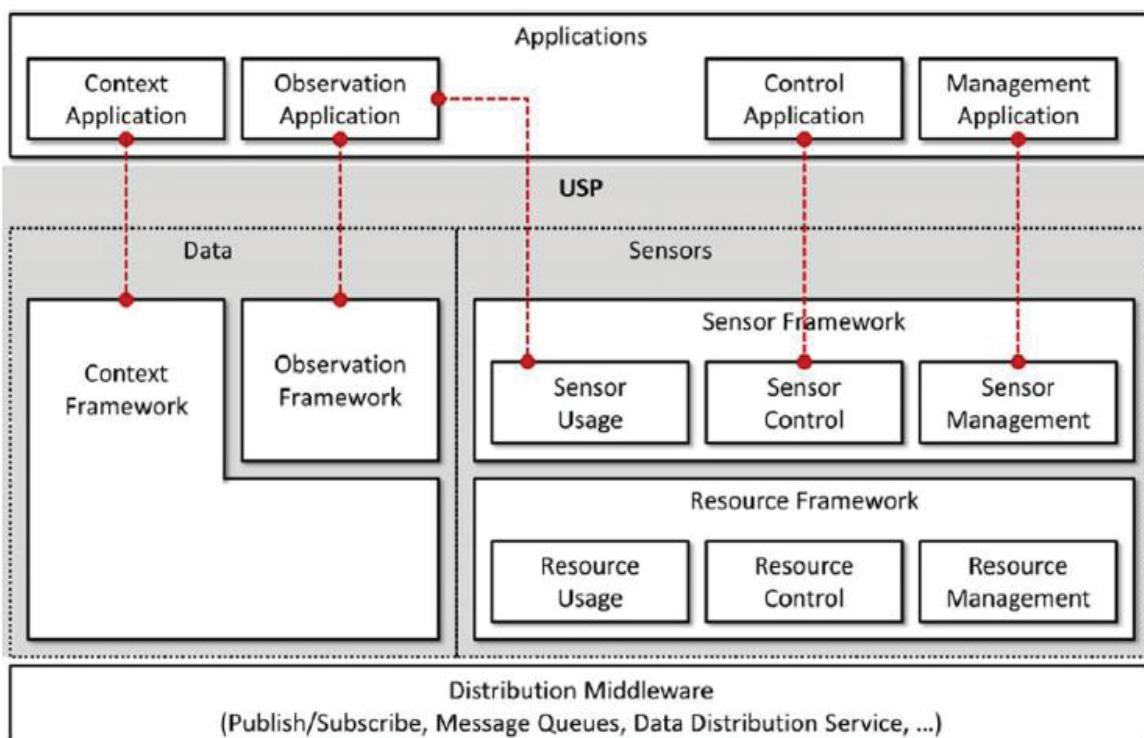
In this context, they refer to the lack of clear definitions of what needs to be done in the field of national critical infrastructure protection. They also point out that functions and competencies are rarely clearly delineated and localized. The fact that most countries do not carry out any independent national analysis of threats is considered to be another shortcoming. Frequently, the American perception of the threat is adopted without change. Often, in these countries, the only analysis done is related to dependence and interdependence; asset analysis activities are limited to the public sector only.

Another important perspective regarding Critical Infrastructures, is that, not depending on the socio-economic field to which relates, they are more and more based on IoT devices, sensors or autonomous systems that communicate to each other in order to deliver data and services. A Critical Infrastructure can include important Data or Logistics nodes that implement Security Policies and may act as Command and Control Centers, but they all depend on the little things that control or measure real devices behaviour.

## 2. Security IoT architectures

In (Ray, 2018) there is presented a synthesis of IoT systems-specific architectures as well:

- architectures based on RFID: EPC, uID, NFC;
- Service-based architectures (SOA) through middleware enhancement;
- Wireless Sensor Network (WSN) based architectures; the most widely used communication protocol is IEEE 802.15.4 but also TCP / IP based protocols such as TinyTCP, mIP, iWIP or USP (Unified Sensor Detection Platform), Figure 1;
- USP architecture stratifies publishing / subscribing, message queues, data distribution services through data-based USP and sensor sensors. Sensors and resource frameworks perform sensory-oriented control and use operations by contextually observing resource efficient management to various top-level applications.
- architectures based on the Supply Chain of Management (SCM) or SCM as a service (IoTMaaS).



**Figure 1:** USP architecture. Source: (Ray, 2018)

The issue of security has always been an area in which research is being carried out continuously (Toma and Popa, 2018). IoT is no exception. In (Ray, 2018) some IoT-based architectures are also presented to address security issues:

- OSCAR (Object-based Security Architecture) supports features such as cache and multicast and provides a mechanism to protect against repeated attacks;
- End-to-End security is based on Datagram Transport Layer Security (DTLS) protocol; the architecture develops the data flow and the underlying communications between the subscriber's certification authority, the gateway, the access server, and the Internet authority. Architecture is analysed from a performance point of view in a health system in (Moosavi, 2018);
- Protection of multimedia traffic is done through the MTSA architecture, which reduces the complexity of multimedia calculations and decreases the size of the actions taken (Naserian et al. 2018).

According to (Kyriazis, 2018) the components of critical infrastructures are services, and the main problem is to not have downtime for these services. In this way all the components and the infrastructure for these services will be secured to assure the continuity and availability of the services.

In (Ahmed, 2017) a comparison between more IoT architectures is made. Five architectures are analysed regarding the Manageability, Security& Privacy, Mobility, Cost-Effectiveness, Efficiency and QoS.

### **3. IoT threats in critical infrastructure**

The number of smart devices is steadily increasing, and the connectivity options are diversifying with the evolution of information technology. The study in Gartner (2016) estimates that in 2020 there will be approximately 21 billion interconnected smart devices. In this context, IoT has become one of the most important modern IT technologies. The term IoT has been mentioned for the first time in a scientific manifestation in 1999 (Ashton, 2009), but the concept of interconnecting intelligent devices incorporated with software, electronics and sensors has emerged more than three decades ago. In 1982, a team at Carnegie Mellon University modified a Coca-Cola machine to be connected over the Internet with a computer where the inventory and bottle temperature were automatically reported. This is considered the first IoT solution (Palermo, 2018).

Solutions to IoT can be classified into two broad categories according to their purpose and beneficiaries: commercial (or consumer) products and industrial solutions (Winder, 2016). Until recently, the interconnection of many smart devices was relatively expensive and complex. Companies and other organizations used IoT, but rarely this technology was adopted by the simple customer. Currently, this technology is accessible to individual users (up to a certain degree). International standards and protocols have been established and products manufactured by different companies have been linked together. Taking these issues into account, people have started to show increased interest in the field of IoT in recent years.

In the last year there have been many cyber-attacks targeting IoT technologies presented in Significant Cyber Incidents (2019):

- **October 2018.** A malware was used to attack a petrochemical plant infrastructure in Saudi Arabia.
- **July 2018.** Security researchers say an Iranian hacking group targeted industrial control systems of US utility companies, Europe, East Asia and the Middle East.
- **July 2018.** Security researchers detect an increase in hacking attempts against Finland's IoT devices during President Trump's summit with Vladimir Putin in Helsinki.
- **June 2018.** The networks of a US contractor have been compromised and the attackers have stolen 614 GB of data on emerging weapons, sensors and communications systems for US submarines.
- **April 2018.** Researchers in the field of computer security demonstrate that hackers in North Korea targeted critical infrastructure, finance, healthcare and other industries across 17 countries using malicious code-like software used in the Sony Pictures attack in 2014.
- **April 2018.** Officials in the United States and the UK issued a joint warning about possible attacks that would deliberately target Western critical infrastructure by compromising Internet routers in households and businesses.
- **April 2018.** The UK's National Cyber Security Centre warns that Russian hackers target critical infrastructure in the UK by infiltrating supply chains.

- **March 2018.** The Baltimore Emergency Services Dispatch System was shut down for 17 hours after a robot attack, forcing the city to return to the classic dispatch centre.
- **March 2018.** The FBI and the Department of Homeland Security issued a joint technical alert to warn against cyber-attacks against US critical infrastructure. The objectives include energy, nuclear, water, aviation and production facilities.
- **February 2018.** A cyber-attack on the Pyeongchang Olympics puts the official Olympic site out of service for 12 hours and disrupts the Wi-Fi network and televisions at the Pyeongchang Olympic Stadium.

These selected incidents are related to IoT technologies. Since IoT technology has begun to be widely used in all areas, it has also been used in critical infrastructures. Critical infrastructure security must also be ensured for information managed through these technologies.

In this paper we aim to identify architectures for the protection of IoT technologies in critical infrastructures.

In (Ueno, 2017) security architectures and mechanism for protecting the security of wide area network are proposed.

Also in (Giles, 2019) the Triton malware is presented. It starts in 2014 in a Saudi plant. In June 2017 the first plant has been shutdown. Very quick after that another shutdown has been caused in August 2017. But these have been made public in December 2017 and now in January 2019 more details has been public. The attack has been conducted in a safety controller named Triconex made by Schneider Electric.

Requirements challenges for the 21th century regarding to the critical infrastructure protection such as: Trust and Collaboration, Usability and Performance, Autonomy and Self-healing, Scalability, Extensibility and Interoperability, Availability and Reliability, QoS, Resilience and Safety-critical are presented in (Alcaraz, 2015).

According to (Zmudzinski, 2019) the number of hacked IoT devices are doubled in 2018 compared with 2017. In 2017, 875.9 intrusions per sensor per day was been reported and in 2018 the number get to 1,702.8.

#### **4. Architecture for identifying devices in an IoT network**

Because within a critical infrastructure system there are several devices and sensors, in this chapter we aim to build an architecture based on which an IoT system can easily recognize the moments when other foreign devices connect to the system. Architecture will analyse the behaviour of devices in the system as and when a specific device is not recognized, it is considered a threat of the critical infrastructure where that IoT system is installed.

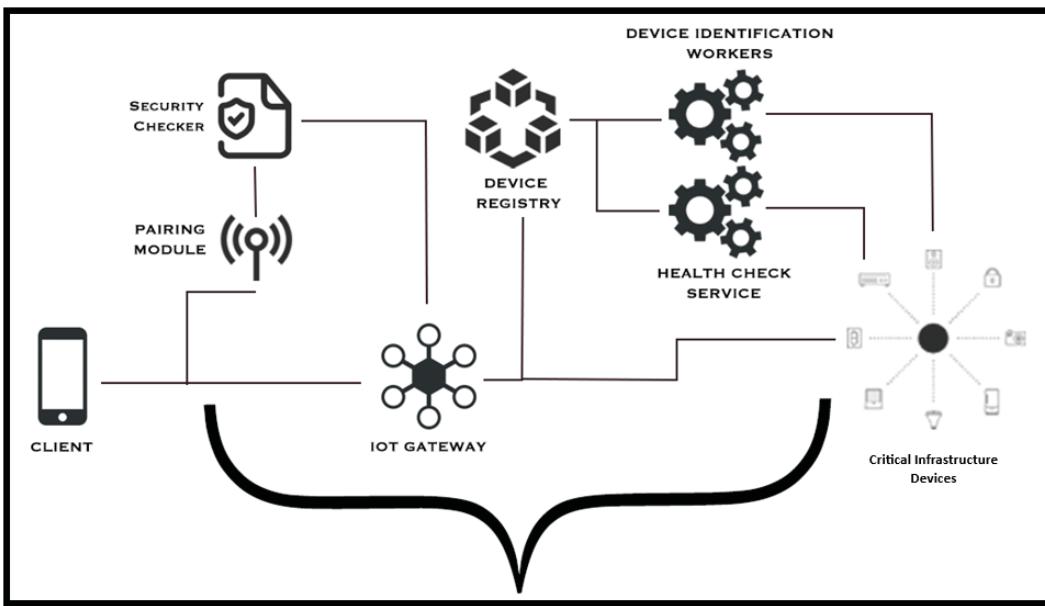
An IoT system consists of four components that interact with each other: people, intelligent objects, processes and technological ecosystem. For this, an appropriate method is needed to ensure reliability and objective achievement, using greater autonomy in threat detection, controlling and responding to attacks. The level of perception consists of all sensors used to collect data within the system. The resulting data is transmitted to the network layer that has the routing and data transmission function to various hubs and IoT devices (Mahmoud, 2015). The level of application, which is the current research level of this study, controls the integrity, authenticity and confidentiality of data.

Digital device capture is the process that, uniquely or with a high statistical probability, associates a physical device with a network entity based on its published attributes and network behaviour. Having a large set of properties, it is basically proven by previous research (Eckersley, 2014) that under certain conditions we can associate a combination of their values with the device, identifying it from a set of similar devices.

In (Ferrando, 2017), the term security of IoT systems is seen as a balance between the intrinsic value of tangible devices and sensors and intangibles, such as services, information and data, which lead to a practical security methodology.

The security of IoT systems is closely related to the detection of anomalies considered to be a security threat. Generally, detecting an abnormality in the use of an IoT system is a change in the overall image of the usage

characteristics. The first method of detecting anomalies is done using knowledge-based detection, supervised or unattended, during Machine Learning independent algorithms.



**Figure 2:** The proposed IoT architecture

The second type of anomaly detection method is done using a space representation. This leads to the description in an n-dimensional representation of each analysed heterogeneous device.

For calculating the distances between two data objects within the space vs. time representation, the Euclidean distance between cluster centroids is used. In this way, the representation of time does not depend on the number of each group, but on the distance used.

The current proposed architecture optimizes the process of detecting anomalies in using the IoT system in terms of space and time consumption. The methodology provides a safe level of use and preserves the benefits of the IoT systems, such as: interconnectivity, services related to things, heterogeneity, dynamic changes, enormous scale, safety and connectivity, described in (Jaafar, 2017).

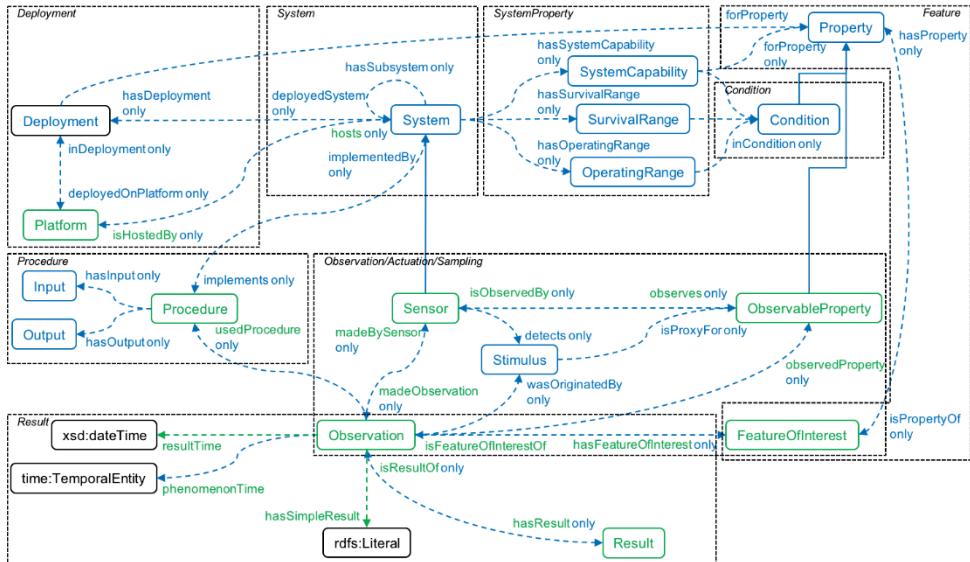
For software development of Smart Devices, a modular architecture is proposed that is scalable by coupling or disconnecting new modules. This approach allows the real-time modification of components without generating dead-time for the system, or the development of new modules, Figure 2.

The proposed architecture for the Smart Network Security device is a Gateway type in which access to all networked devices is accessed through a single access point.

## 5. Security architecture that is using semantic reasoning

One disadvantage of the security mechanisms based on machine learning or behavioural analysis is that they need to identify patterns before they can successfully detect possible threats. If one is able to infiltrate in the critical infrastructure right from its creation point, this type of algorithms will consider the behaviour of the network a normal one. Another weak part of the machine learning based approach is that it detects the security breaches right after they happen, that is why is important for a critical infrastructure to also have a preventive mechanism of detecting vulnerabilities.

Ontologies can play a key role in this area (Doina and Pocatilu, 2014). By adding metadata about the critical infrastructure and saving it in a semantic format, machines can understand the role of different components from the current architecture and how they relate to each other. An example of such an ontology is SSN (Semantic Sensor Network, 2017), the W3C's proposal for semantic modelling the IoT domain. An overview of the SSN ontology with all its classes and properties from an observation perspective is available in figure 3.



**Figure 3:** Overview of the SSN classes and properties (observation perspective). Source: (Semantic Sensor Network, 2017)

Starting from the semantic versions of the CVE database as proposed by (Boja et al., 2018) and (Guo and Wang, 2009), the metadata related to critical infrastructures, saved in an ontology built on top of the SSN, can be linked to the dictionary of common vulnerabilities. In this manner, by running system checks at some time intervals, the person in charge with the infrastructure's security can be warned of possible vulnerabilities that exist within the system. This approach can even go a step forward and provide the person with possible solutions to the identified vulnerabilities based on the data extracted from the CVE (updating the vulnerable version of the software for a specific component for example). Another advantage of a semantic approach is that, if a reasoning engine is used, the ontology can deduce some information about the system without the need to have redundant data, as shown in (Henze et al., 2004). If, for example, a device is using a weather sensor that in fact uses three other sensors (temperature, humidity and wind speed, let's say), then the system can automatically deduce that the device is using a temperature sensor. This can save a lot of time in modelling the system compared to classically approaches and is more appropriate to the humans' way of thinking.

Last, but not least, the semantic approach allows different critical infrastructures to exchange data in a common way. By exposing pieces of information similar to the PODs presented by (Mansour et al., 2016) to the Linked Open Data Cloud, each system can choose what type of information to exchange and with whom.

## 6. Conclusions

Productivity and efficiency is achieved if you make the right decisions, at the right moment and if you have the right data. Industries in various fields are pursuing and reaching this goal by connecting and measuring even the smallest components in their infrastructure, because each piece is as important as the entire puzzle. For example a city potable water delivery system is more cost efficient if it can predict when a water valve will malfunction and change it before the event will happen. This is a real case scenario and companies from different socio-economic fields are interconnecting their infrastructure components to the world wide IoT grid, using the Internet as the backbone. So, the task of securing Critical Infrastructures is expanding away from securing only dedicated servers or Data Centers to managing and protecting a complex and large infrastructure of IoT devices. This research has a proposed a proactive security approach that will detect abnormal behaviours based on known patterns, associated to devices that behave in normal parameters. It can also cope with large and complex IoT architectures for which passive or active security solutions can become too costly or too complex to manage by a single or multiple systems. Another advantage of the proposed architecture is that will continuously learn and adapt to new threats, to new devices, new protocols and new technologies as it is independent from any of the IoT devices implementations.

## Acknowledgements

This paper presents results obtained within the PN-III-P1-1.2-PCCDI-2017-0272 ATLAS project ("Hub inovativ pentru tehnologii avansate de securitate cibernetică / Innovative Hub for Advanced Cyber Security Technologies

"), financed by UEFISCDI through the PN III –"Dezvoltarea sistemului national de cercetare-dezvoltare", PN-III-P1-1.2-PCCDI-2017-1 program.

## References

- Ahmed A., Omar N. M., and Ibrahim H. M., Modern IoT Architectures Review: A Security Perspective, 8th Annual International Conference on ICT: Big Data, Cloud and Security (ICT-BDCS 2017), ISSN 2251-2136, doi: 10.5176/2251-2136\_ICT-BDCS17.30
- Alcaraz, C. and Zeadally, S. Critical Infrastructure Protection: Requirements and Challenges for the 21st Century, In International Journal of Critical Infrastructure Protection (IJCIP), vol. 8, Elsevier Science, pp. 5366, 01/2015
- Ashton, K. (2018) „That ‘internet of things’ thing,” RFID journal, vol. 22, nr. 7, pp. 97-114., 2009.
- Boja, C., Zamfirou, A., Iancu, B., Georgescu, T.M., Cartas, C., Toma, C. (2018). Avant-garde Technology Hub for Advanced Security [Technical Study]
- Choraś Michał, Kozik Rafał, Flizikowski Adam, Hotubowicz Witold and Renk Rafał, Chapter 7: Cyber Threats Impacting Critical Infrastructures, (2016), [https://link.springer.com/chapter/10.1007/978-3-319-51043-9\\_7](https://link.springer.com/chapter/10.1007/978-3-319-51043-9_7)
- Critical Infrastructure Protection: Survey of World-Wide Activities (2004),  
[https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Kritis/paper\\_studie\\_en\\_pdf.pdf.pdf%3Dpublication\\_File](https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Kritis/paper_studie_en_pdf.pdf.pdf%3Dpublication_File)
- Doinea, M. and Pocatilu, P. (2014), Security of Heterogeneous Content in Cloud Based Library Information Systems Using an Ontology Based Approach, Informatica Economică vol. 18, no. 4/2014
- Eckersley, P. (2014) „How Unique Is Your Web Browser?,” Electronic Frontier Foundation, 2014.
- Ferrando R. and Stacey, P. (2017) „Classification of Device Behaviour in Internet of Things Infrastructures: towards distinguishing the abnormal from security threats,” în International Conference on Internet of Things and Machine Learnin, Liverpool, 2017.
- Gartner, I., „Gartner says 6.4 billion connected" things" will be in use in 2016, up 30 percent from 2015.” 2015.
- Guo, M., & Wang, J. A. (2009, April). An ontology-based approach to model common vulnerabilities and exposures in information security. In ASEE Southeast Section Conference.
- Giles, M., (2019), Triton is the world’s most murderous malware, and it’s spreading, Online:  
<https://www.technologyreview.com/s/613054/cybersecurity-critical-infrastructure-triton-malware/>
- Henze, N., Dolog, P., & Nejdl, W. (2004). Reasoning and ontologies for personalized e-learning in the semantic web. Journal of Educational Technology & Society, 7(4).
- Jaafar, F. (2017) „An Integrated Architecture for IoT Fingerprinting,” în IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C), Prague, 2017.
- Kyriazis, Dimosthenis. "Protection of Service-Oriented Environments Serving Critical Infrastructures." Inventions 3.3 (2018): 62.
- Mahmoud, R., Yousuf, T., Aloul F. and Zualkernan, I. (2015) „Internet of things (IoT) security: Current status, challenges and prospective measures,” în 10th International Conference for Internet Technology and Secured Transactions (ICITST), London, 2015.
- Mansour, E., Sambra, A. V., Hawke, S., Zereba, M., Capadisli, S., Ghanem, A., ... & Berners-Lee, T. (2016, April). A demonstration of the solid platform for social web applications. In Proceedings of the 25th International Conference Companion on World Wide Web (pp. 223-226). International World Wide Web Conferences Steering Committee.
- Moosavi, S. R., Nigussie, E., Levorato, M., Virtanen S. and Isoaho, J.(2018) “Performnce Analysis of End-to-End Security Schemes in Healthcare IoT” Procedia Computer Science, vol. 130, p. 432–439, 2018.
- Nazerian, F., Motameni, H. and Nematzadeh, H. (2018), “Secure access control in multidomain environments and formal analysis of model specifications”, Turkish Journal of Electrical Engineering & Computer Sciences, (2018) 26: 2525 – 2540
- Palermo, F. (March 2018). [online]. Available: <https://www.informationweek.com/strategic-cio/executive-insights-and-innovation/internet-of-things-done-wrong-stifles-innovation/a/d-id/1279157>.
- Ray, P., (2018) A survey on Internet of Things architectures, Journal of King Saud University – Computer and Information Sciences, vol. 30, p. 291–319, 2018.
- Semantic Sensor Network Ontology. (2017, December 8). Retrieved January 25, 2019, from <https://www.w3.org/TR/vocab-ssn/>,
- Significant Cyber Incidents, [https://csis-prod.s3.amazonaws.com/s3fs-public/190103\\_Significant\\_Cyber\\_Events\\_List.pdf](https://csis-prod.s3.amazonaws.com/s3fs-public/190103_Significant_Cyber_Events_List.pdf)
- Toma, C., Popa, M. (2018), IoT Security Approaches in Oil & Gas Solution Industry 4.0, Informatica Economică vol. 22, no. 3/2018
- Ueno, Masami, Shingo Kashima, Yuminobu Igarashi, and Masahiro Hori, Mechanisms for protecting the security of a wide area network, [https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201705fa2.pdf&mode=show\\_pdf](https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201705fa2.pdf&mode=show_pdf)
- Winder Ransom, Jubinski, Joseph. (2016) „Internet of Things Examination”, The MITRE Corporation, 2016,  
<https://www.mitre.org/sites/default/files/publications/16-3415-iot-examination.pdf>.
- Zmudzinski A. (2019) Japan: Hacked IoT Devices and Cryptocurrency Networks Doubled in 2018, Online:  
<https://cointelegraph.com/news/japan-hacked-iot-devices-and-cryptocurrency-networks-doubled-in-2018>



# **PhD Research Papers**



# A Conceptual Model for the Development of Cybersecurity Capacity in Mozambique

Martina Jennifer Zucule de Barros<sup>1,2</sup> and Horst Lazarek<sup>1</sup>

<sup>1</sup>Technical University of Dresden, Germany

<sup>2</sup>Eduardo Mondlane University, Maputo, Mozambique

[martina\\_jennifer\\_zucule\\_de.barros1@mailbox.tu-dresden.de](mailto:martina_jennifer_zucule_de.barros1@mailbox.tu-dresden.de)

[horst.lazarek@tu-dresden.de](mailto:horst.lazarek@tu-dresden.de)

**Abstract:** Nowadays the economic development and social well-being of nations is relying more and more on Information and Communication Technologies (ICTs) and the use of cyberspace. This worldwide interconnectedness has brought and continues to bring economic, social, cultural benefits and cybersecurity challenges. Nevertheless, the human factor i.e., the demand for cybersecurity expertise has been recognized as an important element in cybersecurity. Internationally, several standards have been published to help countries developing their national frameworks to invest in cybersecurity capacity building. Developed countries have been making significant efforts to develop their national cybersecurity capacity and build an inclusive and innovative information society where all citizens became not only conscious, but also have the know how to handle cybersecurity issues. On the contrary to developed countries, developing countries (and specifically African countries), cybersecurity is still a challenging process. For instance, in Mozambique there is an evident lack of programs or initiatives related to cybersecurity in general. Therefore, this paper presents some internationally accepted frameworks related to cybersecurity capacity building as well as the current level of cybersecurity capacity building in developed and developing countries such as United States of America (USA), Germany, Mauritius and South Africa. Furthermore, it proposes a new conceptual model for cybersecurity capacity building for Mozambique.

**Keywords:** cybersecurity, capacity building, awareness, education, training, Mozambique

---

## 1. Introduction

The use of cyberspace and ICTs is turning our society into a cyberphysical society where almost all aspects of our daily live rely on cyberspace and ICT infrastructures. This creates huge potential benefits, but this dependence also makes us vulnerable. Worldwide, cybersecurity has become a national priority and several countries have recognized that to ensure secure and prosperous societies and secure their ICT infrastructures cybersecurity capacity building is crucial (De Bruijn & Janssen, 2017). In response, many developing countries have been investing in building cybersecurity competences to defend their national assets from illicit and illegal activities in cyberspace (Hohmann & Pirang, 2017). In contrast to developed countries, many developing countries still lag behind. According to Muller (2015) a developing country “faces challenges in all types of activities connected to cybersecurity capacity building, from human resources development, institutional reform, organizational and adaptation, and in the support provided to increase their access to, and ability to fully benefit from, the Internet and other elements of cyberspace” (Muller, 2015). Pawlak (2014) defines cybersecurity capacity building as “all types of activities (e.g. human resources development, institutional reform or organizational adaptations) that safeguard and promote the safe, secure and open use of cyberspace” (Pawlak, 2014). Moreover, Hohmann et al. (2017) define cybersecurity capacity building as “a set of initiatives that empowers individuals, communities and governments to reap potential gains from investments in digital technologies” (Hohmann et al., 2017).

In 2017, the International Telecommunication Union (ITU) published the *Global Cybersecurity Index* (GCI) report. The document shows the commitment of ITU member states regarding cybersecurity. In Africa, Mauritius ranks 1st. South Africa is also doing well, but Mozambique still lags behind (ITU, 2017). Unfortunately, still yet little has been done. Since the publication of Mozambique’s first national cybersecurity strategy draft the process remains stagnant. In answer to this challenge, this paper proposes a conceptual model for developing the country cybersecurity competences. Our model is sustained by the ITU *Cybersecurity Guide for Developing Countries* (ITU-D, 2009); the *Cybersecurity Capacity Maturity Model* (CMM) developed by the Global Cyber Security Capacity Centre (GCSCC) (GCSCC, 2016), both described in Section 2 and the *Developing Cybersecurity Capacity A proof-of-concept implementation guide* published by RAND Europe (Bellasio, et al., 2018). The rest of this paper is organized as follows: Section 3 presents an overview of cybersecurity capacity building approaches in developed and developing countries. The cybersecurity environment in Mozambique is explored in Section 4. It furthermore presents the our proposed model and discusses the approaches addressed in Section 3. Section 5 concludes the paper.

## **2. International frameworks**

### **2.1 ITU cybersecurity guide for developing countries**

In 2009, the ITU published the *Cybersecurity Guide for Developing Countries* (ITU-D, 2009). It aims to be a tool for developing countries, helping them to better perceive the legal, technical, economical, political, and managerial aspects of cybersecurity in the spirit of the Global Cybersecurity Agenda (GCA). The development of the guide considered the needs of developing and especially least developed countries regarding the use of ICTs for provision of basic services in different sectors, while committed to develop local and increase awareness among all the stakeholders. It is structured in five parts: cybersecurity context, stakes and challenges (1), cyber threats, cyber-attacks and cybercrime issues (2), legal, justice and police approaches (3), technical approach (4), and managerial approach (5) (ITU-D, 2009).

Part one focuses on trends of information society and advances brought by the use of information and communication infrastructures. It emphasizes the need to invest in cybersecurity capacity building and states that “capacity building contributes to the creation of an enabling environment with appropriate policy and legal frameworks, institutional development, including community participation, human resources development and strengthening of managerial systems” (ITU-D, 2009). On the other hand, part two describes a comprehensive approach comprising cybercrime, cyber threats and cyber-attacks. Part three proposes a legal, justice and police approaches associated to cybersecurity and cybercrime issues. Part four focuses on technical approach for cybersecurity presenting the most relevant principles of computer security and specifies the domains of application of cybersecurity. Part five addresses cybersecurity risk management. It helps to understand how to develop a cybersecurity strategy, define a cybersecurity policy and implement security measures. However, the five parts of this guide comprise solutions that several countries have put in place in order to handle cybersecurity (ITU-D, 2009).

### **2.2 GCSCC cybersecurity Capacity Maturity Model for nations (CMM)**

The University of Oxford GCSCC has been developing the CMM. The first version of the model was published in 2014. Through this model the GCSCC wants to increase the scale and effectiveness of cybersecurity capacity building not only in the United Kingdom (UK), but also internationally (GCSCC, 2016). In addition, the center aims to assist countries to improve their cybersecurity capacity in a systematic and substantive way i.e., “self-assess their level of cyber capacity and then directly receive information on how to make improvements” (Muller, 2015). Thus, by helping understanding national cybersecurity capacity, the GCSCC hopes to help promote an innovative cyberspace in support of well-being, human rights and prosperity for all (GCSCC, 2016). Similar to the ITU guide, CMM also comprises five parts named dimensions which are the following: (1) devising cybersecurity policy and strategy; (2) encouraging responsible cybersecurity culture within society; (3) developing cybersecurity knowledge; (4) creating effective legal and regulatory frameworks and (5) controlling risks through standards, organization and technologies. These dimensions “cover the broad expanse of areas that should be considered when seeking to enhance cybersecurity capacity” (GCSCC, 2016). It furthermore defines five stages of advancement such start-up, formative, established, strategic and dynamic used to measure the country progress in cybersecurity capacity. Despite this the GCSCC, intends to continuously enhance the CMM because they aim to make it an universally applicable model reflecting the global state of cybersecurity capacity maturity (GCSCC, 2016).

## **3. Cybersecurity capacity building initiatives around the world**

### **3.1 Developed countries**

#### **3.1.1 USA**

In 2012, the Organization for Economic Development and Cooperation (OECD) published a study where national cybersecurity strategies of OECD member's states were analyzed. This study shows that USA has been making significant investments in national cybersecurity (OECD, 2012). In 2016, the Potomac Institute for Policy Studies published the *United States of America Cyber Readiness at a Glance* (Hathaway et al., 2016). The document states that USA “enjoys advanced levels of cyber-maturity capacities” (Radunovic & Ruefenacht, 2016). The results of this study have been further emphasized by the ITU GCI report which states that USA ranks 2nd as one

of the top ten most committed countries regarding cybersecurity. Moreover, in the Americas region, USA has the highest score for cybersecurity capacity building (ITU, 2017). In USA however, cybersecurity has its roots since 1998, with the establishment of its first university outreach program, Center for Academic Excellence in Information Assurance (CAE-IAE) lead by the National Security Agency (NSA) (Radunovic & Ruefenacht, 2016). The Computer Security Resource Center (CSRC) defines information assurance as “measures that protect and defend information and information systems by ensuring their availability, integrity, authentication, confidentiality and non-repudiation. These measures include providing for restoration of information systems by incorporating protection, detection and reaction capabilities” (CSRC). The Department of Homeland Security (DHS) joined the initiative in 2004, and since then both have been sponsoring CAE-IAE. Moreover, USA has established models for state personnel training. For instance, the Department of Defense (DoD) Policy 8570.1 Information Assurance Training, Certification and Workforce Management (Radunovic & Ruefenacht, 2016). Also by 2016, the National Institute of Standards and Technology (NIST) launched the *National Initiative for Cybersecurity Education* (NICE).

NICE aims “to create an operational, sustainable, and continually improving program for cybersecurity awareness, education, training and workforce development that measurably advances the USA long-term cybersecurity posture” (Paulsen et al., 2012). Thus, to achieve this NICE has defined three strategic goals: (1) raise national awareness about risks in cyberspace; (2) broaden the pool of individuals prepared to enter the cybersecurity workforce; (3) cultivate a globally competitive cybersecurity workforce (NIST, 2016). These goals are organized under four components such as national cybersecurity awareness, formal cybersecurity education, cybersecurity workforce and cybersecurity workforce training and professional development where each of them is led by one or more federal agencies. For instance, national cybersecurity awareness is led by the DHS. Through the *Stop.Think.Connect* campaign, the DHS provides materials related to online safety and protection of personal information targeting several groups (e.g. children, youths, parents) (NIST, 2016). Despite this USA has been carrying out other initiatives. For instance, the government launched the Tech Hire initiative which aims to decrease unemployment through training, develop USA competitiveness (e.g. export cybersecurity products, services, standards, insurance, research or education) and develop cybersecurity capacities in the USA as well as support economic growth (Radunovic & Ruefenacht, 2016).

### **3.1.2 Germany**

According to Hathaway et al. (2016) “today Germany is one of the world’s most technologically advanced telecommunications systems as a result of intensive capital expenditures since its 1990 reunification” (Hathaway et al., 2016). As well as other member states of the OECD, Germany shows advanced levels of cybersecurity capacity building. Furthermore, Germany is considered one of the leading European Union (EU) member states regarding cybersecurity maturity (ITU, 2017). ITU’s Global Cybersecurity Index states that the German Federal Ministry of Education and Research (BMBF) and the Federal Ministry of the Interior (BMI) have signed an agreement on cooperation in information technology (IT) security research. The program covers research and development in new information security technologies (ITU, 2017). However, since 2011, the BMBF has been sponsoring research in cybersecurity as well as supporting a group of educational institutions in this realm (Radunovic & Ruefenacht, 2016). As a result, three centers were established in three different universities: the Center for IT Security, Privacy and Accountability (CISPA) in Saarbruecken, the European Center for Security and Privacy by Design (EC-SPRIDE) in Darmstadt and the Competence Center for Applied Security Technology (KASTERL) in Karlsruhe. Moreover, two big data centers to promote big data driven innovation and industrial applications, science and healthcare were established (Hathaway et al., 2016). On the other hand, the German government has established cybersecurity training programs. For instance, students at the BMBF KASTEL competence center are encouraged to gain certificate as specialist in IT security. Additionally, the Center for Advanced Security Research Darmstadt (CASED) at Technische Universitaet Darmstadt offers courses on security fundamentals to employed professionals and since 2010, the Technische Universitaet Darmstadt offers graduate programs in IT security (Hathaway et al., 2016).

The Fraunhofer Institute Network is another German institution engaged in cybersecurity. The institute aims to boost research and collaboration between academia and the private sector (Hathaway et al., 2016). The Federal Ministry for Economic Affairs and Energy (BMWi – Bundesministerium fuer Wirtschaft und Energie) has developed the IT security in the Economy initiative (IT Sicherheit in der Wirtschaft). The initiative aims to raise awareness of IT security. One of the educative activities is the IT-Sicherheit @ Mittelstand. It is a half-day seminar provided by the Association of German Chambers of Commerce and Industry (Hathaway et al., 2016). The

German Federal Office for Information Security (BSI) is one of the most important Germany's government sectors deeply involved in national cybersecurity issues (BSI, 2017). BSI conducts cybersecurity training initiatives as well as national cybersecurity awareness campaigns. Since 2002, BSI has been conducting national awareness campaigns. Currently, BSI has established a website ([www.bsi-fuer-buerger.de](http://www.bsi-fuer-buerger.de)). It covers topics such as online banking, smartphone security, e-mail coding or social networks including recommendations explained clearly for technical laypeople (BSI, 2017). In addition, the German government provides corporate research and development (R&D) incentives related to cybersecurity in three areas: ICT – detecting and solving cybersecurity incidents; computer technology – working in a digitalized world and e-mobility – value chain proposition (Hathaway et al., 2016).

### **3.1.3 Netherlands**

Netherlands "has become one of the most technologically advanced and highly connected countries in the world" (Hathaway & Spidalieri, 2017). Therefore, the development of cybersecurity capacity became a crucial factor to protect the country ICT infrastructure. As well as Germany, Netherlands is also considered one of the leading EU members states regarding national cybersecurity commitments (ITU, 2017). In 2013, the Hague Security Delta (HSD) was established. The HSD aims to harness Dutch innovation and drive the economic growth through a collaborative work between national entities on cybersecurity and ICT innovation. Currently, the HSD Campus in The Hague "is the national innovation center for security with state-of-the-art labs, education and training facilities" (Hathaway & Spidalieri, 2017). Several Dutch universities are also pursuing various initiatives. For instance, the University of Leiden, Delft University of Technology and The Hague University of Applied Sciences have established a Cyber Security Academy sponsored by the Municipality of The Hague. It is a multidisciplinary research center that offers part-time academic executive graduate programs, short courses and tailored tracks on a wide array of cybersecurity issues (Hathaway & Spidalieri, 2017). In 2016, a partnership between the Dutch government and private and public sector resulted on the establishment of the Dutch Cyber Security Platform for Higher Education and Research "Dcypher" initiative. It was established aimed at enlarge the pool of cybersecurity experts and enhance user's cybersecurity proficiency. Also by 2016, the Netherlands Organization for Applied Scientific Research (TNO) officially opened a cyber-threat intelligence laboratory in the HSD campus. The laboratory aims to test new technologies to improve the early cyber threats detection and information gathering and confidential data exchanges (Hathaway & Spidalieri, 2017).

According to Hathaway et al. (2017) "the Netherlands views R&D, innovation, and collaboration between the golden triangle of business, knowledge institutions, and government bodies as essential to its future" (Hathaway & Spidalieri, 2017). In relation to raising awareness, Netherlands is also doing well. The Dutch government and the National Cybersecurity Center (NCSC) regularly sponsor and participate in multiple cybersecurity awareness campaigns. For instance, the *Alert Online* campaign where several stakeholders collaborate to promote cybersecurity among Dutch citizens, government, public and private sector through workshops, meetings and presentations (Hathaway & Spidalieri, 2017).

## **3.2 Developing countries (African countries, Southern African Developed Community – SADC)**

### **3.2.1 Mauritius**

Mauritius is one of the African countries member states of the ITU. According to Dlamini et al. (2011) Mauritius "has most of the cybersecurity infrastructures in place" (Dlamini, Taute, & Rabede, 2011). ITU's Global Cybersecurity Index report classifies Mauritius as one of the worldwide leading countries regarding cybersecurity commitments. Currently, Mauritius ranks 6th from the top ten most committed countries (ITU, 2017). In relation to cybersecurity capacity building, Mauritius has been doing well. For instance, technical officers of the IT security unit from the Ministry of Technology, Communication and Innovations receive training on information security standards and best practices. On the other hand, the Computer Emergency Response Team (CERT)-MU offers cybersecurity training on digital forensic investigator professional and network forensic (packet analysis) for law enforcement officers (ITU, 2017). Furthermore, Mauritius has developed long relationships with international entities such as the CERT-India and the Japan-CERT. Moreover, Mauritius is a member of the Forum of Incident Response Security Teams (FIRST) (Symantec, 2016).

Despite the Mauritius government has not established cybersecurity education and training centers, several universities offer cybersecurity degree programs. In relation to raising awareness Mauritius is also doing well.

Since 2009, Mauritius has a national cybersecurity awareness campaign called *Sensitization Campaign*. In addition, Mauritius has established the *Cyber Security Mauritius* which aims to educate and enhance public awareness on the technological and social issues facing Internet users and especially online dangers. It targets groups such as kids, parents, home users and organizations. Other awareness initiatives are *MySecureCyberspace* (game), *privacy bird and privacy finder* and *MysecureCyberspace* (portal) (Dlamini, Taute, & Rabede, 2011).

### **3.2.2 South Africa**

South Africa is another African country member of the ITU. ITU's GCI report classifies South Africa belonging to the maturing stage i.e., countries that have “developed complex commitments, and engage in cybersecurity programmes and initiatives” (ITU, 2017). This means that South Africa has been making significant efforts to foster its national cybersecurity capacity. Currently, the development and implementation of R&D plans to promote capacity building is led by the Department of Science and Technology (DST). Various academic institutions such as the Nelson Mandela Metropolitan University (NMMU), Rhodes University, University of Pretoria (UP) and Cape Peninsula University of Technology (CPTU) offer undergraduate and graduate degrees with cybersecurity concentration. On the other hand, institutions such as the Justice College and the South African Judicial Training Institute offers training to prosecutors, judges and magistrates and have included specific training in its curriculum relating to cybercrime investigation, prosecution and evaluation of electronic evidence (Symantec, 2016). Similar to developed countries such as USA and Germany, South Africa has prioritized advanced training programs for state employees provided by external agencies (Symantec, 2016). Thus, the government has established the Cyber Security Institute of South Africa. The institute provides training and services to assist enterprises to effectively manage risks. In 2015, the Cybersecurity Hub was established. It is the central point for collaboration between industry, government and civil society on all cybersecurity aspects in South Africa (ITU, 2017). One of the Hub's actions is to implement national cybersecurity awareness program.

The Department of Telecommunications and Postal Services (DPTS) also conducts cybersecurity awareness initiatives which are being disseminated through mass media (e.g. magazines, billboards, etc.). The Department of Films and Publications has been conducting awareness initiatives related to child harm content over the Internet. Furthermore, the Department of Justice and Constitutional Development has developed certain mechanisms to make the public aware of electronic harassment and measures to help victims of cyber harassment (Symantec, 2016). Moreover, the South African Banking Risk Information Centre (SABRIC) funded by most South African banks also runs cybersecurity awareness initiatives. These initiatives focus mainly on public. It teaches best practices on the use of banks facilities and general security. In addition, it also gives tips on scams and guidelines. All the information is available online (Dlamini & Modise, 2012). Currently, the NMMU, the University of Johannesburg (UJ) and the University of South Africa (UNISA) have established the South African Cyber Security Academic Alliance (SACSAA) (SACSAA, 2011). SACSAA aims “to campaign for the effective delivery of cyber security awareness throughout South Africa to all groups of the population” (SACSAA, 2011). The NMMU and the UJ have also established cybersecurity centers. In 2017, a study conducted by Kritzinger et al. (2017) argues that “one notable area where South Africa leads the way is in its emphasis on academic research in cybersecurity awareness in schools and in the provision of learning and educational materials” (Kritzinger, Bada, & Nurse, 2017).

## **4. Cybersecurity context in Mozambique**

As well as Mauritius and South Africa, Mozambique is also one of the African countries members of the ITU. Cybersecurity in Mozambique is still in its infancy. This is emphasized by the ITU GCI report where Mozambique is classified as initiating stage i.e., countries that have started to make commitments in cybersecurity (ITU, 2017). In 2016, the National Communications Institute of Mozambique (INCM) published the country first national cybersecurity strategy draft (INCM, 2016). The document states that “the national cybersecurity strategy of Mozambique (2017 - 2021) articulates the nation’s coherent and coordinated approach for ensuring a safe, resilient cyberspace that can be fully leveraged by individuals and institutions of Mozambique with confidence” (INCM, 2016). Currently, the INCM is the government entity responsible for handling cybersecurity issues. In 2017, the Ministry of Science, Technology Higher Education and Technical Professional (MCTESTP), the government entity in charge of ICT's published the Electronic Transaction Act. Symantec Cyber Crime & Cyber Security Trends in Africa report states that “according to Mozambican authorities, the laws and regulations that govern areas related to cybersecurity can be found in their Electronic Transaction Act” (Symantec, 2016). However, internationally the development and establishment of national cybersecurity policy or strategy is

considered a best practice to coordinate the nation's cybersecurity activities. In addition, the Internet Infrastructure Security Guidelines for Africa states that "when developing national strategies for Internet infrastructure security, African policymakers should use for essential principles as a guide. These essential principles are awareness, responsibility, cooperation and fundamental rights and Internet properties" (AU, 2017).

To date, the national cybersecurity strategy draft remains stagnant. On the other hand, there have been no concrete developments regarding cybersecurity. For instance, there is no known cybersecurity operational plan neither cybersecurity capacity building initiatives. Furthermore, the strategy says little regarding cybersecurity capacity building. Moreover, the publication of those documents is not being accompanied by the development of an overarching strategy or policy to coordinate and prioritize measures to enhance Mozambique's cyber capacity. Therefore, an overview of the country current situation is given at Table 1. These analyses were based on the five dimensions of the CMM addressed in Section 2.2 and the following five pillars of the ITU GCI report: legal measures, technical measures, organization, capacity building and cooperation (ITU, 2017). From the CMM dimensions, one can notice that the Mozambique's overall stage is start-up while from the ITU GCI pillars the general level is red. Therefore, both elements show that currently, the country commitments in cybersecurity are at the lowest level. Symantec's Cyber Crime & Cyber Security Trends in Africa report states that "the government of Mozambique is currently working with civil society organizations and NGOs to help educate people and raise awareness of cyber related threats and the risk associated with being online" (Symantec, 2016). However, as shown on Table 1 this has not been gathered momentum. In addition, there currently are no cybersecurity models which solely target capacity building. Therefore, this paper aims to fill this gap.

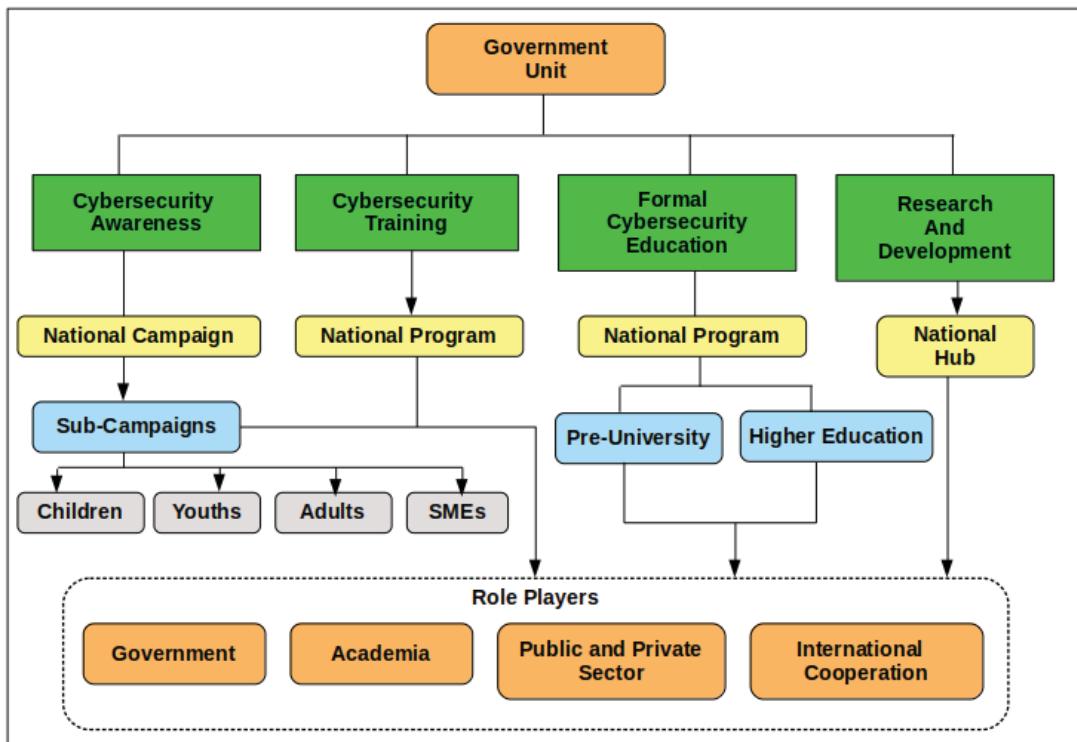
**Table 1:** Overview of cybersecurity commitments in Mozambique

CMM Dimensions	Mozambique's Stage	ITU GCI Pillars	Mozambique's Level
Cybersecurity Policy and Strategy	Start-up	Legal Measures	Yellow
Cyber Culture and Society	Start-up	Technical Measures	Red
Cybersecurity Education, Training and Skills	Start-up	Organizational	Red
Legal and Regulatory Framework	Start-up	Capacity Building	Red
Standards, Organizations and Technologies	Start-up	Cooperation	Yellow
Yellow: medium ; Red: low ; Green: high			

#### **4.1 A new conceptual cybersecurity capacity building model for Mozambique**

In this section, we present the conceptual cybersecurity capacity building model. It consists of four components as depicted in Figure 1. The model is based on internationally accepted frameworks (see Section 2) similar initiatives (see Section 3), identified challenges (see Section 4) and recommendations from the *Developing Cybersecurity Capacity A proof-of-concept implementation guide* (Bellasio, et al., 2018). Cybersecurity awareness is the first component, followed by cybersecurity training, formal education and R&D is the last component. The ITU *Cybersecurity Guide for Developing Countries* states that "capacity building contributes to the creation of an enabling environment with appropriate policy and legal frameworks, institutional development, including community participation, human resources development and strengthening of managerial systems" (ITU-D, 2009). Moreover, to "ensure secure and prosperous societies, countries require a skilled cybersecurity workforce" (Wamala, 2011).

A government unit should be responsible for coordinate the implementation of the model because "although the responsibility for cybersecurity capacity building is shared by many stakeholders, governments often lead coordinating efforts" (Hohmann et al., 2017). Therefore, in this model the government unit plays a crucial role. On the other hand, our model also proposes the role players i.e., stakeholders as one of the model's elements. They will be responsible for handling the implementation of the model's components. In the following subsections, a detailed description of each component is presented.



**Figure 1:** The conceptual cybersecurity capacity building model

#### 4.1.1 National cybersecurity awareness

Cybersecurity awareness "aims to inform and educate users and organizations on how best mitigate cyber threats" (EC, 2018). Therefore, considering the country current context our model proposes the development of national cybersecurity awareness campaign. This campaign aims to provide basic cybersecurity of the risks in cyberspace, foster the development of a national cybersecurity mindset and encourage all online users and organizations to change the behaviour that expose them to online risks (NIST, 2016). In addition, the model also proposes sub-campaigns targeting all audiences (e.g. children, youths, adults and small and medium enterprises (SMEs)) to ensure effective cyber awareness across all layers of the society. Both standards described in Section 2 consider cybersecurity awareness as an important element in national cybersecurity. Furthermore, developed countries as well as developing countries addressed in Section 3, have developed similar initiatives. In developed countries, at least one of the government entities is responsible for conducting national cybersecurity awareness campaigns. For instance, in USA the DHS runs the Stop.Think.Connect campaign. We also suggest the establishment of bilateral and multilateral cooperation. Worldwide, several organizations have been carrying out cybersecurity awareness campaigns. For instance, the *Safer Internet Day* campaign sponsored by the European Commission (EC) or the *European Cyber Security Month* carried out by the European Network and Information Security Agency (ENISA). The Internet Safety Campaign targeting African citizens, Safe and Secure Online, the National Cyber Security Awareness Month, the Safety Internet Day, etc. (Bada & Sasse, 2014; Government, 2011; Bellasio et al., 2018). In addition, cybersecurity awareness and standards guidelines such as ENISA guide and the NIST special publication 800-50 states that any cybersecurity awareness program should follow a life-cycle approach comprising: design, execution and management, evaluation and adjustment (ENISA, 2010) (NIST, 2003). Therefore, this component also considers this element.

#### 4.1.2 National cybersecurity education

The aim of cybersecurity education "is to provide students with sufficient knowledge and skills to pursue a successful career in cybersecurity or an IT-intensive domain and to encourage students to select this career path" (Bellasio, et al., 2018). Therefore, our model also proposes the development of a national program. The goal of this component is to foster cybersecurity through all stages of education in Mozambique. Thus, the national program comprises two approaches: pre-university cybersecurity education and cybersecurity in higher education as illustrated in Figure 1. The first one focuses on primary and secondary levels whereas the second focuses on undergraduate and graduate degrees. To achieve this, at least one government entity should be

responsible for overseeing education initiatives and collaborate with other stakeholders. For example, in USA the National Science Foundation (NSF) and the Department of Education (ED) co-lead cybersecurity education. In Germany, cybersecurity education and R&D is led by the BMBF. In South Africa, cybersecurity education is under the leadership of the DST. Therefore, we suggest that at least one government entity should be responsible for implementing cybersecurity education. On the other hand, we also suggest the establishment of international cooperation. The USA *International Strategy for Cyberspace* states that international cooperation is very essential “to provide the necessary knowledge, training and other resources to countries seeking to build technical and cybersecurity capacity” (House, 2011). Evaluation is an essential element in any cybersecurity education program. Through evaluation it is possible verify the effectiveness of the program, identify areas for improvement and learn from the failures. Thus, this model also considers this factor.

#### **4.1.3 National cybersecurity training**

The goal of this component is to provide the needed knowledge, skills and competences to individuals who work in a professional environment. Our model proposes the development of a national cybersecurity training program as illustrated in Figure 1. We proposed this because cybersecurity training is important to prepare professionals to face the evolving and dynamic context of technology (ITU-D, 2009). Here, it is also important assign a task owner to implement cybersecurity training programs. In all countries presented in Section 3, cybersecurity training is led by at least one government entity. In USA, the DHS is also responsible for cybersecurity training. This is similar in Germany where the BSI also conducts cybersecurity training initiatives. In South Africa, external entities provide cybersecurity training. International cooperation also plays an essential role in cybersecurity training. For example, the International Council of E-Commerce Consultants (EC-Council) offers courses and certification in areas such as Computer Hacking Forensics, Certified Security Analyst etc. Cybersecurity training initiatives should not be static in its development and implementation (Bellasio, et al., 2018). Therefore, this model suggests that regularly cybersecurity training programs should be evaluated.

#### **4.1.4 Research and development**

The lack of initiatives or programs related to cybersecurity also extends to R&D. *The Internet Infrastructure Security Guidelines for Africa* states that “cybersecurity research and development in Africa should be improved to create more usable and current tools for African stakeholders. Because of the international nature of the Internet, information needs to be made available to all stakeholders, not only those within a national border” (AU, 2017). Therefore, our model proposes the establishment of a national cybersecurity research hub to serve as a central point for developing national research and expertise and provide more opportunities for Mozambican citizens. We also suggest that this hub should serve as an international point of contact for collaboration. Through R&D Mozambican citizens will be able to develop new technologies, bring solutions to solve cybersecurity problems and enhance their expertise. This could be through a public-private-partnership (PPP) or international cooperation. For example, in Netherlands the government, business community and academia have launched the Dcypher. The HSD is another example of PPP and according to Hathaway et al. (2015) it is considered “the largest security network in Europe with knowledge bridges to main security networks in the United States, Canada, Singapore, and South Africa ” (Hathaway et al, 2015). On the other hand, in Germany, the BMBF and the Ministry of the Interior have established the IT Security Research program to research and discover new IT security applications (Hathaway et al., 2016). In South Africa, the SACSAA represents an alliance of research groups comprising three universities, NNMU, UJ and UNISA. Nevertheless, ENISA’s report on “Public Private Partnerships” (ENISA, 2017) addresses four types of PPP models such as institutional, goal-oriented, service outsourcing and hybrid PPP (ENISA, 2017). Thus, we suggest the adoption of one these models for develop PPP.

## **4.2 Discussions**

The aim of this study was to develop a cybersecurity capacity building model for Mozambique. The study specifically focused on the analyses of cybersecurity capacity building approaches adopted in countries such as USA, Germany, Netherlands, Mauritius and South Africa as well as recommendations from the ITU framework (ITU, 2017) and the GCSCC CMM framework (GCSCC, 2016). Both frameworks state that cybersecurity capacity building contributes to enhance countries’ cyber resilience and build functional and accountable institutions, appropriate policies and legal frameworks to respond effectively to cybercrime. The analyses of cybersecurity capacity building efforts and commitments indicate that USA, Germany and Netherlands are considered leading sates in cybersecurity and ICT development (ITU, 2017). In USA, Germany and Netherlands ICT is viewed as a key

driver to expand and growth their economies. Furthermore, USA stands at the world-wide leading hi-tech based economy (Harasta, 2013). Germany, USA and Netherlands have established official entities in charge of cyber education, R&D of cybersecurity standards, adopted cybersecurity standards and guidelines and investing in national cybersecurity research and development initiatives. Specifically, USA has founded several multi-agency R&D programs and several federal agencies are leading R&D efforts. Mauritius is the only African state in the top then list of the most committed countries regarding cybersecurity (ITU, 2017). South Africa is classified as maturing stage, whereas Mozambique is initiating stage. This classification is also applied to ICT development. These two countries, expect Mozambique have made significant efforts in ICT and cybersecurity field. Mozambique still has very low levels. This is because, first, there is a lack of political will and lack of government leadership. The second issue is the lack of private and public cooperation in cybersecurity and ICT development. Therefore, these findings reveal that developed countries are better prepared to defend their critical systems and infrastructures on the contrary to developing countries. According to Zareen et al. (2013) developing countries are dependent on developed countries in procurement of ICT products, resources and expertise of development of these products (Zareen et al., 2013). Moreover, Phahlamohlaka et al. (2011) state developing countries have been focusing on becoming more and more connected to the cyberspace neglecting the cybersecurity risks that accompany the connectivity.

## **5. Conclusion**

During the last 15 years, the ICT's have become crucial for the sustainable development of worldwide countries providing endless opportunities to our societies. Hohmann et al. (2017) state that "no country will be able to reap the full potential of ICTs without also building cybersecurity capacity to address the risks associated with connectivity, such as losing trust in digital infrastructures, cybercrime or even threats to national security" (Hohmann et al, 2017). Worldwide, governments as well as international organizations recognize that invest in cybersecurity capacity building is essential. Thus, several governments specifically from developed countries have been making significant efforts to build, advance and sustain their national cybersecurity capacity building initiatives. ITU's GCI report classifies USA, Germany and Netherlands into being in leading stage in the cybersecurity and in the field of ICT development (ITU, 2017). On the other hand, the majority of developing countries still show low levels in cybersecurity and also ICT development and Mozambique is among them. Therefore, in this paper we presented the proposal of a conceptual model for the development of cybersecurity capacity, envisaged in its national cybersecurity strategy draft. The development of this model followed two approaches first, recommendations from international frameworks and secondly, key factors identified from analysis of cybersecurity capacity building approaches made in Section 3. This model has to be implemented. However, as this would go beyond the scope of this paper, we plan to present its implementation in a future work.

## **References**

- AU. (2017). Internet Infrastructure Security Guidelines for Africa. Internet Society African Union (AU).
- Bada, M., & Sasse, A. (2014). Cyber Security Awareness Campaigns Why do they fail to change behaviour. SBS.
- Bellasio, J., Flint, R., Ryan, N., Sondergaard, S., Monsalve, C. G., Meranto, A. S., et al. (2018). Developing Cybersecurity Capacity A proof-of-concept implementation guide. RAND Corporation.
- BSI. (2017, January). Security in Focus Cyber Security Strategy for Germany 2016. Federal Office for Information Security.
- CSRC. (n.d.). Computer Security Resource Center Glossary . Retrieved August 10, 2018, from <https://csrc.nist.gov/>
- De Bruijn, H., & Janssen, M. (2017). Building cybersecurity awareness: The need for evidence-based framing strategies. ScienceDirect, 1-7.
- Dlamini, I., Tauté, B., & Rabede, J. (2011). Framework for an African Policy Towards Creating Cyber Security Awareness. Southern African Cyber Security Awareness Workshop, (pp. 15-31).
- Dlamini, M., & Modise, M. (2012). Cyber Security Awareness initiatives in South Africa: a synergy approach. 7th International Conference on Information Warfare and Security, (pp. 1-10). Seattle.
- EC. (n.d.). Retrieved March 26, 2019, from <https://www.saferinternetday.org/>
- EC. (2018, August 31). Retrieved January 17, 2019, from European Union Institute for Security Studies: <https://www.iss.europa.eu>
- ENISA. (2010, November). The new user's guide: How to raise information security awareness. Retrieved March 25, 2019, from [https://www.enisa.europa.eu/publications/archive/copy\\_of\\_new-users-guide](https://www.enisa.europa.eu/publications/archive/copy_of_new-users-guide)
- ENISA. (2017, November). Public Private Partnerships (PPP) Cooperative models. Retrieved March 26, 2019, from <https://www.enisa.europa.eu/publications/public-private-partnerships-ppp-cooperative-models>
- EU. (2018). Operational Guidance for the EU's international cooperation on cyber capacity building. Luxembourg: European Commission.
- GSCC. (2016). Cybersecurity Capacity Maturity Modelfor Nations (CMM). University of Oxford.

- Government, A. (2011). An overview of international cyber-security awareness raising and educational initiatives. Australian Communications and Media Authority (ACMA).
- Harasta, J. (2013). Cyber Security in Young Democracies. Retrieved April 18, 2019, from [http://www.mruni.eu/lt/mokslo\\_darbai/jurisprudencija/](http://www.mruni.eu/lt/mokslo_darbai/jurisprudencija/)
- Hathaway, M., & Spidalieri, F. (2017, May). The Netherlands Cyber Readiness at a Glance. Potomac Institute for Policy Studies.
- Hathaway, M., Demchack, C., Kerben, J., McArdle, J., & Spidalieri, F. (2016, October). Germany Cyber Readiness at a Glance. Potomac Institute for Policy Studies.
- Hathaway, M., Demchack, C., Kerben, J., McArdle, J., & Spidalieri, F. (2016). United States of America Cyber Readiness at a Glance. Arlington: Potomac Institute for Policy Studies.
- Hathaway, M., Demchak, C., Kerben, J., McArdle, J., & Spidalieri, F. (2015, November). Cyber Readiness Index 2.0 A Plan for Cyber Readiness: A Baseline and an Index. Retrieved March 26, 2019, from <http://www.potomacinstitute.org/academic-centers/cyber-readiness-index?id=29>
- Hathaway, M., Demchak, C., Kerber, J., McArdle, J., & Spidalieri, F. (2016, November). Italy Cyber Readiness at a Glance. Retrieved March 27, 2019, from <http://www.potomacinstitute.org/academic-centers/cyber-readiness-index?id=29>
- Hohmann, M., Pirang, A., & Benner, T. (2017, March). Global Public Policy Institute. Retrieved August 1, 2018, from <http://www.gppi.net/publications/data-technology-politics/article/advancing-cybersecurity-capacity-building-implementing-a-principle-based-approach/>
- House, T. W. (2011, May). Retrieved January 16, 2019, from <https://obamawhitehouse.archives.gov>
- INCM. (2016). Estrategia Nacional de Segurança Cibernetica. Retrieved July 20, 2018, from <http://www.ciberseguranca.org.mz/documentospublicacoes/documentos.html>
- ITU. (2017). Global Cybersecurity Index (GCI). Retrieved May 10, 2018, from <https://www.itu.int/en/ITU-D/Cybersecurity/Pages/GCI-2017.aspx>
- ITU-D. (2009). Cybersecurity Guide for Developing Countries. Retrieved April 20, 2018, from <https://www.itu.int/ITU-D/cyb/publications/2009/cgdc-2009-e.pdf>
- Kritzinger, E., Bada, M., & Nurse, J. R. (2017). A study into the cybersecurity awareness initiatives for school learners in South Africa and the UK. 10th World Conference on Information Security Education (pp. 1-10). IFIP WG.
- Muller, L. P. (2015). Norwegian Institute of International Affairs . Retrieved August 2, 2018, from <https://www.nupi.no/en/Publications/CRISTin-Pub/Cyber-Security-Capacity-Building-in-Developing-Countries-challenges-and-Opportunities>
- NIST. (2003, October). Building an Information Technology Security Awareness and Training Program. Retrieved March 25, 2019, from <https://www.hsdl.org/?abstract&did=460786>
- NIST. (2016, April). National Initiative for Cybersecurity Education (NICE). Retrieved June 20, 2018, from National Institute of Science and Technology: <https://www.nist.gov/itl/applied-cybersecurity/nice/about/strategic-plan>
- OECD. (2012). Cybersecurity Policy Making at a Turning Point Analysing a new generation of national cybersecurity strategies for the Internet economy. Organisation for Economic Cooperation and Development.
- Paulsen, C., McDuffie, E., Newhouse, W., & Toth, P. (2012). NICE: Creating a Cybersecurity Workforce and Aware Public. IEEE Computer and Reliability Societies, 76-79.
- Pawlak, P. (2014). Riding the digital wave The impact of cyber capacity building on human development. Paris: European Union Institute for Security Studies.
- Peter, A. S. (2017). Cyber resilience preparedness of Africa's top-12 emerging economies. ScienceDirect, 49-59.
- Phahlamohlaka, L., van Vuuren, J., & Coetzee, A. (2011). Cyber Security Awareness Toolkit for National Security: an Approach to South Africa's Cyber Security Policy Implementation. Proceedings of Southern African Cyber Security Awareness Workshop (SACSAW), (pp. 1-14).
- Radunovic, V., & Ruefenacht, D. (2016). Cybersecurity Competence Building Trends. Genvena: DiploFoundation.
- SACSAA. (2011, June). South African Cyber Security Academic Alliance. Retrieved July 10, 2018, from <http://www.cyberaware.org.za/>
- Symantec. (2016). Cyber Crime & Cyber Security Trends in Africa. Retrieved July 2, 2018, from <https://www.symantec.com/theme/cyber-security-trends-africa>
- Wamala, F. (2011, September). ITU National Cybersecurity Strategy Guide. Retrieved January 17, 2019, from ITU: <https://www.itu.int>
- Zareen, M. S., Akhlaq, M., Tariq, M., & Khalid, U. (2013). Cyber Security Challenges and Wayforward for Developing Countries. IEEE, 7-14.

# The Peculiarities of Securitising Cyberspace: A Multi-Actor Analysis of the Construction of Cyber Threats in the US (2003-2016)

Noran Shafik Fouad

University of Sussex, Brighton, UK

[n.fouad@sussex.ac.uk](mailto:n.fouad@sussex.ac.uk)

**Abstract:** The rapid development of information and communication technologies rendered cybersecurity an integral aspect of contemporary security discourses and practices in different fields. Yet, despite the obvious intellectual demands of the field, most academic literature on cybersecurity in international relations and security studies remain policy-oriented and under-theorised. One of the few exceptions are studies utilising the Copenhagen school's securitisation theory to studying discourses and practices of cybersecurity, particularly in the US. Nevertheless, the cyber securitisation literature is still limited in its engagement with the complexity of cybersecurity. One important aspect of this limitation is their focus on official and government's discourses; an approach that is not applicable with a multi-stakeholder, privately-dominated cyberspace. This state-centric approach does not reflect the diversity of cybersecurity discourses by highlighting only the militarised, geopolitical narratives, adopted by some policy makers. Besides, it overlooks the nuances in threat perceptions, not just between the private and the public sectors, but also among different agencies inside the government. Therefore, focusing on the US as a case study, this paper will employ the securitisation theory's sectoral analysis for studying the process of securitisation in the field of cybersecurity, using a multi-actor approach which considers the role of several state and non-state actors in producing and managing cybersecurity discourses, and how complex public-private relationships influence cybersecurity policies and practices. The paper uses the method of discourse analysis in studying cybersecurity discourses of the government, private sector, and media in the US, by examining multiple resources, including official policy documents, congressional hearings, and opinion articles. The analysis covers the period from 2003, when the first cybersecurity strategy was announced, until the end of the Obama administration in 2016. The arguments presented by this paper contribute to the theorisation of the complex conceptual and policy problems of cybersecurity and of cyber securitisation processes through an inductive approach that develops an understanding of the logics and politics of security and risk as contextually-bound and sector-dependent.

**Keywords:** cybersecurity, cyberspace, securitisation, US cyber policy, discourse analysis

---

## 1. Introduction

Since the Morris Worm hit the earliest version of the internet (the ARPANET) in 1988, hostile cyber operations have been growing in number and sophistication; ranging from cyber crimes by non-state actors to state-backed cyber operations. At the same time, the range of 'insecure' objects has been widened to include, not only governments, but also individuals, businesses, and most recently, electoral processes. Despite the obvious intellectual demands of the field, most academic literature on cybersecurity remain policy-oriented and under-theorised. One of the few exceptions, however, are studies utilising the Copenhagen school's securitisation theory to studying discourses and practices of cybersecurity, particularly in the US (Eriksson, 2001; Bendrath, Eriksson and Giacomello, 2007; Dunn Cavelty, 2008a, 2008b; Hansen and Nissenbaum, 2009). The significance of these stems from the securitisation theory's explanatory power for understanding how and why a new realm like cyberspace is being constructed as a security sector, and how cyber practices are legitimised when their 'securityness' is accepted by the relevant audiences.

Nevertheless, the cyber securitisation literature remains very limited in its engagement with the complexity of the cyber realm. This can be seen, for instance, in their focus on official and government's discourses; an approach that is not applicable with the multi-stakeholder nature of cyberspace and one that misses the extent to which non-governmental actors produce and manage relevant threat discourses. This state-centric approach is problematic since it does not reflect the diversity of cybersecurity discourses, and because it only highlights the militarised, geopolitical discourses, adopted by some policy makers, that reinforces cyber territoriality over spatiality in a friend-enemy logic. Furthermore, they deal with states as unitary actors in cyberspace; overlooking the nuances in threat perceptions of intelligence communities and military institutions and cyber commands, executive, and legislative branches. Most importantly, by applying the theory's definition of security that is tied to existential threats and exceptional measures, those studies implicitly assume that meanings and practices of security are fixated along sectors. That is, they do not analyse how just as cyberspace has broadened the security agenda, it may have also transformed those meaning and practices beyond the Copenhagen School's conceptual framework.

Therefore, using the US as a case study, this paper will employ the securitisation theory's sectoral analysis for studying the discursive peculiarities of cybersecurity, using a multi-actor approach which considers the role of state and non-state actors in producing and managing cybersecurity discourses, and how complex public-private relationships influence cybersecurity policies and practices. The analysis covers the period from 2003, when the first cybersecurity strategy was announced in the US, until the end of the Obama administration in 2016. The paper primarily aims at answering the following questions: what issues are constructed as threatening in cybersecurity discourses and practices? To what referent objects? By which actor(s)? And targeting which audience(s)? Whom do such discourses and practices empower and/or exclude? In answering these questions, the paper uses the method of critical discourse analysis (CDA) (Fairclough, 2001, 2003), in which discourses are dealt with as both constitutive and constituted. This method is based on a three-dimensional analysis of discursive events: texts and their linguistic analysis; discursive practices, or the interpretation of processes of text production; and the analysis of social and institutional factors that shape discourses, i.e. discourses as social practice. Tracing the evolution of cybersecurity discourses over the study period, the paper employs CDA's concepts of *interdiscursivity* and *intertextual analysis*, which identifies and evaluates the links between texts, discursive genres, and their relations to external environments. Multiple resources are used in this analysis: 1-) seven cybersecurity-related documents issued by the White House, including cybersecurity strategies, presidential directives, and executive orders; 2-) six documents on cybersecurity by the Department of Defence (DoD), including the cyber defence strategy; 3-) four cybersecurity documents by the Department of Homeland Security (DHS); 5-) fifty four cybersecurity-related congressional hearings in the Committee on Homeland security in the House and the Committee on Homeland Security and Governmental Affairs in the Senate. Such hearings should be indicative of the discourses of MPs, executive officials, security experts, and members of the private sector, who are invited to testify in those hearings; and 6-) one hundred ninety-two editorials and opinion articles that has 'cybersecurity' in the major mentions, in the 'national security and international relations section' in US newspapers and wires, retrieved from the Nexis database. Editorials and opinion articles are presumably more explicit on policies and threat perceptions on the subject matter than news reporting. Given the extent of the textual sources involved, qualitative analysis software (NVivo 12) was used in coding the data.

The paper is divided into three sections. It starts first with an exploration of the logics of threats and vulnerabilities in cybersecurity discourses, and how the need for more cybersecurity is legitimised. It particularly focuses on the securitisation of cyber dependency and increased cyber capabilities, threat attribution and the attack logic, as well as the existentiality vs. urgency paradox. The second section examines the constellation of referent objects in cybersecurity discourses, or the objects constructed as being threatened and in need of protection, as well as the cross-sectoral connections they manifest with military, economic, and political security. The third investigates the cybersecurity ecosystem and the complexities of public-private relationships that shape questions of responsibility, liability, and the controversy over state's role.

## **2. The logics of threats and vulnerabilities in cybersecurity**

*"As we know, the genie is out of the bottle, just like nuclear weapons. It can be turned against us. We know what our offensive capability is and it is pretty darn impressive. That capability turned against us, I think is what frightens us, and who would have the motivation to do that." - Representative Michael T. McCaul, congressional hearing (America is Under Cyber Attack: Why Urgent Action Is Needed, 2012, p.45)*

### **2.1 Securitising cyber dependency: Risk society as a security discourse**

In his risk society thesis, Ulrich Beck assumes that we are now living in a 'second modernity', whereby risks can be conceptualised as "a systematic way of dealing with hazards and insecurities induced and introduced by modernisation itself" (Beck, 1992, p. 21). This period of 'reflexive modernity' is marked by the dominant force of the unknown, incalculable, and uncontrollable dangers, which are 'de-bounding' spatially, temporally, and socially (Beck, 2002, p. 41). In this advanced modernity, risks are not the result of an undersupply of technology, but rather from its *overproduction*, which eventually affects everyone, even those who produce and profit from risks (Beck, 1992, pp. 19–23).

Beck's risk society thesis can be seen as a security discourse per se, rather than just an approach to study the objective reality of cyber risks. It can be argued that risk society is a dominant discourse adopted by the majority of actors in constructing cyber threats. In this discourse, cyber dependency is securitised and portrayed as exponential, inevitable, and inherently threatening. Securitising cyber dependency is based on a belief in a

revolutionary *present*: one in which ICTs are revolutionising modern life in ‘unprecedented’ ways, creating a ‘new reality’. This revolutionary present is constructed as a threat as such using two logics. The first legitimises the call for more cybersecurity by highlighting how crucial cyber technologies are for “prosperous economies, vigorous research communities, strong militaries, transparent governments, and free societies” (The White House, 2011, p. 3). Accordingly, an insecure cyberspace means “the gains from computer integration can be wiped out or reversed” (Goldsmith and Hathaway, 2010). The second logic is centred on the increasing vulnerabilities produced by this dependency. It assumes that cyber technologies are growing more complex, and with complexity comes more insecurities and greater risks, because “Complexity is something we can’t change” (Overview of the Cyber Problem, 2003, p. 11).

Thus, the need for security is legitimised by the importance of overcoming dependency-induced threats or to benefit from the fruits of the ICTs and the ‘digital revolution’. Both logics produce multiple assertions in constructing the cyber threat: 1-) increasing cyber dependency and complexity are unavoidable, making the future more threatening than the present; 2-) cyber technology is inherently vulnerable, and thus impossible to fully secure; 3-) calls for ‘more security’ becomes self-justifying, given the understanding that the more cyber dependent the state is, the more inevitably threatened it becomes.

## **2.2 Geography and threat sources: from technical to political attribution**

Since the first cybersecurity strategy was announced in 2003, attribution was not clearly used in cybersecurity discourses, and threats from states and non-state actors were presented on equal footing. Terms like ‘our adversaries’, ‘attackers’, ‘malicious actors’, and ‘America’s enemies’ were used without a clear identification of particular actor(s). However, this situation has been changing gradually ever since to one in which attribution sometimes form the core of the cyber threat perception, with a strong emphasis on nation-states as a threat source, namely Russia, China, Iran, and North Korea. Much of this emphasis on threat attribution is driven from some territorial understanding of cyberspace and the sense of ownership that is found in many political discourses. Phrases like ‘America’s cyberspace’, ‘cyber borders’, and an emphasis on the threat of the ‘foreign’ and the ‘external’ are important examples.

Here, we can differentiate between two types of attribution: *attack* attribution and *threat* attribution. The first is concerned with attacks that have already taken place, while the second is related to the ones that have not, and thus seeks to establish links between future threats/hazards and a particular source. Discourses that attribute the cyber threat to a certain source transfer conventional threat perception to cybersecurity. If the cyber threat is mainly associated with the aforementioned countries that are generally perceived as antagonistic to the US, and if their cyber capabilities are portrayed as exponentially increasing, then the US is automatically threatened. Consequently, the construction of futuristic threat scenarios becomes easier with little needed justification, because threat attribution with traditional enemies is invoked as a *facilitating condition* for securitisation. Although the published reports on attack attribution by the private sector exceed those of the government (Rid and Buchanan, 2015, p. 28), it is the government and some think tanks that focus more on this *threat* attribution. But other than categorisation purposes, what does attribution contribute to the cyber threat construction and why should it be problematised?

Firstly, this aspect of the cyber threat construction takes for granted the need for attribution in cyber defence. In conventional defence strategies, knowing the attack source and their capabilities before defensive or offensive responses is a must. However, cybersecurity is characterised by a high level of asymmetries that render this attribution-specific defence strategies obsolete (Rivera and Hare, 2014, p. 104). Secondly, the cyber threat/hazard attribution overlooks the uncertainties intrinsic to attack-attribution in cybersecurity. For instance, packets used in attacks can be changed before reaching the target and their original addresses can be erased by bots. Thus, ‘to whose benefit’ is not a credible strategy in attribution, because attacks can be implanted by a third-party. And even when they are traced to a certain country, it can be a separate political organisation or individual working for their own interests (Libicki, 2009). Thirdly, defining cyber capabilities is often more a matter of speculation than knowledge. Unlike military arms, cyber offensive tools are not observable, cannot be quantified, and in most cases, they cannot be recognised before an attack actually takes place, because in cyberspace, “offensive capacity correlates with defensive vulnerability” (Schutte, 2012, p. 8). Fourthly, in recent years, the line between offensive and defensive cyber operations is being blurred. Government officials acknowledge that many ‘friendly’ nations maintain an existence on the U.S. networks for information collection. Then who draws the line between the offensive and the defensive? This reinforces the

idea about the cyber threat attribution as a political rather than a technical act, particularly in what the state decides to publish.

### **2.3 The nature of the cyber threat: between existentiality and urgency**

Unlike other security sectors, cybersecurity threats are always perceived in the form of *attacks*, or hostile, purposeful, and deliberate actions by an enemy/adversary against the referent object(s). While this attack logic can still be used occasionally in all sectors, it is the *dominant* one in cybersecurity. All cyber operations, even the ‘defensive’, involve the use of malware by an actor to gain unauthorised access into the target’s system. A vulnerability in a system is not threatening per se if not exploited, and this exploitation requires an adversary’s or another party’s involvement. Although the resemblances with the military sector here are high, one more aspect makes cybersecurity more distinctive: the question of existentiality. According to the securitisation theory, the defining feature of security is the idea of existentiality; i.e. security is concerned with the *survival* of a certain referent object(s), which justifies the calls for urgent responses.

In cybersecurity, despite the existence of existential discourses, the existentiality assumption is not as straightforward as it is in other sectors for multiple reasons. Firstly, the majority of cyber attacks that are seen as the most serious in history were neither objectively existential from a technical viewpoint, nor portrayed as such by the concerned actors. Stealing military, commercial, or personal information can hardly affect the survival of the state, the private sector, or any individual. Similarly, denying customers/citizens access to certain services through denial of service attacks (DOS) does not pose an existential threat to anyone. This does not mean that cyber threats cannot be hyped, exaggerated, or presented in urgent terms, since all those qualities are not essentially linked to existentiality. Secondly, the indirect nature of the majority of cyber attacks and the non-physicality of their consequences, although does not undermine their seriousness and urgency, acts as an *impeding* rather than a facilitating condition to the existentiality assumption. The empirical analysis also proves that existentiality is not the only reason for threats to register in the cybersecurity debate and that it is not a precondition for perceived *urgency*. Generally, cybersecurity is marked by different understandings of *disruptive* and *destructive* implications of cyber threats, and all invoke a certain level of urgency. The majority of discourses emphasise these ‘disruptive’ implications, including huge financial losses that can slow down the economy, loss of productivity and global competitiveness, customers’ loss of confidence in the information infrastructure, etc. Though not portrayed in ‘survival’ terms, these disruptive implications are still perceived as immanent, urgent, and as serious threats to national security.

### **3. The constellation of referent objects in cybersecurity: cross-sectoral connections**

A referent object of security is the object being threatened and the one that security policies aim to protect or secure. In the securitisation theory’s framework, the identification of something as a referent object is always linked to a legitimate claim to existentiality (Buzan, Wæver and Wilde, 1998, pp. 103–104). Nevertheless, applying the same logic on cybersecurity disregards a wide range of important referent objects that lacks this survival quality due to the distinctive nature of cyber threats. Generally, the identification of an exclusive set of referent objects in cybersecurity is not an easy task, because such objects are subject to competing discourses and conflicting interests of multiple actors. As argued by Hansen, cybersecurity is better analysed through the “*competing articulations of constellations of referent objects*” (Hansen and Nissenbaum, 2009, p. 1163). And since the traditional public-private and individual-collective divide is blurred in cybersecurity, it is common to find strong links among them all in the same discourse, of which the following statement is an example: “Our national security, public safety, economic competitiveness, and personal privacy are at risk” (Emerging Cyber Threats to the United States, 2016, p11).

The relationship between cybersecurity and other security sectors can be first examined in the case of critical national infrastructure (CNIs). CNIs are defined as the “public and private institutions in the sectors of agriculture, food, water, public health, emergency services, government, defence industrial base, information and telecommunications, energy, transportation, banking and finance, chemicals and hazardous materials, and postal and shipping” (The White House, 2003). They are usually granted more importance than individual or corporate cybersecurity: “The risks to that infrastructure are greater than the sum of the risks to the individual companies” (Overview of the Cyber Problem, 2003, p13). These infrastructures represent a unique case of the intersection between private and national security, since 85% of them are owned and run by the private sector. Additionally, they represent a constellation of referent objects per se, given their perceived connections to the

'state's security', 'economy', 'way of life', 'lifestyle', 'military operations', 'personal communications', 'public health', etc.

In addition to CNIs, particular intersection with economic security can be seen in portraying 'economic competitiveness', 'business opportunities', 'innovation', 'customers' confidence' as referent objects of cybersecurity, especially against threats of cyber espionage and intellectual property rights theft. Again, here, the public and private are intertwined; it is not just firms that are perceived as threatened, but also the American global 'competitive advantage' and 'economic leadership'. The military as a referent object is also one important component of many cybersecurity discourses, particularly by the executive branch. This was intensified after cyberspace has been declared by the state as a domain of warfare in 2010, just like land, air, sea and space (The White House, 2010; U.S. Department of Defense, 2010). Here, the survival of the armed forces is not necessarily presented as directly threatened as in the military sector, rather it is the survivability of military operations and communications, the military's defence and emergency capabilities, and its ability to utilise cyberspace as a force-multiplier that are perceived as referent objects.

Another very controversial referent object is 'privacy and civil liberties', which usually puts the individual in the centre of cybersecurity discourses. There are two ways in which privacy and civil liberties are constructed as referent objects. The first category of discourses focuses on how privacy is threatened by cyber attacks like identity theft and espionage. The second category questions the negative implications of cybersecurity policies on privacy and civil liberties and criticises the binary of 'security vs. privacy' that is sometimes used to legitimise government practices such as imposed backdoors and surveillance. These criticisms particularly deepened following Edward Snowden's revelation about the NSA's spying practices. As argued in an editorial: "Personal freedom or public safety? In our current environment, it seems one is increasingly taking a back seat to the other" (*The Lowell Sun*, 2015). For instance, voices are divided between a complete support for increased encryption on the basis that it positively affects innovation, the economy, and human rights, and those who oppose it arguing that it can impede law-enforcement processes and facilitate terrorists' communications. It has to be noted here that it is not just the government that tries to legitimise its opposition for increased encryption and its calls for backdoors by prioritising security over privacy. Other voices sometimes also argue that "We cannot achieve privacy without cyber security" (Cyber Security-2009, 2009, p29), and that giving agencies like the NSA more information that affects privacy is "a small price to pay for national, public health and energy security" (Brunner, 2015). These latter discourses implicitly assume that state's security is more important than individual's security, and misses the fact that deliberately weakening encryption may affect the overall security of the system on the long-run, given its links with intellectual property rights and financial transactions (Brantly, 2016).

#### **4. Security actors in the cyber ecosystem: From the co-production of cyber technology to the co-production of cybersecurity**

*"Every computer company you bring into this room will tell you that liabilities will be bad for their industry. Of course they're going to tell you that; it's in their best interests not to be responsible for their own actions. The Department of Homeland Security will tell you that they need money for this and that massive government security program. Of course they're going to tell you that; it's in their best interests to get as large a budget as they can. The FBI is going to tell you that extreme penalties are necessary for the current crop of teenage cyberterrorists; they're trying to make the problem seem more dire than it really is to improve their own image. If you're going to help improve the security of our nation, you're going to have to look past everyone's individual self-interests toward the best interests of everyone." - Bruce Schneier, Counterpane Internet Security, Inc., congressional hearing (Overview of the cyber problem, 2003, p15).*

Classifying cybersecurity actors is not an easy task. The traditional divisions between 'public' and 'private' actors cannot grasp the complexity of the cyber ecosystem and the conflict of interests it is characterised by. This public-private classification gives a false image of a non-existent coherence among the various actors in each category. For instance, on the government side, it is not possible to combine the presidency, DHS, DoD, and the NSA all in one category, since each has their own interests, powers and responsibilities, and technical capabilities in cybersecurity. On the private sector side, a wider division of labour exists among software and hardware vendors, owners and operators of national infrastructures, internet service providers (ISPs), companies running search engines and social media platforms, security firms that provide consultation or insurance services, and all the rest of private sector corporations and entities whose security is essential to the state.

This multiplicity of actors challenges the Copenhagen school's conceptualisation of what constitutes a *securitising actor*. The theory defined the securitising actor as the one who *speaks security* or declares a referent object as existentially threatened by a speech act, which could be any individual or group, not necessarily the state. This is different from 'functional actors', who influence decisions taken to handle a threat without trying to securitise it themselves (Buzan, Wæver and Wilde, 1998, p. 40). Yet, the theory did not go further to consider those entitled with *acting security* or taking the decisions that security speech acts seek to influence. It can be argued that by being silent on who has the power not just to *speak* security but to *act* security, the securitisation theory retains a state-centric perception of security environments. Anyone can *securitise*, but it is the state that makes *security* happen. However, the situation is completely different in cybersecurity. Not only are the private actors significant securitising actors and referent objects of their own, they are in most cases the ones who act security and take the most critical decisions that affect the cybersecurity of the whole nation.

In the cybersecurity debate, there is always a controversy over assigning responsibilities for achieving cybersecurity and determining liabilities for cyber insecurity. Generally, most discourses emphasise the idea of 'collective responsibility', that no one entity can control or achieve cybersecurity without cooperation from all stakeholders, including individual users. Moreover, the government sometimes refer to the private sector as the more capable actor in leading cybersecurity than the government, and that it should form 'the first line of defence' (The Department of Defense, 2015, p. 5).

Accordingly, opinions are divided on whether the government should leave the market forces decide, or should it have a more regulatory role. Proponents of the second view always warn against giving the government the power to micro-manage cybersecurity, which can cripple innovation, slow threat response processes, and harm the economy. The kind of intervention endorsed from this point of view is one that commercialises cybersecurity and support the indirect role of the government in influencing the market as a security customer. However, on the other side, other discourses note the mismatch between national security and commercial security interests: "The challenge is market forces are not designed to respond to national security threats. You cannot make a market case for the Cold War" (Securing America's Future: The Cybersecurity Act of 2012, 2012, p44). The interventions that these arguments call for include: incorporating code integrity clauses in contracts to hold vendors accountable; increasing government's funds for cybersecurity research; forcing businesses to declare when they are subject to cyber intrusions; improving information-sharing with the private sector; among others.

Alas, the most controversial political debate regarding the role of government in cybersecurity is related to the DoD and the NSA. Since its establishment in 2002, the DHS was given the main role of leading the national cybersecurity program following the release of the National Cybersecurity Strategy in 2003 and the establishment of its National Cybersecurity sub-division. Given its civilian nature, there has been no debate on whether it should be involved in cybersecurity or not. The majority of criticism directed towards the agency is usually regarding its effectiveness in securing the .gov infrastructures and cooperating with the private sector. On the other side, the DoD has considered the need to actively operate in cyberspace since 2006, when it regarded cyberspace as an essential component of its military operations. This was followed in 2010 by an official declaration of cyberspace as an 'operational domain' of warfare, and the establishment of the US Cyber Command. Similarly, for many years the NSA has been trying to stretch its prerogatives in cybersecurity policy by emphasising the similarities between military and civilian cybersecurity processes. In fact, this discourse is not just sponsored by the NSA, but also by some security firms and think tanks. According to them, the NSA has more technical acumen to lead national cybersecurity, particularly given what they perceive as a failure of the DHS. Nevertheless, more voices have been criticising this argument and warning from the dangers of militarising cyberspace. They argue that the NSA's role is the main hurdle in information-sharing with the private sector, because everyone fears that the shared information might end up in the NSA's bulk data collection program.

On the question of liabilities for cyber insecurity, disagreements are even deeper. We can distinguish between three discourses in this regard. The first always blames 'America's adversaries' or 'enemies', that are usually 'external' or 'foreign', as argued earlier. It focuses absolutely on blaming the attacker, with implicit assumptions that complete security is impossible. The second focuses on the end-users or customers in a neo-liberal discourse, viewing them as responsible for their insecurities due to their inability to deal with the uncertainties of a risk society. End-users are thus criticised for not updating their systems regularly and configuring them properly. The third discourse focuses on the role of the industry, particularly software vendors. Here, the idea of blaming end-users is criticised, as argued by a think tank representative in a congressional hearing: "cyber crime is the only crime I know of where we blame the victim" (Emerging Cyber Threats to the United States,

2016, p12). Software vendors and private corporations transfer the risk to customers, who suffer the consequences with little or no cost on the side of vendors. Therefore, one solution that is being advocated by a wide-spectrum of actors is imposing liability on the industry, whether software producers or network operators. However, vendors usually use the complexity argument and the idea of the impossibility of vulnerability-free software to get exempted from any sort of liability. Additionally, many vendors argue that imposing liabilities will damage the industry and increase the software cost on customers.

## **5. Conclusion**

This paper used the idea of sectoralisation as presented by the Copenhagen' school securitisation theory to study the peculiarities of cybersecurity discursive construction. It did so by analysing cybersecurity discourses of the government, the private sector, think tanks, and the media over the period from 2003 until 2016 in the US. In discussing the logics of threats and vulnerabilities in constructing the cyber threat, the paper showed how cyber dependency is securitised in a way that legitimises the emphasis on a threatening future, the need for more security, and a certain level of risk acceptance. Threat attribution is also one important characteristic of the cyber threat construction, in which the political may override the technical. Discourses that attribute the cyber threat to Russia, China, Iran, and North Korea have been growing over the years, using the antagonistic relationship between those countries and the U.S. as a facilitating condition for cyber securitisation. Such discourses present the cyber threat as urgent and immanent, but not necessarily always *existential*.

On the referent object of cybersecurity, the paper examined the cross-sectoral connections between cybersecurity and other sectors, particularly the economic, political and military. Again, although critical national infrastructure constitutes the main referent object in the majority of discourses, other objects that lack the existential quality specified by the securitisation theory register in the cyber securitisation discourses. Finally, the paper discussed the multi-stakeholder nature of cybersecurity and how this security is being co-produced by a wide range of actors representing different, and in some cases contradicting, interests. By analysing aspects of power, responsibilities, and liabilities, the paper explored the main controversies in cybersecurity debates over state's role and intervention, the liabilities of the industry, and the responsibilities of end-users. It can be argued that there is no *one* discourse on cybersecurity or cyber threats, and it is a simplification to assume that there is even one discourse that represent each securitising actor, be it the government or the private sector. This diversity explains why the securitisation theory's assumptions and logics can apply on only some but not all discourses of cybersecurity.

## **Acknowledgments**

I would like to thank my PhD supervisors, Prof. Stefan Elbe and Dr. Stefanie Ortmann, for their feedback on the chapter based on which this paper was written.

**Funding:** This paper was written as part of my PhD research, which is funded by the University of Sussex's Chancellor International Research Scholarship.

## **References**

- America Is Under Cyber Attack: Why Urgent Action Is Needed, Hearing before the Subcommittee on Oversight, Investigations, and Management, of the Committee on Homeland Security (Serial No. 112-85), U.S. House of Representatives, 112<sup>th</sup> Cong. (2000).
- Beck, U. (1992) Risk Society: Towards a New Modernity. London ; Newbury Park, Calif: SAGE Publications Ltd.
- Beck, U. (2002) 'The Terrorist Threat: World Risk Society Revisited', Theory, Culture & Society, 19(4), pp. 39–55. doi: 10.1177/0263276402019004003.
- Bendrath, R., Eriksson, J. and Giacomello, G. (2007) 'From "Cyberterrorism" to "Cyberwar", Back and Forth: How the United States Securitized Cyberspace', in Eriksson, J. and Giacomello, G. (eds) International Relations and Security in the Digital Age, pp. 57–82.
- Brantly, A. (2016) 'A Holistic Approach to the Encryption Debate', in Herr, T. and Harrison, R. M. (eds) Cyber Insecurity: Navigating the Perils of the Next Information Age. Rowman & Littlefield, pp. 191–204.
- Brunner, J. (2015) 'Sharing Is Caring: Obama's New Cyber Security Executive Order', The State Press: Arizona State University, 19 February. Available at: <https://www.nexis.com/> (Accessed: 14 February 2018).
- Buzan, B., Wæver, O. and Wilde, J. de (1998) Security: A New Framework for Analysis. Boulder, Colo: Lynne Rienner Publishers.
- Dunn Cavelti, M. (2008a) Cyber-Security and Threat Politics: US Efforts to Secure the Information Age. London: Routledge (CSS studies in security and international relations).

- Dunn Cavelti, M. (2008b) 'Cyber-Terror—Looming Threat or Phantom Menace? The Framing of the US Cyber-Threat Debate', *Journal of Information Technology & Politics*, 4(1), pp. 19–36. doi: 10.1300/J516v04n01\_03.
- Eriksson, J. (2001) 'Cyberplagues, IT, and Security: Threat Politics in the Information Age', *Journal of Contingencies and Crisis Management*, 9(4), pp. 200–210. doi: 10.1111/1468-5973.00171.
- Fairclough, N. (2001) *Language and Power*. 2 edition. Harlow, Eng. ; New York: Routledge.
- Fairclough, N. (2003) *Analysing Discourse: Textual Analysis for Social Research*. London ; New York: Routledge.
- Goldsmith, J. and Hathaway, M. (2010) 'Cybersecurity Changes We Need', *The Washington Post*, 29 May. Available at: <https://www.nexis.com/> (Accessed: 14 February 2018).
- Hansen, L. and Nissenbaum, H. (2009) 'Digital Disaster, Cyber Security, and the Copenhagen School', *International Studies Quarterly*, 53(4), pp. 1155–1175.
- Libicki, M. C. (2009) *Cyberdeterrence and Cyberwar*. Santa Monica, CA: RAND.
- Rid, T. and Buchanan, B. (2015) 'Attributing Cyber Attacks', *Journal of Strategic Studies*, 38(1–2), pp. 4–37. doi: 10.1080/01402390.2014.977382.
- Rivera, J. and Hare, F. (2014) 'The deployment of attribution agnostic cyberdefense constructs and internally based cyberthreat countermeasures', in 2014 6th International Conference On Cyber Conflict (CyCon 2014). 2014 6th International Conference On Cyber Conflict (CyCon 2014), pp. 99–116. doi: 10.1109/CYCON.2014.6916398.
- Schutte, S. (2012) 'Cooperation Beats Deterrence in Cyberwar', *Peace Economics, Peace Science and Public Policy*, 18(3). doi: 10.1515/peps-2012-0006.
- The Department of Defense (2015) 'The Department of Defense Cyber Strategy'. Available at: [http://www.dtic.mil/doctrine/doctrine/other/dod\\_cyber\\_2015.pdf](http://www.dtic.mil/doctrine/doctrine/other/dod_cyber_2015.pdf) (Accessed: 23 March 2017).
- The Lowell Sun (2015) 'Cyber Security at What Privacy Price?', 25 February. Available at: <https://www.nexis.com> (Accessed: 14 February 2018).
- The White House (2003) 'The National Strategy to Secure Cyberspace'. United States Government. Available at: [https://www.us-cert.gov/sites/default/files/publications/cyberspace\\_strategy.pdf](https://www.us-cert.gov/sites/default/files/publications/cyberspace_strategy.pdf) (Accessed: 22 March 2017).
- The White House (2010) 'National Security Strategy of the United States'. United States Government. Available at: [https://obamawhitehouse.archives.gov/sites/default/files/rss\\_viewer/national\\_security\\_strategy.pdf](https://obamawhitehouse.archives.gov/sites/default/files/rss_viewer/national_security_strategy.pdf) (Accessed: 2 April 2017).
- The White House (2011) 'International Strategy for Cyberspace: Prosperity, Security, and Openness in a Networked World'. United States Government. Available at: [https://obamawhitehouse.archives.gov/sites/default/files/rss\\_viewer/international\\_strategy\\_for\\_cyberspace.pdf](https://obamawhitehouse.archives.gov/sites/default/files/rss_viewer/international_strategy_for_cyberspace.pdf) (Accessed: 22 March 2017).
- U.S. Department of Defense (2010) 'Quadrennial Defense Review Report'. Available at: [https://www.defense.gov/Portals/1/features/defenseReviews/QDR/QDR\\_as\\_of\\_29JAN10\\_1600.pdf](https://www.defense.gov/Portals/1/features/defenseReviews/QDR/QDR_as_of_29JAN10_1600.pdf) (Accessed: 2 February 2018).

# E-Health Systems in Digital Environments

Aarne Hummelholm

Faculty of Information Technology, University of Jyväskylä, Finland

[Aarne.hummelholm@elisanet.fi](mailto:Aarne.hummelholm@elisanet.fi)

**Abstract:** As we live in the digital world, people can be provided with more effective treatment methods that allow them to live longer in their home and to live there better. People can be provided with better home care and preventive health care. People can easily carry portable sensors and intelligent devices in their bodies and wrists that relay their vital information to hospital systems in real time, from which healthcare staff can track human vitality even in real time. Although the digital world offers good opportunities to improve healthcare systems and make disease analyses more effective, we must look deeper about that issue. Devices and systems may not work well together. Almost every manufacturer has their own technical solutions and they only work in certain environments. There is a great need for unified concepts and for IT platform solutions in healthcare systems. The technology currently in use is very varied. Standards are developing now days, but they are not yet ready. In addition, the lack of technical and functional requirements for telemedical communications systems and equipment, as well as the requirements for providing secure data transmission in remote medical care. In the news we can see and hear often that there are a lot of medical devices that have damaged the patient's health around the world. And then there are a lot of vulnerabilities and that means security risks, cyber risks and the risks of reliability of data. These risks are associated with IoT devices and sensors, and in the field of data transfer. This document describes telemedicine solutions for the future of society. Includes a brief introduction of hospital equipment in a hospital environment and a patient's home. The main overall is the communication arrangements, consisting of the bio-signal formation of the patient's sensor and the flow of bio-signals to the hospital information systems for analysis and monitoring. This study examines cyber threats and attacks against e-health systems and what that means for patients' health. This study also examines the authenticity, traceability, authentication and protection of privacy.

**Keywords:** healthcare systems, vulnerabilities, cyber threat, cyber-attacks, telemedicine

---

## 1. Introduction

Today, we want to exploit the potential of digitalisation, including in the services of health care, in diagnostics and in the analysis of diseases, in the precautionary and in monitoring the progression of the disease. However, the rapid technological advances underlying digitalisation set their own challenges for technical systems of health care and the whole health services, with their guidelines and regulations.

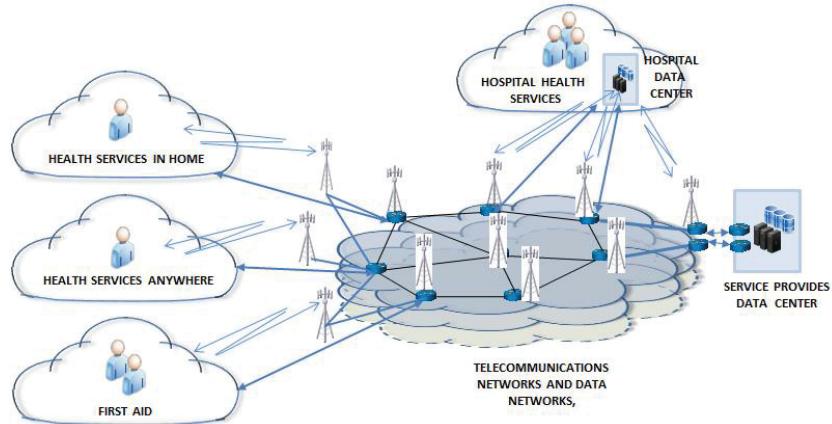
It is imperative that health services are available for 24/7, regardless of time and place, anywhere qualitatively and equitably, not forgetting that patients or the elderly may be in cities or in rural areas long distance from the hospital or from treatment point. In addition, patient spatial data may be used to indicate the person's whereabouts so that it is possible to warn them if necessary, as soon as possible and to get help to the right place.

Health care is looked at in terms of services and quality, but at the same time cost-effectiveness has become increasingly important in the decision-making process. This also affects patient care methods and solutions. The goal is to organise treatments so that patients are in hospital so as short time as possible, and then patients will be sent home if the necessary conditions for arranging home care are in place.

The one aim of digitalisation is to create the conditions for the activities described above. The aim of digital systems is to provide the patient with treatment so that they can be sent to home care without undermining the quality of care or adequate levels, even in the circumstances prevailing at home. All this requires the introduction of new technologies and the integration of different types of transportable equipment, such as IoT, various sensors and actuators, as part of health systems. Together they produce a lot of information from the patient condition, from the elderly's condition and the environment in real time. These large data are analysed in hospital systems and analyses are leading to the necessary treatment-related measures.

As the pace of development has been very rapid and new technology has been introduced very quickly, the international standard work has not been involved in the development process. We often have manufacturer-specific solutions for IoT devices, different sensors and data storage systems in some of the service providers' Data Center, shown in Figure 1. This issues in turn leads to a challenging of connection of IoT devices to smart devices.

Smart devices are then connected to fixed or mobile networks and are used to transfer the patient's bio-signal data to hospital systems. In hospital systems, the information is analysed, and the care staff take the necessary decisions based on analysed results and gives information on management measures to the patients.



**Figure 1:** E-Health top level architecture

Also, for many technical specifications, the situation is variable, and the terms used vary, depending on the speaker, or depending on the time in which the term is used. We talk about the following terms, which vary slightly in importance: Telemedicine, Telehealth, e-Health and m-Health. In this document e-Health and m-Health terms are used. All this diversity of terms and undefined is also a source of challenges for security solutions, privacy and cyber security issues (EU- GDPR).

## 2. Chapter 1: Objective and grouping of paper

The research question is that whether e-Health systems are being used safe in digital environment.

To find the answer to the question, this study seeks to find a model that facilitates cyber threat assessment and threat comparisons and facilitates threat analysis in e-Health environment and systems. The results obtained through the model aims to facilitate the design and implementation of architectural solutions. The model implemented can help in better assessing the cyber-threat scenarios of future telecommunication environments and their impact probabilities e-health systems. The study draws on an e-Health operating environment with top-level architecture, figure 1, where services and infrastructure are grouped into different usage cases. Usage cases are divided into end-to-end communications segments. From these devices we can analyse vulnerabilities and from those we can analyse and define cyber threats to various devices, threats to services and threats to information systems. After that we can more accurately define, evaluate and analyse the cyber threats in whole system and get better overall picture of situation.

As shown in figure 1, there are several other viewpoints. Examination of threats can be made in the various usage cases with patient wearables, sensors and IoT devices including environments devices and e-Health services with them. When a patient or an elderly person is at home or outdoors using an e-Health application on their smart devices, he or she can also use other social services that are provided via telecommunications networks to him. This means that a patient or an elderly person is connected to the e-Health system in hospital or in a home and at the same time to another service or social media service that is currently being offered to people through the Internet. This will be a threat to healthcare systems and possibly to the entire e-Health service when patients or elderly persons exchange of information between the different service segments. The networks of those service providers are used to integrate health data of patients or elderly people to the hospitals systems. We must look carefully about this end-to-end communications path and all devices which are connected to that path. The above-described integration accelerates at all levels of activity, in each region both horizontally and vertically. Analysing the latest technologies and their services and applications in these smart environments will further complicate making cyber threat estimates. These considerations are considered in the selection of the target area for which the final dependency analyses, cyber-threat assessments, risk assessments and analyses are made (Aarne Hummelholm/2018).

Chapter 2 presents the ecosystems and collaborative environment formed by active nodes as well as an architecture model that describes the current operating environments of the hospital, the home, outside home,

telecommunication networks and hospital data center at the general level with their interfaces. The virtual environments management and control and telecommunication networks and data centers is part of this whole. Chapter 3 describes the cyber-threats against the future health care systems and the models which are used to make threat analysis to e-Health infrastructures and services. The chapter 4 deals with making and modelling of threat analyses and chapter 5 conclusions, our solution model and future work

### 3. Chapter 2: Description of the future operating environment

Even today, the situation the hospitals and healthcare centres are the key treatment points where the sick is being examined and cared for. In many situations, access to treatment takes time and frustrates patients or older people. This may be due to a shortage of healthcare personnel, but also long geographical distances may impede medical examinations or access to treatment. Long distances also increase costs and therefore it is not always possible to achieve the cost-effective solution associated with each treatment situation. People may have to travel to a care point only to check their situation, which could have been checked with digital systems without the need for entry. As a result, the development of digital methods of treatment with sensors and IoT devices is strongly developed in order to improve patient care, to look at their condition remotely at home or wherever they are moving in real time. The future healthcare operation environments are presented in figure 2.

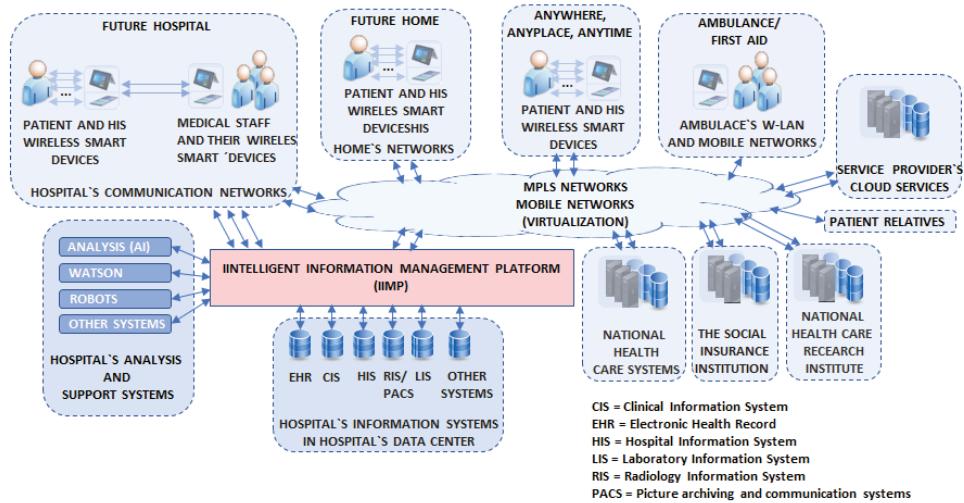


Figure 2; E-Health or m-health operating environment, top level architecture

In order to improve health care, new digital hospitals have been introduced which are more cost-effective than previous traditional hospitals and are geared towards better treatment accuracy and performance in treatment processes and are also more effective in diagnosing diseases. The Digital hospital environment utilizes the digital Hospital Patient Report Systems (EHR), which are the one main systems in this operation (figure 2). In the future hospital, the patient has several digital sensors and IoT devices attached to them to collect their health-related bio-signals, which are sent through the patient's smart device to the hospital Information systems for analysis and follow-up (figure 3).

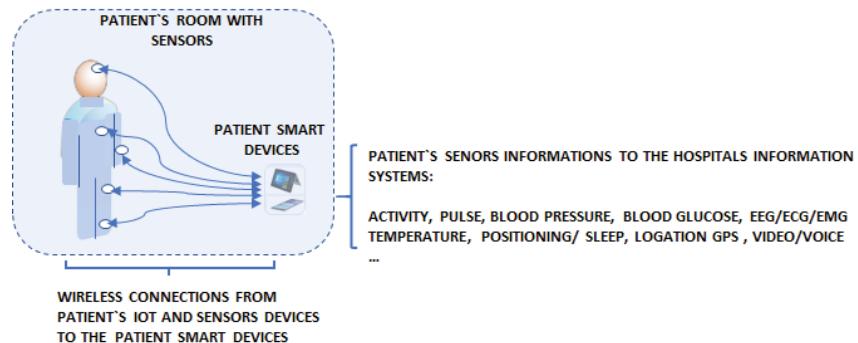
The e-Health or m-Health system taxonomy can be described, describing the data streams and processes when the data is taken by the patient and they become a hospital medical information system where the data is analysed and draws conclusions (Robert S. H. Istepanian).

Patients e-Health or m-Health sensors and hospital healthcare systems taxonomy is composed of the following categories: Health and wellness monitoring, Diagnostic sensors, Prognostic and treatment sensors and Assistive sensors. Each of these categories have their own sub-categories. And these subcategories are further subdivided into sub-categories. Intelligent Data Management Platform (IIMP), (figure 2), types of systems are needed in order to monitor the flow of information from patient's IoT devices and sensors to the hospital health care systems and it also gives possibilities to health care staff to analyse information fast. This also ensures that care staff can find information about patients for analysis in critical situations as well. This type of system can also help you find possible anomalies or changes of data, or if someone has attempted to penetrate or use the data in a way that is not desired.

The hospital has its own local network (LAN) and its own wireless network (W-LAN), through which medical records go to databases in the hospital's data center. The medical data in the hospital databases can be monitored by doctors and medical staff to make the necessary patient care decisions and management measures. The sensors and IoT devices belonging to the infrastructure's automation systems can be also connected to the hospital's local area networks, as well as all the hospital's internal communications systems but this is risk.

This situation means that all clever IoT - devices, sensors and clever terminals would be in the same network systems and possibilities to connect to the hospital's data center. From those, the hospital's networks and hospital's data center have also connections to the internet and to various external information systems such as National Health Care systems and National Social Insurance system (figure 2).

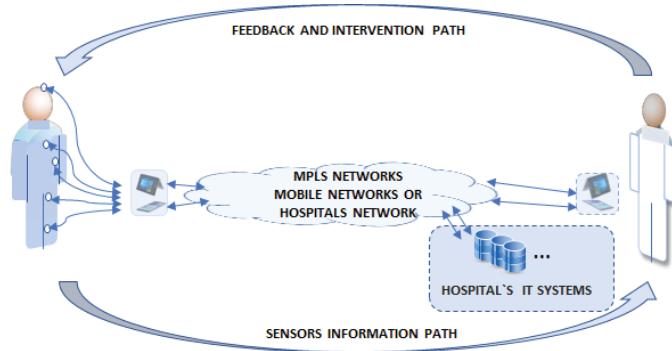
Communication between the patient's smart device and IoT devices and the sensors attached to them is done either through fixed wired connections or wireless connections, including different wireless technical solutions as Bluetooth in different versions, ZigBee, W-LAN, RFID and WiGig, (figure 3). These technologies allow for connection distances ranging from 1 metre to the 100 meters. When the hospital has a wide range of sensors and real estate automation systems that operate in the same frequency bands and when same frequency bands can also be used in hospital staff devices, the emergence of incidents is possible in the form of mutual interference. The new public buildings that are being built under the EU directive can also prevent the operation of wireless communications due to the large damping of the walls and windows of the building and the mobile networks are not working properly indoors (EU -2012/27/EU). In the patient's home may be also lot of sensors and IoT devices, that uses same frequencies than patient's smart device. The inside the patient's home may in this situation be also these frequency interferences.



**Figure 3:** Patient's IoT and sensor devices connections

One important factor in the hospital and in the treatment of patients in the digital environment, figure 2 and 3, is to ensure the proper functioning of communications. When we talk about the human spirit and related issues, it is also necessary to take into account, that digital devices do not operate without electricity.

Ensuring communication is very important when looking at the situation of the patient in rural areas where the supply of electricity may be completely cut-off for hours or even several days due to storms or snow disasters. In this situation, it is very important that the doctor receives information about the patient in one way or another. Doctors and/or medical staff should also be able to monitor the patient's condition at home in a real time situation, so that the necessary steps can be taken in time to guide the patient to necessary measures, (figure 4). Figure 4 shows the flow of patient information from his or her smart devices to the hospital system from which the medical staff receives the information and sends feedback to the patient's smart device. In this way, the exchange of information between the patient or the elderly and the care staff is carried out at a general level, whether in the patient's hospital, at home or outdoors on the town or anywhere. In table 1 we can see the bandwidth needs of the patient's sensors in the communications networks and at what transfer rate data is transferred via telecom networks. The table also shows the delay values that must be reached through communication connections. The values in the table 1 are those obtained from research results, but not yet the actual requirements or recommendations of our nursing systems.



**Figure 4:** A general wireless and fixed m-health monitoring system

**Table 1:** Data rates and bandwidth of key biomedical wireless monitoring, (Robert S. H. Istepanian)

Physiological/Biomedical Parameter	Bandwidth	Rate Latency/Data
ECG (12 leads)	0.1 – 1 kHz	~144 Kbits/<200 ms
EEG (12 leads)	0.1 – 0.2 kHz	~40 Kbits/<300 ms
EMG	0 – 10 kHz	~350 Kbits/<200 ms
Body temperature	0 – 1 kHz	~0.1 kHz
Medical imaging and video streaming data	-	~ > 10 Mbps/<100 ms
Speech and voice	-	~50 – 100 Mbitps/<10 ms
Accelerometer and motion sensing	0 – 0.5 kHz	~30 Kbitps
Blood glucose monitoring	0 – 40 kHz	~1.5 Kbits
Blood pressure	0 – 1 kHz	~15 Hz

#### 4. Chapter 3: Cyber Threats against the future health care systems

As seen in Figure 2, how does the future health care Information System form an extensive and complex package with a range of co-operation requirements and, above all, a highly critical condition for the wellbeing of the patient or the elderly environment. Many wireless technologies are used in the operating environment, and the operating systems do not form a closed set, which it would be possible to have as a separate island without connections to the outside world. In addition, the health registers include the personal data of all citizens and information about their illnesses. These systems are linked to external systems that allow people to send information to those systems. This in turn will raise hackers and cyber players' interest in penetrating one way or another into health systems, because from there is a chance to also have an economic benefit. Cyber attackers can also use system vulnerabilities to harm people who have been selected.

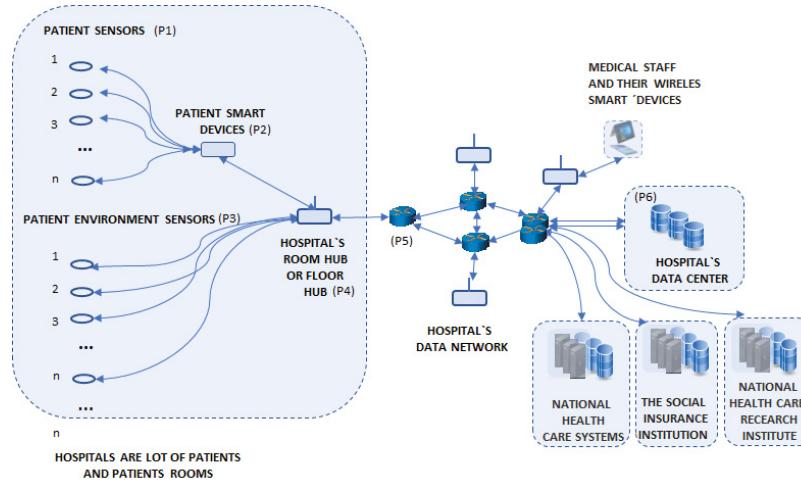
In general, the security threats in wireless health networks include the following issues like:

- monitoring and eavesdropping of patients` vital signs,
- threats to information during transmission,
- routing threats in networks,
- location threats and activity tracking
- denial of service (DoS) threats
- interfere with or inhibit the radio communication of IoT devices and sensors
- using vulnerabilities to get access to the health care services
- attack against to the hospital health care information`s systems
- disrupt or impede the entire hospital's wireless communication and prevent the use of the hospital's daily activities

If attackers know patient location, attacker can follow patient route, where he or she is living, what route or bus or train he or she is using, which kind of car he or she use and so on. Attacker has possibilities to find vulnerabilities in his or her smart device or hospital equipment so, that he can attack against patient's devices or trough those devices, against his or her car and maybe cause an accident and perhaps the patient dies as a result of the accident. After the accident, accident investigators think that the patient received a disease attack

and it occurred for this reason. This can happen especially to VIPs who are in prominent positions if someone wants to inflict damage on them

One of the attackers' goals could be to use these patient Smart Devices in order to get inside the systems from which we paid people grants (figure 5). These systems contain billions of euros in money. When cyber attacker attacking to health systems and hospital systems can it be quickly paralyse the entire society and make people desperate for their future, (Russell Brandon, Sky news, ABC news).



**Figure 5:** A general wireless and fixed m-health monitoring system with environment sensors in hospitals

## 5. Chapter 4: Making and modelling of threat analyses

When looking at figure 2, we see that health care systems are extremely complex and have interfaces in many directions to wireless networks or fixed networks. In addition, there are several service providers, admins and many stakeholders working on the same networks and information systems. The operating limits of the responsible actors must be able to define and instruct the functions so that interoperability is ensured between operators and those using the services. In order to be able to conduct cyber threat assessments and analyses, we must first work with architectural descriptions of the operating environments of the hospital, home, and outside home, with equipment and operating processes. In order to prevent or even reduce the risks described chapter 3, all parts of the health care systems should be segmented into their functional parts and a distinction should be made between them with enough safety mechanism which is a certain type gateway solution. These segments will then be examined also based on use cases. Then we must also look carefully attackers, their capabilities and motivations to disrupting health care systems and organisations, table 2.

Thereafter, data streams can be defined from the bio-signal produced by the patient's sensors to the servers. These data streams can be used to identify the devices and systems associated with each use case situation and to view related dependencies and vulnerabilities. Based on dependencies and vulnerabilities, risk analysis can be defined. These bases, it is possible to make an information and cyber threat, as well as to determine the probabilities of the cyber threats. The detected dependencies, risks, and vulnerabilities can be exported to a table where they are easily extracted for mathematical processing. Dependencies, risks and vulnerabilities can be given a number estimate of the probability of realisation that can be utilised in mathematical review. We can use attack tree models to count cyber threats probabilities. In addition, preliminary analyses of the tabular form can be used, which have already defined, based on sensitivity analysis, a preliminary assessment of each threat. Mathematical processing clarifies the whole and gives a better picture of the threats if they are mathematically presented. It is possible to compare the cases in parallel and to make decisions based on the results. That is worth doing both for the sake of victims and to see the effect of the measure of the test.

In Figure 2, e-Health or m-Health, we can look at the various use cases and, in these cases, to review the service chains, from which we can look for dependencies, risks, vulnerabilities and give them a probability value. These values are put in table 3 which can be used help to calculate the probabilities of each case. Examine the use case in which the patient is hospitalized, figure 5. Figure 5 shows that quite several IoT devices and sensors are connected to the patient's smart device and that the device is connected to a room or floor access point (HUB). The same base station is also connected to various heating, ventilation, air conditioning and cooling systems

(HVAC) systems, etc. Manipulation all these systems can allow an attacker to access and exploit sensitive data stored in hospitals data centers.

**Table 2:** Capabilities and motivations for disrupting health care systems and organisations, example (based on Aurore LE BRIS, Walid EL ASRI, 2016)

		Patient Health		Patient / Hospital Records		National Health Care Systems	
Adversaries / Attackers		Targeted (Specific Victims)	Untargeted (Not Specified)	Targeted (Specific Victims)	Untargeted (Not Specified)	Targeted (Specific Victims)	Untargeted (Not Specified)
Individuals / Small Group					Yes		
Political Group / Hacktivists				Yes			
Organized Crime	Yes			Yes		Yes	Yes
Terrorism / Terrorist org.	Yes	Yes			Yes	Yes	Yes
Cyber Attackers				Yes		Yes	Yes
Nations, States	Yes	Yes	Yes		Yes	Yes	Yes

Making a threat analysis of the whole future hospital system, or just of an entity comprising of service sectors as part of the future intelligent health care systems is a challenging task. Therefore, threat analysis is made based on Figure 5. The entirety consists of Access networks, in a patient room with IoT devices (P1 1...n), patient's smart device (P2), room HUB with HVAC systems (P3), core networks (P4) and data centers networks (P5). The following threat analysis is made for the patient's access network as shown Figure 5, as currently it is subject to major changes and because it involves large quantities of different sensors and IoT devices. The starting point for the analysis is a smartphone interface the IoT devices connected to that, and the connection of the smartphone to the patient's room HUB system and then to the hospital core router. The core router may be virtualized, and it also includes some services in their hospitals' systems own slices. In addition, the HUB system may include firewall functions.

**Table 3:** Threat and risk table model

Ref ID	Org	Functions	Category	Threat	Threat/ Risk	Existing Control	Threat /risk level			Accept/reduce	Recom-mended control	Residual Threat/ Risk			Check Point
							L	C	R			L	C	R	
1 / AH	MC	Identify	Access	Foot-printing	Target Access	IDS /IPS	3	3	8	Reduce	EU- dir.	2	2	3	xx

Where L = Likelihood, C = Consequence, R = Risk

Chapter 3 gives examples of attack mechanisms. How to make aggressive attacks and how attackers try to use vulnerabilities and other mechanisms to gain access to analyse systems and achieve their goals. By doing so the attackers have the possibility to compromise a target device. In these calculations values for these vulnerabilities are got from analysis. The events must be independent and if they are not, we must go to so small entities in order to reach an independent situation with respect to the various functions. The review can be further deepened by examining the vulnerabilities of different OSI layers and by also making analyses of these vulnerabilities.

For probabilistic analysis, defender need estimate the probability of attack success for each node in figure 5, in Attack-Defence Tree (ADT). For the purposes of the review, we define the used notations (Wang, P. (2014)).

**Table 4:** Meaning of notations

Action	Examples	Notation
Attack	Sniffing, enumeration, scanning, ...,	A
Detection	Port scan, information scan, ...,	D
Countermeasure	Analysing of vulnerabilities and to repairing, safeguards put in place, ...,	M

Attacks, probabilistic attack success ( $P(t)$ )

$$P_{1\dots n}(t) = p_{1A\dots n}(t) (1 - p_{D1\dots n}(t)), \text{ to } n \text{ Patient's IoT devices}, \quad (1)$$

$$P_{21}(t) = P_{11}(t)[(p_{2A2}(t) (1 - p_{2D2}(t))], \text{ IoT device to Smart phone}, \quad (2)$$

$$P_{2\dots n}(t) = [(p_{21}(t)) + (p_{22}(t)) + \dots + (p_{2n}(t))], \text{ because, different IoT devices will connect to Smart Phone in different times,} \quad (3)$$

$$P_{3s1\dots n}(t) = P_{3As1}(t) (1 - P_{3ds1}(t)), \dots, \text{ to room connected sensors 1 ... n,} \quad (4)$$

$$P_{4t}(t) = p_{4A1}(t) (1 - p_{4D1}(t)) (1 - p_{4M1}(t)), \text{ HUB}_{(room)} \text{ with attackers, defence and countermeasures} \quad (5)$$

$$P_{4s}(t) = [(p_{3s1}(t)) (p_{3s2}(t)) \dots (p_{3sn}(t))] [p_{4A}(t) (1 - p_{4D}(t)) (1 - p_{4M}(t))], \text{ room sensors are connected to HUB}_{(room)}, \quad (6)$$

$$P_{5t}(t) = p_{5A1}(t) (1 - p_{5D1}(t)) (1 - p_{5M1}(t)), \text{ router}_{(hospital)} \text{ with attackers, defence and countermeasures} \quad (7)$$

$$P_r(t) = [p_{5A1}(t) (1 - p_{5D1}(t)) (1 - p_{5M1}(t))] (P_{4s}(t)) \text{ router}_{(hospital)} \text{ and HUB}_{(room)} \text{ connected} \quad (8)$$

$$P_{6dc}(t) = P_{r1}(t)P_{r2}(t) \dots P_{rn}(t) \text{ hospital's data center router and all hospital router}_{(hospital)} \text{ connected together} \quad (9)$$

The results obtained are then exported to the Threat and Risk Table Model, table 3. The table contains the entities, the activity and category to be considered, the related threat/risk and the controls. Then the table shows the current probability, the resulting consequences and the current risk. In the future it is evaluated how the risk is addressed, what are the recommended controls and remedies with its responsible persons (including organizations). Finally, the equivalent values and checkpoints after remedies are estimated. The table can be done separately for the cyber threat and separately for the risks and furthermore add columns as needed depending on the issues being viewed and related contexts.

## 6. Conclusions, our solution model and future work

A modern hospital has hundreds – even thousands – of workers using laptops, computers, smartphones and other smart devices that are vulnerable to security breaches, data thefts and ransomware attacks. Hospitals keep medical records, which are among the most sensitive data about people. And many hospital's electronics help keep patients alive, monitoring vital signs, administering medications, and even breathing and pumping blood for those in the most critical conditions.

We can say, that anything that is plugged in, whether it has a Wi-Fi connection or not, can be vulnerable to hacking, and lots of medical devices, such as pacemakers and ventilators, are connected to the internet for the benefit of the patients. Pacemakers can connect with a device at home that monitors the rhythms of the heart and are able to send that information to doctors.

Hospitals also have a wide range of support systems and different analytical systems, such as Watson or other analytics systems. In addition, hospitals are increasingly using robots, for example, to dispense medicines, etc. It is very important to protect these systems from external security breaches, data thefts, ransomware attacks, different security attacks, or even cyber-attacks against.

### 6.1 Solution model for smart devices

Because there are lot of security challenges in security solutions, privacy and cyber security issues in now days healthcare systems, new type of smart devices was tested, in which are new type of security solutions. In the

## **Aarne Hummelholm**

hospital patient room patient's and care staff's intelligent devices work together (D2D) and exchange information directly in real time without any other network. Smart devices can work together forming their network, to go outside, and come back again seamlessly. The device prototype works. It tested in the laboratory and on the field (Aarne Hummelholm/2013). The device security system prevents unauthorized persons accessing to the device and the data transmission and the services provided, regardless of whether the patient is traveling, at home or elsewhere outside the hospital, and so on. We must meet the requirements of what the EU-GDPR and EU-MDR directives say. IoT -devices and sensors with platforms in smart devices are not yet tested in security issues (Hanna-Leena Huttunen (2017, 2018)). Research project, Smart medical devices, will be launched soon and it takes into account also EU-GDPR, -MDR and -NIS issues.

### **6.2 Future work**

Artificial intelligence (AI) use needs to be investigated and tested for its ability to protect e-Health's IoT devices, sensors and other health systems so that we can better protect these devices against these malicious software and cyber-attacks.

Because in health care devices are a lot of vulnerabilities and security challenges there, we need to find good architectures for healthcare equipment and systems, to give the requirements to them so that patients and the treatment staff can use them safely in this medical care environment (EU-GDPR/2016, EU-MDR/2017).

One research topic is to measure and test the frequency disturbances in the hospital and patient's home environment and check if there are any possible cases that affect the patient's treatment.

Energy efficiencies is very important research area to investigate it in the health care environment and the health care smart devices.

### **References**

- ABC News, Fears of hackers targeting US hospitals, medical devices for cyber-attacks, Jun 29, 2017.  
Aurore LE BRIS, Walid EL ASRI, State of cybersecurity & cyber threats in healthcare organizations, Essec Business school, 2016.  
Russell Brandom, UK hospitals hit with massive ransomware attack, May 12, 2017  
EU- Energy Efficiency Directive, 2012/27.  
EU-GDPR, The General Data Protection Regulation, 2016/679.  
EU- MDR, The Medical Devices Regulation, 5/2017.  
EU- NIS, Concerning measures for a high common level of security of network and information systems across the Union, 6/2016.  
Ian Armas Foster, NIST's Security Reference Architecture for the Cloud-First Initiative, June 28, 2013  
Aarne Hummelholm, Cyber threat analysis in Smart City environments, ECCWS2018, Oslo, 2018.  
Aarne Hummelholm, Kari Innala, Patent NO.: US 8,606,320 B2, (45) INTELLIGENT BASE STATION, Date of Patent: Dec. 10/2013, PCT Filed: Oct. 21, 2005.  
Huttunen, H. L., Halonen, R., & Koskimäki, H. (2017, September), Exploring use of wearable sensors to identify early symptoms of migraine attack.  
Huttunen, H. L., & Halonen, R. (2018, September), Preferred Biosignals to Predict Migraine Attack.  
Huttunen, H. L., & Halonen, R. (2018), Willingness to Use Smartphone Application Assistant to Support Migraine Treatment.  
Robert S. H. Istepanian, Bryan Woodward, m-Health, Fundamentals and Applications, Wiley, 2017  
ITU-T, Security in Telecommunications and Information Technology, September 2015  
Samant Khajuria, Lene Sorensen, Knud Erik Skouby, Cybersecurity and Privacy Bridging Gap, River Publishers, 2017  
Ramjee Prasad, 5G Outlook, Innovations and Applications, River Publishers, 2016  
Wang, P. Liu, J.C. Threat Analysis of Cyber- attacks with Attack Tree +, 2014

# Undersea Optical Cable Network and Cyber Threats

Aarne Hummelholm

Faculty of Information Technology, University of Jyväskylä, Finland

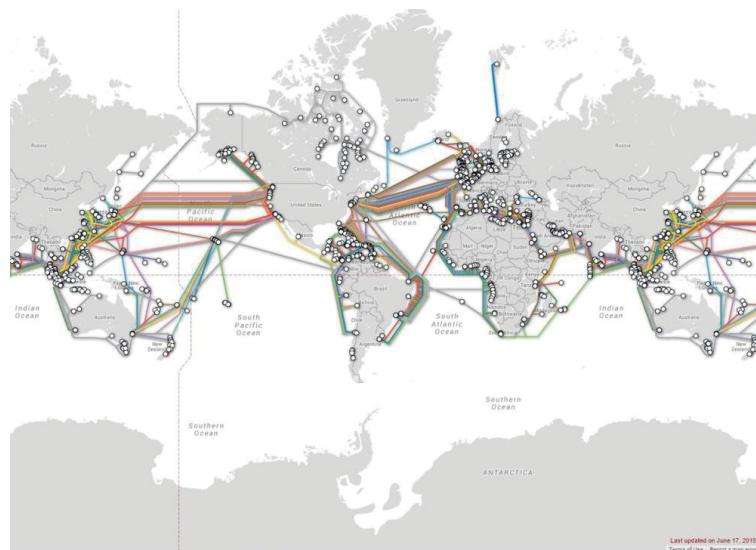
[aarne.hummelholm@elisanet.fi](mailto:aarne.hummelholm@elisanet.fi)

**Abstract:** Almost all services and most of the traditional services are totally dependent on the digital environment. Few users are aware of the revolutionary nature of modern technology. We use day-to-day real-time access to existing digital services in our home country or we use social media (Some) to communicate with friends locally or elsewhere in the world. We can communicate with them in real time with text messages or even through real-time video feed. People have the choice of millions of movies to watch anytime, anywhere. Modern communications connect data centers and data networks of different continents together, enabling real-time communications throughout the world. We can order different goods from all over the world, pay invoices electronically and get the goods delivered to our door. Companies use the same channels of communication for daily communications, trading, sending invitations to tender and transferring money through banks in real time. As a result of the developments described above, people and systems produce huge amounts of data which needs to be processed and stored. However, technical solutions for all new service environments are not yet in line with international standards and their connections to telecommunications and service networks are very diverse. Technically outdated solutions and new technologies are used simultaneously. Future information and communication systems need to be designed and adapted to work in this challenging business environment where security threats and cybercrime are constantly present. Each function has its own service and communication needs depending on the user group. These groups include design and maintenance staff, financial management staff, telecom operators, service provider staff, virtual service providers and operators, administrative agents, citizens, manufacturers, banks, etc. To date no other technology apart from submarine cables systems has had such a strategic impact to our society while at the same time remaining so badly understood by the general population. This means that it is also a very tempting target for hackers and state actors. They seek access to the sea cables and networks connecting continents to each other.

**Keywords:** communication, continents, cybercrime, submarine cables, hackers

## 1. Introduction

We will look first at how different parts of the world are currently connected to each other by submarine optical cables, figure 1.



**Figure 1:** Map of the Worldwide Undersea Submarine Cable Network (Reddit (2017)).

These cables are concentrated in the southernmost seas of the globe, and the terminals for submarine optical cables are located in areas where it is relatively easy to build cable endpoints, including cable communication and energy systems. Each country has its own fiber network that connects cities and the countryside to each other.

Currently submarine optical cable route between Asia and Europe is long. It is very possible that one way or another an undersea submarine optical cable system will break down. Terrorists can deliberately damage them, cyber attackers can penetrate them, natural disasters are a risk and so on. These and many other risk factors are

a good reason for us to design a new submarine optical cable route between Asia and Europe in order to ensure secure communication links between these areas. Figure 2 is a general overview of the Arctic Optical Cable Systems, which combines different regions of Europe, the western parts of Russia, the Siberian Russian regions, areas in the Russian Far East, smart cities in Japan and the border areas of China.



**Figure 2:** The Arctic connect cable system (Jukka-Pekka Joensuu (2018)).

A lot of communication capacity and many new contact points will be needed in the future in order to satisfy the data transfer needs of users, businesses organizations and governments in these areas. These areas require proper and reliable communication links to be able to communicate and use the services offered by the rest of the world. Regional development and joint operations in these northern areas require these links. This new connection gives people in the area real-time access to existing digital services in their home country and they can use it to communicate with their friends either locally or anywhere else in the world.

## 2. Chapter 1: Objective and organization of the paper

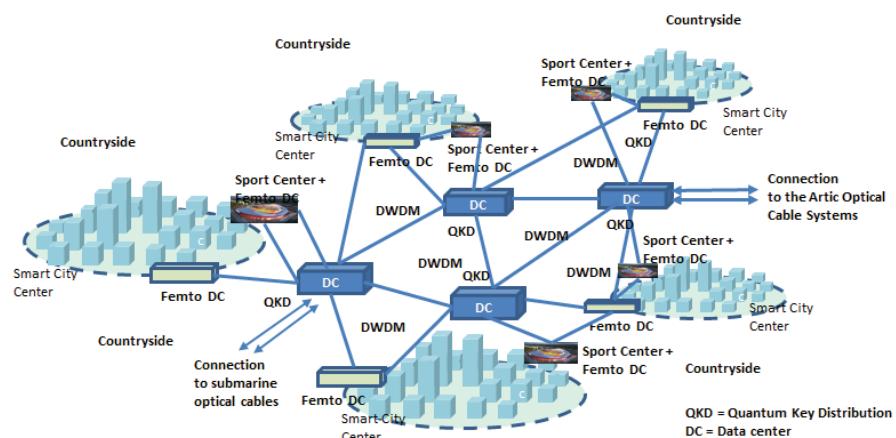
The research question is: whether is it possible to rely on submarine optical cable systems for communication between different continents?

This study seeks a model to facilitate threat assessments and the comparison of threats and to facilitate the threats analysis in these types of ecosystems. In view of the fact that the plan is focused on ocean environments, particularly in the Arctic, in addition to the technical design criteria, also different types of threats such as natural threats, accidents, as well as terrorists and cyber-attack threats must be taken into account. The results obtained through the model will aim to facilitate the design and implementation of architectural solutions. The implemented model can help in improving the assessing the threat scenarios for future submarine optical cables system environments and their impact probabilities. The study utilizes the operating environment as introduced in Figure 2, where different parts of continental services and infrastructure are grouped together. Threat research can be done by submarine optical cable in different segments and assess the types of threats to those segments. The political and commercial aspects of the Arctic region have not been included. The continent's telecommunication networks, submarine optical communications networks and data centers with their services are effectively one entity, through which all future services will be implemented; and, in the future, they will also be working together both in the different continents and between them. The integration described above is accelerating at all levels of activity, in each region and all segments both horizontally and vertically. The nature of these environments will further complicate doing threat estimates of this kind of ecosystems.

Chapter 2 presents the ecosystems and future operating environments of the telecommunication networks, data centers and the submarine optical communications networks at a general level. Chapter 3 describes the natural threats, accidental threats, cyber-threats and dependency analyses which are used to make a threat analysis of the submarine optical communications networks and network infrastructures and the services it provides. Chapter 4 deals with the making and modelling of threat analyses, and Chapter 5 describes the conclusions, solution model for security and future work.

### 3. Chapter 2: Description of the future operating environment and technology

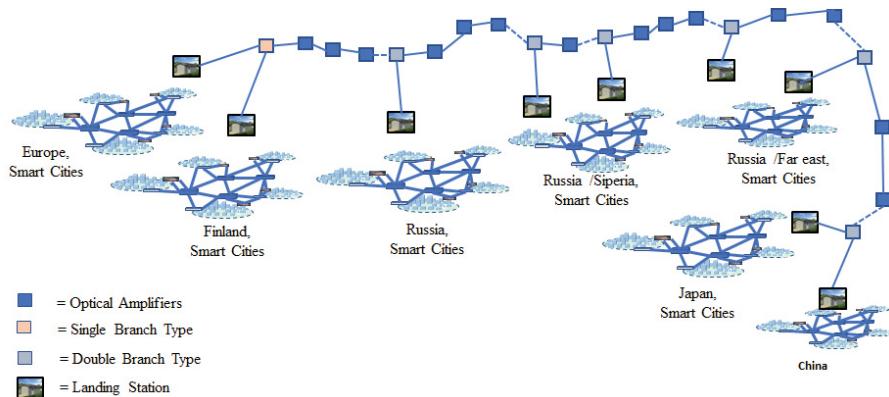
In the smart societies of the future, the amount of information will increase exponentially, as people use their smart devices not only to send messages but also to send real-time videos, to watch movies, and so on. Everything will be done in real time. Therefore, a very large amount of storage space will be required, which in turn means more data centers. Retrieving data from different networks also involves high real-time requirements, so data centers will also need to be as close as possible to the users. Figure 3 shows the smart societies of the future; a high-level architectural description of the smart societies, with large and small data centers, close to the users. Similar structures will be developed in both Europe and in Asia. These European and Asian Smart societies of the future will be connected to each other by a new route, when the Arctic Optical Cable System is installed. The estimated length of the incoming connection length of the Arctic Optical Cable's route is about 18.000-20.000 km (Figure 2). This sets a number of requirements both for the design and the implementation of this submarine optical cable system. Overview of the Arctic Connect cable system are seen in Figure 4, which was examined.



**Figure 3:** Smart Cities in the future environment, top level principle.

The Arctic optical cable route will also be of interest to hackers, terrorists, cyber attackers and state actors, as there is going to be a lot of information crossing from Europe to Asia and vice versa.

Because of the importance and the length of the link, the technical values and requirements that affect the design of the system must be investigated and the result utilized in both the design and the maintenance of the system. When we perform system deployment measurements, those values can be used later to detect the smallest changes in the system and identify any attempts attacking the system. These problems or defects may be caused by natural forces, construction works at sea, terrorists or cyber attackers. All such phenomena cause lesser or greater changes in the measured values associated with the operation of the system. Unfortunately, not all situations can be obtained from identifiable numerical data that can be detected by management and control devices, for example, the tapping of submarine optical cable results in less than a 2% loss of optical signal power and it is thus not so easily correctly detected by using existing technologies. Consequently, even the smallest deviations must be reviewed and analysed. For this reason, we must identify the factors affecting fiber quality that result from the properties of the fibres themselves. This is important to know and take into consideration because it affects the construction of a connection in a number of ways, such as the wavelength and bandwidths supported by the fibres, distance between optical amplifiers from each other, the optical signal levels to be used, etc. As this is also a long-term investment and the submarine optical cable may be in use for more than 25 years, the evolution of technology must also be taken into consideration, in order to anticipate early and timely updates and changes to the systems and equipment. In addition, when transmission speeds are increased in the wavelength by the fibres, there will still be a few boundaries that we can no longer exceed with the current technology. These are the physical limits of single-mode fiber (SMF) and two other limits of optical communications – the Fiber-launched power limit and the Non-linear Shannon limit (Yutaka Miyamoto (2017)).



**Figure 4:** Overview of the Artic connect cable system.

#### 4. The evolution of technology

As the life cycle of the optical cable system we must understand at least in part, the technical evolution that will occurs in optical telecommunication technology and how that will affect these long connections. The optical channel capacity cannot be increased indefinitely, despite the wide optical bandwidth available in the optical range. We can calculate the optical channel capacity that will be able to be achieved (Chesnoy Jose (2016)). We can look at Shannon's definition of the upper limit for transmission capacity (or spectrum efficiency) that can be transported in transport channel as  $C = B \log_2 (1 + SNR)$ , with  $C$  being the capacity in bit/s,  $B$  the bandwidth in Hz, and  $SNR$  being the signal-to noise- ratio. We can also define the single-sided optical noise power spectral density as  $S_n = h\nu/2$ , where  $h = 6.63 \times 10^{-34}$  Js is the Planck value,  $\nu$  is the proton energy  $10^{19}$ J, and the minimum average optical noise power  $P_N = (h\nu/2) B_0$  is proportional to the bandwidth.

Therefore, the optical channel capacity, treated in terms of the optical field, is:

$$C = B_0 \log_2 (1 + 2P_s / h\nu B_0) \quad (1)$$

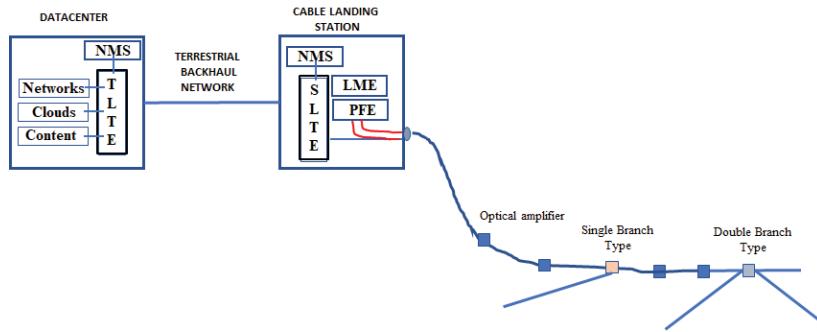
We must also take into account the features of these optical submarine optical cables systems parameters in practical networks to get more accurate information about the functions of the existing system so that we can detect possible intrusion attempts. Since more capacity is needed per fiber pair, new optical signal band, L-band is introduced. The C and L bands form the basic band of future long-distance optical networks ITU-T Manual (2009)). This provides a challenge to network designers to find an optical amplifier with enough bandwidth. In C-band, there are 80 optical channel and also in L-band has 80 optical channels. If we transmit 100 Gbit/s through one wavelength, this means, for examples, that we would have  $80 \times 100$  Gbit/s capacity in use in this kind of network.

#### 5. Long distance submarine optical systems

By 2015 there was a mature product had the capacity of 100 Gbit/s per optical wavelength. Since then ongoing development has continued to find new solutions aimed at increasing wavelength capacity per optical wavelength. As a result of this development, capacities of 200 Mbit/s and 400 Mbit/s are now available. In order to obtain transmission rates 100 Mbit/s or more in submarine optical cables, it would be necessary to install optical amplifiers at about 50 km intervals. This distribution would provide sufficient quality of service across continents.

#### 6. The primary principles of the installation

18.000 km long submarine optical cable system will be installed In the Artic region (Figure 2). There will be few branching points with connections to the continent (figure4). In Figure 5 Data Center NMS means Network Management System, and TLTE means Terrestrial Line Terminal Equipment. Cable Landing Station NMS means Network Management System, LME means Line Monitoring Equipment, PFE means Power Feed Equipment, SLTE means Submarine Line Terminal Equipment. Each cable landing station will be built in the same way. Difference between them will be only how the submarine's optical cables can be brought to the station of course and it depends on the beach area.



**Figure 5:** Subsea Optical Cable System Architecture with Cable Landing Station and Data Center

The main reason for the distance between optical amplifiers is only 50 km is due to the dispersions and non-linear properties of optical cables, the characteristics of the optical amplifiers, noise levels and the characteristics of the fiber. As the modulation techniques of optical data transfer become more complex, we must design systems even more carefully. The distance between the amplifiers is optimized based on the usability and quality of the services. Performance and cost optimization are expressed in terms of cost efficiency €/bit/s. These whole systems also need electrical energy. Energy input to the system can be made of one or more earth points, taking the energy supply protection into account in case of a damage in cable systems.

## 7. Designing submarine optical cable systems

In order to achieve the required usability and quality requirements for very long optical undersea cable connections, the following factors must be taken into account.

Factors affecting fiber connections quality

### 7.1 Attenuation

Attenuation values vary between different wavelength bands and should be smallest in the 1550 nm band, where it is about 0.2 db/km or less.

### 7.2 Dispersions

There are three types of dispersions; Rayleigh Scattering, Chromatic Dispersion and Polarization Mode Dispersion (PMD), which we must take into account. Optical Signal-to-Noise Ratio, OSNR, must also be taken into account as it limits the distance between optical amplifiers.

### 7.3 Impact of non-linearity

Optical fibers also have nonlinear characteristics like; self-phase modulation (SPM), cross-phase modulation (XPM), four-wave mixing (FWM), stimulated Raman scattering (SRS), and stimulated Brillouin scattering (SBS). Those characteristics must be given special attention as the phenomena they cause may be the result of a cyber attacker's action that were not detected by the normal management and control systems of the submarine optical cables.

## 8. Chapter 3: Natural threats, accidental threats and cyber threats

In addition to the technical design criteria, we also need to take into account the various types of threats that will be encountered such as natural threats, accidental threats and cyber or malicious threats. These threats can contribute to prolonging cable routes or partial routes or even altering the originally planned routes. Natural threats include threats such as sharks, earthquakes, landslides, volcanic eruptions, tsunamis, icebergs, sea currents, storm winds and so on. Accidental threats can be caused by everyday work at sea, such as fishing, dragging an anchor, dredging etc. and can damage submarine optical cables, threatening their level of performance. We also need to look at other potential threats since the submarine optical cable routes are long. Many countries have the ability to join (tap) fiber optic cables in order to collect the information being transmitted there in. In every situation, we must always be on the lookout for opportunities to hack into or launch cyber-attacks on submarine optical cables – whether this be “tapping” the lines or other methods such as side channel attacks or side channel spying.

The type of undersea cables types chosen depends on the depth of the sea and the vicinity of the coast in the areas where the above-mentioned threats exist, and the threats are realized. If we consider at cyber attackers' opportunities to join to the optical cable, it would be easiest to penetrate the cable exactly where the cable armoring layers are thinnest. This also means that the attacker must be able to operate deep below the sea level. In practice, only a few large states have the capabilities to do this. When considering this planned undersea optical cable system, with a length of 18,000 km, cyber attackers will be able to connect to undersea optical cables after each optical amplifier, which are deep underwater area. Next, we can consider a submarine optical cable system, based on ITU-T Recommendation G.709, G.971 and G 977. It is divided into a land section, an underwater section, and a second part of the land section. The underwater section has the necessary branching equipment. When we have a 18,000 km long submarine optical cable system in use, we also need Power Feed Equipment for our submarine optical cable systems.

We also need different types of OTDR (Optical Time Domain Reflectometers) to certify the performance of new fiber optics links and detect problems in existing fiber links. It is now possible to use measurement system like COTDR, Coherent Optical Time Domain Reflectometry, for high capacity systems. It is very important to use COTDR, because in 18.000 km submarine optical cable systems we would use different types of fiber cables for compensating dispersion phenomena. This also means that there is very much cable branching and continued points, so it is very difficult to find the smallest deviations in the parameters. It's management and control systems are most critical systems. To identify security challenges in long distance optical cable systems, we must consider Optical Transport Network (OTN) framing and rates, Figure 6, and Optical Transport Network (OTN), OSI layer model, Figure7. Currently, no encryption technology is in use for optical signals. Figure 6 shows the OTN Optical Transport Network (ITU-T, G.709), where the client signal is seen and how the header areas of the different layers are placed in relation to the client signal. Figure 7 can be seen more precisely portraying that what information a cyber attacker could access and use if we do not encrypt these signals. For example, they can change the ROADM, the Reconfigurable Optical Add-Drop Multiplexer, routing in whatever way they want, and either disrupt traffic or drive traffic to a desired connection point, for analysis.

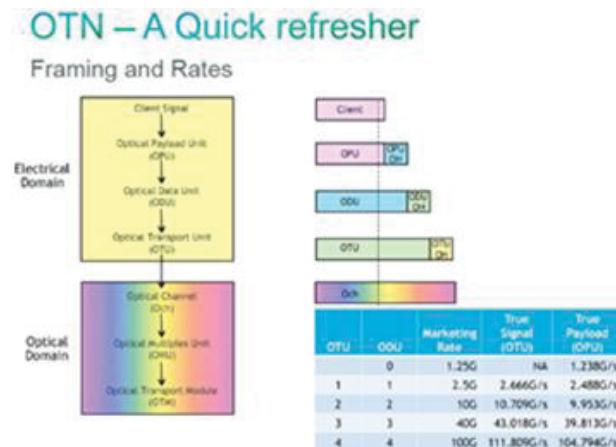


Figure 6: OTN Optical Transport Network, framing and rates, (ITU-T, G.709).

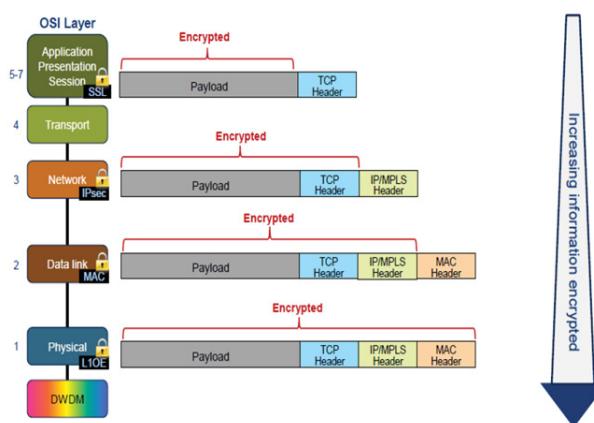


Figure 7: OTN Optical Transport Network, OSI layer and encryption (ITU-T, G.709).

From Table 1 we can see upper level conceptual submarine cable segments` threat matrix which we need to take it account when we are designing and developing undersea submarine optical cables systems.

**Table 1:** Upper level conceptual threat matrix for submarine cable segment, on “Threats to Undersea Cable Communications, September 28, 2017”.

Submarine Cable Segment Threat	Land and Beach Area (Seg.1)	Near Shore Area ~50 m (Seg.2)	Off Shore Area ~ 50 – 100 m (Seg.3)	Continental Shelf ~ 100 – 200 m (Seg.4)	Deep Sea ~ 200 m + (Seg.5)
Natural Threats					
Sharks					
Earthquake					
Landslide					
Volcano					
Tsunami					
Iceberg					
Ocean currents					
Accidental Threats					
Fishing					
Anchor dragging					
Dredging					
Malicious and undersea warfare					
Cyber Attacks					
Vandalism					
Activists					
Theft					
Terrorist					
State-actors					
Undersea warfare					

Threat impact level depicted in colours: Green = Low; Yellow = Medium; Red = High

## 9. Chapter 4: The making and modelling of a threat analysis.

Table 1 illustrates the upper level conceptual threat matrix for submarine cable segments, based on threats to submarine cable communications. We should also note that cyber attackers, hackers and terrorists can use artificial intelligence to enable them to search from vulnerabilities in submarine optical cable systems through which they can penetrate the systems and its services. After doing so they should have the ability to attack Data Centers, different continents. There are many ways that cyber attackers can get inside a submarine optical cables system and to gain access its managements and control systems.

In figure 8 is presented threat probability tree model in the Artic connect cable system, which is used in threat model. Table 1 is divided according to the depth of the submarine optical cable system into different segments and those segments are still divided into different types of categories of threats. Probability of a threat in every segment we can calculate it threat based on information we get from international research reports, from The European Space Agency (ESA), from the Arctic statistics, from sensors and sonars, news concerning on natural or animal cases, accident or injury cases, cyber-attacks etc. how many times they occur and in what areas and at what time of year. This threat probability calculation can be done for the full length of the cable system or just a part of the cable system. In situational picture we need also information from the power supply station's status.

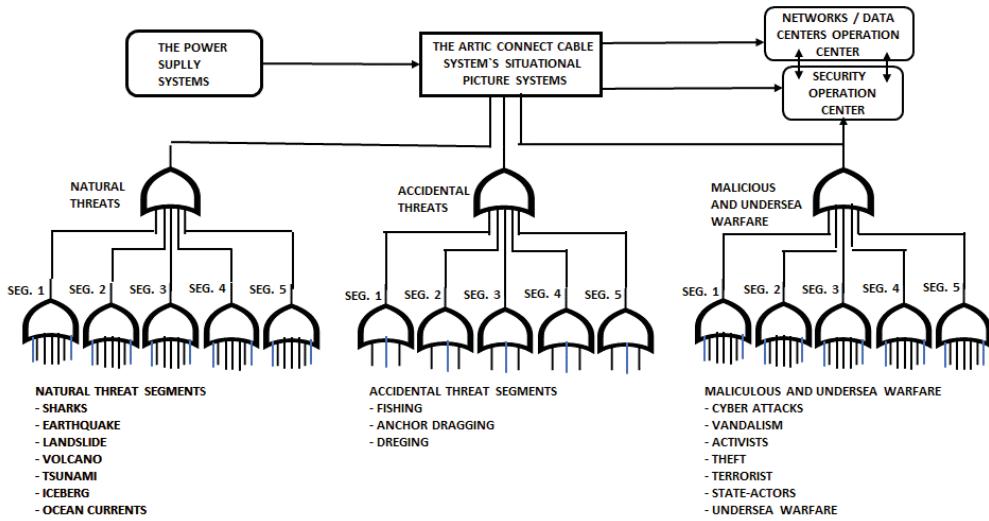


Figure 8: Threat tree model for the Arctic connect cable system, example.

Table 2: Meaning of notations

Action	Examples	Notation
Threats or attack	Sudden event, Accident, Tapping, Eavesdropping, Sniffing, Scanning, ...,	A
Detection	Alarm information, systems management information, international information, ...,	D
Countermeasure	Analysing of threats and vulnerabilities and to repairing, safeguards put in place, ...,	M

Threats ( $P(t)$ ), probabilistic treats or attacks happen.

$$P_{1S1\dots7}(t) = P_{1A1\dots7}(t)(1 - p_{1D1\dots7}(t))(1 - p_{1M1\dots7}(t)), \text{ to 7 different types natural threats.} \quad (1)$$

$$P_{2S1\dots3}(t) = P_{2A1\dots3}(t)(1 - p_{2D1\dots3}(t))(1 - p_{2M1\dots3}(t)), \text{ to 3 different types accidental threats.} \quad (2)$$

$$P_{3S1\dots7}(t) = P_{3A1\dots7}(t)(1 - p_{3D1\dots7}(t))(1 - p_{3M1\dots7}(t)), \text{ to 7 different types malicious and undersea warfare.} \quad (3)$$

$$P_{1S1\dots7}(t) = [(P_{1S1}(t)) + (P_{1S2}(t)) + \dots + (P_{1S7}(t))], \text{ information to the situational picture systems.} \quad (4)$$

$$P_{2S1\dots3}(t) = [(P_{2S1}(t)) + (P_{2S2}(t)) + (P_{2S3}(t))], \text{ information to the situational picture systems.} \quad (5)$$

$$P_{3S1\dots7}(t) = [(P_{3S1}(t)) + (P_{3S2}(t)) + \dots + (P_{3S7}(t))], \text{ information to the situational picture systems.} \quad (6)$$

In figure 8, the situational picture system is for every threat-type own icon, which is telling situation in that segment in the Arctic connect cable system in every part of it. Information from situational system is also send to the security operation center and also network management centers and data center's management systems. That situational picture systems should be in different areas of the Arctic connect cable system for network operators and for service provider their own, because of response time must be fast enough to start for example, in rescue operations. The Arctic connect cable system whole situational picture information must be also in the cable operator's operation center.

The land and beach areas of submarine optical cables systems are the easiest for attackers to penetrate. When using large capacity systems in undersea environment, and new types of modulation technology in those systems, the best possible cable tapping points for cyber attackers are after every optical amplifier in deep underwater area. This offers them various opportunities to obtain large amount of information from different companies, organizations and governments. In this situation cyber-attackers and hackers can obtain IP addresses from these companies', organizations' and governments' and either make DDOS (Distributed Denial of Service attack) attacks against them or use different types ransomware or malware attacks against them. Ransomware attacks are typically carried out by using a Trojan, entering a system through, for example, a vulnerability in a

network service. One possible cyber-attack model is an advanced persistent threat (APT), which is a targeted cyberattack in which an intruder gains access to a network and remains there undetected for a long time. APT attacks typically target organizations such as national defense, manufacturing, and the financial industry, and also companies that deal with high-value information, military plans, and other data from governments and enterprise organizations. The intention of an APT attack is usually to monitor network activity and steal data rather than to cause damage to the network or organization.

Figures 6 and 7 illustrate, if attackers gain access to the submarine optical cable system and there is not encryption system in use, they will also have access to the its management system and thus have the ability to use it to do whatever they want and what suits their purposes. We also need to take into account the power supply system, so that we can be certain that it does not have any vulnerabilities that an attacker can take advantage of in order to attack to our systems.

Considering Figures 3 and 4, Communications Networks in The Future, between different smart cities, we must also look at communications inside the cities, where there are many challenges, presented by the operating environment and heterogeneous telecommunication networks. New devices and systems are seamlessly interconnected there. These systems have expanded into homes, building automation systems, cars and various control and energy systems and people are now using their personal smart devices everywhere. These smart city systems also need their own applications, and their information is stored in a Data Center, shown in Figure 3 and 4. This also means that hackers, terrorists and cyber attackers have many opportunities to find vulnerabilities in this environment, and to attack these Smart City's applications and services. This in turn allows hackers and cyber attackers to attack services and service systems, even on other continents because the Data Center are interconnected.

## **10. Chapter 5: Conclusions, solution model for security and future work**

The system to be built is technically very complicated and will needed many new technical solutions to meet the required transmission rates and usability and quality parameters. This places considerable demands on the management and control of the system as well as on the organization of its maintenance. Changes in social structures take place very quickly and will also affect the implementations and operating models and structures, as well as people's everyday lives and working environments. The current powerful digitalization trend increases the range of services offered and facilitates their easier use. These developments also have a strong impact on the service chains of the provided services, including subcontractors and their subcontracting chains, hardware solutions, service providers and operating models for every part of the service chain on every continent.

Today and in the future modern communications connect data centers and data networks on different continents, enabling real-time communication throughout the world. This type of communications is made possible by undersea optical cable systems, which we use for daily communications. Because submarine cables systems have had such a large strategic impact to our society, they are also a very interesting target for hackers, cyber attackers, terrorist and state actors. They seek to gain access to the information that goes through the networks of these continents which are connected to each other with sea cables. For example, we need to be aware of the possibility of cyber attackers being able to connect to optical fibers, they have the option to change the ROADM routes, which can lead to the communication or disruption of traffic between the entire continent. When considering the cyber security in systems design, we must take into account the upcoming technologies, which means there are more challenges ahead of us. In addition, changes in the cable technology due to dispersion phenomena make their own challenges in detecting intrusion into the cable. We have to be really careful about the design.

## **11. A solution model for security**

Since this new submarine optical cable system (Figure 2), is so long and it is impossible to detect or identify all of the potential attacks against it, it is recommended that an end-to-end encryption system on each wavelength individually at the lowest layer, be put in place. Figure 7 illustrates why this is advisable, as we can see the optical traffic network, the OSI layer and what effect encryption this layer have. When we implemented encryption on the lowest layer, we protect our entire communications systems against various types of attacks. There are currently encryption systems of this type in use, but the capacity to be used may present challenges to those devices and systems. The Quantum encryption system is also currently operational use however such an environment would present challenges regarding the renewal of encryption keys in each optical amplifier.

Individual smart devices are also tested with different types of VPN-encryption's concepts, but that research is still ongoing although it is hoped that the results will be ready by next year.

## **12. Future work**

Because the Arctic connect cable system is a critical system that will be used by many countries, organizations and people for their own purposes, it is essential to study the key issues affecting its functioning. With regard to cyber security, the use of Artificial Intelligence (AI) needs to be investigated and its potential to protect submarine optical cable systems needs to be clarified in order to better protect against malware and cyber-attack.

- With regard to cyber security, the use of Artificial Intelligence (AI) use needs to be investigated, and its potential to protect submarine optical cable systems needs to be clarified in order to better protect against malware and cyber-attacks.
- The possible use of COTDR should be investigated as it is used for searching for faults and can also be used to detect the tapping of cable connections.
- We must study a variety of protection mechanism for submarine optical cables system's because it is an extremely important fiber optic connection between different continents.
- One study area would be different encryption systems, such as quantum encryption or Layer 1 - 2 encryption systems.

## **References**

- Chesney Jose, Undersea Fiber Communication Systems, Elsevier Ltd, 2016.  
Governance for Cyber Security and Resilience in the Arctic, Advanced Research Workshop 27 – 30 January 2019, Rovaniemi, Finland.  
Davenport Tara, Submarine Cables, Cybersecurity and International Law: An Intersectional Analysis, 24Cath. U. J. L. & Tech (2015). Available at: <http://scholarship.law.edu/jlt/vol24/iss1/4..>  
ITU-T, Spectral grids for WDM applications: DWDM frequency grid, G.977 (01/2015)  
ITU-T, G.709/Y.1331, Interfaces for the optical transport network, 6/2016.  
ITU-T, G.971, General features of optical fibre submarine cable systems, 11/2016  
ITU-T, Manual 2009, Optical fibres, cables and systems.  
Jukka-Pekka Joensuu, Navigating the Arctic, 13 th February 2018, <http://asia.blog.terrapinn.com/submarine-networks/2018/02/13/navigating-the-arctic/>.>  
Reddit, Map Of Underwater Cables That Supply The Worlds Internet, 30 September 2017,  
[www.reddit.com/r/MapPorn/comments/73ekox/map\\_of\\_underwater\\_cables\\_that\\_supply\\_the\\_worlds/](http://www.reddit.com/r/MapPorn/comments/73ekox/map_of_underwater_cables_that_supply_the_worlds/)  
Threats to Undersea Cable Communications, September 28, 2017, PUBLIC-PRIVATE ANALYTIC EXCHANGE PROGRAM.  
Ye Yincan, Jiang Xinmin, Pan Guofu, Jiang Wei, Submarine Optical Cable Engineering, Elcevier Inc. 2018.  
Yutaka Miyamoto, Ryutaro Kawaura, Space Division Multiplexing Optical Transmission Technology to support the Evolution of High-capacity Optical Transport Network, June 2017.

# The Current State of Research in Offensive Cyberspace Operations

Gazmend Huskaj

Department of Military Studies, Swedish Defence University, Stockholm, Sweden

[gazmend.huskaj@fhs.se](mailto:gazmend.huskaj@fhs.se)

**Abstract:** Cyber-attacks have increased since the 1988-Morris worm and can target any connected device from any place in the world. In 2010, Stuxnet received a lot of attention as the first cyber-weapon. Its targets were the Iranian nuclear enrichment centrifuges. Nation states are developing cyberspace capabilities to conduct offensive cyberspace operations. Academic researchers have been calling for a more transparent discussion on offensive capabilities and have pointed out the positive impact researchers had during the development of nuclear capabilities. Shrouded in secrecy, the development of offensive capabilities used for operations makes it difficult to conduct research. Therefore, one way to mitigate this is to conduct a systematic review of the current state of research in offensive cyberspace operations. The systematic review method makes it possible to establish certain inclusion and exclusion criteria and systematically go through academic articles to identify the contents, thoughts and research focus of academic researchers. Six scientific databases were queried and 87 articles were read and clustered. The first insight is that, based on the results of the queried databases, research about offensive cyberspace operations is limited. The resulting clusters are a general cluster about cyberspace operations, followed by research in policy, decision-making, governance, capabilities, levels, models, training, deterrence and international affairs. These are then further grouped into: a) general cyberspace operations; b) deterrence; c) international affairs; d) modelling, simulation and training. The article concludes that research into offensive cyberspace operations is maturing as more information is becoming public. Secondly, current research lists some good basic ideas regarding effects which can be achieved through offensive cyberspace operations, how they should be conducted, and related tools, techniques and procedures. However, discrepancies in research efforts exist, with the majority of research coming primarily from the western world. In addition, secrecy and the resulting limited access to information, coupled with research being either too technically focused or too qualitatively focused, show that there still remains room for research in this field. Finally, some directions for future research are examined.

**Keywords:** research in offensive cyberspace operations, cyberspace operations, decision-making, systematic literature review

---

## 1. Introduction

Cyber-attacks have increased since the 1988-Morris-worm (Spafford, 1988) and can target any connected device from any place in the world. Cyber-attacks are defined as actions affecting the confidentiality, integrity and availability of information in information systems. In 2010, Stuxnet was dubbed as the first cyber-weapon, targeting Iranian nuclear enrichment centrifuges (Stark, 2011). “Targeting is the process of selecting and prioritizing targets” (Joint Chiefs of Staff, 2018, p.IV-8). Nation states are developing capabilities to conduct offensive cyber operations. Researchers have been calling for a more transparent discussion on offensive capabilities, pointing out the positive impact researchers had during the development of nuclear capabilities.

The motivation for this article is twofold; firstly, to give readers a description of the current state of research in offensive operations. Secondly, to show where this research is focused. In this article, cyberspace is defined as internetworked information systems. Peterson and Davie (2012) define internetworked as ‘the concept of interconnecting different types of networks to build a large, global network.’ Valacich and Schneider (2010), and Laudon and Traver (2011) have defined information systems (the reader is advised to those authors for the full definitions). Therefore, the long definition of internetworked information systems (i.e. cyberspace) is “hardware and software which are used to create, process, store, retrieve and disseminate information in different types of interconnected networks that build a large, global network, built and used by people.” This definition is grounded and tangible, and there is no better place to provide it than in a literature review. The scope of the article is defined by the inclusion and exclusion criteria.

This article is organised around two research questions:

- (RQ1) What is the current state of research in offensive cyber operations?
- (RQ2) Where is that research focused?

This work contributes to PhD-students who want to explore the topic and decide on a research direction. It also contributes to active researchers and practitioners who want to get insights on the current thoughts and ideas in the area.

The first section describes the method. The second section presents the results followed by discussion and conclusions.

## 2. Method

Major contributions to scientific research are found on databases (Webster and Watson, 2002). The search used the search query “offensive cyber operations.” Searching terms enclosed in quotation marks causes search engines to only present articles containing those exact phrases (Blachman & Peek, 2012). The initial search results were:

Source	Description and Implementation	Results
IEEE Xplore	The database “is the world’s largest technical professional organization dedicated to advancing technology for the benefit of humanity.” The database holds 4,338,558 items at the time of writing.	Four
ScienceDirect	The database is “Elsevier’s leading platform of peer-reviewed scholarly literature” and holds “over 14 million peer-reviewed publications.”	14
Scopus	The database is “the largest abstract and citation database of peer-reviewed literature.”	19
SpringerLink	The database holds “over 10 million scientific documents.” Of these, only 15 were accessible by the author. One of the articles was in German.	54
Web of Science	The database is “your ideal single research destination to explore the citation universe across subjects and around the world.”	9
WorldCat	The database is “the world’s largest library catalog.”	80

Firstly, articles were screened based on their titles and abstracts. Permutations of the articles were removed. The remaining articles were skimmed through their titles and abstracts to evaluate their relevance and quality based on the inclusion and exclusion criteria. The inclusion criteria were published articles, books or chapters in books, conference papers or MSc-theses. The exclusion criteria were articles published prior to year 2000 and non-English articles. Only one article was non-English and is unlikely to skew the results. Next, they were clustered manually into the following research areas:

- General cluster about cyberspace operations
- Policy
- Decision-making
- Governance
- Capabilities
- Levels
- Models
- Training
- Deterrence, and
- International Studies

The author established a review form covering the main argument, key concepts/assumptions, results and cluster/category, and a free text field for major points to make in the discussion. The review form was constructed iteratively.

## 3. Results

This section presents the results of the reviewed articles and some quantitative measures in Table 1. It should be noted that the categorisations are not mutually exclusive, e.g. some articles cover multiple research fields. In addition, even though not all of the 87 articles reviewed are cited below, a reasonably representative sample of articles is included in each category.

**Table 1:** Characteristics of the 87 articles. Note that the categorizations used are not mutually exclusive

General Cyber Operations	49
Deterrence	16
International Affairs	10
Governance	2

Policy	13
Decision-making	5
Capabilities	9
Levels	7
Models & Simulation	7
Training	4

### **3.1 General overview of offensive operations**

This category covers the broader topics regarding offensive operations, followed by targets, vulnerabilities, tactics, techniques and procedures, and proxies.

Eilstrup-Sangiovanni (2017) describes how governments are investing in offensive capabilities: at least 29 governments have dedicated units, and at least 60 more are developing capabilities. Douglass (2012) and Bardin (2015) note which countries are developing offensive capabilities and Kshetri (2016) highlights North Korean offensive operations. Harrison & Herr (2016), Jajodia et al. (2016) and Ottis (2015) note that state and non-state actors, intelligence agencies and military organisations conduct offensive operations. Gompert and Binnendijk (2016) note that offensive operations have certain characteristics making them favourable compared to conventional forces, while Sutton (2013) and Torres (2012) describe the effects of offensive operations in the Georgia 2008-case. Dossi (2018) discusses that offensive operations are conducted through proxies, increasing the difficulty of attribution. Fernandes et al. (2018) and Bardin (2015) discuss US and Israeli investments, while Yeo et al. (2015) and Fernandes et al. (2013) discuss how the US Department of Defense (DoD) has increased the transparency of their capability. Buchanan (2017) compares levels of maturity between countries, and Lin (2010) describes a process for operations. Iasiello (2013) notes how China and Iran use offensive operations to monitor, censor and block information from reaching the public. Finally, Rochetto (2016) notes that terrorist groups lack the skills, tools, resources and determination to conduct offensive operations. Resources include encryption, exploits and staff (Sin et al., 2016).

This category describes targets. Caire (2018) discusses attacking critical infrastructure and transportation systems to bind or deny an adversary's resources. Hart and Klink (2017) argue that offensive operations can enable information operations, while Van der Velt (2017) argues that the Russians weaponised information to affect trust in the US elections. Kallberg and Thuraisingham (2013) identify information systems, networks, infrastructure and industries as possible targets, while their military counterparts are off-limits. Ormrod (2014) also argues for targeting command, control, communications, computers and intelligence, surveillance, and reconnaissance (C4ISR) networks. Porche, et al. (2017) argues for targeting unmanned aerial systems (UAS), adversary units and their devices in a mission area. Uren (2017), Nevill (2016) and Hawkins (2016) disclose that Daesh has been targeted by Australian offensive operations.

This category discusses the use of vulnerabilities for offensive operations. Sigholm and Larsson (2014) discuss exploiting the "Heartbleed Bug" for an offensive operation, while Charlet et al. (2017), Schwartz and Knake (2016) discuss the importance of the Vulnerability Equities Process (VEP), a policy on vulnerabilities. The VEP balances the national security interests of collecting intelligence from foreign nations and terrorist groups, and securing the infrastructure against threats (Schwartz & Knake, 2016). Charlet et al. (2017) note that the government notify software vendor(s) to create a patch once a vulnerability is used for national security purposes.

This category covers tactics, techniques and procedures (TTPs) for offensive operations. Eriksson and Pettersson (2017) compare offensive operations with special operations: small scale, requiring high secrecy and high operator skills, and with high political risk. Grant (2013) notes similarities with special operations in that they operate under government authority and comply with national and international law. Hurley (2017) argues that the nation state, as well as the private and public sectors, can conduct offensive operations. Offensive operations serve to defeat or destroy an enemy, while in the private sector they can "increase shareholder value and market share" (p.19).

This category discusses the use of proxies for offensive operations. Kallberg and Rowlen (2014) note that countries with low cyber-maturity are at risk of being used as proxies. This tactic decreases the risk of attribution to the attacking country but increases the risk for the proxy-nation. Bardin (2015) notes Iran using proxies such

as Hamas and Hezbollah. Borghard and Lonergan (2016) note how Russia used proxies in the 2008-Georgia-case. Forums were used to coordinate attacks targeting Georgian high-value targets. The use of proxies can also reduce costs because offensive capabilities require significant skills to develop and maintain access. In addition, it is easier to provide proxies with computers than weapons.

### **3.2 Deterrence**

This category discusses how offensive operations can increase a nation's deterrence posture. Fischerkeller (2017) notes that offensive operations can generate physical damage, influence adversarial actions, collect intelligence and be integrated into conventional operations to achieve new effects. Eilstrup-Sangiovanni (2017), Lonsdale (2017), and Mazanec (2015) note that a strong offensive capability can achieve enhanced credibility for strategic deterrence.

### **3.3 International affairs**

This category discusses the role of law. Lin (2017) argues that decision-makers will have a more comprehensive understanding of offensive operations if legislation requires military organisations to report to the government on these activities. Smyth (2014) states that offensive operations should adhere to the Law of Armed Conflict. Prescott (2012) cites the DoD policy whereby offensive operations are conducted in the same way as their kinetic capabilities, adhering to "policy principles and legal regimes, including the Law of Armed Conflict" (p.261). Cusumano and Corbe (2018) note that some nations employ illegal and inappropriate means, and misuse international law.

### **3.4 Governance**

Governance constitutes a small part of the literature. Douglass (2012) and Rodriguez (2011) discuss governance at government level and at operational/tactical level. Governance in the US is currently spread across a confusing myriad of competing authorities, crippling the capacity to operate offensively in cyberspace. Rodriguez (2011) argues that offensive operations at operational/tactical level fall within the area of responsibility of the geographic combatant commander.

### **3.5 Policy**

Research on policy in support of offensive operations is limited. A policy on offensive operations assists decision-makers, supports national interests and makes it possible to pursue political goals (Baltrusaits, 2017; Olagbemiro, 2014; Nikitakos & Mavropoulos, 2014). A policy can also clarify any potential future confrontation between nation states (Segal, 2016). It is difficult to promote norms, deter attacks and pursue political goals without a policy. (Johnson, 2014; Nikitakos and Mavropoulos, 2014). Buchanan (2017) discusses the positive impact of U.S. Presidential Policy Directive 20, including how it directs the community to plan, prepare and present cases where offensive operations may be used.

### **3.6 Decision-making**

In decision-making, Grant (2017), Prescott (2013) and Oltramari et al. (2013) discuss how to speed up decision-making processes. This is important due to the speed of cyber-attacks. Another option is to have automated decision-making processes (ADPs) assisting human decision-makers (Prescott, 2013; Oltramari et al, 2013). However, ADPs must have the principles of the Law of Armed Conflict and rules of engagement "incorporated into ADP design processes" (Prescott, 2013, p.3). Finally, Grant (2017) and Uren (2017) discuss the organisation of the offensive capability with related decision-making processes.

### **3.7 Capabilities**

Capabilities are required to achieve desired effects. Buchanan (2017) argues that cyber operations should be conducted early for the development of offensive capabilities. They may be used for operations in preparation of the environment and exploiting system vulnerabilities. This requires intelligence, surveillance and reconnaissance capabilities (Harrison and Herr, 2016). Capabilities in hostile networks should be covert, which Jajodia et al (2016) discuss. McArdle (2016) argues the Chinese and Russians prefer "soft" capabilities because of the level of deniability. Soft capabilities are those just below the threshold of triggering armed responses.

### **3.8 Strategic and tactical levels operations**

This category presents research on levels of operations. Röigas (2018), Lin and Zegart (2017), Lemieux (2015), Andress and Winterfield (cited by Lemieux, 2015), and Cartin (cited by Wilson and Drumhiller, 2015) discuss the strategic level of offensive operations. Lemieux (2015) states they are aligned with defensive and offensive operations; Andress and Winterfield (2015) argue defensive operations are within the dimensions of prevention and deterrence. Cartin (2015) argues the goal is to “influence the perception of one’s security” (p.34). Lin and Zegart (2017) state their focus should be on the long term, while Röigas (2018) notes the strategic role of offensive operations is still unclear. On the tactical level, Lin and Zegart (2017) state offensive operations are focused on a small area with short-term goals, while Lemieux (2015) notes tactical operations as “techniques and practices to secure or penetrate a computer network” (p.2).

### **3.9 Modelling and simulation**

Models can be effective tools to conduct offensive operations. Buchanan (2017) presents an eight-step model while Grant (2012) presents a five-step model. Rochetto (2016) compares a nation state adversary with an adversary group, noting that the group cannot put the same efforts into stealth due to a lack of skills and resources.

### **3.10 Training**

Training, education and certain traits are essential to the conduct of offensive operations. Aybar (2017) describes developing a virtual environment mimicking an adversary’s system to train personnel before deployment. Burke and van Heerden (2016) present a cyber challenge, rigged with sensors, to train operational behaviour. Schweizer et al. (2013) have a course for students on offensive operations. De Souza (2013) discusses certain traits, which personnel designated for offensive and defensive operations should have.

## **4. Discussion**

The current state of offensive cyber operations indicates further opportunities for research. This section discusses some observations while answering the research questions.

### **4.1 General cyber operations**

Governments are conducting offensive operations. Researchers have listed (in alphabetical order) China, Iran, Israel, North Korea, Russia, and the US as countries conducting offensive operations and increasingly investing in that capability. The researchers highlight the importance of capabilities to achieve desired effects, but none of them discusses the fact that capabilities are developed through software development cycles to exploit vulnerabilities in target systems. Ethical dilemmas regarding whether or not to release vulnerabilities exist. The literature highlights how the vulnerability equities process is a solution which has two positive effects. Firstly, it communicates a level of maturity on the authorities conducting offensive operations; secondly, it increases the level of trust between the private and public sectors. The private sector know they will receive information from the government when software flaws are identified. The intelligence services collect information about vulnerabilities in target systems, which may be used as a list of requirements to develop the capability. Developing capabilities may take time. This could be one variable to take into account when deciding whether the operation is at a strategic or tactical level. However, researchers disagree as to what strategic offensive operations are. There may be two reasons; firstly, cyber operations are not fully understood because their possibilities will be revealed as more devices become connected; secondly, because of the nature of offensive operations, as already mentioned above. However, regardless of level, tactics, techniques and procedures for offensive operations require more research. Current research on TTPs is focused mostly on comparing offensive operations with special operations. While this may be true, research explaining and describing TTPs for offensive operations is lacking, especially in the technical-policy area, i.e. describing the impact of technology on policy without using too technology-oriented language. Finally, the decision on whether to hit the target directly or go through proxies needs to be considered. Until now, the use of proxies has been an accepted way to conduct operations, and those most at risk are the countries used as proxies. Therefore, one interesting question is, what areas are off-limits for use as proxies?

## **4.2 Deterrence**

Offensive operations increase a nation's deterrence posture. To do this, offensive operations need to be credible and communicated through the media. Credibility is achieved by conducting operations and "leaking", i.e. releasing information in a controlled fashion, to the media. Intelligence can also be revealed jointly with allies. This may include information on the adversary, their targets, the indicators of compromise, and advice on how to patch any vulnerabilities the adversary has been exploiting.

## **4.3 International affairs**

Decision-makers have a hard time to grasp the abstract world of cyberspace and the affiliated political risk connected with offensive operations. Political risk is mitigated by adhering to international law, the Law of Armed Conflict, and the same policy principles as for kinetic capabilities. To make it tangible, offensive operations may be compared to Special Forces operations.

## **4.4 Policy, decision-making and governance**

The lack of research on policy in offensive operations makes it difficult for decision-makers to know who does what at national level. The only known model for policy in support of operations is PPD20, and is US-specific. A policy in support of offensive operations will likely increase cooperation between competing agencies. In addition, it is likely to have positive effects on deterrence; the level of uncertainty with which the country conducts retaliatory-operations is increased and would-be attackers may be deterred. In addition, it may also increase the speed of decision-making processes. These are required for retaliatory actions. Another way is to use automated decision-making processes for retaliatory actions. However, risks exist. Is the source really the target who attacked or was it a proxy? What are the risks of cascading effects? Will an automatic response lead to escalation? Could an adversary have tampered with the algorithms responsible for automated-decision-making processes? These are some of the issues that have to be considered. Finally, governance is important for effective offensive operations. Governance spread across competing authorities cripples the effectiveness of offensive operations and affects credibility of deterrence. Governance at the operational/tactical level is equally important, if not more so, to achieve the desired effects on adversary targets.

## **4.5 Modelling, simulation and training**

Modelling, simulation and training are important tools to plan, prepare and train for offensive operations. The results show that models differ on the number and order of steps. It may well be so that there is no single model for offensive operations based on context and resources. Training for offensive operations requires virtual environments and courses. However, only one researcher discusses the importance of personality traits for offensive and defensive operations. This area requires more research in order to mitigate the risk of insider threats to the operational security of offensive operations.

## **5. Conclusions and future work**

This article has presented a review of the scientific literature on offensive cyber operations. Six scientific databases were queried, resulting in 180 articles. Of these, 87 articles were reviewed after screening. It is evident that some research describes actors investing, developing and conducting offensive operations. In contrast, less research is focused on governance, training and decision-making. Furthermore, there is an opportunity for research in processes for operations: Lin (2010) describes one. The answer to (RQ1) what is the current state of research in offensive cyber operations is covered by 4.1-4.5. The answer to (RQ2) where research is focused is depicted by Table 1. It is evident that a lot of research is focused on general offensive operations, deterrence and policy. However, there is room for more research on governance, training and decision-making. Overall, there is potential for more research, especially in processes for operations.

## **Acknowledgements**

The author would like to thank Johan Sigholm, PhD, for his valuable inputs to the manuscript.

## **References**

- Baltrusaitis, D. F. (2017). Cyber warfare Cyber War: Do We Have the Right Mindset? In *Handbook of Cyber-Development, Cyber-Democracy, and Cyber-Defense* (pp. 1–22). Cham: Springer International Publishing.  
[https://doi.org/10.1007/978-3-319-06091-0\\_24-1](https://doi.org/10.1007/978-3-319-06091-0_24-1)

## Gazmend Huskaj

- Bardin, J. (2015). Cyber Operations in the Middle East. In F. Lemieux (Ed.), *Current and Emerging Trends in Cyber Operations: Policy, Strategy and Practice* (pp. 97–110). London: Palgrave Macmillan UK.  
[https://doi.org/10.1057/9781137455550\\_7](https://doi.org/10.1057/9781137455550_7)
- Blachman, N., and Peek, J. Quotation Marks Replace the + Operator. Retrieved from  
[http://www.googleguide.com/quote\\_operator.html](http://www.googleguide.com/quote_operator.html)
- Borghard, E. D., & Lonergan, S. W. (2016). Can States Calculate the Risks of Using Cyber Proxies? *Orbis*, 60(3), 395–416.  
<https://doi.org/10.1016/j.orbis.2016.05.009>
- Brantly, A. F. (2016). The decision to attack: military and intelligence cyber decision-making. *The Decision To Attack: Military and Intelligence Cyber Decision-Making*.
- Brantly, A. F. (2016). The Most Governed Ungoverned Space: Legal and Policy Constraints on Military Operations in Cyberspace. *SAIS Review of International Affairs*, 36(2), 29–39. <https://doi.org/10.1353/sais.2016.0018>
- Buchanan, B. (2017). *The Intruder's View. The Cybersecurity Dilemma: Hacking, Trust and Fear Between Nations* (Vol. 1). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780190665012.003.0003>
- Bugeja, J., Jacobsson, A., & Davidsson, P. (2017). An analysis of malicious threat agents for the smart connected home. *2017 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2017*, 557–562. <https://doi.org/10.1109/PERCOMW.2017.7917623>
- Burke, I., & Van Heerden, R. P. (2016). Automating cyber offensive operations for cyber challenges. *Proceedings of the 11th International Conference on Cyber Warfare and Security, ICCWS 2016*, (Marczewski 2013), 65–73.
- Charlet, K., Romanosky, S., & Thompson, B. (2017). It's Time for the International Community to Get Serious about Vulnerability Equities, 0.
- Colloquium, B., & Bruges, C. De. (2010). Technological Challenges for the Humanitarian Legal Framework. *Bruges Colloquium*, (October).
- Eriksson, G., & Petterson, U. (2017). *Special Operations from a Small State Perspective*. (G. Eriksson & U. Pettersson, Eds.). Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-43961-7>
- Fischerkeller, M. (2017). Incorporating Offensive Cyber Operations into Conventional Deterrence Strategies. *Survival*, 59(1), 103–134. <https://doi.org/10.1080/00396338.2017.1282679>
- Flahive, M. P. (2015). *Breaking Bad: Reforming Cyber Acquisition Via Innovative Strategies*. Air Command and Staff College Air University.
- Gompert, D., & Binnendijk, H. (2016). The Power to Coerce: Countering Adversaries Without Going to War.  
<https://doi.org/10.7249/RR1000>
- Grant, T. (2015). Specifying functional requirements for simulating professional offensive cyber operations. *Proceedings of the 10th International Conference on Cyber Warfare and Security, ICCWS 2015*, (March), 108–117.
- Grant, T. (2017). Grant, T.J. (2017). Speeding up Parliamentary Decision Making for Cyber Counter-Attack. In Bryant, A.R., Lopez, J.R. & Mills, R.F. (eds.), *Proceedings*, 12, (March), 152–159.
- Grant, T. J. (2013). Tools and Technologies for Professional Offensive Cyber Operations. *International Journal of Cyber Warfare and Terrorism*, 3(3), 49–71. <https://doi.org/10.4018/ijcwt.2013070104>
- Grant, T., Burke, I., & van Heerden, R. (2012). Comparing Models of Offensive Cyber Operations. *International Conference on Information Warfare and Security*, (January), 108–121.
- Grant, T., Eijk, E. van, & Venter, H. (2016). Assessing the Feasibility of Conducting the Digital Forensic Process in Real Time. *11th International Conference on Cyber Warfare and Security: ICCWS2016*, (March), 146.
- Hart, S. W., & Klink, M. C. (2017). 1st Troll Battalion: Influencing military and strategic operations through cyber-personas. In *2017 International Conference on Cyber Conflict (CyCon U.S.)* (pp. 97–104). IEEE.  
<https://doi.org/10.1109/CYCONUS.2017.8167503>
- Hawkins, Z. (2016). Digital land power: the Australian Army's cyber future, 2016–2018. Retrieved from  
<https://www.aspistrategist.org.au/digital-land-power-australian-armys-cyber-future/>
- Heickero, R. (2015). Russia's Information Warfare Capabilities. In *Current and Emerging Trends in Cyber Operations* (pp. 65–83). London: Palgrave Macmillan UK. [https://doi.org/10.1057/9781137455550\\_5](https://doi.org/10.1057/9781137455550_5)
- Herr, T., & Herrick, D. (2016). Understanding Military Cyber Operations. In R. M. Harrison & T. Herr (Eds.), *Cyber Insecurity* (p. 412). Rowman & Littlefield Publishers.
- Howard, D. (2014). Virtue in Cyberconflict. In L. Floridi & M. Taddeo (Eds.) (Vol. 14, pp. 155–168). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-04135-3\\_10](https://doi.org/10.1007/978-3-319-04135-3_10)
- Hurley, J. S. (2017). Handbook of Cyber-Development, Cyber-Democracy, and Cyber-Defense. <https://doi.org/10.1007/978-3-319-06091-0>
- Irion, K. (2013). *The Secure Information Society*. (J. Krüger, B. Nickolay, & S. Gaycken, Eds.), *The Secure Information Society: Ethical, Legal and Political Challenges*. London: Springer London. <https://doi.org/10.1007/978-1-4471-4763-3>
- Jajodia, S., Subrahmanian, V. S., Swarup, V., & Wang, C. (2016). Cyber deception: Building the scientific foundation. *Cyber Deception: Building the Scientific Foundation*, 1–312. <https://doi.org/10.1007/978-3-319-32699-3>
- Johnson, M. C. (2014). *Refining United States Policy on Offensive Cyber Operations*. Air University.
- Joint Chiefs of Staff. (2018). Joint Publication 3-12: Cyberspace Operations.
- Josang, A. (2014). Potential Cyber Warfare Capabilities of Major Technology Vendors. *Proceedings of the 13th European Conference on Cyber Warfare and Security (Eccws-2014)*, (July), 110–115.
- Kallberg, J., & Cook, T. S. (2017). The Unfitness of Traditional Military Thinking in Cyber: Four Cyber Tenets That Undermine Conventional Strategies. *IEEE Access*, 5, 8126–8130. <https://doi.org/10.1109/ACCESS.2017.2693260>

- Kallberg, J., & Rowlen, S. (2014). African nations as proxies in covert cyber operations. *African Security Review*, 23(3), 307–311. <https://doi.org/10.1080/10246029.2014.924976>
- Kallberg, J., & Thuraisingham, B. (2013). Chapter 19 – From Cyber Terrorism to State Actors' Covert Cyber Operations. *Strategic Intelligence Management*. <https://doi.org/10.1016/B978-0-12-407191-9.00019-3>
- Kshetri, N. (2016). Cybersecurity in South Korea. In *The Quest to Cyber Superiority* (Vol. 2010, pp. 171–182). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-40554-4\\_10](https://doi.org/10.1007/978-3-319-40554-4_10)
- Kshetri, N. (2016). The Quest to Cyber Superiority. <https://doi.org/10.1007/978-3-319-40554-4>
- Land, C. A., & Centre, W. (n.d.). *For Canada's future army*.
- Lee, J.-A. (2015). The Sino-US Digital Relationship and International Cyber Security. In *Current and Emerging Trends in Cyber Operations* (pp. 84–96). London: Palgrave Macmillan UK. [https://doi.org/10.1057/9781137455550\\_6](https://doi.org/10.1057/9781137455550_6)
- Lehto, M. (2015). Cyber Security: Analytics, Technology and Automation, 78, 3–29. <https://doi.org/10.1007/978-3-319-18302-2>
- Lemieux, F. (2015). Trends in Cyber Operations: An Introduction. In F. Lemieux (Ed.), *Current and Emerging Trends in Cyber Operations: Policy, Strategy and Practice*. Palgrave Macmillan.
- Lin, H. (2017). Cybersecurity and Deterrence - A Notification Requirement for Using Cyber Weapons or for Unauthorized Disclosure of a Cyber Weapon.
- Lonsdale, D. J. (2017). Warfighting for Cyber Deterrence: a Strategic and Moral Imperative. *Philosophy & Technology*. <https://doi.org/10.1007/s13347-017-0252-8>
- McArdle, J. (2016). An Assessment of Russian and Chinese Offensive Cyber Operations on U.S. Space Assets. In E. Sterner & J. McArdle (Eds.) (pp. 10–22).
- Nevill, L. (2016). Cyber wrap. Retrieved from <https://www.aspistrategist.org.au/cyber-wrap-145/>
- Nikitakos, N., & Mavropoulos, P. (2014). Cyberspace as a State's Element of Power. In *Cyber-Development, Cyber-Democracy and Cyber-Defense* (Vol. 9781493910, pp. 259–277). New York, NY: Springer New York. [https://doi.org/10.1007/978-1-4939-1028-1\\_10](https://doi.org/10.1007/978-1-4939-1028-1_10)
- Olagbemiro, A. O. (2014). Cyberspace as a Complex Adaptive System and the Policy and Operational Implications for Cyber Warfare. United States Army Command and General Staff College Fort Leavenworth, Kansas.
- Oltramari, A., Lebiere, C., Vizenor, L., Zhu, W., & Dipert, R. (2013). Towards a cognitive system for decision support in cyber operations. *CEUR Workshop Proceedings*, 1097, 94–100.
- Ormrod, D. G. A. (2014). A “wicked problem” - Predicting sos behaviour in tactical land combat with compromised C4ISR. *Proceedings of the 9th International Conference on System of Systems Engineering: The Socio-Technical Perspective, SoSE 2014*, 107–112. <https://doi.org/10.1109/SYSoSE.2014.6892472>
- Ottis, R. (2015). Cyber Security: Analytics, Technology and Automation, 78, 89–96. <https://doi.org/10.1007/978-3-319-18302-2>
- Prescott, J. M. (2012). Direct Participation in Cyber Hostilities: Terms of Reference for Like-Minded States ? 2012 4th International Conference on Cyber Conflicts, 8(May 2011), 251–266.
- Schwartz, A., & Knake, R. (2016). Government's Role in Vulnerability Disclosure: Creating a Permanent and Accountable Vulnerability Equities Process. *Harvard Kennedy School - Belfer Center*, 3, 28.
- Schweitzer, D., Gibson, D., Bibighaus, D., & Boleng, J. (2013). Preparing our undergraduates to enter a cyber world. *IFIP Advances in Information and Communication Technology*, 406, 123–130. [https://doi.org/10.1007/978-3-642-39377-8\\_13](https://doi.org/10.1007/978-3-642-39377-8_13)
- Segal, A. (2016). U.S. Offensive Cyber Operations in a China-U.S. Military Confrontation. *SSRN Electronic Journal*, (March 2016). <https://doi.org/10.2139/ssrn.2836203>
- Sidari, B. D. (2016). *Offensive Cyber Operations: The Need for a Policy to Contend with the Future*.
- Sigholm, J., & Larsson, E. (2014). Determining the utility of cyber vulnerability implantation: The heartbleed bug as a cyber operation. *Proceedings - IEEE Military Communications Conference MILCOM*, 110–116. <https://doi.org/10.1109/MILCOM.2014.25>
- Sin, S. S., Blackerby, L. A., Asiamah, E., & Washburn, R. (2016). Determining extremist organisations' likelihood of conducting cyber-attacks. *International Conference on Cyber Conflict, CYCON, 2016–August*, 81–98. <https://doi.org/10.1109/CYCON.2016.7529428>
- Smyth, V. (2014). The Best Defense is a Good Offense: Conducting Offensive Cyberoperations and the Law.
- Spafford, E. (1988). The Internet Worm Program: An Analysis. Retrieved from <https://spaf.cerias.purdue.edu/tech-reps/823.pdf>.
- Sprengers, M., & van Haaster, J. (2016). Organization of #operations. *Cyber Guerilla*. <https://doi.org/10.1016/B978-0-12-805197-9.00003-6>
- Stark, H. (2011). Mossad's Miracle Weapon - Stuxnet Virus Opens New Era of Cyber War. *Der Spiegel*.
- Subrahmanian, V. S., Mannes, A., Sliva, A., Shakarian, J., & Dickerson, J. P. (2013). Policy Options Against LeT. In *Computational Analysis of Terrorist Groups: Lashkar-e-Taiba* (pp. 157–176). New York, NY: Springer New York. [https://doi.org/10.1007/978-1-4614-4769-6\\_11](https://doi.org/10.1007/978-1-4614-4769-6_11)
- Sutton, W. S. (2013). Cyber Operations and the Warfighting Functions, 32.
- Torres, A. M. (2012). Offensive Cyber is Fires, A Case for MAGTF Integration, 3330(703).
- Witte, J. C. (2015). *The Panacea and the Square Peg: Strategic Fallacies of the Air, Undersea and Cyber Domains*.
- Yeo, S., Birch, A. S., & Bengtsson, H. I. J. (n.d.). The Role of State Actors in Cybersecurity (pp. 217–246). <https://doi.org/10.4018/978-1-4666-9661-7.ch013>.

# Cyber Sanctions: The Embargo of Flagged Data in a Geo-Cultural Internet

Ion Iftimie

CEU, Vienna, Austria

[iftimie\\_ion@phd.ceu.edu](mailto:iftimie_ion@phd.ceu.edu)

**Abstract:** This paper introduces the concept of cyber sanctions, which can be defined as the actual or threatened restriction of digital transactions to affect a behavioral change by the target through the introduction of psychological pressure against its political leaders and populace. It argues that the Internet is currently undergoing a process of fragmentation along geographic and cultural lines and proposes that, while the concept of ‘internet sovereignty’ deals with the country’s choice to control foreign data from coming in or ‘sovereign’ data from going out (self-imposed digital isolation), cyber sanctions deal with senders (powerful states or entities imposing the sanctions) restricting certain ‘flagged’ data from traveling to or from the target (forced digital isolation).

**Keywords:** cyber sanctions, internet fragmentation, internet sovereignty, information warfare, operation glowing

---

## 1. Introduction

Since the inception of the internet, governments have struggled to cope with the digital transformation and to find new ways of controlling online transactions. A study of 606 incidents—Involving 99 countries between 1995 and 2010—of governments that “actually disconnected Internet exchange points or blocked significant amounts of certain kinds of traffic” showed that “39 percent of the incidents occurred in democracies, 6 percent occurred in emerging democracies, 52 percent occurred in authoritarian regimes, and 3 percent occurred in fragile states” (Howard et al., 2011). Over the past few years, however, these attempts have been joined by new laws attempting to shape national and/or regional identity and dimensions of data ownership online. The adoption of “internet sovereignty” (Wu 1996) legislation around the world (Qiang 2019) has started a process of fragmentation and regionalization of the internet along geographic and cultural lines. A series of laws in Russia have been issued since 2014 “waging a campaign to gain complete control over the country’s access to, and activity on, the Internet,” with President Putin even proposing a ‘kill switch’ that would “allow the government to shut down the Internet in Russia during government-defined disasters, including large-scale civil protests” (Duffy, 2015). The European Union General Data Protection Directive (EU regulation 2016/679) has also provided regulatory precedent for the establishment of a regional Digital Single Market “requiring organizational and technical safeguards, such as data pseudonymization, and mandating the designation of a data protection officer in case large-scale and systematic processing of sensitive data occurs (Arts. 37 to 39)” (Marelli and Testa, 2018). In China, Article 37 of its new Cybersecurity Law (CSLaw), which went into effect on 1 June 2017 (Segal, 2018), stipulates that “personal information and other important business data gathered or produced by CII [critical information infrastructure] operators during operations within the mainland territory of the People’s Republic of China, shall be stored within mainland China.” (Heymann et al., 2018). These laws to control digital spaces show that the forced fragmentation of the internet is not just a national or regional phenomenon, but a global one in a “dynamic, complex, multilevel, multimodal, multilateral, and volatile environment” (Carayannis and Dubina, 2014).

The struggle for online dominance is waged not only at national levels, but also “across economic, technical, regulatory, political, and social battlefields” (Hathaway, 2014). The new legislation frameworks have, however, opened the door to new ways of digital censorship and coercion that can be used to control the flow of ‘critical information’. This paper argues that there is a cost associated with censorship. The most frequent expressions of sovereignty across digital infrastructure are “national internet shutdowns, followed by subnational mobile internet, national app/service, subnational internet, national mobile internet, and subnational app/service disruptions;” of these, internet shutdowns alone have cost countries \$2.4 billion in 2015 (West, 2016). If there is a cost that can be associated to disrupting internet access, this implies that censorship can also be used as an instrument of coercion, or as a “cyber embargo” (McNeal, 2006). This paper first looks at the constraints inherent in the application of economic sanctions, and then assesses the feasibility of imposing local/national/regional/ideology-based sanctions online.

## 2. An overview of economic sanctions

Throughout history, economic sanctions have often been viewed by nation states as effective foreign policy tools to solve political, economic, or territorial disputes—despite the fact that only 34 percent of them have been (partially) successful (Hufbauer et al., 2008: 158). Economic sanctions are defined as “the actual or threatened withdrawal of economic resources to affect a policy change by the target” (Chan and Drury, 2000: 2) through the introduction of “psychological pressure against its political leaders and populace” (Eland, 1995: 37). This type of hegemonic behavior can be traced back all the way to 432 B.C., when the Athenian Empire issued the Megarian decree, banning Megarian goods from the Athenian market (Chan and Drury, 2000). Despite political (Doran, 1998; Galtung, 1967; Mansfield, 1994: 116), economic (Neuenkirch and Neumeier, 2015), social (Drury and Peksen, 2014; Weiss et al., 1997), technological (Caruso, 2003), and environmental (van Bergeijk, 1995) arguments against the use of sanctions, they continue to be imposed today by powerful states and entities (individually, or in alliance) against weaker ones. Special attention needs to be given to the debate of whether sanctions should be imposed in a unilateral (Hufbauer et al., 1990; Mitrany, 1925) or multilateral (Barfield and Groombridge, 1998) format. Research shows that the success rate of economic sanctions decreases when sanctions are multilateral, multi-issue, and when no international institution is present (Bapat and Morgan, 2009). Multilateral economic sanctions require 1) a coalition of states agreeing on the specific purpose of the sanctions; 2) a political will to impose the sanctions; and 3) the economic power to impose them (Askari et al., 2003: 28). Based on 174 case studies encompassing 204 observations, senders (the nation states or entities imposing the sanctions) “should be confident that their goals are within their reach, that they can impose sufficient economic pain to command the attention of the target country, that they can follow up economic sanctions with the threat or reality of military force or covert action as necessary, that their efforts will not prompt offsetting policies by other powers, and that the sanctions chosen will not impose insupportable costs on their domestic constituents and foreign allies” (Hufbauer et al., 2008: 178). At the international level, economic sanctions “must be judged as an instrument serving the higher goals of the polity” (Baldwin, 1985: 65), in a fragile context where stakeholders rarely agree on these goals and their perceived costs.

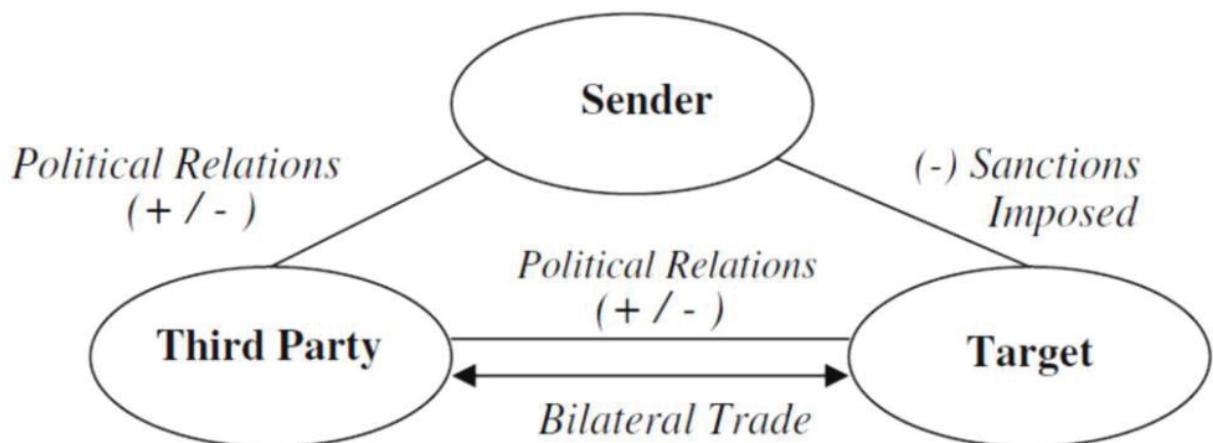
The senders must first determine a joint position clearly delineating which behavior the target (the economically and politically weaker state, company, or individual) is expected to change (Chan and Drury, 2000: 2). The reasons among the senders to impose economic sanctions on the target rarely coincide (Sperandei, 2009), meaning that the joint position will directly impact the political will of each sender. In other words, the salience of the joint position will differ from sender to sender, and is directly proportional with the differential cost of sanctions among the senders in terms of efficacy—defined as “the amount of loss that can be averted versus the cost of taking that action” (Chu, 1966). Hufbauer argued that “stripped to the bare bones, the formula for a successful sanctions effort is simple: The costs of defiance borne by the target must be greater than its perceived costs of compliance. That is, the political and economic costs to the target from sanctions must be greater than the political and security costs of complying with the sender’s demands” (Hufbauer et al., 2008: 50). The success or failures of sanctions, can thus be defined by salience (in terms of relative intensity of interest) and economic power (in terms of relative size and sender leverage), table 1, but may also be dependent on other factors, such as the duration of sanctions, ability to circumvent them or to find alternatives/substitutes to the instruments of coercion, and other unintended consequences.

**Table 1:** Expected outcomes, in terms of salience (relative intensity of interest) and economic power (relative size and sender leverage)

Relative intensity of interest	Relative size and sender leverage		
	TARGET > SENDER	TARGET = SENDER	TARGET < SENDER
TARGET > SENDER	Failure	Failure	Success possible but not likely
TARGET = SENDER	Failure	Indeterminate	Success possible but depends on goal, with modest goals being more achievable than ambitious goals
TARGET < SENDER	Success possible but not likely	Success possible	Success

Source: Adapted from (Hufbauer et al., 2008: 51)

Salience of the joint position is an important element to consider because the cost of imposing sanctions is simply greater for some sender states than the potential benefits brought upon by the sanctions. Winston Churchill once said that “the inherent vice of capitalism is the unequal sharing of blessings” (Zupan, 2011); and this is also the case with the benefits of imposing multilateral economic sanctions. There are times when the costs of imposing sanctions are just too high to bear by the sender. Hufbauer, Schott, and Elliott (Hufbauer et al., 1990: 48) determined that the relative cost to the sender can be classified in one of the four categories: 1) “net gain to the sender” (usually the case when only aid is withheld); 2) “little effect on sender” (when insignificant trade disruptions occur); 3) “modest loss to sender” (when trade is lost, but loss is not substantial); and 4) “major loss to sender” (when loss of trade adversely affects the sender’s economy). Each sender fits in one of the categories listed above; and in the case of multilateral sanctions, they almost never fit jointly under the same category. Because of this, efficacy of the sanctions differs from sender to sender; meaning that some senders and third parties (figure 1), which incur a major loss as a result of the imposed sanctions, are more likely to stop supporting the sanctions.



Source: (Early, 2009: 52)

**Figure 1:** Sanctions trade-offs in the international environment

Another critical element defining the success or failure of multilateral sanctions is the duration of the sanctions. More than half of recorded successes recorded by Hufbauer occurred during the first two years of the sanctions, with probability of success decreasing over time (Dizaji and van Bergeijk, 2013: 722; Hufbauer et al., 2008). Lastly, unintended consequences can also influence the success or failure of multilateral sanctions. The Targeted Sanctions Consortium (TSC) datasets, looking at 63 episodes of UN targeted sanctions (between 1991 and 2013), for example, identified several unintended negative consequences of sanctions, many of which are often not considered in assessing the success or failure of sanctions. These may include (but are not limited to) “legacies of corruption and criminality often left by sanctions, the strengthening of instruments of authoritarian rule, a ‘rally around the flag’ effect, an increase in human rights violations, and their harmful effects on neighboring states” (Biersteker et al., 2018). Too often, the declarations highlighting the success of multilateral sanctions do not account for the great unintended cost and unintended negative consequences to both the target and the senders.

Unilateral sanctions are even harder to impose, because in the case of unilateral sanctions, the sender needs to have a greater amount of economic power to impose sanctions (calculated in % control of the instruments of the sanctions on the target). This economic power over the instruments of coercion has to be strong enough to enable the sender to impose economic sanctions without having to rely on a coalition with conflicting political goals and objectives. Otherwise, there is no guarantee that the target will not simply circumvent the sender by trading with other third parties (Bapat and Morgan, 2009) or find alternatives to the instruments of coercion. Despite this, empirical evidence from 888 cases from 1971 to 2000 (Morgan et al., 2009) “demonstrate convincingly and consistently that multilateral sanctions were less effective than unilateral sanctions” (Bapat and Morgan, 2009). This shows that under certain political, economic, social, technological, and economic conditions unilateral sanctions “can indeed work in terms of influencing the policies of the actor against which they are ostensibly targeted” (Taylor, 2012: 21). Salience, the ability to circumvent sanctions or to find

alternatives, the duration of sanctions, and their unintended negative or positive consequences remain important factors to consider; but the main advantage here is that unilateral senders do not have to agree with other states on the purpose and extent of the sanctions.

### **3. Cyber sanctions: A definition**

To the traditional ‘need to sell’ and ‘need to buy’ economic dimensions of sanctions, digital transformations have added new “need to know” and “need to share” (Best and Jr., 2011; Dawes and Cresswell, 2012; McDermott, 1999) knowledge dimensions. This paper proposes that the ‘need to know’ and ‘need to share’ digital considerations can sometimes be just as salient (in terms of efficacy) as the ‘need to sell’ and ‘need to buy’ economic considerations; meaning that they can be used as instruments of coercion. Technology transfer, for example, has often “served as an important instrument to advance the national security and foreign policy objectives” of powerful states (Yuan, 1996). This has led to emergence of technology transfer sanctions (Tow, 1983), which some have suggested may be “of limited utility and questionable efficacy” (Yuan, 1996) because of a multitude of alternative and/or substitutes. Nevertheless, technology-related sanctions today are not only imposed by nation states and international organizations, but also by “municipalities, small businesses, religious organizations, universities, international financial institutions, unions, and multinational corporations” (Crawford, 1999). Among these, there is no research data analyzing the efficacy of digital or cyber sanctions.

This paper proposes that cyber sanctions can be defined as the actual or threatened restriction of digital transactions to affect a behavioral change by the target through the introduction of psychological pressure against its political leaders and populace. While the concept of “internet sovereignty” (Wu, 1996) deals with country’s choice to control of foreign data from coming in or “sovereign” data (Qiang, 2019) from going out (self-imposed digital isolation), cyber sanctions deal with senders (the countries or international organizations imposing the sanctions) restricting certain “flagged” data (Scales et al., 2011) from traveling to or from the target (forced digital isolation).

#### **3.1 Cyber sovereignty and the geography of cyberspace**

The growing debates for cyber sovereignty and data sovereignty (Baezner and Robin, 2018) challenges the traditional belief that "geography of cyberspace" does not exist (Jensen, 2015). For example, China (Knockel et al., 2018; Kou et al., 2017; Li et al., 2018; Roberts, 2018; Singhi and Liu, n.d.) and Russia (Budnitsky and Jia, 2018; Kerr, 2018; Ognyanova, 2019) are often accused of taking active steps to censor the internet within geographic boundaries, despite the fact that many researchers and legal experts often argue that internet access is a human right (Best, 2004; Mathiesen, 2012; Moodley, 2015; Skepys, 2012; Tăbușcă, 2010; Tully, 2014). An example of self-imposed geographic isolation is the ‘Great Firewall of China’, which has restricted access to data from several U.S. companies, such as Google or Facebook (Lee et al., 2012). This has set precedent for many corporate or government websites that cannot be accessed from IPs belonging to certain geographical regions. While the use of circumvention tools (Mou et al., 2016)—virtual private networks or web proxies—can simply overcome these types of restrictions, they do suggest that the internet is undergoing a process of fragmentation across geographic and cultural lines, opening the door to future cyber sanctions on the internet.

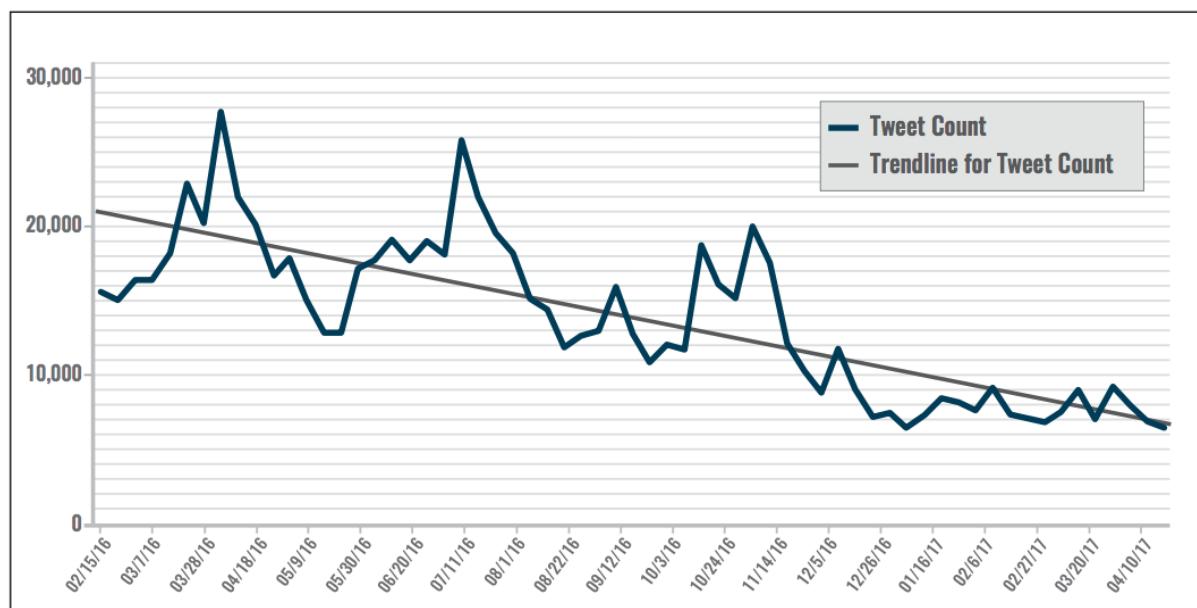
#### **3.2 Cyber sanctions in a geo-cultural internet**

Governments, industry, and even academia—particularly under the guise of protecting intellectual property online—have been known to restrict access to digital data through “patents, copyright, and trade secret laws” (Samuelson, 1990). Nation states and other powerful entities, however, can also restrict data through the use of cyber tools that deny the target’s need to share and need to know. The United States Department of Defense (DoD) 2015 Cybersecurity Strategy, for example, calls for “building capabilities for effective cybersecurity and cyber operations”, to include offensive cyber weapons to be used against terrorist organizations such as ISIL (Miller et al., 2019). According to DoD, these tools are used so that cyber adversaries “lose confidence in their networks, to overload their networks so that they can’t function, and do all of these things that will interrupt their ability to command and control forces” (Lin and Zegart, 2017, 2019). It can be argued that these tools not only target physical networks, but also the entire “socio-geography of terrorism” (Simons and Tucker, 2007), which is composed of both physical nodes and their impact on the radical jihadi subculture (Pisoiu, 2015).

Offensive cyber capabilities have been deemed legal on January 19, 2017 by the General Counsel of the U.S. Department of Defense, who issued a memorandum to U.S. Combatant Commands titled "International Law

Framework for Employing Cyber Capabilities in Military Operations" (Efrony and Shany, 2018). The legal document sets a dangerous precedent, because it rejects the notion of sovereignty in cyberspace and concludes that "state cyber operations that interfere with the integrity of cyber infrastructure without the consent of a territorial State, that intrude into such cyber infrastructure, or perhaps even that alter such systems or their data without effects amounting to force or intervention do not amount to internationally wrongful acts" (Watts and Richard, 2018: 830). Legal experts assert that this may have been the case of a late 2016 operation code-named Glowing Symphony (OGS), where "the United States Cyber Command reportedly acquired administrator passwords to Islamic State (IS) websites. The passwords enabled deletion of digital content, including videos used for recruitment, from cyber infrastructure located in at least five countries outside actively hostile areas of Iraq and Syria.' Similar digital content reportedly resided on cyber infrastructure in as many as 30 other States. Changing the passwords reportedly locked IS administrators out of the websites" (Watts and Richard, 2018: 772).

This study argues that OGS is a perfect example where cyber sanctions, under the defined construct, have already been applied successfully against a terrorist organization. OGS restricted ISIL's need to share not only on networks physically residing in Iraq and Syria (which were controlled by the terrorist group), but also worldwide, where sovereign states (the United States and owners of affected physical infrastructures) acted as senders to effect this denial of service based on national security and socio-cultural grounds. Cyber sanctions imposed during OGS successfully disrupted Islamic State propaganda through content removal from servers residing in multiple countries and through restricting access to physical infrastructure needed to store digital data. This resulted in propaganda efforts being significantly reduced on several global social media platforms, including Twitter (see Figure 2). In effect, "offensive cyber weapons" (Peterson, 2013) acted as instruments of larger cyber sanctions, where the senders (the United States government and its allies) restricted access to physical networks—and thus, the need to know and need to share information online—critical for "recruitment, radicalization, training, communication, tactical use, command and control, fundraising, and cyberattacks" (Espeseth et al., 2013: 91).



Source: (Alexander, 2017: 15)

**Figure 2:** Number of Tweets by ISIL during USCYBERCOM Operation GLOWING SYMPHONY (2016-2017)

Compared to economic sanctions, where sovereignty limitations are clearly defined (often, requiring that economic sanctions are multilateral rather than unilateral), sovereignty does not, however, seem to affect cyber sanctions in similar ways. A legal analysis of OGS illustrates that "cyberspace greatly expands opportunities for States to violate the independence and exclusivity traditionally attendant to sovereignty. By means of its interconnected framework, cyberspace presents States unprecedented access to information and objects on the territory of other States. Cyberspace frees States from many of the geographic and physical restraints that might have previously prevented access. Because of their potential to compromise territorial integrity without significant impact to physical property or immediately proximate impact on persons, cyber operations bring into sharp focus the question whether mere intrusions into territorial property amount to internationally wrongful

acts" (Watts and Richard, 2018: 776). As a result, many of the requirements of imposing unilateral sanctions—such as the physical ownership of the instruments and/or object of coercion—are no longer a requirement, making unilateral sanctions easier to impose in cyberspace.

### **3.3 On unilateral and multilateral cyber sanctions**

The implementation of multilateral economic sanctions often failed because of multiple disagreements between the sender states (Ahrari, 1979: 16), the unwilling to pay the price of imposing the economic sanctions, and the "strategic deficits of prospective weapon users" (Stern, 2006: 1650). This paper argues that challenges of multilateral economic sanctions are amplified in the digital context, specifically because in order for multilateral cyber sanctions to work they require 1) that sender(s) agree on the purpose and cost of the sanctions, which may include privacy trade-offs among other social implications; 2) that sender(s) control the physical infrastructure needed by the target to conduct digital transactions on; and 3) that the need to know and need to share of the target is big enough to effect a change in behavior by the target. Cyberspace also offers an environment where cyber sanctions are easier to impose than economic sanctions in a unilateral setting (despite the growing fragmentation of the internet). For example, in the International Law Framework for Employing Cyber Capabilities in Military Operations memorandum discussed above, the General Counsel assessed that "sovereignty does not prevent States from undertaking a cyber operation against cyber infrastructure used by terrorists in other States even without the consent of the latter State so long as the operation is short of the use of force or intervention" (Watts and Richard, 2018: 829). This would require, however, significant capabilities to access foreign infrastructure without consent and/or discovery, and the willingness to do it, given that it could result in significant international backlash.

From a defense perspective, economic sanctions are perceived as a form of "economic warfare" (Naylor, 2001). This remains true in the case of unilateral and multilateral cyber sanctions. For example, in the United States, the Cyber Command "was devised specifically to make the United States proficient and powerful in this tool of war" (Lin and Zegart, 2017, 2019). But it would be erroneous to limit the foreign policy potential of cyber sanctions to its brute offensive instruments of a state sponsored "denial of service" attacks (Geers, 2010). Both legal and ethical options to restrict access to physical nodes and respective digital nodes exist; but not without serious implications to our understanding of privacy regionally (Tosoni, 2018) and globally (Choo and Sarre, 2015; Shackelford, 2016). The legitimacy and necessity of the cyber sanctions also needs to be given serious consideration by the sender(s), because technological changes could empower the target to circumvent the sanctions. In the end, the trade-offs in terms of privacy-loss by society at large, may lead to little results. The efficacy of the sanctions would also need to be considerable, as cyber sanctions could have a limited lifespan. It can be argued that cyber sanctions work best if the high technological dependency on the need to know and need to share requirements of the target is reinforced by the lack of expedient technological alternatives to replace this dependency (Ngobi, 2018: 19).

## **4. Conclusion**

This paper argues that cyber sanctions in a local/national/regional/ideology context have already been employed. Empirical evidence shows that cyber sanctions can be particularly effective if the digital need to know or to share certain data by the target is big enough, and if the sender(s) have a significant control over the instruments and/or objects of coercion (and their substitutes) needed to fulfill this need. Unlike the case of multilateral economic sanctions, cyber sanctions do not necessarily require senders to work together to restrict access to physical networks (which are owned by other states, and that the target needs in order to fulfill its digital need to know and need to share requirements). While precedent of a "cyber embargo" (McNeal, 2006)—in the form of multilateral cyber sanctions—against a terrorist group (ISIL) has been discussed, with impacts of unilateral cyber sanctions on "cyber-diplomacy" (Hocking, 2005; Pahlavi, 2003; Potter, 2002), further research is needed to determine to what extent cyber sanctions can be used outside of a counter-terrorism framework. Further research is also needed to explore when and how cyber sanctions can be imposed unilaterally against states that are members of the UN Security Council (such as Russia or China, for example) or against rogue states (such as Iran and North Korea).

**Disclaimer:** Although the author served in a senior leadership position for a command responsible for U.S. Department of Defense cyber operations at the time of Operation Glowing Symphony, no access to non-publicly available information on U.S. cyber operations form any part of this Article. The views, assumptions, and

opinions expressed in this article are those of the author and do not necessarily reflect the official policy or position of any agency of the U.S. government.

## **References**

- Ahrari M (1979) Oapec and 'authoritative' allocation of oil: An analysis of the Arab oil embargo. *Studies in comparative international development* 14(1). Springer-Verlag: 9–21.
- Alexander A (2017) Digital Decay? Tracing change over time among English-language Islamic State sympathizers on Twitter. *Program on Extremism*.
- Askari H, Forrer J, Teegen H, et al. (2003) *Economic Sanctions: Examining Their Philosophy and Efficacy*. Greenwood Publishing Group.
- Baezner M and Robin P (2018) *Cyber Sovereignty and Data Sovereignty*. ETH Zurich. Available at: <https://www.research-collection.ethz.ch/handle/20.500.11850/314613>.
- Baldwin DA (1985) *Economic Statecraft*. Princeton University Press.
- Bapat NA and Morgan TC (2009) Multilateral Versus Unilateral Sanctions Reconsidered: A Test Using New Data. *International studies quarterly: a publication of the International Studies Association* 53(4). Narnia: 1075–1094.
- Barfield CE and Groombridge MA (1998) Unilateral sanctions undermine US interests. *The World & I* 13(12). Washington Times Corporation: 92.
- Best ML (2004) Can the internet be a human right. *Human Rights & Human Welfare* 4(1). pdfs.semanticscholar.org: 23–31.
- Best RA and Jr. (2011) *Intelligence Information: Need-To-Know Vs. Need-to-Share*. DIANE Publishing.
- Biersteker TJ, Eckert SE, Tourinho M, et al. (2018) UN targeted sanctions datasets (1991–2013). *Journal of peace research* 55(3). SAGE Publications Ltd: 404–412.
- Budnitsky S and Jia L (2018) Branding Internet sovereignty: Digital media and the Chinese–Russian cyberalliance. *European Journal of Cultural Studies* 21(5). SAGE Publications Ltd: 594–613.
- Carayannis E and Dubina I (2014) Thinking Beyond The Box: Game-Theoretic and Living Lab Approaches to Innovation Policy and Practice Improvement. *Journal of the Knowledge Economy* 5(3). Springer US: 427–439.
- Caruso R (2003) The Impact of International Economic Sanctions on Trade: An Empirical Analysis. *Peace Economics, Peace Science and Public Policy* 9(2). degruyter.com. DOI: 10.2202/1554-8597.1061.
- Chan S and Drury AC (2000) Sanctions as Economic Statecraft: An Overview. In: Chan S and Drury AC (eds) *Sanctions as Economic Statecraft: Theory and Practice*. London: Palgrave Macmillan UK, pp. 1–16.
- Choo KR and Sarre R (2015) Balancing Privacy with Legitimate Surveillance and Lawful Data Access. *IEEE Cloud Computing* 2(4). ieeexplore.ieee.org: 8–13.
- Chu GC (1966) Fear arousal, efficacy, and imminency. *Journal of personality and social psychology* 4(5). psycnet.apa.org: 517–524.
- Crawford NC (1999) Trump Card or Theater? An Introduction to Two Sanctions Debates. In: Crawford NC and Klotz A (eds) *How Sanctions Work: Lessons from South Africa*. London: Palgrave Macmillan UK, pp. 3–24.
- Dawes SS and Cresswell AM (2012) From 'need to know' to 'need to share': Tangled problems, information boundaries, and the building of public sector knowledge networks. *Debating public*. taylorfrancis.com. Available at: <https://www.taylorfrancis.com/books/e/9781466502376/chapters/10.4324/9781315095097-12>.
- Dizaji SF and van Bergeijk PAG (2013) Potential early phase success and ultimate failure of economic sanctions: A VAR approach with an application to Iran. *Journal of peace research* 50(6). SAGE Publications Ltd: 721–736.
- Doran GT (1998) *The Futility of Economic Sanctions as an Instrument of National Power in the 21st Century*. Army War College, Carlisle Barracks, PA. Available at: <https://apps.dtic.mil/docs/citations/ADA343749>.
- Drury AC and Peksen D (2014) Women and economic statecraft: The negative impact international economic sanctions visit on women. *European Journal of International Relations* 20(2). SAGE Publications Ltd: 463–490.
- Duffy N (2015) Internet freedom in Vladimir Putin's Russia: The noose tightens. *AEI Paper & Studies*. The American Enterprise Institute: B1.
- Early BR (2009) Sleeping With Your Friends' Enemies: An Explanation of Sanctions-Busting Trade. *International studies quarterly: a publication of the International Studies Association* 53(1). Narnia: 49–71.
- Efrony D and Shany Y (2018) A Rule Book on the Shelf? Tallinn Manual 2.0 on Cyberoperations and Subsequent State Practice. *The American journal of international law* 112(4). Cambridge University Press: 583–657.
- Eland I (1995) Economic Sanctions as Tools of Foreign Policy. In: Lopez G and Cortright D (eds) *Economic Sanctions, Panacea Or Peacebuilding in a Post-Cold War World?* Boulder, CO: Westview Press, pp. 29–42.
- Espeseth C, Gibson J, Jones A, et al. (2013) Terrorist use of communication technology and social networks. *Technological dimensions of defence against terrorism* 115. IOS Press, Amsterdam: 91.
- Galtung J (1967) On the Effects of International Economic Sanctions, With Examples from the Case of Rhodesia. *World politics* 19(3). Cambridge University Press: 378–416.
- Geers K (2010) The challenge of cyber-attack deterrence. *Computer Law & Security Review* 26(3). Elsevier: 298–303.
- Hathaway ME (2014) Connected Choices: How the Internet Is Challenging Sovereign Decisions. *American Foreign Policy Interests* 36(5). Routledge: 300–313.
- Heymann RAT, Lehmann M and Nimmer RT (2018) Computer Law Review International. beijing-starke.com. Available at: [https://www.beijing-starke.com/sites/starke/files/publications/cri-1.18.daniel\\_albrecht.pdf](https://www.beijing-starke.com/sites/starke/files/publications/cri-1.18.daniel_albrecht.pdf).
- Hocking B (2005) Rethinking the 'new' public diplomacy. In: *The New Public Diplomacy*. Springer, pp. 28–43.

- Howard PN, Agarwal SD and Hussain MM (2011) The Dictators' Digital Dilemma: When Do States Disconnect Their Digital Networks? DOI: 10.2139/ssrn.2568619.
- Hufbauer GC, Schott JJ, Elliott KA, et al. (1990) *Economic Sanctions Reconsidered: History and Current Policy*. Peterson Institute.
- Hufbauer GC, Schott JJ, Elliott KA, et al. (2008) Economic Sanctions Reconsidered 3rd edition. *Peterson Institute Press: All Books*. Peterson Institute for International Economics. Available at: <https://ideas.repec.org/b/iie/pres/4082.html>.
- Jensen ET (2015) Cyber sovereignty: The way ahead. *Tex. Int'l LJ* 50. HeinOnline: 275.
- Kerr JA (2018) *The Russian Model of Internet Control and Its Significance*. Lawrence Livermore National Lab.(LLNL), Livermore, CA (United States). Available at: <https://www.osti.gov/servlets/purl/1491981>.
- Knockel J, Crete-Nishihata M and Ruan L (2018) The effect of information controls on developers in China: An analysis of censorship in Chinese open source projects. *Processing for Internet Freedom*. aclweb.org. Available at: <http://www.aclweb.org/anthology/W18-4201>.
- Kou Y, Kow YM and Gui X (2017) Resisting the Censorship Infrastructure in China. In: *Hawaii International Conference on System Sciences 2017 (HICSS-50)*, 2017. aisel.aisnet.org. Available at: [https://aisel.aisnet.org/hicss-50/dsm/politics\\_in\\_dsm/3/](https://aisel.aisnet.org/hicss-50/dsm/politics_in_dsm/3/) (accessed 2 April 2019).
- Lee J-A, Liu C-Y and Li W (2012) Searching for Internet freedom in China: A case study on Google's China experience. *Cardozo Arts & Ent. LJ* 31. HeinOnline: 405.
- Lin H and Zegart A (2017) Introduction to the special issue on strategic dimensions of offensive cyber operations. *Journal of Cybersecurity* 3(1). Narnia: 1–5.
- Lin H and Zegart A (2019) *Bytes, Bombs, and Spies: The Strategic Dimensions of Offensive Cyber Operations*. Brookings Institution Press.
- Li Y, Yang S, Chen Y, et al. (2018) Effects of perceived online–offline integration and internet censorship on mobile government microblogging service continuance: A gratification perspective. *Government information quarterly* 35(4). Elsevier: 588–598.
- Mansfield ED (1994) Alliances, Preferential Trading Arrangements and Sanctions. *Journal of international affairs* 48(1). Temporary Publisher: 119–139.
- Marelli L and Testa G (2018) Scrutinizing the EU General Data Protection Regulation. *Science* 360(6388). science.scienmag.org: 496–498.
- Mathiesen K (2012) The human right to Internet access: A philosophical defense. *International Review of Information Ethics* 18(12). irie.net: 9–22.
- McDermott R (1999) Why Information Technology Inspired but Cannot Deliver Knowledge Management. *California management review* 41(4). SAGE Publications Inc: 103–117.
- McNeal GS (2006) Cyber embargo: Countering the Internet jihad. *Case W. Res. J. Int'l L.* 39. HeinOnline: 789.
- Miller K, O'Halloran B, Pollman A, et al. (2019) Securing the Internet of Battlefield Things While Maintaining Value to the Warfighter. In: *International Conference on Cyber Warfare and Security*, 2019, pp. 546–553. Academic Conferences International Limited.
- Mitrany D (1925) *The Problem of International Sanctions: By D. Mitrany*. H. Milford, Oxford University Press.
- Moodley A (2015) *Can Internet access be a human right*. ukzn-dspace.ukzn.ac.za. Available at: [http://ukzn-dspace.ukzn.ac.za/bitstream/handle/10413/14286/Moodley\\_Ashailyn\\_2015.pdf?sequence=2&isAllowed=y](http://ukzn-dspace.ukzn.ac.za/bitstream/handle/10413/14286/Moodley_Ashailyn_2015.pdf?sequence=2&isAllowed=y).
- Morgan TC, Bapat N and Krustev V (2009) The threat and imposition of economic sanctions, 1971–2000. *Conflict Management and Peace Science* 26(1). Sage Publications Sage UK: London, England: 92–110.
- Mou Y, Wu K and Atkin D (2016) Understanding the use of circumvention tools to bypass online censorship. *New Media & Society* 18(5). SAGE Publications: 837–856.
- Naylor RT (2001) *Economic Warfare: Sanctions, Embargo Busting, and Their Human Cost*. UPNE.
- Neuenkirch M and Neumeier F (2015) The impact of UN and US economic sanctions on GDP growth. *European journal of political economy* 40. Elsevier: 110–125.
- Ngobi JC (2018) The United Nations Experience with Sanctions. In: Cortright D (ed.) *Economic Sanctions*. Routledge, pp. 17–27.
- Ognyanova K (2019) In Putin's Russia, Information Has You: Media Control and Internet Censorship in the Russian Federation. In: *Censorship, Surveillance, and Privacy: Concepts, Methodologies, Tools, and Applications*. IGI Global, pp. 1769–1786.
- Pahlavi PC (2003) Cyber-diplomacy: A new strategy of influence. In: *Canadian Political Association, General Meeting, Halifax, NS*, 2003. researchgate.net. Available at: [https://www.researchgate.net/profile/Pierre\\_Pahlavi/publication/228739617\\_Cyber-Diplomacy\\_A\\_New\\_Strategy\\_of\\_Influence/links/5906796e4585152d2e967394/Cyber-Diplomacy-A-New-Strategy-of-Influence.pdf](https://www.researchgate.net/profile/Pierre_Pahlavi/publication/228739617_Cyber-Diplomacy_A_New_Strategy_of_Influence/links/5906796e4585152d2e967394/Cyber-Diplomacy-A-New-Strategy-of-Influence.pdf).
- Peterson D (2013) Offensive Cyber Weapons: Construction, Development, and Employment. *Journal of Strategic Studies* 36(1). Routledge: 120–124.
- Pisoiu D (2015) Subcultural Theory Applied to Jihadi and Right-Wing Radicalization in Germany. *Terrorism and Political Violence* 27(1). Routledge: 9–28.
- Potter EH (2002) *Cyber-Diplomacy: Managing Foreign Policy in the Twenty-First Century*. McGill-Queen's Press - MQUP.
- Qiang X (2019) The Road to Digital Unfreedom: President Xi's Surveillance State. *Journal of Democracy* 30(1). Johns Hopkins University Press: 53–67.

- Roberts ME (2018) *Censored: Distraction and Diversion inside China's Great Firewall*. Princeton University Press.
- Samuelson P (1990) Digital media and the changing face of intellectual property law. *Rutgers Computer & Tech. LJ* 16. HeinOnline: 323.
- Scales GW, Elliott J and Norton J (2011) Systems and methods for the remote deletion of pre-flagged data. *US Patent*. 7895662. Available at: <https://patentimages.storage.googleapis.com/3b/85/3e/6d799cf1ebb085/US7895662.pdf> (accessed 23 March 2019).
- Segal A (2018) When China Rules the Web: Technology in Service of the State. *Foreign affairs* 97. HeinOnline: 10.
- Shackelford SJ (2016) Protecting intellectual property and privacy in the digital age: the use of national cybersecurity strategies to mitigate cyber risk. *Chap. L. Rev.* 19. HeinOnline: 445.
- Simons A and Tucker D (2007) The misleading problem of failed states: a 'socio-geography' of terrorism in the post-9/11 era. *Third world quarterly* 28(2). Taylor & Francis: 387–401.
- Singhi N and Liu R (n.d.) Cyber-nationalism in China. *researchgate.net*. Available at: [https://www.researchgate.net/profile/Nethra\\_Singhi/publication/325145087\\_Cyber-nationalism\\_in\\_China/links/5ba19f3d92851ca9ed14b629/Cyber-nationalism-in-China](https://www.researchgate.net/profile/Nethra_Singhi/publication/325145087_Cyber-nationalism_in_China/links/5ba19f3d92851ca9ed14b629/Cyber-nationalism-in-China).
- Skepys B (2012) Is there a human right to the Internet. *J. Pol. & L.* HeinOnline. Available at: [https://heinonline.org/hol-cgi-bin/get\\_pdf.cgi?handle=hein.journals/jpolal5&section=64&casa\\_token=9Hyh7tv1LMQAAAAA:4BQBtp0Fm68vie7QOgMTEK4rJqbA-k45kMg17eqgoFOCKYcuOeAqCbRPeE8T4zta52M3Jm4](https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/jpolal5&section=64&casa_token=9Hyh7tv1LMQAAAAA:4BQBtp0Fm68vie7QOgMTEK4rJqbA-k45kMg17eqgoFOCKYcuOeAqCbRPeE8T4zta52M3Jm4).
- Sperandei M (2009) Between rational choice and historical contingency: The hidden dilemma of multiple objectives in the study of economic sanctions. In: *Conference Papers*, 2009, pp. 1–34.
- Stern R (2006) Oil market power and United States national security. *Proceedings of the National Academy of Sciences of the United States of America* 103(5). National Acad Sciences: 1650–1655.
- Tăbușcă S (2010) The internet access as a fundamental right. *Journal of Information Systems and Operations*. repec.org. Available at: <ftp://ftp.repec.org/opt/ReDIF/RePEc/rau/jisomg/WI10/JISOM-WI10-A21.pdf>.
- Taylor B (2012) *Sanctions as Grand Strategy*. Routledge.
- Tosoni L (2018) Rethinking Privacy in the Council of Europe's Convention on Cybercrime. *Computer Law & Security Review* 34(6). Elsevier: 1197–1214.
- Tow WT (1983) U.S.-Japan military technology transfers: Collaboration or conflict? *Journal of Northeast Asian Studies* 2(4). Springer: 3–23.
- Tully S (2014) A Human Right to Access the Internet? Problems and Prospects. *Human Rights Law Review* 14(2). Narnia: 175–195.
- van Bergeijk PP (1995) The impact of economic sanctions in the 1990s. *The World Economy* 18(3). repub.eur.nl: 443–455.
- Watts S and Richard T (2018) Baseline Territorial Sovereignty and Cyberspace. *Lewis & Clark L. Rev.* 22(3). HeinOnline: 771.
- Weiss TG, Cortright D, Lopez GA, et al. (1997) *Political Gain and Civilian Pain: Humanitarian Impacts of Economic Sanctions*. Rowman & Littlefield.
- West DM (2016) Internet shutdowns cost countries \$2.4 billion last year. *Center for Technological Innovation at Brookings, Washington, DC.* witnessradio.org. Available at: <http://witnessradio.org/wp-content/uploads/2016/10/internet-shutdowns-v-3.pdf>.
- Wu TS (1996) Cyberspace Sovereignty--The Internet and the International System. *Harv. JL & Tech.* 10. HeinOnline: 647.
- Yuan J-D (1996) United States Technology Transfer Policy toward China: Post-Cold War Objectives and Strategies. *Aquatic microbial ecology: international journal* 51(2). SAGE Publications Ltd: 314–338.
- Zupan MA (2011) The virtues of free markets. *The Cato journal* 31. HeinOnline: 171.

# The Transformation of Islamic Terrorism Through Cyberspace: The Case of ISIS

Eleni Kapsokoli

University of Piraeus, School of Economics, Business and International Studies,

Department of International and European Studies, Greece

Laboratory of Intelligence and Cyber-Security

[ekapsokoli@unipi.gr](mailto:ekapsokoli@unipi.gr)

[elenikapsokoli1989@gmail.com](mailto:elenikapsokoli1989@gmail.com)

**Abstract:** Terrorism is a form of political violence that has been with us since antiquity. Over the last few decades, terrorism has been facing the challenges of cyberspace. Interactions and influences between cyberspace and terrorism are large and constantly changing. In order to achieve their goals, terrorist organizations use cyberspace to collect data, raise funds, and conduct propaganda, but also for the purposes of radicalization and operational planning. Scholars have widely discussed whether cyberspace has a transformative effect on the nature of terrorism. In order to examine the transformative or not effect of cyberspace on terrorism, we will analyze the case of Islamic terrorism and, in particular, Islamic State's actions in social media. In recent years, the ISIS has become very active in cyberspace. Specifically, many organizations and individuals are working on the name of ISIS and Jihad in cyberspace.

**Keywords:** terrorism, cyberterrorism, ISIS, Jihad, cyberspace, social media

---

## 1. Introduction

*"We are in a battle and more than half of this battle is taking place in the battlefield of the media (...). We are in a media battle for the hearts and minds of our Ummah"*, Ayman Al Zawahiri, July 2015

The rapid evolution of information and communication technologies (ICTs) as well as the emerging information society have led to a study of the emergence of possible risks in a society. According to the last count, cyberspace users are 4 billion worldwide, making its use important in our everyday life and necessary in order to achieve exchanges of information and data (Kemp, 2018).

Cyberspace is a domain of socio-political and military controversy, in which terrorism also operates (Bryant, 2001). There is also a darker side in cyberspace, which creates more and more challenges to individual and collective security. From the above, arises the question as to what extent and in what way the development of ICTs has affected the phenomenon of terrorism.

According to the existing sources after 9/11, terrorism is constantly evolving in terms of its characteristics and means of execution, but not to its objectives. The symbiotic relationship between terrorism and cyberspace is dominant and grounded (Hoffman, 1998, 56). According to Marshal McLuhan, without communication terrorism does not exist (Torres Soriano, 2008, 2). For terrorism the biggest failure is to be ignored by the target audience (Dowling, 1986, 1). Terrorists exploit the benefits of cyberspace, they learn from their past mistakes and adapt to the needs and circumstances of the present and the future.

The first part of this paper refers to terrorism and cyberspace as well as the connection between them. The second part is devoted to how the ISIS terrorist network acts, changing its action, making it one of the smartest actors in cyberspace. The purpose of the article is to find out if there is some sort of mutation in Islamic terrorism, to highlight the importance of communication strategy in modern terrorism and to analyze how an extremist organization like ISIS is using cyberspace to achieve its goals.

## 2. The utility of cyberspace for terrorist organizations

Based on Clausewitz's dictum that "war is a continuation of politics by other means" (Clausewitz, 1993), likewise it could be argued that cyberterrorism is the continuation of terrorism by other means. Terrorism is trying to gain attention through the publicity generated by its actions. The social media are the main source of information and transmission of such messages (Collins, 1997, 1). In many cases, publicity enables terrorist to produce political effect. Only by reproducing and transmitting terror and fear will terrorists bring about the

desired political change (Hoffman, 2006, 40). Cyberspace is the main means of mutating terrorism in modern times.

Dorothy Denning defines cyberterrorism as a very damaging attack based on the computer as a means, from which various types of attacks or collapse of state or non-state intelligence systems or to intimidate or force governments or societies in order to achieve social and political goals. Moreover, she mentions that in cyberterrorism, there is a convergence of cyberspace and terrorism, where the latter is the means of conducting a terrorist act (Denning, 2006, 2).

Compared with the traditional form of terrorism, cyberterrorism offers several advantages that we will be analyzed below. Traditional terrorists use common methods such as bombs, hijackings and murders (Saint Claire, 2011, 85), which require a lot of financial resources and possibly human losses.

Cyber-actors can have the same political, ideological and religious goals, but behave and act by different means (Heickero, 2014, 3). Gabriel Weimann (Weimann, 2004, 1) emphasizes that cyberterrorism has the potential to affect directly more people than traditional terrorist methods. It is not limited by geographical boundaries, there are no physical obstacles, no effective control and accessibility for future members is free. The actors do not threaten directly their lives as their actions do not take place in the real world, but in a digital world.

The members of the terrorist organizations have adopted a common language of communication, which makes cooperation and connection between them stable and unbroken (Zanini & Edwards, 2001, 7). These groups have become numerous, agile and well-coordinated, making it difficult to identify them. Cyberspace has the following characteristics that make it an ideal tool for terrorist organizations.

First of all, it facilitates the rapid communication and coordination of their actions, real-time talks and exchange of information (e.g creation of weapons and bombs, vital infrastructure, maps, plans, strategies, information for future targets) (Lachow & Richardson, 2007, 2). All of the above actions refer to "Cyber planning" (Timothy, 2003, 114-121), the digital co-ordination of an integrated plan beyond geographical borders that may or may not lead to bloodshed. Effective and asynchronous communication can be done through email. Many email accounts are free, therefore terrorists can hold multiple accounts simultaneously (Conway, 2014, 2).

The use of cyberspace is a means of low-cost and absorbs fewer resources than conventional terrorism. There is a phenomenon of collecting and transferring funds to support and expand the activities of terrorist organizations like Al Qaeda, Hamas, ISIS through the sale of dvd, cd, badges, books (Dean, 2012, 4).

Cyberspace offers small terrorist organizations a voice in order to attract and to communicate with new members. They can create a virtual relationship between the core of the organization and future members, spreading their ideology and goals through YouTube, Facebook, Instagram, Twitter, WhatsApp, Tumblr, Skype and Viber. As a result, terror can be transmitted in various forms and speed (Schori-Liang, 2015). ISIS is an active user of all the above social media tools.

Terrorist organizations use cyberspace to radicalize and recruit means. Cyberspace gives a direct control to the message's content, allowing them to edit the way they perceive themselves and their opponents and works as a "virtual training camp" since there are websites with information for attacks, manufacture of weapons, surveillance and hacking (Stenersen, 2008, 219),

Raising of bandwidth and developing of new software have allowed users to spread quickly information through cyberspace. They operate out of the groups' physical control zones, exploit new technologies and develop new technological solutions when communication and coordination is interrupted. It creates the right space for spreading rumors and fake news that cause more fear to recipients (Schmid & De Graaf, 1998). This fear is also related to the perception and information which the public has on terrorism as a problem (Nacos, 2006).

Cyberterrorism offers partial anonymity and identity protection by using pseudonyms. As a result, the identification and detection of terrorists becomes difficult (Weimann, 2004, 2). For example, the Jihadists use decentralized electronic networks to communicate with their members (Brantly, 2017, 81) and to maintain their identity safe.

The above combination of features makes cyberspace a remarkable strategic asset for terrorists. There is a gap in the field of cyberterrorism research, regarding how terrorists operate successfully without being identified and prevented by security services and intelligence agencies. This success could be attributed to their immediate adaptation to new technological changes.

### **3. The evolution of ISIS: Historical review**

The Islamic State (ISIS) has become active and innovative in cyberspace because of its intensive and systematic use<sup>1</sup>. This may have been surprising to the public, but it did not come as a surprise. As they organized their ground campaign in Syria and North Iraq, former members of the Al-Qaeda group had started an equally effective war of public relations conducting propaganda through social media.

In 2014, ISIS made a dynamic entry into the world, with the establishment of its religious power worldwide under a Caliphate. Caliph Abu Omar Al-Baghdadi proclaimed Caliphate and prayed to God that he would make "Islamic State: A Caliphate according to the prophetic method," that should be based on Islamic law and would try to strengthen and unite all Muslims in every part of its land (McCants, 2015).

The proclamation entitled Allah's promise was published in the twitter account of the Al-tisaam Media Foundation, which was translated in many languages and justifies the foundation of Caliphate (Evans, 2010, 11). The ultimate goal of Caliphate is to strengthen and expand influence and recruitment. The basic means of implementation is brutal violence, traditional tough Muslim views and the power of arms. It tried to create a climate of credibility and establish its legitimacy through cyber propaganda to attract future "Fighters of God" and intimidate the enemies of Islam.

Their ideology, Salafi-Jihadi, emphasizes first on the return to the original beliefs and practices of their pious ancestors (al-salaf al-salih), who were the comrades of the Prophet Muhammad (632), the followers of the comrades and their followers. The primary goal of Islamists is to regenerate the Islamic Ummah, to mobilize and expand the Islamic community against Western society (Bockstette, 2008). The establishment of a caliphate or Islamic state is the means of applying these beliefs and practices. Therefore, they impose their vision of Islam with faith and manifest action and support the conduct of jihad against pagan regimes that do not govern according to God's rules.

### **4. ISIS strategy in social media**

ISIS attracts international interest because of the smart use of mobile technology and social media to achieve its goals (Farwell, 2014, 50). The communication strategy aims to persuade all Islamic supporters that the struggle for the rehabilitation of Caliphate is a religious task. It is presented as a factor of change, a true apostle of a sovereign faith and a means of enforcing social justice. With these means it can successfully transmit powerful images of its relentless and fearless warriors such as beheadings and executions for intimidation and construct relevant narratives.

A turning point in ISIS growth was Halifi's first speech that addressed all Muslims for the establishment of the Caliphate. The recruitment of "mercenaries" reached unprecedented numbers (Al-Qarawee, 2015). Those who are usually convert into radical Islam, are mainly young people who perceive the situation in Syria and Iraq as an attack on their coreligionists and consider their obligation to defend them. They are unsatisfied with their lives and as outcasts, search for the meaning of life in the establishment of the Caliphate and are attracted by the military triumph of ISIS.

Islamic terrorists who have been in Europe in recent years are second-generation immigrants and have the following characteristics. They are usually not deeply religious, they have a criminal record, they originate from Muslim families and have embraced a radical version of Islam in prison. It is noteworthy that a strong religious identity prevents terrorism, therefore, we cannot attribute the rise of terrorism to religion. Thus, religion, as the sole motive to join a terrorist organization, is overestimated. They perceive ISIS as an idealized organization in which they will enjoy power, they will fight for a noble cause and develop a strong sense of belonging. There are

---

<sup>1</sup> It is estimated that 98% of territorial belonging to the Caliphate has been liberated by the jihadist yoke. American forces estimate that the Islamic State has been limited to 2% of the maximum of the territories it held from " Islamic State and the crisis in Iraq and Syria in maps" (28/3/2018) <https://www.bbc.com/news/world-middle-east-27838034>

different profiles of terrorists, such as young girls and boys, students, entrepreneurs, and family leaders (Scott & Spaniel, 2016, 35).

ISIS recruits members with an academic background in the fields of information technology, international relations, political science and communication studies. This enables ISIS to use modern techniques and strategies. Their narrative targets all Muslims regardless of nationality, sex and language. In addition, the increasing number of Western fighters reduces the feeling of alienation to the new members. These young hackers and computer-geeks have turned ISIS into a major cyber threat (Scott & Spaniel, 2016, 11). ISIS by exploiting Al Qaeda's communication heritage, by utilizing new communication technologies and the ability of its members to use online communication techniques, is practicing cyber jihad (Hoffman & Schweitzer, 2015).

ISIS aims two main target audiences and therefore differentiates its strategic communication, in terms of content and language. (Bockstette, 2008). The first audience consists of the aspiring fighters that ISIS is trying to manipulate and convert. The second one includes their enemies, the unbelievers and the western society that ISIS is trying to influence and prevent them from participating in opposing regimes (Rose, 2014).

## **5. ISIS actions in cyberspace**

ISIS's cyber toolkit includes among others, online videos, posts and hashtags. According to Watts, "they want to send a message back to their communities that they are involved", therefore their social media enables foreign observers to monitor their actions and strategy (Farwell, 2014, 50).

ISIS had created a 13-minute video titled "There is no life without jihad", that includes testimonies from British and Australian fighters who reject the Middle East borders after W.W I. This video strengthens their position in social media and inspires them to conduct jihad. Due to an Android and PC application titled "Dawn" which was available on Google Play Store from April till June 19, 2014, every time a twitter account closes, another account will appear within one minute (START Report, 2014). Thanks to "Dawn", members of ISIS can make their own posts and by tweeting the same hashtags, they can potentially radicalize thousands of users (Marks, 2014).

A hashtag is a way of creating groups of conversations in an uncontrolled discussion in a forum. ISIS used hashtags to communicate with a wider audience. This is exactly what happened with the 2014 World Cup. Exploited the crowd's love for football to spread through tweets containing hashtags such as #Brazil2014, #ENG, #France and #WC2014 to gain access to millions of searches and accounts in order to recruit new members (Farwell, 2014, 51).

According to CNN, from June 2014 to February 2017, ISIS conducted or inspired more than 140 terrorist attacks in 29 countries other than Iraq and Syria, which caused 2.043 deaths and thousands of injuries (Lister, Sanchez, Bixler, O'Key, Hogenmiller & Tawfeeq, 2017). An important factor in organizing and disseminating these achievements was the internet (Gordon, 2003, 11).

ISIS has combined its physical attacks with a series of cyberattacks. A recent example is that after the terrorist attack on the French newspaper Charlie Hebdo, cyberattacks were made on 19.000 French websites (Griffin, 2015). Prior to the terrorist attack on ISIS in France on November 13, 2015, many investigations had recorded its action in social media and mainly on twitter, namely 46.000 profiles of Islamic terrorists, while the actual number is 90.000 (Berger J.M. & Morgan J., 2015, p.7).

The objectives of the ISIS communication strategy have great interest. ISIS is not interested in demonstrating a leadership and organizational trend in terrorist actions against the West, but it wants to inspire all Muslims to act autonomously. Many unorganized terrorist groups act individually in the name of the Islamic State. The communication strategy of the organization is very focused on the activation and recruitment of Muslims<sup>2</sup>. As observed in "Flames of War", their propaganda contains messages against Western capitalism.

---

<sup>2</sup> Indicatively, in 2015, the organization numbered 25.000 supporters in Syria and Iraq, of which 4.500 are from Europe and North America and the rest come from 80-90 different countries, proving the geographical impact. Talbot D., «Fighting ISIS Online: The lonely efforts to counteract ISIS's mastery of social media», MIT Technology Review, 30/9/2015, <https://www.technologyreview.com/s/541801/fighting-isis-online>

The contents of their messages include captions of detainees, abductees and battlefield scenes. An example of this is the 22-minute video, in which a Royal Jordanian Air Force pilot prisoner, Muath Al Kasaesben, was burning in a cage or videos released on 23 June 2015 with spy executions, drowning in cages, blasting of passengers in a car and explosions of beheadings.

A strategic disadvantage is the fact that unauthorized individual fighters spread, via their smartphones, messages with violent content. Such material would provide support to the organization's opponents and could be used to discredit their overall effort (Dean, June, 2014). Indicative of the above is the fact that the State Department of America has created a video that hampers the organization's actions for recruiting as it shows its barbarity and suggests that potential members will face a gruesome death.

ISIS's communication strategy has been criticized by Al-Qaeda, which did not support images of Muslims killing Muslims, seeing them as counterproductive. The disclosure of non violent pictures of ISIS killers to embrace animals was a good trick, but the emotional impact of ISIS fighters' images showing the blood of Muslims and other innocents is likely to fail.

## **6. Cyber-terrorist groups of ISIS**

The unprecedented growth of ISIS in terms of know-how and technical support, was achieved by the contribution of supporting terrorist groups. These groups acted independently in cyberspace and chose different targets for cyberattacks, going against the core values of ISIS. Such attacks were not recognized by the organization, as it often led to conflicting messages. The most efficient pro-ISIS hacker groups are the following: Cyber Caliphate Army (CCA), United Caliphate Cyber (UCC) and Rabitat Al Ansar (League of Supporters).

Cyber Caliphate Army (CCA) was one of the first cyber-terrorist organizations to support the Islamic state and was represented by Junaid Hussain, who was a British hacker and made several cyberattacks (Alkhouri, Kassirer & Nixon, 2016, 3). One important action is that it has gained illegal access to a large number of websites, such as that of NATO, the British Ministry of National Defense and Facebook accounts (e.g. Mark Zuckerberg). Hussain became one of the leaders of Islamic Caliphate's propaganda and hired hackers whom he recruited to do cyberattacks and recruitment. The most significant cyberattacks reported, were at the twitter and YouTube accounts of the United States Central Command (US CENTCOM), where their user's profile pictures were changed with the images of a cover man and the message "Je suis ISIS".

Sons of Caliphate Army (SCA) was founded in 2016 as a subgroup of the Cyber Caliphate (Nance & Sampron, 2017, 69). Their most important action is the release of a video called "Flames of Ansar," which threatens Mark Zuckerberg founder of Facebook and Jack Dorsey founder of Twitter, who respectively contributed to the suspension of 10.000 Facebook accounts and 5.000 twitter accounts (Manusaga, 2016, February 23).

United Cyber Caliphate (UCC) was the most coordinated and basic team of ISIS<sup>3</sup>, created in 2016 by the merger of several hacking teams (Cyber Caliphate Army, Kalashnikov E -Security Team, Sons Caliphate Army) and was announced by a twitter account with the hashtags #CaliphateCyberArmy #SonsCaliphateArmy #KalashnikovTeam #Team #UnitedCyberCaliphate (Nance & Sampron, 2017, 74; Alkhouri, Kassirer & Nixon, 2016, 18). One of their most important actions is the campaign of #Gazwa:Reloaded, where on April 18, 2016 published 10.000 personal accounts. On May 2, 2016, the UCC published a list of 1.543 personal accounts of the most important crusaders in Texas (Memri Lab, 2016, May 3). On June 7, 2016, another list was released, naming 8.318 people from 21 countries that had to be killed in the name of the Islamic State (Nance & Sampron, 2017, 75). Some of these people are employees of IBM, Microsoft and Barclay Bank among others. Likewise, on July 21, they published private information on 2.461 New York's residents that had to be killed in the name of (Memri Lab, 2016, July 21).

Finally, Rabitat Al-Ansar, which was a part of larger pro-Islamic State media collective called the Media Front was not always known as a cyber unit. For almost one year, the group acted in support of ISIS as jihadi propaganda media unit, releasing articles and jihadi material (Alkhouri, Kassirer & Nixon, 2016, 12). The campaign also included English phrases and threats for attacks on U.S.A, trying to terrorize the citizens and also

<sup>3</sup> Nevertheless, there are speculations that the UCC may be a front of the Foreign Intelligence Service of the Russian Federation (SVR), so not really connected to ISIS. Allen I. (20/6/2016), " Islamic State's online army is a Russian front, says German intelligence" <https://intelnews.org/2016/06/20/01-1921/>

inviting them to Islam. Continuing its preparation for what was to come, the group also worked to mobilize ISIS's online supporters and translators, via designated Twitter accounts. This unit released videos containing actions and operations against the infidels and US forces in Iraq, translated texts in Arabic for education and recruitment purposes. In 2015, cyber-attacks against US accounts began, with the hashtag #WeWillBurnUsAgain, which became 15.000 times retweeted (Bora, 2015, November 4).

## **7. Countering ISIS**

Countering ISIS requires a combination of efforts at both the national and international level. The EU has undertaken strong legislation regarding censorship. The latter is viewed as a key tool to fight extremism, but adds little to the fight against terrorism. Terrorists resemble hydra<sup>4</sup>; cut off one head and two more shall take its place. The case of ISIS had become one of the most socially-mediated conflicts. ISIS is the first terrorist organization that globalized terrorism by enabling its followers to join "electronic jihad" and the "digital caliphate". Current technologies include firewalls, password protection systems, key encryption, intrusion detection systems and access control lists. States and international organizations are gradually realizing the need to construct a counter-narrative.

States are slow to embrace change and any national regulation will most likely be vague, focusing on existing companies rather than on future technologies. A censorship-based approach will force terrorists to explore new technologies or develop counter-measures to escape identification (Isaacson, 2018, p.2). One example of countering ISIS, is the Center for Strategic Counterterrorism Communications (CSCC) of the State Department, which is publicly highlighting misinformation by exposing hypocrisy behind the ISIS rhetoric. The State Department uses twitter accounts in English, Arabic and other languages to identify terrorists and find tactics and strategies (Plett Usher, 2014).

Preventing the dissemination of terrorist content will not prevent terrorist organizations from creating propaganda, terrorists will search for new outlets or means to spread the content as the states tries to remove it. Censorship alone is insufficient to prevent supporters from accessing content. The practicing of censorship may have a boomerang effect. States will frame the political opponents as terrorists and thereby delegitimize them.

## **8. Conclusion**

In recent years the battlefield has changed. It has become a global battlespace with no geographical boundaries between its opponents. Terrorist groups such as Al Qaeda, Hezbollah and recently ISIS have digitalized their way of communication, which allows them to operate outside the national territories. Also, they are keen on exploiting every available technological capability, as well as their opponent's shortcomings.

Social media have a dual function. On the one hand, they facilitate the social interaction between people who create, share or exchange information and ideas within virtual communities and networks. On the other hand, there is a dark side of social media, where user's information serve indirectly and directly the actions of terrorist organizations. User's profiles function as a gigantic database that supplies terrorists with the necessary information for recruitment and radicalization. The only thing that separates a terrorist organization from a potential member is a friend request, a hashtag or following the friend or the group. Platforms like Facebook and Twitter have begun a policy of banning terrorist's accounts, which is insufficient.

It is oxymoron that ISIS, despite its harsh religious beliefs that envisions a traditional Islam - which probably never existed - has nevertheless adopted a strategy that fully utilizes the opportunities offered by online media for the purposes of propaganda, recruitment, communication and education, bypassing conventional methods. Thus, the organization tries to protect its identity and ideology from the use of cyberspace, but at the same time uses the latter as a means of achieving its goals.

Terrorists will never stop using traditional ways and means. Cyberterrorism has not replaced conventional terrorism. On the contrary, the former acts as a facilitator for the development and expansion of the latter. It could be argued that Islamic terrorism has gone through a "metamorphosis" process. Cyberspace modernized terrorist organizations and enabled them to act with greater safety and anonymity (Green, 2015). ISIS is certainly not the first extremist organization following the wave of modernization and the use of new technologies to

---

<sup>4</sup> The Lernaean Hydra or Hydra of Lerna, more often known simply as the Hydra, is a serpentine water monster in Greek mythology.

draw everyone's attention. Modern terrorists have weaponized social media via their keyboards. ISIS's goals, know-how, strategy and methods have been left behind as a legacy for the present and future terrorist groups.

A thorough study of ISIS's social media strategy could be a means of deterring and countering likeminded terrorist organizations. Countering Islamic terrorism will require focused efforts in order to defame and disorganize the terrorist groups among Muslims, while simultaneously taking actions to eliminate terrorism. This will be successful only if both public and private sector cooperate and develop specific measures that aim to weaken terrorist actions and by adopting a coherent strategy narrative.

## **Acknowledgements**

This work has been partly supported by the University of Piraeus Research Center.

## **References**

- Alkhouri L., Kassirer A. & Nixon A., (2016), *Hacking for ISIS: The emergent cyber threat landscape*, Flashpoint
- Al-Qarawee, H.H., (2015), *The Discourse of ISIS: Messages, propaganda and indoctrination* at Maggioni M. & Magri P., *Twitter and Jihad: the communication strategy of ISIS*, Milano, ISPI
- Allen I. (20/6/2016), "Islamic State's online army is a Russian front, says German intelligence"  
<https://intelnews.org/2016/06/20/01-1921/>
- Berger J.M, (June, 2014), *How ISIS games Twitter*, The Atlantic,  
<https://www.theatlantic.com/international/archive/2014/06/isis-iraq-twitter-social-media-strategy/372856/>
- Berger J.M & Morgan J. (2015), *The ISIS Twitter Census Defining and describing the population of ISIS supporters on Twitter, The Brookings Project on U.S. Relations with the Islamic World*
- Bockstette C., (2008), *Jihadist Terrorist use of strategic communication management techniques*, George C. Marshall Center, Garmisch-Partenkirchen
- Bora K, (2015, November 4), *ISIS Supporters Start #WeWillBurnUSA Again Twitter Campaign of Threats, Referring To 9/11 Attacks»*, International Business Times, <https://www.ibtimes.com/isis-supporters-start-wewillburnusagain-twitter-campaign-threats-referring-911-1878317>
- Brantly A., (2017), *Innovation and Adaptation in Jihadist Digital Security*, Survival
- Bryant R. (2001), *What kind of space is cyberspace?*, Minerva- An Internet Journal of Philosophy,  
<http://www.minerva.mic.ul.ie//vol5/cyberspace.html>
- Clausewitz V., (1993), *On War*, Everyman's Library, 1993
- Collins B., (1997), *The Future of Cyberterrorism*, Crime and Justice International <http://www.crime-research.org/library/Cyberter.htm>
- Conway, M., (2014), *Reality Check: Assessing the (Un) likelihood of Cyberterrorism*. In T.M. Chen, *Cyberterrorism: Understanding, Assessment and Response*, New York, NY: Springer
- Dean G., (2012), *The Dark side of social media, review of online terrorism*, Griffith University
- Denning D.E, (2006), *A view of cyberterrorism five years later*, Calhoun: The NPS Institutional Archive
- Dowling R.E, (1986), *Terrorism and the media: a rhetorical genre*, Journal of Communication, Willey Online Library
- Evans R., (2010), *From Iraq to Yemen: Al -Qaida's Shifting Strategies*, CTC Sentinel
- Farwell J.P, (2014), *The Media Strategy of ISIS*, Survival, Vol 56, No. 6
- Gordon S., (2003), *Cyberterrorism?*, Symantec Security Response
- Greene K., (2015), *ISIS: Trends in terrorist media and propaganda*, Cedarville University  
[https://digitalcommons.cedarville.edu/international\\_studies\\_capstones/3/ Accessed](https://digitalcommons.cedarville.edu/international_studies_capstones/3/)
- Griffin A., (2015, January 15), *Charlie Hebdo: France hit by 19.000 cyberattacks since Paris shootings in unprecedented hacking onslaught*, Independent <https://www.independent.co.uk/life-style/gadgets-and-tech/news/charlie-hebdo-france-hit-by-19000-cyberattacks-since-paris-shootings-in-unprecedented-hacking-9980634.html>
- Heickero R., (2014), *Cyber Terrorism, electronic Jihad*, Strategic Analysis
- Hoffman A. & Schweitzer Y., (2015), *Cyber Jihad in the service of the Islamic State*, Strategic Assessment, Vol. 18
- Hoffman B. (1998), *Inside Terrorism* Columbia University Press, 1998
- "Islamic State and the crisis in Iraq and Syria in maps" (28/3/2018), <https://www.bbc.com/news/world-middle-east-27838034>
- Isacson Z. (2018, September 2), "Combating Terrorism Online: Possible actors and their roles", Lawfare  
<https://www.lawfareblog.com/combating-terrorism-online-possible-actors-and-their-roles>
- Kemp S. (2018), *Digital in 2018: World's Internet users pass the 4 billion mark*, Blog: We are social,  
<https://wearesocial.com/blog/2018/01/global-digital-report-2018>
- Lachow I. & Richardson C., (2007), *Terrorist use of the internet: the real story*, National Defense UNIV Washington D.C
- Lister T., Sanchez R., Bixler M., O'Key S., Hogenmiller M. & Tawfeeq M., (2017), *ISIS goes global: 143 attacks in 29 countries have killed 2,043*, CNN, <https://edition.cnn.com/2015/12/17/world/mapping-isis-attacks-around-the-world/>
- Manusaga S, (2016, February 23), *Apple vs FBI: Bill Gates, Mark Zuckerberg, John McAfee and more taking stands*, Times, Los Angeles

**Eleni Kapsokoli**

- Marks P., (June 25, 2014), *How ISIS is winning the online war for Iraq*, New Scientist, <https://www.newscientist.com/article/dn25788-how-isis-is-winning-the-online-war-for-iraq/>
- McCants W., (2015), *The ISIS Apocalypse: The History, Strategy, and Doomsday Vision of the Islamic State*, St. Martin's Press
- MEMRI CYBER & JIHAD LAB, (2016, May 3), *In Kill List On Telegram, Pro-ISIS Hacking Group Posts Names Of 'Most Important Crusaders In Texas'*, <http://cjlabs.memri.org/lab-projects/monitoring-jihadi-and-hacktivist-activity/pro-isis-hacking-group-releases-kill-list-with-2461-nyc-residents-personal-details/>
- MEMRI CYBER & JIHAD LAB, (2016, July 21), *Pro ISIS Hacking Group releases kill list with 2,461 NYC Residents Personal Details*, <http://cjlabs.memri.org/lab-projects/monitoring-jihadi-and-hacktivist-activity/pro-isis-hacking-group-releases-kill-list-with-2461-nyc-residents-personal-details/>
- Nacos B.L, (2006), *Terrorism and media in the age of global communication*, Hamilton D.S (Edit), *Terrorism and International Relations*, Center for Translatic Relations, Washington D.C
- Nance M. & Sampron C., (2017), *Hacking ISIS: the war to kill the cyber Jihad*, Skyhorse Publishing, New York
- Plett Usher B., (2014, November 28), "The U.S. state department's YouTube 'digital jihad,'" BBC News Magazine, <http://www.bbc.com/news/magazine-30045038>
- Rose S., (2014), *The ISIS propaganda war: a hi-tech media jihad*, The Guardian, <https://www.theguardian.com/world/2014/oct/07/isis-media-machine-propaganda-war>
- Saint Claire S., (2011), *Overview and Analysis on Cyber Terrorism*, School of Doctoral Studies, European Union
- Schorri-Liang C., (2015), *Cyber Jihad: Understanding and countering Islamic State Propaganda*, GCSP Policy Paper
- Schmid A.P & De Graaf J., (1998), *Violence as communication-insurgent terrorism and the western news media*, Sage Publications
- Scott J. & Spaniel D., (2016), *The Anatomy of Cyber Jihad: Cyberspace is the new great equalizer*, Institute for Critical Infrastructure Technology
- START Report, (2014), *The Islamic state of Iraq and Levant: Branding, leadership culture and lethal attraction*, National Consortium for the study of terrorism and responses to terrorism, A department of homeland security science and technology center of excellence
- Stenersen A., (2008), *The Internet: A Virtual Training Camp?*, Terrorism and Political Violence
- Threat tactics report: Islamic State of Iraq and the Levant (2014) TRADOC G-2 ACE Threats Integration Intelligence Community Directive Number 203: Analytical Standards
- Timothy T., (2003), *Al Qaeda and the Internet: The Danger of Cyber planning*, Spring
- Torres Soriano M.R (2008), *Terrorism and the mass media after al Qaeda: a change of course*, Athena Intelligence Journal
- Weimann G., (2004), *www.terror.net: How Modern Terrorism Uses the Internet*, United States Institute of Peace
- Zanini M. & Edwards S., (2001), *The networking of terror in the information age* from the book *Networks and Netwars: The Future of Terror, Crime, and Militancy* by John Arquilla & David Ronfeldt, RAND

# Protecting the Besieged Cyber Fortress: Russia's Response to Cyber Threats

Martti Kari

University of Jyväskylä, Finland

[martti.j.kari@jyu.fi](mailto:martti.j.kari@jyu.fi)

**Abstract:** The Information Security Doctrine of the Russian Federation (RF) defines the threat to information security as a complex of actions and factors that represent a danger to Russia in the information space. These threats can be information-psychological (i.e., when the adversary tries to influence a person's mind) or information-technical (i.e., when the object of influence is the information infrastructure). The information infrastructure of the RF is a combination of information systems, websites, and communication networks located in the territory of the RF, or those used as part of international treaties signed by the RF. A cyber threat is an illegal penetration or threat of penetration by an internal or external actor into the information infrastructure of the RF to achieve political, social, or other goals. Cyber threats against Russia are increasing and becoming more diverse. The Russian assessment of the cyber threat contains the same besieged fortress narrative as the country's other threat assessments do. In this narrative, Russia is surrounded by hostile states and non-state actors in cyberspace. The sources of the cyber threat are Western intelligence services, terrorists, extremist movements, and criminals. To protect itself against cyber threats, Russia is increasing its digital sovereignty by preparing to isolate the Russian segment of the Internet, RUNET, from the global Internet. Russia is also improving the protection of its critical information infrastructure. To protect itself against cyber threats but also to monitor the opposition, Russia has increased surveillance of RUNET and banned user anonymity. Russia is also making an effort to replace imported information and communication technology (ICT) with Russian production. This paper discusses Russia's defense against cyber threats. After the introduction, the paper begins with a description of the Russian cyber threat perception. The main section then discusses Russia's response to this threat. This study uses grounded theory, an appropriate method for this subject because little theoretical and structured information has, to date, been published on the Russian response to cyber threats. The study data are drawn from official Russian documents such as strategies, doctrines, laws, and presidential decrees.

**Keywords:** Russia, cyber threat, cyber defense, cyberspace

## 1. Introduction

According to Russian authorities, the formation of cyberspace as a domain of warfare poses a threat to the Russian Federation's (RF) national interests (PP-2796, 2014) in the information space. According to the Doctrine of Information Security of the RF, *Information space* is a complex of information, objects of informatization, information systems, networks, and information technology. *Informatization* refers to social, economic, and technical processes for adopting and expanding information technology in society and throughout the country as well as to secure access to information resources. Information space includes subjects creating, generating, and processing information; subjects developing and using information technology; or subjects managing information security. It also includes mechanisms regulating the information relations in society. (UP-646, 2016.)

The threat to information security has two dimensions. First, it can be *information-psychological*, which is aimed at influencing the human mind, including its moral and intellectual world, social policy, psychological orientation, and the ability to make decisions. Second, the threat can be *information-technological*, which influences information technology systems (Kamyshev, 2009). The Russian concept of the information-technological threat corresponds to the Western concept of cyber threat. According to the Russian definition, cyberspace<sup>1</sup> is a limited part of the information space. Cyberspace is an environment formed by a set of communication channels on the Internet and other networks, the technological infrastructure that ensures their functioning, and any form of human activity carried out through their use. A cyber threat to Russia is an illegal penetration or threat of penetration by an internal or external actor into the information infrastructure of the RF to achieve political, social, or other goals. Cyber security is a complex of conditions under which all components of cyberspace are protected from all threats and undesirable impacts (SBRF, 2013b).

The increased interest in cyberspace as a domain of warfare has also heightened the need for theoretical studies to assess the cyber threat perceptions of different states and their responses to these threats. Although much non-academic information has been published about Russian offensive cyber capabilities and operations, only a limited amount of information has been published about the country's cyber threat scenarios and defensive

<sup>1</sup> киберпространство

cyber capabilities. However, there is enough information in official Russian legal documents to collect at least a satisfactory picture of the Russian perception of cyber threats and Russia's response to those threats. To protect itself against cyber threats, Russia is increasing its digital sovereignty by preparing to isolate the Russian segment of the Internet, RUNET, from the global Internet. Russia is also improving the protection of its critical information infrastructure. As a further means of protection against cyber threats but also as a way to monitor the opposition, Russia has increased surveillance of RUNET and banned user anonymity. In addition, Russia is making an effort to replace imported information and communication technology (ICT) with Russian production.

This paper examines Russia's defense against cyber threats. After the introduction, there is a description of the Russian cyber threat perception. The main section then discusses Russia's response to this threat. This study uses grounded theory, an appropriate method for this subject because little theoretical and structured information has, to date, been published on the Russian response to cyber threats. The study data are drawn from official Russian documents such as strategies, doctrines, laws, and presidential decrees.

## **2. Russian cyber threat assessment**

The National Security Strategy of the Russian Federation (UP-683, 2015) describes the world as polycentric, where the use of force in international politics is increasing. The West tries to maintain its position by containing Russia (UP-640, 2016). This confrontation between Russia and the West has extended to the information space as well because Western countries are using ICT against Russia to achieve their geopolitical goals (UP-683, 2015). The Kremlin sees the international arena as a battlefield, where the battle to disrupt Russia's digital sovereignty is waged every day (Sinovets, 2016). Digital sovereignty<sup>2</sup> means Russia's rights independently determine internal and geopolitical interests in the digital space (Yarovaya, 2013). Russian national interests – such as sovereignty, territorial integrity and constitutional order – are threatened through cyberspace by Western states, but also by terrorists and criminals. Western countries' preparations for information warfare and aspirations to change cyberspace into a war zone threaten Russia's strategic interests in the cyber environment (UP-646, 2016).

President Putin (2016) has stated that because of the risks inherent to digitalization, Russia has had to strengthen its defenses against cyber threats targeted, for example, at Russian infrastructure, the country's financial system, and the state's leadership and management. The aim of the United States is "to destroy strategic balance, to change the balance of power in such a way not just to dominate but to dictate their will to anyone" (Putin, 2015). The USA uses its technological superiority to dominate the information space (UP-646, 2016).

According to President Putin, the Soviet Union was a besieged fortress constantly under threat of attack by the West (Aron, 2008). After the annexation of Crimea, Kremlin's besieged fortress narrative has become one of the primary means for Putin's regime to maintain power (Kolesnikov, 2016). The besieged fortress view can also be seen in Russia's cyber threat perception, in which Russia describes itself as a besieged fortress in cyberspace. The number and severity of dangers and threats have increased in cyberspace, and those threats are shifting to the internal sphere of the RF (PP-2796, 2014). Vladislav Surkov, the First Deputy of Russian Presidential Administration from 1999 to 2011 and one of the main ideologists of the Kremlin, highlighted internal threats and stated in 2004 that "the enemy is at the gate, and not only at the gate because in the besieged fortress there is a fifth column...sponsored by foreign states" (Ovtsharenko, 2004).

The Military Doctrine of Russia (PP-2796, 2014) defines military danger as interstate or internal relations characterized by a combination of factors that can, under certain conditions, lead to a military threat. Such a threat can emerge in these relations when there is a real possibility of the emergence of military conflict between the opposing parties or by the high degree of readiness of a state, a coalition of states or separatist or terrorist organizations to use military force or armed violence. According to the Military Doctrine, military dangers and military threats are expanding to the information space as well as to the internal sphere of the RF. In modern conflicts, information warfare is used as a part of warfare and the enemy is impacted throughout their entire area of operation, including the global information space (PP-2796, 2014).

The Information Security Doctrine of Russia (UP-646, 2016) includes the same visions of an aggressive West discussed in the National Security Strategy and Military Doctrine. Some states are using their technological superiority to dominate the information sphere and to achieve military and political goals. An unbalanced division

---

<sup>2</sup> For more on Russian Digital Sovereignty, see Kukkola, Ristolainen & Nikkarila, 2017

of responsibilities in running the Internet between the states increases this technological superiority. This prevents the safe functioning of RUNET, because actors outside Russia can block Russia's access to the Internet and destabilize the functioning of RUNET (SBRF, 2012).

The targets of cyber threats in Russian threat perception can be divided into four categories: the national interests of the RF, the information resources of the RF, the information infrastructure of the RF, and the Russian Armed Forces. The national interests of the RF are the inviolability of its constitutional order, sovereignty, independence, national and territorial integrity, and consolidating the RF's status as a leading world power (UP-640, 2016).

One of the threats to Russian national interests in cyberspace is a lack of competitive ICT and the inadequate use of information technology in the production and research and development of future technologies. This technological backwardness in ICT has created a dependence on foreign information technology. Such underdevelopment weakens Russia's cyber defenses, facilitates cyber intelligence operations in Russia, and gives Western special services an opportunity to influence Russia's information resources (UP-683, 2015; UP-646, 2016). The use of foreign ICT challenges Russia's information security management.

The Draft of the Information Security Doctrine 2015 stated that Russia is lagging behind the leading foreign states in the development of competitive information technology, including supercomputers (PUP-1, 2015). In 2013, Russia was at least three to five years behind the USA in ICT (Eliseev, 2013) and five-and-a-half years behind the USA in supercomputing technology (Moukin, 2013). This technological inferiority strengthens the Russian perception of its strategic vulnerability in cyberspace.

The exploitation of cyberspace by foreign intelligence services against Russia and the possibility of cyberspace attacks on the Russian information resource and information infrastructure have increased. Attacks against objects of its critical information infrastructure are becoming more complex, more frequent, more coordinated (UP-646, 2016), and these attacks can have a destructive impact on the infrastructure. Terrorists and extremists are among those creating means to have this kind of destructive impact (UP-203, 2017). These threats can result in a loss of control, the destruction of infrastructure, irreversible negative change (or destruction) of the economy of the country or an administrative-territorial unit or a significant, long-term deterioration in the safety of the population living in these territories (SBRF, 2012b).

Foreign special services, terrorist organizations, and extremist movements are also targeting the information infrastructure and information resources of the Russian Armed Forces (PP-2796, 2014). The main targets of possible cyberspace exploitation and attacks include strategic missile warning and defense systems, air and space defense forces, and strategic missile forces. Attackers may try to weaken the defense capability of these strategically important systems and forces (SBRF, 2013b; PP-2796, 2014). During a pre-war period and in the first phase of any hostilities, the mobilization of the Russian Armed Forces and the deployment of wartime troops to operational areas are potential targets of cyberspace attacks. The logistical systems supporting mobilization and strategic deployment would also be targets of cyberspace attacks before the outbreak of a war (SBRF, 2012; PP-2796, 2014).

### **3. Defense against cyber threats**

The main means of Russian response to cyber threats are improved protection of the critical information infrastructure of the Russian Federation (CIIRF), a pivot to digital sovereignty by isolating RUNET from the global Internet, increased surveillance of RUNET, banning user anonymity online and the replacement of ICT imports with Russia's own ICT production.

One of Russia's national interests in the information sphere is to ensure the sustainable and uninterrupted functioning of the CIIRF (UP-646, 2016). The concept of the CIIRF was discussed already in the Russian Information Security Doctrine in 2000, hereinafter ISD 2000 (PP-1895, 2000). ISD 2000 started to debate the protection of the CIIRF, about which the core question has been the roles and responsibilities of different state authorities in information security (IS) management of the CIIRF. After ISD 2000, the protection of the CIIRF took almost two decades to organize because of the power struggle over IS management between the Federation Security Service (FSB), the Federal Service for Technical and Export Control (FSTEC), and the Russian Armed Forces, and because of the clarification of the responsibilities of private companies and other legal entities for protection.

In 2013, President Putin signed a decree on the creation of a state system for detecting, preventing, and eliminating the consequences of computer attacks on the information resources of the Russian Federation, hereinafter the GosSOPKA<sup>3</sup> Decree (UP-31, 2013). The GosSOPKA system is a combined, territorially distributed complex that includes authorities and means for detecting, preventing and eliminating the consequences of computer attacks on the CIIRF as well as for responding to other incidents. The GosSOPKA Decree of 2013 assigned the IS management related to cyberattacks to the FSB, but the question of the comprehensive protection of the CIIRF remained unresolved until the CII Security Law in 2017. After two drafts of a law for the security of the CIIRF, one in 2006 and the other in 2013, President Putin signed the Law on the Security of the Critical Information Infrastructure of the Russian Federation (FZ-187, 2017), hereinafter the CII Security Law, in July 2017. Its purpose is to define the CIIRF along with the organizational and legal basis of the IS management of the CIIRF to ensure its stable functioning when targeted by computer attacks.<sup>4</sup>

The critical information infrastructure of the Russian Federation (CIIRF) includes objects of critical information infrastructure as well as the telecommunication networks used to organize the interaction of these objects. The objects of the CIIRF are information systems, information and telecommunication networks, and automatic control systems operating in the following sectors: defense, healthcare, transport, communications, credit and finance, energy and fuel, nuclear, rocket and aerospace, mining, metallurgical, and chemical. The threats to the CIIRF include unauthorized access, destruction, modification, blocking, copying, provision, and dissemination of information about an object of the CIIRF (FZ-187, 2017).

In December 2017, it was confirmed that the FSB, which was tasked to create the GosSOPKA system in 2013, would also be the authority to operate GosSOPKA (UP-620, 2017). The processes implemented in the GosSOPKA framework are detecting, attributing, and responding to computer attacks; eliminating the consequences of computer attacks on the information resources of the RF; assessing the IS management situation and cyber threats; and the collection and analysis of information about computer attacks and computer incidents (SBRF, 2014; UP-620, 2017).

The FSB established and operates the National Coordination Center for Computer Incidents (NCCCI) and regional and territorial IS operations centers (SOC). The GosSOPKA SOCs will be established in the Russian Federation on the federal district<sup>5</sup> as well as the subject level.<sup>6</sup> The SOCs can be operated by the FSB, or they can be departmental or corporative SOCs. The common tasks of SOCs include collecting and analyzing information about computer attacks and computer incidents, responding to threats, and eliminating the consequences of computer incidents in information resources (UP-31, 2013).

The Federal Service for Technical and Export Control of the Russian Federation (FSTEC) is a federal executive body charged with ensuring the security of the CIIRF, countering technical intelligence, and the technical protection of information as well as a specially authorized body in the field of export control (UP-569, 2017). The identification and categorization of the objects of the CIIRF are the first steps in the process of securing and protecting it. The categorization of these objects is a process during which a subject in the CIIRF evaluates and categorizes the significance of a CII object according to the instructions of the FSTEC. Significant objects are placed into Category I, II or III. The categorization (i.e., the assigning of a category number to each object) is based on the social, political, economic, and environmental significance of the object for ensuring the country's defense, state security, and law and order. Category I is for the CIIRF's most significant objects.

After the categorization, the FSTEC specifies requirements to ensure the security of critical CIIRF objects as well as requirements to establish security systems and ensure the functioning of these objects. The FSTEC also includes requirements to ensure the security of information and telecommunications networks which are assigned to one of the three categories of significance and which, in cooperation with the Ministry of Telecom and Mass

---

<sup>3</sup> GosSOPKA is an abbreviation of the Russian phrase "state system for detecting, preventing and eliminating the consequences of computer attacks."

<sup>4</sup> A *computer attack* is defined as the targeting of software and/or hardware in CII facilities (i.e., the telecommunication networks used to organize the interaction of such objects), with a view to violating and/or terminating their operation and/or creating a security risk that is handled by such objects information.

<sup>5</sup> A federal district is a grouping of the federal subjects for governing by federal governmental agencies. There are eight federal districts in Russian Federation.

<sup>6</sup> The subjects of the Russian Federation are the main administrative divisions in Russia.

Communications of the Russian Federation, are included in the registry of significant CIIRF objects. For the banking and finance sector, the FSTEC sets requirements in consultation with the Central Bank of the Russian Federation. The subject of the CIIRF is obliged to follow FSTEC instructions and establish security arrangements corresponding to the CIIRF object's category of significance. The FSTEC is authorized to evaluate the security arrangements of the objects included in the registry (FZ-187, 2017).

The Kremlin considers digital sovereignty one of the country's main national interests in cyberspace. To secure digital sovereignty, Russia is developing RUNET, a national system of the Internet (UP-646, 2016), the functioning of which should be stable and safe in peacetime, in the event of a direct threat of aggression, and in wartime (UP-646, 2016). This entails that it would be possible to disconnect RUNET from the global Internet (Eliseev, 2013). The Ministry of Communications' Information Society program aims to have 99% of RUNET traffic transferred inside Russian borders by 2020. Part of this plan is to duplicate 99% of RUNET's critical infrastructure within Russia (Meduza, 2016).

In December 2018, the State Duma started to discuss draft legislation to improve Russia's digital sovereignty and to ensure the sustainable operation of RUNET in the case of cyberattacks and other aggressive actions from abroad. The draft names the United States as Russia's main cyber threat and states that Russia must take measures to secure the long-term and stable functioning of RUNET and to improve the reliability of Russia's Internet resources (PZF 608767-7, 2018).

The idea of the draft is to create a Russian national system for .ru and .rf domains, and develop a Russian IP-routing system in a way that a minimum amount of Russian Internet traffic would cross the Russian border and be transferred through foreign exchange points and servers outside Russian borders. The Federal Service for Supervision of Communications, Information Technology and Mass Media (Roskomnadzor) develops requirements and rules for actors that run or maintain the Internet in Russia. These actors are Internet providers, the owners of communication lines that cross Russia's national borders, the owners of technological communication networks, the owners of anonymous system numbers, and the owners of traffic exchange points (PZF 608767-7, 2018). Russian Internet providers are required to install technical equipment to counter threats to the RUNET. With this equipment, Roskomnadzor would block banned online resources in Russia and monitor compliance with the new traffic routing rules and the use of the new national domain name system. New monitoring equipment would be provided to Internet service providers (ISP) free of charge, subsidized by Roskomnadzor and the Digital Society program.

Roskomnadzor will establish a traffic-exchange registry. Service providers and companies would be forbidden from using Internet exchange points that are not on the registry. The exchange points would be banned from connecting to companies that do not comply with regulations and rules on the use of the Internet. Roskomnadzor will establish a federal agency called the Center for Monitoring and Managing Public Communication Networks. The tasks of this center are to control Internet regulations, collecting information from Russian companies about, for example, their network infrastructures, and their IP addresses, operating the internet exchange registry, and adjusting the country's traffic routing. According to the draft, the system's efficiency will be checked and improved through regular exercises, participation in which would be mandatory (PZF 608767-7, 2018).

The Russian Armed Forces have their own military intranet, which is a closed IT network specially protected against external cyberattacks. This intranet is called the Closed Data Transmission Segment (CDTS)<sup>7</sup> and it is not connected to the global Internet. The computers of CDTS are protected against, for example, connections by uncertified USB drives and external hard drives. The system has its own e-mail service, which allows the transfer of sensitive information, including secret and top secret documents (Tass, 2016).

Increased surveillance of RUNET is part of the RF's struggle against internal threats. The FSB has a mandate to monitor RUNET traffic. The tool for FSB Internet surveillance is the System for Operative Investigative Activities (SORM).<sup>8</sup> Since the 1990s, the operational capabilities of SORM systems have been improved from SORM 1 to SORM 3. SORM 1 collected mobile and fixed line telephone calls. SORM 2 began collecting Internet traffic. SORM

---

<sup>7</sup> Закрытый сегмент передачи данны (ЗСДП)

<sup>8</sup> Система технических средств для обеспечения функций оперативно-розыскных мероприятий

3 collects all kinds of communication on social networks, Wi-Fi, e-mails, Internet traffic, mobile calls, and voice-over-Internet. SORM 3 was introduced into operative use in 2014 (Soldatov and Borogan, 2015). ISPs are required to provide the FSB with statistics on all Internet traffic that passes through their servers. ISPs are also required to install SORM devices on their servers, routing all transmissions in real time through the FSB's local offices (PP-538, 2005).

Two laws were signed in 2017 to ban user anonymity on RUNET. Owners of virtual private network (VPN) services and Internet anonymizers are prohibited from providing access to websites banned in Russia. Roskomnadzor has authorization to block sites that provide instructions on how to circumvent government blocking (FZ-276, 2017). Companies registered in Russia as "organizers of information dissemination," including online messaging applications, are prohibited from allowing unidentified users. Those companies are required to identify their users by their cell phone numbers, and the government is tasked with elaborating the identification procedure. Mobile applications that fail to comply with requirements to restrict anonymous accounts will be blocked in Russia (FZ-241, 2017).

The information security of Russia is characterized by a lack of competitive information technology. The level of dependence of Russian industry on western ICT is high. One of the ways to correct Russia's technical backwardness in ICT and protect it against cyber threats is to develop the country's own IT sector by improving its research, development, and production of information (UP-646, 2016). To improve the security of its information infrastructure, Russia has to replace imported ICT software and equipment with Russian-made counterparts and lay the foundation for technological independence in ICT production (UP-203, 2017). President Putin (2018) stated that Russia needs to build its own digital platforms, ones that should be compatible with the global information space. The ISD 2000 (PP-189, 2000) had already identified the backwardness of Russian ICT as one of the main threats to the country's information security. Over the past decade, however, Russia has not managed to reduce the lead of Western countries in this area.

#### **4. Conclusion**

The Russian assessment of the cyber threat against it contains the same besieged fortress narrative as the country's other threat assessments do. Hostile state and non-state actors are surrounding Russia in cyberspace and cyber threats against the country are increasing and becoming more diverse. To protect itself against these cyber threats, Russia has taken operational, technical, and legal actions. The most important of these are improved protection of the CIIRF, preparations to isolate RUNET from the global Internet, intensified surveillance and the ban of user anonymity on RUNET, and the aspiration to replace imported ICT with Russian-produced ICT.

Russia is also making significant efforts to increase its digital sovereignty. It is possible that Russia will manage to create technical and operational readiness to at least partly isolate RUNET from the global Internet by the end of 2020. Russia is also improving the protection of its critical information infrastructure. The definition of the CIIRF and the division of responsibilities between authorities to protect it were confirmed by legislation in 2017 and the implementation phase has now started. The National Coordination Center for Computer Incidents (NCCCI), along with part of the regional and territorial IS operations centers, are now operational.

For Russia, the most difficult question in responding to cyber threats is that the country is lagging behind the leading foreign countries in the development of competitive information technology, including supercomputers, and this gap strengthens the Russian perception of its strategic vulnerability in cyberspace. For almost twenty years, Russia has tried, without success, to replace imported ICT software with Russian-made counterparts, and it seems that they will not succeed in the near future either. Russia is attempting to compensate for this lack mainly by isolating RUNET and by protecting the CIIRF.

#### **References**

- Eliseev I (2013) I shot digital cannon. Rossiyskaya Gazeta No 6085 (109) May 23. (in Russian)  
<https://rg.ru/2013/05/23/ashmanov.html>
- FZ-187 (2017). Federation Law of the RF 187 on the Security of Critical Information Infrastructure of the Russian Federation. (in Russian), <https://rg.ru/2017/07/31/bezopasnost-dok.html>
- FZ-241 (2017) Federal Law of the RF 241 "On Amendments to Articles 101 and 154 of the Federal Law" On Information, Information Technologies and Information Protection" (in Russian) <https://rg.ru/2017/08/04/informacia-dok.html>
- Meduza (2016) Russia's Communications Ministry plans to isolate the RuNet by 2020. May 13, 2016. Available at:  
<https://meduza.io/en/news/2016/05/13/communications-ministry-plans-to-isolate-runet-by-2020>

- PZF 608767-7 (2018) Draft of Law On Amendments to Certain Legislative Acts of the Russian Federation. (in Russian)  
<http://www.lexfeed.ru/law/608767-7>
- Kamyshev, E. (2009). *Информационная безопасность и защита информации*. Information Security and Protection of Information, (in Russia), <http://window.edu.ru/resource/033/75033/files/InfoBesop.pdf>
- Kolesnikov A (2016) Do Russians Want War?. Carnegie Moscow Center. Available at: [http://carnegieendowment.org/files/Article\\_Kolesnikov\\_2016\\_Eng-2.pdf](http://carnegieendowment.org/files/Article_Kolesnikov_2016_Eng-2.pdf)
- Kukkola, J; Ristolainen, M & Nikkarila, J-P (2017). GAME CHANGER Structural transformation of cyberspace  
<https://puolustusvoimat.fi/documents/1951253/2815786/PVTUTKL+julkaisuja+10.pdf/5d341704-816e-47be-b36d-cb1a0ccae398/PVTUTKL+julkaisuja+10.pdf.pdf>
- Moukin, G. (2013). Supercomputing Gap Seen as Threat to Economy. The Moscow Times. November 28, 2013.  
<https://themoscowtimes.com/articles/supercomputing-gap-seen-as-threat-to-economy-29999>
- Ovtsharenko Y (2004) Deputy Head of the Presidential Administration Vladislav Surkov: Putin is strengthening the state, not himself. (in Russian) <https://www.kompravda.eu/daily/23370/32473/>
- PP-1895. (2000). Information Security Doctrine of the Russian Federation. <http://base.garant.ru/182535/>
- PP-2796. (2014) Military doctrine of the Russian Federation, (in Russian), <https://rg.ru/2014/12/30/doktrina-dok.html>
- PUP-1. (2015). Information Security Doctrine of the Russian Federation (draft). <http://www.worldinwar.eu/information-security-doctrine-of-the-russian-federation-draft/>
- Putin, V. (2015) Meeting of the Valdai International Discussion Club. : <http://en.kremlin.ru/events/president/news/50548>
- Putin, V. (2016) President's Speech to the Federal Assembly, (in Russian) <http://kremlin.ru/events/president/news/53379>
- PP-538 (2005) [Decree of the Government of the Russian Federation of August 27, 2005 N 538 (ed. Of September 25, 2018 "On Approval of the Rules for Interaction of Communication Operators with Authorized State Bodies Conducting Operational-Investigation Activities. (in Russian) [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_55326/](http://www.consultant.ru/document/cons_doc_LAW_55326/)
- Putin V (2018) President's Speech to the Federal Assembly]. (in Russian) <http://kremlin.ru/events/president/news/56957>
- SBRF. (2012) The main directions of the state policy in the field of ensuring the security of automated systems for managing production and technological processes of critical infrastructure facilities of the RF, (in Russian),  
<http://www.scrf.gov.ru/security/information/document113/>
- SBRF. (2013) The concept of cybersecurity strategy of the Russian Federation (Draft), (in Russian), <http://council.gov.ru/media/files/41d4b3dfbdb25cea8a73.pdf>
- Sinovets P (2016) From Stalin to Putin: Russian Strategic Culture in the XXI Century, Its Continuity, and Change. Odessa. Mechnikov National University. <http://www.davidpublisher.org/Public/uploads/Contribute/57eb1fe5a12bc.pdf>
- Soldatov A, Borogan I (2015) The Red Web. New York: Public Affairs
- Tass (2016) In the Russian Federation developed the military Internet for the safe exchange of secret information  
<https://tass.ru/armiya-i-opk/3715422>
- UP-31 (2013) Decree of the President of the Russian Federation of January 15, 2013 N 31c Moscow "On the establishment of a state system for detecting, preventing and eliminating the consequences of computer attacks on information resources of the Russian Federation" (in Russian) <https://rg.ru/2013/01/18/komp-ataki-site-dok.html>
- UP-203. (2017) The Strategy for the Development of the Information Society in the Russian Federation for 2017-2030, (in Russian), <http://www.kremlin.ru/acts/bank/41919>
- UP-640. (2016) Foreign Policy Concept of the Russian Federation (in Russian) [http://www.mid.ru/foreign\\_policy/official\\_documents/-/asset\\_publisher/CptlCkB6BZ29/content/id/2542248?p\\_p\\_id=101\\_INSTANCE\\_CptlCkB6BZ29&\\_101\\_INSTANCE\\_CptlCkB6BZ29\\_languageld=ru\\_RU](http://www.mid.ru/foreign_policy/official_documents/-/asset_publisher/CptlCkB6BZ29/content/id/2542248?p_p_id=101_INSTANCE_CptlCkB6BZ29&_101_INSTANCE_CptlCkB6BZ29_languageld=ru_RU)
- UP-646. (2016) Doctrine of Information Security of the Russian Federation, (in Russian), <https://rg.ru/2016/12/06/doktrina-infobezobasnost-site-dok.html>
- UP-683. (2015) The National Security Strategy of the Russian Federation, (in Russian), <http://pravo.gov.ru/proxy/ips/?docbody=&nd=102385609>
- Yarovaya M (2013) Igor Ashmanov: "Today information domination is the same as air superiority]. May 1, 2013. (in Russian) <https://ain.ua/2013/05/01/igor-ashmanov-segodnya-informacionnoe-dominirovaniye-eto-vse-ravno-chto-gos-podstvo-v-vozduxe>

# Government Efforts Toward Promoting IoT Security Awareness for end Users: A Study of Existing Initiatives

Kautsarina and Bayu Anggorojati

Universitas Indonesia, Depok Indonesia, Indonesia

[kautsarina61@ui.ac.id](mailto:kautsarina61@ui.ac.id)

[bayuanggorojati@cs.ui.ac.id](mailto:bayuanggorojati@cs.ui.ac.id)

**Abstract:** Due to the ease and cost of developing IoT devices as well as to the high adoption rate of smart connected things, the IoT ecosystem is expected to grow continually. Some IoT applications are already on the market, such as in a smart home, wearable, connected vehicle, medical and healthcare, smart grids, and so on. However, the increasing use of IoT, especially in the individual domain, causes the opening of security vulnerabilities. Various studies have attempted to identify security issues on IoT. Many studies indicate that data privacy is one of the primary considerations in IoT because of the high possibility of creating security risks, such as unauthorized access, tapping, data modification, data forgery and so on. Some IoT services and applications provide personal and sensitive information openly and can be misused because of the leakage of data to third parties. Several studies state that the enforcement mechanism of IoT is still inadequate. Besides, there are issues such as identification problems, authentication and authorization, and cross-device dependencies. From the various vulnerabilities, threat sources have been identified by previous researchers apart from bad manufacturers and outside parties, namely IoT users themselves as owners of devices that intentionally or unintentionally provide access to sensitive information. From that case, user awareness becomes a critical aspect of the IoT ecosystem. Some countries are known to have prepared themselves with strategies and master plans in dealing with the IoT era. This study seeks to identify what efforts have been made by the government and other stakeholders in promoting IoT security for end-users in various countries and what best practices can be learned from existing initiatives, and what aspects might still need further research.

**Keywords:** government efforts, IoT ecosystem, security awareness, end user

---

## 1. Introduction

Human errors can be the result of negligence, accident, or deliberate action(Nicholson *et al.*, 2016). IBM report (2015) states that 9 out of 10 information security incidents caused by some human error(IBM Security, 2015). IoT can bring convenience to people, but if it cannot ensure the protection of personal privacy, private information may leak at any time. So the security of IoT cannot be ignored. With the widely spreading of IoT adoption, it will provide a more extensive wealth of information; thus the risk of information exposure will increase. In IoT cases, the exploitation of privacy and security harms users. For example, connected devices installed on the customers home of Singapore telco company is suspected hit by DDoS attacks on October 2016 (Yu 2016). It harmed the users because these devices were allowing remote attackers to gain access and control the devices(Yu, 2016).

Considering what happens with smartphone technology that sometimes captures information without the consent or knowledge of the user(Contos, 2015), there is every reason for users of IoT devices to worry about their privacy. Creating and maintaining information security requires the application of technical security controls, but also the form of administrative, procedural and managerial control(Tsohou *et al.*, 2008). Therefore, security compliance is not possible without addressing the human issues of information security with proper awareness and training(Bresz, 2004). Relevant government regulations and standards are currently lacking. Neither is their sufficient awareness among consumers to demand privacy and security solutions (Conventus Law, 2017).

It is essential that privacy and security be part of the design process at each stage of the design and development process. However, also it brings up the issue of the lifecycle of an IoT system and how it is maintained after deployment. Departing from this, the need to provide information about what users need to pay attention to and what needs to be done to secure smart devices is very critical.

There have been previous studies that conducted a literature review of measuring the information security awareness of smartphone users(Sari and Candiwan, 2014). There is also a systematic literature review with Kitchenham to measure cybersecurity awareness(Rahim *et al.*, 2015). Consumer awareness also become one of the key themes in the discussion of recommendation on industry reports for government action(Tanczer *et al.*,

2018). However, no one has done a literature review of the methods and approaches used to increase the security awareness of IoT users on existing IoT initiatives. To fulfill the gap, the authors conduct this study.

The aim of this study is gathering the existing knowledge and gains understanding regarding various methods and approaches for improving user security awareness and compliance in IoT. The increasing use of IoT, especially in the individual domain, causes the opening of security vulnerabilities, so it is critical to continuously enhance the security awareness culture and transform this culture into actual security conscious behaviors. This study will contribute to providing insight for those who want to improve the security awareness of IoT user in their ecosystem.

Professionals and scholars express the need to increase information security culture and a focus on the human factors involved in ICT to counter security risks(Alhogail and Mirza, 2010; Wamala, 2011; Glaspie and Karwowski, 2018). Extensive literature reviews quite convince that human to represent a significant threat to security. The review of existing IoT initiatives conducted with the goals of finding answers to the following research question:

*RQ: What approaches or best practices used by current IoT initiatives that address end-user awareness issues in the IoT ecosystem?*

This paper is consist of four sections and organized as follows. The first section presents an introduction to this study. Section 2 provides the scope of the research and the methodology of the literature review used throughout the review process. Section 3 presents the results of the review process and discusses existing studies that found in the previous section to answer the research question. Section 4 concludes the paper and offers some future works.

## **2. Methodology**

### **2.1 Information sources and inclusion criteria**

This study has used a systematic review to ensure that both the search and the retrieval process have been accurate and impartial. A systematic review is defined as a research technique that attempts to collect all empirical evidence in a particular field, to assess it critically and to obtain conclusions that summarize the research(Okoli and Schabram, 2010). A review protocol is describing each step of the systematic review, including eligibility criteria, was therefore developed before beginning the search for literature and the data extraction.

### **2.2 Information sources and inclusion criteria**

Authors are collecting kinds of literature from relevant sources that provided in official government or association websites and also search on electronic databases such as Science Direct, ProQuest, and Scopus with periods 2011-2018. It considered that this period would allow the retrieval of a current number of studies on the topic and detect the research trend for this topic. Types of literature that including on this study is Articles (journal, magazine, newspaper); Grey literature (conference proceeding, white paper, government document, and published report). The following inclusion criteria were used: articles published in English (IC1) and articles that deal with user IoT initiatives (IC2). Only documents are written in English (IC1) were included since the Scientific Community favors this language in the publication of research studies. Finally, IC2 was added to answer the research question. Authors also scanned the reference lists included in articles to ensure that this review would be more comprehensive.

### **2.3 Study selection**

The study selection was organized in the following four phases:

- 1. The search for a document from website and electronic databases. This phase was performed by using the following search string: ("smart things" OR "internet of things" AND "initiatives" OR "program" OR "strategic plan" OR "master plan" AND "security" AND "awareness" AND "methods" OR "approaches"), which adapted to the databases' search engines.
- 2. Exploration of title, abstract and keywords of identified articles and selection based on eligibility criteria.

- 3. Complete or partial reading of articles that had not been eliminated in the previous phase to consider whether they should be included in the review, by the eligibility criteria.
- 4. Scanning the reference lists of articles to discover new studies which were then reviewed as indicated in phases 2 and 3, but these articles had to satisfy the inclusion criteria.

### 3. IoT initiatives

European Union and the US are recognized to have significant efforts on IoT initiatives since 2007. Nevertheless, IoT initiatives in the US is scattered. From the results of the IDC survey on G20 countries, several countries in the Asia-Pacific region are known to have good IoT developments, namely China, Australia, Japan, Republic of Korea, Singapore and Indonesia. Based on the list of the countries, we focused on investigating IoT initiatives in these countries. Besides, we explore other efforts relevant to IoT security awareness. Summary of IoT initiatives reviewed in this study shown in Table 1.

**Table 1:** List of reviewed IoT initiatives

Issuing Institution	Name of initiatives	Issuing Date	Target of initiatives
<i>Government Initiatives</i>			
European Research Clusters on the Internet of Things (IERC)	The Internet of Things 2012 New Horizon	2012	EU-countries stakeholders
U4IoT consortium	U4IoT User Engagement for Large Scale Pilots in the Internet of Things	January 2017	LSPs Consortia; end-user
European Union Agency for Network and Information Security (ENISA)	Baseline Security Recommendations for the Internet of Things in the context of Critical Information Infrastructures	November 2017	IoT experts, software developers, and manufacturers; information security experts; Security solutions architects; Chief Information Security Officers (CISOs); Critical Information Infrastructure Protection (CIIP) experts
UK-Department for Digital, Culture, Media, and Sport	Mapping of IoT Security Recommendations, Guidance, and Standards to the UK's Code of Practice for Consumer IoT Security	October 2018	Working group member agencies
UK- DCMS & the National Cyber Security Centre	Code of Practice for Consumer IoT Security	October 2018	The device manufacturer, IoT service providers, Mobile application developers, retailers
UK-DCMS	Consumer Guidance for Smart Devices in the Home(for Digital, 2018)	October 2018	End-user/consumer
Finnish Strategic Centre for Science, Technology, and Innovation	IoT Strategic Research Agenda (IoT-SRA)	2012	Relevant stakeholder of the Finnish ICT industry
US Department of Homeland Security	Strategic Principles for Securing the Internet of Things	November 2016	IoT developers, manufacturers, service providers, industrial and business level consumers
US National Telecommunications and Information Administration (NTIA)	Incentives and Barriers to Adoption of IoT Update Capabilities	November 2017	IoT producers, governments, industry policymakers, researchers, civil society advocates

***Kautsarina and Bayu Anggoro Jati***

<b>Issuing Institution</b>	<b>Name of initiatives</b>	<b>Issuing Date</b>	<b>Target of initiatives</b>
US National Institute of Standards and Technology (NIST)	(Draft) Considerations for Managing the Internet of Things (IoT) Cybersecurity and Privacy Risks	September 2018	organization
US NIST	Interagency Report on the Status of International Cybersecurity Standardization for the Internet of Things (IoT)	November 2018	Working group member agencies
US National Cyber Security Alliance	Data Privacy Day	January 2019	End-user
Australian Communications and Media Authority (ACMA)	The Internet of Things and the ACMA's areas of focus	2015	Relevant stakeholders
IoT Alliance Australia	Strategic Plan to Strengthen IoT Security in Australia	September 2017	Relevant stakeholders
China State Council	Guiding Opinions of the State Council on Promoting the Orderly and Healthy Development of the Internet of Things	February 2013	Relevant stakeholders
China Ministry of Industry and Information Technology (MIIT)	13th Five Year Plan Development Plan for the Internet of Things	December 2017	Relevant stakeholders
South Korea	Master Plan for Building the Internet of Things (IoT)	Mei 2014	Relevant stakeholders
Japan	Cybersecurity Strategy	June 2018	Relevant stakeholders
Singapore Government Technology Agency	Smart Nation Sensor Platform	August 2017	Relevant stakeholders
Indonesia Ministry of Communication and Information Technology	(Draft) Masterplan IoT	2018	Relevant stakeholders
<i>Other Initiatives</i>			
Online Trust Alliance (OTA)	Smart Device Purchase & Setup Checklist	December 2015	Buyers, users
OTA	Smart Home Checklist: Maximizing Security, Privacy & Personal Safety	March 2017	Buyers, sellers, renters
OTA	IoT Security & Privacy Trust Framework v2.5	October 2017	Developers, purchasers, and retailers
Internet Society (ISOC)	Top tips for the Internet of Things security and privacy	June 2018	Consumers
I Am The Cavalry	Hippocratic Oath for Connected Medical Devices	January 2016	Healthcare stakeholders
Open Web Application Security Project (OWASP)	IoT Security Guidance	2018	Manufacturers, developers, enterprises, consumers
Broadband Internet Technical Advisory Group (BITAG)	Internet of Things (IoT) Security and Privacy Recommendations	November 2016	IoT device manufacturers

Issuing Institution	Name of initiatives	Issuing Date	Target of initiatives
Industrial Internet Consortium (IIC)	Industrial Internet of Things Volume G4: Security Framework	2016	Owners, operators, system integrators, business-decision makers, architects and any stakeholder with interest in security
Cloud Security Alliance (CSA)	Security Guidance for Early Adopters of the Internet of Things	April 2015	Business consumers

### **3.1 Examples of EU-countries IoT initiatives**

Internet of Things (IoT) has become a concept identified by the European Commission as one of the pillars supporting future infrastructure(European Commission, 2009). In Europe, the academic research work in IoT was mainly performed in different EU-funded seventh Programme Framework (FP7) projects. European Research Cluster on the Internet of Things (IERC) was founded and funded under FP7 in 2009, to utilize the research achievements better and to provide a place to share the lessons and experiences from different projects. Currently, IERC comprises around 30 EU-funded projects. The IoT Initiative (IoT-I) represents the first serious attempt in building a unified IoT community in Europe.

*User engagement for Large Scale Pilots in the IoT(U4IoT),* is an initiative to ensure that end-user rights, related to data protection, are fully respected. To socialize General Data Protection Regulation (GDPR) awareness, EU developed a web-based game about privacy concept to support end-user engagement in the five Large Scale Pilots (LSPs) on the Internet of Things (IoT) financed by the European Commission and other partners. The target groups of users are LSP consortia and the end-users of the 5 LSPs who will be exposed to the IoT pilots, with the assumption that they have different levels of education, such as farmers, event organizers, engineers, physicians, and so on, and also the broad public in general.

The Finnish government has many efforts to build an IoT ecosystem, and also has defined that cybersecurity is a critical part of the Internet of Things. Through their programs such as Cyber Security Cluster and Finnish Information Security Cluster (FISC), they prepare to enhance education in the cyber security field. Finnish Funding Agency for Technology and Innovation (TEKES) develop IoT Strategic Research Agenda(Finnish Strategic Centre for Science Technology and Innovation, 2011). End-user adaptation becomes a critical part of the IoT ecosystem because users adoption of the first services will ease the adoption path for other services as user literacy of IoT services improves.

As well as the UK, the review showed that the UK develops many initiatives to support IoT secure ecosystem. The Code of Practice sets out practical steps for IoT manufacturers and other industry stakeholders to improve the security of consumer IoT products and associated services(Department for Digital Culture Media and Sport, 2018). To provide comprehensive guidance, they also provide the mapping for reference and tool to understand the relationship between the code of practice and existing sources from industry or association.

### **3.2 US IoT initiatives**

In April 2008, the U.S. National Intelligence Council (NIC) published a conference report on “Disruptive Civil Technologies – Six Technologies with Potential Impacts on U.S. Interests out to 2025”, and one of the technologies was IoT(National Intelligence Council, 2008).

Then, the US Department of Homeland Security released Strategic Principles for Securing the Internet of Things in 2016(U.S. Department of Homeland Security 2016). The principles are designed to improve the security of IoT across the full range of design, manufacturing, and deployment activities. Widespread adoption of these strategic principles and the associated suggested practices would dramatically improve the security posture of IoT. These principles are intended to be adapted and applied through a risk-based approach that takes into account relevant business contexts, as well as the particular threats and consequences that may result from incidents involving a network-connected device, system, or service. US Senate also introduced Act to provide minimal cybersecurity operational standards for Internet-connected devices purchased by Federal agencies(the Senate of the United States, 2017).

### **3.3 Examples of Asia-Pacific countries IoT initiatives**

People are defined as one of IoT enablers on the ACMA report(Street and Vic, 2015). Access in the IoT environment is predicated on the technical proficiency of users in being able to use devices, identify sources of services and manage digital information to ensure that citizens in Australia can productively engage with IoT opportunities. Skills that are expected to become more critical in the IoT environment will include some technical proficiency, as well as managing security settings, identity and authentication requirements, and the digital footprint created by IoT communications.

IoT Alliance Australia (IoTAA) build Strategic Plan to Strengthen IoT Security in Australia on 2017(IoT Alliance Australia, 2017), to achieve effective industrial practices in cybersecurity for IoT in Australia and set an agenda for collaborating with government, industry and global partners in IoT security. In the document, it briefly said about consumer need about security awareness:

*"Education for the consumer on why security is important and how they can use the security principles to guide their decision-making process (think Heart Foundation tick). We need to specify who will be responsible for this. This could be device manufacturer (e.g., Fitbit, home security device manufacturer), or, could be platform operator (e.g., ISP or other service providers), or, could be government/industry body (e.g., Dept. PM&C, IoTAA). By identifying who should be responsible, we articulate the standards that need to be met." (IoT Alliance Australia, 2017).*

The report (IoT Alliance Australia, 2017) also stated that a consumer education scheme might involve the following steps: 1)general public education regarding the importance of internet security; 2) promotion of certification schemes and signs related to certification to consumers; 3) publication of information and media engagement regarding case identified weakness in IoT devices and the fallout from cyber-attacks.

Meanwhile, on China, the government have produced annual studies of the IOT industry's weaknesses to guide policymaking, and they offer authoritative assessments of the barriers that continue to serve as a drag on the industry's growth. To address the development needs of the IoT industry, China's government has proposed and enacted policy solutions that include: Developing "special action plans" to promote technologies that can drive innovation in specific IoT applications, deepen understanding of network applications, and promote a model of "healthy sustainable development" for the Internet of Things.

In Japan' Cybersecurity strategy report(NISC, 2018) noted that policy approaches towards achieving improving socio-economic vitality and sustainable development is creation of secured IoT Systems that includes promoting new business harnessing the secured IoT systems with the idea of 'Security by Design' and establishing comprehensive guidelines for IoT systems security in the various areas such as energy, automotive and medical industries. The report also states about policy approaches towards building a safe and secure society for the people. Includes promoting security measures for general users such as security alerts and tips, and also promoting local community-based outreach and awareness raising activities.

Based on South Korea IoT Masterplan(Ministry of Science ICT and Future Planning, 2014), the country surely had a proper plan to establish a safe IoT infrastructure. Within, 2014 South Korea develop an 'Information Security Roadmap for the Internet of Things' and establish a cooperative framework with other countries such as EU, US, and Japan for prompt responses to and analysis incidents based on information sharing. Besides, the country will establish pilot projects to testing security functions and capacities within the IoT Innovation Center and promote the embedding of security measures within IoT product and services such as healthcare and home electronics areas, since in the planning stage. The country also will expand the development of IoT security and nurture IoT information security coordinators. They also intend to apply the concept of 'Privacy by Design' from the planning stage and develop privacy enhancement technologies (PET).

While Singapore turned out to have had many interesting programs to socialize the benefits of IoT for users since 2015, for example in the National Steps Challenge program to encourage Singaporeans to be more physically active with the help of wearable steps tracker and Healthy 365 App (Smart Nation Singapore, 2015). The government is fully aware of privacy and security concerns of IoT implementation, so they involved programmers from differing groups to This country currently focus on accelerating the deployment of sensors and other IoT devices toward Smart Nation Sensor Platform (GovTech Singapore, 2017).

### **3.4 Other initiatives**

According to the review, we found that there are sufficient sources of references for security and privacy concerns for IoT device manufacturers and developers. For instance, Industrial Internet Consortium through the report released Security Framework as a reference that guides for improving organizational approaches on creating reliable Industrial IoT systems(Industrial Internet Consortium, 2016). The Online Trust Alliance (OTA), an Internet Society initiative, has developed many useful resources such as an IoT Trust Framework(Internet Society, 2017) that can be used as a risk assessment. On the context of end-user, OTA provides easy to follow the smart home checklist to guide the user about what to review regarding their connected devices on smart home(Online Trust Alliance, 2017). Besides, Internet Society sharing top tips to for IoT users to protect their security and privacy through learning to 'shop smart,' update the devices, review the privacy settings, the good-practice of a password and turn on encryption(Internet Society, 2018b).

Through the document titled "IoT Security for Policymakers," Internet Society provides guiding principles and recommendations for governments to consider when addressing IoT security (Internet Society, 2018a). The government must play an active role to protect the user by facilitating more optimal security and privacy by providing clarity on the existing laws of exclusion that can apply to the IoT context. As with the misleading ban on conventional products, IoT manufacturers are prohibited from making deceptive statements about the security of IoT products or services offered. Retailers should also share responsibility and not sell IoT products with known critical safety and security defects.

Since 2014, the Open Web Application Security Project (OWASP) also take attempts to help consumers purchase secure products in the Internet of Things to use(OWASP, 2014). They provide necessary level guidance from the consumer perspective. Even they claimed that the guideline is not a comprehensive list of considerations, but ensuring that these fundamentals are covered will significantly aid the consumer in purchasing a secure IoT product.

There also an initiative called I Am The Cavalry movement that was built in response to concerns about cybersecurity risks that might threaten public safety, which its efforts are focused on cybersecurity issues relating to medical, automotive, home electronics and public infrastructure. On January 2016, the movement released 'Hippocratic Oath for Connected Medical Devices' that describes commitment from the perspective of medical devices capabilities that preserve patient safety, as well as trust in the process of care delivery itself(I Am The Cavalry, 2016).

BITAG on their report noted that end users of IoT device rely upon the IoT supply chain, from manufacturer to retailer, to care about their security and privacy(BITAG, 2016). So that BITAG recommends that the IoT supply chain takes the specified steps for end-user awareness, smart devices should have a privacy policy that is clear and understandable, and manufacturers should provide explicit methods for consumers to determine whom they can be asked for further question and support, also plans to contact consumers to inform something important such as software vulnerabilities or other issues.

### **3.5 Best practices**

#### **3.5.1 Communicating privacy concerns**

Public awareness become one of the aspects of the technologies discussed are critical concerning privacy, and they will have social implications related to acceptance of the technology. Various initiatives recommend promoting security best practices and guiding principles to guide the design, deployment, and use of IoT devices and services. Besides, foster a culture of security among key stakeholder, including the users.

Various technological approaches are offered related to data protection, privacy and security concerns. One initiative defines that different access levels, control policies, and mechanisms to guarantee that no identification of personal data is possible by unauthorized clients/operators must be carefully interpreted and applied. Besides, and depending on the use case or scenario, "opt-in" paradigms (in which users must voluntarily express and confirm their awareness and willingness to share personal data) should be incorporated as much as possible.

So that, Users' privacy concerns about the accessibility and use of information captured by IoT devices and sensors is a significant challenge, and users need to be assured that they must be enabled to understand and manage and control the exposure of their private and sensitive data. The problem also includes legislation ensuring individual privacy rights for what kind of surveillance and information the authorities can employ and access. Besides, privacy awareness still on the track of 2016-2020 research themes on IERC. The current situation means there are even more initiatives to expects to handle IoT privacy issues.

So do in China. Strengthen security awareness of the Internet of Things is one of the fundamental principles from the State Council. China has an objective in creating a sound development environment with an effort to establish and improve a policy system that is conducive to the promotion of IoT applications, innovation incentives, and orderly competition, and promptly promote the development of laws and regulations on information security and privacy protection. Another principle is to strengthen protection management and ensure information security. Improve the information security management and data protection level of the Internet of Things, strengthen the research and development of information security technology, promote the construction of information security assurance system, establish and improve the supervision, inspection and security assessment mechanism, and effectively safeguard the information collection, transmission, processing and application of the Internet of Things.

### *3.5.2 A form of security awareness approaches*

There are several forms of approaches known in the review. *UK' Consumer Guidance for Smart Devices in the Home* provides such an awareness poster and checklist to enhance the user security awareness of their smart devices. While on the US, there are attempts put on 'Secure IoT Devices' logo or notation for products that comply with security best practices, so that would create product differentiation in a crowded market for IoT devices. Interestingly, EU efforts toward the Denmark government (the Alexandra Institute, 2011) and the Finnish government creates IoT comics(Kiravuo et al., 2016) to approach IoT security awareness on youth age. Another exciting effort, the EU toward U4IoT develops privacy games to enhance privacy awareness of end-users. There is, however, no one-size-fits-all solution for mitigating IoT security risks. So, various specific guidelines should be provided equally across the diversity of IoT devices. According to the review, we found user security awareness guideline related to smart home and smart devices. The clear step of practice that easy to follow by non-technical users are should be minded as the assumption that users using IoT devices have various level of education.

## **3.6 The implication for Indonesia context**

As said by ITU that governments generally have lots of freedom in adopting national regulations, it is found that several countries have their priorities. However, follow best practices is always advisable. The implication is in the Indonesian context; this review will be reference material for developing guidelines that can be useful for increasing security awareness of smart device users. The review showed that users' IoT security awareness can be developed in any form, such as factsheets, checklists, games even comics. These kind of methods are known adopted in the field of forming security awareness(Chmura, 2017). Regarding to developing Indonesia IoT master plan, various methods of security awareness should be well planned to be featured on the initiative, so that it will be clearer in its application and measurement of results to find out the impact of applying the method on enhancing user security awareness of IoT devices.

## **4. Conclusion and future work**

IoT initiatives in several countries discussed have different focus and priorities. Related to efforts to build security awareness for users of IoT devices, it is known that the form of guidance aimed at IoT device users as a reference is still limited. However, several countries provide excellent recommendations to follow. The UK offers a reasonably comprehensive example in developing IoT user security awareness strategies. We also realized that there is no one-size-fits approach to solving security awareness. Therefore, the findings of this review need to be further elaborated and discussed with relevant stakeholders to the responsibility of information security awareness. The implication is in the Indonesian context, this review will be reference material for developing guidelines that can be useful for increasing security awareness of smart device users. The evaluation must also be done to be able to measure the success of a security awareness program for end-users. Due to no attempt was made to evaluate the use of this guideline in the document reviewed, then in the next study, the proposed

direction compiled from various stakeholder discussions would then be tested at end users to measure the effectiveness of this guideline in increasing user security awareness.

## **References**

- Alhogail, A. and Mirza, A. (2010) 'Organizational Information Security Culture Assessment', pp. 286–292.
- BITAG (2016) *Internet of Things (IoT) Security and Privacy Recommendations*. Available at: [https://www.bitag.org/documents/BITAG\\_Report\\_-\\_Internet\\_of\\_Things\\_\(IoT\)\\_Security\\_and\\_Privacy\\_Recommendations.pdf](https://www.bitag.org/documents/BITAG_Report_-_Internet_of_Things_(IoT)_Security_and_Privacy_Recommendations.pdf).
- Bresz, F. P. (2004) 'People-Often the Weakest Link in Security, But One of the Best Places to Start,' *Journal of Health Care Compliance*, 6(4), pp. 57–60.
- CAICT & DG CONNECT (2016) *EU-China Joint White Paper on Internet-of-Things*. Available at: <http://iot6.eu/white-papers>.
- Chmura, J. (2017) 'Forming the Awareness of Employees in the Field of Information Security,' *Journal of Positive Management*, 8(1), p. 78. doi: 10.12775/jpm.2017.006.
- Contos, B. (2015) *Security and the Internet of Things - are we Repeating History?*, CSO.
- Conventus Law (2017) *Asia Pacific - 2017 Predictions: Criminals Harness IOT Devices As Botnets To Attack Infrastructure*. / *Conventus Law, Conventus Law*. Available at: <http://www.conventuslaw.com/report/asia-pacific-2017-predictions-criminals-harness/> (Accessed: 20 December 2018).
- Department for Digital Culture Media and Sport (2018) 'Code of Practice for Consumer IoT Security,' Gov.UK, (October). Available at: <https://www.gov.uk/government/publications/secure-by-design/code-of-practice-for-consumer-iot-security%0Ahttps://www.gov.uk/government/publications/secure-by-design>.
- European Commission (2009) *Internet of Things Strategic Research Roadmap*. Available at: [http://www.internet-of-things-research.eu/pdf/IoT\\_Cluster\\_Strategic\\_Research\\_Agenda\\_2009.pdf](http://www.internet-of-things-research.eu/pdf/IoT_Cluster_Strategic_Research_Agenda_2009.pdf).
- Finnish Strategic Centre for Science Technology and Innovation (2011) *Internet-of-Things Strategic Research Agenda*. doi: 10.1109/cisti.2016.7521526.
- for Digital, D. (2018) 'Consumer Guidance for Smart Devices in the Home.' Available at: [www.cyberaware.gov.uk](http://www.cyberaware.gov.uk).
- Glaspie, H. W. H. W., and Karwowski, W. (2018) 'Human factors in information security culture: A literature review,' in Nicholson, D. (ed.) *Advances in Intelligent Systems and Computing*. Springer International Publishing AG, pp. 269–280. doi: 10.1007/978-3-319-60585-2\_25.
- GovTech Singapore (2017) *Smart Nation Sensor Platform Factsheet*. Available at: [https://www.tech.gov.sg/files/media/speeches/2017/05/Factsheet\\_Smart\\_Nation\\_Sensor\\_Platform.pdf](https://www.tech.gov.sg/files/media/speeches/2017/05/Factsheet_Smart_Nation_Sensor_Platform.pdf).
- I Am The Cavalry (2016) *Hippocratic Oath for Connected Medical Devices*. Available at: <https://www.iamthecavalry.org/domains/medical/oath/>.
- IBM Security (2015) 'IBM 2015 Cyber Security Intelligence Index', *IBM Security Managing Security Services*, p. 24. doi: SEW03039-USEN-02.
- Industrial Internet Consortium (2016) 'Industrial Internet of Things Volume G4 : Security Framework', *Industrial Internet Consortium*, V1(IIC:PUB:G4:V1.0:PB:20160919), pp. 1–173.
- Internet Society (2017) *IoT Security & Privacy Trust Framework v2.5*. Available at: [https://www.otsalliance.org/system/files/files/initiative/documents/iot\\_trust\\_framework6-22.pdf](https://www.otsalliance.org/system/files/files/initiative/documents/iot_trust_framework6-22.pdf).
- Internet Society (2018a) *IoT Security for Policymakers*. Available at: <https://www.internetsociety.org/resources/2018/iot-security-for-policymakers/>.
- Internet Society (2018b) 'Top tips for the Internet of Things security and privacy.' Internet Society, p. 8. Available at: <https://www.internetsociety.org/wp-content/uploads/2018/06/2018-IoT-Consumer-Tips.pdf>.
- IoT Alliance Australia (2017) *Strategic Plan to Strengthen IoT Security in Australia*. Available at: <http://www.iot.org.au/wp/wp-content/uploads/2016/12/IoTAA-Strategic-Plan-to-Strengthen-IoT-Security-in-Australia-v4.pdf>.
- Kiravuo, T. et al. (2016) 'Internet of Things Coming Soon to Your Home.' Clarified Studio, pp. 1–69. Available at: [www.clarified.info](http://www.clarified.info).
- Ministry of Science ICT and Future Planning (2014) *Master Plan for Building the Internet of Things ( IoT )*.
- National Intelligence Council (2008) *Six Technologies With Potential Impacts on the US Interests Out to 2025*. Available at: <https://fas.org/irp/nic/disruptive.pdf> (Accessed: 30 December 2018).
- Nicholson, D. et al. (2016) *Advances in Human Factors in Cybersecurity*.
- NISC (2018) *Cybersecurity Strategy (Provisional Translation)*. Available at: <https://www.nisc.go.jp/eng/pdf/cs-senryaku2018-en.pdf>.
- Okoli, C. and Schabram, K. (2010) 'A Guide to Conducting a Systematic Literature Review of Information Systems Research,' *Working Papers on Information Systems*, 10(26), pp. 1–51. doi: 10.2139/ssrn.1954824.
- Online Trust Alliance (2017) *Smart Home Checklist Maximizing Security, Privacy and Personal Safety*. Available at: [https://otalliance.org/system/files/files/initiative/documents/smart\\_home\\_check\\_list\\_3-17.pdf](https://otalliance.org/system/files/files/initiative/documents/smart_home_check_list_3-17.pdf).
- OWASP (2014) *OWASP Internet of Things Top Ten Project, OWASP IoT Project*. Available at: [https://www.owasp.org/index.php/Category:Threat\\_Agent](https://www.owasp.org/index.php/Category:Threat_Agent) (Accessed: 20 November 2017).
- Rahim, N. H. A. et al. (2015) 'A systematic review of approaches to assessing cybersecurity awareness,' *Kybernetes*, 44(4), pp. 606–622.

- Sari, P. K., and Candiwan, C. (2014) 'Measuring Information Security Awareness of Indonesian Smartphone Users', *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 12(2), p. 493. doi: 10.12928/TELKOMNIKA.v12i2.2015.
- Smart Nation Singapore (2015) *Smart Nation Progress*. Available at: <https://www.smartnation.sg/why-Smart-Nation/smart-nation-progress> (Accessed: 24 December 2018).
- Street, C. and Vic, M. (2015) 'The Internet of Things and the ACMA's areas of focus Emerging issues in media and communications Occasional paper,' (November).
- Tanczer, L. et al. (2018) *Summary literature review of industry recommendations and international developments on IoT security: PETRAS IoT Hub*. Available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/686090/PETRA\\_S\\_Literature\\_Review\\_of\\_Industry\\_Recommendations\\_and\\_International\\_Developments\\_on\\_IoT\\_Security.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/686090/PETRA_S_Literature_Review_of_Industry_Recommendations_and_International_Developments_on_IoT_Security.pdf).
- The Alexandra Institute (2011) 'Inspiring the Internet of Things.' the Alexandra Institute.
- The Senate of the United States (2017) *Internet of Things (IoT) Cybersecurity Improvement Act*. Available at: <https://www.congress.gov/bill/115th-congress/senate-bill/1691>.
- Tsouhou, A. et al. (2008) 'Investigating information security awareness: Research and practice gaps,' *Information Security Journal*, 17(5–6), pp. 207–227. doi: 10.1080/19393550802492487.
- U.S. Department of Homeland Security (2016) *Strategic Principles for Securing the Internet of Things*.
- Wamala, F. (2011) 'ITU National Cybersecurity Strategy Guide,' *Chemistry & ...*, p. 122. doi: 10.1017/CBO9781107415324.004.
- Yu, E. (2016) 'Singapore telco says DDoS attacks behind two recent outages,' *ZDNet*, 26 October.

# Cyber Entanglement: A Framework for the Study of U.S.–China Relations

Karine Pontbriand

University of New South Wales, Canberra, Australia

[k.pontbriand@student.adfa.edu.au](mailto:k.pontbriand@student.adfa.edu.au)

**Abstract:** This paper explores entanglement in cyberspace as a conceptual framework for the analysis of 21st century International Relations. In quantum physics, entanglement refers to the property of sub-atomic particles to behave as a single non-separable entangled system. Using quantum entanglement as a metaphor to understand the relationships between states in cyberspace, this paper argues that entanglement has become a condition *sine qua non* of the current international system, where particles (in terms of states and non-state actors alike) are entangled and are thus in a situation of *unavoidable interaction*. Focusing on the relationship between two “entangled giants”, China and the United States (U.S.), this research paper will discuss the implications of what could be considered “cyber entanglement” on the interactions between states. The U.S.-China relationship in cyberspace is at the centre of an important academic and public debate. China is currently seen in the United States as one of their biggest cyber threats, and International Relations theorists and practitioners have warned that the two states are heading towards a Thucydides’ trap, where war might become inevitable. However, the two great powers, despite their competing national interests, are entangled by the connective infrastructure of cyberspace. This paper’s fundamental argument is two-fold. First, despite divergences between China and the United States on different cyber-related and economic issues, cyber entanglement means that tensions are less likely to escalate to the level of armed conflict. Second, cyber entanglement drives states toward collaboration. This paper proposes that any framework for the analysis of the behaviour of the U.S. and China in cyberspace has to assume that the relationship is impacted, and shaped, by entanglement.

**Keywords:** cyberspace, entanglement, international relations, China, United States

---

## 1. Introduction

In 2017, the American technology giant Microsoft announced that it was working with the Chinese government to develop a special edition of its Windows 10 operating system (OS). China had previously banned Microsoft Windows 8 on government computers in 2014, allegedly because of security and surveillance concerns (Kai, 2014). Therefore, Microsoft decided to produce a Chinese version of Windows 10 that included more management and security controls, in accordance with the Chinese government’s requirements (Myerson, 2017). The company developed the new version in partnership with the Chinese state-owned corporation China Electronics Technology Group (CETC). In other words, Microsoft, a company headquartered in the United States (U.S.), created an OS currently used by the Chinese government, and provides Information Technology (IT) services for Chinese customers through support and deep collaboration with Chinese companies, such as Lenovo, who was one of the first Chinese partners to have devices that come preinstalled with Windows 10 China Government Edition (Myerson, 2017).

The fact that a giant U.S. tech company like Microsoft is working very closely with the government of the world’s second economy and biggest player in Asia has an important effect on the current international system. It is an illustrative example of the concept introduced in this paper. The connective nature of cyberspace has transformed the international system, creating a new paradigm where the enmeshment between nation-states goes beyond interdependence: this paradigm, that I conceptualize as “cyber entanglement”, has implications for states relations in cyberspace. The U.S.-China bilateral relationship is a clear example. In an article for *Foreign Affairs*, former Australian prime minister and Chinese expert Kevin Rudd argued that China and the U.S. need a common conceptual framework of what their relationship is ultimately about in order to help navigate through the “Thucydidean dilemma” that they are confronting (Rudd, 2018). This paper argues that cyber entanglement could be this common conceptual framework. Moreover, it stresses that China and the United States are deeply entangled in cyberspace, which explains why they are more likely to collaborate despite their competitive interests.

This paper will first expand on the concept of quantum entanglement and on its utility as a metaphor to explain states interaction in cyberspace. It will then introduce the conceptual framework built on cyber entanglement and explain the potential of this framework to analyse bilateral relationships between nations. Finally, the paper will use the China-U.S. case-study to illustrate what are the potential impacts of cyber entanglement on bilateral relations.

## 2. The concept of entanglement in quantum physics

The notion of entanglement in quantum physics refers to the property of quantum particles to behave as a single non-separable entangled system (Raimond et al, 2001, 565), regardless of spatial distance or even time. The term was coined by Schrödinger who illustrated that when two systems “enter into temporary physical interaction due to known forces between them, and when after a time of mutual influence the systems separate again, then they can no longer be described in the same way as before” (Schrödinger, 1935, 555). Entanglement manifests a correlation between entities who once they have come into contact maintain contact even kilometres away. “Spooky action at a distance”, as referred to by Albert Einstein (Hamilton, 2017), also means that the state of each particle cannot be described independently of the state of the other. Rather, the state of the particle must be described for the system as a whole, and the “state of one particle is determined by a measurement performed on the other”(Raimond et al, 2001, 565). Therefore, to be entangled is not simply to be intertwined or involved with another, but to even lack an independent existence (Barad, 2007). For instance, in an entangled relationship, the separation of the two particles is not possible, because the particles no longer have an individual, independent nature. Quantum entanglement is “a calling into question of the very nature of two-ness, and ultimately of one-ness as well” (Barad, 2010, 251).

Outside of the cyber-domain, Adesso uses the phenomenon of love as a metaphor to explain how entanglement can be used to describe social interactions. “In our metaphor, nobody of the two lovers is complete on its own” (Adesso, 2007). The two individuals are non-separable halves of the same entangled entity and complement each other only when taken together, as a couple. This metaphorical description of love serves as a good transition for the introduction of the utility of entanglement in order to better understand social phenomena, such as international relations. This paper uses the metaphor of quantum entanglement to extend the understanding of IR as it relates to cyberspace. Metaphors and analogies impact how people and policymakers understand cyberspace and its underlying concepts. “Metaphors are very useful when initially creating shared understanding and building knowledge” (Ormrod and Turnbull, 2016, 284). However, this paper does not assume that it is a direct analogous tool, but simply that the quantum entanglement metaphor offers a useful conceptualization of inter-states relations in cyberspace.

## 3. The metaphor: Entanglement in international relations and cyberspace

The concept of entanglement has been previously borrowed from quantum physics and applied to the academic discipline of International Relations (IR). In IR, particles can refer to state and non-state actors, such as individuals, international organizations and corporations. Quantum entanglement of sub-atomic particles can therefore be used as an image to study macro-level relations like inter-state relations (Hamilton, 2017). Consequently, the notion of entanglement in IR postulates that in order to describe the condition of an international phenomenon, we need to assume the interconnection of the particles together, as well as with and within the global system, and try to explain how that affects the interaction. According to Montgomery (2016, 105), adopting a quantum entanglement perspective “allows for phenomena that are highly interdependent, simultaneous and geographically separated”. For instance, the analogy is particularly useful to study global processes that transcend national boundaries, such as cyberspace (Montgomery, 2016). Montgomery’s argument is that the academic discipline of International Relations should move away from notions of independent and isolated units and instead recognize the “quantum complexity of social reality” (Montgomery, 2016, 105).

With regards to states interactions in cyberspace more specifically, entanglement has been proposed as a deterrence strategy for cyber conflict (Nye, 2017; Brantly, 2018; Foerster, 2012; Schwartz, 2011). Deterrence can be understood as dissuading people by making them believe that the costs of their actions to them will exceed the benefits (Nye, 2017). In *Deterrence and Dissuasion in Cyberspace*, Nye (2017, 58) refers to entanglement as “the existence of various interdependences that make a successful attack simultaneously impose serious costs on the attacker as well as the victim”. Potential adversaries will therefore refrain from attacking because they have something highly valuable to lose and thus the benefits of the status quo are greater than the risk of losing what they have. For example, Nye suggests that a Chinese cyber-attack on the U.S. power grid, which imposes great costs on the U.S. economy, would conversely induce damages for China as well because the two countries are economically interdependent (Nye, 2017).

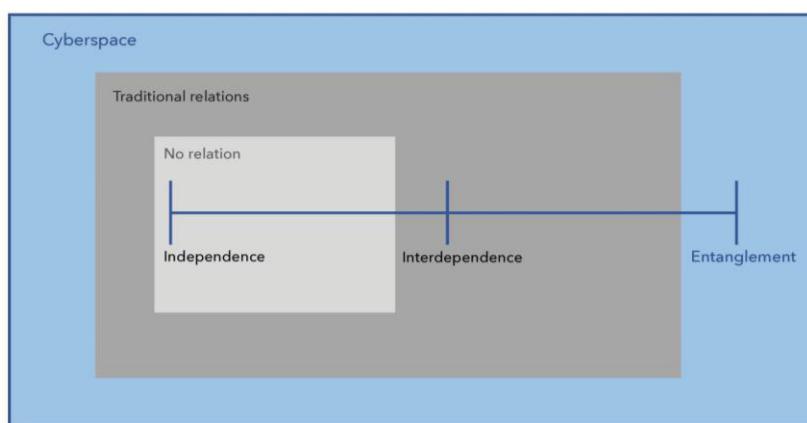
However, this application of the notion of entanglement remains anchored in the traditional paradigm of IR theory. Entanglement here really only means “interdependence by another name”, as put by Harknett (2017).

As a central concept of the academic discipline of IR, interdependence posits that economic exchanges and communications between states will decrease the incentives for war and instead foster greater stability and peace (Bolt and Shearn, 2014). But history has demonstrated the limits of economic interdependence theory, and its critics use World War I as evidence that economic ties may not prevent devastating war (Nye, 2017). In the current era, where global transactions rely on digital networks, it is however possible to argue that cyberspace has both broadened and deepened interdependencies between nations. The emergence of global Internet enabled supply chains is a good example of the new level of interdependence between countries and companies for economic benefits (Brantly, 2018; Friedman, 2007). The connective nature of cyberspace, understood as the global digital electronic telecommunications including the entirety of information and communication systems and devices (Deibert, 2017), has transformed the international system, creating a new paradigm where states are part of a deeply integrated system characterised by entanglement. For instance, the Internet relies on the connection between networks distributed around the world. Its architecture is decentralized (Singer and Friedman, 2014) and cooperatively constructed (Lindsay, 2017). To ensure the circulation of information, actors (such as Internet service providers, telecom carriers, programmers, etc.) need to adopt common standards: code, procedures, routines, programs, protocols, files, folders. Consequently, incentives for cooperation are built into cyberspace and they increase as more economic, political and social activities rely on information technology (IT) (Lindsay, 2017).

Therefore, as argued by Harknett (2017), deterrence is the wrong framework for analysing current cyber dynamics, and thus a new line of research outside the deterrence paradigm is needed. “Interconnectedness means that national security actors are in constant contact with other players and, unlike in strategic environments in which deterrence might succeed, it suggests that strategy must assume that contact and action are never absent” (Harknett and Nye, 2017, 198). Furthermore, entanglement in cyberspace can be understood not as the result of a nation’s deterrence strategy, but rather as an *inherent characteristic* of the cyber age. Actors are indeed in constant interaction because of the entanglement created by cyberspace.

#### 4. The cyber entanglement conceptual framework

The following section introduces a proposed cyber entanglement conceptual framework which provides a lens through which relationships between certain states can be better understood in the digital age. As explained previously, the characteristics of cyberspace, its connective and integrated nature, have had an immense impact on the relationships between the “particles” of the international system. States’ bilateral relationships can be organized on a spectrum ranging from absolute independence, where there is zero relation between states, and absolute entanglement, where a country’s behaviour is linked to the behaviour of another. This spectrum is shown by the Figure 1 below.



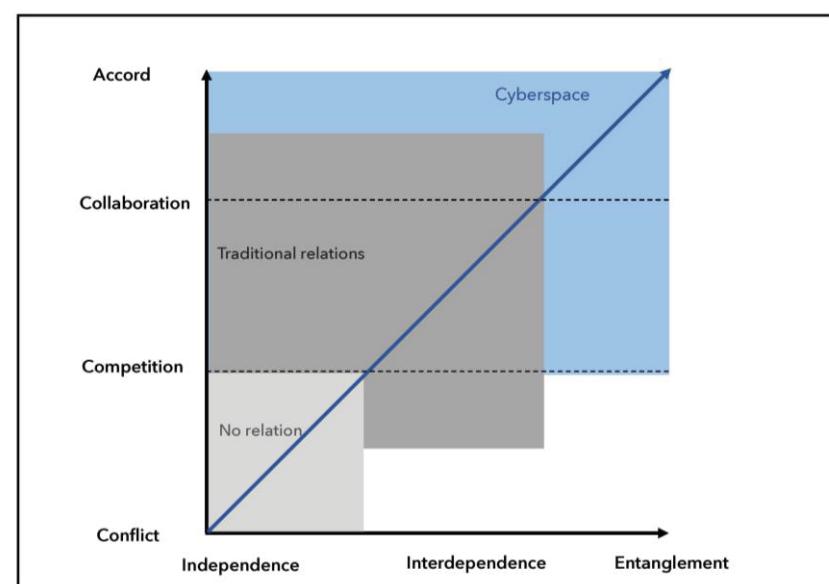
**Figure 1:** Cyber entanglement spectrum

For example, the relation between North Korea and the United States would be situated more closely to the left end of the spectrum (independence) considering that they have limited traditional relationships. Then, as states share traditional relationships such as economic exchanges, diplomatic interactions or infrastructure sharing (e.g. roads, railways, power grids), they move along the spectrum towards a condition of interdependence. Subsequently, establishing deeper relations through cyberspace will move them along the spectrum towards the state of entanglement. The Canada-U.S. relationship for instance would be situated closer to the right end of

spectrum (entanglement), because of the deep integration of the two states' infrastructures, economies, militaries and societies in general (Government of Canada, 2018). Austin (2018b) has contended that cyberspace has created a new form of deep mutual dependence in which the "mingling of interests and activities is so profound, that it is called entanglement". For instance, Internet traffic and the World Wide Web, ICT infrastructures (such as undersea cables), the establishment of technology supply chains and co-ventures and other forms of economic partnerships between foreign technology companies all contribute to deepening entanglement between states. Moreover, as suggested by Lindsay (2017), the development of digital technology is inherently cooperative, and few countries have the ability to design and produce technological equipment entirely by themselves (Lysne, 2018), or at least without information previously provided by/accessed from another state. As a result, international cooperation in cyberspace is essential, generating in turn a high dependency between states' cyber industrial complex. Going back to the quantum metaphor, it means that entangled states become almost inseparable.

Therefore, I define cyber entanglement as a state of deep integration between two or more particles of the international system where complete separation is rendered impossible because of their various, pervasive and untraceable interdependencies. In order to appropriately describe and understand the relationship between two states, and even adversaries, one needs to evaluate and appreciate the existence of cyber entanglement and the impact it has on the overall relationship.

This paper argues that the notion of interdependence, anchored in traditional IR theory, offers a foundation to the study of inter-state relations, but that entanglement extends upon and enhances the IR concept of interdependence, by implying an unbreakable bond between entangled entities. The premise of economic interdependence theory is that economic exchanges create incentives for collaboration and decrease the motivations for war. Similarly, the premise of this framework is that there is a correlation between cyber entanglement between countries and their behaviour: cyber entanglement creates incentives for collaboration and diminishes the likeliness of armed conflict. The relationship can still be characterized by economic, political and military competition, but cyber entanglement ultimately drives the two parties toward collaboration. Figure 2 below illustrates this idea as if it were a straight-line correlation. However, entanglement must not be understood as the direct continuation of interdependence. Rather, entanglement and its impact on inter-states relations is qualitatively different from interdependence. Basic IR interdependence theory suggests that states have the choice to cease their interactions when they see fit, for example in times of conflict. Cyber entanglement assumes on the contrary that states are in a situation of unavoidable interaction because of the characteristics of cyberspace, which in turn makes war unlikely. The very constitution of cyberspace restrains conflict, and forces states to collaborate even while competing (Lindsay, 2017). Therefore, this paper proposes that any framework for the analysis of the U.S.-China interactions in cyberspace has to assume that the relationship is impacted by cyber entanglement.



**Figure 2:** Correlation between entanglement and conflict

The vertical axis is a reproduction of the spectrum of conflict-accord introduced by Shambaugh (2013) to conceptualize the U.S.-China relationship. It refers to the type of relationship between two states spanning from conflict towards general accord. In Figure 2, “conflict” refers to armed conflict understood as a situation where there are hostilities, which may include or be limited to cyber operations, between two or more states (Schmitt, 2017), and where weapons (kinetic or cyber) are used to cause damage or destruction to objects or injury or death to persons (Schmitt, 2017, 452). Conversely, the term “accord” in this spectrum refers to total and complete harmony.

The horizontal axis reproduces the cyber entanglement spectrum introduced earlier, where the starting end corresponds to total independence – no relation – and the final end corresponds to entanglement – no separation. The graphic illustrates that as entanglement increases, so does collaboration. In the “No relation” zone, there is a greater risk for conflict. The “Traditional relations” zone refers to states which share relations, such as trade and military alliances, to a level that can reach and extend beyond interdependence. This, by extension, increases the level of collaboration. Finally, the “Cyberspace” zone corresponds to cyber entanglement between states, and thus generates a relationship where states could potentially reach the greatest level of accord. In summary, the two-pronged hypothesis at the heart of this paper is that cyber entanglement means that tensions are less likely to escalate to the level of armed conflict, and that it instead drives states to collaborate. The next section will expand on the U.S.-China case-study to demonstrate this correlation.

## **5. The U.S.-China case**

The United States is threatened by China’s technological and economic growth. In a report to the congressional committees, the Government Accountability Office (GAO) affirmed that China’s global expansion is a long-range emerging threat facing the United States. China is perceived as an adversary, especially with regard to their cyber capabilities (U.S. Government Accountability Office, 2018). In 2018, President Donald Trump initiated a “trade war” with China, imposing tariffs on Chinese goods and thus creating a situation of “severe economic competition” between the two powers. This competition has the potential to generate a Thucydides’ trap (Rudd, 2018), with the two states heading towards war. The Peloponnesian reference is used to explain the likelihood of conflict occurrence between a rising power and a currently dominant one (Allison, 2017). However, despite their competing national interests, cyber entanglement makes war between China and the U.S. unlikely.

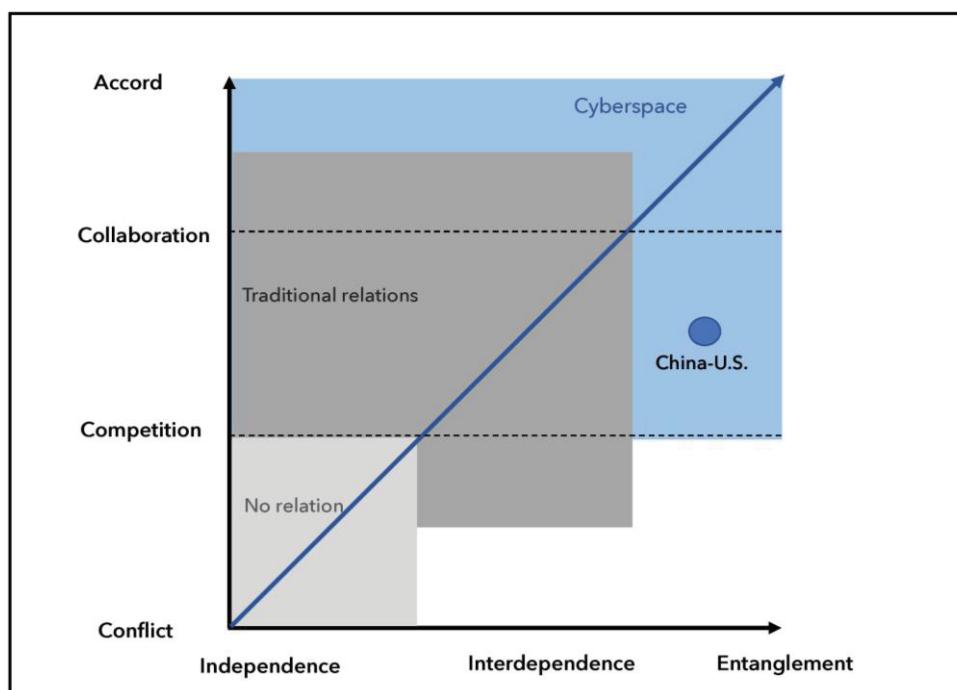
First, the two powers’ technology sectors are highly interconnected, and political differences between the United States and China did not previously stop the two countries’ economies from becoming “increasingly intertwined” (Moak and Lee, 2015). According to the U.S. Bureau of East Asian and Pacific Affairs (2018, August 22), despite the current tensions, China remains the third largest export market for U.S. goods, while the United States is the largest Chinese export market. China and the U.S. profit greatly from their trade relations, particularly in the ICT sector. American technology companies need China’s market to export their products and services, while China sees the technology supply chains as integral to its economic development (Inkster, 2016). For instance, Qualcomm, Intel, and Apple have a significant share of their global revenue coming from China (Bergsten et al, 2014). In 2016, China imported \$228 billion worth of integrated circuits, accounting for over 90 percent of its consumption (Segal, 2018). Additionally, China is also an enormous supplier for American firms. Out of seven major American IT manufacturers an average of 51 percent of shipments to these companies between September 8, 2012, and September 7, 2017 originated in China (Beeny et al, 2018). In the case of Microsoft, 73% of its imports came from China (Beeny et al, 2018).

Second, entanglement is increased by the fact that almost all critical sectors in both countries rely on digital technologies, which are produced, installed, operated and owned largely by foreign companies. For example, the American enterprise Cisco has a market share in China of more than 50 percent in information infrastructure in critical sectors such as banking, defence, and government (Bergsten et al, 2014). In a paper by the Rand Corporation on the relationship between China and the United States in cyberspace, it was reported that “Chinese representatives cited numerous ways in which their country was dependent on U.S. capabilities: their credit card and airline reservations systems were housed in the United States; their emergency communications depended on a U.S. corporation; their offices depended on Microsoft” (Harold et al, 2016, 50). Moreover, in 2013, the state-sponsored *China Economic Weekly* published an article stating that eight large U.S. IT firms, called the “guardian warriors”, had become indispensable to China’s information infrastructure – Cisco, IBM, Google, Qualcomm, Intel, Apple, Oracle and Microsoft (Bergsten et al, 2014; Tiezzi, 2015). Some of the firms’

projects in 2013 included “Cisco’s upgrades of the People’s Bank of China’s Intranet, IBM’s facilitation in building the Yunnan province police bureau’s database, and Microsoft’s improvements to China Eastern Air’s information technology” (Rosen, 2013). Therefore, Chinese leaders know that the country’s cyber sector is (and will remain in the coming decades) highly dependent on technology transfer from the most advanced countries, including the United States (Austin, 2018a). China and the U.S. also share the ownership of essential physical elements of the global communication system, such as submarine cables. For example, the New Cross Pacific (NPC) cable system is owned and/or operated by a consortium consisting of China Mobile, China Telecom, China Unicom, Chunghwa Telecom, KT, Microsoft, Softbank Telecom (2019). These cables are spread around the globe and connect countries and continents (Sechrist, Vaishnav, Goldsmith et al, 2012), creating even more entanglement between nations.

Finally, one of the main counterarguments with regard to entanglement can be based on the principle of cyber sovereignty, particularly articulated in the Chinese discourse (Lu, 2015; Xi, 2015; Inkster, 2016; Klimburg, 2017; Fang, 2018). However, despite its cyber sovereignty rhetoric, China believes that there is more to gain than to lose by being connected to the Internet (Austin, 2018a). Chinese president Xi Jinping expressed it in a 2016 speech on cybersecurity: “cybersecurity is open rather than closed” and “it can only improve if we strengthen international exchanges, cooperation and interaction”(Austin, 2018a, 6). Both the United States and China are highly invested in the current global system enabled by information technologies and their interconnection; therefore, they have no incentives to “disconnect”(Lindsay, 2015). Furthermore, cyber entanglement presupposes that separation is impossible, and so cyber sovereignty – understood as total independence and control over a portion of the Internet – is highly improbable because it does not account for the inherent decentralized and entangled infrastructure of cyberspace. “Even China’s leaders realize that our systems are so entangled – the word “entangled” is correct in a literal sense – that they cannot be separated” (Schwartz, 2011, 224). In other words, states cannot hope to enjoy the benefits of interconnexion while separating themselves from the very infrastructure that enables it. As put by Lindsay (2015, 40), “China cannot credibly commit to abide by its own norm of internet sovereignty. And second, China benefits from the current system”.

To summarize, the incentives for collaborative behaviour are much greater than those for hostility. Cyber entanglement between China and the United States means that a destructive conflict is unlikely, and that the two powers are instead driven toward collaboration. Shambaugh (2013) argues that the U.S.-China relationship is situated in the middle between competition and cooperation, never achieving accord but avoiding conflict. Accordingly, in the graphic illustration, their relation would be situated in the “Cyberspace” zone, as shown below.



**Figure 3:** U.S.-China cyber entanglement

This analysis suggests that China will refrain from launching any potentially destructive cyberattack that would amount to the physical destruction of infrastructure or injury/death of a person against an infrastructure co-owned and operated by a Chinese company. China is not likely to conduct any action that could have a negative impact on its own interests, hence why cyber entanglement is one of the best arguments against conflictual behaviour between China and the U.S.

## **6. Conclusion**

This paper argued that the quantum entanglement metaphor provides a useful conceptualization of inter-state relations in the digital age. States are increasingly entangled in cyberspace, which means that are deeply integrated through various interdependencies, such as technological trade and the sharing of hardware infrastructure and software. Their own prosperity and stability depend on their access to cyberspace and on their exchanges with each other through the means of information and communication technologies (ICTs). Cyber entanglement, the conceptual framework introduced in this paper, fosters a better understanding of the U.S.-China relationship in cyberspace. Notwithstanding their competing national interests, the two great powers are entangled together in, and because of, cyberspace. The cyber entanglement conceptual framework could therefore be useful for leaders and policymakers to navigate their tensions and make their way out of Thucydides' trap. Future work includes the development of specific detailed IR examples and case studies to support the creation of a mature Cyber Entanglement Conceptual Framework.

## **Acknowledgements**

I would like to acknowledge Prof Greg Austin and Dr David Ormrod for their guidance and contribution to this paper and overall mentorship. Many thanks also to my colleagues Ms Rhiannon Neilsen, for her helpful contribution to editing, and Ms Lena Sentker for her precious help with the graphic representations.

## **References**

- (2019) *Submarine Cable Map* [Online], PriMetrica, Inc., Available: <https://www.submarinecablemap.com/>.
- Adesso, G. (2007) The social aspects of quantum entanglement, *Ordint la Trama*, No. 56.
- Allison, G. (2017) *Destined for war: Can America and China escape Thucydides's trap?*, Houghton Mifflin Harcourt, Boston, New York.
- Austin, G. (2018a) *Cybersecurity in China: The Next Wave*, Springer, Cham, Switzerland.
- Austin, G. (2018b) *Opportunity, Threat and Dependency in the Social Infosphere*, Unpublished manuscript.
- Barad, K. (2007) *Meeting the Universe Halfway : Quantum Physics and the Entanglement of Matter and Meaning*, Duke University Press.
- Barad, K. (2010) Quantum Entanglements and Hauntological Relations of Inheritance: Dis/continuities, SpaceTime Enfoldings, and Justice-to-Come, *Derrida Today*, Vol 3, No. 2, pp 240–268.
- Beeny, T., Bisceglie, J., Wildasin, B. and Cheng, D. (2018) Supply Chain Vulnerabilities from China in U.S. Federal Information and Communications Technology. , U.S.- China Economic and Security Review Commission Interos Solutions, Inc.
- Bergsten, C. F., Hufbauer, G. C. and Miner, S. (2014) *Bridging the Pacific : toward free trade and investment between China and the United States*, Peterson Institute for International Economics, Washington, D.C. .
- Bolt, P. J. and Shearn, B. (2014) Cyberpower and cross-Straits security, In: CHU, M. M. & KASTNER, S. L. (eds.) *Globalization and security relations across the Taiwan strait : in the shadow of China*, Routledge.
- Brantly, A. (2018) Conceptualizing Cyber Deterrence by Entanglement, *SSRN Electronic Journal*.
- Bureau of East Asian and Pacific Affairs (2018, August 22) *U.S. Relations With China* [Online], U.S. Department of State, Available: <https://www.state.gov/r/pa/ei/bgn/18902.htm>.
- Deibert, R. (2017) Cyber Security, In: CAVELEY, M. D. & BALZACQ, T. (eds.) *Routledge Handbook of Security Studies*, 2nd ed, Routledge, New York.
- Fang, B. (2018) Cyberspace sovereignty : reflections on building a community of common future in cyberspace. Singapore: Springer.
- Foerster, S. (2012) Strategies of deterrence, In: JASPER, S. (ed.) *Conflict and cooperation in the global commons: a comprehensive approach for international security*, Georgetown University Press, Washington DC.
- Friedman, T. (2007) *The World Is Flat: A Brief History of the Twenty-First Century*, Picador/Farrar, Straus and Giroux, New York.
- Government of Canada (2018) *Canada-United States relations* [Online], Available: [https://international.gc.ca/world-monde/country-pays/united\\_states-etats\\_unis/relations.aspx?lang=eng](https://international.gc.ca/world-monde/country-pays/united_states-etats_unis/relations.aspx?lang=eng).
- Hamilton, S. (2017) Securing ourselves from ourselves? The paradox of entanglement in the Anthropocene, *Crime, Law and Social Change*, Vol 68, No. 5, pp 579-595.
- Harknett, R. J. and Nye, J. S. (2017) Is Deterrence Possible in Cyberspace?, *International Security*, Vol 42, No. 2, pp 196-199.
- Harold, S. W., Libicki, M. C. and Cevallos, A. (2016) *Getting to Yes with China in Cyberspace*, RAND Corporation.

- Inkster, N. (2016) *China's Cyber Power*, Routledge, Abingdon.
- Kai, J. (2014). Why China Banned Windows 8, *The Diplomat*, 28 May.
- Klimburg, A. (2017) *The Darkening Web*, Penguin Press, New York.
- Lindsay, J. R. (2015) The Impact of China on Cybersecurity: Fiction and Friction, *International Security*, Vol 39, No. 3, pp 7–47.
- Lindsay, J. R. (2017) Restrained by design: the political economy of cybersecurity, *Digital Policy, Regulation and Governance*, Vol 19, No. 6, pp 493-514.
- Lu, W. (2015) Cyber Sovereignty Must Rule Global Internet, *The Huffington Post* [Online].  
[https://www.huffpost.com/entry/china-cyber-sovereignty\\_b\\_6324060](https://www.huffpost.com/entry/china-cyber-sovereignty_b_6324060).
- Lysne, O. (2018) The Huawei and Snowden questions: Can electronic equipment from untrusted vendors be verified? Can an untrusted vendor build trust into electronic equipment? Cham, Switzerland: Springer Open.
- Moak, K. and Lee, M. W. N. (2015) *China's economic rise and its global impact*, Palgrave Macmillan, New York.
- Montgomery, A. H. (2016) Quantum mechanisms: Expanding the boundaries of power, space, and time in global security studies, *Journal of Global Security Studies*, Vol 1, No. 1, pp 102-106.
- Myerson, T. (2017) Announcing Windows 10 China Government Edition and the new Surface Pro, *Microsoft Blog* [Online].  
<https://blogs.windows.com/windowsexperience/2017/05/23/announcing-windows-10-china-government-edition-new-surface-pro/>.
- Nye, J. S. (2017) Deterrence and Dissuasion in Cyberspace, *International Security*, Vol 41, No. 3, pp 44–71.
- Ormrod, D. and Turnbull, B. (2016) The cyber conceptual framework for developing military doctrine, *Defence Studies*, Vol 16, No. 3, pp 270-298.
- Raimond, J. M., Brune, M. and Haroche, S. (2001) Manipulating quantum entanglement with atoms and photons in a cavity, *Reviews of Modern Physics*, Vol 73, No. 3, pp 565-582.
- Rosen, D. H. (2013) Eight Guardian Warriors: PRISM and Its Implications for US Businesses in China, *Rhodium Group* [Online]. <https://rhg.com/research/eight-guardian-warriors-prism-and-its-implications-for-us-businesses-in-china/>.
- Rudd, K. (2018) How to Avoid an Avoidable War: Ten Questions About the New U.S. China Strategy, *Foreign Affairs*.
- Schmitt, M. N. (2017) *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, Cambridge University Press, New York, NY.
- Schrödinger, E. (1935) Discussion of Probability Relations between Separated Systems, *Mathematical Proceedings of the Cambridge Philosophical Society*, Vol 31, No. 4, pp 555-563.
- Schwartz, P. (2011) Scenarios for the Future of Cyber Security, In: LORD, K. M. & SHARP, T. (eds.) *America's Cyber Future: Security and Prosperity in the Information Age*, Center for a New American Security, Washington, D.C.
- Sechrist, M., Vaishnav, C., Goldsmith, D. and Choucri, N. (2012) The Dynamics of Undersea Cables: Emerging Opportunities and Pitfalls., In: HUSEMANN, E., & LANE D. (ed.) *30th International Conference of the System Dynamics Society*. St. Gallen, Switzerland.
- Segal, A. (2018) When China Rules the Web: Technology in Service of the State, *Foreign Affairs*, Vol 97, No. 5, pp 10-18.
- Shambaugh, D. (ed.) (2013) *Tangled Titans : The United States and China*, Rowman & Littlefield Publishers, Inc., Lanham, Maryland
- Singer, P. W. and Friedman, A. (2014) *Cybersecurity and cyberwar: what everyone needs to know*, Oxford University Press, New York.
- Tiezzi, S. (2015). New report highlights China's cybersecurity nightmare, *The Diplomat*, February 18.
- U.S. Government Accountability Office (2018) National Security: Long-Range Emerging Threats Facing the United States As Identified by Federal Agencies.
- Xi, J. (2015) President Xi Jinping delivers a keynote speech at 2015 World Internet Conference.

# Effects of Cyber Domain in Crisis Management

Jussi Simola and Martti Lehto

University of Jyväskylä, Faculty of Information Technology, Finland

[juhemisi@student.jyu.fi](mailto:juhemisi@student.jyu.fi)

[martti.j.lehto@jyu.fi](mailto:martti.j.lehto@jyu.fi)

**Abstract:** There is fundamental need in EU-level to develop common alarm procedures and emergency response models with preventive functions which work well from local to national level and from national to international level. European Public Protection and Disaster Relief (PPDR) services such as law enforcement, firefighting, emergency medical and disaster recovery services have recognized that lack of interoperability of technical systems limits cooperation between the PPDR authorities. Also, the military (MIL) and critical infrastructure protection (CIP) faces similar challenges. Recent major accidents have indicated that lack of human resources affects to disaster recovery. PPDR-actors cannot start operations, if there is a human factor preventing the flow of information. Preventing a domino effect after a disaster may be delayed. There is a need to understand how public safety authorities can act in a preventive manner so that a potential accident or offense can be prevented in advance. This paper's goal is to find out main factors which affect to implementing of the next generation hybrid emergency response system for critical infrastructure protection. Early detection of any threat and rapid response to neutralize the threat may help to save human lives and vital functions before any disaster occurs. By comparing present emergency response processes to the next generation Smart hybrid emergency process model, it can be found effects and factors which prevent to implement this architecture. For example, legislation, organizational changes, lack of using cyber dimension and emergency procedures effects to combine different kind of PPDR -functions. Cyber dimension as a part of situational awareness raises its value for the continuity management. For traditional purposes, PPDR services are being seen as separate physical operational functions. This study proposes to solve the problems of development needs through technical, organizational and structural alternatives. The main issue regarding dividing reliable decision support information to decision-makers is related to at which point in chain-reaction a human action is more harmful than useful. It has been seen in earlier empirical studies that human activities may prevent to manage functions of essential emergency response procedures during a disaster. It's necessary to create emergency response model, that will be functionally capable and modern combining cyber and physical elements in a right proportion.

---

**Keywords:** critical infrastructure protection, cyber-physical threats, emergency response, PPDR, continuity management

## 1. Introduction

European decision-makers like politicians have recognized, that it's not enough to start emergency response procedures in traditional way at the scene of an accident or a catastrophe. Nowadays hybrid attacks against critical infrastructure are based on combination of different kind of threats. Human factors, technological communication problems and lack of interaction between different PPDR actors show challenges at the scene of an accident. It's necessary to take into account these things before starting to build the next generations emergency response model.

Thanks to the rapid development of information systems, national legislation has also faced new challenges. On the other hand, practiced policy in Europe has been based on the image of the world that free movement between countries should be facilitated in Europe. The obstacles to free movement were reduced in the Schengen area until the terrorism that came with the Middle East refugee wave forced the European decision-makers to change the political lines. Terrorist attacks in the United States, Australia, France, Belgium, Germany, Sweden and Finland have changed the weighting of security issues.

The EU's internal and external border control have been intensified and the conditions for asylum applications have been revised. EU information systems projects have become increasingly multinational. Security has been perceived as a common EU affair, no longer a separate national task. The importance of legislation is emphasized when building common IT structures and platforms for information systems. National legislation may become an obstacle, especially in situations where other partner countries have implemented laws that support new IT solutions. The outline of the paper is as follows. After the introduction section 2 presents theoretical framework and central concepts of the paper. Section 3 handles research background, objectives and methods. Section 4 handles findings. Section 5 include discussion and section 6 conclusions.

## **2. Theoretical framework and literature review**

In the future it's not enough to develop separate technological solutions for critical infrastructure protecting. In EU-level there is a need to reach common situational picture when cross-boarding threat like cyberattack has occurred. Smart nations or European union needs cooperation between smart cities, because without smart cities smart nation cannot form. Thus smart information systems are being developed, it's important that there is already infrastructure where to connect the system. Every smart city should be construct from a long-term view. Smart city needs urban built environment. This case study aims to find out those factors which affect to implementation of the Hybrid Emergency Response Model. There are separate situation centers, emergency response centers and organizations fighting against cyber threat's, but there is no common emergency response model for all kind of hybrid-threats. The author of this research has innovated next generation emergency response model (Simola & Rajamäki, 2017). It's necessary to research things that are setting barriers to implementation process.

The proposed cross-boarding intelligent emergency management system will provide next generation emergency response model for state decision-makers and PPDR-authorities. The model will combine different data sources, analyze them and produce predictive emergency actions before an alarming accident has occurred. Developed Hybrid Emergency Response -model is one kind of concept which can be expanded to the maritime surveillance environment.

### **2.1 Data protection regulation in EU countries**

The EU General Data Protection Regulation (GDPR) harmonize data privacy laws across Europe. The law is technology neutral and applies to both automated and manual processing if the data is organized in accordance with pre-defined criteria (European Commission, 2016b). The purpose is to protect and empower all EU citizens data privacy and to reshape the way organizations across the region approach data privacy. GDPR applies to all businesses offering goods and/or services to EU. That means that the organizations do not have to reside in the EU area or even in Europe. If you are holding private information about an EU citizen whom you provide services, GDPR applies (European Commission, 2016b).

Personal data cover e.g. name, address, email address, an internet protocol address, location data on a mobile phone and a cookie ID, the advertising identifier of your phone. In some cases, there is a specific sectoral legislation regulating for instance the use of location data or the use of cookies. Directive presents mostly a continuation of earlier Data Protection Directives efforts (European Commission, 2016b).

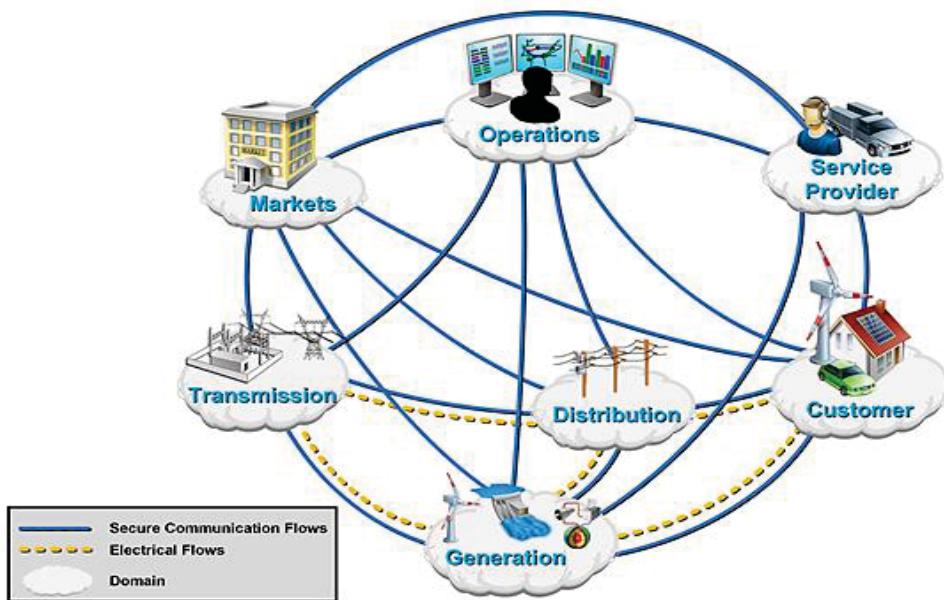
Eu directive named the ePrivacy 2002/58 has been amended by Directive 2009/136, which introduces several changes, especially in what concerns cookies, that are now subject to prior consent. The directive does not apply to issues concerning criminal law and state security, public security and defense. The interception of data is covered by the new EU Data Retention Directive the purpose of which is to amend E-Privacy Directive (IBP, 2014)

The EU Data Protection Directive 2016/680 or Law Enforcement Directive regulates on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data. This proposal applies cross-border and national processing of data by member states' competent authorities for the purpose of law enforcement. This comprise e.g. the prevention, investigation, detection and prosecution of criminal offences, the safeguarding and prevention of threats to public security (European Commission, 2016a).

### **2.2 Central concepts**

#### **2.2.1 Smart city, nation and infrastructure**

Internet of Things connects systems, sensors and actuator instruments to the broader internet. IOT allows the things to communicate, exchange control data and other necessary information while executing applications towards machine goal (Electrical Technology, 2016). Fig. 1. Illustrates secure communication flows, electrical flows and different domains (Updated NIST Smart Grid Framework 3.0, Feb 2014).



**Figure 1:** Interaction of actors in different smart grid domains

Cybersecurity risks should be addressed as organizations implement and maintain their smart grid systems (National Institute of Standards and Technology, 2014). A smart grid system may consist of information technology which is a discrete system of electronic information resources organized for the collection, processing, maintenance, use, sharing, dissemination or disposition of information. A smart grid system may also consist of operational technologies (OT) or industrial control systems (ICS) like SCADA systems, distributed control systems (DCS), and other control system configurations (CHONG & KUMAR, 2003; National Institute of Standards and Technology, 2014). Industrial Internet of Things (IIoT) collects data from connected devices (i.e., smart connected devices and machines) in the field or plant and then processes this data using sophisticated software and networking tools. The entire IIoT requires a collection of hardware, software, communications and networking technologies (Electrical Technology, 2016).

Critical Information Infrastructure means any physical or virtual information system that controls, process, transmits, receives or stores electronic information in any form including data, voice or video that is vital to the functioning of critical infrastructure. Critical infrastructure (CI) includes energy production, transmission and distribution networks, ICT systems, networks and services (including mass communication), financial services, transport and logistics, water supply, construction and maintenance of infrastructure, waste management in special circumstances. That smart network will integrate information and communication technologies with the power-delivery infrastructure (Ahokas, Guday, Lyytinen, & Rajamäki, 2010; Ministry of the Interior, 2016).

### 2.2.2 Sensors for monitoring, buildings, bridges and other structures

The development of more robust and advanced smart sensors could help provide valuable information about the health of various structures, including bridges, tunnels, buildings like shopping malls and water distribution systems. Sensors can provide valuable insight on the structural health and condition of bridges or buildings. In the future building can monitor the activities of all individuals inside the building. In the future buildings, bridges and shopping malls are part of smart city and smart grid (NIST, 2012).

### 2.2.3 Location based sensors

Retailers of malls may use indoor or/and outdoor navigation technologies to provide location-based services using mobile “push” notifications to provide advertisements. Technologies are currently available to not only locate a customer but are also be able to establish history of a path taken by a typical customer during the day (Kini & Suomi, 2018; Rachel, 2013). Advertisement networks are able to locate and custom-deliver an advertisement to customer with or without customer’s permission. With this technology it is possible to provide personalized marketing based on the consumer’s location. If mobile users give permission (opt-in) to the companies whose brand, products and services they like, companies send them personalized advertisements when they are shopping (Yiu, Jensen, Møller, & Lu, 2011).

#### *2.2.4 Cyber infrastructure and cyber physical systems*

The term cyber-physical systems (CPS) was coined by Helen Gill at the National Science Foundation in the U.S. to refer to the integration of computation with physical processes. In CPS, embedded computers and networks monitor and control the physical processes. CPS are enabling next generation of “smart systems” like advanced robotics, computer-controlled processes and real-time integrated systems (Lee & Seshia, 2015). Cyber Infrastructure Includes electronic information, communications systems, services and the information contained in these systems and services. Information and communications systems and services are composed of all hardware and software that process, store, and communicate information or any combination of all of these elements. Processing includes the creation, access, modification and destruction of information. Storage includes paper, magnetic, electronic, and all other media types (National Institute of Standards and Technology, 2014). According to Franke and Brynielsson (2014), cyber situational awareness is a subset of situational awareness, i.e., cyber situational awareness is the part of situational awareness which concerns the “cyber” environment. Such situational awareness can be reached, e.g. by the use of data from IT sensors (intrusion detection systems, etc.) that can be fed to a data fusion process or be interpreted directly by the decision-maker (Franke & Brynielsson, 2014). Communications include sharing and distribution of information, e.g. computer systems; control systems (e.g., supervisory control and data acquisition—SCADA); networks, such as the Internet; and cyber services (e.g., managed security services) which are part of cyber infrastructure.

#### *2.2.5 Cyber and hybrid threats*

According to DHS & Office of Emergency Communications (2016) cyber threats can be illustrated in many ways. Potential Risks to emergency response system components may be formed from devices or equipment, network infrastructure and connections or data applications and services. In spear-phishing attack means that a criminal finds a webpage for his target organization that supplies contact information for the company. Using available details to make the message seem authentic, the criminal drafts an email to an employee on the contact page that appears to come from an individual who might reasonably request confidential information, such as a network administrator. The email asks the employee to log into a false page that requests the employee's username and password or click on a link that will download spyware or other malicious programming (Rouse, 2017). Data breaches mean data has stored on user device and it is accessed, manipulated or stolen. Users may download malicious software “malware” (e.g., botnets, viruses, spyware, trojans and rootkits). It is called “Man-in-the-middle attack” when wireless link between the user device and the tower may be susceptible and allow attackers to steal data or monitor conversations. In Denial-of-service (Dos) attack, criminals overload towers or other key network resources with requests for network access, damaging or destroying the operability of the targeted infrastructure and straining the capacity and resiliency of the network. Insider threats: Employees or other authorized personnel may produce insider threats when they use their access to steal, corrupt, or destroy data. In malicious applications attackers create applications that appear to be safe but allow them to steal, corrupt or modify data, eavesdrop on conversations, or acquire data on the location of victims and/or first responders (DHS & Office of Emergency Communications, 2016). Hybrid threat means for example combination of different kind of physical and cyber threats.

#### *2.2.6 PPDR services*

The term “Public Protection” is used to describe critical public services that have been created to provide primary law enforcement, firefighting, emergency medical and disaster recovery services for the citizens of the political subdivision of each country. The term Public Safety and Disaster Response, within certain regions, can also be construed as PPDR. The military (MIL) and critical infrastructure protection (CIP) are also included in the term (Baldini, 2010).

The Emergency Response Centre Administration provides emergency response center services throughout Finland. The duty of the Emergency Response Centre Administration is to receive emergency calls from all over the country for the rescue, police and social and health services; handle communications relating to the safety of people, property and the environment; and relay the information they receive to the appropriate assisting authorities or partners (National Emergency Number Association (NENA) and the Association of Public-Safety Communications Officials (APCO), 2016; The emergency response act, 2010).

### **2.2.7 Emergency response management information systems**

Traditional Emergency Response System should consist of at least basic components like a database, data analysis capability, normative models and interfaces. E.g. personnel in Emergency Response Center use Emergency Response system. It is one kind of DSS system. Decision support systems are used to track key incidents and the progress of responding units, to optimize response activities and to act as a mechanism for queuing ongoing incidents (Ashish et al., 2007; Endsley, 1988; Endsley, 1995).

Situation center means the place where PPDR authorities make decisions to allocate resources to the right proportion. The words Command and Control individually and collectively mean different things to different communities (Alberts & Hayes, 2006). C2, situation center or Emergency Operation Room is a physical or virtual location designed to support emergency response, business continuity and crisis communications activities. PPDR authorities meets at the C2 -room to manage preparations for an impending event or manage the response to an ongoing incident. By gathering the decision makers together and supplying them with the most current information can be made better decisions (Ashish et al., 2007). In systems engineering, monitoring means a process within a distributed system for collecting and storing state data. A PPDR monitoring station is a workstation or place in which sensor information accumulates for end users who need it. Monitoring systems include information collection, analysis and provision for end-users, which is front-deployed knowledge. Government Situation Centre ensure that the state leaders and central government authorities are kept informed continuously (Ministry of defence, 2010).

### **2.2.8 Open Source Intelligence as a part of the HERM**

OSINT is defined as the systematic collection, processing, analysis and production, classification and dissemination of information derived from sources openly available to and legally accessible by the public in response to particular government requirements serving national security. It is any unclassified information, in any medium, that is generally available to the public, even if its distribution is limited or only available upon payment (Glassman & Kang, 2012; Morrow & Odierno, 2012; Nurmi, 2015).

## **3. Research background, objectives and methods**

At present public safety authorities (PPDR) do not use cyber dimension in their daily routine at all. The problem is that public safety authorities have separate Cyber security organizations with own administrations. Organizations which have responsibilities for cyber security operations are separated from PPDR services. As a part of FICORA, The National Cyber Security Centre Finland (NSCS-FI) produce information of Cyber threats for stakeholders, but that data does not reach e.g. emergency response centers or situation centers. Separate organizational cyber security functions, methods and procedures prevent effective response for cyber physical threats. Combining Open Source Intelligence data (Morrow & Odierno, 2012) and traditional intelligence sources overall situational awareness arises. Hybrid threats need coordinated hybrid responses, therefore also a cyber situational picture is needed.

### **3.1 Method and process**

#### **3.1.1 Case study research strategy**

Empirical approach helps to understand PPDR authorities' entity. Choosing a case study research strategy enables investigation of interaction between the different factors. The multimethodological approach consists of four case study research strategies: theory building, experimentation, observation and systems development (Nunamaker, Minder Chen, & Purdin, 1991). Yin (2014) identifies five components of research design for case studies: (1) the questions of the study; (2) its propositions, if any; (3) its unit(s) of analysis; (4) the logic linking the data to the propositions; and (5) the criteria for interpreting the findings. This case study is carried out with the guidance of Yin (2014). This research concentrates in sources of scientific publications, collected articles and literary material.

#### **3.1.2 Analyzing vulnerabilities**

We can divide research-area in four sections; local, regional, national and European level. Local PPDR-area consists of one city or municipalities, regional area is wider area including organizations like regional administration with PPDR-authorities, cities and municipalities. The focus of the research is on the protection of

critical infrastructure at local and regional level and how the current EMS system could be developed to be able to respond to the future challenges of cross-border cooperation between PPDR authorities. This is an important question, because there is a common need to develop interoperability between information systems within European Union member countries. Firstly, next generation emergency response system should work in lowest local level before it can be connected to the next level.

We have used combination of different methodologies to find out those factors which affects to introduction of the next generation emergency response model. The Framework by National Institute of Standards and Technology focuses on using business drivers to guide cybersecurity activities and considering cybersecurity risks as part of the organization's risk management processes. The Framework will help an organization to understand, align and prioritize its cybersecurity activities with its mission requirements, risk tolerances and resources. The Tiers or levels provide a mechanism for organizations to view and understand the characteristics of their approach to managing cybersecurity risk, which will help in prioritizing and achieving cybersecurity objectives (DHS & Office of Emergency Communications, 2016; National Institute of Standards and Technology, 2018).

A prescriptive metric known as Technology Readiness Level (TRL) has been used mainly by NASA, the DoD, the DoE and the Department of Homeland Security to address the readiness of systems under development. TRLs have been adapted for biomedical systems, modeling and simulation technologies, learning systems and software intensive systems, among others. Additional readiness levels (RLs) were developed to meet specific needs. Proliferation does emerge as a problem when the tendency is to add new undefined RLs that did not have the quality control in their construction as the original (Perseus, 2013).

To address integration, another metric called Integration Readiness Level (IRL) was introduced by the Systems Development & Maturity Laboratory (SysDML) at Stevens Institute of Technology. The introduction of an IRL to the assessment process not only provides a check as to where a technology is on an integration readiness scale but also presents a direction for improving integration with other technologies. Combining both TRL and IRL scales it is possible to form a knowledge base on the technological maturity level of the emergency response services infrastructure. Tier levels 1-3 are used instead of 1-9 in this research. Two emergency response systems were compared with each other; present system and the next generation hybrid emergency response model.

Three main categories have been chosen to classifications:

- Legislation concerning the smart hybrid model
- Technological maturity Level
- Readiness Level of organizational and political view

The approach of the research is at the local and regional level and it includes the intelligent city area with its authorities and operational functions of situation centers and emergency response information system.

## **4. Results**

### **4.1 Emergency response model for critical infrastructure protection**

The highest state decision-makers, such as members of the Finnish government or highest public safety officers must understand digital entity of the environment where citizens are living. As figure 2 illustrates, formation of cyber-physical threats is gathered from different sources and separate organizations handle those threats. There is no common preventive cyber functionalities or connection between emergency response administration and National Cyber Security Centre Finland which acts under the Finnish Communications Regulatory Authority.

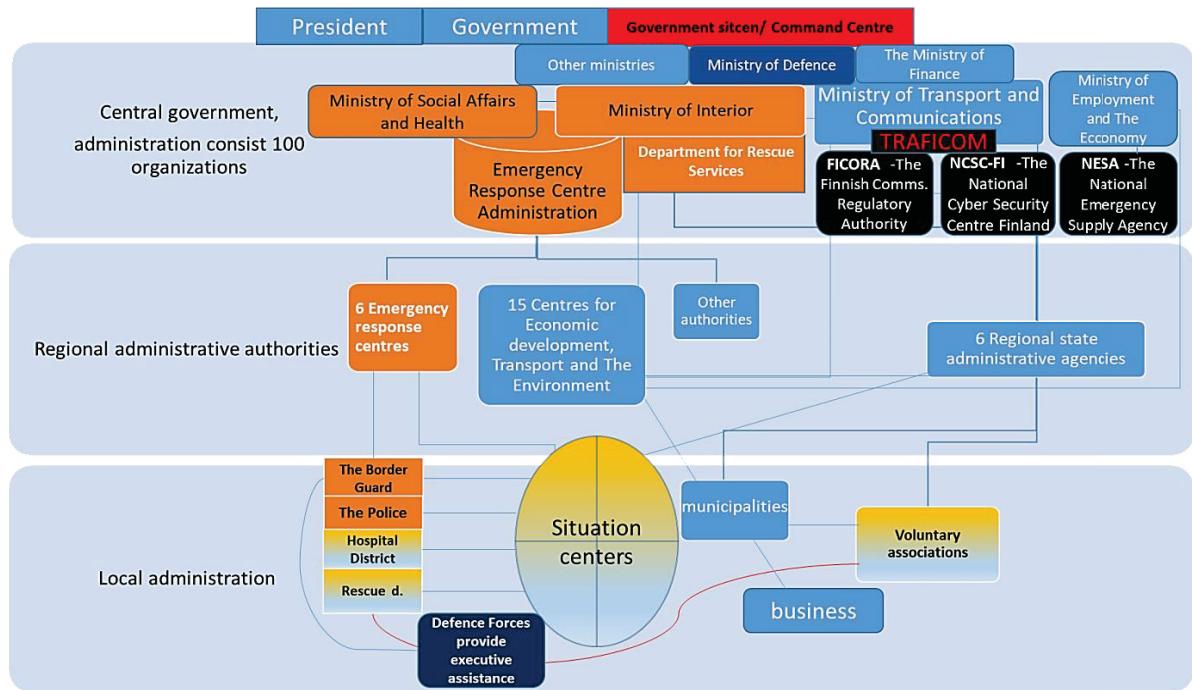


Figure 2: Organizations responsibilities of cyber security

Hybrid emergency response model as a part of smart society and smart city will create a secure framework with efficient procedure to identify and assess national and cross-boarding threats in critical infrastructure. It will provide efficient decision support solution for decision-makers and PPDR authorities how to protect critical infrastructure, but there are fundamental factors which prevent to start implementation of the system. When present national and European development concerning developing next generation emergency response model are taken into consideration the difference of new and the old model illustrated as fig.3.

Maturity level	Low	Med	High	
	1	2	3	
<b>Low (red 1) presents that the maturity level of the system does not correspond the research area. Medium (green 2) presents that the maturity level is average. High (blue 3) presents that the maturity level of the system is ready for the implementation.</b>				
Research areas	Present system	Next gen. HERM syst		
European legislation	1		2	
Legislation concerning technology	3		1	
Legislation concerning privacy issues	3		2	
<i>Legislation concerning the smart hybrid model</i>	sum	7		5
Technological maturity	2		3	
Smart city maturity	1		3	
Maturity of organizational Integration	1		2	
Opportunities to use smart devices	1		3	
Opportunities to integrate sensor tech.	2		3	
Maturity to integrate it-systems	1		3	
Operational reliability	2		2	
<i>Technological maturity Level</i>	sum	10		19
Organizations maturity level	3		1	
The political readiness at national level	3		1	
European policy	1		2	
<i>Readiness Level of organizational and political view</i>	sum	7		4
<b>Total</b>	24		28	

Figure 3: Maturity level of emergency response systems

Firstly, there are organizational factors which prevent to implement new system. Those factors are closely related to legislation, because there is PPDR administration like emergency services which act under the municipalities. Emergency Response Centre acts under Ministry of Interior. On the other Coast guard acts under the Ministry of Interior, but Defense forces act under The Ministry of Defence. There is a lot to do that the operating environment would be favorable to the next generation emergency response system. In Finland the legislation does not give permission for law enforcement to trace citizens digital behavior in real time. Tools like OSINT, Geo-targeting, Geo-fencing with Wi-Fi, Cell Towers and Beacons create a privacy-restricting advertisement and surveillance circuit that aims to trace consumer behavior. These tools are possible to use only with new Hybrid Emergency Response model. Therefore, maturity level for using mobile technologies is so low.

## **5. Discussion**

In democratic society, it must be taken into consideration that privacy concerns and public safety functions both effect to our quality of life. No one wants to live in an environment where citizen's rights and responsibilities are unclearly defined. Important things for us, such as the data privacy issues, can be more relieved on the grounds that the "common good" requires it. How can we then define the common good? This issue has been controversial in Europe. Determining the public interest or limiting the need to protect society has sometimes caused difficulties. The problem is related to situations where protected legal intresses are incompatible. Government agents, utility executives, policymakers and technology providers must agree about a common goal and take actions to accelerate the process towards final deployment, legal and organizational barriers have to be removed. Given the scale of the effort required and the enormity of the challenges ahead, collaboration among different sectors is essential and should be developed through various channels in order to ensure and accelerate the success of the future smart control centers. In a society where the limits of public and private commercial players have become obscured, the risks are also increasing. Citizens should be able to trust decision-makers, authorities, and society that they do not have to constantly think about what kind of digital footprints they are left behind in any department store control unit. As a single datum, separate information of human life is not significant, but if data is combined from the different sources, the position of a citizen as a manager of his or her own life may change significantly.

## **6. Conclusions**

The new intelligence legislation package proposed by the Finnish government would include provisions on the principles of intelligence activities. If the legislation package will be approved, it is expected to enhance the ability of the PPDR authorities to respond on major national and international hybrid threats, because it also allows wider use of new decision support system technologies. It requires clarification of common rules. In other words, in a public place, e.g. in shopping centers privacy protection should be facilitated if citizen accept common rules which have been created in the form of legislation.

People have been irritated by the fact that people's behavior has been collected much more widely, what has been told and uses that are not known. Therefore, it might be important to look at the big picture of the protection of critical infrastructure. What kind of elements can be included in the framework which protect the vital functions of society? When all the things we do leave some data to tracking systems, people have the right to know what information is collected and for what purpose it has been collected. Perhaps even more important thing is to know who is the holder of the personal data and what is the storage time of the data.

The next generation hybrid model will integrate existing surveillance systems and networks with new ones and it based on active operations and automated functions. There is a need to strengthen the entire intelligence ecosystem in maritime and inland. There is also a need in command and control functions to design a combination of a new kind of hybrid sensor technology that uses location based solutions and OSINT tool in order to detect threats in advance because common cyber situational picture is needed. Location based intelligence is an applicable emergency response tool for public safety authorities in shopping malls and in city areas. The presented hybrid model will offer an updated emergency response management model to PPDR services. Effective cooperation between public safety authorities needs a common technology for all authorities and organizational cooperation requires a common infrastructure and clearer and faster connections.

A dynamic cyber-physical infrastructure is needed in order to respond to a rapidly evolving alert situation. The local and state level PPDR -atmosphere can no longer be separated in the traditional sense. Threats have

changed into combinations of threat types and, as a consequence, public safety organizations like the Police or Finnish Border Guard must be able to prevent new kinds of hybrid threats and respond to them. Improving the flow of information between the public sector and citizens, including volunteer associations, is also a relevant part of this framework. It must be possible to prevent and respond faster to the realization of threats. Municipal actors relying on municipal technical resources is not sustainable because cooperation between the Police, Finnish Border Guard and emergency services has developed. A modelling platform for a smart emergency response model can lead to important new results. The cyber domain can be used as a powerful dimension to enhance data fusion to more accurate overall situational awareness. By processing raw data on anomalous behavior in advance, PPDR services can use smart emergency response functions before any threats have occurred.

## **References**

- Ahokas, J., Guday, T., Lyytinen, T., & Rajamäki, J. (2010). Secure and reliable communications for SCADA systems. Paper presented at the *International Journal of Computers and Communications*, 6(3)
- Alberts, D. S., & Hayes, R. E. (2006). *UNDERSTANDING COMMAND AND CONTROL. DoD command and control research program*. Center for Advanced Concepts and Technology (ACT).
- Ashish, N., Kalashnikov, D. V., Mehrotra, S., Venkatasubramanian, N., Eguchi, R., Hegde, R., & Smyth, P. (2007). Situational awareness technologies for disaster response. In H. Chen, E. Reid, J. Sinai, A. Silke & B. Ganoz (Eds.), *Terrorism informatics: Knowledge management and data mining for homeland security*. Springer.
- Baldini, G. (2010). *Report of the workshop on "interoperable communications for safety and security" with recommendations for security research*. (No. JRC60381). Publications of Office of the European Union. doi:10.2788/19075
- CHONG, C., & KUMAR, S. (2003). Sensor networks: Evolution, opportunities and challenges. Paper presented at the *IEEE*, 91(8) 1247-1256.
- DHS, & Office of Emergency Communications. (2016). *Cyber risks to next generation 911*. Department of Homeland Security.
- Electrical Technology. (2016). Internet of things (IOT) and its applications in electrical power industry. Retrieved from <http://www.electricaltechnology.org/2016/07/internet-of-things-iot-and-its-applications-in-electrical-power-industry.html>
- The emergency response act 692/2010, (2010).
- Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. Paper presented at the *Proceedings of the Human Factors Society 32nd Annual Meeting*, 97-101.
- Endsley, M. R. (1995). Toward a theory of situation awareness. *human factors*, (37), 32-64.
- EU data protection directive 2016/680, Directive U.S.C. (2016a).
- General data protection regulation (EU) 2016/679, Regulation U.S.C. (2016b).
- Franke, U., & Brynielsson, J. (2014). *Cyber situational awareness: A systematic review of the literature*. *Computers & security* (pp. 18-31-46) doi: 10.1016/j.cose.2014.06.008
- Glassman, M., & Kang, M.Ju. (2012). Computers in human behavior; intelligence in the internet age: The emergence and evolution of open source intelligence (OSINT).28(2), 673-682.
- IBP. (2014). *European union cyber security strategy and programs handbook. strategic information and regulations*. Washington DC, USA: International Business Publications.
- Kini, R., B, & Suomi, R. (2018). Changing attitudes toward location-based advertising in the USA and Finland, *journal of computer information systems*.58 doi:10.1080/08874417.2016.1192519
- Lee, E., Ashford, & Seshia, S., Arunkumar. (2015). *Introduction to embedded systems, A cyber-physical systems approach* (2nd ed.) Lee & Seshia.
- Ministry of Defence. (2010). *Security strategy for society, government resolution*. Helsinki: Ministry of Defence;
- Ministry of the Interior. (2016). *National risk assessment 2015*. Helsinki: Ministry of the Interior.
- Morrow, J., & Odierno, R. (2012). *Open-source intelligence, ATP 2-22.9, army techniques publication*. ( ). Washington: Headquarters, Department of the U.S. Army.
- National Emergency Number Association (NENA) and the Association of Public-Safety Communications Officials (APCO). (2016). *NENA/APCO next generation 9-1-1 public safety answering point requirements*. ( ). USA: NENA and APCO. Retrieved from [https://www.nena.org/resource/resmgr/Standards/NENA-APCO-REQ-001.1.1-2016\\_N.pdf](https://www.nena.org/resource/resmgr/Standards/NENA-APCO-REQ-001.1.1-2016_N.pdf).
- National Institute of Standards and Technology. (2014). *Guidelines for smart grid cybersecurity national institute of standards and technology, volume 1 - smart grid cybersecurity strategy, architecture, and high-level requirements*. U.S. Department of Commerce. doi:10.6028/NIST.IR.7628r1
- National Institute of Standards and Technology. (2018). *Framework for improving critical infrastructure cybersecurity*. (No.1.1). NIST.
- NIST. (2012). *Cyber-physical systems: Situation analysis of current trends, technologies and challenges*. ( ). Maryland, USA: National Institute of Standards and Technology.
- Nunamaker, J., Minder Chen, J. R., & Purdin, T. (1991). Systems development in information system research. (3), 89-106.
- Nurmi, P. (2015). *OSINT - avointen lähteiden internet-tiedustelu*. Helsinki: Aalto yliopisto.

- Perseus. (2013). *Protection of European seas and borders through the intelligent use of surveillance D26.12 working document: Assessment report*.
- Rachel, M. (2013). MIT Technology review. Retrieved from <https://www.technologyreview.com/s/510491/every-step-you-take-tracked-automatically/>
- Rouse, M. (2017). Spear phishing. Retrieved from <https://searchsecurity.techtarget.com/definition/spear-phishing>
- Simola, J., & Rajamäki, J. (2017). Hybrid Emergency Response Model: Improving Cyber Situational Awareness. Paper presented at the *16th European Conference on Cyber Warfare and Security*, University, College, Dublin, Ireland. 442-451.
- Yin, R. K. (2014). *Case study research, design and methods* (5th ed.). Thousand Oaks: Sage Publications.
- Yiu, M., L., Jensen, C. S., Møller, J., & Lu, H. (2011). Design and analysis of a ranking approach to private location-based services. *ACM Transactions on Database Systems*, 36(2), 10:1-10:42. doi:10.1145/1966385.1966388

# A Data Security Framework for Cloud Computing Adoption: Mozambican Government Cloud Computing

Ambrósio Patrício Vumo<sup>1</sup>, Josef Spillner<sup>2</sup> and Stefan Köpsell<sup>3</sup>

<sup>1</sup>Universidade Eduardo Mondlane, Maputo, Mozambique

<sup>2</sup>Zurich University of Applied Sciences, School of Engineering, Service Prototyping Lab, Switzerland

<sup>3</sup>Technische Universität Dresden, Department of Computer Science, Germany

[ambrosio\\_patrício.vumo@mailbox.tu-dresden.de](mailto:ambrosio_patrício.vumo@mailbox.tu-dresden.de)

[josef.spillner@zhaw.ch](mailto:josef.spillner@zhaw.ch)

[stefan.koepsell@tu-dresden.de](mailto:stefan.koepsell@tu-dresden.de)

**Abstract:** Cloud computing has become one of the fastest growing areas in information technology (IT). This technology offers several benefits such as cost reduction, improving information and communication technologies (ICTs) to organizations as well as increased flexibility. Due to these features, several developed and developing countries, specifically African countries have adopted and are still adopting this technology. However, one of the biggest challenges in cloud computing is related to data security and privacy. This paper provides a concise analysis related to data security and privacy in the cloud computing environment. It also addresses recommendations provided by international organizations such as the National Institute of Standards and Telecommunications (NIST), International Telecommunication Union (ITU), and European Union Agency for Network and Information Security (ENISA) on security in cloud computing environment. In Mozambique, the government has been implementing cloud computing. However, currently, there is a lack of a framework for data security and data protection. Therefore, this paper proposes a framework for data security and data protection for Mozambican government cloud computing and other cloud computing providers.

**Keywords:** cloud computing, data security, privacy, privacy by design, Mozambique

---

## 1. Introduction

Over the past 20 years, the Government of Mozambique has been adopting several initiatives to integrate ICTs into public services. One of these was the development of the national ICT policy lead by the ICT Policy Commission under the supervision of the Prime Minister. In 2000, the national ICT was approved by the Council of Ministers and its ICT policy implementation strategy was adopted (UTIPI, 2010). One of the goals outlined in the ICT policy implementation strategy was the establishment of electronic government (e-Gov). The main goal of the e-Gov is to support the reform of the public sector, to improve the performance of the public sector and minimize cost-effectiveness through the use of ICTs. As a result of the e-Gov strategy in 2006, by the Council of Ministers, several initiatives have been carried out such as the implementation of the government network (GovNet), government portal, capacity building, state financial administration system, national system of civil registration system, biometric driving license, motor registration systems, criminal registration system, biometric identity card (ID) and passports (IST-Africa, 2016).

In 2016, the National ICT Institute (INTIC), led the implementation of the Mozambican government cloud computing. Currently, the infrastructure is located at Maluana national ICT park in Maputo and is based on cloud service infrastructure provided by Microsoft i.e., Windows Azure solution. According to ENISA (ENISA, 2015), cloud computing “drives the vast spectrum of both current and emerging application, products, and services, and is also a key technology enabler for the Future Internet”. This new paradigm, unless aligned with well-defined strategic initiatives and compliance with: i) data protection; ii) regulations and legislations; and iii) management and control, the cloud computing will not make valuable contributions. In Mozambique, despite cloud computing being implemented through the e-Government National Strategy at Maluana Data Center, there is lack of national regulations and legislations specially related to data security and data privacy in this environment. Indeed in 2017, a study conducted by Vumo et al (Vumo A, 2017), shows that several government websites as well as public and private websites present several vulnerabilities related to data security. In 2018, the Mozambican government established a new government entity to be in charge of the e-Gov, the National Institute for e-Gov (INAGE). However, since the establishment of this entity little has been done related to cloud computing specifically, data security and data protection. Other countries such as Egypt, United States of America (USA) and Canada (Government of Canada, 2018) have developed national frameworks covering the security of data in the cloud computing environment. For instance, in 2015, the European Union (EU) through ENISA

published the security framework for governmental clouds (ENISA, 2015) and in 2018, the USA published its cloud security guidance (USADHS, 2018). In Mozambique, data security and privacy in cloud computing environment is still only hinted at in some of the national legislations such as the electronic transaction act and the labour law. Therefore, this study analyzes approaches related to data security and privacy in cloud computing environments. Furthermore, it also addresses recommendations provided by international organizations such as NIST, ENISA and ITU. Moreover, this study proposes a conceptual data security framework for cloud computing in Mozambique. The framework addresses several aspects related to the cloud environment including security and privacy requirements and the auditing process. The rest of this paper is structured as follows: in section 2, the research approach is addressed. Section 3 provides a brief overview of related works. Section 4 focuses on data security and issues in cloud computing. Section 5 discusses the cloud computing environment in Mozambique. Section 6 proposes the framework and in Section 7 conclusions and future work is presented.

## **2. Research methodology**

### **2.1 Research approach**

The main goal of this paper is to propose a data security framework in cloud computing for Mozambique in the form of an artifact. Therefore the design science research (DSR) paradigm was conducted. DSR uses artifact design and construction to develop new knowledge (Ken Peffers, 2017). DSR comprises six activities; problem identification and motivation, objectives of a solution, design and development, demonstration, evaluation and communication. The first phase focuses on the definition of the research problem and justification of the solution. The second phase addresses the definition of the goals derived from the phase one. Phase three focuses on the development of an artefactual solution. Phase four demonstrates the efficacy of the artifact. Phase five observes how well the artifact solves the problem and finally phase six publishes it to relevant audiences (Ken Peffers, 2017). Furthermore, a desk research approach was conducted to explore the available literature in cloud computing. Moreover, the privacy by design (PbD) approach proposed by Cavoukian (Cavoukian, 2012) (Cavoukian A, 2012) and recommendations from internationally accepted cloud computing frameworks were used. According to Cavoukian (Cavoukian, 2012)(Cavoukian A, 2012), PbD comprises seven foundational principles described as best practices: i) proactive not reactive; preventative not remedial; ii) privacy as the default ; iii) privacy embedded into design; iv) full functionality - positive-sum, not zero-sum; v) end-to-end security - full life cycle protection; vi) visibility and transparency; and vii) respect for the user privacy.

## **3. Related work**

Many researchers such as (Bhandari A, 2016), (Skolmen D, 2015) and (Youssef A, 2012) have conducted studies regarding data security issues and challenges in cloud computing environment. All three authors have proposed frameworks to protect data in cloud computing through privacy and security requirements. However, a study conducted by IDRBT (2013) presents additional requirements to ensure the security of data in cloud computing environment such as laws, regulations and guidelines issued by the regulation authorities (IDRBT, 2013).

## **4. Background**

### **4.1 Cloud computing definition, deployment models and services models**

One of the most widely used definitions related to cloud computing was provided by NIST. NIST defines cloud computing as “a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources such as networks, servers, storage, applications and services that can be rapidly provisioned and released with minimal management effort or service provider interaction (NIST, 2017).” On the other hand, the ITU Telecommunication sector (ITU-T) defines cloud computing as “a paradigm for enabling network access to a scalable and elastic pool of shareable physical or virtual resources with on demand service provisioning and administration (ITU, 2017)”. According to NIST (2017), a cloud model comprises five essential characteristics: on-demand self-service, broad network access, resource pooling, rapid elasticity and measured service). Furthermore, cloud deployment models is defined as a way in which cloud computing can be organized based on the control and sharing of physical or virtual resources namely: i) private cloud; ii) public cloud; iii) community cloud; and iv) hybrid cloud (ITU, 2017). Moreover, cloud computing service models can be classified according to services delivery (one or more capabilities offered via cloud computing). Thus, there are three classic clouds or well-known and commonly used service models in the cloud paradigm namely: Software as a Service - SaaS, Platform as a Service - PaaS and Infrastructure as a Service - IaaS (ITU-T, 2012) (NIST, 2017).

## **4.2 Cloud computing as regulated activity**

Cloud computing refers to the ability to access and manipulate information stored on remote servers, using any device with network or Internet access (NIST, 2017). According to Cloud Security Alliance (CSA, 2011), multi-tenancy is considered an important part of cloud computing. Despite multi-tenancy offering many benefits, it also has many challenges because more than one tenant can be in one physical machine. Because of this feature, they could attack each other (Aldossary S, 2016). However, this facility has been influencing and continues to influence governments, companies and individuals to use or to increasingly adopt cloud computing (ITU-T, 2012). Despite the benefit of cloud computing which offers the advantage of reducing cost through the sharing of computing resources such as server, network and storage, there are changes related to the way information is managed, especially where personal data processing is concerned. Thus, without knowledge of the physical location of the servers or how the processing of personal data is configured, the end-user consumes cloud computing services without any information about these processes. Hence, data in the cloud computing are easier to manipulate and there is lose of control. For instance, storing personal data on a server somewhere in the cyberspace could pose a major threat to individual privacy (ITU-T, 2012). Therefore, because of these issues it is crucial to regulate cloud computing activities. Moreover, it is important to ensure that when the processing of personal data is managed by a third party specific regulations must be used to ensure compliance with the law.

## **4.3 Data security issues and challenges in cloud computing**

Shaikh (2012), states that in the cloud computing environment, a cloud provider creates, deploys and manages the resources, application and services. Multi-tenancy and virtualization are the key features to make efficient utilization of the existing resources and application. Through virtualization, a single server, computing facility, data center and operating system can host many users. Therefore, the majority of cloud service providers use this facility to provide their services to several users. However, this concept i.e., resource sharing poses a major threat to individual data security and privacy. The CSA defines data security in cloud computing as the process of controls and technologies used to enforce information governance. Hence, data security consists on: i) detecting and /or preventing data migration to the cloud computing; ii) protecting data moving to (and within) the cloud computing; and iii) protecting data in the cloud (CSA, 2011). Over the past years, many researchers such as Albugmi et al. (2016), Shaikh and Sasikumar (2012), Behl and Behl (2012), Shariati et al. (2015), Bokhari et al (2016) and Aldossary and Allen (2016) have been conducting research related to data security and privacy protection issues in cloud computing environment.

# **5. Data security and cloud computing in Mozambique**

## **5.1 Cloud computing in Mozambique**

Currently, the Government cloud solution is based on Microsoft Windows Azure solution and vCloud solution provided by Vodacom as mentioned in Section 1. Nevertheless, some Internet service providers and organizations in public and private sectors are also customers of global IT companies such as Amazon and Microsoft. As mentioned in Section 1, INAGE was established to oversee the implementation of e-Gov. The datacenter located at Maluana in Maputo province is considered the country's primary site where all government cloud infrastructure is located nationally. Aside from this, Mozambique is still not adequately progressing in relation to cloud computing. In developed countries, for instance, EU members countries such as UK, Germany, France cloud computing is anchored in their national strategies (ENISA, 2015). In addition, they have also adopted best practice guides for public bodies (ENISA, 2015). Mozambique still does not have a national strategy covering cloud computing neither data security and privacy or adopted internationally accepted standards and compared to other countries in the region such as South Africa, Mauritius and Tanzania, Mozambique is still lagging behind.

## **5.2 Data security and regulation in Mozambique**

The constitution of the Republic outlines some aspects related to the protection of personal data. Pursuant to Article 41 of the constitution of the Republic of Mozambique, all individuals are entitled to intimacy of their private life (Neves L, 2016). Further, Article 71 of the Constitution grants all individuals the right to privacy, prohibiting the use of electronic means for recording and processing individually identifiable data in respect of political, philosophical or ideological beliefs, of religious faith, party or trade union affiliation or private lives (Neves L, 2016). Despite data protection being referred to by the Constitution of the Republic, regulated by

Labour Law and the recently approved electronic transaction which also provides some elements related to the processing of personal data (sectorial Laws), these laws provide only some degree of legal data protection. Therefore, to date Mozambique still does not have a full regime for data protection issues (Miranda ASARL, 2017). Some issues are still not clear, e.g., the limit time of data storage and what will happen if any user decides to exit a provider (intermediary service provider) and move to another. Moreover, it is not clear if once the user has completed the exit process the right to be forgotten is achieved, i.e. none of the client's data remains with the provider. The electronic transactions law does not specify the definition of sensitive information and data categorization (Henriques F, 2017) (Neves L, 2016). In this context, there are no special rules for determining: i) types of personal data and data classification; ii) the storage of cookies or equivalent; iii) requirement to store personal data inside the jurisdiction and iv) the international transfer of personal data which is very important for cloud computing environment (Henriques F, 2017)(Neves L, 2016).

## **6. The proposed data security framework for cloud computing environment in Mozambique**

### **6.1 Why does Mozambique need data security framework for cloud computing?**

Since 1998, the Mozambican government has been conducting initiatives to integrate ICTs into public services and during the last five years the government has started to pay attention to the integration of cloud computing services into public services. Internationally, several organizations such as NIST, ITU, and ENISA advise governments to develop national strategies related to cloud computing and these strategies should include aspects related to data security and privacy. In Mozambique, there is lack of a national strategy covering the adoption of cloud computing and data security. In the EU, ENISA has developed a Security Framework for Governmental Clouds. The document serves as guidance to EU members and helps states move "towards a seamless and more secure adoption of cloud computing" (ENISA, 2015). Mozambique has also not followed recommendations provided by organizations such as ITU and the African Union (AU) although Mozambique is a member country. Thus, this framework is significantly important as a strategic initiative and will play a key role as a guideline to implementing data security in the cloud computing environment. Figure 1 shows the proposed framework. It has five components: i) organization privacy and security responsibilities; ii) cloud computing environment; iii) security requirements; iv) security mechanisms; and iv) cloud audit. A detailed description of each of them is presented in the following subsections.

### **6.2 Organization privacy and security responsibility**

An organization privacy and security responsibility component provides the roles of officers in defining, establishing and integrating security mechanisms based on security policies, standards and laws as described below:

- The Chief information security officer, Chief privacy officer and compliance officer should perform security and privacy responsibilities using the Information Security Standards such as: ISO 27000 Series, COBIT and NIST800 Series (M, 2013) in conformity with national laws, National ICT policy, the national security strategy draft and recommendation by (Bowen P, 2002), (Cavoukian, 2012), (CSA, 2011), (Cavoukian A, 2012) as well as the Information Security Governance activities as proposed by (Osca Rebollo, 2015) . Therefore, the main activities are: i) Chief information security officer manages, establishes, defines and affirms the security and privacy posture in the organization; and ii) Chief privacy officer is the data protection specialist who defines data and privacy policies.
- To guarantee a secure cloud computing environment the compliance officer should perform security mechanisms based on security requirements proposed in this framework and standards developed by internationally recognized organizations (M, 2013).

### **6.3 Security requirements**

According to (Firesmith, 2004), security requirement is a quality requirement specifying a required amount of security in terms of a system-specific criterion and a minimum level of an associated quality measure that is necessary to meet one or more security policies. In this regard, the proposed framework would be based on following security requirements (confidentiality, integrity, availability, authentication, access control and non-repudiation) which must be used to deploy security mechanisms (Firesmith, 2004) (Yu X, 2010).

## 6.4 Security mechanisms

Security mechanism is an architecture mechanism that helps fulfill one or more security requirements and/or reduces security vulnerabilities (Firesmith, 2004). The following subsections present the security mechanisms to be observed.

### 6.4.1 Network security

A cloud computing can be public or private depending on the accessibility of services. Service and applications are accessed from remote locations in a cloud environment. Continuous availability of cloud service without any disruption, denial of service, and other attacks are network security issues. Therefore, the following are the key aspects that must be considered in the network security level:

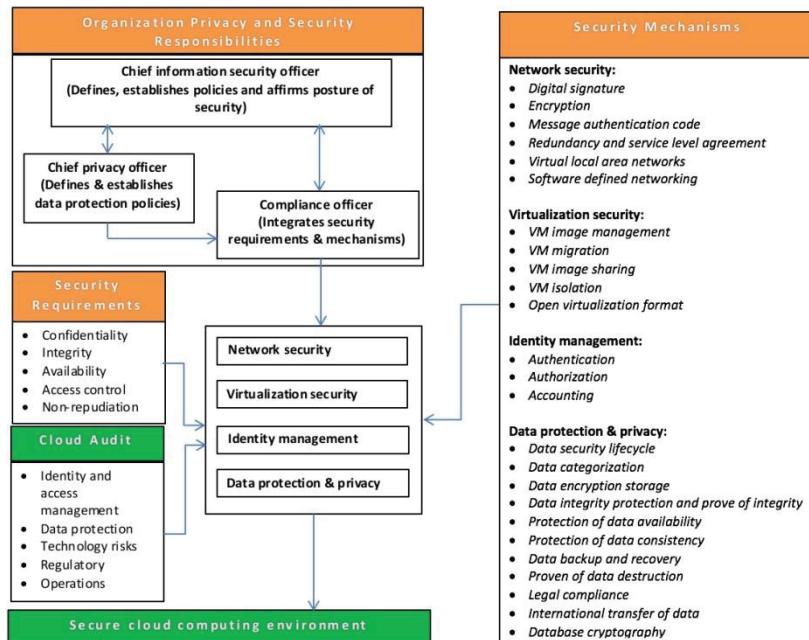


Figure 1: Proposed data security framework

- Encryption:** Confidentiality is the protection of transmitted data from passive attacks. With respect to the content of data transmission, several levels of protection can be identified. The broadest service protects all user data transmitted between two users over a period of time.
- Digital signature and Message authentication code:** Digital signature and Message Authentication Code (MAC) also known as a key hash function are used to provide message integrity and message authentication respectively. The use of recommended secure hash algorithms and public key length is very important (Ristic I, 2016).
- Redundancy and service level agreement:** X.800 and RFC 2828, define availability as the property of a system or a system resource being accessible and usable upon demand by an authorized system entity, according to performance specifications for the system. A variety of attacks can result in the loss of or reduction in availability. Ensuring availability through redundancy and Service Level Agreement contract between consumer and provider specifying a list of requirements for the entire duration of the service is very important (Singh A, 2017).
- Virtual Local Area Networks (VLANs):** VLANs leverage existing network technology implemented in most network hardware. VLANs are extremely common in enterprise networks, even without cloud computing. They are designed for use in single-tenant networks (enterprise data centers) to separate different business units and functions. VLANs are not designed for cloud-scale virtualization or security and should not be considered, on their own, an effective security control for isolating networks (Rich Mogull, 2017).
- Software Defined Networking (SDN):** A more complete abstraction layer on top of networking hardware, SDNs decouple the network control plane from the data plane. There are multiple implementations, including standards-based and proprietary options. Depending on the implementation, SDN can offer much

higher flexibility and isolation. Implemented properly, and unlike standard VLANs, SDNs provide effective security isolation boundaries (Rich Mogull, 2017).

#### **6.4.2 Virtualization security**

The virtualization technology gives a number of hardware partitions. There are three types of virtualization approaches in cloud computing systems virtualization namely (Ashalatha R, 2017): i) server virtualization; ii) network virtualization; and iii) storage. Thus, the following security mechanisms must be observed in virtualization environment:

- **VM image management** (Luo S, 2011) (Schoo P, 2010): VM image is a special type of file/data format which is used to instantiate (create) a virtual machine within the virtual environment. Therefore, the confidentiality and integrity of VM image is of great importance when VM is under bootstrapping or migrating.
- **VM migration** (Luo S, 2011) (Aldossary S, 2016) (Schoo P, 2010): VM migration is a vulnerable process that is easily attacked. Thus, when a VM machine is going to migrate somewhere such as in some network's hosts or other network's hosts, the security mechanisms should be taken in account.
- **VM image sharing** (Aldossary S, 2016) (Schoo P, 2010): VM can be instantiated from a VM image. A shared image repository can be used to share VM images or a user can have his own VM image. Since there is a repository for sharing VM images, some malicious users could take advantage of this feature in order to inject a code inside a VM.
- **VM isolation** (Aldossary S, 2016) (Schoo P, 2010): When VM Machines run in some hardware (host), they share all components (e.g.: processor, memory and storage). Isolating VM logically to prevent one from intervening with another is not enough since they are sharing some hardware in some host. Therefore, data may leak. In other perspectives of security this is serious issue. Hence, isolation should be at the level of VM machine and hardware.
- **Open virtualization format** (DMTF, 2018): Open Virtualization Format (OVF) is a hypervisor neutral, efficient, extensible and open specification for packaging and distribution of virtual machines in the form of virtual appliances. This requirement is necessary in case of interoperability between cloud computing providers.

#### **6.4.3 Identity anagement**

Shaikh (Shaikh R, 2012) considers that identities are generated to access a cloud service. Each user uses his identity for accessing a cloud service. Unauthorized access to cloud resources and applications is a major issue. A malicious entity can impersonate a legitimate user and access a cloud service. Thus, Identity Management System for providing authentication, authorization and accounting is an issue for both provider as well as user in a cloud computing environment:

- **Access control:** When data is outsourced to the cloud, which is untrusted because it is in a domain where security is not managed by the data owner, data security has to be given more attention. When more than one entity wants to share data, a mechanism to restrict the access to data should be put in place to restrict who can access that data. The techniques which keep data content confidential and keep unauthorized entity from accessing and disclosing the data by using access control while permitting only authorized entities to share those data is very important (Aldossary S, 2016).
- **Authentication** (Yu X, 2010) (Singh A, 2017): The purpose of authentication is to prevent unauthorized access to user's data. Both the users and cloud service providers must be authenticated before using the data; this requirement will surely reduce the risk of information leakage. There are various technologies to authentication, such as passwords, certificates and biometrics.
- **Authorization:** Authorization provides some permission to a subject to perform certain operations, if the subject is authenticated. In order to complete the authentication and authorization, first authentication is done then authorization is provided. The insufficient authorization, session/credential prediction, session expiration, and session fixation can lead to access a protected area beyond their privileges (Singh A, 2017).
- **Accounting:** The security goal that generates the requirement for actions of an entity to be traced uniquely to that entity. This supports non-repudiation, deterrence, fault isolation, intrusion detection, prevention, after-action recovery and legal action. This process must take into account responsibility for protecting and managing the appropriate use of that information through legal requirements.

#### **6.4.4 Data protection and privacy**

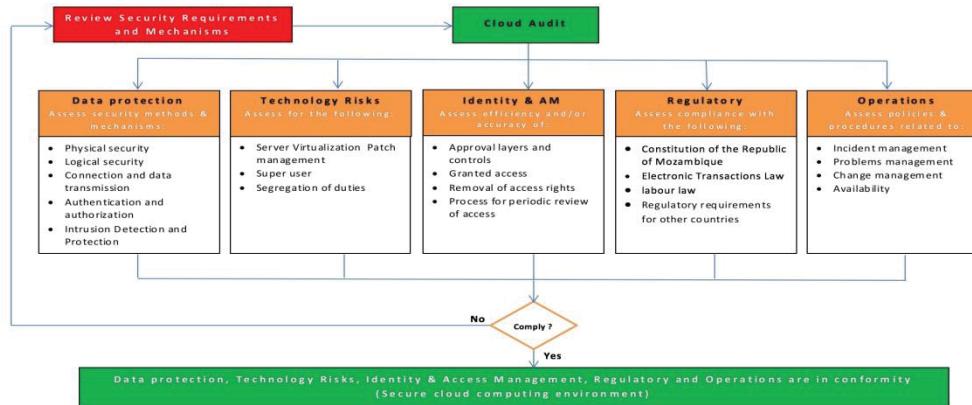
In this proposed framework, the Chief information security officer and Chief privacy officer have responsibilities to define and establish data security, privacy protection policies and security policies. Thus, the requirements for data security that must be considered in cloud computing are following:

- *Data security lifecycle*: chief privacy officer and chief security officer must use the data security lifecycle to identify security exposures and determine the most appropriate control (CSA, 2011).
- *Data categorization*: This mechanism is carried out for a variety of purposes. The key goal of data categorization is to define high level categories to determine which security controls may apply and in which cloud computing deployment model can be deployed. Al-Sayid et al (2013) present some examples of data categorization based in organization sensitive data (Digital Guardian, 2017) (Government of Canada, 2018).
- *Data encryption storage*: As data is in third-party cloud computing, the biggest threat is the information leak. Thus, when data is encrypted and stored in this environment, the risk of information leakage reduced. In this case, asymmetric and symmetric algorithms encryption technologies must be used to encrypt data (Yu X, 2010).
- *Data integrity protection and prove of integrity*: This mechanism is used to ensure that the data is correct before storing. On the other hand, the data proof of integrity is a service provided by the cloud service provider, which allows data owners to check whether the data stored in the cloud is complete or not (Yu X, 2010).
- *Protection of data availability*: In the cloud environment, data may be stored in different devices. If some devices or nodes fail, it could lead to data unavailability. Therefore, the cloud service providers must ensure data availability (Yu X, 2010).
- *Protection of data consistency*: Users always want to use data that is in the newest status. The consistency of data is hard to keep up in share mode (Yu X, 2010).
- *Data backup and recovery*: The easiest way to reduce data security risk is to conduct data backup and recovery. The data backup work can either operate in local devices or nodes or in remote. The local backup mode can quickly be executed but has problems of single point of failure; the remote backup mode also has such problems but the corresponding backup and recovery work is more complex (Yu X, 2010).
- *Proof of data destruction*: when the user wants to delete the data stored in cloud this means data must be destroyed completely. Therefore, it is necessary to provide data destruction proof to make sure that the owner's data has indeed been destroyed (Yu X, 2010).
- *Legal compliance*: According to the Cloud Security Alliance, cloud computing with their deployment model creates new dynamics in the relationship between an organization and its information, involving the presence of third-parties. This creates practical challenges in understanding how laws apply to the different parties under various scenarios in the cloud computing environment. It is important to consider legal issues, especially those around data might collection, storage and processing. There will likely be state, national or international laws to consider to ensure legal compliance (Bruunette G, 2018).
- *International transfer of data*: To carry out the cross border of data related to an individual, regulations must identify what data could be allowed and restricted or prevent personal data from leaving the country from a privacy perspective.
- *Database cryptography*: To secure the host level, database encryption is a key strategy to protect the contents of data within the database. Singh (2015) argues that the main idea behind this is that in the case of an intruder somehow getting to the system's database; due to encryption he should not be able to misuse the data in the database (Singh P, 2015).

#### **6.5 Auditing cloud computing**

Mohammad et al (Mohammad M, 2018), define the cloud computing auditing process as a consolidation of different International Standards on Auditing and audit methodology which can be performed by internal or external auditors. Therefore, the main objectives of cloud computing audit are: i) to mitigate risks introduced by the cloud computing environment; to evaluate the efficiency of controls related to the cloud; and iii) to continuously improve internal processes, procedures and tools. Thus, proposed cloud computing audit deemed within five relevant areas are proposed by KPMG (Stephens, 2013) (Yati Nurhajati, 2016): i) identity and access

management (AM); ii) data protection; iii) technology risks; iv) regulatory; and v) operations. Figure 2 shows cloud computing auditing process. For more detail about dimensions risks associated with these areas were presented by (Stephens, 2013) (Yati Nurhajati, 2016).



**Figure 2:** Cloud computing auditing process

## 7. Conclusion and future work

Despite the fact that cloud computing provides several services to individuals and organizations alike, there are also many problems that need to be addressed and one of them is related to how information is managed. As discussed in Section 4 data security and privacy protection is the main concern in the cloud computing environment. This is due to the characteristics of the cloud infrastructure because cloud computing services are often owned by a third party. Moreover, without knowledge of the physical location of the servers by users, user's data in the cloud computing environment are easier to manipulate leading to a loss of control. In this paper, we have analyzed the state-of-the-art of cloud computing. We have analyzed many issues related to the security of data and reviewed some frameworks for secure cloud's proposed by authors such as (Bhandari A, 2016), (Skolmen D, 2015), (Youssef A, 2012) and (IDRBT, 2013). Based on these analyses we have proposed a conceptual framework to help in the implementation of a trusted cloud computing environment in Mozambique that satisfies data security and privacy requirements. In particular, cloud computing auditing process has been presented to verify deployment of data protection and privacy in order to establish a trusted cloud computing environment. To the best of our knowledge there are no articles or research works on data security and privacy protection in the cloud computing environment in Mozambique. As a further work, we plan to implement this framework. However, we believe that more efforts should be done by the Mozambican government to provide trusted, safe and protected cloud computing environment to its citizens and the society in general.

## References

- Albugmi A, A. M. (2016). Data security in Cloud Computing. Fifth International Conference on Future Generation Communication Technologies. IEEE.
- Aldossary S, A. W. (2016). Data Security, Privacy, Availability and Integrity in Cloud Computing: Issues and Current Solutions. International Journal of Advanced Computer Science and Applications. IEEE.
- Al-Sayid N, A. D. (2013). Database Security Threats: A Survey Study. International Conference on Computer Science and Information Technology. IEEE.
- Ashalatha R, A. J. (2017). Network Virtualization System for Security in Cloud Computing. International Conference on Intelligent Systems and Control. IEEE.
- Behl A, B. K. (2012). An Analysis of Cloud computing Security Issues, Information and Communication Technologies. Information and Communication Technologies (WICT). IEEE.
- Bhandari A, G. A. (2016). A Framework for data security and storage in Cloud Computing. International Conference on Computational Techniques in Information and Communication Technologies. IEEE.
- Bokhari M, S. O. (2016). Security and Privacy Issues in Cloud Computing. International Conference on Computing for Sustainable Global Development (INDIACOM). IEEE.
- Bowen P, H. J. (2002). Information Security Handbook: A Guide for Managers. Gaithersburg, MD 20899-8930: NIST Special Publication 800-100.
- Bruunette G, M. R. (2018, May 23). Retrieved May 23, 2018, from <https://downloads.cloudsecurityalliance.org/assets/research/security-guidance/csaguide.v3.0.pdf>
- Cavoukian. (2012). Operationalizing Privacy by Design: A Guide to implement Strong Privacy Practices. IEEE Technology and Society Magazine. IEEE.
- Cavoukian A. (2012). Privacy: Front and Center. IEEE Security and Privacy. IEEE.

- Chirammal H, M. P. (2016). Mastering KVM Virtualization: Dive in the cutting edge technique of Linux KVM virtualization and build the virtualization solutions your datacenter demands. Mumbai: Packt Publishing.
- CSA. (2011). Security Guidance for Critical Areas of Focus in Cloud Computing V3.0. Cloud Security Alliance. Cloud Security Alliance.
- Digital Guardian. (2017, November 26). Retrieved November 26, 2017, from <https://digitalguardian.com/resources/data-security-knowledge-base/data-protection-101>
- DMTF. (2018, March 13). Retrieved March 13, 2018, from <https://www.dmtf.org/standards/ovf>
- Government of Canada. (2018, September 10). Retrieved September 10, 2018, from <https://www.canada.ca/en/treasury-board-secretariat/services/information-technology/cloud-computing/government-canada-cloud-adoption-strategy.html>
- Governo de Mozambique. (2017, December 26). Retrieved December 26, 2017, from <http://www.portaldogoverno.gov.mz/por/Media/Files/Lei-de-Transacoes-Electronicas>
- Henriques F, L. M. (2017, December 26). Retrieved December 26, 2017, from [https://uk.practicallaw.thomsonreuters.com/w-009-5516?transitionType=Default&contextData=\(sc.Default\)&firstPage=true&bhcp=1%2c](https://uk.practicallaw.thomsonreuters.com/w-009-5516?transitionType=Default&contextData=(sc.Default)&firstPage=true&bhcp=1%2c)
- IDRBT. (2013). Cloud Security Framework for Indian Banking Sector. Reserve Bank of India. An IDRBT Publication.
- INTIC. (2018, March 18). Retrieved March 18, 2018, from <http://www.intic.gov.mz/por/Informacao/Noticias/INTIC-deixa-funcao-implementadora-para-INAGE>
- IST-Africa. (2016). Report on ICT Initiatives and Research Capacity in IST-African Partner Countries. IST Africa. IST-Africa Consortium.
- ITU. (2017, December 23). Retrieved December 23, 2017, from <https://www.itu.int/rec/T-REC-Y.3500-201408-I>
- ITU. (2017, December 15). Retrieved December 15, 2017, from [https://www.itu.int/en/ITU-D/Cybersecurity/Pages/Country\\_Profiles.aspx](https://www.itu.int/en/ITU-D/Cybersecurity/Pages/Country_Profiles.aspx)
- ITU-T. (2012). Privacy in Cloud Computing, ITU-T Technology Watch Report. ITU. ITU.
- Kelley B, e. a. (2016). Securing Cloud Containers Using Quantum Networking Channels. IEEE International Conference on Smart Cloud. IEEE.
- Lebanidze E. (2017, December 18). Retrieved December 18, 2017, from [https://www.owasp.org/images/8/83/Securing\\_Enterprise\\_Web\\_Applications\\_at\\_the\\_Source.pdf](https://www.owasp.org/images/8/83/Securing_Enterprise_Web_Applications_at_the_Source.pdf)
- Luo S, e. a. (2011). Virtualization Security for Cloud Computing Service. International Conference on Cloud and Service Computing. IEEE.
- M, R.-O. (2013). The Complete Reference: Information Security (2nd ed.). (M.-H. Companies, Ed.) New York: McGraw-Hill Education.
- Miranda ASARL. (2017). Privacy/Data Protection Law: How Much Disclosure does Growth Need? Annual Conference of the African Bar Association. Miranda Alliance.
- MSG. (2017, December 15). Retrieved December 15, 2017, from <http://www.managementstudyguide.com/desk-research.htm>
- Neves L, T. J. (2016). Data Protection in Mozambique: Inception Phase. Springer International Publishing AG. Springer International Publishing AG.
- NIST. (2017, December 12). Retrieved December 12, 2017, from <https://www.nist.gov/publications/nist-definition-cloud-computing>
- NIST800-125. (2011). Guide to Security for Full Virtualization Technologies. NIST Special Publication 800-125.
- Oscar, R., Daniel, M., & Medina, E. (2015). Introducing a Security Governance Framework for Cloud Computing. The Computer Journal 58 (10), 10.
- Ristic I. (2016). OpenSSL cookbook: A guide to the most frequently used OpenSSL features and commands (2 ed.). London: Feisty Duck Limited.
- Schoo P, e. a. (2010). Challenges for Cloud Networking Security. International Conference on Mobile Networks and Management. Springer.
- Shaikh R, S. M. (2012). Security Issues in Cloud Computing: Survey. International Journal of Computer Application (0975-8887). IJCA Journal.
- Shariati S, e. a. (2015). Challenges and Security Issues in Cloud Computing from two perspectives: Data security and privacy protection. International on Knowledge-based Engineering and Innovation. IEEE.
- Shrivastwa A, S. S. (2015). Learning Openstack: Set up and maintain your own cloud-based Infrastructure as a Service (IaaS) using Openstack. MUMBAI: Packt Publishing.
- Singh A, C. K. (2017). Database Security Using Encryption. Journal of Network and Computer Application. ACM.
- Singh P, K. K. (2015). Database Security Using Encryption. International Conference on Futuristic Trend in Computational Analysis and Knowledge Management. IEEE.
- Singh S, J. Y. (2016). A Survey on Cloud Computing Security: Issues, threats and solutions. Journal of Network and Computer Applications. ELSEVIER.
- Skolmen D, G. M. (2015). Protection of personal Information in the South Africa Cloud Computing Environment: A Framework for Cloud Computing Adoption. Information Security for South Africa (ISSA). IEEE.
- Smith J. (1998). The Book (2 ed.). London: The Publishing company.
- Splaine S. (2002). Testing Web Security-Assessing of Security of Web Sites and Application (1 ed.). Indiana: Wiley Publishing.

***Ambrósio Patrício Vumo, Josef Spillner and Stefan Köpsell***

- Sun Y, e. a. (2014). Data Security and Privacy in Cloud Computing. International Journal of Distributed Sensor Networks. Hindawi Publishing Corporation.
- UTIPI. (2010). eGovernment Interoperability Framework for Mozambique. Ministerio de Ciencia e Tecnologia. Maputo: Ministerio de Ciencia e Tecnologia.
- Vumo A, S. J. (2017). Analysis of Mozambican Websites: How do they Protect their Users? Information Security for South African (ISSA). IEEE.
- Wu H, e. a. (2010). Network Security for Virtual Machine in Cloud Computing. International Conference on Computer Sciences and Convergence Information Technology. IEEE.
- Youssef A, A. M. (2012). A Framework for Secure Cloud Computing. International Journal of Computer Science Issues. IEEE.
- Yu X, W. O. (2010). A View About Cloud Data Security From Data Cycle. Computational Intelligence and Software Engineering. IEEE.



# **Masters Research Papers**



# A Technical Overview on the Usage of Cloud Encryption Services

**Daniel Carvalho, João Morais, João Almeida, Pedro Martins, Carlos Quental and Filipe Caldeira**

**Informatics Department Polytechnic Institute of Viseu, Portugal**

[estgv14798@alunos.estgv.ipv.pt](mailto:estgv14798@alunos.estgv.ipv.pt)

[estgv16054@alunos.estgv.ipv.pt](mailto:estgv16054@alunos.estgv.ipv.pt)

[estgv16620@alunos.estgv.ipv.pt](mailto:estgv16620@alunos.estgv.ipv.pt)

[pedromom@estgv.ipv.pt](mailto:pedromom@estgv.ipv.pt)

[quental@estgv.ipv.pt](mailto:quental@estgv.ipv.pt)

[caldeira@estgv.ipv.pt](mailto:caldeira@estgv.ipv.pt)

**Abstract:** As the number of successful attacks to ICT infrastructures raises, encryption plays a relevant role in maintaining data security and privacy, thus protecting confidential information. The existence of a large amount of unencrypted data, for instance, multiple files stored in smart cities control centers, support the urgency to minimize the dangers of this situation effectively. With the advance and integration of systems within smart cities, several critical files are available to multiple services and apps. For example, pictures, citizen's documents, city services information. Several of these files contain personal and/or critical information and are stored unencrypted, mainly depending on access control systems to maintain their security. With the growing number of services available in smart cities aiming to increase the information available to citizens, this problem should be quickly addressed. In this context, this paper proposes the adoption of widely available cloud services to maintain encrypted data. It is described how to set up encryption in the download and upload of files implementing Azure, Google Cloud, or Amazon Web Services platforms, plausibly and straightforwardly. The discussion includes the processes to generate, maintain and use encryption keys within groups of user/services and also how the storage service automatically encrypts and decrypt the data. The paper presents a comparative discussion about the selected platforms, including also actual price comparisons. From the experimental work, the results show that, among the platforms tested on our experimental scenario, Amazon was the most natural platform to integrate the encryption with other services.

**Keywords:** cloud, encryption, smart cities protection

---

## 1. Introduction

The cloud is now a reliable and straightforward option to guarantee a vast variety of services, including the storage of data files, keys, secrets and many more that are applied by all types of companies. Platforms such as Amazon Web Services (AWS) Key Management Service (KMS), Azure Key Vault, and Google Cloud Platform (GCP) KMS, allow to implement services (Campagna, 2015), (Diogenes, 2016), (Steven Porter, 2018) aiming to, for instance, when a user works with files in the cloud, ensure that there is a certain level of security in its use (Ramakrishnan, 2017).

These services were evaluated under the following process: first, keys are created using the Key Management Services from all the mentioned platforms and defined the identity and access permissions to the services. The next step is to manage the Storage service available in each platform and enable the encryption in the storage of, for instance, pictures, citizen's documents and city services information using the previously created key. Following, the storage encryption needs to be set up. This process differs from platform to platform. In AWS S3 (J Varia, 2014), (Amazon, 2012) various options exist to set up encryption, one of them is the creation of a data bucket without encryption and the other a bucket with encryption using the customer managed keys created in AWS KMS (Anthes, 2010). In Azure Blob Storage and GCP Cloud Storage, there is a default level encryption that is defined by the respective platforms, yet there is also the possibility to use the keys created in service Azure Key Vault and GCP KMS (Qian, 2009). This way, by integrating two services within the same platform, it is possible to evaluate the availability of the encrypted files existing in the storage services.

As expected, users without permission to access the necessary services to decrypt and access the files, get a 403-forbidden page state while users who otherwise have permission can access and read/write the data.

In the next sections, the main steps taken to achieve the result are described, and other observations we perceived analyzing the services which are documented as the following: Section II describes the services that, in Azure, are all agglomerated in one major service named Key Vault. This Vault contains the management of keys, secrets, and certificates. These are found in different components in the AWS and Google Cloud platforms.

Section III explains the steps taken to implement and integrate the services with the common objective (upload and download of encrypted files). In section IV it is discussed how the services compare to each other, the advantages and disadvantages of each platform including a price comparison to set up a primary encrypted data storage.

## **2. Related work**

While on Azure Key Vault (KV) the services are all centralized in one facility, on Amazon Web Services (AWS) and Google Cloud Platform (GCP) the services appear distributed in single components. For example, Secrets in KV can be found in Amazon Secrets Manager where both allow the management and retrieval of secrets used by the users and applications. In the Google Cloud Platform, there is only the possibility of storing the secrets in Cloud Storage (GPC service) (Steven Porter, 2018).

KV Certificates can also be found in yet another service known as AWS Certificate Manager where users can provision certificates for their domain and, if set, AWS CM and KV Certificates can manage the renewal of the certificates automatically generated previously. It is also available the option to establish a secure managed infrastructure within an organization allowing the issuing and revocation of the certificates defined in the private certificate authority (Authority, 2018), (O'Boyle, 2017).

One important component that integrates with these services and allows to oversee the deployment of infrastructure in the cloud, is the Identity and Access Management known in all these platforms as IAM (J Varia, 2014), (Riti, 2018). This service allows the secure management and access to other services in the platform through permission boundaries on which are set permission policies.

Another cloud service that implements key management is Alibaba Cloud Key Management Service that also guarantees confidentiality, integrity, and availability of keys, performs envelope encryption, and the readiness to integrate it with various Alibaba Cloud storage services to ensure the security of the stored data (Cusumano, 2010).

As Dan O'Boyle said “Cloud Based Key Management Services (Google KMS, Azure Key Vault, Amazon KMS) provide encryption keys as a service. KMS create a centralized access control list. Using a KMS, it is possible to centralize secrets, removing them from local libraries. Key rotation can be automated, often making a KMS more secure than local key management practices” (O'Boyle, 2017).

As will be noted in this paper AWS KMS is a versatile and user-friendly service. Albert Anthony affirmed “AWS KMS has various benefits such as importing your own keys in KMS and creating keys with aliases and description. You can disable keys temporarily and re-enable them. You can also delete keys that are no longer required or used.” (Anthony, 2018).

Google Cloud Key Management Service is also a sound choice according to Steven Porter, 2018), “(...) can use KMS to encrypt and decrypt data stored virtually anywhere”.

Azure Key Vault also provides excellent security. “Azure Key Vault provides a secure service for Azure applications and services (...)” (Zoiner Tejada, 2018).

## **3. Experimental setup**

All Amazon Key Management Service, Azure Key Vault and Google Cloud Key Management Service were tested following a procedure that allows integrating the service with others, inside the same platform. The storage in the cloud seems more and more of a solution not only in enterprise-level designs but also to regular users for the convenience of data access in multiple devices and data redundancy. Within Amazon Web Services the service chosen to integrate was Amazon S3 (Amazon Simple Storage Service), on Microsoft Azure the chosen service was its closest correspondent, Azure Blob Storage and on Google Cloud Platform the chosen service was Cloud Storage. Having multiple services that implement close to the same functionality in different platforms, the next step is to use the Key services to create keys, discuss the available options offered by the platforms in terms of encryption, the ease in the management of the account keys, what actions can be executed with the keys and the users defined to administrate and use them.

### 3.1 Amazon KMS

Initially, users are defined in AWS Identity and Access Management service where access policies are set allowing the users to use the S3 service, so they can upload/download files to the storage bucket. In the Key Management Service, it is created a key in which are going to be granted privileges to administrators and users from IAM, so they can administer the key (enabling, disabling, rotation and other operations) and use the key (encrypt and decrypt data) respectively (OGÍGAU-NEAMTIU, 2015). In the S3 it is created a bucket that will use the previously generated key to encrypt and decrypt the data that a user stores and requests to access.

#### 3.1.1 Identity and access management – user setup

In the service Identity and Access Management (IAM), two users were created with friendly names to understand better the context in which they are applied. One user is named “Admin” which has privileges over the management of the keys, and the other user is named “User” and has privileges to use the key to encrypt and decrypt data (Lewis, 2013). During the creation of the previous users, the appropriate permissions are set, allowing only, in this context, consents to the KMS and AWS Storage S3 service. Each user, depending on the options set in the creation process, gets a sign-in link that allows the user to login into the AWS platform.

#### 3.1.2 Key management service setup

The Key Management Service from AWS only uses symmetric encryption and uses AES (Advanced Encryption Standard)-GCM (Galois/Counter Mode) algorithm with 256-bit encryption keys (Campagna, 2015). The KMS has three types of keys that can be better understood on Table 1 (it refers to the keys from a user perspective).

**Table 1:** Different types of CMKs

Type of CMK	Can view	Can manage	Used only for my AWS Account
Costumer managed CMK	Yes	Yes	Yes
<i>3.1.3 AWS managed CMK</i>	Yes	No	Yes
AWS owned CMK	No	No	No

AWS Owned CMKs (Costumer Master Keys) are not visible and can't be managed in any way by the user in KMS or any other location. These keys are owned and managed by AWS and are set to be used in various AWS accounts.

AWS Managed CMKs are keys that are created, managed and used automatically in the usage of other services of the platform and in which the user wants to setup encryption, one example being cloud data storage, but do not want to be managing/maintaining another service to keep the data encrypted. In this case, the service sets up a key that is managed by the service taking an alias referring to the service that was integrated with (e.g., aws/s3) and is only available for the user to keep track and view data referring to the key such as:

- ARN - (Amazon Resource Name) that uniquely identify resources set up in the Amazon Web Services platform;
- Alias - Display name that is given to the key;
- Origin (AWS KMS, External ...);
- Status (Enabled, Disabled ...);
- Description – Provides additional information on the key such as using a S3 service “Default master key that protects my S3 objects when no other key is defined”;
- Creation Date (e.g. 2018-11-18 17:00 GMT).

AWS Customer Managed CMKs are created and owned by the user(s) that manage the service. Creating this type of Key allows the user to enable, disable, rotate the key, schedule its deletion and to set up policies for its use.

AWS Customer managed CMK's can encrypt/decrypt up to 4KB of data (often adopted to encrypt other data keys as a method of envelope encryption). To encrypt files being uploaded to the S3 service this is the type of key that is created for testing (it is necessary to have in mind the location in which the key is being created in order to be available in the region where the bucket is located).

During the creation process, the first step is to set up an alias and description for the key, so it can better describe the key future use. During this first step, it is also possible to define if the key being created is to be generated by the KMS service or to import a 256-bit symmetric key from the user infrastructure.

The second step is optional, and it defines tag keys and values to categorize the key. It allows for a better identification and to keep track of the costs that the key is generating.

In the third step, the key administrators are defined and may also be set the option that allows them the deletion of the key. In this step, the administrator of the key was set to be the IAM user Admin.

In the fourth step, the key usage permissions are set, defining which users can use the key to encrypt and decrypt data. It is also possible to add external AWS accounts, so they may have access to the key data in their account. This may require setting up more conditional policies to the external account that can be a key administrator or key user.

The fifth step shows all the information and key policies set in the previous steps in a JSON (JavaScript Object Notation) object and concludes the key creation process.

The CMK created can also be configured to automatically rotate every defined period, improved guarantying security (instead of the extensive reuse of the same encryption key) practice over data encryption. If the automatic CMK rotation is enabled, which amounts one more dollar to the monthly bill, AWS KMS generates new cryptographic material for the Costumer managed CMK one time a year (the period is not alterable). It also guarantees that data encrypted with older versions of the key is still available by saving the older cryptographic material maintaining its key ID, ARN, policies, and permissions. Furthermore, if the key rotation is a responsibility of the CMK administrators, they control any CMK rotation and define the schedule (e.g., 1 year, 1 year and a half etc.). These facts imply that the user must create another CMK and decrypt old data with the old key and re-encrypt it with the new CMK, though it also allows using imported keys from their infrastructure.

#### *3.1.4 S3 - storage service setup*

At first, a bucket is created in AWS S3 (Simple Storage) (J Varia, 2014), (Amazon, 2012) named “availabledata”. It will contain the files that may or may not have critical information stored. When setting up the bucket, one important detail to have in mind is setting the bucket location in the same location that the KMS key is defined or the keys may not appear. In the bucket creating options it may be enabled the default encryption checkbox which allows the use of an AWS owned CMK, AWS Managed CMK and the Costumer managed CMK. Presently let's assume that not all the data will be encrypted (default encryption disabled) so some data of the bucket may be public, and some will be critical information that only users that have permission to use the key will be able to read. Using the sign-in link from the IAM service, the user “User” logs in into his account and accesses the S3 service which has the “availabledata” bucket visible. After accessing this bucket, a file was uploaded without encryption with, supposed, public domain data and an encrypted file, encrypted with the key created in the KMS service (chosen in the encryption section when uploading a file), which contains critical data.

The file uploaded without encryption will be available for public access while the file uploaded with encryption will have a scope of users that can use the key to read the data. At this moment the IAM user “User” can read the encrypted file because he has usage permissions over the key that was used to encrypt the file. Let's assume that a user, for whatever reason (e.g., leaves the company), and should not continue to have permission to access the data. In this scenario, in the KMS service, the usage permission is removed from the policies to the user “User” so he can no longer have access to the data. From this moment on, until the user has been given usage permissions over the key again, he will experience an HTTP 403 Access Denied code error trying to access the file.

#### *3.1.5 Usage log – AWS CloudTrail*

The Key usage can be visualized at the service CloudTrail (J Varia, 2014) from Amazon Web Services Platform which allows for an Administrator to track the operations made by users with permissions over the key. A log in the CloudTrail includes information aiming to identify what happened, when and where, namely: the AWS region, error code (Access denied – Error got in the previous operation), event name (decrypt – Operation that

was trying to be executed), the time that the event was triggered, the type of user, and the user that attempted to complete the operation.

## **4. Google Cloud Platform KMS**

Google Cloud Platform KMS service (GCP KMS) is used to create a keyring where keys can be stored (Steven Porter, 2018), with these keys, it is possible to encrypt files stored in the GCP Cloud Storage and thus have more security.

### **4.1 GCP – cloud management service**

GCP keyring service allows to store various keys. When the key is created, it is necessary to define some parameters such as, type of key (Symmetric or Asymmetric key), levels of protection (Software or Hardware Security Module) and set the key rotation time, which consists of creating a new version of the key. With this/enabling rotation it is offered more security. There are two types of available keys:

- **Symmetric Keys:** Are always created with the same algorithm (AES-256) and the user cannot choose other algorithm to create the key, that is the only algorithm that GCP KMS uses for this type of keys.
- **Asymmetric Keys (Beta):** Unlike the Symmetric keys, in this case when we create an asymmetric key, we can choose what algorithm we want to use. Here are available four algorithms, nevertheless, the algorithm recommended by Google is RSA\_DECRYPT\_OAEP\_3072\_SHA256. Note that, up to the date this paper, the asymmetric key is in beta version.

#### *4.1.1 Levels of protection*

To create a key, it is necessary to choose a level of protection this can be of Software type or Hardware Security Module type. After creating a key, the level of protection cannot be changed. There are two levels of protection:

- **Level of Protection by Software:** The cryptographic operations are run in a Software; this protection has lower cost in versions keys and operations with the key.
- **Level of Protection by HSM (Beta):** The cryptographic operations are run in a Hardware security module; this level of protection has higher cost comparatively with the Software.

It is possible to share keys with other users, this sharing is done through Gmail accounts, at the same time is also possible to define Identity and Access Management (IAM) users.

## **4.2 GCP – cloud storage**

This service works with intervals, i.e., to save the files it is necessary to create an interval and inside this interval are stored the files. During the interval creation process, it is necessary to choose a key type that will be used, that can be of the type “Key Managed by Google” or “Key Managed by Client”, (Key Managed by Client are the default keys created in GCP KMS). When using “Key Managed by Client”, it is important to note that, the interval must be in the same region where the key was created on KMS, e. g., if the interval is saved in Europe, the key also needs to be created and saved in Europe, so that it can be used to encrypt the interval (Qian, 2009).

To share files stored in GCP Cloud Storage, it is possible to give access to other users through the service GCP Identity & Access Management (GCP IAM), and directly with Gmail accounts. In both cases it is possible to define policies for Identity & Access Management (IAM).

## **5. Azure key vault**

This service allows to generate or import Keys, generate or import secrets that can be passwords, API Keys, Keys generated by Key Vault amidst others and generate or import digital certificates (OGIGAU-NEAMTIU, 2015). Regarding the generation of keys, after being generated, they can encrypt stored files in the Blob Storage, which is Microsoft Azure File Storage service. In this file storage it is possible to encrypt files through keys generated by Blob Storage Service.

### **5.1 Key vault – generated keys**

The key vault (Freato, 2015), (Diogenes, 2016) uses the asymmetric key methods. Once a key is generated it is necessary to keep in mind the type of algorithm that was used, in general RSA (Rivest-Shamir-Adleman) and EC

(Elliptic Curve). The size of key depends on selected configurations, as well as, the key activation date, and the key expiration date. After the key is generated, it is possible to view some settings of key like the key type, key size, and can also be defined the key permitted operations: encrypt, decrypt, sign, verify, wrap key and unwrap key. There are two types of available keys:

- **"Soft" keys:** A key processed in software, by the Key Vault, but is encrypted using a system key that is in an HSM. Clients may import an existing RSA or EC (Elliptic Curve) key, or even request for the Key Vault to generate one.
- **"Hard" keys:** A key processed in an HSM (Hardware Security Module). These keys are protected in one of the Key Vault HSM Security Worlds (there is one Security World per geographic location to maintain isolation). Clients may import an RSA or EC key, in soft form or by exporting from a compatible HSM device.

It is possible to share Key Vault with other Azure accounts through IAM (Identity and Access Management), where the policies of access to the service are also defined. The owner of service can assign to other azure client different rules being the most important ones:

- **Owner:** allows to manage everything, including access to resources;
- **Contributor:** Lets you manage everything except access to resources;
- **Reader:** Lets you view everything, but not make any changes.

## 5.2 Blob storage

Once created, blob storage allows stored Blobs, files, create rows and create tables that can be used in another Microsoft Azure service (Azure Cosmos DB).

In Blob Storage service it is necessary to define the encryption method, which by default, uses Managed Microsoft Keys for Azure Blobs, or it can be generated by Azure's Key Vault Service, in this case, using asymmetric keys (Calder, 2011), (Redkar, 2009).

After defining the encryption method, it is possible to share access to Blob being necessary to generate a SAS (Shared Access Subscription) and connection string. This connection string must be shared with the clients with whom the Blob will be shared. The SAS and connection string are generated through a key that is created within the own Blob, being that the Signing Key, SAS and connection String can be saved in Key Vault secrets to remain safe.

Both Azure Key Vault and Blob Storage have the possibility to check Logs generated by the service in the Activity Log section.

## 6. Results comparison

### 6.1 Import and export keys

#### AWS KMS

In AWS Key Management Service, it is not possible to export keys, the keys created in the KMS stay within it, so it guarantees their security, enables the consistent enforcing policies and providing centralized logs of usage over the keys. Although it is not possible to export keys it is possible to import them from the user infrastructure, but it needs to be a 256-bit symmetric key.

#### GCP Cloud KMS

It is not possible to export a key from GCP KMS because all the encryption and decryption must be made inside GCP Cloud KMS. This helps preventing misuse and allow the GCP Cloud KMS register the activities of the keys. It is also not possible import a key for GCP Cloud KMS.

#### Azure Key Vault

In Azure Key vault it is possible to import keys, and the supported key file must be pfx, pem or byok. It is not possible to export the key, but it is possible to export a backup of the key. The intent of this backup is to allow the client to reuse the keys already stored in a key vault in another key vault. It is important to note that this

operation is only available inside the same client account, e. g., it is not possible to export a backup key from one client and import it into another client.

## 6.2 Available encryption levels

Table 2, compares side by side the available encryption levels offered by each service.

**Table 2:** Key encryption comparison

	AWS KMS	Azure Key Vault	GCP KMS
<b>Symmetric Key</b>	AES – GCM	-	AES and GCM
<b>Asymmetric Key</b>	-	RSA and EC	RSA_DECRYPT_OAEP_3072_SHA256.
<b>Data Size</b>	4KB	2048bit RSA block	64KB
<b>Unwrap Keys</b>	RSA-OAEP and RSA-PKCS	RSA-OAEP and RSA-PKCS	-
<b>Sign/Verify</b>	-	RSA-PPS AND RSA-PKCS	-

- AWS allows the storage of files with and without encryption (Three types of keys AWS Owned CMKs, AWS Managed CMKs, AWS Costumer Managed CMKs);
- Azure allows the storage of files with Managed Microsoft Keys or Key Vault Keys;
- GCP allows to store files with key managed by Google or key managed by Google Cloud KMS;
- AWS Ensures redundancy and availability of keys within the same region;
- Ensures redundancy and availability of keys within the same region and for a secondary zone;
- GCP Ensures redundancy and availability of keys within the same region;

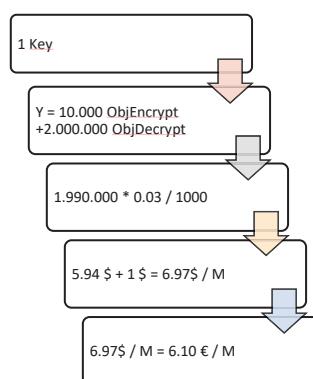
## 6.3 HSM keys price

The price of the HSM generated keys are:

- Key creation: 1\$
- Key rotation: 1\$
- Price for each 10000 operations run over the key: 0.03\$

The price of the keys is related with their generation in the HSM (Hardware Security Module).

For AWS KMS and Azure Key Vault there is a free tier regarding the previously analyzed type of keys that sets the price to one dollar per month if the number of operations executed with that key does not exceed 20000 operations per month in which the monthly price ends being only the price of the key (1\$).



**Figure 1:** Pricing example

Figure 2 shows other pricing example, where the number of operations is higher than the free tier can offer. In this example with one key and over 2.000.000 operations run over that key (10.000 encryption operations plus 2.000.000 decrypting operations), minus the free tier operation count, the price of the operations alone is 5.94\$ plus the 1\$ fee of the active key, sums up to 6.94\$ (approximately 6.10€).

## 7. Conclusions

In this paper it is possible to compare and verify inscription services that are important when dealing with any type of stored data on the cloud, critical or non-critical. This information may be of various types such as, pictures, citizen's documents and city services information. Several of these files contain personal and/or critical information and most of the times are stored unencrypted, mainly depending on access control systems to maintain their security.

At the level of comparison between these services it is possible to affirm that, the platform that provides the best QoE (Quality of Experience) is de AWS (Amazon Web Services), following is GCP (Google Cloud Platform) and lastly is Azure Key Vault. During the experimental setup the services in AWS Platform and GCP took the least time to setup, research and effort to get up and running, but the AWS CloudTrail provides more user-friendly interface.

Regarding the levels of protection in the creation of the keys the AWS only provides HSM (Hardware Security Modules) while in Google and Azure exist the possibility to choose between HSM and Software. Both AWS KMS and Azure Key Vault only generate symmetric keys while in GCP it is possible to generate symmetric and asymmetric keys, as such GCP is the better choice since it offers more options.

When uploading files in AWS S3 (Simple Storage) it is possible to store encrypted and unencrypted data, by contrast in GCP and Azure it is only possible to store encrypted data.

The costs that the services are similar, because the price to create, rotate and execute operations over the key are the same.

## Acknowledgements

This article is a result of the CityAction and BlueEyes project, respectively, CENTRO-01-0247-FEDER-017711, and 02/SAICT/2016, supported by Centro Portugal Regional Operational Program (CENTRO 2020), under the Portugal 2020 Partnership Agreement, through the European Regional Development Fund (ERDF), and also financed by national funds through FCT Fundação para a Ciência e Tecnologia, I.P., under the project UID/Multi/04016/2016. Furthermore, we would like to thank the Instituto Politécnico de Viseu for their support, and the students from the course of "Complements of Operating Systems" 2018, for their support."

## References

- Amazon, E. C. (2012, November). Amazon web services. *Amazon web services*.
- Anthes, G. (2010). Security in the cloud. *Communications of the ACM. Security in the cloud. Communications of the ACM*.
- Anthony, A. (2018). AWS: Security Best Practices on AWS. *AWS: Security Best Practices on AWS*.
- Authority, P. C. (2018). AWS Certificate Manager Private Certificate Authority. *AWS Certificate Manager Private Certificate Authority*.
- Calder, B. W. (2011). Windows Azure Storage: a highly available cloud storage service with strong consistency. *Windows Azure Storage: a highly available cloud storage service with strong consistency*.
- Campagna, M. (. (2015). Aws key management service cryptographic details. *Aws key management service cryptographic details*.
- Cusumano, M. (2010). Cloud computing and SaaS as new computing platforms. *Cloud computing and SaaS as new computing platforms*.
- Diogenes, Y. S. (2016). Microsoft Azure Security Infrastructure. *Microsoft Azure Security Infrastructure*.
- Freato, R. (2015). Microsoft Azure Security. *Microsoft Azure Security*.
- J Varia, S. M. (2014). Overview of amazon web services. *Overview of amazon web services*.
- Lewis, G. A. (2013, January). Role of standards in cloud-computing interoperability. *Role of standards in cloud-computing interoperability*.
- Murty, J. (2008). Programming amazon web services: S3, EC2, SQS, FPS, and SimpleDB. O'Reilly Media, Inc.
- O'Boyle, D. (2017). Your Secrets in Cloud-Based Key Management Services. *Your Secrets in Cloud-Based Key Management Services*.
- OGÂAU-NEAMTIU, F. (2015). CRYPTOGRAPHIC KEY MANAGEMENT IN CLOUD COMPUTING. *Defense Resources Management in the 21st Century. CRYPTOGRAPHIC KEY MANAGEMENT IN CLOUD COMPUTING. Defense Resources Management in the 21st Century*.
- Qian, L. L. (2009, December). Cloud computing: An overview. In IEEE International Conference on Cloud Computing. *Cloud computing: An overview*. In IEEE International Conference on Cloud Computing.

- Ramakrishnan, R. S.-S. (2017, May). Azure data lake store: a hyperscale distributed file service for big data analytics. Azure data lake store: a hyperscale distributed file service for big data analytics.
- Redkar, T. G. (2009). Windows azure platform. Windows azure platform.
- Riti, P. (2018). Identity and Access Management with Google Cloud Platform. In Pro DevOps with Google Cloud Platform. Identity and Access Management with Google Cloud Platform. In Pro DevOps with Google Cloud Platform.
- Shrivastava, M. a. (2015). Implementing Advanced Encryption Standard (A-AES) to Secure Data on the Cloud. Implementing Advanced Encryption Standard (A-AES) to Secure Data on the Cloud.
- Steven Porter, T. H. (2018). Google Cloud Platform for Developers: Build Highly Scalable Cloud Solutions.
- Zoiner Tejada, M. L. (2018). Developing Microsoft Azure Solutions . Developing Microsoft Azure Solutions .

# Description Logics and Axiom Formation for a Digital Forensics Ontology

Dagney Ellison, Adeyemi Richard Ikuesan and Hein Venter

University of Pretoria, South Africa

[dagneye@hotmail.com](mailto:dagneye@hotmail.com)

[aikuesan@cs.up.ac.za](mailto:aikuesan@cs.up.ac.za)

[heinventer@gmail.com](mailto:heinventer@gmail.com)

**Abstract:** Knowledge management in the fields of digital forensics and security exist in diverse models owing to the peculiar function of each knowledge area. A system of knowledge that functions upon a set of assumptions, facts and rules is called a knowledge base and is built upon the framework of an ontology. Ontology has its roots in philosophy and deals with the question of being and what is true. In this instance, an ontology implemented as a knowledge base requires statements of truth (also referred to as axiom) as the foundation of the knowledge and for reasoning. Such explicit concepts are expressed as Description Logics and are thus useful for reasoning. However, existing digital forensic ontologies as well as applicable knowledge management models are devoid of explicit sets of axioms and conceptualization that can be used to develop a generic knowledge base for digital forensics and security. This study thus developed a set of foundational axioms and conceptualization which comprises Description Logics that can be reliably leveraged to address this limitation. Description Logics is used as the formal language for representing and reasoning about the knowledge in the domain of digital forensics and digital security as this research presents axioms to be used in conjunction with research carried out in the paper entitled An Improved Ontology for Knowledge Management in Security and Digital Forensics. The outcome of this theoretical research provides a set of axioms patently useful to any objective model where reasoning over digital forensics knowledge is required. The paper discusses the origins of Description Logics and describes their usefulness in reasoning - particularly in ontology reasoning for digital forensics and security. Using inference procedures on the description logics, implicit knowledge about concepts and individuals can be inferred resulting in additional axioms.

**Keywords:** ontology, axioms, description logics, digital forensics, knowledge base

---

## 1. Introduction

The growing trend in the knowledge area of information security is such that several stakeholders frequently coin abstractions to represent a body of knowledge. This process has attracted the need for a formal methodology for knowledge management in information security in general. Research methods in the area of knowledge management in information security, particularly digital forensics, have explored the concepts of classification, knowledge representation, taxonomies, models, and frameworks (Adeyemi, Razak and Azhan, 2012; Ikuesan and Venter, 2018), and more recently, ontologies (Ellison, Venter and Ikuesan, 2017). Ontology as an approach to knowledge representation can be used to represent specific knowledge domain or a generic abstraction of multiple domains. However, the fundamental challenge in modeling a specific or generic domain using ontologies is the crafting of appropriate connections between objects in the domain and the logical sequence of the knowledge represented. In ontologies, these challenges are often referred to as axiom formation and description logics. Whilst axioms can be defined as a logical assertion that depicts the underlying conceptualization/theory, description logics depicts a logical formalism that can be used to formulate knowledge representation languages. Fundamentally, both axiom formation and description logics are contingent on the development of a formal language that can be used to explicitly relate object and logical sequences (Noy and McGuinness, 2000). Such a language can, therefore, be used by domain experts to harmonize and annotate the seemingly disintegrated compositions of a knowledge domain. As highlighted in (Noy and McGuinness, 2000), one common objective for the development of such formal explicit standardization of a domain is to enhance the common understanding of the structure of information in that domain.

Language is a method of communication used by humans and computers alike and is a requirement for reasoning. A language is built upon terms and structure (*language / Definition of language in English by Oxford Dictionaries*, no date) and can be described according to its rules. When one speaks, one is limited to the constraints of that language if one wants to be understood. In the same way, languages defined for computer reasoning must contain rules and restrictions. In a particular domain, the content of what the language describes is the knowledge of that domain. Digital forensics is comprised of terms and facts that can be modeled with axioms and description logics and therefore processed by a computer. Reasoning against the domain would be

possible based upon the axioms and logic applied to the domain in order to represent the knowledge of the domain. This is then modeled in the form of an ontology.

Ultimately, the knowledge within digital forensics would be able to be queried and results returned according to the defined constraints of the language. Similar works which reveal axioms for a digital forensic ontology are presented in section 2. The term given to a system developed on top of an ontology is a knowledge base and is expounded upon in the background section, section 3. Section 4 entails the contribution of axioms for the domain of digital forensics. Section 5 discusses the set of axioms and finally, section 6 concludes this paper.

## 2. Related work

Oltramari, Cranor, and Walls (Oltramari *et al.*, 2014) present an ontology in their paper about creating an ontology for situational awareness in cyberspace by describing the relationships between the different aspects of a potential digital incident with respect to security policy implementations, the attacker and the defender. Their ontology, therefore, consists of the components: Role, Requirement, Policy, Artifact, Action, Task, Plan and Abstract Quality. Some concepts are from the DOLCE-SPRAY base ontology upon which their ontology is built. Each concept was extended to provide a further classification for the description logics. The paper presents a list of description logics for a section of the ontology describing the necessary conditions between the components such as the restriction on what a defensive operation is and an attacker as depicted in figures 1 and 2 below.

DEF_OP $\equiv \exists \text{ hasParticipant. DEFENDER}$ $\sqcap \exists \text{ executes. MISSION_PLAN}$ $\sqcap \exists \text{ hasParticipant. COUNTERMEASURE}$ $\sqcap \exists \text{ hasRequirement. DEF_REQ}$
--

**Figure 1:** Excerpt of description logics for ‘defensive operation’ in the ontology presented by Oltramari et al.

ATTACKER $\equiv \forall \text{ exploits. VULNERABILITY} \sqcap \exists \text{ uses. THREAT}$
---

**Figure 2:** Excerpt of description logics for ‘attacker’ in the ontology presented by Oltramari et al.

Additionally, Herzog, Shahmehri, and Duma (Herzog, Shahmehri and Duma, 2007) present an ontology for the purpose of creating “an ontology that provides a general overview, contains detailed domain vocabulary, allows queries, supports machine reasoning and may be used collaboratively”. Although the axioms of this ontology are not revealed in the paper, there is an emphasis on vocabulary and its significance in reasoning. The ontology also aims to “provide a general overview of the domain of information security” like the ontology used in this paper.

Herzog et al. illustrate reasoning over their ontology by means of inference rules and therefore are able to deduce results such as the threats which violate the confidentiality of data using a reasoner such as Pellet of FaCT. The inference used in the ontology is for the purpose of “sorting and categorizing threats and countermeasures according to security goals, assets and defense strategies”. Furthermore, the paper goes as far as to demonstrate the effective querying of the ontology with the popular query language SPARQL (SPARQL Protocol and RDF Query Language).

## 3. Background

Natural language used when describing components of digital forensics is undefined. The language of digital forensics must be defined in order for a knowledge-based system to reason over the facts of the domain of digital forensics. A knowledge-based system, therefore, comprises of axioms which represent components of the knowledge of the domain which are made explicit. The explicit knowledge contained within the system, when reasoned over, results in implicit knowledge being revealed. The knowledge contained within the knowledge base can be extended and reused indicating that the system would not grow old and become outdated as the information it contains can always be kept up to date and therefore always be relevant. A knowledge base is a type of database of information but with different capabilities to that of a regular relational database management system. A knowledge base is built upon a framework known as an ontology. An ontology is the interconnectedness of data of the domain linked through concepts and relationships between the concepts defined by the ontology developer. The idea of having linked items of data is what gives the ontology value as the knowledge base is then able to return relevant bits of data as opposed to returning entire documents containing one small component of relevant data.

The idea of an ontology stems from philosophy, as Zúñiga (Zúñiga, 2001) states, an ontology “serve as the bridge for the coming together of information systems and philosophy”. Axiom formation in the context of reasoning is a component of philosophy along with logical reasoning. It is upon the theory of logic that ontologies are able to be made useful for reasoning and querying.

### 3.1 Description logics

Logic is the subject of reasoning against laws of truth. Propositional logic deals with the *propositions* (statements) of natural language and how the propositions are connected to each other (Levin, 2017). A proposition can be true or false and can be used to determine the validity of an argument by the use of truth tables or inference rules (O'Regan, 2017).

First-order logic extends the propositions of propositional logic into *predicates*. This means that multiple propositions can be represented by a predicate as predicates can take variables. In this way, first-order logic allows for new facts to be ascertained via deductive reasoning (O'Regan, 2017). Description logics are *decidable fragments* of first-order logic. Decidable means a statement can be computed as being true or false. The fragment is a subset of first-order logic acquired by enforcing syntactic restrictions on first-order logic. Description logics are formal languages constructed for the use of representing facts and performing reasoning. Description logics are comprised of the following three types of elements: individuals/constants, concepts/unary relations and roles/binary relations. The interactions of these three elements define the language and this is known as the description logic syntax. The semantics is based on the predicate logic.

There are various types of description logic languages available which can be categorized based on its particular composition. Symbols represent the permissible features of the language such as  $\mathcal{AL}$ .  $\mathcal{AL}$  is the base language also known as the Attributive Language and is typically extended to  $\mathcal{ALC}$  where  $C$  is used as a shorthand notation representing the constructs of  $\mathcal{UE}$  (concept union and full existential quantification). A description logic  $\mathcal{AL}$  is made up of the constructs: atomic negation, concept intersection, universal restriction, and limited existential quantification (Sikos, no date). This is represented in the last four rows of Table 1, where  $A$  is an atomic concept,  $C$  and  $D$  are concepts, and  $R$  is an atomic role.

The semantics is defined according to interpretation  $I$  which consist of the set  $\Delta^I$  known as the domain of  $I$ . The interpretation function  $\cdot^I$  assigns an atomic concept  $A$  to a set  $A^I \subseteq \Delta^I$ , an atomic role  $R$  to a binary relation  $R^I \subseteq \Delta^I \times \Delta^I$  (Obitko, no date)

**Table 1:** Syntax and semantics of the  $\mathcal{AL}$  description logic

	Syntax	Semantics
Atomic concept	$A$	$A^I \subseteq \Delta^I$
Atomic role	$R$	$R^I \subseteq \Delta^I \times \Delta^I$
Top concept	$T$	$\Delta^I$
Bottom concept	$\perp$	$\emptyset$
Atomic negation	$\neg A$	$\Delta^I \setminus A^I$
Concept intersection	$C \sqcap D$	$C \cap D$
Limited existential quantification	$\exists R.T$	$\{a \in \Delta^I \mid \exists b.(a,b) \in R^I\}$
Universal restriction	$\forall R.C$	$\{a \in \Delta^I \mid \forall b.(a,b) \in R^I \Rightarrow b \in C^I\}$

$\mathcal{ALC}$  is the Attributive Language with Complement. It is an extension of  $\mathcal{AL}$  including the constructs for concept union and *full* existential quantification. This is represented in Table 2.

**Table 2:** Syntax and semantics of the  $\mathcal{ALC}$  description logic

	Syntax	Semantics
Atomic concept	$A$	$A^I \subseteq \Delta^I$
Atomic role	$R$	$R^I \subseteq \Delta^I \times \Delta^I$
Top concept	$T$	$\Delta^I$
Bottom concept	$\perp$	$\emptyset$
Atomic negation	$\neg A$	$\Delta^I \setminus A^I$
Concept union	$C \sqcup D$	$C \cup D$
Concept intersection	$C \sqcap D$	$C \cap D$
Full existential quantification	$\exists R.C$	$\{a \in \Delta^I \mid \exists b.(a,b) \in R^I \wedge b \in C^I\}$
Universal restriction	$\forall R.C$	$\{a \in \Delta^I \mid \forall b.(a,b) \in R^I \Rightarrow b \in C^I\}$

Further constructs may be added to the ALC description logic such as role hierarchy, functional properties, nominals, and inverse properties.

A description logic language is made up of a combination of available syntactical rules. The choice in rules is often motivated by the trade-off between the expressivity of the language – the measure of what can be expressed with the given constructs – and the performance during classification and reasoning. As asserted in (Krötzsch, Simančík and Horrocks, 2013), “the best balance between expressivity of the language and complexity of reasoning depends on the intended application”. However, there are additional considerations such as the rigor – the extent of the ontology’s satisfiability and consistency – and the semantics – the intended meaning (Stevens, 2001).

### 3.2 Reasoning

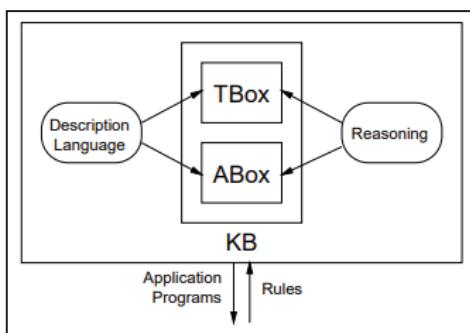
Ontologies are built on an open world assumption (OWA). An OWA is in opposition to the CWA (closed world assumption) which assumes that what is not known must ultimately be *false*. In contrast, with the OWA, the result *false* would not be returned when the outcome of reasoning or a query is, in fact, *unknown* (Sequeda, no date). Therefore, to get from a world where everything is possible (OWA) to a world where reasoning can occur over predefined precepts and return a true or false response, restrictions on the knowledge are required. The restrictions specify the knowledge according to what is possible. These restrictions define what is not possible (facts that are excluded or forbidden).

In the OWA, for questions posed against the ontology that the ontology is unable to explicitly answer, the ontology will not be able to return a decided true or false response. Information that is missing is assumed to be unknown as opposed to false (Keet, 2013). Unless a statement is explicitly defined or can be inferred from existing knowledge, the ontology is unable to answer and therefore the OWA is incomplete by default. This, therefore, gives room for extending the knowledge available to the knowledge-based system. In contrast, a Closed World Assumption applies when there is complete information of a domain.

The reasoning is the computation of inference within the ontology (Krötzsch, Simančík and Horrocks, 2013) and involves functions on the ontology as a whole as well as functions on the relationships between individuals within the ontology based on class definitions. The purpose of a reasoner is to infer implicit knowledge from the given ontology model. Logical consequences are deduced from a defined set of axioms (the ontology classes and relationships) and assertions (where individuals are placed within the ontology classes). The main goals of reasoning are subsumption (subclass checking), ontology consistency, class satisfiability, realization (instance checking) and conjunctive query answering (Meditkos, no date). Ontology consistency is based upon whether or not the models of the ontology contain a contradiction – if so, the ontology is inconsistent. Class satisfiability considers whether or not the description of the class is broken.(Sattler, Stevens and Lord, 2013)

### 3.3 Knowledge base

A knowledge base is formally represented as  $K = \langle A, T \rangle$ . In other words, a knowledge base is made up of two components known as a TBox and an ABox. The TBox is the set of axioms that express terminological knowledge. The ABox is the set of axioms that assert facts by means of assigning an individual to a particular class/concept (Paptaixarhis *et al.*, 2009). It is upon the TBox and ABox that reasoning is carried out. This is illustrated in figure 3.



**Figure 3:** Basic architecture of a knowledge-based system (Baader and Nutt, no date)

### 3.4 General axioms for a digital forensics knowledge base

The axioms presented in this paper are based upon the ontology presented in (Ellison, Venter and Ikuesan, 2017). In this ontology, high-level elements of the domain of digital forensics are represented. In order to make the axioms of the ontology more specific, four selected high-level concepts are classified further, namely Platform, Device, Discipline, and Evidence. In addition, the legal component of digital forensics was also considered and is presented in relation to evidence.

The axioms have been developed in such a way as to consider restrictions on abstract concepts of digital forensics. The full extent of implemented description logics cannot be observed in an image of the ontology, however, the hierarchy for each concept has been captured and included for convenience. This is then followed by the language components and restrictions for each concept. The hardware facet of digital forensics, in general, is not elaborated in this paper but only the types of evidence and platforms for software. In the next subsections, concepts such as platform, device, evidence, discipline, and legal are presented. The extension of the conceptual depiction and the corresponding axioms are logically presented.

### 3.5 Platform concept

The classification of the platform concept is presented in Figure 2. This is the first concept to be considered as it is the smallest component of a computer, abstractly, in terms of hardware or software. The first sub-classification is, therefore, hardware and software as both can be referred to as a platform. For the sake of interest, hardware would have sub-classes of items such as memory (Flash, RAM, etc.). The software concept is immediately classified further into four general software components of a computer – the operating system, applications, utilities, and services. Each software platform is then classified into different types.



**Figure 1:** Extension of the platform concept

The developed axioms that represent the platform concept, as highlighted in Figure 4, are further present in Table 3. Hardware and software are presented as distinct items. Only three levels of classification are presented as this is sufficient for the general axioms presented in this manuscript. However, Figure 4 does present examples of further classification for interest.

**Table 3:** Axioms for the platform

1	HARDWARE ⊑ PLATFORM
2	SOFTWARE ⊑ PLATFORM
3	HARDWARE ⊑ ~ SOFTWARE
4	SOFTWARE ⊑ ~ HARDWARE
5	MEMORY ⊑ HARDWARE

6	PERSISTENT_MEMORY ⊑ MEMORY
7	NON-PERSISTENT_MEMORY ⊑ MEMORY
8	OPERATING_SYSTEM ⊑ SOFTWARE
9	SERVICE ⊑ SOFTWARE
10	UTILITY ⊑ SOFTWARE
11	APPLICATION ⊑ SOFTWARE

### 3.6 Device concept

Device forensics is a major component in both software and hardware platform. A comprehension of the composition of a device and what type or category of evidence to extract from such a device presents a milestone to different forensics stakeholders. The classification of devices used in digital forensics is presented in Figure 5. The devices were first classified according to where they are deployed or used. Therefore, the first sets of subclasses are network devices, personal devices, shared devices, peripheral devices, and other devices. Network devices comprise those that enable and facilitate network connections. Personal devices are those that belong to a person such as a mobile phone or a personal computer. Shared devices consist of communication devices and computers that provide a service to a group of people. Peripheral devices are devices which are or can be connected to a computer such as a keyboard or external memory. Other devices are devices such as games consoles.



Figure 2: Extension of device concept

The only axiom presented for this concept is the general statement that a device, regardless of what it is, is made up of both a hardware and a software component. This logical depiction is shown in Table 4.

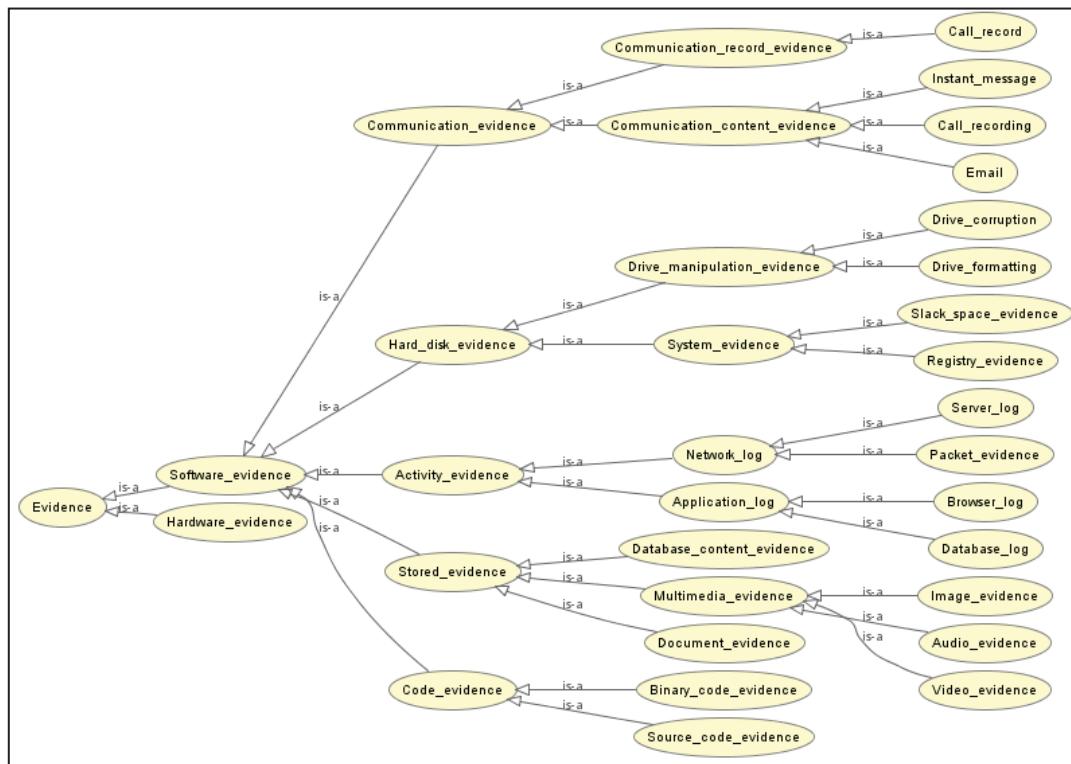
Table 4: Axiom for device

12	DEVICE ≡ ∃hasHardware.HARDWARE ⊓ ∃hasSoftware.SOFTWARE
----	--

### 3.7 Evidence concept

The question of what constitutes evidence is a major factor in a digital investigation, and the challenge varies from one stakeholder to another. Evidence can be generally defined as either hardware or software. The classification of types of Evidence is presented in Figure 6. The classification of hardware evidence is omitted at this stage. The software evidence types are classified into communication evidence, hard disk evidence, activity evidence, stored evidence and code evidence. These types of evidence are based upon the nature of the

evidence – where and how they exist. Communication evidence is the message content in whichever medium as well as the record of the message metadata such as time, date, duration, and location if applicable. Hard disk evidence is evidence either of manipulation of the hard disk or evidence of tampering with the registry, for example. Activity evidence is the record of events which took place in a particular application or over some network. Stored evidence is evidence that exists in a type of file on a device. Finally, code evidence is evidence of source code or binary code being modified.



**Figure 3:** Extension of evidence concept

The axioms of the types of evidence according to platform and device are further presented in Table 5. The first set (13 - 27) are the axioms of subsumption depicted in Figure 6. The second set is the axioms which describe the relationship of types of evidence with regards to platform and device, as highlighted in Table 6.

**Table 5:** Subsumption axioms for evidence

13	SOFTWARE_EVIDENCE ⊑ EVIDENCE
14	HARDWARE_EVIDENCE ⊑ EVIDENCE
15	COMMUNICATION_EVIDENCE ⊑ SOFTWARE_EVIDENCE
16	HARD_DISK_EVIDENCE ⊑ SOFTWARE_EVIDENCE
17	ACTIVITY_EVIDENCE ⊑ SOFTWARE_EVIDENCE
18	STORED_EVIDENCE ⊑ SOFTWARE_EVIDENCE
19	COMMUNICATION_RECORD_EVIDENCE ⊑ COMMUNICATION_EVIDENCE
20	COMMUNICATION_CONTENT_EVIDENCE ⊑ COMMUNICATION_EVIDENCE
21	DRIVE_MANIPULATION_EVIDENCE ⊑ HARD_DISK_EVIDENCE
22	SYSTEM_EVIDENCE ⊑ HARD_DISK_EVIDENCE
23	APPLICATION_LOG ⊑ ACTIVITY_EVIDENCE
24	MULTIMEDIA_EVIDENCE ⊑ STORED_EVIDENCE
25	CODE_EVIDENCE ⊑ SOFTWARE_EVIDENCE
26	BINARY_CODE_EVIDENCE ⊑ CODE_EVIDENCE
27	SOURCE_CODE_EVIDENCE ⊑ CODE_EVIDENCE

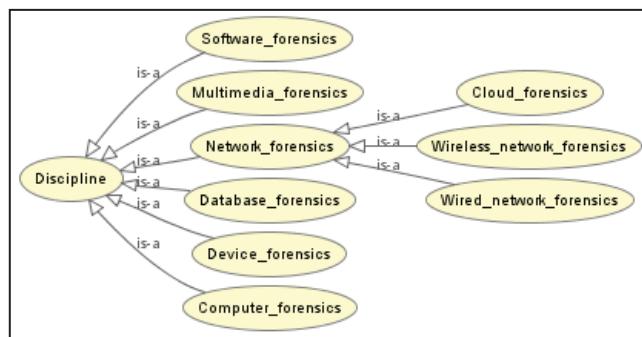
**Table 6:** Axioms for evidence

28	EMAIL_EVIDENCE ⊑ COMMUNICATION_CONTENT_EVIDENCE ∩ ∃isFoundIn.EMAIL_CLIENT
29	INSTANT_MESSAGE_EVIDENCE ⊑ COMMUNICATION_CONTENT_EVIDENCE ∩ ∃isFoundIn.IM_CLIENT
30	MULTIMEDIA_EVIDENCE ⊑ MULTIMEDIA_EVIDENCE ∩ ∃isFoundIn.(APPLICATION ∪ MEMORY) ∩ ∃isFoundOn.(PERSONAL_DEVICE ∪ SHARED_DEVICE ∪ EXTERNAL_STORAGE)

31	$\text{DATABASE\_EVIDENCE} \sqsubseteq \text{STORED\_EVIDENCE} \sqcap \exists \text{isFoundIn}.\text{DATABASE\_SERVICE}$
32	$\text{DOCUMENT\_EVIDENCE} \sqsubseteq \text{STORED\_EVIDENCE} \sqcap \exists \text{isFoundIn}.\text{(APPLICATION} \sqcup \text{MEMORY)}$
33	$\text{DATABASE\_LOG} \sqsubseteq \text{APPLICATION\_LOG} \sqcap \exists \text{isFoundIn}.\text{DATABASE\_SERVICE}$
34	$\text{BROWSER\_LOG} \sqsubseteq \text{APPLICATION\_LOG} \sqcap \exists \text{isFoundIn}.\text{(DATABASE\_SERVICE} \sqcup \text{BROWSER)}$
35	$\text{NETWORK\_LOG} \sqsubseteq \text{ACTIVITY\_EVIDENCE} \sqcap \exists \text{isFoundIn}.\text{(NETWORK\_SERVICE} \sqcup \text{NETWORK\_DEVICE)}$
36	$\text{CALL\_RECORDING} \sqsubseteq \text{COMMUNICATION\_CONTENT\_EVIDENCE} \sqcap \exists \text{isFoundIn}.\text{DATABASE\_SERVICE} \sqcup \exists \text{isFoundOn}.\text{CELLULAR\_DEVICE}$
37	$\text{CALL\_RECORD} \sqsubseteq \text{COMMUNICATION\_RECORD\_EVIDENCE} \sqcap \exists \text{isFoundIn}.\text{DATABASE\_SERVICE} \sqcup \exists \text{isFoundOn}.\text{CELLULAR\_DEVICE}$

### 3.8 Discipline concept

Figure 7 presents the classification of Disciplines in digital forensics. Discipline is sub-classified into the six classes software forensics, multimedia forensics, network forensics, database forensics, device forensics, and computer forensics. Only network forensics was classified further into cloud forensics, wireless forensics, and wired forensics.



**Figure 4:** Extension of discipline concept

The set of axioms below describes the different disciplines of digital forensics according to evidence types, platform, and device. Some types of evidence are restricted to only one type of evidence whereas others are restricted to a few. Some types of evidence are restricted to devices in general whereas some are restricted to only specific platforms. Where the evidence type contains its own restriction with regards to device or platform, these restrictions are not included for these axioms.

**Table 7:** Axioms for discipline

38	$\text{MULTIMEDIA\_FORENSICS} \sqsubseteq \text{DISCIPLINE} \sqcap \forall \text{concerns}.\text{MULTIMEDIA\_EVIDENCE}$
39	$\text{DATABASE\_FORENSICS} \sqsubseteq \text{DISCIPLINE} \sqcap \forall \text{concerns}.\text{(DATABASE\_CONTENT\_EVIDENCE} \sqcap \text{DATABASE\_LOG})$
40	$\text{DEVICE\_FORENSICS} \sqsubseteq \text{DISCIPLINE} \sqcap \forall \text{concerns}.\text{HARDWARE\_EVIDENCE} \sqcap \exists \text{isFoundOn}.\text{DEVICE}$
41	$\text{SOFTWARE\_FORENSICS} \sqsubseteq \text{DISCIPLINE} \sqcap \forall \text{concerns}.\text{CODE\_EVIDENCE} \sqcap \exists \text{isFoundIn}.\text{SOFTWARE}$
42	$\text{NETWORK\_FORENSICS} \sqsubseteq \text{DISCIPLINE} \sqcap \forall \text{concerns}.\text{NETWORK\_EVIDENCE} \sqcap \exists \text{isFoundOn}.\text{DEVICE}$
43	$\text{COMPUTER\_FORENSICS} \sqsubseteq \text{DISCIPLINE} \sqcap \forall \text{concerns}.\text{HARD\_DISK\_EVIDENCE} \sqcap \exists \text{isFoundIn}.\text{(UTILITY} \sqcup \text{OS)}$

### 3.9 Legal concept

The five concepts for the legal aspect of digital forensics are Case, Jurisdiction, Geolocation, Act, and Section. The axioms given in tables 8-10 depict the relationships between the five concepts. A case is made up of evidence, belongs to jurisdiction and has relevant Acts. A location also has jurisdiction but also has the subclasses Country and State. And finally, an Act has some Sections.

**Table 8:** Axioms for case

44	$\text{CASE} \sqsubseteq \exists \text{hasEvidence}.\text{EVIDENCE} \sqcap \exists \text{hasJurisdiction}.\text{JURISDICTION} \sqcap \exists \text{hasAct}.\text{ACT}$
----	--

**Table 9:** Axioms for geolocation

45	$\text{COUNTRY} \sqsubseteq \text{GEOLOCATION}$
46	$\text{STATE} \sqsubseteq \text{COUNTRY}$
47	$\text{GEOLOCATION} \sqsubseteq \exists \text{hasJurisdiction}.\text{JURISDICTION}$

**Table 10:** Axioms for Act

48	$\text{ACT} \sqsubseteq \exists \text{hasSection}.\text{SECTION}$
----	---

#### **4. Discussion**

The natural language of digital forensics is not useful for machine processing as it has no predictable structure. Therefore, automated processing requires a structure to be introduced into the language and this is in the form of logic in order for a true or false response to be achieved. A true or false response is the result of any logical argument given a set of axioms except in the open world assumption where the result may be unknown if not true and not explicitly set to false.

A language is made up of a set of rules which define what is and what is not permissible. The basic subset of permissible rules for the common ALC language allows for subsumption, negation, intersection, union, existential quantification and universal restriction. Based on this, the general axioms for a knowledge base in digital forensics is developed.

The axioms introduce the beginnings of a formal structure for the language of digital forensics which a computer can understand and process against. Only the software aspects of this are represented in this manuscript. Although, the hardware is not totally omitted but is limited to device classification. The axioms relate firstly a device to its platform components which are represented by describing a device as a subset of the intersection of hardware and software for reason that there exists no device that does not consist of both hardware and software. The next set of axioms related different types of evidence to devices and platforms. Some types of evidence are particular to some platform or some device and where this is a restriction, this is reflected in the axioms. Finally, the disciplines within digital forensics are related to the types of evidence to which they pertain to, along with the specific platform or device on which they may be found. With all these axioms in place, the task of reasoning may be conducted in order to infer new axioms and determine the state of classes within the ontology and the state of the ontology. Furthermore, a query language layered on top of the ontology would then be able to make use of the restrictions in order to determine the true or false outcomes of statements and therefore return a meaningful set of data based on the axioms.

#### **5. Conclusion**

A set of general axioms was provided for the domain of digital forensics for the purpose of a knowledge base. The axioms are constructed for the purpose of a general ontology on digital forensics and in this context make facts about digital forensics explicit allowing for reasoning to be conducted over the domain for the purpose of information retrieval. Only in the work by Oltramari did the author find similar axioms within the field of digital forensics. Therefore, the axioms presented in this paper add to the somewhat limited set of axioms for the field that can be found. As the axioms in this paper concern only a small portion of the entire field of digital forensics, one could enhance these axioms by including further concepts such as techniques, tools, and legal cases. One could also further classify the concepts that were enhanced in this paper.

#### **References**

- Adeyemi, I., Razak, S. and Azhan, N. (2012) 'Identifying critical features for network forensics investigation perspectives', *arXiv preprint arXiv:1210.1645*.
- Baader, F. and Nutt, W. (no date) 'Basic Description Logics'. Available at: <https://www.inf.unibz.it/~franconi/dl/course/dlhb/dlhb-02.pdf> (Accessed: 14 January 2019).
- Ellison, D., Venter, H. S. and Ikuesan, A. (2017) 'An Improved Ontology for Knowledge Management in Security and Digital Forensics', in *16th European Conference on Cyber Warfare and Security (ECCWS 2017)*. Dublin, pp. 725–733.
- Herzog, A., Shahmehri, N. and Duma, C. (2007) 'An Ontology of Information Security', *International Journal of Information Security and Privacy*, 1(4), pp. 1–23. doi: 10.1145/508171.508180.
- Ikuesan, A. R. and Venter, H. S. (2018) 'Digital forensic readiness framework based on behavioral-biometrics for user attribution', in *2017 IEEE Conference on Applications, Information and Network Security, AINS 2017*. Miri, Malaysia: IEEE Comput. Soc, pp. 54–59. doi: 10.1109/AINS.2017.8270424.
- Keet, C. M. (2013) 'Open World Assumption', in *Encyclopedia of Systems Biology*. New York, NY: Springer New York, pp. 1567–1567. doi: 10.1007/978-1-4419-9863-7\_734.
- Krötzsch, M., Simančík, F. and Horrocks, I. (2013) *A Description Logic Primer* \*. Available at: <https://arxiv.org/pdf/1201.4089.pdf> (Accessed: 29 January 2019).
- language | Definition of language in English by Oxford Dictionaries* (no date). Available at: <https://en.oxforddictionaries.com/definition/language> (Accessed: 10 January 2019).
- Levin, O. (2017) *Discrete Mathematics: An Open Introduction*. 2nd edn.
- Meditkos, G. (no date) 'Rule-based Applications on Top of Ontologies Architectures and Challenges'. Available at: <https://www.iti.gr/iti/files/document/seminars/MeditkosSeminar.pdf> (Accessed: 7 February 2019).
- Noy, N. F. and McGuinness, D. L. (2000) 'Ontology Development 101 : A Guide to Creating Your First Ontology', pp. 1–25.

- O'Regan, G. (2017) 'Propositional and Predicate Logic', in. Springer, Cham, pp. 109–135. doi: 10.1007/978-3-319-64021-1\_6.
- Obitko, M. (no date) *Syntax and Semantics - Introduction to ontologies and semantic web - tutorial*. Available at: <http://www.obitko.com/tutorials/ontologies-semantic-web/syntax-and-semantics.html> (Accessed: 5 February 2019).
- Oltramari, A. et al. (2014) 'Building an Ontology of Cyber Security', in *CEUR Workshop Proceedings*, pp. 54–61. Available at: [http://ceur-ws.org/Vol-1304/STIDS2014\\_T08\\_OltramariEtAl.pdf](http://ceur-ws.org/Vol-1304/STIDS2014_T08_OltramariEtAl.pdf) (Accessed: 8 March 2017).
- Paptaxiarhis, V. et al. (2009) 'Developing Rule-Based Web Applications', in *Handbook of Research on Emerging Rule-Based Languages and Technologies*. IGI Global, pp. 371–392. doi: 10.4018/978-1-60566-402-6.ch016.
- Sattler, U., Stevens, R. and Lord, P. (2013) '(I can't get no) satisfiability', *Ontogenesis*. Available at: <http://ontogenesis.knowledgeblog.org/1329> (Accessed: 7 February 2019).
- Sequeda, J. (no date) *Introduction to: Open World Assumption vs Closed World Assumption - DATAVERSITY*. Available at: <https://www.dataversity.net/introduction-to-open-world-assumption-vs-closed-world-assumption/> (Accessed: 18 February 2019).
- Sikos, L. F. (no date) *The ALC Description Logic*. Available at: <https://www.lesliesikos.com/alc-description-logic/> (Accessed: 5 February 2019).
- Zúñiga, G. L. (2001) 'Ontology', in *Proceedings of the international conference on Formal Ontology in Information Systems - FOIS '01*. New York, New York, USA: ACM Press, pp. 187–197. doi: 10.1145/505168.505187.

# **Self-Directed Learning Tools in USAF Multi-Domain Operations Education**

**Nathaniel Flack and Mark Reith**

**Department of Electrical and Computer Engineering, Air Force Institute of Technology,  
Dayton, USA**

[nathaniel.flack@afit.edu](mailto:nathaniel.flack@afit.edu)  
[mark.reith@afit.edu](mailto:mark.reith@afit.edu)

**Abstract:** The United States Air Force (USAF) is aggressively pursuing transformation in the areas of Multi-Domain Operations (MDO) and enterprise education and training. The drive for these changes goes up to the highest levels of US Air Force and Department of Defense leadership motivated by a rapidly evolving world. These evolutions are forcing military organizations to educate personnel rapidly and more effectively while adapting to new threats from multiple contested domains. Advancement in these two areas requires new information systems enabled by modern information technology to empower rich collaboration and innovation leveraging current operational experience and industry best practice. This paper explores the challenges facing MDO education in the 21<sup>st</sup> Century and defines elements of a potential solution drawing from Self-Directed Learning theory, proven commercial technology, and the new USAF Continuum of Learning construct. Extending the framework and cloud-based learning system created by the Air Force Institute of Technology called the Cyber Education Hub™ we propose a solution called the Multi-Domain Operations Hub. This new environment would enable content consumption, sharing, and creation as well as collaboration and innovation among members across all five warfighting domains (land, sea, air, space, and cyber) and all United States military branches. By utilizing the Topic Map and Knowledge, Skills, and Abilities Tree concepts integral to the existing Cyber Education Hub, the proposed solution will use the elements most attractive to Air Force functional communities while replacing cyber-specific elements with those relevant to MDO. The paper concludes by offering possible research questions to inform the development and implementation of the MDO Hub while proposing a potential human subjects research experiment to test the effectiveness of the system.

**Keywords:** multi-domain operations (MDO), multi-domain command and control (MDC2), US Air Force education and training, self-directed learning (SDL), cyber education hub™ (CEH™), MDC2 card game, gamification

---

## **1. Introduction**

In the past two years the United States Air Force (USAF) has aggressively pursued innovation in two broad areas. The first, and most pressing, is a shift to a “Multi-Domain Operations (MDO)” mindset. The USAF Chief of Staff, General David Goldfein (2018) says this transformation is vital to prepare for future warfare “that will require [the US Military] to defend against and attack foes on land and sea as well as in the air, space and cyberspace.” Specifically, Goldfein is pushing the USAF to “master Command and Control of the multi-domain battle,” which is also referred to as Multi-Domain Command and Control (MDC2). The second area is an overhaul of its education and training paradigm, shifting from multi-month face-to-face programs to a more modular, agile, and on-demand structure. Given the size of the force and the complexity of the proposed changes, both of these initiatives will be largely unsuccessful without the utilization of innovative technology to provide elevated collaboration and engagement from members across the force. The 2018 US National Defense Strategy states, “Today, every domain is contested—air, land, sea, space, and cyberspace.” Therefore, these transformations are vital to the future success of America’s military in future conflicts.

When he was appointed to his current position in 2016, Goldfein made Command and Control (C2) in a multi-domain context one of his top three priorities. He wrote, “The changing national security environment also requires us to examine who we sense, decide, and act rapidly and in concert across all domains – or to put it another way, master command and control of the multi-domain battle” (Goldfein, 2017). His message to all Airmen and the defense industrial and technological base is that the dominance the USAF enjoys today in the air, space, and cyber domains is not good enough. All the capabilities in these domains must be integrated, along with the sea and land domains, to create new and dramatic effects. Alberts and Hayes (2006) write, “New C2 Approaches are the fulcrum of an Information Age transformation of the Department of Defense (DoD) and understanding Command and Control is among the most important and urgent tasks we have on the critical path to transformation and the ability to meet 21st century mission challenges.” In a recent December 2018 article, Goldfein again highlighted the need for a shift to a multi-domain mindset ensuring future technology to be able to quickly gather information from multiple domains and then “just as quickly direct military actions will have the decisive advantage in battle.”

At the same time, many have called for changes to military education and training strategies. The former Secretary of Defense, James Mattis, is one of these voices. He emphasizes that the US Armed Forces need to “be prepared to deal with technological, operational, and tactical surprise, which requires changes to the way we train and educate our leaders and our forces...” (Mattis, 2017). Air Education and Training Command (AETC), the USAF’s Major Command dedicated to recruiting, training, and educating its members, is responding to these calls by transforming the way they think about education and training and forging new information technology tools to support that transformation. Given the push for more MDC2 innovation, from leaders like the Secretary of the Air Force, the Honorable Heather Wilson, changes in education and training should focus on creating a system to facilitate MDO collaboration and develop innovative solutions. Another key leader in the MDC2 arena is Col Jeffry Burdett, the 505<sup>th</sup> Training Group commander. Shortly after leading one of the first major MDC2 exercises for the USAF in 2017 he stated, “The Air Force needs a mechanism for tracking operational-level C2 experience” (Caputo, 2017). This represents another angle of MDO that could be fulfilled by the innovative use of technology.

In summary, USAF leadership is asking for solutions that will advance the ball on MDO by providing innovative and relevant solutions engineered through collaboration of a diverse team while tracking MDC2 expertise across the force. In response, this paper proposes a potential solution that extends a prototype learning environment designed for cyber education currently under development at the Center for Cyberspace Research (CCR) at the Air Force Institute of Technology (AFIT) and tailoring it for MDO education. The new learning environment will fuse innovative ideas from all levels of the force, lessons learned from real-world experience, and the vision of current and future leaders to shape MDO and MDC2.

## **2. Describing the current challenges**

In an article titled “Rethinking USAF Cyber Education and Training” (Reith et al., 2018), describe the current challenges involved in cyber education in the US military. In response to these challenges they proposed a framework and prototype system called the Cyber Education Hub™ (CEH™). Their framework was created to address significant challenges in the realm of cyber education. First, the ubiquity of cyber in every functional community creates a “scalability and breadth problem.” Cyber education needs to reach to all users, but also contain information specific to their functional community. Second, because cyber is a man-made and man-manipulated environment, it changes at a higher rate than other science and technology fields, creating a “currency problem.” Third, the size and interconnected nature of cyber leads to a “complexity problem.”

### **2.1 Intersection with cyber education challenges**

To varying degrees, MDO and MDC2 education face the same challenges as the cyber domain. First, MDO is inherently broad covering military operations in land, sea, air, space, and cyber. Education solutions must leverage resources, knowledge, and experience from every warfighting domain and make it available to the rest of the DoD. This requires a connection between all learners across various domains also creating a scalability and breadth problem. At a basic level, forces that want to integrate to achieve a common goal must know the general capabilities of each and commanders over multi-domain forces need to know the capabilities under their control to know how best to employ them. Second, MDO must be able to adapt quickly to new information, also creating a currency problem. Goldfein writes, “The changing national security environment also requires us to examine how we sense, decide, and act rapidly and in concert across all domains.” Because the operations themselves must be adaptable and broadly coordinated, education methods to teach these principles must also be ready to change as new information surfaces. Last, the interweaving of capabilities from multiple domains with personnel from many organizations, including national partners and coalitions, produces immense complexity. MDO requires warfighters who understand and execute their function when needed, but also know how they fit into the broader mission so that they can integrate effectively and innovate when necessary without creating unnecessary risk. Harris (2018) attests, “Planners and operators of one domain must have not only the skills to perform their own missions, but they must also understand how planners and operators of other domains assure or even challenge their mission accomplishment.”

### **2.2 Specific MDO education challenges**

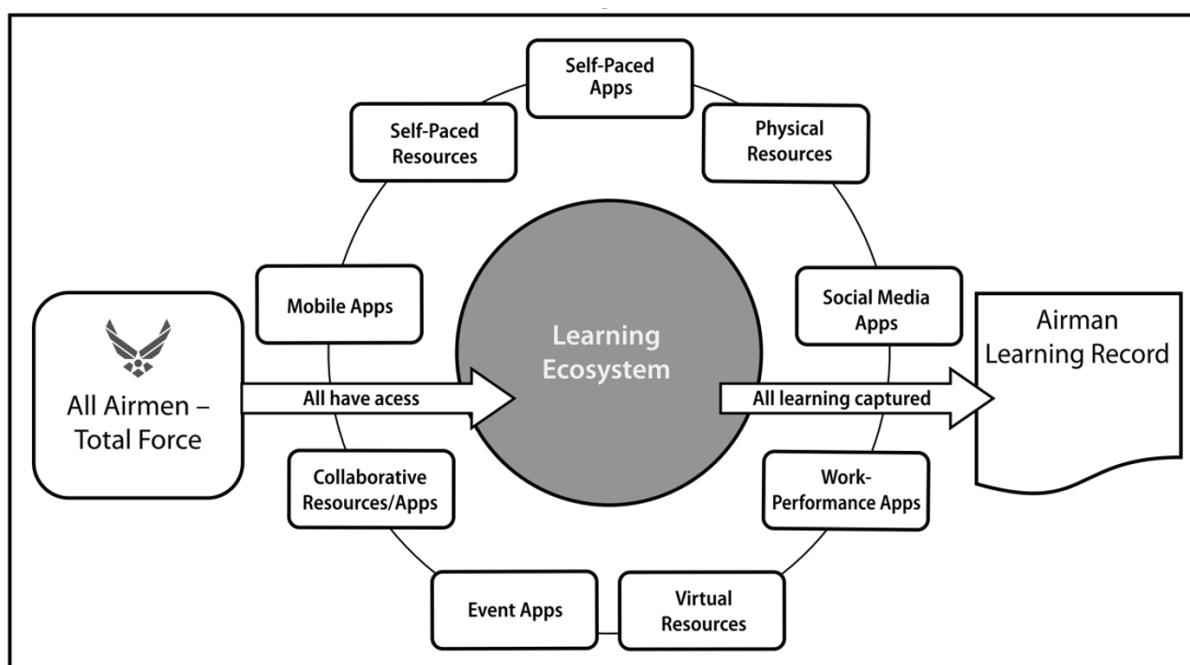
In addition to these three challenges, specific characteristics of MDO need to be considered when designing an appropriate education system. First, in the US Military, MDO is vaguely defined. It is a nebulous term that will require the community to define and redefine as understanding and practice evolve. Developing such a

definition requires space where ideas can be formulated and then evaluated in such a way that the best ones rise to the top. This characteristic also necessitates the incorporation of lessons learned from the operational communities on the front lines of Multi-Domain conflict to ensure that educational content is not divorced from real-world operations. Second, MDO education must facilitate learning by integrate differing communities effectively. MDO contributors will intrinsically have different backgrounds, which must be valued in the MDO context. However, a common operational lexicon must be established to enable communication and collaboration between various organizations and services. Therefore, unique approaches are needed to identify domain-specific and domain-agnostic terms and highlight them throughout the content. For example, each warfighter will need to learn the basic terms used in all other domains and express their domain's capabilities in a way that is understandable by all. Melding of various communities will be challenging considering that some, like Nuclear Command, Control, and Communications are not immediately considered in the MDO conversation but will need to be a part of the solution moving forward. Finally, MDO education requires a combination of theory and practice. It is not sufficient to only discuss solutions, but an environment is needed where ideas and strategies can be tested and refined. Interactive content is needed that engages participants to explore MDO in order to discover pitfalls and identify winning approaches.

These challenges must be met with innovative solutions to transform three key areas: (1) the way people think, (2) the processes used to execute defensive and offensive operations, and (3) the technology used across a wide spectrum of applications. In part, the current transformations in USAF education and training appears to be moving the USAF toward a solution.

### **3. Lifelong learners and USAF education**

Through a construct called the Continuum of Learning (CoL), AETC is transforming the way that Airmen will conduct education and training. Roberson and Stafford (2017) describe how the CoL will move learning from the classroom to where it is needed right now, where the mission is executed. This means a shift away from instructor-led learning to self-learning and online courses. According to AETC's leadership, the overall goal of these efforts is to create learning effectiveness by encouraging and supporting life-long learning. Figure 1 depicts the concept of AETC's Learning Ecosystem explained by Lt Gen Roberson and Dr. Stafford in their description of the CoL.



**Figure 1:** The USAF learning ecosystem envisioned by Roberson and Stafford (2017).

The CoL will change the way the USAF approaches education and training by providing modularized, blended, and competency-based learning that may be mandated by a training authority or accessed by the learner "on-demand". Furthermore, the Learning Ecosystem will track an individual's learning experiences serving as the centralized record of what an Airmen knows and what he or she can do. More information and explanation are provided in Roberson and Stafford (2017).

The change to focus on producing lifelong learners will advance both MDO and cyber education. However, another shift is needed away from Airmen consuming content based on mandated timelines to a Self-Directed Learning (SDL) model where individuals are empowered to take control of their own learning. This kind of shift has the potential to produce large numbers of warfighters who can both execute the mission today and adapt to overcome future challenges.

SDL, a term coined by Knowles in the 1960s, describes a method of learning that puts the responsibility for learning on the shoulders of the learner. Knowles (1975) writes, in SDL “individuals take the initiative, with or without the help of others, in diagnosing their learning needs, formulating learning goals, identifying human and material resources for learning, choosing and implementing learning strategies and evaluating learning outcomes” (As quoted in Hase and Kenyon, 2000). Hase and Kenyon (2000) take Knowles’ analysis a step further by focusing on self-determined learning across the spectrum of the education and learning lifespan. Their analysis also takes into the rapidly changing world where learning needs to be immediate and learning methods must be flexible.

MDO education needs to encourage students to become self-directed learners in order to take advantage of the enhanced collaboration and speed of innovation produced by modern technology, especially social media and the Web 2.0 revolution, typified by the speed of information exchange. This technology offers any organization the ability to effectively deliver education and training on a grand scale and create agile tools to provide rich collaboration and innovation.

#### **4. SDL in military education and training**

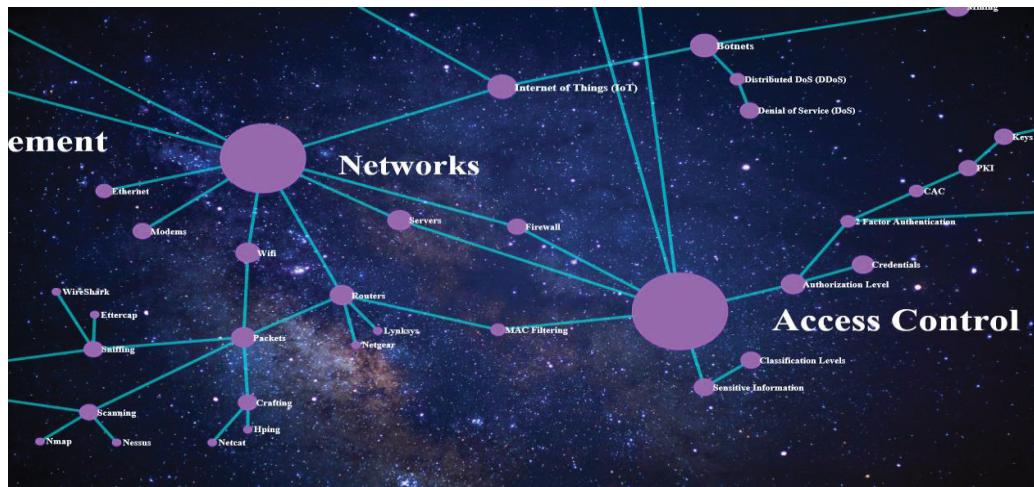
The military is facing the same challenges as the wider world stemming from digital transformation and interconnectivity which complicates military communities, work centers, and battlefields. The application of the principles of SDL should be a key consideration in growing a military force ready to operate in a 21st century environment. However, there are several aspects of SDL that will require evaluation and testing before implementation in the military. America is very proficient at training warfighters to succeed in operational contexts that are well-known and relatively predictable. Through repetition and discipline, military members are taught how to fight according to well-established tactics, techniques, and procedures. The military should not discard this training and discipline but should emphasize it throughout a warfighter’s training. In this context, SDL models will seem to work against the current culture in military environment where reproduction is highly valued. In almost all US military organizations each member can fill multiple roles so that the mission continues despite losses. This is necessary for conducting military operations in a wartime environment and requires multiple individuals to have the same baseline training to perform a given task. This is a strength of the US Armed Forces enabled by its rigorous training programs.

However, in today’s complex operational environment of MDO, warfighters will face challenges where the solution is unknown. The military must take the same standardized fighting force and train warfighters to adapt and overcome when they face never-before-seen challenges. In many cases, training alone will not be sufficient, but will require education and the ability to rapidly learn in new contexts. This requires that warfighters know *how they learn*, a key focus of SDL, so that when these new challenges arise, they know how to apply their knowledge, skills, and abilities to win. In these cases, the self-directed learner will have the awareness and abilities to form new ideas and build new combinations of military capabilities based on the situation, creating operational advantages. Therefore, as members of the military advance in their education, training, and operational experience they should be given more freedom to direct their own learning. This will create warfighters who are poised to meet both today’s and tomorrow’s challenges.

#### **5. Proposed strategy, framework and analysis**

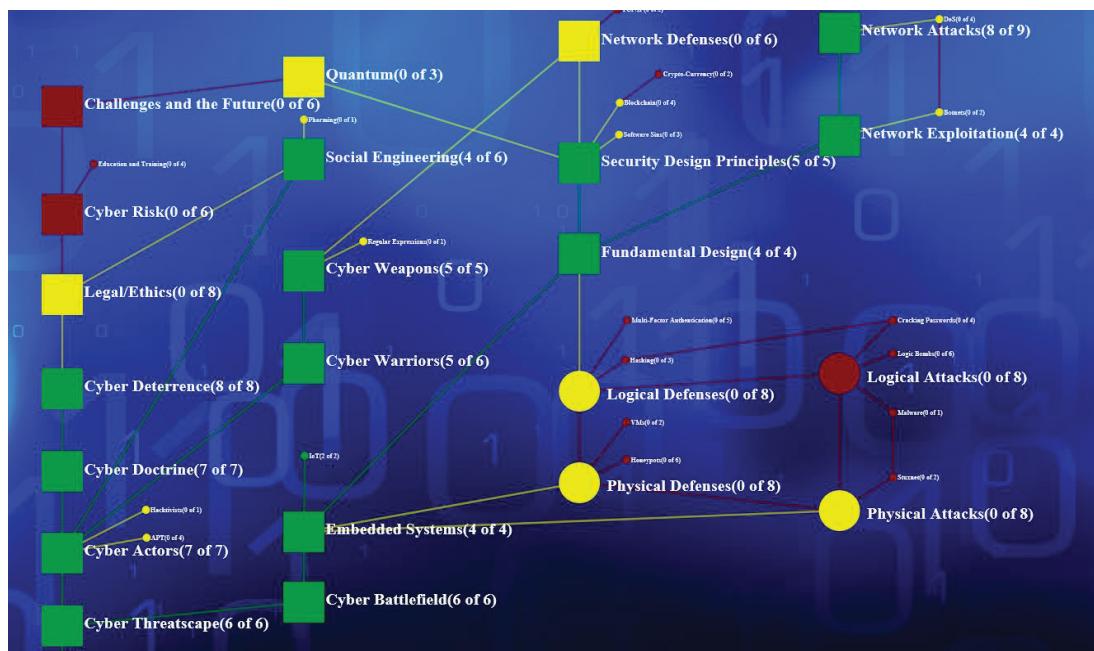
The CEH™, introduced above, implements facets of SDL in a user-centered learning platform tailored for the military environment and compatible with the AF CoL construct. Reith et al. (2018) sought to replicate content sharing sites such as YouTube, Netflix, etc. and enhance user participation by applying gamification elements, which is a foundational aspect of their design. Deterding et al. (2011) define gamification as “the use of game design elements in non-game contexts.” While current research shows leans toward the positive effect of gamification (Koivisto and Hamari, 2019), further research is needed to prove in the context of military cyber and MDO education and training. Ongoing research on the CEH™ shows positive initial results from study participants and USAF operational units (Tomcho, 2019).

The CEH™ utilizes a Topic Map to maintain the relationships among cyber concepts and tries to capture the essential elements and topics of cyberspace. Tomcho and Reith (2018) define the Topic Map as a web of cyber concepts that reveals how various topics are connected. A portion of the cyber Topic Map in the CEH™ is shown in Figure 2. A similar Topic Map would enhance MDO education by showing users where gaps exist in their current understanding while providing an easy way to fill that gap with accessible and relevant content. Users are encouraged to explore the areas of the map, with which they are not familiar. A Topic Map will also give users a graphic view of the broad MDO landscape, which should lead to a better grasp of the complexity and breadth challenges inherent in MDO.



**Figure 2:** Current cyber Topic Map in the CEH™.

Additionally, Tomcho et al. (2018) explain the use of Knowledge, Skill, and Ability Trees (KSA Trees) to present challenges and learning goals to a user and track their progress. KSA Trees, shown in Figure 3, guide learners through a subset of resources from the CEH™ with a specific goal in view. Various shapes and colors designate certain requirements and progress and some content is inaccessible until required progress is made. KSA Trees will also benefit MDO education by guiding learners through pre-selected material in a logical manner. This material could be selected and organized by operational leaders, career field managers, training professionals or MDO practitioners to create learning tracks that have proven effective for large audiences or those deemed essential by superiors. Another benefit is that KSA Trees will track a learner’s progress toward a goal and may motivate some learners to consume more content. KSA Trees will also designated the knowledge, skills, and abilities currently valued by the MDO community orienting new or inexperienced members more quickly.



**Figure 3:** A KSA Tree used as part of an AFIT graduate course

Following the CEH™ model, this paper proposes a similar education hub for MDO with a unique focus on MDC2 called the *Multi-Domain Operations Hub (MDO Hub)*. By leveraging past development, design, and research as well as current experimentation on user engagement, the new education system design can focus on elements directly tied to MDO education. This new system would deliver education and training content to warfighters across multiple services and organizations in the DoD focused on synergizing operations from the five major domains.

By design, the MDO Hub would be closely tied to the CEH™ as cyber is one of the five domains and affects or enables the communication and connectivity between the other domains. Additionally, integrating cyber operations into current C2 systems and determining how cyber acts as both a supported and supporting capability is another immediate challenge facing the DoD. However, simply creating an area in the CEH™ for MDO education will not accomplish the goal of creating a system for warfighters from all domains to interact because it may be too cyber heavy limiting involvement from other domains. Design decisions will need to balance the environment across organizations and domains removing barriers to entry for potential learners and content contributors. The current system enjoys significant input by members of the USAF and in its current design is geared for the USAF members, therefore further analysis is required to determine the best design for a DoD-wide audience.

As mentioned above, one of the challenges of MDO education is that it requires a combination of theory and practice. Education resources must move beyond books, articles, and videos to provide capabilities for users to implement the knowledge they are acquiring and practice the skills necessary to be a warfighter who can understand and implement MDO in their context. This includes hands-on, interactive, and collaborative content accessible to all users. One MDO Hub feature will be a built-in digital card game, called the MDC2 Card Game, designed to allow players to build and test strategies as well as show how capabilities from different domains can be packaged to create advantages on the battlefield. The MDC2 Card Game, designed by Lin and Reith (2018), is under evaluation to determine how best to utilize the game to teach MDO concepts. The game has many potential learning objectives. One of objectives identified by Lin and Reith (2018) is that cyber capabilities are not magic dust to be sprinkled on operations at the very end but require steps in the kill chain just like kinetic capabilities. It is designed for two players to battle each other with a hand-selected deck of 40 cards covering military capabilities from cyber, land, space, and air domains. However, enhancements could be made in the digital version to allow individuals to test strategies against automated opponents given tailored scenarios or have 3 or more individuals or teams play at once.

Other key elements of the MDO Hub will be a way to track a user's experience and engagement in the MDO community. By tracking a user's content consumption and creation, contributions, acquired skills and abilities (such as success in the MDC2 Card Game) and operational experience, the MDO Hub could be a single environment for tracking C2 experience and providing robust training. This would fulfill the need, expressed by Burdett, for MDC2 experience tracking and include much more than just years of experience or positions held but also track what the users have recently consumed and contributed to the MDO community.

## **6. Evaluation strategy and future work**

Our strategy for evaluating this proposed environment includes leveraging research conducted using the CEH™ as well as feedback from DoD personnel from all domains and those already engaged in MDO education to answer the research questions detailed below.

(1) What effect will a virtual, collaborative learning system with integrated features like a Topic Map, KSA Trees, and serious games have on user engagement, comprehension, and engagement in an MDO context? This research question will require further exploration into the current avenues of MDO education across the DoD to create a comparative experiment. A potential experiment includes using the same content for two courses given to similar participants but provide one class access to the content through the MDO Hub. This human subject research experiment would use a KSA Tree duplicating the content of the selected course and an MDO Topic Map with additional content. Data collection would focus on learner actions and engagement through environment data and participant surveys. Additional optional content consumed, number of participant comments, and amount of content added to the system could be collected for comparative analysis. One potential barrier could be the availability and selection of an existing MDO course willing to help facilitate this research.

(2) What MDO content should be provided to learners and what fields should be included in an MDO Topic Map? Future research will summarize and analyze current MDO education courses and solutions across the DoD for best practices. This research could uncover other hands-on or role-playing elements that would be beneficial to the military to produce effective MDO warfighters. Additionally, this research question would consider the subjects or tasks most appropriate for KSA Trees. Research opportunities exist to test the response from DoD personnel from multiple domains and services to various elements of the MDO Hub. User preferences, content views, the order content is consumed, and perceived usefulness would provide insight into the most effective content. Research in this area could explore the best way to integrate the MDC2 Card Game into MDO education and the perceived value of the serious game from different communities.

(3) How should the MDO Hub be organized and structured? How does industry and other militaries use SDL and organize content. This research question would explore what specific “channels” should be presented to the user on the MDO Hub landing page. Many similar commercial sites define general categories such as “Popular” and “Recently Added” but other designators should be added that are specific to MDO. How these are defined and used by the underlying system to present certain content to users will flow out of future research of current MDO training. Decisions made will affect how deep learners delve into domains other than their own operational domain. For example, how much depth does a cyber operator need to be able to understand how to integrate their capabilities with sea or land operations? This research will answer questions concerning what content should be used in an MDO system vice content that is too detailed and should be moved to a system reserved for a specific domain.

(4) What efforts are needed to bolster collaboration on solutions for defensive MDO? How does industry and other militaries approach MDO? What implications does MDO have for installation defense? The AF Chief of Staff has described the benefits of MDC2 as overwhelming enemy forces by executing operations for multiple domains at a speed at which they cannot react quickly enough. However, more research is needed to best implement MDO in the arena of military installation defense. Military establishments may face coordinated threats from multiple domains and must be ready to respond. Future research in this area could identify key transformations that bring together single-domain-focused organizations to examine defensive MDO at the tactical and operational levels. Timing of this research could prove beneficial as the USAF is in the midst of deploying mission defense teams across the force to focus on local installation and critical mission defense.

## **7. Conclusion**

In response to sustained calls for change, the USAF is currently conducting two key transformations in the areas of MDO and education and training. These are happening at a unique time in history when technology is available to effectively and efficiently gather and create education and training content to enable the development of large populations to help prepare them to adapt to future challenges. Multi-domain operations must be fueled by multi-domain education. Addressing the intersection of these two topics, this paper proposed a digital learning environment, called the MDO Hub, to begin to address the concerns and suggested the creation of an experimental prototype providing future research opportunities. The proposed system draws heavily from the framework proposed by Reith et al. surrounding the CEH™ platform. Both their system and the one proposed here utilizes concepts from SDL and successful commercial tools and should be used to shape further research and development of the AF CoL. Furthermore, the paper proposed a strategy for evaluating the MDO Hub digital learning environment through a human subjects research experiment. Finally, it proposed four possible research questions probing the most effective methods to educate an agile and effective fighting force with the mindset and experience to wage multi-domain warfare anywhere in the world.

**Disclaimer:** The views expressed are those of the authors and do not necessarily reflect the official policy or position of the Air Force, the Department of Defense, or the U.S. Government.

## **References**

- Alberts, D. S., and Hayes, R. E. (2006) *The Future of Command and Control - Understanding Command and Control*. CCRP Publication Series. Vol. 71. <https://doi.org/10.1353/jmh.2007.0051>.
- Caputo, V. J. (2017) “505<sup>th</sup> Command and Control Wing Hosts Multi-Domain C2 Joint Exercise.” AF.MIL. <https://www.505ccw.acc.af.mil/News/Article-Display/Article/1272361/505th-ccw-hosts-multi-domain-c2-joint-exercise>.

**Nathaniel Flack and Mark Reith**

- Deterding, S., Dixon, D., Khaled, R., and Nacke, L. "From Game Design Elements to Gamefulness: Defining Gamification." In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, pp. 9-15. ACM, 2011.
- Eddins, J. M., Jr. (2018) "Byte-Size Learning." Airman Magazine. May 8, 2018. Accessed December 31, 2018.  
<http://airman.dodlive.mil/2018/05/08/bytessizelearning/>.
- Goldfein, D. (2018), "US Air Force Chief of Staff: Our military must harness the potential of multidomain operations." Defense News. <https://www.defensenews.com/outlook/2018/12/10/us-air-force-chief-of-staff-our-military-must-harness-the-potential-of-multidomain-operations>.
- Goldfein, D. (2017) "CSAF Focus Area: Enhancing Multi-Domain Command and Control. . . Tying it All Together." [http://www.af.mil/Portals/1/documents/csaf/letter3/CSAF\\_Focus\\_Area\\_CoverPage.pdf](http://www.af.mil/Portals/1/documents/csaf/letter3/CSAF_Focus_Area_CoverPage.pdf)
- Harris, A. (2018) "Preparing for Multidomain Warfare: Lessons from Space/Cyber Operations." *Air & Space Power Journal* 32, no. 3: 45-61.
- Hase, S. and Kenyon, C. (2000) "From Andragogy to Heutagogy," *Ulti-BASE In-Site*.
- Knowles, M. S. (1975) *Self-Directed Learning: A Guide for Learners and Teachers*. Oxford, England: Association Press.
- Koivisto, J. and Hamari J. (2019) "The Rise of Motivational Information Systems: A Review of Gamification Research." *International Journal of Information Management* 45: 191–210. doi:10.1016/j.ijinfomgt.2018.10.013.
- Lin, A. and Reith, M. (2018) Multi-Domain Command & Control Card Game Instructions. Version 1.0. Working Paper.
- Mattis, J. (2017) "Secretary of Defense Jim Mattis' House Armed Services Committee Written Statement for the Record," [online], [http://www.politico.com/f/?id=0000015c-9f04-d0\\_70-a57d-fffe4c600001](http://www.politico.com/f/?id=0000015c-9f04-d0_70-a57d-fffe4c600001).
- Mattis, J. (2018) Summary of the 2018 National Defense Strategy of The United States of America.  
<https://dod.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf>.
- Pomerleau, M. (2018) Air Force Begins to Roll Out Special Cyber Defense Teams. Fifth Domain: Cyber.  
<https://www.fifthdomain.com/dod/air-force/2018/12/27/air-force-begins-to-roll-out-special-cyber-defense-teams>.
- Reith, M., Trias, E., Dacus, C., Martin, S., and Tomcho, L. (2018) "Rethinking USAF Cyber Education and Training," Proceedings of the 13<sup>th</sup> International Conference on Cyber Warfare and Security, pp 439-447.
- Roberson, D.L., Stafford, M.C. (2017) "The Redesigned Air Force Continuum of Learning: Rethinking Force Development for the Future," Air University Press, Curtis E. LeMay Center for Doctrine Development and Education.
- Tomcho, L., Reith, M. (2018) "Engaging Airmen with Cyber Education and Training." *Journal of The Colloquium for Information System Security Education*, Winter.
- Tomcho, L., Lin, A., Reith, M., Long, D., Coggins, D. (2018). "Applying Game Elements of Cyber eLearning: An Experimental Design." Air Force Institute of Technology.
- Tomcho, L. Personal correspondence. Used with permission. 15 January 2018.

# Detecting Advanced Persistent Threat Malware Using Machine Learning-Based Threat Hunting

Tien-Chih Lin, Cheng-Chung Guo and Chu-Sing Yang

National Cheng Kung University, Tainan, Taiwan

[dange0.tw@gmail.com](mailto:dange0.tw@gmail.com)

[jjguo@crypto.ee.ncku.edu.tw](mailto:jjguo@crypto.ee.ncku.edu.tw)

[csyang@ee.ncku.edu.tw](mailto:csyang@ee.ncku.edu.tw)

**Abstract:** Malware has always been a threat to computer users. In particular, advanced persistent threats (APTs), which target companies and organizations, often cause considerable losses to victims. Anti-virus software cannot effectively stop APTs from exploiting targets. This occurs because APTs excel at using rootkits to hide their tracks and use code obfuscation to impede efforts to analyze their systems. Moreover, APTs customize malware for every victim. Therefore, signature-based anti-virus software cannot detect malware that is not known from earlier information. Rather than passively waiting for an attack and then extracting the signature of malware, threat hunting, which is an active defensive concept, should be used. The system proposed in this paper focuses on threat hunting, which detects a threat at an early stage and enables immediate response. In addition, we used dynamic analysis to understand the behavior of the process and used machine learning to classify malicious behavior against the user. By using XGBoost as the classifier, ten-fold cross-validation yielded an f1 score higher than 0.99 and the classifier could successfully classify real-world malware programs.

---

**Keywords:** threat hunting, machine learning, event correlation, malware detection, advanced persistent threats

## 1. Introduction

Advanced Persistent Threats (APT) have threatened businesses and organizations for many years. These complex and multistage attacks on specific organizations have long incubation periods and whenever possible they deliberately hide their tracks (Navarro, Deruyver, Parrend, 2018). In addition, fileless malware (often used in APT) is a technique that disguises a code or script as data and uses a legitimate program such as PowerShell or cmd to execute scripts (Velasco, Duijn, 2018). In a normal system, these PowerShell functions are regularly called. Therefore, it is difficult to determine if a PowerShell script is legal or malicious.

Techniques used to detect malware can be broadly classified into two categories: signature-based detection and anomaly-based detection (Idika, Mathur, 2007). Signature-based detection (Sahu, Ahirwar, Hemlata, 2014), which involves using anti-virus with approaches such as the YARA rule ("YARA - The pattern matching swiss knife for malware researchers"), exhibits considerable accuracy in detecting known malware. However, fileless malware can easily bypass such detection. Furthermore, detection of malware that has not been reported earlier is difficult (Idika, Mathur, 2007). In case of an APT attack, the attacker usually creates new malware or modifies existing malware. Therefore, signature-based detection cannot protect the victim. To resolve such problems, it is necessary to make changes in detection approaches. Anomaly-based detection (Idika, Mathur, 2007) is divided into learning and detection phases. Normal behavior is observed during the learning phase and abnormal behavior is detected during the detection phase. The ultimate goal of the malware remains the same; malware programs use several techniques to confuse anti-virus software. For example, ransomware (Al-rimy, Maarof, Shaid, 2018) encrypts files to facilitate extortion of the victim, and a backdoor maintains a secret tunnel to the attacker's command and control (c&c) servers. Although several obfuscation techniques exist, the behavior of all malware programs is similar. Through anomaly-based detection, 0-day attack methods can be devised to protect the victim from APT attacks. This study used anomaly-based detection to classify the behavior of the program.

To strengthen network security, the IT industry has proposed many concepts, such as traditional passive defense and active defense. Passive defense involves passively waiting for events to occur and then providing incident response. Active defense involves active detection of threats. Cyber threat hunting has been promoted in recent years for developing active defense strategies. The main objectives are to defend against internal network threats, perform related search and identification, and enhance the understanding of hostile forces.

The system proposed in this study was based on threat hunting and was divided into three steps. The first step involved generating logs on the endpoints by actively using Sysmon (Russinovich, Garnier) and sending them to the monitoring platform. Sysmon records process behaviors during its runtime and can be considered as a type

of dynamic analysis. Compared with static analysis, dynamic analysis can accurately record the behavior of the process, and is not affected by code obfuscation. In the second step, we correlated the logs to reconstruct the behavior of the process. Most malware programs distribute functions, such as “drop file,” “upload file,” and “connect with c&c,” to avoid being detected by anti-virus programs. In this step, we correlated the behaviors to enable a comprehensive understanding of events. The third step involved classifying the reconstructed process behaviors by using a machine learning algorithm to determine malicious processes. The proposed method can be used to observe behaviors that cannot be observed through static analysis and to classify variant malware that cannot be classified using signature-based detection.

The paper is structured as follows. Section 2 summarizes related studies in the field. Section 3 presents the method of our system. Then, we validate our system and present the results in Section 4. Finally, Section 5 is the conclusion of the paper.

## **2. Related studies**

Malware detection is divided into two categories: signature-based detection and anomaly-based detection. The disadvantages of signature-based detection have been presented in the Introduction. In this study, we focused on anomaly-based detection. Several studies have focused on detection of malware by incorporating machine learning and other data mining techniques. These assessments are anomaly analysis and can be further divided into static analysis and dynamic analysis.

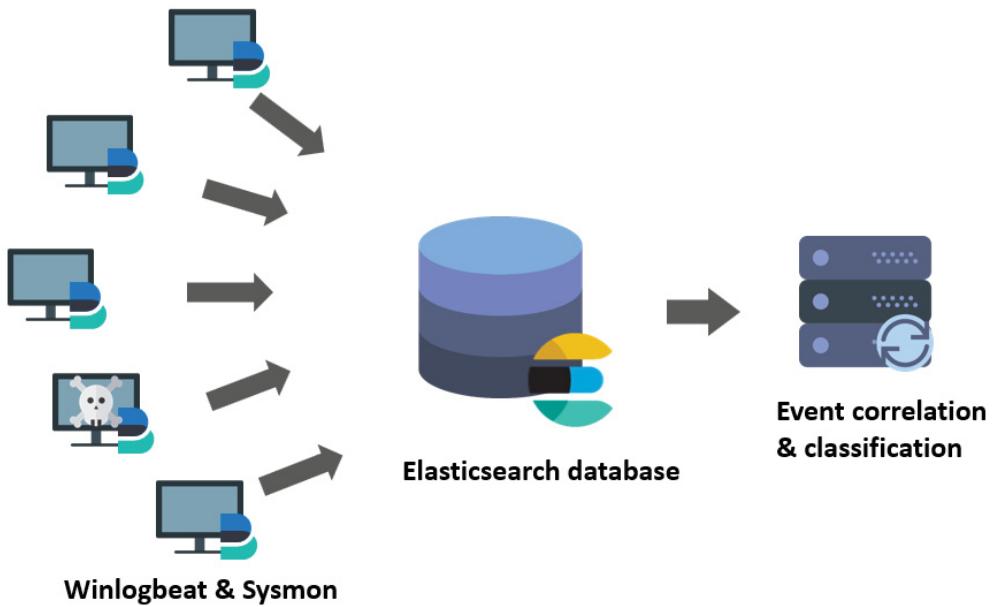
In static analysis, researchers only focus on the static nature of binary executable files, such as PE header and byte sequences. In (Lyda, Hamrock, 2007), experiments revealed that most malware programs are encrypted and exhibit packed behavior, and these behaviors resulted in higher entropy than that of a benign software. (Reddy, Dash, Pujari, 2006), (Reddy, Pujari, 2006), (Santos et al, 2009) and (Bolton, Anderson-Cook, 2017) have used the n-gram method to determine special sequences of instructions and assembly code. They have used other machine learning classification algorithms to classify benign and malicious software. However, (Moser, Kruegel, Kirda, 2007) presents a disadvantage of static analysis. First, when the malware uses code obfuscating techniques, static analysis cannot analyze malware action. Second, commercial software organizations encrypt and pack executable files because of copyright concerns, therefore, not all encrypted or packed files can be categorized as malware. Dynamic analysis can compensate for these shortcomings.

In dynamic analysis, researchers focus on malware behavior. Malware is tested in a special environment and its action after execution is observed. (Anderson et al, 2011) used special hardware called Ether to collect runtime instructions for malware and represented it by using Markov chains. Then, support vector machines were used to calculate the distance of various Markov chain kernels and achieve a classification effect. (Willems, Holz, Freiling, 2007), (Shukla, 2008), (Yoshioka, Matsumoto, 2010), (Blasing et al, 2010) and (Guarnieri, Tanasi, Bremer) have used sandbox technology to execute malware in secure and independent environments to observe the behavior of malicious programs safely. However, (Bayer et al, 2009), (Harris, Stutz, Lynch, 2017) and (Branco, Barbosa, Neto, 2012) have indicated that most malware programs stop or change behavior if any sandbox or virtual machine execution is detected.

Dynamic analysis that relies on a particular environment requires that malware samples be sent to a specific environment for execution and observed. However, in practice, it is impossible to perform environment analysis on malware samples without executing the file every time, which is not feasible in terms of performance and efficiency. Therefore, in this paper, an agent will be installed on the endpoint to collect real-time event logs and transfer them to a back-end database. When malware is executed in a real-world environment, the back-end machine uses machine learning techniques to clarify the attack and ensure the IT department immediately responds.

## **3. Method**

In this section, we describe the design of our system. The system collects the event logs from endpoint devices and classifies them at the back-end classifier. The system architecture is presented in Figure 1, and we can divide the system into two components.

**Figure 1:** System architecture

The first component of the system is log collection. Winlogbeat ("Winlogbeat: Analyze Windows Event Logs | Elastic") and Sysmon (Russinovich, Garnier) were installed at endpoint devices. Sysmon is a system monitor which monitors and logs system activity to the Windows event log. Sysmon contains information about process creation, network connections, and other sensitive process activities, as listed in Table 1. After Sysmon logs an event, Winlogbeat sends the event data to Elasticsearch ("Open Source Search & Analytics · Elasticsearch"). Elasticsearch is a search engine that can store, search, and analyze large volumes of data. Our scenario consisted of several types of events with different information formats. For example, at the process creation event, Sysmon logs attributes such as the process' image and MD5. However, at the network creation event, Sysmon logs attributes such as the IP of source and destination. The high flexibility and high expandability of Elasticsearch render it best choice for our study.

The second component consisted of event correlation and classification. The input of this component consists of the logs stored on Elasticsearch. Each log generated by Sysmon and stored in Elasticsearch is independent. For example, a process A generates several logs by using Sysmon. Although those logs are generated by process A, they do not affect other logs and are independent. However, each log cannot be viewed separately because the behavior of these log records is typical for any benign process, and, therefore, logs should be correlated. After correlating the logs, we can infer the complete behavior of the process and assess whether it is a malicious process. The workflow is depicted in Figure 2. Each module is described in subsequent sections.

**Table 1:** Sysmon events

Event ID	Tag
Event ID 1	Process creation
Event ID 2	A process changed a file creation time
Event ID 3	Network connection
Event ID 4	Sysmon service state changed
Event ID 5	Process terminated
Event ID 6	Driver loaded
Event ID 7	Image loaded
Event ID 8	CreateRemoteThread
Event ID 9	RawAccessRead
Event ID 10	ProcessAccess
Event ID 11	FileCreate
Event ID 12	RegistryEvent (Object create and delete)
Event ID 13	RegistryEvent (Value Set)
Event ID 14	RegistryEvent (Key and Value Rename)
Event ID 15	FileCreateStreamHash

Event ID	Tag
Event ID 17	PipeEvent (Pipe Created)
Event ID 18	PipeEvent (Pipe Connected)
Event ID 19	WmiEvent (WmiEventFilter activity detected)
Event ID 20	WmiEvent (WmiEventConsumer activity detected)
Event ID 21	WmiEvent(WmiEventConsumerToFilter activity detected)

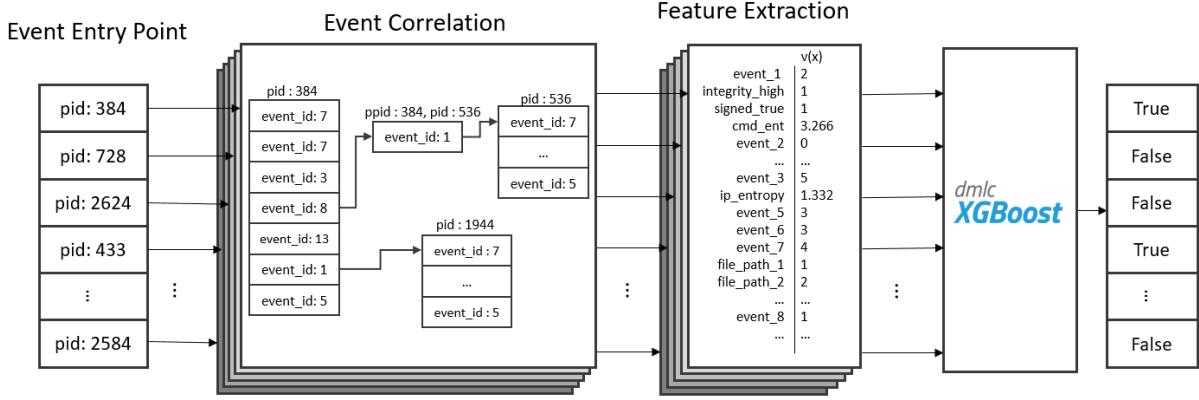


Figure 2: Workflow of event correlation and classification

### 3.1 Event entry point

The process identifier (pid) and globally unique identifier (guid) of every event are recorded in Sysmon event log to record information about which process created an event. The process creation event is the first Sysmon event for every process in the operating system. Therefore, all the process creation event logs should be collected. Furthermore, we collected each pid and guid for each event entry point. The process creation event logs can be collected by querying all the event logs of the “process creation” type from Elasticsearch.

### 3.2 Event correlation

Before we correlated event logs, all event types were classified into two categories: correlation events or information events. Event logs with record relationship information were categorized in the correlation event class. For example, a process creation event log that recorded a pid and a parent process identifier (ppid) in its log could be a correlation event. Other events (such as the image loading event which records the image to be loaded) could be considered as information events. The correlation event class is presented in Table 2, and the events not shown in Table 2 are information events.

Table 2: Event correlate rules

Event ID	Tag	correlation condition
Event ID 1	Process creation	pid
Event ID 8	CreateRemoteThread	pid
Event ID 10	ProcessAccess	pid
Event ID 11	FileCreate	image name
Event ID 15	FileCreateStreamHash	image name

The event entry points obtained in the previous step were used to track and correlate. First, we used pid and guid to query the event log and then obtained information about all the event logs created using this process. If an event log belonged to a correlation event class, then we used the event to determine the next event. For example, by tracing process A we realized that a process creation event occurred. This creation event indicated that process A created another process (called process B here). We defined process A to be the parent process, which corresponded to ppid in this event log; process B as the child process, which corresponded to pid in this event log. Thus, we used the pid of process B for tracing its event log and correlated it with process A.

Finally, each event entry point was assigned to a process tree, as depicted in Figure 2. This process tree indicated the behavior of the original process and became the input to the next module.

### 3.3 Feature extraction

After building the process tree, we started extracting features from the process tree. The main features extracted from the tree are presented in Table 3, which is divided into three categories.

**Table 3:** Feature table

Feature	Feature type
event_1	Event counter
integrity_high	Event counter
signed_true	Event counter
cmd_ent	Entropy value
event_2	Event counter
...	
event_3	Event counter
dst_ent	Entropy value
event_5	Event counter
event_6	Event counter
event_7	Event counter
file_path_1	Important file path
file_path_2	Important file path
...	
event_8	Event counter
event_9	Event counter
event_10	Event counter
event_11	Event counter
key_path_1	Important file path
...	

#### 3.3.1 Event counter

The first category was the event counter. Sysmon consists of 22 event types. We calculated the occurrences of these 22 event types from the process tree we obtained in the previous step. For example, if a process generates four process creation event logs during its life cycle, then in its feature table, the value of process creation event counter is 4. We used event counters because various valid processes or malware processes have different event counter distributions. For example, a ransomware program may scan all files on the disk and encrypt them. Thus, in this case, a large number of FileCreate events can occur.

Besides the 22 different event types, there are also other important information records in the log will be counting occur times. Such as integrity\_high and signed\_true are the information record in Process creation event, both of them are very important information, so there will be isolated features to counting the number of the occurrence of these events.

#### 3.3.2 Entropy value

The second category is entropy calculation. This feature consists of two scenarios; the first determines whether the command line message recorded by the process creation event uses code obfuscation, and the second determines whether the network connection event has a horizontal scan.

In the first scenario, we used the command line information recorded in the process creation event. A command line input is created when a process creation event occurs, and this command line reflects the behavior of the process. This information is useful for determining an abnormal operation. In most fileless malware programs, the command line is encrypted with code obfuscation. This results in an abnormally high entropy when the command line executable is parsed as a character string. For example, the command line executable in Figure 3 was extracted from real-world malware. Most of the commands were encoded using base-64 to evade anti-virus checks. The entropy of this command line executable was 4.0733, which was higher than that in normal software.

```
C:\Users\...\befjhdbef.exe: 6|8|7|6|6|0|7|9|6|9|4
KUhCPjcyLS4rLxwpS048Skk9NC0XK0g9TVFJUKRAPjQsGic9Q01UQjs3KiwuMTAAkUNCOz
cnHC1IS0k
+VTxLWUBAnyoyMSw5GSZNPE5QPUTZT1JFNGJrcGoyKC1tcsm81PjxPRSVNSUotOkdKJUVIP
kgaKUNFQD1CRT41GSk+MTYrLy4wKygrGileKzQrKzMqGCg
+LT0mKBomQC41JisaLz0sNyQtGidJTE1ETjpOVkxMQU87PVk2Fy1HTkk8Tj10Xz5MRjg5G
idJTE1ETjpOVko7RT43QmltZftlISwoSWxpJSsnLSUzKDEeLCpCbmBmWlyEsKCstMiUrJyJ
vLxonP1I/X05JRjRkbmxpNCkvXSgwZ18rLV9sZHQuLWw1X2Znb15gcmh1biVqX2wobV90b
mdZaXEoXXFFGi8
+Tz9WP0c8RUNIRTYXKT9LTUXYPExPUEo/STkqGChOQkFHQ1NGUV1NS0Y3IChNRzQuGic9T
Ss9GSZMTEpQQUY/WVc
+Qz1GST9BRjtBRU5JRjQcKUFMWUxVR0tDREE3bGtvXyAoST9LUUxGQkhBX05KP01bPj1ST
TcyGSZCQEa/UDYrG19CS1k7Vug5RkM9Xz5FPU1VSkw
+PjdmlWmNtXBwpPEhRSExI0D5WRUo1KjEwLiorKCguKyYsMBovSThMOEhGPUVZQ05MSztDS
Ddsal29fIChLQ0RBNyktMCoxKy8sLDmAJz1JUU5FRjs7W05BRj83NCgvKSksKy4jLTQxKTE
vLiZKRQ==
```

**Figure 3:** Example of command line information

In the second scenario, we used the source and destination IP records of the network connection event. This feature detects horizontal scans because horizontal scans are the best method to scan intranet architecture or spread a virus. Most malware programs attempt these scans. The method of calculation of entropy is as follows: First, we collected the source IP and destination IP of the network behavior generated by the program. Then, we extracted the network IDs of these IPs with the mask “255.255.255.0”. For example, when the mask is “255.255.255.0”, the network IDs of “192.168.0.1” and “192.168.0.100” are “192.168.0”. We obtained entropy values from these network IDs. For example Table 4 depicts two IP access lists. The first list is a classic example of horizontal scanning; the entropy of the network ID is 0.4101. The next list is an example of a normal network connection behavior; the entropy of this network ID is 1.3322. This explains the difference between a normal network connection and a horizontal scan.

**Table 4:** Example of network behavior

Behavior	IP access list
Horizontal scanning	‘192.168.0.1’, ‘192.168.0.2’, ‘192.168.0.3’, ‘192.168.0.5’, ‘192.168.0.7’, ‘192.168.0.9’, ‘8.8.8.8’
Normal using	‘52.197.231.80’, ‘172.217.19.206’, ‘216.58.212.163’, ‘172.217.168.238’, ‘172.217.19.206’

### 3.3.3 Important file path counter

The final category involves counting the number of visits to particular file paths. These file paths consisted of the file paths such as “C:\windows” and a few registry key paths . We discovered that a few malware programs load or scan files were located at specific file paths. We observed that some malware programs used the tools under the path “C:\Windows\system32\wbem”. These tools are provided by Microsoft, and have powerful functions in the computer. These tools could be meaningful features, so we compiled statistics regarding the file paths that malware often visited. Therefore, we detected file paths that malware often visited, then selected the top ten most frequently visited file paths and registry key paths in each event types as features.

## 3.4 XGBoost

(Chen, Guestrin, 2016) proposed that XGBoost, also known as “Extreme Gradient Boosting”, can be used for supervised learning problems. XGBoost is derived from gradient boosted trees. Therefore, the algorithm can be used for decision tree ensembles. This algorithm constructs trees for classifying one input because a single tree is not sufficient. Therefore, the ensembles model calculates the sum of prediction from every output tree. In this study, we used it as a binary classifier. The input was the feature vector, the output was whether the process was malicious or benign.

## 4. Result

In this section, we demonstrated experimental validation of the proposed method. We implemented our system in Python. The XGBoost algorithm was implemented using the Python XGBoost module (“XGBoost Documentation — xgboost 0.81 documentation”) and verified using the sklearn module (“scikit-learn: machine learning in Python — scikit-learn 0.20.2 documentation”).

Our victim environment was built on a virtual machine with Windows 7 Service Pack 1 64-bit operating system equipped with Intel Core i7-3770 3.40 GHz CPU, 1.46GB RAM. The Elasticsearch database, event correlation, and classification server were built on a Ubuntu 16.04 64-bit operating system equipped with Intel Core i7-7700, 3.60 GHz CPU, and 16GB RAM.

#### 4.1 Experimental data

Experimental data was generated by using 3,265 different malware programs. This malware was detected in 2014–2017, and included adware, backdoors, bots, and ransomware, etc. Each malware program was executed on the victim environment for 3 mins; Sysmon generated event logs that were sent by using Winlogbeat.

In addition to observing malware behavior, we collected benign software event logs from normal usage. The collection target included unzip tools, communication tools, and Windows system daemons, etc.

#### 4.2 Ground truth validation

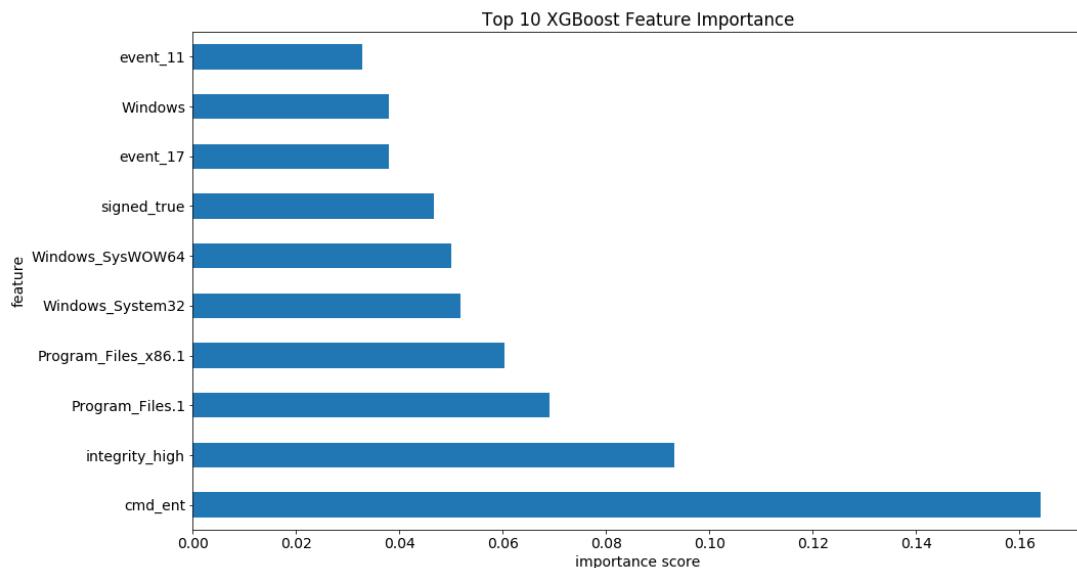
We used ten-fold cross-validation to test the robustness of our system. The validation divided the entire data into ten parts, iteratively selected nine parts for training, and the assigned one part for testing. The comparison with different models is shown as Table 5. There are four hidden layers in Neural Network, each of them has 128 nodes with ReLu as the activation function. Show as result, XGBoost gets the highest score in all model. The scores are the mean score of the evaluation method by 10 fold cross-validation.

**Table 5:** Result of evaluation

Evaluation method\Model	XGBoost	SVM	Decision Trees	Neural Network
Accuracy	0.98 (+/- 0.02)	0.90 (+/- 0.06)	0.97 (+/- 0.04)	0.96 (+/- 0.08)
Precision	0.98 (+/- 0.02)	0.89 (+/- 0.06)	0.98 (+/- 0.03)	0.97 (+/- 0.06)
Recall	0.99 (+/- 0.01)	0.98 (+/- 0.02)	0.98 (+/- 0.02)	0.96 (+/- 0.08)
F1 score	0.99 (+/- 0.01)	0.93 (+/- 0.03)	0.98 (+/- 0.02)	0.94 (+/- 0.16)
ROC_auc	1.00 (+/- 0.01)	0.91 (+/- 0.10)	0.97 (+/- 0.04)	0.97 (+/- 0.06)

#### 4.3 Feature importance

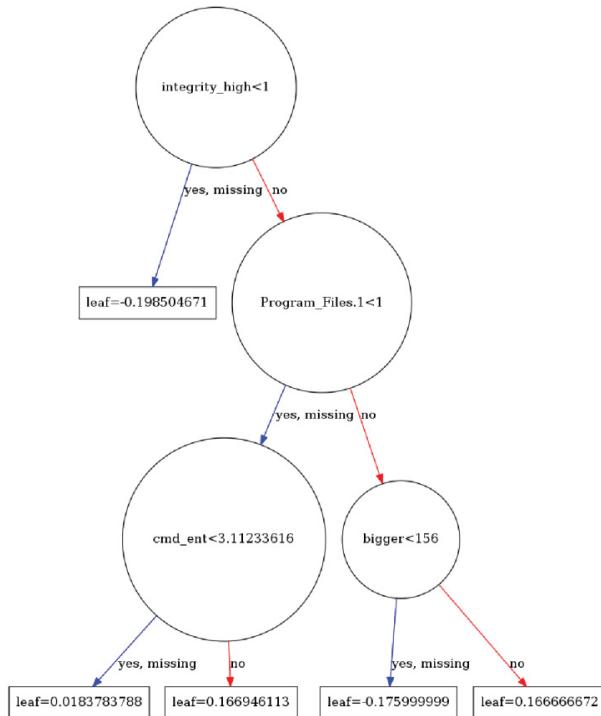
XGBoost is one of the decision tree ensembles algorithm. The principle of a decision tree is to use this feature as a condition to separate data from different labels. And there is a way to measure the importance of this feature, that is, after separation by features, the higher the purity of the data, the more important of the features. Then, the feature importance in XGBoost will be the averaged feature importance across all the decision trees in the XGBoost model. Figure 4 shows the feature importance of XGBoost model. As the figure shows, cmd\_ent is one of the most important feature.



**Figure 4:** Top 10 XGBoost feature importance

#### 4.4 XGBoost feature importance

Figure 5 shows the structure of one of the decision trees in XGBoost model. Compared to the black-box, it is much easier to know the decision rules in XGBoost, and it is helpful to determine whether the classification is robust. Each decision tree will output a score at the leaf node, and the total output will be the sum of every leaves score.



**Figure 5:** One of decision trees in XGBoost

## 5. Conclusion

The main contribution of this paper is that we designed a system that can classify malicious and benign software in real time. We leverages Sysmon and Winlogbeat to generate event logs and transfer the logs to the database. The performance costs are lightweight at the endpoint computers. Then we can classify these logs on a back-end server without affecting performance of the endpoint computers. At the back-end, we proposed an event correlation method and extracted features after correlation. Then, we classified the feature vector by XGBoost and obtained F1 score greater than 0.99 in the 10-fold cross-validation.

## Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions on the paper.

This work was supported in part by the Ministry of Science and Technology of Taiwan, under Contracts MOST 108-2218-E-006-035.

## References

- Al-rimy, B. A. S., Maarof, M. A. & Shaid, S. Z. M. (2018), "Ransomware threat success factors, taxonomy, and countermeasures: A survey and research directions", *Computers & Security* 74, 144–166.
- Anderson, B., Quist, D., Neil, J., Storlie, C. & Lane, T. (2011), "Graph-based malware detection using dynamic analysis", *Journal in computer Virology* 7(4), 247–258.
- Bayer, U., Habibi, I., Balzarotti, D., Kirda, E. & Kruegel, C. (2009), "A view on current malware behaviors", in "LEET".
- Harris, M. D., Stutz, D. & Lynch, V. K. (2017), "Mitigation of anti-sandbox malware techniques". US Patent App. 14/929,851.
- Blasing, T., Batyuk, L., Schmidt, A.-D., Camtepe, S. A. & Albayrak, S. (2010), "An android application sandbox system for suspicious software detection", in "2010 5th International Conference on Malicious and Unwanted Software(MALWARE 2010)", IEEE, pp. 55–62.
- Bolton, A. D. & Anderson-Cook, C. M. (2017), "Apt malware static trace analysis through bigrams and graph edit distance", *Statistical Analysis and DataMining: The ASA Data Science Journal* 10(3), 182–193. .
- Branco, R. R., Barbosa, G. N. & Neto, P. D. (2012), "Scientific but not academic overview of malware anti-debugging, anti-disassembly and anti-vmtechnologies", *Black Hat*.

- Chen, T. & Guestrin, C. (2016), "Xgboost: A scalable tree boosting system", in "Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining", ACM, pp. 785–794.
- Guarnieri, C., Tanasi, A. & Bremer, J., "Cuckoo Sandbox - Automated Malware Analysis", Cuckoosandbox.org, (2018). [Online]. Available: <https://cuckoosandbox.org/>. [Accessed: 21- Jan- 2019].
- Idika, N. & Mathur, A. P. (2007), "A survey of malware detection techniques", Purdue University48.
- Velasco, L. & Duijn, R. (2018). "Fileless-Threats-Analysis-and-Detection", Dearbytes.com [Online]. Available: <https://www.dearbytes.com/wp-content/uploads/2018/08/Fileless-Threats-Analysis-and-Detection.pdf>. [Accessed: 28- Dec- 2018].
- Lyda, R. & Hamrock, J. (2007), "Using entropy analysis to find encrypted and packed malware", IEEE Security & Privacy5(2).
- Moser, A., Kruegel, C. & Kirda, E. (2007), "Limits of static analysis for malware detection", in "Computer security applications conference, 2007. Twenty-third annual", IEEE, pp. 421–430.
- Russinovich, M. & Garnier, T. "Sysmon - Windows Sysinternals", Docs.microsoft.com, (2018). [Online]. Available: <https://docs.microsoft.com/en-us/sysinternals/downloads/sysmon>. [Accessed: 30- Dec- 2018].
- Navarro, J., Deruyver, A. & Parrend, P. (2018), "A systematic survey on multi-step attack detection", Computers & Security. "Open Source Search & Analytics · Elasticsearch", Elastic.co, 2018. [Online]. Available: <https://www.elastic.co/>. [Accessed: 21- Jan- 2019].
- Reddy, D. K. S., Dash, S. K. & Pujari, A. K. (2006), "New malicious codedetection using variable length n-grams", in "International Conference on In-formation Systems Security", Springer, pp. 276–288.
- Reddy, D. K. S. & Pujari, A. K. (2006), "N-gram analysis for computer virusdetection", Journal in Computer Virology2(3), 231–239.
- Sahu, M. K., Ahirwar, M. & Hemlata, A. (2014), "A review of malware detectionbased on pattern matching technique", International Journal of Computer Science and Information Technologies5(1), 944–947.
- Santos, I., Penya, Y. K., Devesa, J. & Bringas, P. G. (2009), "N-grams-basedfile signatures for malware detection.", ICEIS (2)9, 317–320.
- "scikit-learn: machine learning in Python — scikit-learn 0.20.2 documentation", Scikit-learn.org. [Online]. Available: <https://scikit-learn.org/stable/>. [Accessed: 21- Jan- 2019].
- Shukla, J. (2008), "Application sandbox to detect, remove, and prevent mal-ware". US Patent App. 11/769,297.
- Yoshioka, K. & Matsumoto, T. (2010), "Multi-pass malware sandbox analysiswith controlled internet connection", IEICE transactions on fundamentals of electronics, communications and computer sciences93(1), 210–218.
- Willems, C., Holz, T. & Freiling, F. (2007), "Toward automated dynamic mal-ware analysis using cwsandbox", IEEE Security & Privacy5(2).
- "Winlogbeat: Analyze Windows Event Logs | Elastic", Elastic.co, 2018. [Online]. Available: <https://www.elastic.co/products/beats/winlogbeat>. [Accessed: 21- Jan- 2019].
- "XGBoost Documentation — xgboost 0.81 documentation", Xgboost.readthedocs.io. [Online]. Available: <https://xgboost.readthedocs.io/en/latest/>. [Accessed: 21- Jan- 2019].
- "YARA - The pattern matching swiss knife for malware researchers", Virustotal.github.io, (2018). [Online]. Available: <https://virustotal.github.io/yara/>. [Accessed: 28- Dec- 2018].

# A Standardization Guide for Homomorphic Encryption Applications for Digital Forensic Investigations

Nnana Mogano

Department of Computer Science, Faculty of Engineering, Built–Environment and Information Technology, University of Pretoria, South Africa

[nnana.mogano@gmail.com](mailto:nnana.mogano@gmail.com)

**Abstract:** In recent years, cloud computing has emerged as an important paradigm in the computing environment due to its many advantages both to individuals as well as corporates. Some of these advantages include cost efficiencies, increased storage space and many other benefits associated with handing over responsibility to a third party for access to data and information as and when needed. The recent advancement in technology has resulted in the growing need to perform more functions, either than storage, in the cloud whilst maintaining the security and functionality aspect of cloud computing. Homomorphic encryption has been hailed as the holy grail of modern computing, it allows specific types of computations to be carried out on the encrypted data without having to decrypt beforehand as is the case with regular encryption methods. In this case, the client would need to provide the private key to the server for decryption prior to a desired manipulation of data, which might affect the confidentiality and privacy of the stored data. Recent literature highlights the importance of a homomorphic encryption scheme that is able to maintain security and still remain efficient in performing tasks. There are numerous homomorphic cryptosystems that have been proposed but not all have been proven to be practical in specific real-life scenarios. Although there have been few attempts, the lack of standardization has led to many experts in the field merely providing an intuitive approach to the implementation of homomorphic encryption for specified applications as a result attracting some resistance in the information technology industry. Also, this gap has brought about concerns regarding data and information security thus hindering the wide adoption of homomorphic encryption. This paper aims to provide a guide to the implementation of the existing homomorphic encryption schemes across various applications to enable a digital forensic investigation process to follow in order to address the highlighted concerns in a manner that can be measurable and mitigated. A matter analysis of empirical studies available on these applications will be presented.

**Keywords:** homomorphic encryption (HE), cloud security, cryptosystems, fully homomorphic encryption (FHE), digital forensic investigation, digital forensic evidence

---

## 1. Introduction

The rapidly growing requirement for secure, convenient and efficient technology has led to the optimization of conventional encryption schemes. Most importantly, this requirement highlights the need to perform various functions on data in a manner that ensures privacy. Following the invention of public key cryptography by Diffie and Hellman (Diffie and Hellman, 1976) numerous researchers in the field have strived to propose a cryptosystem that is able to process data in the cloud without having to decrypt it. It was not until 1978 when Rivest, Adleman and Dertouzos (1978) introduced the term homomorphism in addressing the concept of computing on encrypted data. Homomorphism, in mathematics, refers to the mapping of one data set into another while preserving relationships between elements in both sets. In cryptography, the data for homomorphic encryption scheme retains the same structure when given inputs to perform a function over a domain. Homomorphic encryption introduces a mechanism that allows the computation of arbitrary functions over encrypted data without the use of the decryption key. The mechanism, unlike conventional encryption methods, ensures that the data is protected not only while in transit but also offers protection of the computation of the data. The data in its simplest form is referred to as the plaintext. The data or the plaintext in encrypted form, called the ciphertext, is where the computation is carried out. Typical user functions such as storage, retrieval, searching and processing are carried out in the cloud server environment without compromising on the security. In the context of homomorphic encryption, these functions are reduced to addition and multiplication.

Homomorphic encryption as an optimization to conventional encryption methods offers a strengthened security approach due to its functionality of computing on encrypted data. Although most of the existing homomorphic encryption schemes are efficient in performing one function or the other, there is yet still to exist a “one-size-fits-all” scheme that is able to perform arbitrary function in the most efficient manner possible. While more secure than conventional encryption methods, the question on the practical aspect of these cryptosystems has led to their slow paced adoption in the global arena within the field. Challenges such as limitations with regards to the operations and number of iterations for such operations has contributed to the hindrance towards wide adoption of this advanced technology. Furthermore, optimization of some existing schemes are rather

impractical due to the time they take to compute functions. Another major challenge is that there is no standard on the implementation for homomorphic encryption making it difficult for alignment and accountability within the industry.

A lot of business owners and investors are concerned about the security aspects associated with the use of the cloud. Security in homomorphic encryption becomes compromised as a result of the ciphertext growth as the depth of the circuit increases. In order to influence the wide adoption of homomorphic encryption, the capabilities and the limitations of the existing homomorphic encryption schemes have to be outlined in order to identify opportunities for standardization. This paper is structured as follows: Section 2 delves into the importance of homomorphic encryption by providing a snapshot of incidents from serious security breach cases involving some well-known companies followed by an explanation of what homomorphic encryption is in Section 3. Section 4 of this paper goes into detail regarding similar research to the proposed effort. Section 5 details the proposed model. Finally, Section 6 draws conclusion as well as highlights opportunities for further digital forensic investigation tools and procedures as research in future in a more detailed manner for a practical fully homomorphic encryption scheme.

## **2. The importance of homomorphic encryption**

Homomorphic encryption has been heralded as the Holy Grail in modern technology. The technology provides a functionality that allows computations to be performed on encrypted data without the need to decrypt the data first. Various homomorphic encryption schemes have been introduced that are able to provide this functionality to a limited extent. The true breakthrough came in 2009 when Craig Gentry, for his thesis (Gentry, 2009), proposed a different type of homomorphic encryption scheme – one that is able to compute arbitrary functions on encrypted data and unlimited number of operations – called fully homomorphic encryption. The need for such a scheme had been evidenced by the security related issues synonymous with the use of the cloud.

On the 4<sup>th</sup> of February 2015 (Khan et al., 2016), a healthcare insurance company, Anthem, Inc. disclosed that criminal hackers had broken into its servers and potentially stolen over 37.5 million records that contain personally identifiable information from its servers, even worse other related companies were affected. Digital forensics investigators assigned to the case found that the hackers had smuggled data out in a cloud-based file sharing service. Regular data protection methods such as firewalls and other related security measures were inefficient due to the complexities of the cloud such as the distributed nature of the infrastructure, that is, its scalability. Pandi et al. (2018) outlines other related challenges regards to the cloud environment. Even though Anthem Inc. was not required by the law to have their data encrypted, the company suffered major losses in terms of revenue within this period including reputational damage to their image. In the scenario, the reason for the unencrypted data was mainly for marketing purposes and the company felt that conventional encryption alone still would not have ensured that their clients' data was secured. It is little that the company did not accept any accountability on the matter. The frequent use of Anthem Inc.'s data within the company as well as their data exchanges with other companies meant that data was continually being decrypted. Homomorphic encryption would have eliminated this problem.

An infamous similar occurrence was in April 2011 (Thomas, 2011) when Sony's PlayStation network was hacked into and had its user's accounts breached leaking credit card information. Victims of such incidents even after either resetting their passwords across various other platforms or terminating their service contracts very often become victims of other related crimes in the future. The impact of having personally identifiable information in the wrong hands such as hackers exposes the victims to a potential lifelong battle of constantly guarding against similar violations. Following the Sony incident in the same year (Walters, 2014), Dropbox was found to be storing unencrypted user files. The discovery caused a major uproar and in protest, users closed down their accounts as a result of the company having not encrypting their confidential files. The cloud can still perform meaningful computations on data even though it is encrypted with homomorphic encryption. Homomorphic encryption would have provided a better solution in ensuring that the security standards for these companies' clients are upheld. An enforced standard would have provided the legal framework onto which all these companies would have had to adhere to. Furthermore, the standard would have also provided guidelines on how to prevent and deal with such situations.

### 3. Homomorphic encryption

Homomorphic encryption is a form of encryption that permits computation on ciphertexts, generating an encrypted result which, when decrypted matches the result of the operations as if they had been performed on the plaintext. In abstract algebra, this phenomenon is referred to as homomorphism. The performed functions are called mappings between the plaintext space and the ciphertext space.

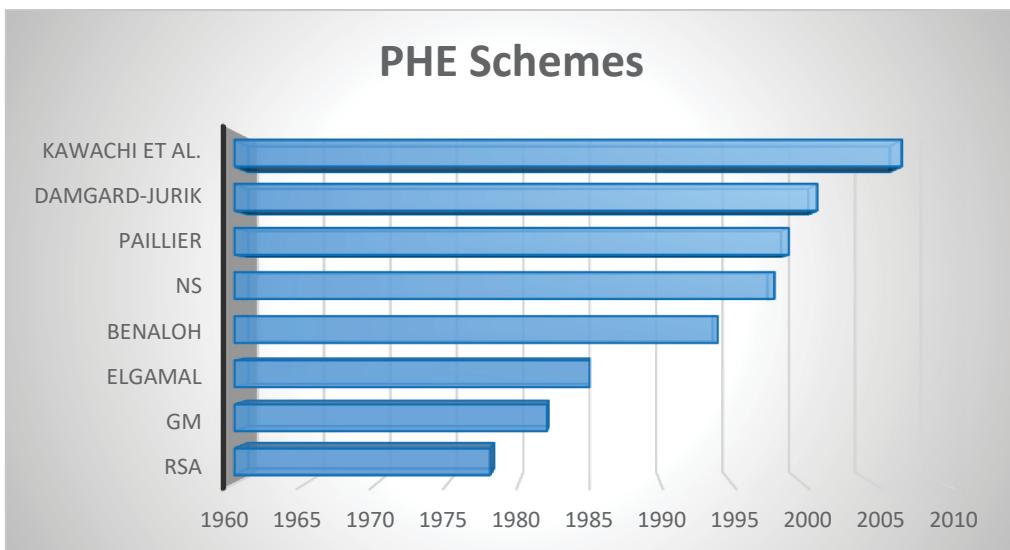
*Definition:* An encryption scheme is called homomorphic over an operation ' $\oplus$ ' if it the following equation hold  

$$E(m_1) \oplus E(m_2) = E(m_1 \oplus m_2), \forall m_1, m_2 \in M \quad (1)$$

where  $E$  represents the encryption algorithm and  $M$  is the set of all messages.

Homomorphic encryption schemes are based on well-defined mathematical structures which are hard to solve and for this reason makes them generally more secure. Homomorphic encryption schemes are categorized as partially homomorphic (PHE), somewhat homomorphic and fully homomorphic. With regards to partially homomorphic encryption schemes, only one operation on encrypted data can be performed, addition or multiplication but not both. A somewhat homomorphic encryption (SWHE) scheme permits more than one operation to be performed but can only support a limited number of addition and multiplication operations. When both addition and multiplication can be performed under any function the encryption scheme is said to be fully homomorphic (FHE). Properties of homomorphic encryption include the requirement for **semantic security**: as with regular encryption homomorphic encryption is said to be secure if no adversary has an advantage, more than 50%, in guessing whether a given ciphertext is the encryption of one of two messages. For this property, the requirement is for encryption to be randomised so that an adversary is not able to tell two encryptions of the same message from one another. **Efficient decryption:** For any function to be carried out by the server, this property guarantees that the decryption runtime is not dependent on the functions that were evaluated on the ciphertexts. And lastly, **compactness:** the growth of the ciphertext must be independent of the complexity of the homomorphic encryption operation. That is, for any arbitrary set of operations, the ciphertext does not expand in length. The compactness property guarantees that homomorphic operations on the ciphertexts do not expand the length of the ciphertexts.

The use of FHE over SWHE or PHE, or whichever one of the homomorphic encryption schemes is largely dependent on the functions that one would like to perform. Complex functions would work most efficiently under the fully homomorphic encryption scheme. In reality, most of the current homomorphic encryption schemes practically used in the cloud are somewhat or partially homomorphic. The functional requirements for both these types of cryptosystems is not as complex compared to that for fully homomorphic schemes. As a result, these homomorphic encryption schemes have 'manageable' ciphertext growths from the addition and multiplication functions performed. The compactness property has to be fulfilled in order to have efficient results or decryption. All other homomorphic encryption schemes, following the first few PHE schemes, where introduced as optimizations of conventional encryption methods with the aim to eventually be able to perform an unlimited number of arbitrary functions on encrypted data while circumventing the known challenges within the technology. Several useful PHE and SWHE schemes existing that are in practice in real life scenarios. Some prominent examples of PHE schemes within literature that paved the way towards the idea of a FHE scheme are the RSA (Rivest et al, 1978); the Goldwasser-Micali or GM (Goldwasser and Micali 1982); the ElGamal (ElGamal, 1985); Benaloh (Benaloh, 1994); Naccache and Stern (Naccache and Stern, 1998); Okamoto and Uchiyama (Okamoto and Uchiyama ,1998) ; Paillier ( Paillier 1999); Damgård and Jurik (Damgård and Jurik, 2001) and Kawachi et al. (2007). Each of these cryptosystems have greatly idealized PHE in some away and hence some were adopted for implementation in real-life applications. For instance, the RSA homomorphic encryption scheme has been used to secure internet (digital signatures), banking and online credit card transactions. The ElGamal HE scheme has been used for the implementation for Hybrid systems. One other application area that has been successfully implemented is the online e-voting systems. These are some examples of real-life applications that prove the efficient implementation of homomorphic encryption under specific functional requirements.



**Figure 1:** A timeline of well-known PHE schemes

With regards to SWHE cryptosystems, examples include the Yao (1982); Sander et al. (1999); the Boneh-Goh-Nissim or BGN (Boneh et al., 2005); Ishai and Paskin (2007) just before Gentry's FHE scheme a couple of years later in 2009. These SWHE schemes were proposed as a result of the performance issues identified within the first PHE schemes. The BGN supports arbitrary number of additions and one multiplication by keeping the ciphertext size constant. SWHE schemes provide the next-level towards achieving the first FHE scheme. In 2009, Gentry's PhD thesis (Gentry, 2009) gave rise to the first FHE scheme. Its security is based on the Euclidean lattices. Gentry's HE scheme is an optimization of the earlier proposed schemes, however the scheme has some computational overhead. The runtime for output is unfeasible for real-life scenarios. Gentry has however proposed the concept of bootstrapping to try overcome this shortfall. Mainly due to the long runtimes associated with the implementation of this FHE scheme many new schemes and optimizations have followed his work in order to address the overhead. New approaches to obtain a new FHE scheme is mostly based on the problems on lattices as a more reliable security structure. More complex than the existing mathematical structures, lattices offer a better solution particularly with the ciphertext noise reduction. A lattice, in mathematics, is the linear combinations of independent vectors or basis vectors. Other implementations towards a FHE schemes have been proposed following Gentry's ground breaking introduction of the first theoretical implementation of a FHE scheme based on different mathematical structures such as the FHE scheme over integers, Learning with Error (LWE) and the Ring Learning with Error (RLWE) approaches. Essentially, the goal for homomorphic encryption is to have a HE scheme that is both efficient in producing the correct results and also be secure within an acceptable agreed threshold. The security of fully homomorphic encryption schemes is generally evaluated by the earlier mentioned properties: semantic security, compactness and efficient decryption.

#### 4. Related work

In recent years there has been a significant amount of efforts in progress towards the standardization of the implementation for fully homomorphic encryption. The process, although very much progressive, is still at its infancy stage. Most applications follow the SWHE scheme approach, where parameters are set to allow the evaluation of limited depth circuits, making the computations practical, and avoiding costly operations. The idea for a practical FHE scheme is theoretical at this stage and as a result, most literature is aimed at coming up with more practical solutions to the existing challenges. Nonetheless, the proposed standardization approach seeks to leverage on these gaps to galvanize the community toward a common purpose to standardization of Homomorphic Encryption. Some of the standardization workshops that have been held are the Microsoft collaboration with New Jersey Institute of Technology and National Institute of Standards and Technology in July 2017(Chase et al, 2017). The outcome yielded white papers were drafted by three working groups at the workshop: standards for API design, security and the applications for HE. Shortly thereafter, the MIT Stata Centre workshop was held. The goals of the meeting were to approve a draft standard for parameter selection for homomorphic encryption and revisit the initial API standard draft. Keynote speakers within the HE community were invited to facilitate the 'meeting-of-the-minds' session towards the standardization effort. Several other

HE standardization workshops attended by a group of subject matter experts and researchers from all across the world have been held.

Amongst other aspects, the impact of standardization implies the need for agreed thresholds on security for homomorphic encryption schemes since the main issue hampering the adoption of these cryptosystems is on their security. Current efforts provide a detailed analysis of the known state of security of these schemes and then recommend a wide selection of parameters to be used at various security levels. Chen et al. (2017) use the known security attacks and their estimated running times in order to make parameter recommendations. Also, additional features of these encryption schemes which make them useful in different applications and scenarios are described. In cryptography, security attacks are distinguished into (1) cipher-text only attacks, (2) known-plaintext attacks, and (3) chosen plaintext attacks (Diffie and Hellman, 1976). In the instance of an attack, the security requirement for semantic security will have been violated. Message digest algorithms can be used in order to ascertain if there has been any tampering with the original message. With this algorithm, the same number is produced for a given input and also provides a different number for different inputs. An exact copy will have the same message digest as the original but if a file is changed even slightly it will have a different message digest from the original (Casey, 2011). This mechanism can also be applied to detect growth of the ciphertext length which often result in incorrect evaluated output. In their paper, Chen et al. (2011) encourage the consideration of stronger security requirements beyond semantic security as other forms of severe attacks are not addressed by the semantic security guarantee, countering them requires additional measures beyond the use of homomorphic encryption. The proposal of the standardization approach for a digital forensic investigation is seeking to provide assurance of resolution to an investigative process in a security breach incident.

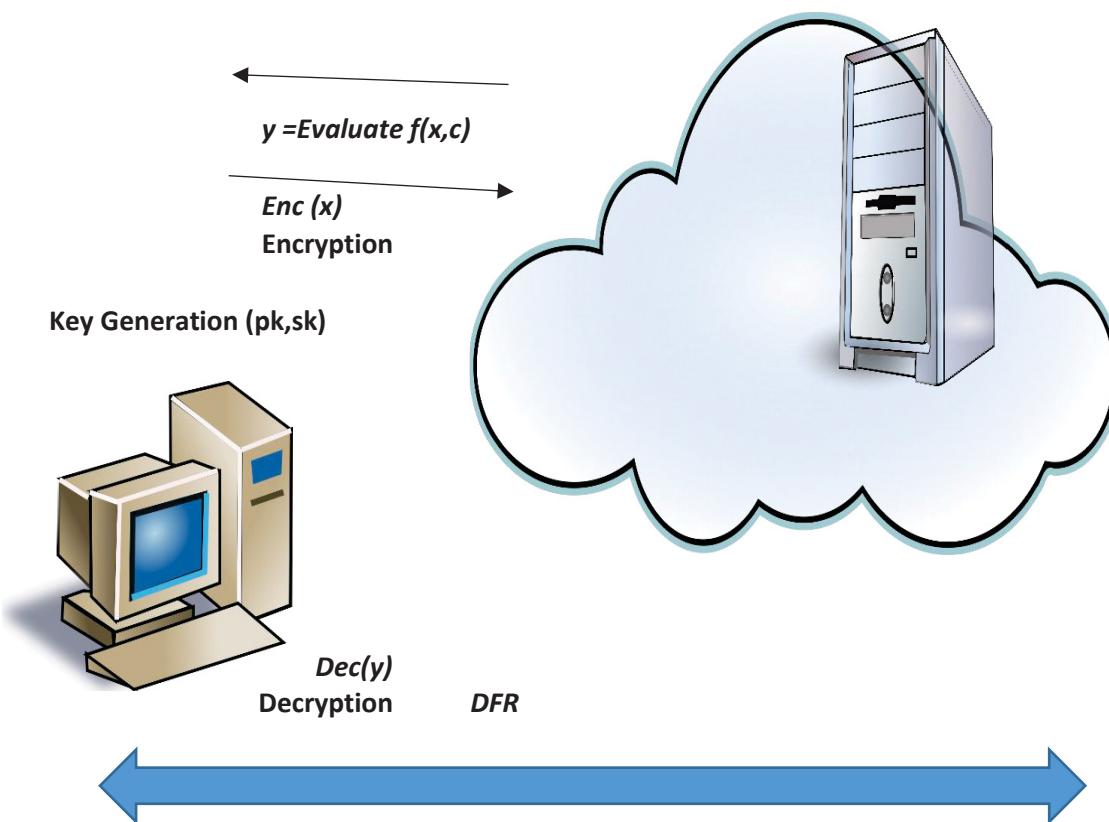
## **5. Proposed model**

The field of digital forensics deals with crimes that are committed in the electronic or digital domains. Within the cyberspace, various law enforcement agencies have witnessed all sorts of cybercrimes that have led to victims either losing their hard earned investments, lured to their untimely deaths or having these criminals involved in money laundering scams. All these types of cybercrimes have highlighted the need for digital forensic investigations in order to bring these criminals to book.

The Digital Forensic Research Workshop (DFRWS) (DFRWS Workshop, 2001) defines digital forensics to be the use of scientifically derived and proven methods towards the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence derived from digital sources for the purposes of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorised actions shown to be disruptive to planned operations.

Digital forensic investigation in the cloud present investigators with a number of challenges due to the distributed nature of the cloud. Digital evidence is usually distributed on many computers making the collection of data related to the potential crime unfeasible. Casey (2011) defines digital evidence as any stored or transmitted data obtained from electronic devices that could support or refute the theory of how a digital data incident occurred or that addresses critical elements of the offense such as intent or alibi. Digital evidence is the essence of digital forensic investigation, investigators need to ensure the integrity and preservation of digital evidence throughout the entire investigation process leading up to presentation of the findings of the process. Evidence integrity refers to the all the checks that are performed to show that the evidence has not been altered from the time of collection therefore authenticating the originality of the collected data. In forensic science, the use of message digests and hash values are mechanisms that assist in ensuring integrity of digital evidence. Earlier, we have identified that stronger additional security measures may need to be enforced beyond the use of homomorphic encryption, especially for fully homomorphic encryption schemes. Many types of digital evidence found in clouds will likely be similar to that found in conventional digital forensic investigations, including Office application documents, emails and images. Grispos, et al. (2012) provide an evaluation of current applications that can be used to generate digital evidence in the cloud. A number of logging mechanisms are used by many cloud service providers for tracking use within their services. The Message Log Search, a service from Google, allows administrators to make queries on email messages. Investigators can use this tool to find information logs such as emails, identification of email recipients, the date the email was sent and IP addresses.

Digital forensic science can be categorised as proactive, reactive and coactive. For the proposed model, the proactive approach to digital forensic science will be sort in an attempt to prepare for - and to avoid criminal attacks by hackers in the cloud for existing homomorphic encryption schemes. Grobler et al. (2010) mentions that proactive digital forensics, also referred to as digital forensic readiness, is a means of detecting potential criminal acts. The likelihood for the occurrence of severe security attack with regards to homomorphic encryption is if the scheme is symmetric unlike asymmetric cryptosystems. In the cloud environment, many Cloud Service Providers (CSPs) lack the infrastructure to gather logs and often intentionally keep this information away from their clients. Trenwith and Venter (2013) mention that the CSPs also have full control over the sources of evidence and also the companies' assets, making investigations by corporate security teams difficult if not impossible. There is significant literature coverage on the challenges of the implementation of efficient digital forensic investigation in the cloud. The process can however be simplified by the application of digital forensic readiness (DFR) in the cloud. The proposed solution emphasizes the implementation of a digital forensic readiness process in the cloud. Various standards exist for DFR, Rowlingson (2004) provides a more detailed ten step process for digital forensic readiness which is preferred herein for that reason. To apply DFR efficiently to cloud environments requires the collection of evidence in such a manner that the integrity of the evidence is maintained throughout the process of collecting, transporting and storing of evidence. The collected evidence has to be preserved through the investigation process.



**Figure 2:** Digital forensic readiness (DFR)

## 6. Conclusion

Current homomorphic encryption schemes are distinguished by their efficient implementation across various applications. The adoption of this technology in the global arena present opportunities in different industries such as health care, medicine and financial sectors to protect data and client privacy but first homomorphic encryption will have to be standardized. The standardization will most likely be by multiply stakeholders such as corporate bodies and government agencies. In the IT environment, without the existence of a practical implementation of a FHE scheme guaranteeing an elevated security level with no functional limitations, there is still a need for CSPs or organisations to have additional security measures in place that give 'extra security' in

the unfortunate event of a security breach. The proposed model identifies digital forensic readiness procedures as one of those additional precautionary measures, conventional digital forensic readiness procedures are used as they have proven to be effective. There is still opportunity to explore other measures for these cryptosystems. Importantly, is the agreement on security levels for varying parameter sets for homomorphic encryption schemes. Ultimately, this effort will also have legal implications on a global scale. Although there may not be a 'one size fits all' HE scheme for real-life applications literature identifies some homomorphic encryption schemes that are best suited for specific applications based on their complexity and amount of time required for the computation. Essentially, the right scheme is the one that fits your constraints in the way that is acceptable. These constraints may be acceptable for other applications over the other. Time, memory and security could vary as critical amongst different applications. The trade-offs between these constraints ultimately dictate the best homomorphic encryption scheme. Perhaps optimizations of the FHE scheme will lead to a scheme that is able to solve all possible arbitrary functions with minimum if any overhead.

## References

- Benaloh, J., 1994, May. Dense probabilistic encryption. In *Proceedings of the workshop on selected areas of cryptography* (pp. 120-128).
- Boneh, D., Goh, E.J. and Nissim, K., 2005, February. Evaluating 2-DNF formulas on ciphertexts. In *Theory of Cryptography Conference* (pp. 325-341). Springer, Berlin, Heidelberg.
- Brakerski, Z., Gentry, C., & Vaikuntanathan, V. (2012). (Levelled) fully homomorphic encryption without bootstrapping. In *ITCS*, pages 309–325.
- Canetti, R., Goldreich, O., & Halevi, S. (1998). "The Random Oracle Methodology Revisited". Symposium on Theory of computing (STOC). ACM. pp. 209–218. Retrieved 2007-11-01.
- Casey, E. (2011). *Digital Evidence and computer crime Forensic science computers and the internet*, 3<sup>rd</sup> ed. Elsevier Inc.
- Chase, M., Chen, H., Ding, J., Goldwasser, S., Gorbunov, S., Hoffstein, J., Lauter, K., Lokam, S., Moody, D., Morrison, T. and Sahai, A., 2017. Security of homomorphic encryption. *HomomorphicEncryption.org, Redmond WA, Tech. Rep.*
- Damgård, I. and Jurik, M., 2001, February. A generalisation, a simplification and some applications of Paillier's probabilistic public-key system. In *International Workshop on Public Key Cryptography* (pp. 119-136). Springer, Berlin, Heidelberg.
- Diffie, W. & Hellman, M. (1976). *New Directions in Cryptography*. *IEEE Transactions on Information Theory*, Vol 22, No 6, November 1976, pp. 644-654.
- ElGamal, T., 1985. A public key cryptosystem and a signature scheme based on discrete logarithms. *IEEE transactions on information theory*, 31(4), pp.469-472.
- Garfinkel, S.L. (2010) Digital forensic research: The next 10 years. *Digital investigation: The International Journal of Digital Forensics & Incident Response*, 7. 64 – 73.
- Gentry, C. (2009) A fully homomorphic encryption scheme. PhD thesis, Stanford University. From <http://crypto.stanford.edu/craig>
- Grispos, G., Storer, T., & Glisson, W.B. (2012). Calm Before the Storm: The Challenges of Cloud Computing in Digital Forensics. *International Journal of Digital Crime and Forensics*, Volume 4, Issue 2, Pages 28-48
- Grobler, C.P., Louwrens, C.P. and von Solms, S.H., 2010, February. A framework to guide the implementation of proactive digital forensics in organisations. In *2010 International conference on availability, reliability and security*(pp. 677-682). IEEE.
- Hemdan, E. & Manjaiah, D.H. (2015). *International Journal of Innovative Research in Computer and Communication Engineering*. Available from [http://www.ijrcce.com/upload/2015/sacaim/1\\_101.pdf](http://www.ijrcce.com/upload/2015/sacaim/1_101.pdf)
- K. Thomas. "Sony Makes it Official: PlayStation Network Hacked," *PC Computing*, pp. 12, April 2011.
- Kawachi, A., Tanaka, K. and Xagawa, K., 2007. SERIES C: Computer Science.
- Khan, S.I. and Latiful Hoque, A.S.M., 2016. Digital Health Data: A Comprehensive Review of Privacy and Security Risks and Some Recommendations. *Computer Science Journal of Moldova*, 24(2).
- Naccache, D. and Stern, J., 1998, November. A new public key cryptosystem based on higher residues. In *ACM conference on Computer and communications security* (pp. 59-66).
- National Institute of Justice (2001), 'Electronic Crime Scene Investigation A Guide for First Responders'. From <http://www.ncjrs.org/pdffiles1/nij/187736.pdf>
- Okamoto, T. and Uchiyama, S., 1998, May. A new public-key cryptosystem as secure as factoring. In *International conference on the theory and applications of cryptographic techniques* (pp. 308-318). Springer, Berlin, Heidelberg.
- Paillier, P., 1999, May. Public-key cryptosystems based on composite degree residuosity classes. In *International Conference on the Theory and Applications of Cryptographic Techniques* (pp. 223-238). Springer, Berlin, Heidelberg.
- Pandi, Gayatri & K H Wandra, Dr & D Scholar, Ph. (2018). CLOUD FORENSIC FRAMEWORKS, CHALLENGES, STATE OF ART AND FUTURE DIRECTIONS.
- R. Rowlingson Ph.D, "A ten step process for forensic readiness," *International Journal of Digital Evidence*, vol. 2, 2004.
- Reith, M., Carr, C., & G. Gunsch. (2002). "An example of digital forensic models", *International Journal of Digital Evidence*, vol.1 no.3, pp. 1-12
- Rivest, R., Adleman, L., & M. Dertouzos. (1978). On data banks and privacy homomorphisms, In: *Foundations of Secure Computation*, New York: Academic Press, pp.169–179

***Nnana Mogano***

- Rivest, R.L., Shamir, A., & L. Adleman. (1978). A Method for Obtaining Digital Signatures and Public Key Cryptosystems," Communications of ACM, Vol. 21, pp. 120-126
- Sander, T., Young, A. and Yung, M., 1999. Non-interactive cryptocomputing for nc/sup 1. In *40th Annual Symposium on Foundations of Computer Science (Cat. No. 99CB37039)* (pp. 554-566). IEEE.
- T.C. DFRWS Workshop. (2001) "A road map for digital forensic research: dfrws technical report," DTR-T00-01 Final, Tech. Rep
- Taylor, M., Haggerty, J., Gresty, D. & Hegarty, R. (2010). Digital evidence in the cloud computing systems. *Computer Law & Security Review*, 26(3). 304-308.
- Trenwith, P.M. & Venter, H.s. (2013). Digital forensic readiness in the cloud. *Information Security for South Africa 2013* IEEE. 1-5. 10.1109/ISSA.2013.6641055.
- Walters, R., 2014. Cyber attacks on US companies in 2014. *The Heritage Foundation*, 4289, pp.1-5.
- Yao, A.C., 1982, November. Theory and application of trapdoor functions. In *23rd Annual Symposium on Foundations of Computer Science (SFCS 1982)* (pp. 80-91). IEEE.

# Fake News Detection Using Ensemble Machine Learning

Potsane Mohale and Wai Sze Leung

University of Johannesburg, Johannesburg, South Africa

[216048751@student.uj.ac.za](mailto:216048751@student.uj.ac.za)

[wsleung@uj.ac.za](mailto:wsleung@uj.ac.za)

**Abstract:** Consuming news from social networks has become the new normal. In theory, the rapid rise of social media can be seen to have a positive impact in promoting active citizenship. Unfortunately, social networks can also prove to be a considerable threat. Just as their platforms promote easy access to rapid and low-cost dissemination of information, social media is also fertile ground for the spreading of misinformation. As a result, governments now face the reality of having to bolt the proverbial barn door while the fake news horse is already free, running amok. Given fake news' ability to deceive, cause instability, and spread propaganda, governments must ensure that there are measures in place that will allow them to effectively deal with what qualifies as a form of cyber warfare. It is therefore quite critical to be able to detect fake news on social media to mitigate the potential negative effects. Detecting fake news on social networks can be quite a challenge as they are often written to masquerade as real news. To effectively detect fake news on social networks, one may need to exhaust all the auxiliary information. Furthermore, the sheer amount at which fake news is seen to propagate means that the process of identifying and shutting down the fake news articles cannot be left to human means alone. Ensemble machine learning makes use of a set of classifiers whose individual decisions are aggregated by weighted voting to improve predictions, decrease variance and bias. In this paper, we present an ensemble machine learning model which determines the truth probability of given statements from social networks by considering all the relating metadata. The proposed system makes use of five different classifiers to improve the detection of fake news on social networks. We trained and tested the model using a fake news dataset with the results of our experiment yielding fake news detection accuracies averaging 80%. Results are shown to improve significantly when the feature selection stage of the training process includes more attributes of the dataset.

**Keywords:** fake news, social networks, machine learning, classification, ensemble learning

---

## 1. Introduction

Automated, accurate detection of fake news on social media is becoming a necessity amidst the backdrop of an interconnected society. The rise of social media has brought about the existence of additional information sources which has often acted as a catalyst to promote democracy. Research has even indicated that more people are encouraged to vote as a result of greater exposure to news content (Markoff, 2012). Unfortunately, the propagation of news articles running unchecked in social networks has also led to the proliferation of fake news.

Fake news refers to news whose meaning and context has been deliberately modified to coerce a certain ideology on the readers, mostly on real events taking place in the world at any point in time (Mohale & Leung; 2018, Nurse et al., 2014; Catillo et al., 2013). The excessive spreading of fake news can have serious negative impacts on society because it leads to an imbalance of the news ecosystem authenticity. Fake news is often used by propagandists and politicians to convey their political messages and manipulate readers to accept certain biased and distorted facts. Fake news impacts how their readers react to their surroundings, and often triggers confusions, distrust and impeded the ability to differentiate between what is right and what is not effectively.

As we spend an increasing amount of time on social networking platforms, most people tend to consume news from these platforms rather than the traditional news outlets, reasons being that it is faster, more convenient and less expensive to consume news on social networks in comparison to television or newspapers (Shu et al., 2017). For example, at any point, 85% of Twitter trends are topics based on current affairs (Vosoughi et al., 2017; Yang et al., 2015). It is also easy to share and engage with other people on the news we read on social networks. This potential for rapid propagation and far-reaching information raise a lot of information quality assurance and management challenges for social network platforms.

Most of the leading social networks currently gauge information credibility manually. However, this is not efficient due to the large amounts of data on platforms (Liu & Xu, 2016). For example, an average of 7901 tweets are made on Twitter every second while on Facebook there are approximately 51000 comments being posted every minute (Mohale & Leung, 2018). Fake news on social networks has become a part of our everyday life; we often come across news reporting on political violence, health crisis, and intolerance amongst people of different

races, culture and ethnic groups. Spreading fake news on social networks can cause a great deal of distress to the people, as most rely on information from these platforms for making decisions concerning medical issues, shopping choices and how they react during disasters or political unrests. Therefore, automatic and effective detection of fake news on social networks is very important in order to ensure the social well-being of the communities and general national security.

Detecting fake news on social networks presents a huge challenge because fake news is written in a way that intentionally misleads the readers into believing that it is real authentic news. Therefore, it is necessary to consider auxiliary information like user profile involved, and the user engagement on such a post in order to come up with an effective detection mechanism. Taking advantage of this auxiliary data is also a big problem since the data on social networks can be unstructured, incomplete and comes in very big volumes (Kwon et al., 2014; Saez-Trumper, 2014). The fake news content propagated on social networks is very diverse, comes with different writing styles, and sometimes refers to true sources in an incorrect context in order to support false claims (Shu et al., 2016; Roy et al., 2018).

In this paper, we present an ensemble machine learning model comprised of six independent classifiers. We trained five different classifiers using the Liar data sets and aggregated the performance of each classifier before feeding their outputs to the main classifier and applying an adaptive boosting algorithm to improve accuracy. We anticipate that our model can become more effective than most existing solutions which rely on a single classifier due to reduced variance. An ensemble model provides us with an improved accuracy because it eliminates several problems posed by unstructured data on social networks such as confidence estimation, non-stationary distributions and missing features (Zang & Ma, 2012).

The rest of this paper is organized as follows. In section 2 we present a brief literature study on fake news on social networks and describe the psychological background of fake news consumption. In section 3 we discuss the work that has been done by other notable researchers to try to automate the detection of fake news on social networks, and we also outline what is lacking from these approaches. In section 4 we discuss in greater detail what ensemble machine learning is, and how it can combat the current inaccuracies of models being used to detect fake news on social networks. Sections 5, 6 and 7 describe our ensemble machine learning solution for fake news detection on fake news and provide an analysis of the results achieved. Section 8 concludes our work and provides a discussion on future work.

## **2. Literature review**

Historically, fake news has existed for a long time and has become more invasive owing to the development of communication tools which allow rapid transmission of information over a very short period (Simons, 2018). The use of propaganda and misinformation has existed since news began to be circulated by the printing press, long before the establishment of modern journalism integrity rules and ethical codes (Rubin et al., 2016; Bakir & McStay, 2018). Humans have never been any good at differentiating between fake and real news due to our mental vulnerabilities. The main vulnerabilities that make humans susceptible to consuming fake news are the confirmation bias and the naïve realism (Shu et al., 2017; Kumar & Shah, 2018).

Confirmation bias is the tendency of people to only look out for information which is in line with their pre-existing beliefs, overlooking the other facts (DiFranzo & Gloria, 2017). Naïve realism is the theory that news consumers believe that their perceptions about reality is the only accurate and rational views. Social networks polarize users' views, triggering bias and naïve realism in a more disinterred manner by means of an echo chamber effect (Shu et al., 2017; Kumar & Shah, 2018).

### **2.1 Echo chamber effect on social networks**

The echo chamber effect, which is also known as bubbles, is a feature of social networks which makes these platforms a fertile ground for publishing fake news. Bubbles are groups of users who consciously or unconsciously consume the same set of news on social networks. This new paradigm of information consumption changes the way people are exposed to news. In such bubbles, users are selectively exposed to news that are compatible with their own ideologies. This creates a psychological challenge of users being unable (or perhaps unwilling) to dispel fake news. The echo chamber effect facilitates a situation where people believe fake news based on two factors, social credibility and frequency heuristics (Roy et al., 2017; Zhang et al. 206).

Social credibility means that consumers of news are more likely to believe a news article if others also believe in it, especially when there is no way of verifying the validity. Frequency heuristics means that readers will believe news they are frequently exposed to, even if it is fake news. In an echo chamber, users consume and share similar content with one another. The echo chamber leads to homogenous groups of people with limited information cycles (Shu et al., 2016; Mohale & Leung, 2018; Conroy et al., 2015). Echo chambers are believed to trap humans easily. Psychology explains that humans have a tendency of ignoring facts which tax their brains, and so would rather stay in a cluster of like-minded people and seek information which conforms to their own biases (Niclewicz, 2017, The Economist, 2016).

## **2.2 Fake news as a security threat**

Fake news is slowly becoming a national security problem that can lead to dangerous physical situations and threaten national peace and stability. The threat caused by spreading fake news on social networks has led to several European governments to consider the implementation of laws that hold people accountable and organization accountable for the dissemination of inaccurate information (Vasu et al., 2018). In this section, we outline a few real-life examples of how fake news is being used to destabilize national security.

One example of this is a fake story which went viral in Russia claiming that a German soldier had forced himself on a local girl in Lithuania. This story caused much commotion as the chairman of the NATO Military Committee felt that Russia was undermining the NATO troops deployed in the Eastern flank countries (Reuters, 2017). Germany's agency for domestic security further warned against cyber misinformation which put government official and members of parliament in danger. Another example of how fake news can put human lives in danger is the fake story which claimed that a paedophile ring linked to Hilary Clinton was operating from a restaurant in Washington D.C. Despite being debunked, this story generated over a million tweets during the period leading to the US presidential elections (Shu et al., 2016). Following this story, one man armed with an automatic rifle stormed into the restaurant in question in December 2016 and started shooting. When interrogated, the man explained that he had wanted to investigate what he had come across on social networks (Niclewicz, 2017).

Further fake news stories have led to social tensions emerging since. There have been unverified claims of Russian influence in the outcome of the 2016 US presidential elections. These claims have gained interest and traction, playing a major role in sowing division and leading to US citizens questioning the credibility of their country's democracy and the trustworthiness of their leaders. The propagation of fake news stories on social networks was at its peak during the US elections, with fake news stories were being more engaged than real news, according to Niclewicz, (2017). Twenty popular fake news stories were shared and received about 9 million reactions on Facebook alone. The echo chamber effect adds to the distortion of reality on social networks by reinforcing emotions and creating polarized societies.

## **3. Related work**

Fake news on social networks is not a new phenomenon. It has been well-studied in different disciplines, such as journalism and human psychology. However, it is only in recent years where research in detecting fake news using computational means has gained momentum. Automation of fake news detection has been mostly researched from two perspectives: analysing the spread of fake news; and analysing the content of fake news (Kwon et al., 2017). Inspired by the previous research into this topic, our previous work also focused on studying the most common features of fake news on social networks and creating a ranking of the most appropriate features to feed to a classifier for training.

Castillo et al. (2011) focused on the way credible and non-credible information propagates on Twitter by analysing posts related to currently trending topics. They grouped features of content on Twitter into four groups, the message attributes, the user details, the topics itself and the propagation dynamics of the tweet in question, feeding this to a classifier to determine the credibility of the information. Their results reveal that user characteristics and propagation features are the most important to consider when automating the detection of fake news on social networks. Similarly, Rubin et al. (2016) also worked toward a mechanism which employs the Rhetorical Structure Theory to identify the differences in network dynamics of different stories.

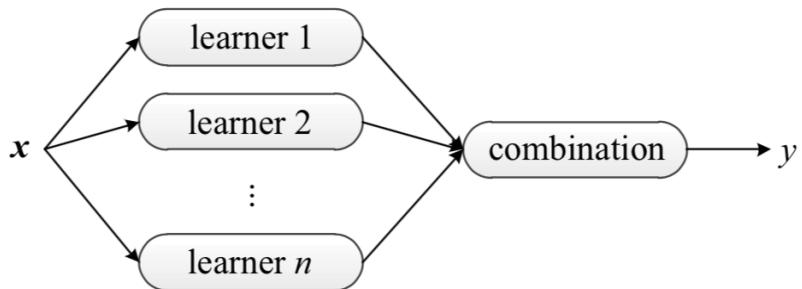
Vosoughi et al. (2017) developed a model to automatically verify rumours on Twitter by examining the linguistic styles and characteristics of people involved in the network propagation dynamics by training a Hidden Markov Model on rumours collected on Twitter over real-world events. Their model achieved a very high accuracy of

75% in predicting rumours on Twitter. However, this solution does not include other social networks and does not consider other languages (other than English) when dealing with the linguistic features of a rumour.

In his work, Bajaj (2017) built classifiers which predict whether the given news content is fake or not, by purely focusing on the Natural Language Processing mechanisms. He used a dataset from Kaggle and a Signal Media News dataset and evaluated the different classifiers based Gated Recurrent Unit, Bi-directional Short term and Long-term Memory. Similarly, Roy et al. (2018) also developed several deep learning models to detect fake news and classified them into two categories, pre-defined and fine-grained. They developed models based on the Convolutional Neural Network and Long-term Short-term Memory and fed the representatives from each of these to Multi-layered Perceptrons. Their results gave out a 45% accuracy, which is not the worst for an unsupervised model (Shu et al., 2017).

#### 4. Ensemble machine learning

Ensemble learning, which is also known as committee based learning or multiple classifier algorithms, are methods which train multiple classifiers to solve one problem, by combining their results. Ensemble methods combine a set of learners, in contrast to the ordinary approach which constructs a single classifier from the training data (Zhou, 2012; Zhang & Ma, 2012).



**Figure 1:** Basic overview of an ensemble-based learning method (Zhou, 2012)

Ensemble learning methods contain several classifiers called base learners which are generated by training data using a base learning algorithm. Most ensemble models make use of a single learning algorithm to come up with homogenous base learners, while some models may use multiple base learning algorithms to produce heterogeneous learners (Dietterich, 2000). Ensemble models address numerous machine learning problems such as feature selection, confidence estimation, error correction and missing data. There are three basic stages engaged in when building an ensemble machine learning model: data sampling, classifier training, and classifier combination, as briefly discussed below (Zhou, 2012; Zhang & Ma, 2012):

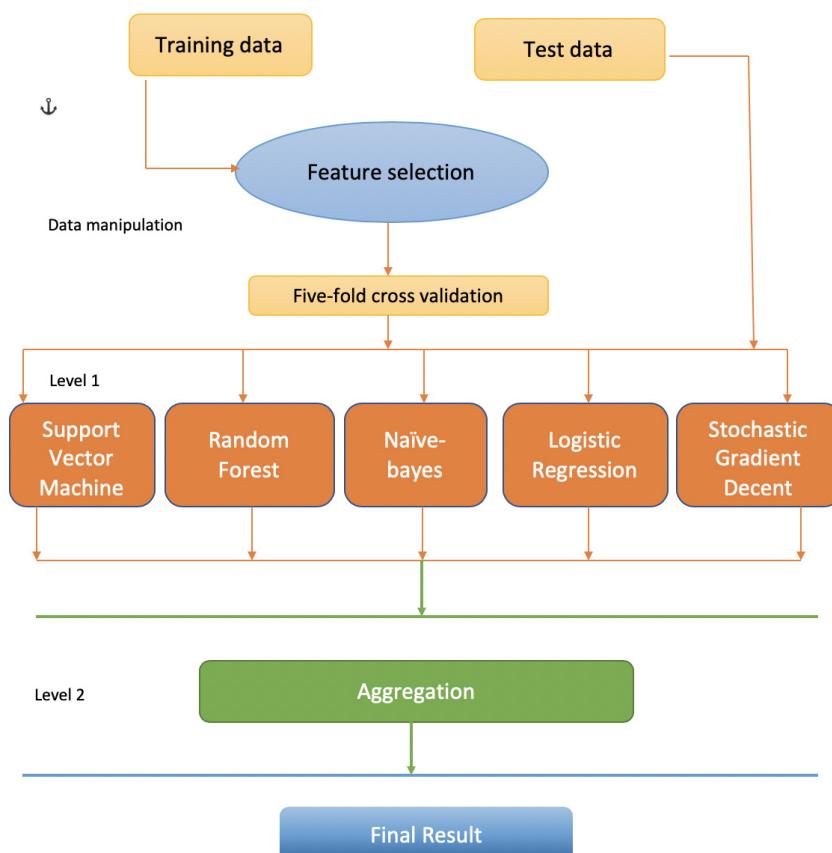
- (a) Data sampling – it is essential to select an appropriate sampling strategy for an ensemble, as different sampling strategies lead to different performances. For example, sampling from a distribution which collects misclassified samples lead to a boosting algorithm while using replicas of the training data on different base learner leads to bagging.
- (b) Training classifiers – bagging, boosting and stack generalization are the most commonly applied methods for training classifiers in an ensemble model.
- (c) Classifier combination – there are different rules for combining the results of each classifier of an ensemble, the most common one being a simple majority voting. Voting based combinations can be applied after training each classifier, but in some cases, an extra training step may be required before voting is done. Voting comes in three different ways, unanimous voting where all classifiers agree to one solution, simple majority voting, where more than half of the classifiers should agree and plurality voting where the solution with the highest number of votes is considered the ultimate one.

It is important to note that combining different classifiers does not necessarily mean that the final output will be better than the best performing classifier in the ensemble. Instead, it reduces the likelihood of choosing a classifier with the poorest performance.

Ensemble methods are very robust and accurate as they provide a very intuitive approach to incremental learning. With an ensemble learning, we can solve the problem of data fusion which is common when attempting to detect fake news on social networks where data comes from different sources (Dietterich, 2002). Ensemble methods can be applied in this scenario by training different classifiers on data from different sources. Data on fake news on social networks often poses the problem of missing features and incomplete data, and ensemble methods can handle this exceptionally well by training different classifiers with different subsets of the whole feature set. Ensemble methods further allow easy determination of confidence estimation in the general performance, which is almost impossible with single classifiers (Lee et al., 2010; Valentini, 2002).

## 5. Methodology

In this paper, we propose an ensemble learning model for detecting fake news on social networks using an adaptive boosting algorithm. Our solution consists of two levels: the first level trains the base classifiers and feeding their predictions to the main classifier as illustrated in Figure 2. To train the main classifier, we used five-fold cross-validation by randomly splitting the data set into subsets and using the prediction from the five base classifiers.



**Figure 2:** Proposed ensemble-based learning method overview

This stacking mechanism guarantees a stable ensemble method by running numerous iterations over the training stage and optimally weighing the majority votes of the base classifiers (Sikora, 2016). In each iteration, our stacking algorithm trains the base classifiers, selects the one with the lowest error count, and feeds the chosen classifier's predictions as part of the training set for the main classifier. During each iteration, the classification error on each base classifier is substantially reduced. This process results in having a final classifier with much more accuracy than the base classifiers.

## 6. Implementation

### 6.1 Data and feature selection

We trained and tested five models on the publicly available “Liar: A benchmark dataset for fake news detection” (Wang et al., 2017) data set which contains thirteen variables for training, testing and validation. This set consists

of about 12800 labelled statements collected from a fact-checking website, PolitiFact.com. The content in this dataset is annotated using six different hand-curated annotations. Table 1 describes each of these annotations. The Liar dataset contains 10269 training statements, 1283 testing statements and 1284 validation statements.

**Table 1:** Liar dataset annotations

Label	Description
True	The statement is valid
Mostly true	Most of the content is true
Half-true	The statement could either be true or false
Barely-true	Most of the content is false, but the statement could be true depending on context
False	The content contains a false statement
Pants-fire	The content contains blatant lies

Before we trained the classifiers on the data, we performed the feature extraction process using the methods from the python Sci-kit library. We used the term-frequency (tf-tf weighing) to encode the features and picked columns, 1 (the identity of the statement), 2 (the label of the statement), and 5 (the author's details) for training, testing and evaluation.

## 6.2 Training and classification

The three features we extracted are fed into five classifier we built using the Python Sklearn library functions. We used the Support Vector Machine, Random Forest, Naive-bayes, Logistic Regression and the Stochastic Gradient Decent classifiers. We manipulated the dataset and took advantage of the different classifiers by randomly dividing the training data into five disjoint subsets. Each of the classifiers is trained on of the derived subsets and tested independently using the rest of the test data. The results of each of the classifiers are taken to the second stage of the model and fed into the main classifier as features for the second training and testing. This cross-validation committee helps our model to effectively reduce the bias that comes with generalization as a result of overfeeding a single classifier (Sikora, 2015).

## 7. Results and analysis

To evaluate the performance of our ensemble model, we measure the precision, F-score, and recall. Precision is the number of correct positive predictions against the total number of positive predictions. Since our dataset is highly skewed, we easily reach a higher precision score by having fewer positive predictions. Recall is the number of true positives in the number of observations. The F-score combines precision and recall which also gives us the overall performance of our model. These metrics enable us to evaluate the performance of our model different perspectives.

To confidently test our hypothesis, we observed the performance of the five classifiers without the level 2 stage and aggregated their performance. We also went on to observe the performance of the whole model, with all stages taking place, and we noticed a major difference in performance after fitting the initial result to the main model.

**Table 2:** Average performance of each classifier

Label	Precision	Recall	F-score
True	0.74	0.37	0.49
Mostly true	0.56	0.63	0.6
Half true	0.55	0.72	0.59
Barely true	0.68	0.56	0.62
False	0.53	0.75	0.63
Pants-fire	0.79	0.51	0.62
Average	<b>0.65</b>	<b>0.6</b>	<b>0.60</b>

Table 2 reports on the precision, recall and F-score of the six labels in the data set for the five classifiers before level 2 training and testing. On average the five classifiers managed a 60% accuracy on the Liar data set. The recall is promising, and we believe it could get better if we train the model with more attributes of the user characteristics and if we were to consider the network dynamics of the statements in the data set. We also observed that the classifiers get caught up between similar annotations like true and mostly-true, where it often makes the mistake of swapping the decisions incorrectly. This error could be due to the limited set of our training data.

Finally, we observed the performance of the whole model after taking the level 1 results as input and training the main classifier using the rest of the provided test data. We can note that the stacking achieves a very much higher performance, and this is due to the initial splitting of the data set which reduced overfitting and generalization on the level 1 classifiers. The results do confirm our hypothesis that stacking does improve the results of automated detection of fake news on social networks. Table 3 shows the performance of the whole model after allowing the stacking to take place.

**Table 3:** Overall performance of the model

Label	Precision	Recall	F-score
True	0.83	0.66	0.79
Mostly true	0.77	0.81	0.84
Half true	0.75	0.89	0.82
Barely true	0.82	0.72	0.83
False	0.77	0.93	0.84
Pants-fire	0.88	0.7	0.84
Average	<b>0.8</b>	<b>0.76</b>	<b>0.82</b>

## 8. Conclusion and future work

In this paper, we tackle the problem of fake news on social networks by using ensemble machine learning. We first discussed the problems caused by fake news on social networks, how it affects people individually and the broader society and how it becomes a national security problem by giving real-world examples. We further went on to give a brief overview of how automation is currently being used to detect fake news on social networks and analysed the possible improvements in these techniques. We presented an ensemble-based model for detection of fake news as our way of improving on the currently existing approaches. We trained and tested our model on the Liar dataset, which is appropriate because it has been carefully annotated and split into training, testing and verification parts.

The testing results show that our model has an average recall of 80%, which is very high when compared to some of the currently existing solutions. We believe that our model can perform better when trained and tested on data sets which include a few more attributes of content on social networks such as complete user profile, a complete history of the statements published by the user, and auxiliary information such as time and location from where the statements were made. Further, our model presently works only on annotated structured data whereas most data on social networks can be noisy and highly skewed.

In the future, we would like to extend our work by developing a model for detecting most current trends on social networks and filtering such fake news based on current fake news trends. We believe that this will be very important, especially in times of disaster where people are most vulnerable to falling into the trap of believing fake news.

## References

- Bajaj, S., 2017. The Pope Has a New Baby!. *Fake News Detection Using Deep Learning*, Stanford.
- Bakir, V. and McStay, A., 2018. Fake news and the economy of emotions: Problems, causes, solutions. *Digital Journalism*, 6(2), pp.154-175.
- Castillo, C., Mendoza, M. and Poblete, B., 2013. Predicting information credibility in time-sensitive social media. *Internet Research*, 23(5), pp.560-588.
- Conroy, N.J., Rubin, V.L. and Chen, Y., 2015, November. Automatic deception detection: Methods for finding fake news. In *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community* (p. 82). American Society for Information Science.
- Dietterich, T.G., 2000, June. Ensemble methods in machine learning. In *International workshop on multiple classifier systems* (pp. 1-15). Springer, Berlin, Heidelberg
- Dietterich, T.G., 2002. Ensemble learning. *The handbook of brain theory and neural networks*, 2, pp.110-125.
- Kumar, S. and Shah, N., 2018. False information on web and social media: A survey. *arXiv preprint arXiv:1804.08559*.
- Kwon, S., Cha, M., Jung, K., Chen, W. and Wang, Y., 2013, November. Aspects of rumor spreading on a microblog network. In *International Conference on Social Informatics* (pp. 299-308). Springer, Cham.
- Lee, B.K., Lessler, J. and Stuart, E.A., 2010. Improving propensity score weighting using machine learning. *Statistics in medicine*, 29(3), pp.337-346.
- Liu, Y. and Xu, S., 2016. Detecting rumors through modeling information propagation networks in a social media environment. *IEEE Transactions on computational social systems*, 3(2), pp.46-62.

- Markoff, J. (2012) "Social Networks Can Affect Voter Turnout, Study Says", [online], The New York Times, <https://www.nytimes.com/2012/09/13/us/politics/social-networks-affect-voter-turnout-study-finds.html>
- Mohale, P. and Leung, W.S., 2018. Extrapolation of aspects of fake news on social networks. African Conference on Information Systems.
- Niklewicz, K., 2017. Weeding out fake news: an approach to social media regulation. European View, 16(2), pp.335-335.
- Nurse, J.R., Creese, S., Goldsmith, M. and Rahman, S.S., 2013, July. Supporting human decision-making online using information-trustworthiness metrics. In International Conference on Human Aspects of Information Security, Privacy, and Trust (pp. 316-325). Springer, Berlin, Heidelberg.
- Nurse, J.R., Agrafiotis, I., Goldsmith, M., Creese, S. and Lamberts, K. (2014). Two sides of the coin: measuring and communicating the trustworthiness of online information. Journal of Trust Management, 1(1), 5.
- Pomerantsev, P., 2016. cited in 'The Post-Truth World: Yes, I'd Lie to You'. The Economist, 10.
- Rubin, V., Conroy, N., Chen, Y. and Cornwell, S., 2016. Fake news or truth? using satirical cues to detect potentially misleading news. In Proceedings of the Second Workshop on Computational Approaches to Deception Detection (pp. 7-17).
- Roy, A., Basak, K., Ekbal, A. and Bhattacharyya, P., 2018. A Deep Ensemble Framework for Fake News Detection and Classification. arXiv preprint arXiv:1811.04670.
- Saez-Trumper, D., 2014, September. Fake tweet buster: a webtool to identify users promoting fake news on twitter. In Proceedings of the 25th ACM conference on Hypertext and social media (pp. 316-317). ACM.
- Sikora, R., 2015. A modified stacking ensemble machine learning algorithm using genetic algorithms. In Handbook of Research on Organizational Transformations through Big Data Analytics (pp. 43-53). IGI Global.
- Simons, G., 2018. Fake News: As the Problem or a Symptom of a Deeper Problem? Обрáz (pp. 33-44).
- Shu, K., Sliva, A., Wang, S., Tang, J. and Liu, H., 2017. Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1), pp.22-36.
- Valentini, G. and Masulli, F., 2002, May. Ensembles of learning machines. In Italian Workshop on Neural Nets (pp. 3-20). Springer, Berlin, Heidelberg.
- Vasu, N., Ang, B., Teo, T.A., Jayakumar, S., Raizal, M. and Ahuja, J., 2018. Fake news: National security in the post-truth era. RSIS.
- Vosoughi, Soroush, Mostafa'Neo Mohsenvand, and Deb Roy. "Rumor gauge: Predicting the veracity of rumors on Twitter." ACM Transactions on Knowledge Discovery from Data (TKDD) 11, no. 4 (2017): 50.
- Wang, W.Y., 2017. " liar, liar pants on fire": A new benchmark dataset for fake news detection. arXiv preprint arXiv:1705.00648.
- Yang, Z., Wang, C., Zhang, F., Zhang, Y. and Zhang, H., 2015, September. Emerging rumor identification for social media with hot topic detection. In Web Information System and Application Conference (WISA), 2015 12th (pp. 53-58). IEEE.
- Zhou, Z.H., 2012. Ensemble methods: foundations and algorithms. Chapman and Hall/CRC.
- Zhang, C. and Ma, Y. eds., 2012. Ensemble machine learning: methods and applications. Springer Science & Business Media.

# Synthetic Data Generation With Machine Learning for Network Intrusion Detection Systems

Marvin Newlin, Mark Reith and Mark DeYoung

Air Force Institute of Technology, Dayton, USA

[marvin.newlin@afit.edu](mailto:marvin.newlin@afit.edu)

[mark.reith@afit.edu](mailto:mark.reith@afit.edu)

[mark.deyoung@afit.edu](mailto:mark.deyoung@afit.edu)

**Abstract:** Machine learning is becoming an integral part of cybersecurity today, particularly in the area of network anomaly detection. However, machine learning techniques require large volumes of data to be effective. Although there are some datasets available for training Network Intrusion Detection Systems (NIDS), many of them are outdated or do not contain enough useful information for training/classification of NIDS. Therefore, generating synthetic data that is realistic is imperative for training effective intrusion detection systems. Currently, the most common methods for generating synthetic data are simulation or emulation through a software package like OPNET, and then machine learning is used to analyze the dataset for correctness. This paper argues for an approach to utilize machine learning to develop models in order to generate the datasets themselves for NIDS, which is an approach that is not commonly used. In this paper, we discuss some of the well-known available datasets, the features that make up a good dataset, the reasons for utilizing generative modeling to synthesize network data and lay out a basic approach to developing generative models for synthetic data by leveraging machine learning.

**Keywords:** synthetic data generation, network intrusion detection system, machine learning

---

## 1. Introduction

Network Intrusion Detection Systems (NIDS) utilize machine learning to detect anomalies and intrusions into network systems. As with all machine learning, datasets are a crucial aspect of the machine learning process for training, classification, and testing purposes (Abt and Baier, 2014a). NIDS are a critical aspect to network security to prevent systems from being infected by malware or subjected to attack. However, these systems all are negatively impacted by the fact that there are very few good and reliable datasets available for use (Sharafaldin, Habibi Lashkari and Ghorbani, 2018) (Małowidzki, Berezi and Mazur, 2017).

One block to the availability of datasets is the privacy issues that arise with utilizing real data. Anonymizing real data is difficult and often leads to only analysing parts of the network data that have been discovered, or removing the actual payload data from the traffic (Macía-Fernández *et al.*, 2017). Another block to dataset availability is the general difficulty in obtaining network data. This can be due to agreements set in place to prevent use of real data, which relates back to the privacy issues with real data utilization. Additionally, some types of network anomalies like certain types of malware do not exist or are very hard to find in real network data so using that data for anomaly detection may not work as intended (Ricks, Tague and Thuraisingham, 2018).

In this paper we examine several of the datasets that are commonly used and their deficiencies. We also discuss the aspects that a “good” data set provides as described by Gharib *et al.* (Gharib *et al.*, 2017) and others. We then discuss what datasets have been recently produced and the features they provide. We then discuss how to solve the problem of lack of datasets through with generative modelling and machine learning to develop synthetic datasets.

## 2. Available datasets

There are a handful of available commonly used datasets for NIDS. In this section, we analyse the currently available datasets and discuss their features and limitations.

- **DARPA (Haines *et al.*, 2001):** This was one of the first available datasets for intrusion detection and was developed in 1998. This dataset includes a variety of traffic types, including browsing, e-mail, FTP, and Telnet. It also contains several attacks such as buffer overflows, Rootkit, and DoS. The main problem with this dataset is that it is so old at this point that it does not contain data that is useful for NIDS evaluation today (Sharafaldin, Habibi Lashkari and Ghorbani, 2018).
- **KDD99 (KDD Cup 1999 Data, 1999):** KDD99 is a modified version of the DARPA dataset. It was generated by analysing the tcp-dump section of the DARPA dataset. As a result, it contains many of the same issues that

DARPA has but does include some new attacks like Neptune-dos, satan, rootkit, and buffer overflow. The main problem with this dataset is that it merges normal network traffic with the traffic from the attacks and so it suffers from a large number of redundant records (Gharib *et al.*, 2017).

- LBNL/ICSI Enterprise Trace (**LBNL/ICSI Enterprise Tracing Project - Project Overview, 2005**): This dataset is composed of network traffic from a medium sized network. The data does not contain payload information and was very heavily anonymized so that identification of IP addresses could not take place (Gharib *et al.*, 2017).
- CDX (US Military Academy 2009) (**Sangster *et al.*, 2009**): This dataset was generated from network warfare competitions that took place at the US Military Academy in 2009. It contains some normal network traffic such as email and DNS. It also contains traffic from exploits such as Nikto and WebScarab. However, due its data being generated from a competition, it suffers from a more limited range of activity than a normal network would have, i.e. it has too much malicious traffic and not enough normal, mundane traffic to be useful. (Sharafaldin, Habibi Lashkari and Ghorbani, 2018).
- CTU-13 (Czech Technical University 2011)(**García *et al.*, 2014**): This dataset was developed by researchers at the Czech Technical University who studied the signatures of 13 different types of malware from a real network and contained machines that were infected with the malware and machines that were not so that they could capture live traffic (Yavanoglu and Aydos, 2018).

### **3. Features of a “good” dataset**

Since we are discussing a lack of good datasets, we must define what makes up a good dataset. There a handful of works that have been published on the characteristics of a good dataset and they have several commonalities. They are: currency, comprehensive configuration, labels, traffic diversity, attack diversity, ground-truth, and anonymity (Sharafaldin, Habibi Lashkari and Ghorbani, 2018) (Małowidzki, Berezi and Mazur, 2017) (Gharib *et al.*, 2017) (Bowen *et al.*, 2016).

*Currency* is an important characteristic of a good dataset. Given the constantly evolving nature of the cyber world, an old dataset is not representative of the types of traffic both normal and malicious that is seen on today's networks (Małowidzki, Berezi and Mazur, 2017). This is also an important feature since the NIDS being developed need to be trained on traffic they are likely to encounter in a deployed scenario. *Comprehensive configuration* refers to a complete network setup. Since real networks all contain a diverse range of devices like PCs, servers, switches, routers, modems, and firewalls, a good dataset should contain traffic from these types of devices (Gharib *et al.*, 2017). *Labels* means that the data should contain classifiers built in so that it is easy to determine the type of traffic (e.g. normal or malicious) and if it is malicious what kind of traffic (botnet, malware, etc) (Sharafaldin, Habibi Lashkari and Ghorbani, 2018). This is an important requirement when dealing with machine learning. Having a labelled dataset means that we can more accurately and efficiently train machine learning algorithms in a supervised learning environment (James *et al.*, 2013, 26).

*Traffic diversity* means that there should be a range of protocol traffic contained within the dataset. This means HTTP(S), FTP, DNS, SMTP, SSH and other common protocols (Mon Divakaran *et al.*, 2016) (Gharib *et al.*, 2017). *Attack Diversity* is having a variety of malicious activity in the data so that the models can be trained on the signatures of many different attacks. Examples of these include DoS, Browser based, DNS, backdoors, and many other different kinds (Gharib *et al.*, 2017). Similar to currency, a wide range of attack vectors is important for NIDS training since they might be exposed to a variety of attacks when deployed in real world networks. *Ground-truth* means knowing the true positive and negatives contained within the data. Since any anomaly detection method is prone to false positives and negatives, it is imperative to know exactly what the issues within the data are (Bowen *et al.*, 2016). This is also important in the machine learning realm because this feature allows for assessment of model accuracy. Model accuracy is an important performance measure when dealing with machine learning algorithms. We can use time-to-accuracy, the runtime to reach a specified accuracy, as a performance measure to evaluate and improve our algorithm (James *et al.*, 2013, 29).

Finally, *anonymity* is an important aspect of the datasets because the majority of entities on the internet want to protect their privacy. Therefore, anonymity in the dataset means that the data doesn't identify any particular entity (Sharafaldin, Habibi Lashkari and Ghorbani, 2018). Maintaining the anonymity of a dataset allows us to extract the useful features and hopefully allow access to a larger amount of data. This is a key point in why generating synthetic data is useful (Abt and Baier, 2014b).

Additionally, it goes without saying that a dataset should contain both normal, legitimate traffic and malicious traffic. There are a couple of ways to accomplish this. *Correlated* traffic is network traffic where malicious and legitimate traffic are collected together. *Unrelated* traffic is where the malicious traffic is generated separately from the legitimate traffic and is then mixed in with the legitimate traffic (Małowidzki, Berezi and Mazur, 2017).

#### **4. Approaches for synthetic data generation**

There are two main approaches currently used for generating synthetic network data. The first is simulation and the second is emulation. Simulation involves leveraging network simulation software to generate the network traffic to be captured. Thus, this method uses all software to accomplish the data generation (Ricks, Tague and Thuraisingham, 2018). Emulation combines the software approach of simulation with the physical lab approach of setting up a multi-node network on the same host machine. This allows for generation of traffic that is almost the same as real world traffic but allows for the flexibility of simulation (Ricks, Tague and Thuraisingham, 2018) (Gharib *et al.*, 2017).

One problem that arises from straight simulation is an abstraction from real world elements. In order to completely simulate traffic, certain assumptions must be made that do not necessarily hold true for real data generated from physical machines. The advantage to emulation is that the traffic generated is more realistic while simulating the lower physical and link layers (Ricks, Tague and Thuraisingham, 2018).

One recent dataset generated by Sharafaldin *et al.* used the following approach. They set up a testbed network using emulation and then performed a series of different attacks at different times to generate the necessary variety of data. They then utilized several different feature extraction techniques to determine the most significant features of the generated data and then fed that into their model to predict the attack flows. They then evaluated their data with machine learning based on the evaluation framework established by Gharib *et al.* against the other data sets to show that it had all the necessary features (Sharafaldin, Habibi Lashkari and Ghorbani, 2018).

Ideally, rather than generating traffic from a simulation or emulation alone, we can utilize machine learning from big data to help us generate a model for realistic data which we can then use to train NIDS. To do this, we could take a small amount of data that would be generated through simulation/emulation or analyse real network data that has been made available. Then we would utilize a variety of machine learning techniques to extract the important features of the data and then based on those features, generate more data which we can then feed to the NIDS for training. From there we can continuously re-evaluate the data and feature selection to improve the accuracy of the generated data.

Implementing this model, although it sounds straightforward, is not an easy task. One of the main underlying ideas of synthetic data generation is called the Generative Adversarial Network (GAN) introduced in 2014 by Ian Goodfellow (Goodfellow *et al.*, 2014). The GAN is composed of a generator and a discriminator. The purpose of the generator is to analyse the data and then generate a synthetic output that mimics the input data. The output from the generator is then sent to the discriminator, whose job it is to determine if the input is real or if it came from the generator. The response from the discriminator is then fed back into the generator, which improves its generation based on the feedback from the discriminator. This game continues until the output from the generator converges and the discriminator cannot distinguish between the generator and the real data (Goodfellow *et al.*, 2014).

The original idea of GAN as proposed by Goodfellow *et al.* works well for continuous data (e.g. images or audio). However, when dealing with text or categorical data, it tends to not work as well, and network data like packet traces fall into the discrete category. To this end, some improvements to GAN have been made. The Wasserstein GAN (WGAN) was introduced in 2017 and utilizes a different distance measurement than the original GAN to measure the “distance” between outputs. This approach solves some of the divergence and training issues that the original GAN framework suffered from (Arjovsky, Chintala and Bottou, 2017). Some additional improvements have been made on the WGAN to improve its functionality on discrete data. *Improved Training of Wasserstein GAN’s* discusses the limitations that WGAN produces with weight clipping and demonstrate some of its flaws. To alleviate this problem, they propose a new method of a gradient penalty to maintain the 1-Lipschitz property of the Wasserstein GAN without the instability produced by weight clipping. In their evaluation of this method, they found that this improvement on WGAN helped improve its performance on discrete data.

It is important to utilize a variety of techniques within the feature selection because many attacks are hard to discover based on a single algorithm, so a variety helps in discovering the elements of an attack (Agrawal and Agrawal, 2015). Additionally, once the feature selection is complete, we would compare the generated data to the real data that we started with through machine learning. This comparison then allows us to readjust our features to make them more realistic before feeding them in the NIDS.

Outside of the GAN framework, an approach was successfully demonstrated for synthetically generating log data by Wurzenberger *et al.* in 2016. They utilized log line clustering and a Markov Chain approach to analyse the characteristics of real log files and then generate their own synthetic log files with those same characteristics. In their evaluation, they generated a synthetic network event sequence log and ran it through an anomaly detection system (ADS) called AECID (Automated Event Correlation for Incident Detection). They found that for the purposes of testing the ADS, the real log file and the synthetically generated log file performed nearly the same (Wurzenberger *et al.*, 2016). Ideally, this approach could be modified to work with actual network traffic so that we then have an alternate method of generating synthetic traffic other than simulation or emulation.

## **5. Conclusion and future work**

This paper addresses the problem of generating synthetic data for Network Intrusion Detection Systems. The problem arises from a lack of available “good” datasets that can be used for the machine learning tools for NIDS. Available datasets and their limitations are explored along with the features that make up a “good” dataset. The approaches for generating the synthetic data are then discussed along with the elements that are used to model and then generate the synthetic data, along with a discussion of an example where synthetic log data was successfully generated. In the future, work to improve on the algorithms for data analysis and the methods of determining the “closeness” of the synthetic data to the real data will be necessary for advanced research in this area.

## **6. Disclaimer**

The views expressed are those of the authors and do not necessarily reflect the official policy or position of the Air Force, the Department of Defense, or the U.S. Government.

## **References**

- Abt, S. and Baier, H. (2014a) ‘A Plea for Utilising Synthetic Data when Performing Machine Learning Based Cyber-Security Experiments’. doi: 10.1145/2666652.2666663.
- Abt, S. and Baier, H. (2014b) ‘A Plea for Utilising Synthetic Data when Performing Machine Learning Based Cyber-Security Experiments’, in *Proceedings of the 2014 Workshop on Artificial Intelligent and Security Workshop - AISeC ’14*. New York, New York, USA: ACM Press, pp. 37–45. doi: 10.1145/2666652.2666663.
- Agrawal, S. and Agrawal, J. (2015) ‘Survey on Anomaly Detection using Data Mining Techniques’, *Procedia - Procedia Computer Science*, 60, pp. 708–713. doi: 10.1016/j.procs.2015.08.220.
- Arjovsky, M., Chintala, S. and Bottou, L. (2017) ‘Wasserstein GAN’. doi: 10.2507/daaam.scibook.2010.27.
- Bowen, T. *et al.* (2016) ‘Enabling reproducible cyber research-four labeled datasets’, *Proceedings - IEEE Military Communications Conference MILCOM*, pp. 539–544. doi: 10.1109/MILCOM.2016.7795383.
- García, S. *et al.* (2014) ‘An empirical comparison of botnet detection methods’, *Computers and Security*, 45, pp. 100–123. doi: 10.1016/j.cose.2014.05.011.
- Gharib, A. *et al.* (2017) ‘An Evaluation Framework for Intrusion Detection Dataset’, *ICISS 2016 - 2016 International Conference on Information Science and Security*, (Cic), pp. 0–4. doi: 10.1109/ICISSEC.2016.7885840.
- Goodfellow, I. J. *et al.* (2014) *Generative Adversarial Nets*. Available at: <http://www.github.com/goodfeli/adversarial> (Accessed: 20 February 2019).
- Haines, J. W. *et al.* (2001) ‘1999 DARPA Intrusion Detection Evaluation: Design and Procedures’. Available at: <http://www.dtic.mil/docs/citations/ADA387747> (Accessed: 30 October 2018).
- James, G. *et al.* (2013) *An Introduction to Statistical Learning, Springer Texts in Statistics*. New York, NY, USA: Springer. doi: 10.1007/9781461471387.
- KDD Cup 1999 Data (1999). Available at: <http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html> (Accessed: 30 October 2018).
- BNL/ICSI Enterprise Tracing Project - Project Overview (2005). Available at: <https://www.icir.org/enterprise-tracing/Overview.html> (Accessed: 30 October 2018).
- Maciá-Fernández, G. *et al.* (2017) ‘UGR’16: A new dataset for the evaluation of cyclostationarity-based network IDSs’, *Computers & Security*, 73, pp. 411–424. doi: 10.1016/j.cose.2017.11.004.
- Małowidzki, M., Berezi, P. and Mazur, M. (2017) ‘Network Intrusion Detection : Half a Kingdom for a Good Dataset’, *ECCWS 2017 16th European Conference on Cyber Warfare and Security*, pp. 1–6. Available at: [https://www.wiaw.pl/art\\_prac/2015/Network\\_Intrusion\\_Detection.pdf](https://www.wiaw.pl/art_prac/2015/Network_Intrusion_Detection.pdf) (Accessed: 29 October 2018).

**Marvin Newlin, Mark Reith and Mark DeYoung**

- Mon Divakaran, D. et al. (2016) 'REGENT: A Framework for Realistic Generation of Network Traffic'. doi: 10.17706/IJCCE.
- Ricks, B., Tague, P. and Thuraisingham, B. (2018) 'Large-scale realistic network data generation on a budget', in *Proceedings - 2018 IEEE 19th International Conference on Information Reuse and Integration for Data Science, IRI 2018*, pp. 23–30. doi: 10.1109/IRI.2018.00012.
- Sangster, B. et al. (no date) *Toward Instrumenting Network Warfare Competitions to Generate Labeled Datasets*. Available at: [www.whitewolfsecurity.com](http://www.whitewolfsecurity.com) (Accessed: 30 October 2018).
- Sharafaldin, I., Habibi Lashkari, A. and Ghorbani, A. A. (2018) 'Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization', in *Proceedings of the 4th International Conference on Information Systems Security and Privacy*, pp. 108–116. doi: 10.5220/0006639801080116.
- Wurzenberger, M. et al. (2016) 'Complex log file synthesis for rapid sandbox-benchmarking of security- and computer network analysis tools', *Information Systems*. Elsevier Science Ltd., 60(C), pp. 13–33. doi: 10.1016/j.is.2016.02.006.
- Yavanoglu, O. and Aydos, M. (2018) 'A review on cyber security datasets for machine learning algorithms', in *Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017*. doi: 10.1109/BigData.2017.8258167.

# Human Factors of Cyber Operations: Decision Making Behind Advanced Persistence Threat Operations

**Veikko Siukonen**

**Finnish Defence Research Agency, Riihimäki, Finland**

[veikko.siukonen@mil.fi](mailto:veikko.siukonen@mil.fi)

**Abstract:** Cybersecurity headlines have been dominated by cyber espionage operations conducted by Advanced Persistent Threat (APT) groups. The rising volume and sophistication of attacks have given researchers more data to analyze these operations, but the reports and studies tend to focus on technical aspects. The indicators, evidence, and forensics of cyber espionage are of course relevant, but they lack a broader understanding of the procedures behind the attacks — Humans make decisions in cyber espionage operations. The research on Advanced Persistent Threat (APT) operations fails to factor in the human, focusing their analysis on technical aspects. This paper identifies and dissects the decision-making process behind APT-operations. The purpose is to widen the perspective from a technical point of view to a more comprehensive understanding of the adversary's operation as a whole. The Intrusion Kill Chain (IKC), introduced by Lockheed Martin Corp., is used as a theoretical framework for APT operations and its phases. IKC defines seven phases in a cyber attack, from early reconnaissance to actions on the objective. The decision-making process is analyzed through John Boyd's OODA loop framework (Observation, Orientation, Decision, and Action). The decisive points and decisions that adversary needs to make in different phases of APT-operation are deduced from publicly available reports. This paper suggests that by understanding the adversary's decision-making process and identifying the decisive points in the different phases of an APT-operation, a defender can improve their capabilities to recognize and react to ongoing attacks. The ability to understand the adversary as a decision maker enhances preventive or reactive courses of action.

**Keywords:** cyber and information campaigns, advanced persistent threat, cyber kill chain, decision-making process, OODA loop

## 1. Introduction

The number of cyber attacks, sophistication, and effects have increased significantly in recent years. In particular, Advanced Persistent Threat (APT) have become one of the most significant security threats for organizations and individuals. (Gardiner, Cova & Sishir 2014, 3; Lemay, Calvet, Menet & Fernandez 2017).

An APT is an advanced long-term threat, where the term advanced refers not only to the use of sophisticated attack methods and malware but also to the context. Persistent refers to the longevity of the attack operation, its persistence. To be labeled an APT, the threat requires a high level of expertise both in the development of the necessary malware and delivery methods and in the planning and implementation of the operation. An APT attack is characterized by the fact that they perform intrusions in succession, developing their operations with successes and failures until they finally reach their goals. An attack on a particular organization may take months to several years. APT groups prioritize concealment to a high degree in all stages of their operations and select the target organizations for their attacks carefully. (Bhatt, Yano & Gustavsson 2014, p. 390; Hutchins, Clopperty & Amin 2010, pp. 1-2; Gardiner, Cova & Sishir, 2014, p. 5)

Known APT operations include, for example, an attack that utilized the Stuxnet malware to slow down Iran's nuclear program in 2010 (see Symantec 2013), an operation directed at the Ukrainian power grid in December 2015 (see SANS 2016), and government-sponsored Russian hacking groups implicated in the breach of Democratic party affiliated actors during the 2016 presidential election in the United States (see US CERT 2016). All of these operations required long-term preparation, high-level of expertise, and significant resources. The operations show how the purpose of an APT operation may not only be silent intelligence gathering but also to create physical impacts on a target.

APT attack operations are often conducted by large groups, sometimes sponsored by a nation state and have significant resources. It is challenging to attribute ATP attacks to specific actors because the attackers try to conceal their actions from start to finish. Attempts to identify an attacker include gathering evidence of the language used to write malicious code, metadata of malware, time stamps for program compilation, identification of attack background motivation, the target of the attack, used IP addresses, and server infrastructure. (FireEye 2017). Published analyses are technical by nature and do not look at the human factor affecting the operations: the decision-making process. Phases of the decision-making process occur in cognitive,

information and physical domains (Alberts, Garstka, Hayes, & Signori 2001, p. 132). It is important to notice that the technical indicators address the physical domain and therefore cognitive domain is often disregarded.

APT groups are continually changing their behavioral patterns and attack methods to prevent disclosure (Ussath, Jaeger, Feng, & Meinel 2016, p. 1). Also, APT groups develop operational security by utilizing reports of revealed operations. If an attack is not detected, it is impossible to react to them. Preventive measures, detection, response and recovery capabilities are required to protect against APT attacks, so in essence, all protection attempts require the ability to understand and model the threat. Typically, organizations respond to identified threats with both technical and administrative solutions.

This paper is a summary the authors master's thesis. Thesis examines the APT attack operation from the decision-making process. Exploring published and technically analyzed operations from a decision-making perspective increases understanding of the requirements to conduct APT attacks. By recognizing these operational requirements and identifying the steps in the decision-making process at the various stages of the operation, measures may be developed to protect against attacks, detect them, react to them in time, and recover from attacks. Also, the management perspective makes it possible to identify and examine human factors at different stages of the attack.

In this paper, the seven-step Intrusion Kill Chain (IKC) model presented by Hutchins, Clopperty & Amin (2010) has been selected to model the APT attack. The model has been widely used in various studies (see e.g. Bhatt et al. 2014), in technology companies (see e.g. Microsoft 2014) and governmental organizations (see e.g. The Department of Homeland Security National Cybersecurity and Communications Integration Center 2017). The use of the model is based on systematic gathering and analysis of threat information.

The decision-making process is viewed through the stages of the OODA loop, presented by John Boyd (1927-1997). The sub-section on the OODA loop is complemented by the presentation of Mica Endsley's (1995) Situation Awareness (SA) theory, as a high-quality SA is essential to the decision-making process.

This paper is a qualitative study that explains the decision-making factors at different stages of the APT attack, using theory-driven content analysis. The purpose of the research is to find out how and where the decision-making processes occur in an APT operation. Decision-making during the APT operation is considered to be bound to the stages of the OODA loop from the point of view of the attacker. The main research question is: How do the phases of the OODA loop occur at different stages of the APT attack?

## **2. Background**

The decision-making process of an APT attack is analyzed in this article. The progress of the attack is modeled via the IKC and the progress of decision-making process during the attack is modeled by the use of the OODA - loop. These shall be introduced below.

### **2.1 Intrusion kill chain model**

Several different modeling techniques have been developed to examine APT attack operations, such as Attack Graph (Phillips & Swiler 1998; Meicong Li, Huang, Wang & Fan, 2016), Diamond Model (Caltagirone, Pendergast & Betz 2013), and Intrusion Kill Chain (Hutchins, Clopperty & Amin 2010). Also, several different security companies have developed their models for describing APT attacks. For example, the F-Secure model (see F-Secure 2015) divides the attack into four stages. Although there are several alternatives, their ultimate goal is the same: divide the operation into smaller successive phases, identify their characteristics and possible indicators to detect and respond to an attack.

The Intrusion Kill Chain model is based on the principle of succession, requiring the previous step to be successful before the next one can begin. By identifying and interrupting this chain at any stage of an attack, the attacker is prevented from moving to the next stage of the operation and thus achieving the goal of the attack. The kill chain methodology is based on U.S Department of Defense targeting doctrine that defines the steps of this process as find, fix, track, target, engage, assess (U.S. Department of Defense 2007).

The goal of the IKC model is to identify and interrupt the attack as early as possible. Identifying, analyzing and tracking attack-specific indicators at different stages of an attack is of utmost importance. An indicator is any

information that refers to an intrusion attempt or its preparation. The phases of the model are shown in Figure 1.



**Figure 1:** Seven phases of Intrusion Kill Chain (Hutchins, Clopperty & Amin 2010, p. 4)

In the *reconnaissance* phase, the attacker plans the mission. The attacker collects as much information as possible from the target organization to determine which objects and vulnerabilities it may use to enter into the target system.

The goal of the *weaponization* phase is to create malware that enables the attacker to gain a foothold in the target system. Typically it exploits some vulnerability in the target system.

The execution of the operation starts at the *delivery* phase. Its goal is to get the malicious program delivered into the target system. Examples of delivery methods include email attachments, web pages, or removable storage media.

After the malware is delivered to the target system, *exploitation* triggers the attacker's code. Most often, exploitation targets an application or operating system vulnerability, but it could also more simply exploit the users themselves or leverage an operating system feature that auto-executes code.

In the *installation* phase, malware is installed in the target system. Malware often includes a hidden backdoor or a rootkit program that allow an attacker to access and operate the target system. Installation of malware on the target system allows the adversary to maintain presence inside the environment.

In the *command and control* (C2) phase, the attacker establishes a C2 channel to the target system. Often, the malware contacts the attacker-configured C2 server after the target system is compromised. Malware enables an attacker to have persistent "hands on the keyboard" access to the target network.

*Action on objective* means that the attacker has completed the first six phases and the attacker's actions have not been revealed to the target organization. The attacker is now able to operate in the target system to reach its objectives. Typically, this objective is data exfiltration which involves collecting, encrypting and extracting information from the victim environment.

## 2.2 OODA loop

The importance of decision-making grows the greater and more complex an operation becomes. APT operations are extensive, long-term, multi-stage, and often require significant resources. In this study, the decision-making process is described via Boyd's OODA decision-making cycle.

Planning and executing a mission requires leadership that necessitates decision-making. Decision-making relies on humans and is linked to the organization's management system, which consists of people, equipment, messaging facilities, management infrastructure, and operating model. The functions of the management system are enabled by the organization's information system. An information system consists of people, data processing equipment, data transfer devices and software (Alberts & Hayes 2006, p. 47). It should be recognized, that decision-making is carried out within the organizational management system, which defines the power and responsibilities of the organization's actors.

The OODA decision-making cycle is complemented by Mica Endsley's (1995) theory of understanding, where Endsley defines high-quality understanding central to the decision-making process. Endsley's theory seeks to explain what aspects affect the understanding of a situation.

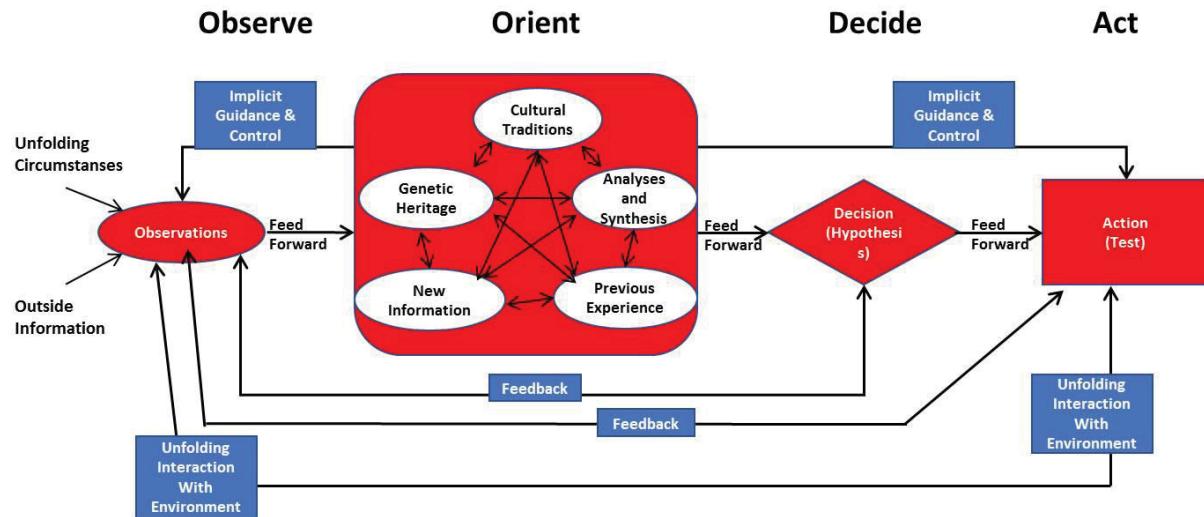
The OODA loop has been widely used as a framework for leadership and decision-making both in military doctrines and in the corporate world. (Grant & Kooter 2005, p. 1; Osinga 2007, 3-5) The use of the OODA loop

in cyber warfare and military operations had been studied and estimated to be suitable for the context (Lehto 2017, 18).

Criticism of Boyd's theory (e.g., Alberts et al. 2001; Osinga 2007, 5-6) is often directed toward the simplistic decision-making cycle, and to the fact that Boyd did not publish actual academic research on his theory.

The underlying science to Boyd's theory has been studied by Frans Osinga (Osinga 2007) and looking deeper into Boyd's decision-making model, and detailed breakdown of the phases, the criticism of the simplistic nature of the decision-making process does not seem to hold water. The four-stage decision-making cycle is a comprehensive, pragmatic and inclusive model that takes into account both the context-related and the human-centered dimension of the decision-making process.

Figure 2 shows Boyd's cybernetic double loop based on the graphic presentation of *The Essence of Winning and Losing* (1995, p. 4).



**Figure 2:** John Boyd's OODA loop. (Boyd 1995, p. 4)

The decision-making model, OODA loop, consists of four successive steps: Observation, Orientation, Decision and Action (Boyd 1995, p. 4). The initiative remains with the entity that can make and execute their decisions faster than the opponent. (Brehmer 2005, pp. 2-3; Osinga 2007, pp. 1-2).

In the observation phase, information about the operating environment is collected using available resources, consisting of people, systems, and methods available for data collection. (Boyd 1995, p. 4; Osinga 2007, pp. 230-233) In the orientation phase, analysis and synthesis of the current situation are prepared as the basis for decision making. Analysis and synthesis of the prevailing situation arise through interpretation. The purpose of the orientation phase is to produce a reasoned analysis of the current situation as the basis for decision-making, and as such the importance of the orientation phase for other stages of the process is central. (Boyd 1995, p. 4; Osinga 2007, pp. 230-233) At the decision-making phase, the most justified and appropriate measure for the current situation is chosen to be implemented. The decision on the action provides the basis for action. (Boyd 1995, p. 4; Osinga 2007, pp. 230-233) The action decided in the decision phase is either implemented as such or tested for its functionality. After the act-phase, the loop continues back to the observation phase in order to observe the impact of the actions in the operating (Boyd 1995, p. 4; Osinga 2007, pp. 230-233).

The theory of Boyd's decision-making process has been developed in the context of military leadership and decision-making. Implementing a decision-making cycle that is faster than the opponent guarantees the superiority of information over the opponent and enables the achievement of the mission's goals and defeating the opponent. Complex organizations may have multiple OODA loops running simultaneously with lower level loops usually working faster than the upper level. (Endsley & Jones 1997, pp. 11-12)

### 2.2.1 Situation awareness in decision-making

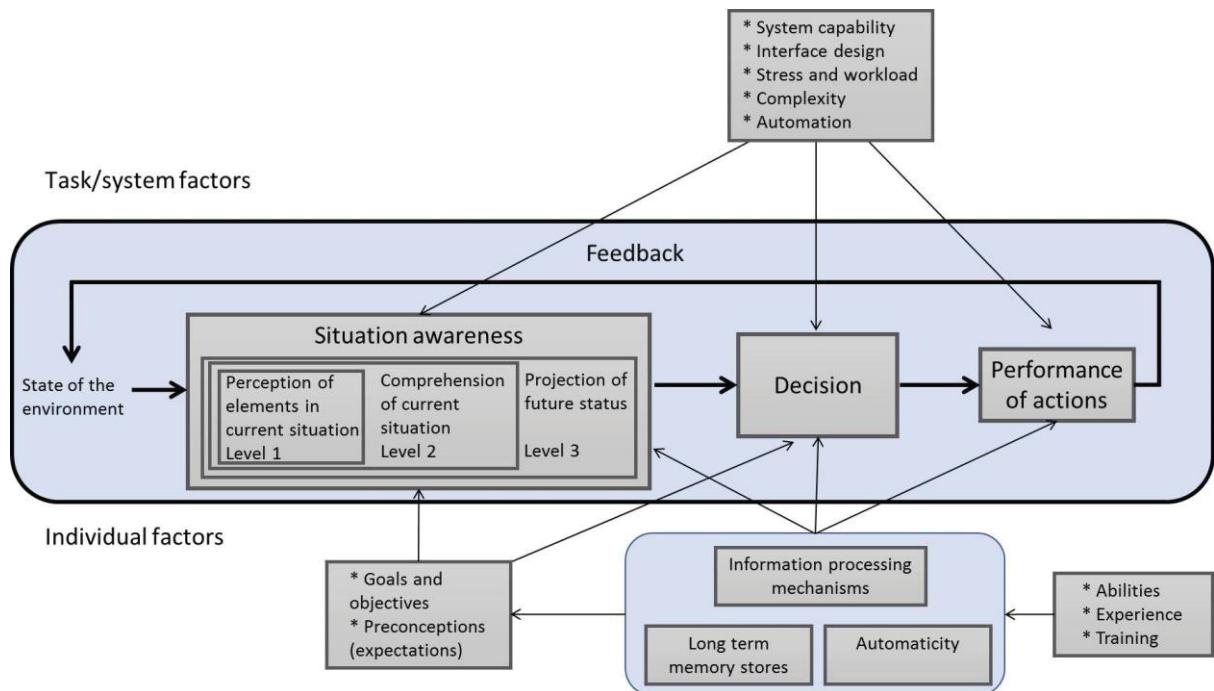
In the orientation phase, the criteria for decision-making are created as a result of observations collected from the operating environment and the subsequent analyses conducted based on these observations. Current situational picture, situation awareness (SA) and understanding the overall situation create conditions for decision-making (Endsley 1995).

Understanding of the situation is based on both the experience and the knowledge base of the interpreter and the information obtained from the current situation. It is strongly linked to the ability of the interpreter to interpret the situation. The understanding also includes the ability to identify and control the element of uncertainty. (Alberts et al. 2001, p. 139)

The concepts of SA and understanding are closely related and are often understood as identical. Alberts et al. (2001, 19) have compared situation awareness and understanding of the situation in a military environment. They state that situational awareness focuses on what is known about the past and the present, while perception is about how the situation is or can develop, and how different activities affect the evolving situation. In this comparison, situational awareness is strongly future-oriented.

According to Endsley's SA theory, (1995), situation awareness is the perception of the relevant elements around, in terms of time and place, while understanding is their meaning in the prevailing context, and the ability to assess the situation in the near future. At the first level, the actor deals with information relevant to the situation. On the second level, the operator combines the information obtained with the given task, objectives and context. On the third level, the actor can utilize all the information and evaluate the development of the situation and future activities. (Endsley 1995, pp. 36-37).

The model of situation awareness in a dynamic decision-making process is shown in Figure 3. The model takes into account both individual and systemic factors.



**Figure 3:** Model of situation awareness in dynamic decision making (Endsley 1995, pp. 35)

There are many similarities with the orientation phase of the OODA loop and the SA model. Both emphasize context-relatedness and recognize the central role of the individual in creating understanding about the situation. A high-quality SA allows an actor to make and implement the right decisions for the operation. At the same time, a high-quality SA requires sufficient skills, training, and experience from the individual.

How these models can be combined and utilized in APT context shall be described in the next chapter.

### 3. APT described by the combination of IKC and OODA

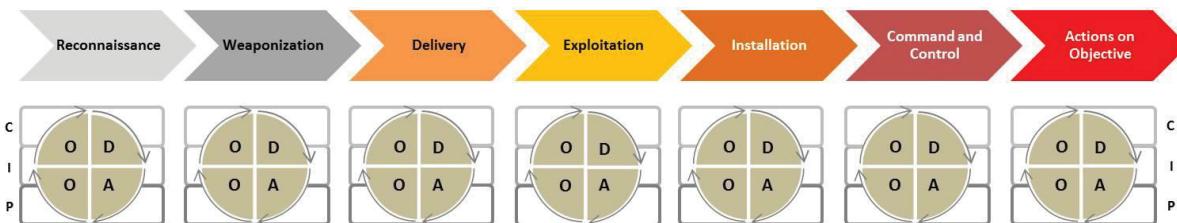
This paper posits that the steps of the OODA loop are executed in every phase of an APT operation. The review can be carried out from the perspective of the organization behind the APT-operation and the perspective of the individuals involved. In this study, the perspective is limited to the operating organization's perspective.

As described in the previous chapter, the orientation phase creates the criteria for decision-making. Interpreting the information in the orientation phase in the context of false or incomplete SA will inevitably lead to a bad outcome. For the APT operation, this would mean unmasking and interrupting the operation before it reaches its goal.

The leadership and decision-making of a challenging and multi-stage APT operation must be based on a high-quality understanding of the situation. If decisions are made by an incomplete or inaccurate understanding of the situation, the probability of the operation being revealed and interrupted is increased. Therefore, the personnel involved in the operation must have sufficient skills, training, and experience to create a high-quality understanding of the environment. Personnel must not rely only on operational capability. Creating an understanding of the situation is a continuous process, and it occurs at all stages of an APT attack.

Each phase of an APT attack requires a review of the steps in the decision-making process. In this study, the implementation of the steps of the APT attack is examined through the four stages of the OODA loop. The objective of the review is to identify the key points of decision-making and decision-making.

Figure 4 depicts the framework combining the IKC and OODA models.



**Figure 4:** Model that combines Intrusion kill chain model and OODA loop

The definition for the acronyms shown under each attack phase on Figure 4 are as follows: left upper corner Orientate (O), right upper corner Decide (D), right bottom corner Action (A), left bottom corner Observe (O). On the left and right side of the figure acronyms are Cognitive (C), Information (I) and Physical (P). Orientation and decision phases take place mainly in the cognitive domain, while the observation and action phase appears within the information and the physical domain (Alberts et al. 2001, p. 132).

### 4. Results: Analysis of APT-operations

This study is qualitative research that seeks to respond to research tasks and questions through theoretical content analysis. The nature of the study is exploratory. The different phases of the APT attack and the decision-making process are under review. Decision-making is viewed from the perspective of the attacker through the stages of the OODA loop. The purpose of this chapter is to describe the implementation of the research and the method used and to justify why these methods have been selected.

The nature of the research subject justifies the choice of an indirect, qualitative research strategy and a constructivist paradigm. APT-threat groups and their decision-making process cannot be directly approached, as this would require direct participation in an ongoing APT-operation. There are no established manuals that guide the APT-group's activities. Due to their covert and criminal nature, a direct approach is not possible.

The steps of the APT attack and the related decision-making process can be approached indirectly. In this study, the review focuses on reports of actual operations. The research material consists of detected and reported attacks, so the phenomenon is analyzed through interpretation. The selection of the reports for analysis is based upon the following criteria:

- The group and the operations it has performed meet the characteristics of an APT operation.
- The APT group and its operations have been reported and analyzed by several security companies.

- The reports used are sufficiently comprehensive and contain meaningful content from the theoretical framework.

The data for analysis was collected using a previous survey (Lemay et al. 2017) and supplemented with advanced information search on the APT28 group. The final data for analysis consisted of the following reports published by security companies:

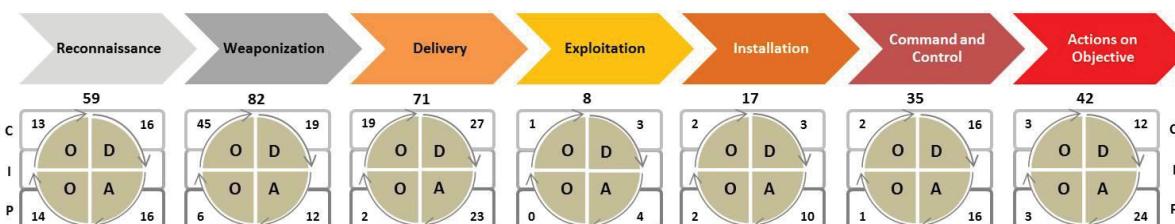
- APT28: a window Into Russia's Cyber-espionage operations? (FireEye 2014).
- Operation pawn storm using decoys to evade detection (TrendMicro 2014).
- Microsoft security intelligence report volume 19 (Microsoft 2015).
- APT28 under the scope of a journey into exfiltrating intelligence and government information (Bitdefender 2015).

In this study, theory-driven content analysis was selected as the analytical method. From the material, a set of 302 reduced expressions, or phrases, was identified, and they were selected as units of analysis. These units were then clustered into the IKC phases, classified under the OODA-loop in identified classes, and finally compiled into a numerical presentation that combined the clustering and classification.

The formation of classes was guided by both material and research questions for theoretical analysis. Research questions had a significant impact on the emergence of classes as they combine research theory and material. The formation of classes was guided by the abductive logic of reasoning.

In clustering the units of analysis, the APT attack phases were used. The simplified expressions were grouped under the attack phase that was best described by the expression. The insertion of the phrase into the attack phase was not always unambiguous, and in some exceptional cases, the expression was associated with up to three different phases.

The final classification of the material was done by identifying the simplified expressions within each step of the attack to the stages of the OODA loop. In this way, the material was structured into a meaningful form. The results of the analysis are shown in Figure 5. The figure shows how many reduced expressions in each step of the attack have been identified and how they are divided into four phases of the decision-making process.



**Figure 5:** A graphic depiction of the results of the analysis.

The number of reduced expressions in each phase of the attack is shown in the figure: reconnaissance (59), weaponization (82), delivery (71), exploitation (8), installation (17), command and control (35) and actions on objective (42). The first three and last two phases of the IKC seem to be the most visible and vital phases according to the quantity. The numerical review provides an indicative picture of how the expressions are divided into different stages of the attack and how the steps in the decision-making process occur.

One obtains how the quantity (45) of the expressions in the orientation in weaponization (82) phase is substantial (see Figure 5). Consequently, the importance of orientation is emphasized when compared with other phases. The mission's objective, the information about the target, and the organization's knowledge and skills determine the organization's ability to execute its operation. According to the observed material the organization's efficiency consists of the expertise and experience of its individuals, as well as the sufficiency of the organization itself.

The impact of the orientation on weaponization phase on future phases is evident. During this phase the characteristics of the malware, delivery method, command, and control infrastructure, and encryption functionality are considered. Coordination of activities between different organizations and groups was also identifiable during the orientation phase.

The main result of the review is that: all stages of decision-making take place at each stage of the attack. It is particularly interesting to look at the orientation and decision phases of the decision-making process because they take place in the attacker's cognitive dimension. One could argue that the first three phases of the attack are especially important, and that various type of information can be extracted to understand APT groups more comprehensively.

The results also indicate, that an attacker is prone to rely more heavily on the cognitive domain than the physical domain in the phase of weaponization, and to a lesser extent in the phase of delivery. This is a non-trivial observation, as weaponization is arguably usually considered as a mainly technical task that fits an attack method into vulnerability.

Actions taken in the physical domain can be observed by the defender. The actions reflect the decisions that the attacker have made during each phase of the attack. Information about the orientation behind the decisions can be extracted from the reports concerning the attacks trough interpretation.

## **5. Conclusion and discussion**

This study argues, based on the analysis of the decision-making process and intrusion kill chain method, that it is possible to identify the human factors of cyber operation. Each phase of an APT operation requires decision making, and humans make these decisions. This is self-evident, but by examining the backgrounds of a decision-making process and applying it into an APT-operation, it is possible to gain understanding about the attacker's cognitive domain.

Action and observation occur in the physical domain, and different analytical and protective frameworks, such as the IKC, rely on observing indicators in this realm. However, these models are nearly blind to any cognitive factors in the attacker's decision-making process.

This paper suggests that by understanding the adversary's decision-making process and identifying the decisive points in the different phases of an APT-operation, a defender can improve their capabilities to recognize and react to ongoing attacks. The ability to understand the adversary as a decision maker enhances preventive and reactive courses of action. The results also raise the question, if the current protective measures are cost-effective. Is it possible to affect the cognitive domain of the attacker more easily, than focus on the physical domain by hardening the defending cyber infrastructure? Multi-layered cyber-defense systems could benefit from understanding the adversary further from a decision-making point of view, and exploit detectable weaknesses.

This study has demonstrated a framework that combines Intrusion Kill Chain model and OODA loop. This framework allows the defender to go beyond observations of specific indicators of the APT-operation in the physical domain. By using this framework the defender can extract information from the adversary's cognitive domain. A better understanding of the adversary's decision-making process in the cognitive domain can provide a chance for the defender to take back the initiative and disturb the attacker's decision-making process.

One possible future course of action could be to develop deception methods based on the results of this study. Is it possible to dissuade the attack from continuing by introducing something else than an impenetrable wall? Is it possible to lead the attacker to conduct wrong decisions by implementing protective measures that are directed against attacker cognitive domain? In order to address these challenges more study is required. In practice, this could involve methods of not just preventing exploitation, but also reliable means of detecting and affecting attacker's cognitive weaknesses.

## **References**

- Alberts, D., Hayes, R., (2006) *Understanding Command and Control*. The Command and Control Research Program (CCRP) series. [online] [www.dodccrp.org/files/Alberts\\_UC2.pdf](http://www.dodccrp.org/files/Alberts_UC2.pdf) [Accessed 8 December 2016].
- Alberts, D., Garstka, R., Hayes, R., Signori, D., (2001) *Understanding Information Age Warfare*. The Command and Control Research Program (CCRP) series.
- Bhatt P., Yano E. & Gustavsson P. (2014) "Towards a Framework to Detect Multi-Stage Advanced Persistent Threats Attacks", *8th International Symposium on Service-Oriented System Engineering*. pp. 390-395.
- Boyd, J.R. (1995) The Essence of Winning and Losing. [Online], [www.danford.net/boyd/essence.html](http://www.danford.net/boyd/essence.html) [Accessed 9 October 2017].

**Veikko Siukonen**

- Brehmer, B. (2005) "The Dynamic OODA Loop: Amalgamating Boyd's OODA Loop and the Cybernetic Approach to Command and Control" *10th, International Command and Control Research and Technology Symposium*. [Online], [www.dodccrp.org/events/10th\\_ICCRTS/CD/papers/365.pdf](http://www.dodccrp.org/events/10th_ICCRTS/CD/papers/365.pdf) [Accessed 7 October 2016]
- Endsley, M.R. (1995) "Toward a Theory of Situation Awareness in Dynamic Systems". *Human Factors*, Vol. 37, No. 1, pp. 32–64.
- Endsley, M. R. & Jones, W. M. (1997). *Situation Awareness, Information Dominance & Information Warfare*. Belmont, MA: Endsley Consulting.
- FireEye (2017) "Advanced Persistent Threat Groups", [Online], <https://www.fireeye.com/current-threats/apt-groups.html> [Accessed 4 July 2017].
- Gardiner, J., Cova, M. & Sishir, N. (2014) "Command & Control: Understanding, Denying and Detecting", [Online], <https://arxiv.org/ftp/arxiv/papers/1408/1408.1136.pdf> [Accessed 10 October 2017].
- Hutchins, E., Clopperty, M. & Amin, R. (2010) "Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains", [Online], Lockheed Martin Corporation, <https://www.lockheedmartin.com/content/dam/lockheed/data/corporate/documents/LM-White-Paper-Intel-Driven-Defense.pdf> [Accessed 26 July 2017].
- Lemay, A., Calvet, J., Menet F. & Fernandez, J. (2017) "Survey of publicly available reports on advanced persistent threat actors" *Computers and Security* Vol. 72, pp. 26-59.
- Lehto, M. (2018). "The Modern Strategies in the Cyber Warfare" In *Cyber Security: Power and Technology*. Springer, Jyväskylä.
- Microsoft (2016). "Microsoft Cloud Red Teaming", [Online], <https://gallery.technet.microsoft.com/Cloud-Red-Teaming-b837392e> [Accessed 16 August 2017].
- Osinga, F. (2007) *Science, Strategy and War: The Strategic Theory of John Boyd*, Routledge, London and New York.
- Symantec (2013). "Stuxnet 0.5: The Missing Link", [Online], [http://www.symantec.com/content/en/us/enterprise/media/security\\_response/whitepapers/stuxnet\\_0\\_5\\_the\\_missing\\_link.pdf](http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/stuxnet_0_5_the_missing_link.pdf) [Accessed 6 September 2017]
- SANS Industrial Control Systems Security (2016) "Analysis of the Cyber Attack on the Ukrainian Power Grid", [Online], [https://ics.sans.org/media/E-ISAC\\_SANS\\_Ukraine\\_DUC\\_5.pdf](https://ics.sans.org/media/E-ISAC_SANS_Ukraine_DUC_5.pdf) [Accessed 1 Septemper 2017]
- US CERT, (2016) "Joint analysis report: RIZZLY STEPPE – Russian Malicious Cyber Activity", [Online], [https://www.us-cert.gov/sites/default/files/publications/JAR\\_16-20296A\\_GRIZZLY%20STEPPE-2016-1229.pdf](https://www.us-cert.gov/sites/default/files/publications/JAR_16-20296A_GRIZZLY%20STEPPE-2016-1229.pdf) [Accessed 6 September 2017]
- US-CERT (2017) "The Department of Homeland Security: National Cybersecurity and Communications Integration Center", [Online], [https://ics-cert.us-cert.gov/sites/default/files/documents/Destructive\\_Malware\\_White\\_Paper\\_S508C.pdf](https://ics-cert.us-cert.gov/sites/default/files/documents/Destructive_Malware_White_Paper_S508C.pdf) [Accessed 1 September 2017]
- U.S. Department of Defense. (2007) "Joint Publication 3-60 Joint Targeting", [Online], [https://www.aclu.org/files/dronefoia/dod/drone\\_dod\\_jp3\\_60.pdf](https://www.aclu.org/files/dronefoia/dod/drone_dod_jp3_60.pdf) [Accessed 6 Septemper 2017]
- Ussath, M., Jaeger, D., Feng, C., & Meinel, C. (2016) "Advanced persistent threats: Behind the scenes", *Annual Conference on Information Science and Systems*, ss. 181-186.

# **Non Academic Papers**



# Thoughts About a General Theory of Influence in a DIME/PMESII/ASCOP/IRC2 Model

Thorsten Kodalle<sup>1</sup>, Char Sample<sup>2</sup>, David Ormrod<sup>3</sup> and Keith Scott<sup>4</sup>

<sup>1</sup>Command and Staff College of the German Armed Forces, Germany

<sup>2</sup>ICF Inc., Columbia, USA

<sup>3</sup>University of New South Wales, Australia

<sup>4</sup>De Montfort University, UK

[thorstenkodalle@bundeswehr.org](mailto:thorstenkodalle@bundeswehr.org)

[charsample50@gmail.com](mailto:charsample50@gmail.com)

[drdave@linux.com](mailto:drdave@linux.com)

[jkscott@dmu.ac.uk](mailto:jkscott@dmu.ac.uk)

**Abstract:** The leading question of this paper is: "How would influence warfare ("iWar") work and how can we simulate it?" The paper discusses foundational aspects of a theory and model of influence warfare by discussing a framework built along the DIME/PMESII/ASCOP dimension forming a prism with three axes. The DIME concept groups the many instruments of power a nation state can muster into four categories: Diplomacy, Information, Military and Economy. PMESII describes the operational environment in six domains: Political, Military, Economic, Social, Information and Infrastructure. ASCOP is used in counter insurgency (COIN) environments to analyze the cultural and human environment (aka the "human terrain") and encompasses Areas, Structures, Capabilities, Organization, People and Events. In addition, the model reflects about aspects of information collection requirements (ICR) and information capabilities requirements (ICR) - hence DIME/PMESII/ASCOP/IRC2. This model was developed from an influence wargame that was conducted in October 2018. This paper introduces basic methodical questions around model building in general and puts a special focus on building a framework for the problem space of influence/information/hybrid warfare takes its shape in. The article tries to describe mechanisms and principles in the information/influence space using cross discipline terminology (e.g. physics, chemistry and literature). On a more advanced level this article contributes to the Human, Social, Culture, Behavior (HSCB) models and community. One goal is to establish an academic, multinational and whole of government influence wargamer community. This paper introduces the idea of the perception field understood as a molecule of a story or narrative that influences an observer. This molecule can be drawn as a selection of vectors that can be built inside the DIME/PMESII/ASCOP prism. Each vector can be influenced by a shielding or shaping action. These ideas were explored in this influence wargame.

**Keywords:** modeling and simulation, M&S, DIME, PMESII, ASCOP, wargaming, serious gaming, gamification, information warfare, influence warfare, weaponization of everything, perception

---

## 1. Preface

We live in a world where people are dying due to the return of measles (The Economist 2018) after the spread of misinformation in social networks about the dangers of vaccinations (Godlee et al. 2017). A growing group of people believe the world is flat (Burdick 2018). Grounded in the scientific method, this article is a contribution to the Human, Social, Culture, Behavior (HSCB) models targeting the HSCB community (Dean S. Hartley III 2018). A goal of this paper is an academic, multinational and whole of government influence wargaming community in order to develop the understanding and capabilities of influence and counter-influence activities across that community. On an intermediate level "perhaps the most basic purpose of a model is to increase the user's comprehension of something" (Dean S. Hartley III 2018). In essence "all models are wrong, but some are useful" (Box 1976).

## 2. Introduction

### 2.1 The scientific method

We are using the scientific method to tackle the problem and phenomenon of "influence" (and particularly such related concepts as "weaponized information" (Pomerantsev, P. and Weiss, M. 2014), "fake news" (Sample and Justice, Connie & Darraj, Emily 2018) and "cognitive hacking" (Cybenko et al. 2002)) in unconventional warfare/information warfare/hybrid warfare/operations in grey zones. The basic investigative framework (as described by North (2018)) is as follows:

- We observe the world around us (especially the phenomenon of hybrid warfare, information operations and the weaponization of everything))
- We ask a question about what we see (how would influence warfare work and how can we simulate it?)
- We construct a hypothesis that could answer our question (e.g. we write an influence warfare concept)
- We think of a way to test our hypothesis (e.g. construct an influence wargame prototype)

- We run experiments to see if our hypothesis's prediction was correct (e.g. we play the wargame)
- We draw a conclusion from the experiment (e.g. we conduct an after-action review (AAR) after playing the wargame)
- We communicate our results (e.g. we write this article)
- We refine, alter or reject our hypothesis (e.g. we are developing version 2.0)

As an observer of the world we need to be aware of the mental model we are using for filtering and categorizing our observations and avoid several pitfalls. We will evaluate our observations either implicitly with biases, and we might be aware of the limitations of thinking fast and thinking slow about a problem (Kahneman 2012). But even if we read Daniel Kahneman, there are still several mistakes we can make and fallacies we might commit as outlined by Michael Shermer (Shermer 2002). This is an old problem that haunts philosophy for over two millennia. One of the most recent books on skeptical and critical thinking skills (Novella et al. 2018) is supported by a more than 13-year-old skeptical podcast with over 630 episodes (Bellucci 2018). We must consider the role of cultural biases. (Fiske and Taylor 2013) noted that analysts can detect others' biases but are blind to their own. This project sought to mitigate some biases by using a multi-national team. While the team consisted of participants from seven Western nations, we explicitly questioned and examined potential bias to address it. For this first version of the game we acknowledge the existence of our western biases (Hofstede et al. 2010; Nisbett 2003), but due to temporal and fiscal restraints we are unable to fully mitigate these biases.

## **2.2 Modelling the world**

We seek to choose an explicit way of thinking and we want to pay attention to the way we structure our sensory data input and the way we evaluate it. Therefore, we built a model. Applying Scott E. Page's "many models thinker approach" (Scott E. Page 2018) we recognize that one model might solve many problems and one problem might be solved by many models. The model we built (a wargame) is one such model of the world.

This would be a basic sequence for answering a research question using a modelling and simulation approach (Barth et al. 2012) follows:

- Formulate the question (How do we wargame influence or how do we gamify influence for military usage?)
- Identify relevant elements of the target system (DIME/PMESII/ASCOP/IRC<sup>2</sup> planning levels (tactical, operational, strategic and time level (instant, short term, medium term, long term)
- Choose model structure (e.g. a multilevel and multidimensional wargame rigid and flexible elements)
- Implement model (e.g. version 1.0)
- Run and analyze model (we did run version 1.0 from 15.-18. October 2018)
- Communicate results (we write this article and seek validation through the review process)

At this point we run into a chicken and egg problem. What is first: the model or the question? From a philosophical point of view, the model exists a priori (Kant 2015). Dietrich Dörner contends that additional models might be helpful (Dörner 2017).<sup>1</sup> There are many models available to run simulations on a modern crisis, from the matrix wargame (Curry and Price 2017) to computational simulations of unconventional conflicts (Hartley III 2017). We are moving on a wide spectrum from educational wargaming to Courses of Action (CoA) analysis and we might retrieve educational benefits for training (e.g. how to gamify training) and organizational insights into capability building. For an excellent introduction into the wide range of possibilities to use wargaming as a method the British Ministry of Defence (MoD) published the "Wargaming Handbook" in 2017 (Development, Concepts and Doctrine Centre 2017).

## **2.3 Framing the question**

This paper will describe the "problem space" and the "solution space" following an explanation of the modelling and simulation approach, recognizing the existence of unsolved philosophical ontological and epistemological problems (Corazon 2018). They may not be the same. Einstein observed that certain complex problems on a specific level of complexity can only be solved on a higher level of complexity (Papathanassiou 2019). Computers

---

<sup>1</sup> Aristotle thought the egg and the chicken were both unchanged, always there in an infinite cosmos. That's when philosophy got it totally wrong. From a scientific point of view the egg from which the chicken hatched was laid by a proto chicken. That's called evolution (North 2018).

in the information age are a perfect example. With computers we introduced a new level of complexity and we solved a lot of problems of the industrial age (e.g. by crunching vast amounts of numbers in a very short time) but we also created a whole new set of problems specific to the information age (like transformational pain disrupting existing markets and developing new markets) – which might only be solved on a higher level of complexity (e.g. with artificial intelligence).

Developing our understanding of the problem is critical to solving it. “If I had an hour to solve a problem, I'd spend 55 minutes thinking about the problem and 5 minutes thinking about solutions.” (Garson O'Toole 2018a). In a similar paradigm, “Give me six hours to chop down a tree and I will spend the first four sharpening the axe.” (Garson O'Toole 2018b). I would like to spend a certain amount of time to sharpen our perspective on the problem. As a general frame for discussing any problem I adapted this matrix with lead questions from a course by the Teaching Company on the art of debate (Atchison 2016), which seems to be a good starting point in relation to the title.

Argumentation Matrix - Leading Questions				
Topic (generic)	Problem Status <i>Conjecturalis</i>	Definition Status <i>Definitivus</i>	Quality Status <i>Qualitativus</i>	Responsibility Status <i>Translationis</i>
<b>Problem</b>	Do we have a problem? Do we have many problems on different levels, or a meta-problem?	Is the problem sufficiently described/defined?	Is it a big problem? Do we need to act (now or later or not at all)?	Did the "right people" describe/define the problem?
<b>Cause</b>	What is the cause? Is there a single cause? Are there more than one causes?	Is the cause sufficiently described/defined?	Are the causes qualified by their strength of influence?	Did the "right people" make the cause analysis?
<b>Solution</b>	What is the solution? Is there a single solution? Are there many solutions?	Is the solution sufficiently described/defined?	Are solutions prioritized?	Did the "right people" came up with the solution?
<b>Cost</b>	How much does the solution cost?	Are the costs sufficiently described/defined?	What is the relation between cost and problem? Is the problem actually worth solving?	Did the "right people" calculate the costs?

**Figure 1:** Argumentation matrix – Adapted from (Atchison 2016)

We move through this matrix from top left to bottom right, proceeding row by row. While we may still be confused, we at least will be confused on a higher level. This matrix should be helpful answering and reflecting on any problem by framing the issues. The “Responsibility – Status Translationis” column will specifically shed light on any hidden agenda some stakeholder might have. After all, the definition of a problem and the authority to select terms will frame the problem in a certain way and might elevate a path to a certain preferred solution. For example, the inability to frame the violent acts of homegrown extreme right activist as “terrorism” may prevent the use of certain tools in the toolbox provided by the “War on Terror” (Reitman 2018).

This paper focuses on the first two cells and the respective lead questions:

- Do we have a problem?
- Do we have many problems on different levels, or do we have a meta-problem?
- Is the problem sufficiently described/defined?

On the first question (do we have a problem?) we would argue that we do indeed have a problem since man is not a purely rational being, the much-vaunted *homo economicus*. In cyber space the main problems are humans. Kevin Mitnick, one of the most infamous hackers of all time, has said, “There is no Microsoft patch for stupidity or, rather, gullibility.”<sup>2</sup> Cognitive hacking/human engineering is the main reason for 90% of all successful cyber-attacks (Kelly 2017) (Proofpoint 2018)<sup>3</sup> – tricking humans into doing something irrational, or performing an action that counters the existing security policy. Humans represent a poorly understood component in the cyber environment, and the emotional nature of human decision-making suggests accurate predictions of decisions are challenging. In spite of this, data can be shaped to provide information in a manner in which the emotional choice and the “logical choice” can appear to be the same (Cybenko et al. 2002; Sample 2017).

<sup>2</sup> Thinking about Cybersecurity: From Cyber Crime to Cyber Warfare. Paul Rosenzweig, published by THE GREAT COURSES, 2013 <https://www.thegreatcourses.co.uk/courses/thinking-about-cybersecurity-from-cyber-crime-to-cyber-warfare.html> last visited on 13. November 2018. See also: <https://www.proofpoint.com/sites/default/files/pfpt-uk-tr-the-human-factor-2018.pdf> last visited 17. March 2019

<sup>3</sup>There is a podcast completely devoted to explore the fallibility of human behavior <https://thecyberwire.com/podcasts/hacking-humans.html> last visited on 17. March 2019

### **3. Building the model**

#### **3.1 A general string theory of influence**

For an in-depth elaboration on the second (do we have many problems on different levels, or do we have a meta-problem?) and third question (is the problem sufficiently described/defined?) we will develop a multidimensional DIME/PMESII/ASCOPE/ICR<sup>2</sup> model incorporating a tactical, operational and strategic level on a short, medium and long effect time scale. The DIME model captures the main elements of national power, *Diplomacy, Information, Military, Economy*. The PMESII concept shows the interactions among the *Political, Military, Economic, Social, Informational* and *Infrastructure* domains. ASCOPE stands for *Area, Structures, Capabilities, Organizations, People and Events* (characteristics of civil considerations during a military campaign Intelligence Collection Requirements (ICR) are part of an Intelligence Collection Plan (ICP). ICR also stands for Information Capability Requirements – in this context they will be referred to as ICR<sup>2</sup>.

The advantage of the model presented within this paper is its modularity. We can focus on specific aspects without losing the holistic approach.<sup>4</sup> The modular construction leaves room to attach new aspects or cut existing ones to reduce complexity.

*Influence* is used as a cross discipline concept, in the same way that physics impacts many domains. New terms are also often created to describe new phenomenon. For example, “LikeWar” is used to describe the weaponization of social media (Singer and Brooking 2018). This is part of the “weaponization of everything” (Mousavizadeh 2015) and a theory of influence should be described in a framework which is capable of incorporating this. Broadly, influence relates to an effect a message, narrative, person or object can have on another’s behavior. In the context of this paper influence is bounded by the DIME/PMESII/ASCOPE/ICR<sup>2</sup> dimensions.

#### **3.2 Starting point**

The DIME/PMESII paradigm is my starting point (Hartley III 2017).<sup>5</sup> Although “The origins of the DIME/PMESII paradigm are unclear” (Hartley 2017) and acknowledging its limitations, the DIME/PMESII model is a useful way to start such a model, reflected across many military doctrines. “Essentially, all models are wrong, but some are useful” (Box 1976). There are alternatives to DIME and PMESII. An expansion would be MIDLIFE (*Military, Informational/Intelligence, Diplomatic, Legal (Law Enforcement), Infrastructure, Finance*) as a concept of national power. Some argue, that the linear structure of the PMESII structure only reveals the “what” and not the “why” of complex systems (Hartley III 2017).

#### **3.3 The DIME concept**

The DIME concept groups the many instruments of power a nation state can muster into four easy to remember categories. Two categories collect the soft power instruments or levers of power: Diplomacy and Information. Diplomatic power rests on negotiations and agreements. Information power lies in gaining information from others and in controlling the information desired by others. Two categories collect the hard power instruments or levers of power: Military and Economy. Military power is an obvious component. In a broader understanding this includes also the police and other instruments/actors from the executional power of a nation state. Economic power is also an obvious component. The idea of framing military and economic power as “hard power” instruments and diplomatic power and information power as “soft power”, however, narrows the view of the application of these instruments. “I suppose it is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail” (Maslow 1977). Maslow’s “law of the hammer” allows us to identify a cognitive bias, seeing military and economic levers as somehow ‘harder’ than diplomacy or information (Maslow 1977).

#### **3.4 The PMESII concept**

The five domains; political, military, economic, social, information and infrastructure; can be understood as part of an operational environment (Hartley III 2017). The following list (according to Hartley III) is not

<sup>4</sup> This was one of the goals of the influence wargaming exercise in October 2018: develop a wargaming framework specifically suited to influence

<sup>5</sup> Dean S. Hartley III also provides access to selected projects regarding tackling complex system on his website: <http://drdeanhartley.com/HartleyConsulting/hartley2.htm> last visited 17. March 2019

comprehensive. The primary components of the political domain are governance and the rule of law. The primary components of the military domain are conflict, government and security (intelligence services would be considered part of the military domain). The primary components of the economic domain are agriculture, crime, energy, finance, governmental economic actions and jobs. The primary components of the social domain are basic needs, education, health, movement and safety. The primary components of the informational domain are general information items, media, opinions and information operations. The primary components of the infrastructure domain are business infrastructure, social infrastructure, energy infrastructure, government infrastructure, transportation infrastructure and water infrastructure (Hartley III 2017).

The PMESII concept combines the domain interactions into a system and creates a framework for operational design and joint planning. The planner has a frame of reference for collaboration with inter organizational and multinational partners to determine and coordinate actions, fostering a comprehensive approach and providing a stepping stone for a Whole of Government-approach. PMESII supports the identification of Centre of Gravity (CoG) on different levels, operational CoG and strategic CoG (Hartley III 2017).

### 3.5 The ASCOPE concept

ASCOPE is used in a Counter Insurgency (COIN) environment to analyze the cultural and human environment or what is sometimes referred to as “human terrain” and encompasses areas, structures, capabilities, organization, people and events. Understanding ASCOPE is seen to identify the causes of an insurgency. It aims to provide the “who”, “what”, “when”, “where”, “why”, and “how” of the environment. By combining DIME/PMESII with ASCOPE and ICR<sup>2</sup> we hope to compensate for the limited functions of the standalone application of the PMESII concept.

PMESII/ ASCOPE	Political	Military	Economic	Social	Information	Infrastructure
<b>Areas</b>	District Boundary, Party affiliation areas	Coalition/LN bases, historic ambush/ED sites	bazaars, shops, markets	parks and Other meeting areas	Radio/TV/news papers/where people gather for word-of-mouth	Irrigation networks, water tables, medical coverage
<b>Structures</b>	town halls, government offices	police HQ Military HHQ locations	banks, markets, storage facilities	Churches, restaurants, bars, etc.	Cell/Radio/TV towers, print shops	roads, bridges, power lines, walls, dams
<b>Capabilities</b>	Dispute resolution, Insurgent capabilities	security posture, strengths and weaknesses	access to banks, ability to withstand natural disasters	Strength Of local& national ties	Literacy rate, availability of media/phone service	Ability to build/maintain roads, walls, dams)
<b>Organization</b>	Political parties and Other power brokers, UN	what units of military, police, insurgent are present	Banks, large land holders, big businesses	tribes, clans, families, youth groups, NGOs/GOS	NEWS groups, influential people who pass word	Government ministries, construction companies
<b>People</b>	Governors, councils, elders	Leaders from coalition, LN and insurgent forces	Bankers, landholders, merchants	Religious leaders, influential families	Media owners, mullahs, heads of powerful families	Builders, contractors, development councils
<b>Events</b>	elections, council meetings	lethal/nonlethal events, loss of leadership, operations, anniversaries	drought, harvest, business open/close	holidays, weddings, religious days	IO campaigns, project openings, CIVCAS events	road/bridge construction, well digging, scheduled maintenance

**Figure 2:** PMESII/ASCOPE matrix<sup>6</sup>

### 3.6 ICR as part of an ICP

The ICP is the systematic process used by most modern armed forces and intelligence services to meet intelligence requirements through the tasking of all available resources to gather and provide pertinent information within a required time limit. The creation of a collection plan is part of the intelligence cycle.<sup>7</sup>

### 3.7 ICR2 – Information Capability Requirements

ICR<sup>2</sup> can contain different elements, being open to adding new elements. In an influence activity or wargame, we can implement Physical Destruction, Presence Posture Profile, PSYOPS, Key Leader Engagement, Cyber Operations, Electronic Warfare, Public Affairs. Kinetic effects should always be part of our consideration. A bullet through the brain disrupts a human thought pattern – literally and permanently. It is like an old fashion attack

<sup>6</sup>Adopted from an open source planning template:  
<https://www.trngcmd.marines.mil/Portals/207/Docs/wtbn/MCCMOS/Planning%20Templates%20Oct%202017.pdf?ver=2017-10-19-131249-187> last visited on 17. March 2019

<sup>7</sup> See: <https://fas.org/irp/doddir/army/fm34-2/Appa.htm> last visited on 17. March 2019

on the power grid, this does not need a sophisticated hack of an Industrial Control System (ICS) - blowing up a power generator or putting an axe through a power line will achieve the same results.

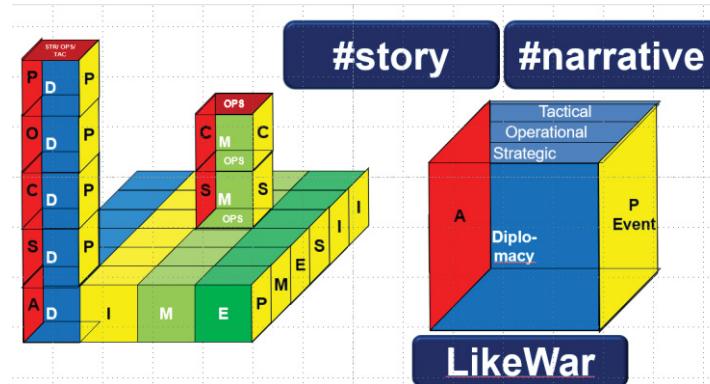
### 3.8 Perception is reality

One basic idea of the Grand Theory of Influence is “perception is reality”; this is also a leading idea of constructivism (Prawat and Floden 1994) and constructive perception (Sternberg and Mio 2007). This is the main reason why cognitive hacking works (Cybenko et al. 2002; Thompson 2004), exploiting the difference between perception and reality, or in other terms, the highlighting of similarities between perception and deeply held values and beliefs. From a classical physics point of view, an action (or event – like in ASCOPE), takes place in time and space and is the result of specific causes and leads to certain effect (in either a deterministically or probabilistically way). However, the perception from an observer is biased by her sensors (hardware and wetware (Gazzaniga et al. 2019)) and her analytical software for attributing meaning to certain events. And there is a lot that can go wrong, when a human brain tries to make sense of the world, about causes and about future effects. Usually, humans do not commit formal logical fallacies, but there are a lot of informal logical fallacies a human can commit (McCandless 2018)).

## 4. The DIME/PMESII/ASCOPE/ICR2 model

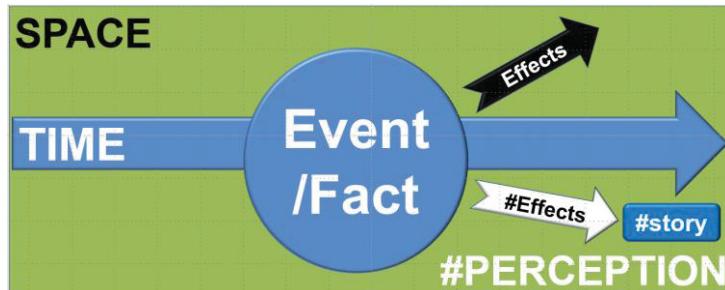
### 4.1 General description

The following model is the result of implementing an information warfare concept throughout a wargame in October 2018, featuring a whole of government approach (WoG) and reflecting on the DIME, PMESII and ASCOPE concepts. This multidimensional model can be considered as a rectangular prism with the DIME categories on the X axis, the PMESII domains on the Z and the ASCOP (**not** the E for Event) on the Y axis. This would result in a prism with  $4 \times 5 \times 5 = 100$  cubes. These cubes can have a tactical, operational or strategic “spin” (just like a quark); they are the building blocks inside a “space-time” continuum for a story or narrative.



**Figure 3:** Prism deconstruction into cubes with “spin”

An event (as a result of an action) takes place in this three-dimensional analytical framework.

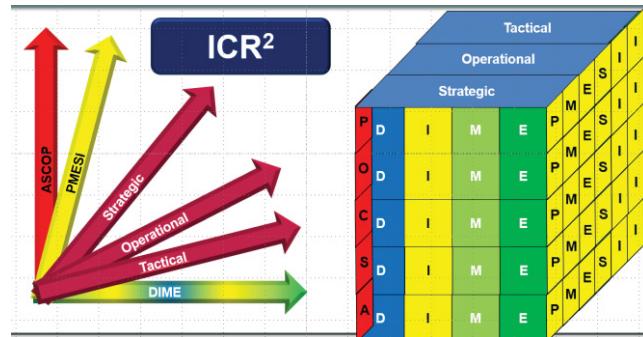


**Figure 4:** The event

This event might have a strategic, operational and/or tactical effect which can be considered instantaneous, short-term mid-term or long-term. Any event might be considered as a nonlinear vector crossing through different cubes inside this rectangular prism, creating specific ICRs.

If you map this event inside the prism any domain/categories that is relevant for the event can be imagined as a point in this cube. This point can be identified in the prism according to coordinates on the axis. There are overlaps possible, creating more than one point in different categories, on the same axis. Ultimately the mapped-out points would look like an ordered network with specific nodes.

This kind of approach is demonstrated in Eric Berlow's TED Talk on "Simplifying complexity" in which he demonstrates the transformation of the COIN Afghanistan Stability Dynamics Map<sup>8</sup> into a 3D visualization<sup>9</sup>. The 2D approach is sometimes confusing for the average person, like almost any stock-flow diagram in system dynamics. The 3D rectangular prism approach provides the space to map out all the connections in 3D while reflecting the visualization that humans experience in the physical world. It is both a framework and a model.



**Figure 5:** Complete prism dimensions

## 4.2 The perception field

Humans build their opinions based on certain perceptions. Douglas Adams and Terry Pratchett toyed with this idea of humans constructing their world view based on total reinterpretation of reality. Douglas Adams famously introduced the "SEP Field" (Somebody Else Problem Field) (Adams 1982) and Terry Pratchett introduced Death as a literally skeletal Grim Reaper who humans reified into a normal person because the human mind could not endure the truth (Pratchett 2013). We are perfectly fine with optical illusions/distortions/biases, but we have difficulties in accepting cognitive biases. The Monty Hall Paradox shows a hardwired bias concerning dependent and independent probability. Humans are bad at probability and Marilyn vos Savant started a nationwide controversy in the U.S. in 1990 trying to explains this most provocative of the counterintuitive problems in her column "Ask Marilyn" (Vos Savant 1997, 1996).

There is a lot that can go wrong in communications. Based on Schulz von Thun's communication model (four sides or channels of a message/communication) (Piel 2016), a sender cannot transmit a purely factual message without also signaling about her relationship with the receiver. She also sends information about herself and may send an implicit call for action to the receiver. This is on a very fundamental level the basic principle of deterrence (D):  $D = \text{Capability} \times \text{Commitment} \times \text{Communication}$  (another variation:  $D = \text{Capability} \times \text{Resolve} \times \text{Signaling}$ ). On an even higher level a narrative might have four sides: "Your Side", "My Side", "The Truth" (the reflection in mainstream media) and "What Actually Happened" (the actual truth) (Leitner and Rieger 2013).

Influence can be understood as a force that is transforming this perception field in a certain direction and pulling or pushing the "molecule" into another form with different properties. We might want to shield certain properties of this molecule and we might want to shape other parts. There are certain forces in the information field that have a specific attraction. Terry Pratchett speaks of "Narrativium", a specific element on Discworld on which humans run and that is an essential part of any story (Pratchett et al. 2000). Some stories are bound to happen and there are specific story lines and structures that have a strong attraction to the reader. For example, the hero's journey (Campbell 2003) or the 'from rags to riches' trope (Al-Fahim 1998). Working along these strong attractive narrative lines of a story might grant a lot of traction in any information/influence operation.

<sup>8</sup> [http://www.visualcomplexity.com/VC/project\\_details.cfm?index=788&id=788&domain=](http://www.visualcomplexity.com/VC/project_details.cfm?index=788&id=788&domain=) last visited on 17. March 2019

<sup>9</sup> [https://www.ted.com/talks/eric\\_berlow\\_how\\_complexity\\_leads\\_to\\_simplicity?language=en](https://www.ted.com/talks/eric_berlow_how_complexity_leads_to_simplicity?language=en) last visited on 17. March 2019

### **4.3 To SHAPE and/or to SHIELD**

To demonstrate the effectiveness of this model as a proof of concept we need to be able to describe an existing event or action within this cube and we need to predict certain events, actions or effects by creating new particles. For example, a **Diplomatic** action such as a G20 summit delegation making a statement at a forum criticizing a competing nation may be matched with an **Economic** action such as a sanction, will result in an **Economic** effect impacting at the **Strategic** level impacting the **Capabilities** of the affected nation (in example in reducing their ability to access the global financial system). This particle has two DIME dimensions (D and E), one PMESII effect (E), one ASCOP aspect (C) and a strategic spin and would be just two potential cubes (**DIME/PMESII/ASCOP** and **DIME/PMESII/ASCOPE** both with a strategic spin) out of many. In addition to being two cubes in the model, these are also events, with a broader narrative and additional effects that will radiate outwards, like a stone thrown into a pond. Other cubes may be affected, perceptions may be altered, narratives might be affected. Or the stone may simply sink and add to the pond floor.

## **5. Conclusion and future work**

The leading question of this paper is: "How would influence warfare ("iWar") work and how can we simulate it?". The model presented within this paper provides a method for structuring the problem of influence without seeking to reduce the complexity of the problem and diminish our capacity to solve difficult challenges. This model embraces complexity but provides it with structure.

This article is aimed at the HSCB community and one goal is to establish a community of academic, multinational and whole of government influence wargamer. Every reader is invited to establish contact and reach out to the authors. Our "whole of society" multinational wargaming community is already growing.

Future work includes the creation of an ordered influence network. Being modular, the expansion of this model will be tested for analytical value (in example from DIME to MIDLIFE). Benchmark for the actual usage of this modular approach will be the applicability for effect-based planning (e.g. in future all domain operations in urban terrain/mega cities). The "iWar" concept relies on actions according to "SHIELD" or "SHAPE" relevant influence inside the "narrative molecule". The authors intend to develop this approach further and apply it to a specific problem. This will happen in the preparation for the next version of the influence wargame. This work will also fuel the work of NATO Research Task Group 129 "Gamification of Cyber Defence/Resilience" in a coming workshop of All-Domain-Cyber-Wargaming in June 2019. The model will also be tested for educational and training value.

## **References**

- Adams, Douglas (1982): Life, the Universe, and Everything. Hitchhiker series bk. 3. New York: Random House (Hitchhiker series bk. 3).
- Al-Fahim, Mohammed (1998): From rags to riches. A story of Abu Dhabi. London: I.B. Tauris.
- Atchison, Jarrod (2016): The Art of Debate. Edited by The Great Courses. Wake Forest University. Available online at <https://www.thegreatcourses.co.uk/courses/the-art-of-debate.html>, checked on 20th March 2019.
- Barth, Rolf; Meyer, Matthias; Spitzner, Jan (2012): Typical Pitfalls of Simulation Modeling - Lessons Learned from Armed Forces and Business. JASSS. Available online at <http://jasss.soc.surrey.ac.uk/15/2/5.html>, checked on 20th March 2019.
- Bellucci, Joel (2018): The Skeptics Guide through to the Universe. <https://www.facebook.com/theskepticsguide>. Available online at <https://www.theskepticsguide.org/>, checked on 20th March 2019.
- Box, George E. P. (1976): Science and Statistics. In Journal of the American Statistical Association 71 (356), pp. 791–799. DOI: 10.1080/01621459.1976.10480949.
- Burdick, Alan (2018): Looking for Life on a Flat Earth. What a burgeoning movement says about science, solace, and how a theory becomes truth. Available online at <https://www.newyorker.com/science/elements/looking-for-life-on-a-flat-earth>, checked on 20th March 2019.
- Campbell, Joseph (2003): The hero's journey. Joseph Campbell on his life and work. First New World Library edition. Edited by Phil Cousineau. Novato, California: New World Library (The collected works of Joseph Campbell).
- Corrazen, Raul (Ed.) (2018): Ontology: Theory and History. Available online at <https://www.ontology.co/>, checked on 20th March 2019.
- Curry, John; Price, Tim (2017): Modern crises scenarios for matrix wargames. Kindle Edition: History of Wargaming Project.
- Cybenko, G.; Giani, A.; Thompson, P. (2002): Cognitive hacking: a battle for the mind. In Computer 35 (8), pp. 50–56. DOI: 10.1109/MC.2002.1023788.
- Dean S. Hartley III (2018): HARTLEY CONSULTING. Solving Complex Operational and Organizational Problems. Available online at <http://drdeanhartley.com/HartleyConsulting/hartley2.htm>, checked on 20th March 2019.

- Development, Concepts and Doctrine Centre (2017): Wargaming Handbook. Available online at [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/641040/doctrine\\_uk\\_wargaming\\_handbook.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/641040/doctrine_uk_wargaming_handbook.pdf), checked on 20th March 2019.
- Dörner, Dietrich (2017): Die Logik des Misslingens. Strategisches Denken in komplexen Situationen. 14. Auflage. Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag (Rororo, 61578 : Science).
- Fiske, Susan T.; Taylor, Shelley E. (2013): Social cognition. From brains to culture. 2nd edition. Los Angeles, London, New Delhi, Singapore, Washington, DC: Sage.
- Garson O'Toole (2018a): I Would Spend 55 Minutes Defining the Problem and then Five Minutes Solving It – Quote Investigator. Available online at <https://quoteinvestigator.com/2014/05/22/solve/>, checked on 20th March 2019.
- Garson O'Toole (2018b): To Cut Down a Tree in Five Minutes Spend Three Minutes Sharpening Your Axe – Quote Investigator. Available online at <https://quoteinvestigator.com/2014/03/29/sharp-axe/>, checked on 20th March 2019.
- Gazzaniga, Michael S.; Ivry, Richard B.; Mangun, George R. (2019): Cognitive neuroscience. The biology of the mind. Fifth edition: redesigned illustrations and anatomical figures. New York, London: W.W. Norton & Company.
- Godlee, Fiona; Smith, Jane; Marcovitch, Harvey (2017): A vaccine crisis in the era of social media. In BMJ (Clinical research ed.) 342. DOI: 10.1136/bmj.c7452.
- Hartley III, Dean S. (2017): Unconventional Conflict. A Modeling Perspective. Cham, s.l.: Springer International Publishing (Understanding Complex Systems). Available online at <http://dx.doi.org/10.1007/978-3-319-51935-7>.
- Hofstede, Geert; Hofstede, Gert Jan; Minkov, Michael (2010): Cultures and organizations. Software of the mind ; intercultural cooperation and its importance for survival. Rev. and expanded 3. ed. New York: McGraw-Hill.
- Kahneman, Daniel (2012): Thinking, fast and slow. London: Penguin Books.
- Kant, Immanuel (2015): Die drei Kritiken. Kritik der reinen Vernunft (1781/87) ; Kritik der praktischen Vernunft (1788) ; Kritik der Urteilskraft (1790). Köln: Anaconda.
- Kelly, Ross (2017): Almost 90% of Cyber Attacks are Caused by Human Error or Behavior. <https://www.facebook.com/chieffexecutivergroup>. Available online at <https://chieffexecutive.net/almost-90-cyber-attacks-caused-human-error-behavior/>, checked on 20th March 2019.
- Leitner, Felix von; Rieger, Frank (2013): Alternativlos! Folge 29. In Alternativlos Folge 29 reden wir über die Manipulation unserer Weltbilder am Beispiel der Idee, dass unsere Gesellschaft immer egoistischer werde. Available online at <https://alternativlos.org/29/>, checked on 20th March 2019.
- Maslow, Abraham H. (1977): Die Psychologie der Wissenschaft. Neue Wege d. Wahrnehmung u.d. Denkens. München: Goldmann (Goldmann-Sachbücher, 11131).
- McCandless, David (2018): Rhetorical Fallacies – A list of Logical Fallacies & Rhetorical Devices with examples — Information is Beautiful. Information is Beautiful. Available online at <https://informationisbeautiful.net/visualizations/rhetorical-fallacies/>, checked on 20th March 2019.
- Mousavizadeh, Nader (2015): The weaponization of everything: Globalization's dark side. Available online at <http://blogs.reuters.com/great-debate/2015/09/24/the-weaponization-of-everything-globalizations-dark-side/>, checked on 20th March 2019.
- Nisbett, Richard E. (2003): The geography of thought. How Asians and Westerners think differently ... and why. New York: Free Press. Available online at <http://www.loc.gov/catdir/description/simon041/2002032178.html>.
- Novella, Steven; Novella, Bob; Novella, Jay; Bernstein, Evan; Santa Maria, Cara (2018): The skeptics' guide to the universe. How to know what's really real in a world increasingly full of fake. London: Hodder & Stoughton.
- Papathanassiou, Manolis (2019): Quotes by Albert Einstein. Available online at <https://best-quotations.com/authquotes.php?auth=127>, checked on 20th March 2019.
- Piel, Lisa (2016): Vier Seiten einer Nachricht. Das Kommunikationsmodell nach Friedemann Schulz von Thun. [S.I.]: GRIN PUBLISHING.
- Pomerantsev, P. and Weiss, M. (2014): The Menace of Unreality. How the Kremlin Weaponizes Information, Culture and Money". In The Interpreter. Available online at [http://www.interpretermag.com/wp-content/uploads/2014/11/The\\_Menace\\_of\\_Unreality\\_Final.pdf](http://www.interpretermag.com/wp-content/uploads/2014/11/The_Menace_of_Unreality_Final.pdf), checked on 20th March 2019.
- Pratchett, Terry (2013): Reaper man. London: Gollancz (Discworld, the Death collection).
- Pratchett, Terry; Stewart, Ian; Cohen, Jack S. (2000): The science of Discworld. Great Britain: Ebury Press.
- Prawat, Richard S.; Floden, Robert E. (1994): Philosophical perspectives on constructivist views of learning. In Educational Psychologist 29 (1), pp. 37–48. DOI: 10.1207/s15326985ep2901\_4.
- Proofpoint (2018): Proofpoint The Human Factor 2018 2018. Available online at <https://www.pbwcz.cz/Reporty/pfpt-us-wp-human-factor-report-2018-180425.pdf>, checked on 20th March 2019.
- Reitman, Janet (2018): U.S. Law Enforcement Failed to See the Threat of White Nationalism. Now They Don't Know How to Stop It. Edited by New York Times. Available online at <https://www.nytimes.com/2018/11/03/magazine/FBI-charlottesville-white-nationalism-far-right.html>, checked on 20th March 2019.
- Sample, Char (2017): Cognitive Hacking: Recognizing and Countering 21st Century Deception. Available online at [https://www.softbox.co.uk/pub/CSP17\\_Char\\_Sample\\_Workshop.pdf](https://www.softbox.co.uk/pub/CSP17_Char_Sample_Workshop.pdf), checked on 20th March 2019.
- Sample, Char; Justice, Connie & Darraj, Emily (2018): A Model for Evaluating Fake News. Available online at <https://www.hSDL.org/?view&did=818918>, checked on 20th March 2019.
- Scott E. Page (2018): Model Thinking | Coursera. Edited by Coursera. University of Michigan. Available online at <https://www.coursera.org/learn/model-thinking>, checked on 20th March 2019.

***Thorsten Kodalle, Char Sample, David Ormrod and Keith Scott***

- Shermer, Michael (2002): Why people believe weird things. Pseudoscience, superstition, and other confusions of our time. First Holt Paperbacks edition. New York: St. Martin's Griffin.
- Singer, P. W.; Brooking, Emerson T. (2018): Likewar. The weaponization of social media. Boston, New York: Houghton Mifflin Harcourt.
- Sternberg, Robert J.; Mio, Jeff (2007): Cognitive psychology. International student ed., 4. ed., [Nachdr.]. Belmont, Calif.: Thomson Wadsworth.
- The Economist (2018): Anti-vax fears drive a measles outbreak in Europe. <https://www.facebook.com/theeconomist>. Available online at <https://www.economist.com/europe/2018/08/25/anti-vax-fears-drive-a-measles-outbreak-in-europe>, checked on 20th March 2019.
- Thompson, Paul (2004): Cognitive hacking and intelligence and security informatics. In Dawn A. Trevisani, Alex F. Sisti (Eds.): Enabling Technologies for Simulation Science VIII. Defense and Security. Orlando, FL, Monday 12 April 2004: SPIE (SPIE Proceedings), pp. 142–151.
- Vos Savant, Marilyn Mach (1997, 1996): The power of logical thinking. Easy lessons in the art of reasoning-- and hard facts about its absence in our lives. 1st St. Martin's Griffin ed. New York: St. Martin's Griffin.

# Success Factors and Pitfalls in Security Certifications

**Helvi Salminen**

**Gemalto Oy, Finland**

[helvi.salminen@gemalto.com](mailto:helvi.salminen@gemalto.com)

**Abstract:** The modern society is increasingly dependent on information systems and would turn into chaos if the availability, integrity and confidentiality of systems were lost. We have learnt to trust these systems. Various security certification schemes have been introduced in order to establish the required trust. Certifications are used as “quality labels” – a valid and reliable assurance of security. Certifications are based on a predefined set of criteria, and can be granted e.g. to products and organizations. But is a certification - compliance with a set of security requirements – assurance of a good level of security? The validity and reliability of certifications as security assurance method can be challenged in many ways. Has the development process of the security criteria been biased? Is the set of criteria based on risks relevant to the certified target? Is the auditor accreditation process well managed? Is the security auditor's competence sufficient? Can a regulatory audit be passed with a security theatre? The main problem in the certifications is often the distorted balance of the interests of different stakeholders. There is also a common misunderstanding that the tightest criteria or tightest interpretation always results in best security – in reality it isn't often so. The above stated questions present either success factors or pitfalls – depending on the answers in each specific case. Auditing criteria, auditor's competence and organizational culture are important contributors to the certification's validity and reliability as security assurance. The author of this paper has more than 20 years of experience in security certifications and audits - mainly as auditee but also as auditor, and has experience in applying several security standards and frameworks. This experience enriched by discussions with other security professionals and auditors forms the basis for analyzing the validity and reliability of audits as security assurance method. Several examples are included in the paper. The conclusion of the analysis is that in spite of their limitations security certifications are a useful method for security assurance. However, it is important to understand their weaknesses and take a closer look at the certified entity when dealing with matters of vital importance.

**Keywords:** certification, assurance, security, audit, reliability

---

## 1. Why security certifications?

Information and communication systems have become an integral part of our technical environment from personal use cases to critical infrastructure. It is difficult to find an area of human activity where information systems don't have a role. Use of cash has largely been replaced with payment cards and other electronic payment methods. Modern business models with complex interdependencies would not exist without efficient real-time information processing capacity. Large automated factories and power plants rely on complex process control systems ... The list of activities which would be in serious difficulties or stop if the information system does not offer its services is getting longer and longer.

These systems are also trusted the task to handle and control valuable assets on personal level, in companies and public administration. When we trust our assets to an entity we naturally want to be sure that they are handled properly. It is not possible for us as individuals, often also representatives of a company or public administration, go and see ourselves to ensure that our assets are safe. How can a person that his money does not disappear as result of a bank's system malfunction, internal fraud or bank robbery utilizing system weaknesses? What about sensitive healthcare system weaknesses? How can we be sure that national secrets are not leaking to hands of hostile forces?

Security certifications have been created to establish trust in the increasingly important information systems and the activities which depend on these systems. Certification by an independent third party is meant to be assurance that the integrity, availability and confidentiality of the assets trusted to the certified entity are maintained.

## 2. Accreditation and certification as an infrastructure of trust

In order to provide reliable certifications the certification process itself must be well designed and implemented. The organizational structure of certification can be seen as an infrastructure of trust. The two main processes which establish the trust are *accreditation* and *certification*.

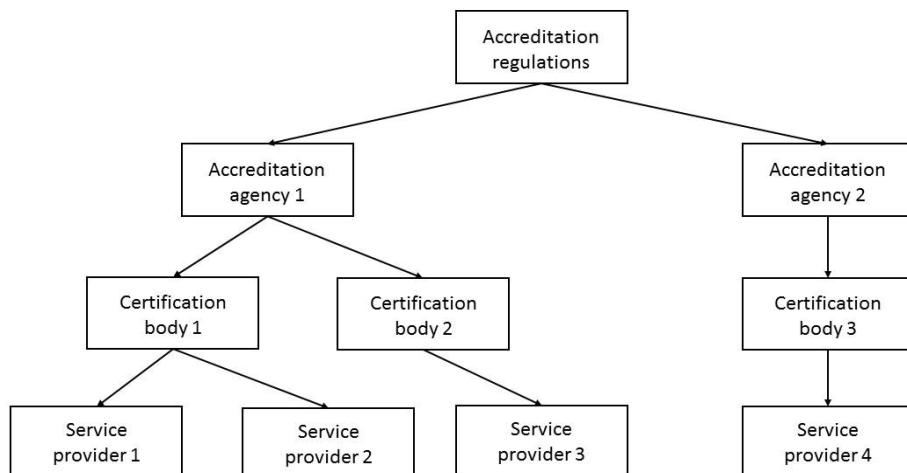
Oxford (2019) defines the term *accreditation* as “The action or process of officially recognizing someone as having a particular status or being qualified to perform a particular activity.”

EU (2008) states that in European Union member states there can be only one organisation providing accreditation services. If there is no national accreditation in a member state, a conformity assessment body can request accreditation from another member state. The regulation also defines strict criteria to ensure that the accreditation bodies operate in a rigorous and transparent manner.

According to EU (2008) "The particular value of accreditation lies in the fact that it provides an authoritative statement of the technical competence of bodies whose task is to ensure conformity with the applicable requirements." To ensure the impartiality of the accreditation bodies they must operate independently of commercial conformity assessment bodies.

There is sometimes even among certified organisations misunderstanding of the roles of accreditation and certification. Finas (2016) article clarifies the objectives of these two processes and also their main differences. *Accreditation* is a process which aims at demonstrating the reliability of the results and/or the credibility of certificates issued by a certification body. Accreditation is granted for the scope in which the certification body is able to demonstrate the competence. *Certification* is assessment of conformity with requirements which most cases are stated in standards.

The organizational structure of the accreditation-certification can be described as follows:



**Figure 1:** Accreditation and certification as infrastructure of trust

*Accreditation regulations* are on top of the trust infrastructure.

*Accreditation agency* evaluates the competence of the certification of the certification body.

*Certification body* assessed the service provider's conformity with the requirements stated in a given standard and issues the certificates. The assessment can also be done by a separate accredited inspection body.

*Service provider* implements the requirements.

Service providers 1, 2, 3 and 4 are certified based on the same security standard by an accredited certification body. The accreditation-certification process is designed to ensure that the requirements defined in the security standard are met in a comparable manner by all four certified service providers. So a service user should feel confident to select any of the four service providers – even if there are three different certification bodies and two accreditation agencies involved.

### **3. Types of security certifications**

Finas (2016) explains that there are three main targets of certifications: systems, products and personnel. There are several certification schemes, and also many different types of targets can be included in each of the three main categories.

As this paper is concentrating on information security related topics, the following two types of certification schemes are analysed in this paper: security evaluation of products and certification of security management systems. These two categories form a significant part of information security related certification activities.

### **3.1 Product evaluation schemes**

The purpose of product evaluation is to ensure that the product meets the security requirements established in a security standard. Examples of security evaluation schemes are Common Criteria and FIPS 140-2.

#### **Common Criteria**

*Common Criteria* (CC) is an arrangement established to ensure that evaluations of information technology products and protection profiles are performed consistently. Once performed according to a well defined procedure by competent assessor, the evaluation can be trusted and thus eliminates the need to duplicate the evaluation.

In the Common Criteria process the target of evaluation is the product or system being assessed. Production profile (PP) specifies the security requirements for a category of products. Security target (ST) specifies the security properties of the TOE, and Security Functional Requirements (SFR) specify individual security functions which may be provided by a product. Common Criteria presents a standard catalogue of such functions. Security Assurance Requirements (SAR) are definitions of what security measures are in place during development and evaluation of a product.

An important concept is Evaluation Assurance Level (EAL) which describes the depth and rigor of evaluation. The higher the EAL, the more assurance requirements are applicable during the development of a product. Common Criteria certification can ensure that claims about the security attributes of the evaluated product were independently verified. In other words, products evaluated against a Common Criteria standard exhibit a clear chain of evidence that the process of specification, implementation, and evaluation has been conducted in a rigorous and standard manner.

Common Criteria MSSR (2017) expands the product evaluation to the environment in which the product is developed. MSSR is applied also to the environment in which the evaluated product is produced. The purpose of production audit is to ensure that the product is correctly handled in the production phase and is the one which has the evaluation label.

#### **FIPS 140-2**

Whereas CC is a generic evaluation scheme applicable to a wide range of products, *FIPS 140-2* (2002) defines security requirements for cryptographic modules. The standard increasing, qualitative levels for security: Level 1 being the first level, level 4 being the highest.

The standard's requirements cover in detail interfaces, operational aspects, physical security and cryptographic functions, including random number generation. Design assurance is also included in the standard.

Many other standards, including PCI, require FIPS 140-2 evaluated cryptographic equipment to be used.

### **3.2 Security management system certification**

Security management system certification is done against a management system standard. There are several standards for this purpose. This chapter includes a brief presentation of a few standards.

ISO27001 (2013) is a management system standard which claims to be applicable to all organizations regardless of type, size or nature. The management system and the Annex A normative controls are defined in a generic manner – *what* instead of *how* – which ensures wide applicability.

ISO14298 (2013) is a security management system standard for security printing which a specific business process. The management system is quite similar to ISO27001. The Annex A normative controls, though specific

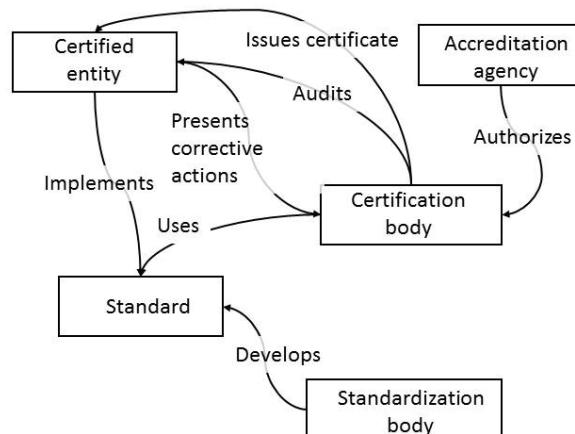
to the security printing business, are defined in a rather generic manner as the ISO27001 controls. The number of controls is limited. In security printing business most organizations are required to apply additional controls defined by Intergraf (2018).

PCICP-P (2017) and PCICP-L (2017) are security requirements standards for payment card production. These standards do not include a clear definition of management process. However, they include e.g. requirements for security role allocation, security reporting and evidences of application of required controls. The standards also define several controls in a very detailed manner, explaining not only *what* to do but also *how* to do it. However, as they are not product specific, they can be considered to be in the management system standard category.

Katakri (2015) is Finnish national security auditing criteria. It has been created to be used as an auditing tool for authorities. Facility Security Clearance in Finland is granted based on an audit done using the Katakri criteria. Katakri includes criteria for physical, administrative and information security. Katakri is not a requirements definition standard, but contains questions and implementation examples and guidelines for three different security levels. In addition Katakri refers to other documents which can state detailed security requirements e.g. for handling EU classified information.

### **3.3 Certification process**

The following diagram presents the actors involved in a certification process. It is slightly simplified, but gives an idea of the complexity certifications. In the picture the certification body conducts the audit. However, the audit can also be performed by a separate accredited inspection body.



**Figure 2:** Actors in security certifications

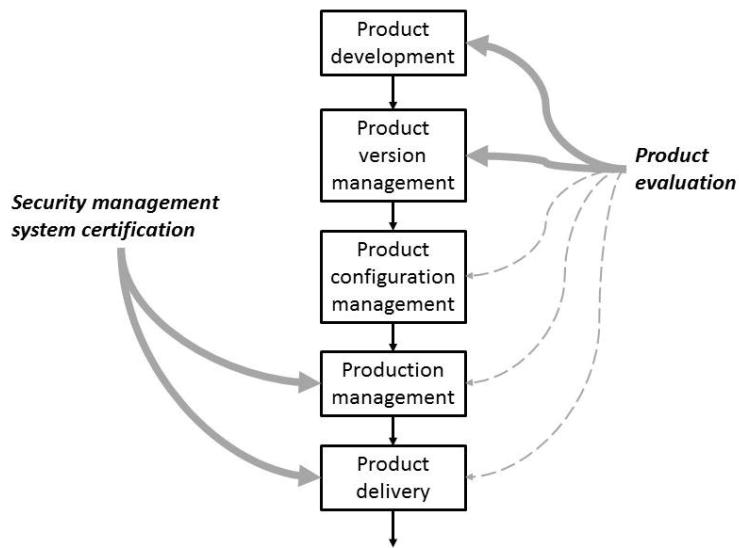
## **4. Pitfalls in the security evaluation and certification processes**

### **4.1 Possible gaps in the product evaluation – production certification coverage**

Product evaluation and security management system certification often form a chain of trust. Product evaluation is designed to ensure that a product meets a defined security criteria on a required level of trust. Product evaluation process can also be extended to cover the production process.

The scope of security management system certification in a production environment typically covers production and product delivery.

The following picture shows the coverage of the product evaluation and security management system certification processes in a typical case. These assurance processes have a strong involvement in most parts of the entire chain of operations. However, the product configuration management – which is the configuration of the product in the production phase, where the evaluated product can be a part of a more complex product – may not be fully covered by either of the assurance processes. So there is a potential weak point and the assurance of integrity of the product throughout the entire life cycle may not be sufficient.



**Figure 3:** Combined evaluation - certification process

#### **4.2 Possible weaknesses in some standards used for security management certification**

Security management system standards are designed to ensure that there are sufficient security controls in place. The standards or their application in practice, in spite of their usefulness, have some weaknesses. The following table summarizes the strengths and weaknesses of some standards. Similar problems are encountered also in standards which are not listed in the table.

**Table 1:** Strengths and weaknesses of security management certification standards

Standard	Strengths	Weaknesses
ISO27001	Wide applicability. Well defined security management system process with management involvement. Control objectives defined for mandatory controls. Strong link to risk management in control selection and implementation.	Lacks concreteness – certification can be obtained having different security levels. Scoping – restricted scope can give false impression of the security in the organization. Comparability of two certified organizations
ISO14298	Well defined security management system process with management involvement. Control objectives defined for mandatory controls. Standard applicable to a specific activity – suitability for the purpose.	Set of mandatory controls is limited and lacks many controls important in the security printing process. Scoping – restricted scope can give false impression of the security in the organization. Comparability of two certified organizations.
ISO14298 with Intergraf ICR	<i>Adds to the ISO14298 basic version</i> Strong link to risk management in control selection and implementation Well defined set of controls specific to the specific activity. Concreteness of the control requirements	The Intergraf ICR is not publicly available Some controls have a detail level which is not clearly justified based on risks. Limited auditor base – only two accredited auditing bodies.
PCICP Physical and logical security	Comprehensive set of controls defined for a specific activity (payment card production and personalization) Concreteness of the controls	Some controls have a detail level which is not clearly justified based on risks – link to risk management is not clearly defined. No classification of nonconformities. Possibility to use compensating controls is not defined in the standards. Link to management processes and management involvement is very thin.

Standard	Strengths	Weaknesses
Katakri	Wide applicability Definition of security levels for information having different protection needs	Lack of clarity of requirements. Scoping – restricted scope can give false impression of the security in the organization. Comparability of two certified organizations. External references to standards done in a way which does not clearly state if the controls are mandatory or optional.
All above	Provide a systematic approach to security and include useful control definitions. Obligation to have external audits ensures that there is a review of security by an independent party.	Restricting the scope of the certification can give false impression of the security or the organisation. Complexity of the standards make it difficult to find auditors whose competence covers all areas of the standards. Number of controls makes it difficult to verify compliance in the limited time reserved for audits. Implementation and audit costs.

#### **4.3 Whose interests are best served**

There are several interested parties in the accreditation, evaluation and certification processes: accrediting agencies, certification bodies, auditing bodies, organizations being audited, companies doing business with certified organizations, individuals using the certified services or products etc.

Ideally, it is the interest of all parties involved that the security measures are defined and implemented in a way which reduces everyone's risk. Risk awareness can increase as result of the certification process, and systematic approach with management support improves significantly the security culture and solutions. However, in practical situations many other types of factors may influence the certification process and its results.

Many of the potential problems have their roots in economy. An organisation being audited can value the low cost of the certification more than risk management. The knowledge of the different auditors' competencies and audit approach can then lead to select the auditor who causes "less trouble" – low cost of the audit and less reported nonconformities. This can, however, be short-sighted and in the long run significantly increase risk of the organization's and other interested parties.

Certification bodies and auditors are in competition with each other. Sometimes the competition is done with lower prices which can mean less time spent at the target organization and as consequence significant weaknesses may not be noticed and corrected.

In some cases auditors are instructed to make extremely strict interpretations which can lead to major nonconformities without a significant risk and very high cost of audits.

#### **4.4 More can be less**

Often people think that the stricter the control requirements and their interpretation, the better the resulting security. However, sometimes strict approach and extreme detail level can result in less security. This can happen at least in two types of situations.

When the control requirements and their interpretation are very strict, achieving compliance requires significant resources which adds the costs. When the cost of control implementation and the risk being managed with the controls are not in balance, there can be several types of negative responses. Maybe the most significant one is that the scope of the certification is reduced to minimum permitted by the certification scheme, leaving out activities which would benefit from the certification. Strictness can also lead to lack of controls elsewhere where there is higher risk, or implementation of controls which may seem sufficient but do not really reduce risk.

Detailed control requirements without the possibility of compensating controls can result in inconsistent security. The requirements standards do not necessarily include the control definitions which are most suitable for the target. Compliance for the sake of compliance and not for reducing risk can gradually undermine the organization's risk management culture.

## **5. Credibility of certifications**

It is in the interest of all parties involved to maintain the credibility of certifications. Trusted certifications eliminate the need to do additional comprehensive audits of a certified product or organisation. At its best the accreditation-certification process is a self-correcting mechanism which results in better security culture and risk management, and emerging problems are solved more efficiently.

Accreditation process is important in establishing and maintaining the credibility of certifications. It is important that accreditation of certification and audit bodies and individual auditors is done properly. And it is equally important to monitor the performance regularly. Feedback of the auditees should be an important part of the monitoring process. The process should as well as possible ensure that the results of two audits using the same requirements standards in the same target don't significantly differ from each other.

In some certification schemes (e.g. FIPS 140-2 and PCI Card Production Standards) the scope is well defined by default. Many other standards, however, enable restriction of the scope which leads to the situation in which the same certification can be obtained with significantly different scopes even in the same type of organisations. In these cases it is important that all relevant important parties are aware of the scope.

Generic standards which are applicable to different types of organisations and processes have the advantage of flexibility for control selection. However, this makes it difficult to compare two certified organisations even in the case when the scope of the certification is similar. This type of setup requires the auditors to have more competence than in case of standards with detailed control statements. And the situation is even more challenging for an interested party who wants to compare the two organisations – which often results in the need to perform additional audits to verify that the specific security requirements of the interested party are met.

As the technology is changing rapidly, it is very difficult to create good standards with detailed control definitions except for some specific areas. Some of the required controls may be outdated soon after the publication of the standard. If the possibility of compensating controls is not included in the standard, which often is the case, compliance may be in conflict with security.

The standardisation process itself also has an impact in the credibility. International standards are created using a procedure in which a lot of parties are involved. The standardisation process reduces significantly the possibility that a single interested party can include its own requirements in the standard. The creation and update of standards typically takes a long time, even years, which makes it difficult to react quickly to changes in technology.

Security culture of a certified organisation is a key factor in the credibility of the certifications. In the best case the certification process increases top management involvement in security management and improves risk awareness and risk management culture. In this case there is a good base for trust. "Security label with less cost" approach works in the opposite direction.

## **6. Conclusions**

Despite their limitations security certifications have many significant benefits. They offer a level of assurance that the security needs of different interested parties are fulfilled. The key factors are common standards, audits performed by accredited auditors, and naturally the impact of the certification in the auditee's security culture. However, there are some weaknesses which could be addressed to increase the positive impacts of the certification.

In the certification standards better definition of control objectives, inclusion of the possibility to use compensating controls and well defined and documented link to risk management would be important. It would also be important to have well documented auditing guidelines (or instructions) which are also available to the auditee.

In a complex environment where several security frameworks are applied and several regulatory audits are performed, it would be important to ensure that no gaps remain in the chain of trust. This would require new type of thinking and cooperation of different certification schemes.

**Disclaimer:** This paper is based on the author's experience and does not represent opinion of any institution.

## References

- CC, <https://www.commoncriteriaportal.org/>
- CC JIL (2017). *Minimum Site Security Requirements*. Joint Interpretation Library, December 2017.
- EU (2008). *REGULATION (EC) No 765/2008 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL, setting out the requirements for accreditation and market surveillance relating to the marketing of products and repealing Regulation (EEC) No 339/93*
- Finas (2016). *Accreditation and certification; objectives and the main differences*.  
<https://www.finias.fi/sites/en/topical/articles/Pages/Accreditation-and-certification;-objectives-and-the-main-differences.aspx>
- Finas (2018). *Accreditation is a global issue*.  
[https://www.finias.fi/sites/en/topical/articles/Pages/blog\\_Accreditation\\_is\\_a\\_global\\_issue.aspx](https://www.finias.fi/sites/en/topical/articles/Pages/blog_Accreditation_is_a_global_issue.aspx)
- FIPS 140-2 (2002). *Security Requirements for Cryptographic Modules*. Information Technology Laboratory, National Institute of Standards and Technology. March 2002.
- Intergraf (2018), *Certification Requirements for ISO 14298 Management for Security Printing Processes*. Intergraf 2018
- ISO (2012), *Conformity assessment -- Requirements for the operation of various types of bodies performing inspection*, International Organization for Standardization
- ISO 27001 (2013), *ISO/IEC 27001:2013, Information technology. Security techniques. Information security management systems. Requirements*. International Organization for Standardization
- ISO 14298 (2013), *ISO 14298:2013(E), Graphic technology — Management of security printing processes*, International Organization for Standardization
- NSA-FI (2015). *Katakri 2015 – Information security audit tool for authorities*. National Security Authority of Finland
- Oxford (2019) Definition of accreditation in Oxford online dictionary.  
<https://en.oxforddictionaries.com/definition/accreditation>
- PCICP-P (2017), *Payment Card Industry (PCI) Card Production and Provisioning - Physical Security Requirements Version 2.0*, PCI Security Standards Council
- PCICP-L (2017), *Payment Card Industry (PCI) Card Production and Provisioning - Logical Security Requirements Version 2.0*. PCI Security Standards Council

# Instilling Digital Citizenship Skills Through Education: A Malaysian Perspective

**Ramona Susanty, Zahri Yunos, Mustaffa Ahmad and Norriza Razali**

**CyberSecurity Malaysia, Selangor, Malaysia**

[ramona@cybersecurity.my](mailto:ramona@cybersecurity.my)

[zahri@cybersecurity.my](mailto:zahri@cybersecurity.my)

[mus@cybersecurity.my](mailto:mus@cybersecurity.my)

[norrizan@tech-capacity.com](mailto:norrizan@tech-capacity.com)

**Abstract:** The increasing presence of technology in the learning environment poses a new challenge to keeping children safe in the digital world. In line with CyberSecurity Malaysia's overarching mission of creating and sustaining a safer cyberspace to promote national sustainability, social well-being and wealth creation, the CyberSAFE™ program has been implemented focusing on outreach and capacity building. However, the unprecedented development of social media and the increasing access to smart phones present a new challenge to parents and educators alike. While technology serves as learning stimulants, the potential threats caused by excessive exposure to technology and its inappropriate content, can have lasting harmful effects on young learners. The potential threats of technology have created a sense of urgency amongst all stakeholders responsible for delivering digital citizenship education especially for K 12 students, as a way to prepare students as technology users with the appropriate norms and responsible users of technology. This presentation shares CyberSecurity Malaysia's experience as the impetus in establishing the digital citizenship initiative with key stakeholders that deliver various education programs. A highlight will be a discussion on the design thinking approach adopted throughout various engagement sessions with key stakeholders. Towards the objective of instilling digital citizenship amongst K 12 students through a sustainable and scalable model, the presentation will cover four key areas which are empathizing teachers and parents' challenges with instilling digital citizenship, identifying the solution, developing the resources and teacher and parent development programs, establishing the implementation strategies and the impact assessment.

**Keywords:** CyberSAFE™, digital citizenship, education, teachers, parents, students

---

## 1. Background

The Internet Users Survey by MCMC (2017) states that there were 24.5 million internet users in Malaysia with 96% active in texting and 89% regular social network sites visitors. Global statistics 2021 forecast for social media users is 3.2billion or 43% of the world's population.

As more and more individuals interact digitally—with content, one another, and various communities, the concept of digital citizenship becomes increasingly important. Digital citizenship can be described as the self-monitored habits that sustain and improve the digital communities you enjoy or depend on (Preston et al. 2018).

Digital citizenship is a concept which helps teachers, technology leaders and parents to understand what students/children/technology users should know to use technology appropriately. It can be incorporated in teaching and learning as a teaching tool to prepare students or technology users for a society full of technology. Digital citizenship is the guideline on the norms of appropriate, responsible use of technology.

## 2. Literature review

There is a bulk of studies and articles discussing the risks, issues and challenges of excessive and unmanaged exposure to technology. This section of the paper highlights the risks, issues and challenges globally, best practices and the development in this area within Malaysia.

### 2.1 Risks – harmful effects of unregulated and unprotected online interaction

There is a growing concern when more and more young children are exposed and interacting on digital platforms. Social psychologists have made several observations on the pitfalls of unregulated and unprotected interactions with digital platforms by young users. Aleksandra (2018) for instance, highlights a number of social psychological effects around emotions and social intelligence, attention, addiction and brain chemistry, texting and multitasking, and *sharenting* (where parents' internet behaviour such as posting their children's photos are mirrored by their children).

## **2.2 Issues – young and adult users lack awareness and low concerns**

A European Union study (Smahel D et al. 2015) across seven countries present some key findings on the pioneering effort in Europe on the exploration of young children (0–8) and their families' experiences with new technologies. Below are highlights of the key findings that are potential risks:

### *2.2.1 Children's perspectives*

- Children have little awareness of what internet is, what 'online' means, what risks they can encounter or the benefits they can gain.
- Children learn from observing their parents who, in most cases, are unaware of their children's mirroring habit.
- Children use digital technology individually rather than socially.
- Children use their parents' device which increases risks of problematic experiences with pop ups and in-app purchases.

### *2.2.2 Parents' perspectives*

- Parents see digital technologies as positive but challenging at the same time in their control and regulation.
- Parents can see risks of economic consequences, incidental inappropriate content, health or social impacts. Encountering violence and strong language seems of greater concern to parents than sexual content or unwanted contact.
- Parents are clearer about the risks than benefits in fostering creativity, imagination, social skills, knowledge acquisition, hand-eye coordination and educational provision for future.
- Parents are concerned about the exposure to violence and inappropriate content and in app purchase risks but seem to underestimate the wider risks of the use of technologies by their children.
- Parents use restrictive strategies. They set rules to limit children's access to digital technology mainly through time limits and restrictive condition of use (a short selection of games or videos, strictly off-line, passwords).

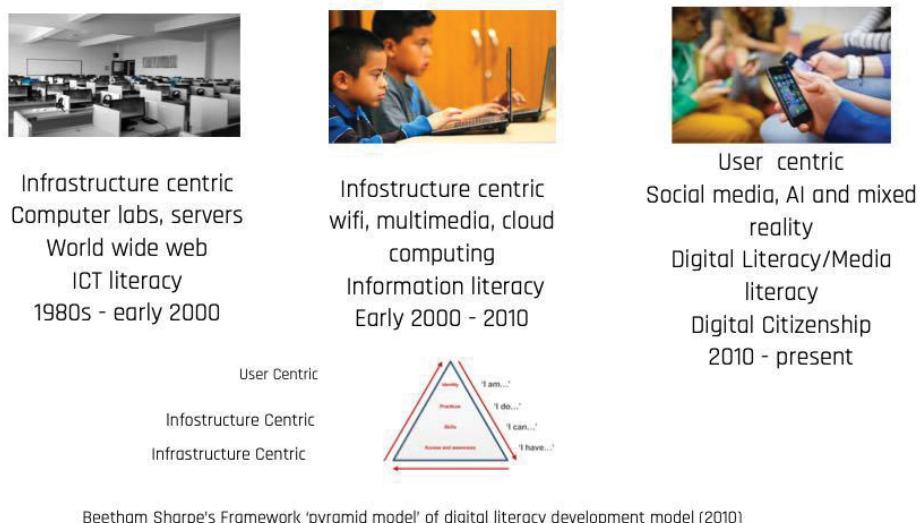
The study (Smahel D et al. 2015) shows that children remain naïve about the appropriate conduct online and tend to model after adults' behaviour with devices. The study also concludes that there is evidence of gaps in parental knowledge relating to online risks. The recommendations generally point to strategies to help parents, carers and educators on how to support, supervise, mediate, and capitalise benefits of digital platforms for children in learning and acquiring digital and critical thinking skills in collaboration with schools. The education sector is identified as having a critical role in instilling the self-regulatory behaviour amongst children through learning activities.

The national survey on cybersafety (Yunos et al. 2017) among school children in Malaysia reveals that 40% of school children do not know how to protect themselves online, >40% said online safety is important yet take low levels of protection, 70% are not concerned with the invasion of privacy or anonymity of the person they interact with, and more than 70% identified themselves with various forms of online harassment, namely calling others mean names, posting improper messages and inappropriate photos.

It can be concluded from these studies that young users as well as parents or carers are ignorant or are not concerned about the harmful effects of low levels of online protection and irresponsible conduct when interacting online.

## 2.3 Challenges – slow rate of digital citizenship adoption

### 2.3.1 Beetham Sharpe's framework and technology phases



The Beetham Sharpe's framework 'pyramid model' of digital literacy development model suggests that with the prevalence of social media, the users are at the centre as they are uploading, sharing, collaborating and expressing themselves (Sharpe & Beetham 2010).

This is a departure from the infrastructure and infostructure phases. The infrastructure phases up until the early 2000 saw organizations initiating scaled acquisition and deployment of hardware and the initial development of the world wide web. ICT literacy was the focus. Cybersecurity concerns were centred on securing the hardware from exposure to malware. The infostructure phase shifted from hardware to Wi-Fi, multimedia, cloud computing and software. Information literacy began to gain traction as the internet and multimedia convergence enabled users access to information from devices, anytime anywhere. Cybersecurity concerns at this stage was still centred on the safety of the hardware and software. Users' protection began to gain increased focus but at slower rate especially in the case of developing countries, like Malaysia.

With the current phase of social media and other emerging technology trends such as augmented reality, virtual reality, mixed reality, chatbots, blockchain, and wearable tech, the safety of users is the main concern. However, the safety of users is no longer confined to online security. Issues which are social in nature began to be recognized as threatening the positive digital environment for users.

Promoting positive conduct that builds upon users' protection, respect and learning is becoming a movement. Users' own inappropriate conduct, exposure to other users' misconduct and to ill intended content such as fake news, cyberbullying, child grooming, clickbait, carry potential social harms at both the societal and individual levels. Yet, this movement has not really gained wide recognition amongst authorities, educators and parents in Malaysia. Digital literacy and media literacy are interchangeably used as digital citizenship. Yet, it is vital for users to analyse, verify and identify types of content online and independently protect themselves and self-regulate their behaviour when online.

## 2.4 Global best practices

The US has made significant progress in the development and implementation of digital citizenship in education. Various organizations with core interest in educational technology such as Digital Citizenship Institute and International Society for Technology in Education (ISTE) to name a few, have led the way with digital citizenship resources. Schools and school districts choose and design their own digital citizenship programs often using

available resources for lessons and teacher professional development from these organizations. Digital citizenship is often taught as a subject in schools that have decided to implement it.

The Digital Citizenship Institute (2018) identifies nine digital citizenship themes which are digital access, digital commerce, digital communication, digital literacy, digital etiquette, digital law, digital rights and responsibility, digital health and wellness, and digital security. These themes are also clustered along three principles, respect, educate and protect (REP).



The Australian New South Wales (2018a), Department of Education makes digital citizenship part of the curriculum for all Australian New South Wales government schools. Students are taught on how to be safe and responsible digital citizens, appropriate technology use, and how to identify and avoid online dangers such as cyberbullying, inappropriate websites, grooming, losing control of personal pictures, viruses and other security issues. The Australian New South Wales Education Department's Digital Citizenship website and the Office of the eSafety Commissioner provide digital citizenship resources for teacher professional learning and lessons.

The key competencies and values of the New Zealand Curriculum have encouraged the teaching of cybersafety and digital citizenship (2018c). The approach has been student centred and supported by professional development and educational resources led by New Zealand's leading internet safety body Netsafe which aligns with the international shift in policy and practice away from protective safety-based approaches towards more holistic and strengths-based solutions, adopting effective online safety approaches that balance protective and promotional activity.

The UK Department for Education (2016) published 'Keeping Children Safe in Education', an updated statutory guide for schools and colleges in May 2016, for adoption from 5th September 2016. Supporting schools with these legal requirements is Childnet and the Safer Internet Centre made up of three UK charities bringing together young people and professionals working with young people, including the police, schools, youth workers and groups. These charities conduct programs and provide resources addressing digital citizenship for young people, teachers and professionals, and parents and carers.

Singapore launched the Digital Readiness Blueprint (2017) that signifies the commitment of the Singapore Government and along with their partners in the private and people sectors to ensure that people are at the centre of Singapore's Smart Nation efforts. Aligned with the Smart Nation efforts, Singapore MOE has also implemented a Cyber Wellness framework to instil the principles of "Respect for Self and Others", "Safe and Responsible Use", and "Positive Peer Influence" to guide students in making thoughtful decisions in their online activities.

The DQ Institute in Singapore (2018d), a public-private-civic-academic coalition in association with the World Economic Forum aims to bring quality digital intelligence education to every child. The Digital Intelligence Quotient, is a set of critical life skills in this digital age. DQ is the sum of technical, mental and social skills empowering individuals to deal with these challenges of living in a digital world and thrive. The DQ Institute provide resources covering digital citizenship for educators and parents.

In summary, the digital citizenship programs in many developed countries are implemented mainly through education in partnership with private organizations and other government agencies responsible for internet safety. In the US, Australia and New Zealand, digital citizenship is largely taught as a subject of its own or integrated within the IT subject. Resources and professional learning for educators are made available for teachers, leaders and parents. NSW, Australia is an example of a state-wide implementation of digital citizenship to state-run schools.

## **2.5 Local digital citizenship initiatives**

Apart from CyberSecurity Malaysia's CyberSAFE™ program, digital citizenship related programs in the country are sporadic. MCMC is initiating a Public Awareness on Internet Safety Campaign with 'Klik dengan Bijak' ("Click Wisely") as the main theme which is aiming for safety, security and responsibility. children, youth and parents and guardians. MOE's programs are in close collaboration with CyberSAFE™. Digi is the only Telco that is active in programs that carry digital citizenship messages.

CyberSecurity Malaysia through the CyberSAFE™ Program is already a piece of evidence that Malaysia too is in the same direction with the rising global digital citizenship movement (Ab Hamid et al. 2018). Through the CyberSAFE™ program, CyberSecurity Malaysia has an array of programs that are considered pioneering efforts in the region for initiating digital citizenship especially on internet safety and security.

The targeted groups, kids, youth, parents, organizations, and community are already identified as recipients of various programs notably the CyberSAFE™ in schools and the National ICT Security Discourse (NICTSeD). The CyberSAFE™ program has also developed various materials such as cyber tips, posters, cyber tools, games and quizzes, newsletter, and video.

With the imminent issues associated with the exponential rise of social media users in the country, it is imperative for the digital citizenship component within the CyberSecurity Malaysia's CyberSAFE™ program to extend and widen the emphasis of safety and security on social media while strengthening and elevating the internet safety and security focus.

The National Digital Citizenship initiative suggests the strengthening and scaling of the existing CyberSAFE™ program such that it addresses the threats of not only internet but the widening social media focus. In this way, the CyberSAFE™ program not only becomes more comprehensive but is also progressing and aligning with the current and future development of technology.

As widely practised in developed countries, education is the key vehicle to widely and efficiently impart the digital citizenship skills to the citizens, especially to the younger generation.

## **3. Methodology**

The methodology employed to develop the National Digital Citizenship Initiative (NDCI) relies on an engagement model approach that is combined with an empirical research. Analysis and inputs are recorded and documented for reference and guidance in the form of a framework referred to as the National Digital Citizenship Initiative Framework.

Key activities consist of focused groups and workshops to engage key stakeholders who are potential education delivery agents and research work through analysis of online documentary evidence and reports and exploration.

The design thinking methodology was employed for the engagement workshops with the main objective of identifying the root cause for the low scalability and sustainability of existing programs. Five key activities of design thinking were conducted to assist participants to deep dive into the existing contexts and challenges. The activities are:

- Deriving inspiration to understand the problem or the opportunity;
- Empathizing their direct audience to understand users' wants and needs, what might make their life easier and more enjoyable and how technology can be useful for them;
- Ideation for idea generation alternating between divergent and convergent thinking
- Prototyping to turn their ideas into something concrete
- Testing for getting feedback from users

Multiple engagements with key stakeholders, the Ministry of Education, MARA and National Parent Teacher Association, Government Linked Companies, NGOs, TELCOs, and tech industry offer verification and bring added value to the empirical evidence.

The research work provides empirical evidence from local and global practices. Combined with the engagement model approach, the NDCI's deployment model, content, programs, structure, objective, target audience, measurable indicators, and impact study are formulated.

#### **4. Findings**

The empirical evidence suggests that developed countries are already integrating digital citizenship in their teaching and learning activities in their public education system. With the imminent rise of social media and technology, education and agencies in Malaysia are already working with CyberSecurity Malaysia on some programs.

While relevant programs have been implemented, the lack of scalability and sustainability of programs remain as challenges. For education agencies, since digital citizenship is a non-core business, they lack expertise, budget and dedicated manpower to support implementation.

As for CyberSecurity Malaysia, there have been a lot of success stories in terms of pioneering CyberSAFE™ programs for raising awareness, conducting studies, developing content, establishing networks with agencies and global positioning. Moving forward, CyberSecurity Malaysia is seeking an approach that will minimize the constraints of scalability and sustainability.

Below are five areas that CyberSecurity Malaysia needs to address:

- No direct access to students, the main audience for digital citizenship. This is the main root cause for scaling and sustainability issues. Identifying and investing on agents with direct access will help scale and sustain.
- Face to face programs require huge resources to scale to the large volume of 450 thousand teachers, 10,000 schools and 5 million students that are spread across the country. An efficient approach needs to be considered.
- CyberSAFE content is intended to be delivered as a subject in itself is perceived as an added burden for teachers. Education agencies accept content that can be integrated in the teaching and learning process of other subjects in the main curriculum or within the co-curriculum. This requires training for teachers.
- Content awareness is low as it is mainly channelled through the website portal. Current users have low tendencies to search websites for content and expect links to content to be pushed through social media platforms.
- Content is varied in terms of programs and audience. A consolidated approach for a single overriding communication objective and less divergent audience segmentation are two areas for consideration.

#### **5. The solution**

##### **5.1 Rationale**

Findings from various global and local studies pointing to socio psychological risks on users, challenges emanating from users' lack of awareness and don't care attitude, and challenges due to the slow rate of response from relevant authorities, educators and parents, have created a sense of urgency on the need to establish a national digital citizenship initiative in order to raise awareness and protect citizens, especially children.

## **5.2 Why education?**

The global digital citizenship movement in many countries consider education as the most effective means to disseminate and build digital citizenship skills amongst the younger generation. CyberSecurity Malaysia shares this view that education is an effective and efficient channel to reach out to the end target audience – children.

Education as a promotional platform of digital citizenship skills and knowledge for the following reasons:

- Children are vulnerable to digital threats.
- School children are in the formative years where skills and knowledge acquired will be retained and shape their minds and personality.
- Schools provide direct access to children in their formative years.
- Digital tools are increasingly integral to teaching and learning activities. Learning while using the digital tools is a meaningful way of learning about digital citizenship.
- Digital citizenship can be an extension to the ICT subject or the various co-curriculum activities.

## **5.3 Deployment model**

There are two main components of the deployment model, which are partnership with education agencies and implementing programs.

### *5.3.1 Partnership with education agencies*

As practiced in most developed countries, the implementation of digital citizenship in schools is mainly carried out through partnerships between education departments and other government agencies responsible for internet safety and private organizations, such as NGOs and charities. The implementation of the NDCI requires coordinated efforts by multiple agencies led by CyberSecurity Malaysia. The collaborative approach with agents is an efficient way to scale and sustain the digital citizenship initiative at the national level. This is especially important to manage the resources at CyberSecurity Malaysia and to address CyberSecurity Malaysia's limited direct access to end users.

The approach is to leverage education. Education, as discussed above, is the most viable promotional platform of digital citizenship skills and knowledge. The implementation of digital citizenship in schools is mainly carried out through partnerships between education departments and other government agencies responsible for internet safety and private organizations, such as NGOs and charities having direct access to parents and other key stakeholders.

In the context of CyberSecurity Malaysia, the various departments and agencies within the Ministry of Education are agents with direct access to the teachers, the implementers and students, the beneficiaries. Other providers of public education, such as MARA, are considered agents. Private organizations and NGOs having direct access to parents and other key stakeholders are also agents. Collaborating with agents is an efficient way to scale and sustain the digital citizenship initiative at the national level.

### *5.3.2 Programs*

Four broad programs are identified as key to achieve sustainability and scalability in an efficient manner.

- Content Development and Delivery: The first aim is to review the current content and identify content that supports digital citizenship and the gap based on the requirements of the nine digital citizenship themes. Content development is a coordinated effort with the relevant public education agencies in Malaysia.
- Promotional Campaign: The campaign is essential to inform, persuade, or remind stakeholders, end users and the general public of the digital citizenship initiative. The goals are to enhance Cybersecurity's public image, improve knowledge and skills, galvanize the public, and garner support.
- Stakeholder Engagements: This program contains a series of engagements with officials at the relevant agencies and centrally organized meetings or forums involving decision makers with the presence of the sponsor. There needs to be a structure for sponsor, champions and implementer for the partnership with agents to be successful.

- Global Networking: Networking with global organizations that are spearheading the digital citizenship movement in different countries is a part and parcel of the initiative. This global networking is meaningful for making connections for sharing ideas and knowledge with similar organizations, build opportunities for Cybersecurity to share the digital citizenship initiative in education and acquire new approaches.

#### **5.4 Measurable indicators for impact assessment**

ISTE Digital Citizen standards (2018e) is a reference for the framework to measure the impact of the implementation of digital citizenship in schools. As a guideline to the Digital Citizen Standard, ISTE states that "Students recognize the rights, responsibilities and opportunities of living, learning and working in an interconnected digital world, and they act and model in ways that are safe, legal and ethical."

For this, ISTE further breaks down the indicators as the following:

- Students cultivate and manage their digital identity and reputation and are aware of the permanence of their actions in the digital world.
- Students engage in positive, safe, legal and ethical behaviour when using technology, including social interactions online or when using networked devices.
- Students demonstrate an understanding of and respect for the rights and obligations of using and sharing intellectual property.
- Students manage their personal data to maintain digital privacy and security and are aware of data-collection technology used to track their navigation online.

### **6. Conclusion**

While technology serves as learning stimulants, the potential threats caused by excessive exposure to technology and its inappropriate content, can have lasting harmful effects on young learners. The potential threats of technology have created a sense of urgency amongst all stakeholders responsible for delivering digital citizenship education especially for K 12 students. In taking Malaysia to the next level of cyber security, CyberSecurity Malaysia is poised to tackle issues beyond internet safety and security or 'digital literacy.' With the national digital citizenship program in alliance with education agencies, instilling digital citizenship skills through education will, in the long run, yield youths who are civic minded, who will not become victims and will not victimize others when online.

Inclusion of digital citizenship initiative education framework in the CyberSAFE™ Program will certainly contribute to the national sustainability, social well-being and wealth creation of cyber security ecosystem in Malaysia. The framework is envisaged be the baseline for developing more content, which could help parents and teacher to educate K 12 students as technology users with the appropriate norms and responsible users of technology. In digital citizenship; identifying the solution, developing the resources and teacher and parent development programs, establishing the implementation strategies and the impact assessment will be the key factors for digital citizens to sustain the social well-being in the future.

### **References**

- Ab Hamid, R., Yunos, Z. & Mustaffa, A., 2018. Cyber Parenting Module Development For Parents. In *12th Annual International Technology, Education and Development Conference 2018 (INTED2018)*. pp. 9620–9627.
- Aleksandra, 2018. Digital Childhood: Kids and Social Media. A Social Media and Psychology Blog. Available at: <http://socialmediapsychology.eu/2018/02/01/digital-childhood-kids-and-social-media>.
- Anon, 2018a. Digital Citizenship. *State of NSW, Department of Education and Training*. Available at: <http://www.digitalcitizenship.nsw.edu.au/documents/DCImplementation.pdf>.
- Anon, 2018b. Digital Citizenship Institute. Available at: <http://digcitinstitute.com/2018/09/27/our-digital-citizenship-community-mindset-digcitcommunity/>.
- Anon, 2018c. Digital Citizenship Modules. *Ministry of Education New Zealand*. Available at: <http://elearning.tki.org.nz/layout/set/lightbox/Technologies/Infrastructure/Digital-citizenship-systems-and-infrastructure-for-BYOD-and-GAFE>.
- Anon, 2017. Digital Readiness Blueprint. *Ministry of Communications and Information, Singapore*. Available at: <https://www.mci.gov.sg>.
- Anon, 2018d. DQ Institute of Singapore. Available at: <https://www.dqinstitute.org>.
- Anon, 2016. Gov. UK Statutory Guidance: Keeping Children Safe In Education. Available at: <https://www.gov.uk/government/publications/keeping-children-safe-in-education-2>.

- Anon, 2018e. ISTE Standards for Students. Available at: <https://www.iste.org/standards/for-students>.
- MCMC, 2017. Internet Users Survey. Statistical Brief Number Twenty One. *Malaysian Communications and Multimedia Commission*. Available at: <https://www.mcmc.gov.my/skmmgovmy/media/General/pdf/MCMC-Internet-Users-Survey-2017.pdf>.
- Preston, C. et al., 2018. Towards Tomorrow's Successful Digital Citizens: Providing the Critical Opportunities to Change Mindsets.
- Sharpe, R. & Beetham, H., 2010. Understanding Students' Uses of Technology for Learning: Towards Creative Appropriation. In *Rethinking Learning For A Digital Age: How Learners Are Shaping Their Own Experiences*. Routledge, pp. 85–99.
- Smahel D et al., 2015. Young Children (0-8) and Digital Technology. *Young Children (0-8) and Digital Technology: A Qualitative Exploratory Study Across Seven Countries*. Available at: [https://www.researchgate.net/profile/David\\_Smahel/publication/283398879](https://www.researchgate.net/profile/David_Smahel/publication/283398879).
- Yunos, Z., Ab Hamid, R.. & Mustaffa, A., 2017. Cyber Security Situational Awareness among Students: A Case Study in Malaysia. *World Academy of Science, Engineering and Technology, International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, 11(7), pp.1654–1660.

# Smart Citizens Wanted! How to act Responsibly With Data Security and Privacy?

Christine Ziske<sup>1</sup> and Ulf Ziske<sup>2</sup>

<sup>1</sup>Kikusema AB, Mariestad, Sweden

<sup>2</sup>KikuSema GmbH, Berlin, Germany

[christine.ziske@kikusemail.de](mailto:christine.ziske@kikusemail.de)

[ulf.ziske@kikusemail.de](mailto:ulf.ziske@kikusemail.de)

**Abstract:** The modern citizen is part of a worldwide digitalised system, in which (s)he is in a relationship with Third Parties, like, for example, e-commerce providers, Internet service providers, Trust centres and with the national government or other countries. The emerging concept of a Smart City requires a smart security, which is permanently evolving. This fact in turn requires smart citizens. The available resources of citizens when acting in their own interest cannot keep pace with the accelerated development of ICTs. Personal security will be reached by taking individual responsibility, by being the active party. We are convinced that only the competence of a human being is qualified to be a primary factor. But how to take responsibility? Today's authentication is based on Multi-factor Authentication (MFA) using combinations of passwords, software tokens, hardware tokens and biometrics. These currently available methods will be discussed and evaluated. The use of an eID is shown in different countries. As a user or consumer, one has to authenticate oneself several times a day. One has to use 'strong' passwords and is advised not to store these. Our concept 'Pass Sign' supports users in dealing with passwords or memorized secrets by augmenting their abilities. Thinking in patterns and remembering patterns is one part of human behaviour. Why not use a graphical interface by drawing an image on a virtual 'Compass Rose' and applying global things such as colours, directions, and patterns? Thereby, the ordinary keyboard with 94 characters can be replaced with a virtual key-board of the 65,535 characters from the UTF16 character set. The passwords are not stored at all. There is a demand for responsible action from all parties involved in the authentication process, including the citizen. That's why the citizen must be supported not only in the qualified use of passwords but also throughout the entire authentication process. The concept of the 'Five New Protocols' are aimed at improving the whole authentication process. Our concepts have the potential to be an innovative "survival kit" for smart citizens.

**Keywords:** human factor, graphical interface, authentication, memorized secrets, passwords, augmented intelligence

## 1. Setting the scene

According to Sun Tzu: "Whoever defeats the enemy without battle really understands warfare" (Sunzi, 2011). We must bring together all the tools already available for strong authentication and secured processes in an easy but secure way to avoid breaches of citizens' privacy and property. The password is not dead despite many predictions and the variety of alternative solutions. Users need many different passwords. The average amount is about 100 to 200 accounts per user. It is difficult to find unbreakable passwords and keep them secure without writing them down. We are convinced that the human factor is an essential factor within the authentication process and needs to be supported and augmented by tools. Thinking in patterns is very typical and easy for humans. That was the basis for our work. In addition, we are convinced that there is a need to think in different ways instead of repeating the ideas of authentication already existing for decades.

## 2. Smart citizens of a Smart City

There is a variety of definitions of what is meant by a Smart City. The Smart City should meet new challenges in terms of services of general interest, the efficient organization of urban transport or the supply of energy as well as the ensured access to health care and education. To address these challenges information and communication technologies (ICT) should make an important contribution. The Smart City is a concept which links the major challenges of rapidly rising urbanization with the application of technologies to a vision of a sustainable city of the future (Beinrott, 2015). Today the Smart City concept is being tested and implemented by technology companies primarily in major cities around the world. The city is not smart but the people who live in it are smart. Smart citizens are not simply inhabitants of a Smart City. Eventually, it's about the optimal usage of technology by humans: "(...) the smart city is how citizens are shaping the city in using this technology, and how citizens are enabled to do so. "(Schaffers et al, 2012, p.99). The citizen should be considered as an active user of technologies and should be given appropriate tools to deal with these technologies. In addition to online banking or online trading, the smart citizen will be faced by many more topics such as Internet of Things, Cloud Computing, Spatial Data Infrastructure, Artificial Intelligence, and e-Government. In order to use all these offered services, one has to authenticate countless times a day. The number of accounts to authenticate and

authorize has increased dramatically. The citizen must be empowered to meet these demands, because insecure authentication promotes cybercrime. The smart citizens need a smart booster to strengthen their own abilities.

### **3. Authentication today**

#### **3.1 Available solutions**

The National Institute of Standards and Technology NIST (2011) states that living in a smart city based on modern ICTs puts citizens in front of the request to authenticate their digital identities to different parties many times a day. They must be able to utilize efficient, easy-to-use, and secure solutions.

Generally, the three factors, namely, knowledge ('something you know'), possession ('something you have'), and biometrics ('something you are') are the basis for authentication. Knowledge-based authentication are passwords or more precisely 'Memorized Secrets'. Memorized Secrets include passphrases and PINs as well as passwords.

*"Authentication based on possession that means the use of One Time Passwords (OTPs) in the shape of hardware- and software tokens or the use of digital certificates based on Private Key Infrastructures (PKIs). The security of authentication is increased, but the use of PKIs is costly and time consuming" (Forst, 2014, p.7).*

Biometric methods include the recognition of physiological characteristics as fingerprint, face, iris, palm vein, and DNS or behavioral characteristics as writing behavior, lip movement, voice, and gear.

Authentication by biometrics is based on the evaluation of probabilities and not sufficiently protected against facsimiles. Biometric characteristics contain more than the information required for authentication, e.g. on gender, ethnic origin or state of health. The utilization is a convenient way for citizens but poses risks of fraud and restriction of privacy.

Today's authentication is based on Multi-factor Authentication using combinations of passwords, software tokens, hardware tokens and biometrics. Very common is the usage of the smartphone as a possessing factor in combination with the use of different sensors, e.g. fingerprint scanner. As a result, the smartphone will become an exceptionally valuable item that has to be protected against misuse.

Governments around the world legitimize digital solutions called electronic identification (eID) to handle the authentication of citizens' identity. Apart from online authentication the option to sign electronic documents were given. The usage varies in different countries.

In Sweden, 97.5 percent of the population uses the Mobile BankID for authentication and signing in many areas of their everyday lives. The focus is on the access to e-government services, within the fields of tax, health insurance, pension fund or education. The spread in use is increasing within the private sector too, in online banking and especially for the cashless transfer of small amounts. "3.3 billion transactions were registered in 2018" (Wemnelli, 2018, p.2). Since 2003 this type of eID has been available for PCs and since 2011 for mobile devices. In Austria, citizens are given access to public services online with a Citizen Card or Mobile Phone Signature. "Both alternatives can be used for providing evidence of identity and for the creation of legally valid signatures in online procedures. Numerous web services from both, the private and the public sector can be used e.g. tax assessment, insurance-data retrieval, criminal-record certificate, application for pension and child allowance" (Ziske, 2018, p33).

The German National Identity Card was introduced in 2010. It can be used for strong authentication on the Internet and to sign digital documents. Statista (2017) states that the number of users is below 15 percent. Only public services are offered. In the private sector, the eID is not used.

*"In Estonia over 90 percent of citizens file their tax returns digitally. Based on a Citizen Card practically every service, e.g. driver's license issue or parliamentary election can be handled electronically" (Deutschland Funk, 2018, p.1).*

The mobile payment app from Alibaba as a kind of digital ID card is used by 520 million users in China. Other 980 million Chinese have linked their ID card with the WeChat account. Both should be usable in all situations, but unfortunately also for monitoring the citizens. (Corum, 2018)

Since September 2018, there is a European-wide recognition requirement for the various eIDs. In future, EU-citizens can use their own eID in other EU countries to enroll at universities, register their trade, submit tax returns or apply for vehicle registration. This is a good contribution towards supporting the global smart citizen, who does not always belong to the nationality of the city in which (s)he lives.

### **3.2 Evaluation**

"The smartphone is becoming more and more the default device for Multi-factor authentication. It is used for receiving TANs by SMS or authenticator apps, for fingerprint or face recognition or as an NFC-device to detect additional hardware tokens, e.g. Citizen Cards. The price for the increased security is paid by the disclosure of your own telephone number" (Ziske, 2018, p.44). There is no anonymity anymore. The smartphone must not be lost, the phone number should not be changed. The increased protection reduces privacy. Furthermore, second factors of authentication are included without the consent of the user. The geolocation, specifically the IP address of a citizen is included as a second factor by Global Players, like Amazon, Facebook, and LinkedIn. Banks check typing behavior without user's agreement. The combination of behavioral and environmental factors is becoming more common. The legitimizing of eIDs by governments is a step in the right direction but the spread in all areas of life progresses too slowly. This is especially true in the private sector where ordinary logins with passwords only, are still in use. There must be solutions that work already today. There are only a few applications suitable for secure usage in a smart city. The use of the eID is combined with a PIN-code. PIN-codes are also passwords but composed of digits. This means the password is not dead. The password remains the controlling element. It means the citizen must be continually supported to handle the use of qualified passwords. Ordinary password manager or password vaults are not the solution because they require a master password.

How to handle this? Nobody can wait for the perfect final authentication infrastructures in a Smart City because the concept of a Smart City is already being introduced gradually. The citizen needs a 'survival kit' now to gradually improve authentication. The safest way of active participation is based on the cognitive abilities of human beings. That starts by using many different and complex password and time-limited passwords.

### **4. Boosting the human factor**

An essential component is the 'human factor'; the conscious and intentional human act. That's a very controversial topic. It is important to consider whether and when the human factor should be included or excluded. We believe that security can only be achieved through the involvement of the persons concerned.

One of the six design principles for military ciphers stated by Kerckhoffs (1883) is:

- it must be possible to communicate and remember the key without using written notes, and correspondents must be able to change or modify it at will.

These days this key is the password of a user. The access to user accounts and the protecting of data is usually based on password authentication.

The huge amount of attacks against passwords, like brute force (Kelley, et al., 2012), dictionary (Bonneau, et al., 2012) and social engineering makes it essential to use secure passwords. "At this point the user's ability to create 'secure passwords' came into the focus" (Bonneau, 2010, p.11).

A large range of studies have appeared on the subject of wrong or predictable user behaviour (Yan, et al., 2004; Taneski, Hericko and Brumen, 2014; Ur, et al., 2015) and how about choosing secure passwords (Schneier, 2014; Shay, 2014).

In addition, users need to cope with different password implementations at services on the Internet state (Horsch, et al., 2017). A lot of password requirements, which are fixed rules regarding password length and the allowed characters were postulated. These requirements are highly diverging between services (AlFayyadh et al, 2012).

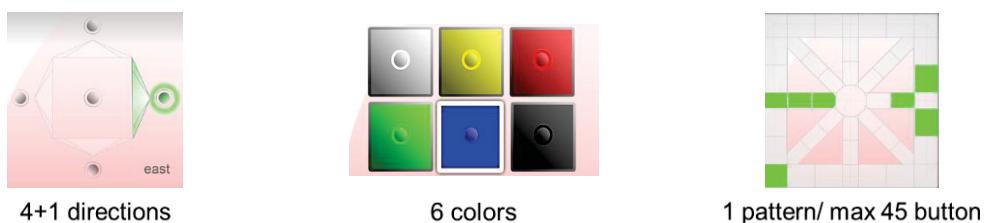
Last June the NIST-Guidelines on secure passwords were completely revised. It is no longer talking passwords but 'Memorized Secrets'. Memorized Secrets should be interpreted to include passphrases and PINs as well as passwords. Now the user would have the option of assigning 64-character long passwords of any characters, including the space. The Memorized Secrets should be simple and easy to remember. For example, the password

could be a sentence. And there is no need longer to change passwords regularly (Grassi et al, 2017). The new NIST-Guidelines on Digital Identities made some studies obsolete. (Bonneau, 2012) All this leaves the user in a real dilemma. Per se the user is completely on her/his own and needs practical help. Because we are convinced that the human factor and the Memorized Secret are the essential factors within the authentication process we were seeking tools suitable for supporting users and augmenting their abilities.

#### 4.1 Boosting by the concept of the Pass Sign

What if there was the possibility to handle this huge number of logins or accounts in an easy way? The solution could be an innovative graphical interface with an extremely good memorability. An image only existing in the user's mind will be needed. The image is called Pass Sign. We've developed an application to implement this concept. The APP FabulaRosa provides you with complex and long passwords by drawing an image on a virtual 'Compass Rose' and applying universal things such as colours, directions, and patterns. The 'Memorized Secret' in the form of passwords is not stored at all; it is generated in the moment you draw the image on the screen.

##### 4.1.1 FabulaRosa and Entropy



**Figure 1:** FabulaRosa- interface

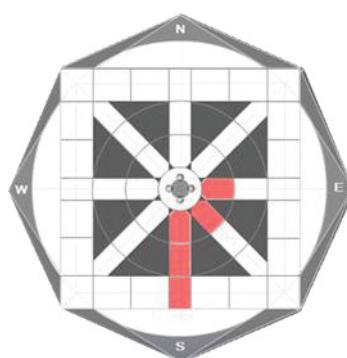
You can choose 5 directions, 6 colours and 45 buttons that equals 30 levels.

30 levels by 45 buttons equals 1350 possibilities to activate a button. That represent a combination amount of  $65,535 \cdot 1350 = 1.7^{e+6502}$ .

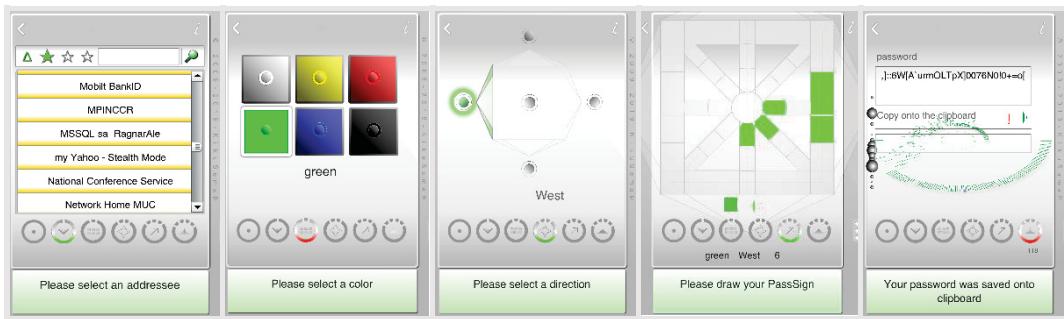
In other words, the extension of the Key Space, rather called Pass Space, is 65,535 with the power of 1350, which equals 1.7 with the power of  $e+6502$ . The entropy is  $1,350 \cdot \log_2(65,535) = 21,599.9703$ .

The calculation is based on the work of Claude Shannon (Roch, 2009) who introduced the term 'Entropy' on the advice of John von Neumann as a mathematical criterion of the sufficient randomness.

*"You should call it entropy. Nobody knows what entropy really is, so in a debate you will always have the advantage"* (von Neumann, 1940/41).



**Figure 2:** FabulaRosa - pattern interface



**Figure 3:** FabulaRosa – screens of the app

#### 4.1.2 *FabulaRosa* – What do you get

FabulaRosa provides regular passwords for ordinary logins as well as extremely long and complex passwords for high security applications. The passwords could have a maximum length of 850 with characters from the UTF16 character set which includes 65,535 characters. The passwords can be customized to requirements of the addressee that means the recipients of the login.

fhn1mkYXncfl9QAAZAp9akxW9vAp5AJ2LoMqAh1wAjMOAjVbS7yAuvW4RwEqsc

俞劄戛讐得鲲奢乖三京淨剗 | 蜈亩嫡

**Figure 4:** FabulaRosa – examples for passwords

#### 4.1.3 *FabulaRosa* – what are the algorithms behind

Behind the layout of ‘FabulaRosa Engine’ is a hidden universe, enriched with a so-called Pass Space with 1350 spheres. Each addressee has its own universe. The FabulaRosa Algorithm can be used for creating passwords and for encryption. The FabulaRosa Algorithm represents the engine of the whole concept and consists of the following parts.

- *The Pattern You Draw – The Pattern:* You can select 45 buttons, 6 colours and 5 directions. 6 colours by 5 directions that equals 30 levels. 30 levels by 45 buttons equals 1350 possibilities to activate a button. All buttons are linked with each other and each button correlates with a so-called *Mass Number*.
  - *The Addressees You Allocate – The Mass Number Algorithm:* A so-called *Alias Addressee* is created for each addressee. The Mass Number will be calculated by a certain algorithm based on this Alias Addressee.
  - *The User You Are:* Each user has its own so-called *Individual Security Constant* which ensures that two users using the same <pattern> will NOT use the same <password>.

The user gets an individual key set. These keys are transformed into the Individual Security Constant.

The password is calculated with a length of 1350 characters using the UTF16 character set.

The password is customized to the requirements of the addressee regarding the feasible length and the character set.

The whole app can be protected against unauthorized access by the so-called *Safety Start Algorithm*.

The entire user data, including all the information about the addressees are encrypted. One can't proceed within the app without drawing an additional pattern.

## **5. Boosting the whole authentication process**

The boosting of the Human Factor needs to be extended from qualified use of passwords to mastering the entire authentication process. This must be done qualitatively and quantitatively. The smart citizen will be faced by more and more parties. Perhaps the construction of a Smart City will soon resemble the Tower of Babel. By analogy, the different languages are now the different security services. Our general point of view is that the user has to be self-reliant concerning using modern technology. Today's authentication via smartphones becomes more and more essential for the user, both in private or public spaces, whether using trustworthy or untrustworthy devices. But how to handle authentications issues in unsure public environments and how to enforce the entire authentication process? Furthermore, the objective should be that the unauthorized access to users' sensitive data does no longer lead to serious consequences. How to realise that? Users' sensitive data must be stored and transmitted only with encryption. Each privacy data value must be encrypted by itself. The use of Memorized Secrets must be time-based otherwise the authentication process will be blocked.

### **5.1 Boosting by the Five New Protocols**

The following are Five New Protocols:

Protocol 1 – QR Code  
Protocol 2 – Stealth Mode  
Protocol 3 – Secret Based Login  
Protocol 4 - Encrypted User Data  
Protocol 5 – Multi Instance Authentication/Scrambled Secrets

These were developed to overcome current issues within the authentication process.

Firstly, because passwords can be intercepted and cracked the Stealth Mode was developed.

Secondly, because the addressees, e.g. web hosts probably protect user credentials insufficiently, sensitive data can be misused. That's why the idea of transferring 'content free strings' was born. Finally, there is a need to overcome the so-called Keyhole Security which means only locking the door of your safe by using a simple 'True-False-Authentication-Mechanism'.

There is the problem of authentication in public spaces with untrusted devices. This problem could be solved by Protocol 1. Between personal smartphones and public devices, a password is transferred by QR-Code.

By using Protocol 1 the real password or its hash is transferred. This disadvantage is solved by applying Protocol 2, the Stealth Mode. The password is changed within a certain frame so that it cannot be identified during the valid period.

With Protocol 3 the attack of the password transfer is avoided. Only a 'content-free string' will be sent to the user. During the handling of the login by the engine of FabulaRosa the string will be scrambled.

Protocol 4 enables the user to encrypt and decrypt complex data locally by using the engine of FabulaRosa. Complex data could be users' sensitive data, files, and applications.

The idea behind Protocol 5 or Scrambled Secrets is protecting your 'values' themselves by encrypting. This can be combined by using a Multi-Instance-Mode. Up to eight instances which could be a mixture of different technical factors with different simultaneous authorities can be involved in the process.

These protocols were already developed in 2010. Only the QR-Code is used by many users in different apps. The mass of data breaches has not caused service providers on the Internet to introduce real innovations, such as the 'Stealth Mode'. The concept of 'content free string' can improve the Cloud security. Most users' sensitive data is kept carelessly and not even encrypted in the providers' clouds. Scherschel (2019) states recently that trivial vulnerabilities in five of the world's largest web hosts endanger the users' credentials of 7 million domains. If the data only makes sense when converted back by the user, the hackers will not be able to exploit the intercepted data.

Innovations are being hesitantly introduced. The change of mindset and its implementation is an even more cumbersome task. Innovations must prevail against the ‘sluggish user’ who prefers simplicity rather than security. It is not easy to put Sun Tzu’s principles of being proactive into practice.

The following figure illustrates firstly the status of the implementation and that the security level will increase from protocol to protocol.

Secondly, it shows that the security level increases as the user participates more actively and/or third parties, like cloud service provider, web hosts or e-Government services invest in securing the entire authentication process.

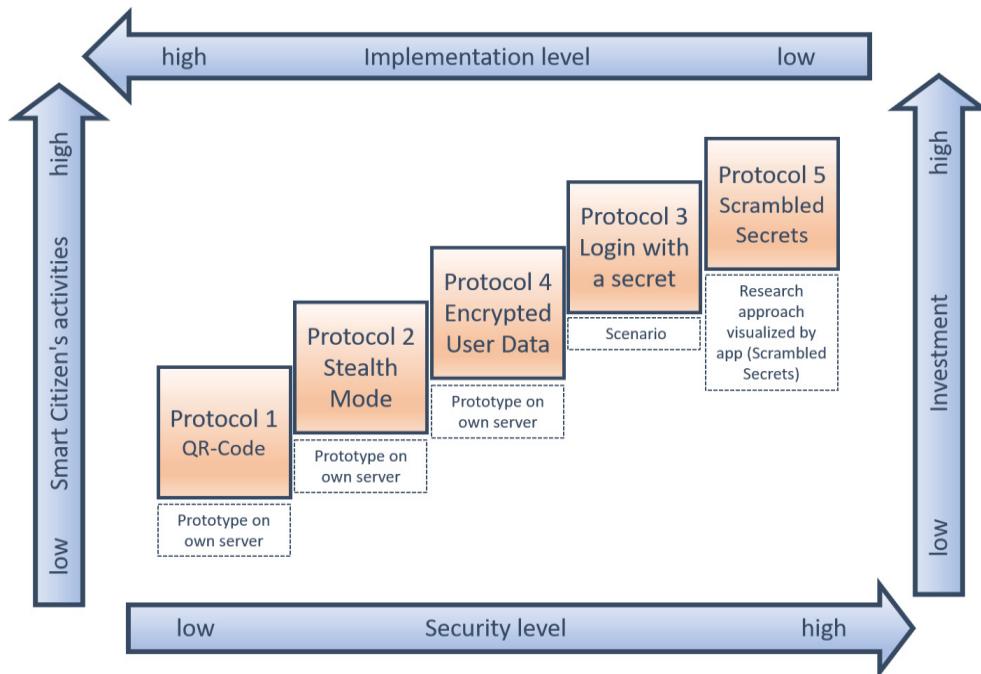


Figure 5: The Five New Protocols

## 6. Conclusions

The vision is to create a ‘citizen-centric authentication system’ in which personal data is handled securely by using solutions which will be Ease-of-use but making it more difficult for criminals to compromise online transactions. The citizens must have the confidence that their own digital identity, privacy and property is adequately protected.

The citizen must have the trust in the Smart City and all its technical and organizational components. Just as one climbs into a plane or a lift with the confidence that it will not crash. The complexity of the technical systems and the individual interests of all smart citizens must be considered.

Several authentication methods were offered today. Smartphones are often used within Multi-factor authentication methods or for the handling of eIDs. Biometrics and behavior are increasingly involved at the risk that privacy is infringed. Most of these solutions are dependent on the finally use of PIN-codes or, at best on Memorized Secrets.

The knowledge-based factor ‘something you know’ remains the determining factor of authentication.

We are aware that the human factor is both – the weakest and the strongest link in the chain.

But we are convinced that only the competence of a human being is qualified to be the primary factor in authentication and that the human factor has to be strengthened. Personal security will only be reached by taking individual responsibility, by being the active party.

Our concepts of ‘Pass Sign’ and ‘The Five New Protocols’ will empower the smart citizen to act responsibly with data security and privacy.

Our vision is that once adopted, our concepts and apps can help to assure that security, privacy, interoperability and ease of use will be significantly improved for all parties, which are involved in the authentication process. We are looking for committed partners to walk together on this path.

The News Needs Friends!

## References

- AlFayyadh, B., Thorsheim, P., Jøsang, A. and Klevjer, J. (2012) Improving usability of password management with standardized password policies. In *Proceedings of Sécurité des Architectures Réseaux et Systèmes d'Information*, Cabourg.
- Beinrott V. (2015) Bürgerorientierte Smart City Potentiale und Herausforderungen. [online] Available at: <<https://www.zu.de/institute/togi/assets/pdf/TOGI-150302-TOGI-Band-12-Beinrott-Buergerorientierte-SmartCity-V1.pdf>> [Accessed 16 December 2018].
- Bonneau, J. (2012) The science of guessing: analyzing an anonymized corpus of 70 million passwords. In: *Proceeding of the IEEE Symposium on Security and Privacy*. San Francisco, 2012, IEEE. pp 538–552.
- Bonneau, J., Herley, C., Van Oorschot, P.C. and Stajano, F. (2012) The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. In: *Proceeding of the IEEE Symposium on Security and Privacy*. San Francisco, 2012, IEEE. pp 553–567.
- Bonneau, J. and Preibusch, S. (2010) The password thicket: technical and market failures in human authentication on the web. In: Proceedings of the Ninth Workshop on the Economics of Information Security (WEIS) Cambridge, 2010.
- Corum, C. (2018) *Chinese digital ID comes to Alibaba's payment app*. [online] Available at: <<https://www.secureidnews.com/news-item/chinese-digital-id-comes-to-alibabas-payment-app/?tag=email>> [Accessed 11 June 2018].
- Deutschland Funk (2018) *Zwischen digitaler Moderne und Sowjetvergangenheit*. [online] Available at: <[https://www.deutschlandfunk.de/estland-zwischen-digitaler-moderne-und-sowjetvergangenheit.724.de.html?dram:article\\_id=408828](https://www.deutschlandfunk.de/estland-zwischen-digitaler-moderne-und-sowjetvergangenheit.724.de.html?dram:article_id=408828)> [Accessed 13 July 2018].
- Forst, C. (2014) *Sichere Authentifizierung – Teil 1: Klassische Methoden*. [online] Available at: <<https://conplore.com/sichere-authentifizierung-teil-i-klassische-methoden>> [Accessed 29 November 2018].
- Grassi, P.A., Fenton, J.L., Newton, E.M., Perlner, R.A., Regenscheid, A.R., Burr, W.E. and Richer, J.P. (2017) *Digital Identity Guidelines. Authentication and Lifecycle Management*. (NIST Special Publication 800-63, Version 1.0.2, pp 13–15) [online] Available at: <<https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-63b.pdf>> [Accessed 10 August 2017].
- Horsch, M., Braun, J. and Buchman, J. (2017) Password Assistance. In: *Proceedings of Open Identity Summit*. Bonn: Köllen Druck+Verlag GmbH., 2017.
- Kelley, P.G., Komanduri, S., Mazurek, M. L., Shay, R., Vidas, T., Bauer, L., Christin, N., Cranor L.F. and Lopez, J. (2012) Guess Again (Again and Again): Measuring Password Strength by Simulating Password-Cracking Algorithms. In: *Proceeding of the IEEE Symposium on Security and Privacy*. San Francisco, 2012, IEEE.
- Kerckhoffs, A. (1883) La cryptographie militaire. *Journal des Sciences Militaires*. 9, pp.5–38 & pp.161–191.
- NIST: National Institute of Standards and Technology (2011) *National Strategy for Trusted Identities in Cyberspace, Enhancing Online Choice, Efficiency, Security, and Privacy* [online] <<https://www.nist.gov/sites/default/files/documents/2016/12/08/nsticstrategy.pdf>> [Accessed 21 December 2018].
- Roch, A., (2009) *Claude E. Shannon: Spielzeug, Leben und die geheime Geschichte seiner Theorie der Information*. Berlin: Gegenstalt Verlag.
- Schaffers, H., Komninos, N., Pallot, M., Trousse, B., Tsarchopoulos, P., Posio, E., Fernandez, J., Hielkema, H., Hongisto, P., Almirall, E., Bakici, T., Lopez Ventura, J. und Carter, D. (2012), *Fireball - Landscape and roadmap of future internet and smart cities*. [online] Available at: <[https://hal.inria.fr/file/index/docid/769715/filename/FIREBALL\\_D2.1\\_M24.pdf](https://hal.inria.fr/file/index/docid/769715/filename/FIREBALL_D2.1_M24.pdf)> [Accessed 3 January 2019].
- Scherschel, F. (2019) *Triviale Hoster-Sicherheitslücken gefährden 7 Millionen Domains*. [online] Available at: <<http://www.heise.de/-4275552>> [Accessed 4 January 2019].
- Schneier, B. (2014) *Choosing secure passwords* [online] Available at: <[https://www.schneier.com/blog/archives/2014/03/choosing\\_secure\\_1.html](https://www.schneier.com/blog/archives/2014/03/choosing_secure_1.html)> [Accessed 23 March 2018]
- Shay, R., Komanduri, S., Durity, A.L., Huh, P.S., Mazurek, M.L., Segreti, S.M., Ur, B., Bauer, L., Christin, N. and Cranor, L.F. (2014) Can long passwords be secure and usable? In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '14. New York, NY: ACM, pp 2927–2936. [online] <http://doi.acm.org/10.1145/2556288.2557377>.
- Statista (2017) *Für welche Zwecke haben Sie die Online-Ausweisfunktion des neuen Personalausweises (nPA) bereits eingesetzt?*, [online] Available at: <<https://de.statista.com/statistik/daten/studie/777662/umfrage/nutzung-der-online-ausweisfunktion-des-npa-in-deutschland/>> [Accessed 5 January 2019]
- Sunzi. (2011) *Die Kunst des Krieges*. Translated by H. Eisenhofer. Hamburg: Nikol Verlagsgesellschaft.

- Taneski, V., Hericko, M. and Brumen, B. (2014) Password security – No change in 35 years? In: *Proceedings. 37th International Convention on Information and Communication Technology, Electronics and Microelectronics MIPRO*. Opatija, 2014, IEEE.
- Ur, B., Segreti, S.M., Bauer, L., Christin, N., Cranor, L.F., Komanduri, S., Kurilova, D., Mazurek, M.L., Melicher, W. and Shay, R. (2015) Measuring real-world accuracies and biases in modeling password guessability. In: USENIX, *Proceedings of the 24th USENIX Security Symposium*. Washington, D.C., 2015.
- Von Neumann, J. (1940/41) Conversation, between John v. Neumann and Claude Shannon, occurred between fall 1940 to spring 1941 at the Institute for Advanced Study, Princeton, New Jersey. [online] Available at: <http://www.eoht.info/page/Neumann-Shannon+anecdote> [Accessed 01 May 2018].
- Wemzell, M. (2018) Statistik BankID – användning och innehav – fördjupning. [online] Available at: <https://www.bankid.com/assets/bankid/stats/2018/statistik-2018-12.pdf> [Accessed 13 December 2018].
- Yan, J.J., Blackwell, A.F., Anderson, R.J. and Grant, A. (2004) Password Memorability and Security: Empirical Results. *IEEE Security & Privacy*, 2(5), pp 25-31.
- Ziske, C. (2018) Ist das Passwort tot? Sichere Authentifizierungsverfahren. Lecture at DITACT women's IT Studies, Salzburg 20th August 2018. [online] Available at: [http://ccc.ziske.de/wp-content/uploads/2018/08/9Password\\_ditact2018.pdf](http://ccc.ziske.de/wp-content/uploads/2018/08/9Password_ditact2018.pdf) [Accessed 19 December 2018].

**Work  
in  
Progress  
Papers**



# Towards Automated Threat-Based Risk Assessment for Cyber Security in Smarthomes

Pankaj Pandey<sup>1</sup>, Anastasija Collen<sup>2</sup>, Niels Nijdam<sup>2</sup>, Marios Anagnostopoulos<sup>1</sup>, Sokratis Katsikas<sup>1, 3</sup> and Dimitri Konstantas<sup>2</sup>

<sup>1</sup>Norwegian University of Science and Technology, Gjøvik, Norway

<sup>2</sup>University of Geneva, Switzerland

<sup>3</sup>Open University of Cyprus, Nicosia, Cyprus

[Pankaj.Pandey@ntnu.no](mailto:Pankaj.Pandey@ntnu.no)

[Anastasija.Collen@unige.ch](mailto:Anastasija.Collen@unige.ch)

[Niels.Nijdam@unige.ch](mailto:Niels.Nijdam@unige.ch)

[Marios.Anagnostopoulos@ntnu.no](mailto:Marios.Anagnostopoulos@ntnu.no)

[Sokratis.Katsikas@ntnu.no](mailto:Sokratis.Katsikas@ntnu.no)

[Dimitri.Konstantas@unige.ch](mailto:Dimitri.Konstantas@unige.ch)

**Abstract:** Cyber security is a concern of each citizen, especially when it comes to novel technologies surrounding us in our daily lives. Fighting a cyber battle while enjoying your cup of coffee and observing gentle lights dimming when you move from the kitchen to the sitting room to review your today's running training, is no longer science fiction. A multitude of the cyber security solutions are currently under development to satisfy the increasing demand on threats and vulnerabilities identification and private data leakage detection tools. Within this domain, ubiquitous decision making to facilitate the life of the regular end-users is a key feature here. In this paper we present a Risk Assessment Model (RAM), originating from Negative to Positive approach, to automate the threat-based Risk Assessment (RA) process, tailored specifically to the smart home environments. The calculation model application is demonstrated on derived threat-triggered evaluation scenarios, which were established from analysing the historical evidence of data communication within the smarthome context. The main features of the proposed RAM are identification of the existing risks, estimation of the consequences on possible positive and negative actions and embedding of the mitigation strategies. The application of this modelling approach for automation of RA would lead to a deep understanding on the extent to which decision making could be automated while tracking and controlling the cyber risks within the end-user's accepted risk level. Through the proposed RAM, common factors and variables are extracted and integrated into a quantified risk model before being embedded in the automated decision making process. This research falls within the GHOST (Safe-Guarding Home IoT Environments with Personalised Real-time Risk Control) project, aiming to provide a cyber security solution targeted at the regular citizens.

**Keywords:** risk assessment, IoT, security, smarthome

---

## 1. Introduction

The goal of the GHOST project (Collen *et al*, 2018) is to provide a cyber security solution targeted at the non-expert citizens by raising their awareness and understanding of the security risks associated with all aspects of cyber security from threats and vulnerabilities identification and personal data leakage detection up to making informed decisions affecting their cyber-physical smart home. GHOST aims to transform smart home occupants' decisions into reliable automated security service, promoting user-friendly end-user habits through usable security.

The Risk Assessment (RA) is a central functionality of the GHOST software implementation focused on the context-aware real-time threat protection. It gathers information about the current risks, analyses in real-time current network traffic flows and correlates them with the normal behaviour of the smart home. RA is responsible for determining at multiple stages in the processing of the data what the current Risk Level (RL) is. This RL is associated with a particular action a device or an end-user is about to take. RA validates real-time communication context using device behaviour profiles, entailing the processing of the communication context properties. The fusion of the RLs accepted according to user preferences and of typical behaviour stored in security patterns allows an automatic decision making, where RLs matching and comparison indicates the appropriate security action: allowing or blocking the whole communication stream, or propagating the intervention to the user interface for the end-user's approval or correction.

The structure of this paper is as follows. The recent advancements in the field of Behaviour Analysis (BA), Risk Prediction and Estimation (RPE) and Mitigation Techniques (MT) are presented in Section 2. Section 3 explains

the Risk Assessment and Modelling (RAM) approach, whereas the calculation of the RLs is demonstrated in Section 4. The application of RAM in a selected scenario is presented in Section 5. Finally, conclusions and directions for further work are summarised in Section 6.

## 2. Related work

Schiefer (2015) demonstrates the challenges that RA poses in a smart home installation due to the heterogeneous nature of the IoT devices. The spectrum of the threats for smart homes is twofold, namely privacy and security related. However, in most cases, the attacks are targeting both aspects. Unfortunately, the biggest problem still relies in primitive security settings that are ignored by unaware users. According to (Sivaraman, Habibi Gharakheili and Fernandes, 2017), multiple security incidents involving IoT devices exploit primitive attack vectors, such as the use of default passwords or weak communication protocols. The most notorious example is the break out of the Mirai botnet (Bertino and Islam, 2017), that took over at least 100,000 IoT devices. From the above, it is evident that a non-expert user has no way to perceive the full picture of the potential risks involved in the smart home she is living in, and that an automatic security risk monitoring solution is essential.

**Behaviour Analysis:** One of the approaches widely used in proactively managing security incidents is BA. In the case of smart home security, BA can be applied directly on any existing network at the router/gateway entry/exit point of any smart home installation. In terms of the approaches used in BA, Machine Learning is the most common method used for anomaly detection. For example, Saad et al (2011), successfully identified malicious behaviour on the network by comparing application of several existing ML classifiers. Zhao et al, (2013) expanded the existing method with the use of the decision trees, allowing zero-day detection of the involvement in botnet activities. The framework proposed by Nari and Ghorbani (2013), aimed at detecting malware, is using behaviour graphs, improving the accuracy and false positive detection by incorporating graph attributes.

**Risk Prediction and Estimation:** In Kitchin and Dodge (2017) provide a risk overview for the case of smart cities. This survey can be considered the closest on the risk analysis, vulnerability and MT identification in the field of Cyber-Physical System (CPS) security. There, the authors determine five main vulnerability categories: a) Weak software security and data encryption, b) Use of insecure legacy systems and poor ongoing maintenance, c) Many inter-dependencies and large and complex attack surfaces, d) Cascade effects and e) Human error. The same categories are also applicable to the case of a smart home environment. Furthermore, Almohri et al, (2017) suggest to incorporate threat modelling for RA directly at the IoT device design stage, distinguishing three main approaches: attacker-, system- and asset-centric (Martins et al, 2015). Rao et al, (Rao et al., 2018) present a very promising approach, based on the execution time of the processes in a CPS environment. This approach is the closest to the work in GHOST, in terms of dynamic real-time RA.

**Mitigation Techniques:** Current research in the MT does not spread much further than providing generic recommendations for formal risk evaluation processes. The closest work presented in (Kitchin and Dodge, 2017), provides guidelines for smart cities environment. The authors recognise three main categories of MT: a) Security by design, b) Traditional security mitigation, and c) Formation of the core security teams within the administrative staff supporting infrastructure installations. However, no further dynamic and automatic solutions are presented in the relevant literature.

## 3. Proposed risk assessment model

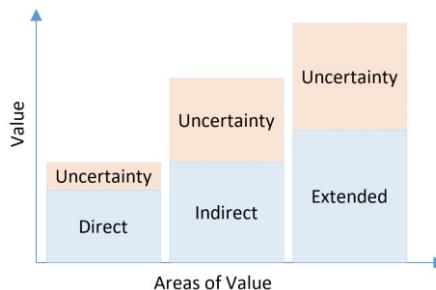
The approach taken for RA in GHOST involves the use of predefined Risk Levels (RL). “Negative to Positive Model” (ISO/IEC TR 27016:2014, 2014) was adapted for RL definition relying on four-dimensional correlation between values and activities. This model assesses risk on the basis of the cost (or benefit) associated with the option to either proceed with an action or not and turns negative values (cost) to positive (yield/return). Use of this model in our case results into the definition of four RLs, as shown in **Table 1**.

**Table 1:** Risk level definitions

	Question	Example
RL <sub>1</sub>	What will the positive value be if an activity is done?	compliance with privacy laws thus at the lowest level of risk in failing the compliance
RL <sub>2</sub>	What will the positive value be if an activity is not done?	collecting anonymised user information thus at a slightly higher level of risk in the event of failure of anonymisation technique and/or data theft

	Question	Example
RL <sub>3</sub>	What will the negative value be if an activity is done?	collecting personal information and sharing the data with unauthorised third party
RL <sub>4</sub>	What will the negative value be if an activity is not done?	not anonymising the user data and paying penalty for the misuse of the data

The Basic Value Model (BVM) (*ISO/IEC TR 27016:2014*, 2014) is used to estimate the positive or negative value involved in each RL. The principle of BVM which is based on three different characteristics is shown in Figure 1.



**Figure 1:** Principle basic value model

With reference to the BVM, the following definitions apply:

**Direct Values:** direct economic values, such as failure of a device, or direct investment based on an occurrence which could be active or passive.

**Indirect Values:** the additional and more intangible values gained or lost, having a greater uncertainty and as such they can be within ranges. For example, the unavailability of services due to DDoS attacks or increased administrative tasks.

**Extended Values:** reflect the values affected by the direct and indirect values and can be significantly huge and are also affected by other factors, such as impact on society and/or the GHOST network as a whole. Extended values of items such as brand or reputation are often difficult to quantify. Extended values are mostly negative but may also be positive as a consequence when information security is applied.

Addressing the four RLs and corresponding questions (**Table 1**) in combination with the principle BVM, led to the creation of a balance board to assure coverage of all risk relevant aspects. The potential duplication of the values related to the same activity is consecutively handled by using a simple balance table as shown in **Table 2**. Note that the factor C (Cost) is not applicable to the formulation used, but is part of the original BVM model.

**Table 2:** Balance table for net values

	Base	Activity	Positive Value Activity Done	Positive Value Activity Not Done	Negative Value Activity Done	Negative Value Activity Not Done	Net
Ref			A	B	C	D	
1	A possible activity to change the current situation	Activity "XY" done	Value	Not Applicable	Cost	Not Applicable	A1-B1
2	The possible activity not done	Activity "XY" not done	Not Applicable	Value	Not Applicable	Cost	B2-D2

#### 4. Risk exposure calculation

Estimation of risk exposure at different RLs is based on incorporating a multitude of Influence Factors (IF). Their listing along with the current integration status in the RAM is outlined in **Table 3**.

**Table 3:** Types of influencing factors

Type of IF	Description	Status	Reasoning
Physical	Sum of the tangible assets that comprise the GHOST network	Yes	Devices, sensors, or any IoT assets in a smart home
Customer/User	Smart home residents/owner	No	Perception factor, to be quantified

Type of IF	Description	Status	Reasoning
Societal	Perception that the society in general has about an appliance/device in the GHOST network and network as a whole	No	Perception factor, to be quantified
Reputational	Perception that competitors, suppliers, customers, government and other stakeholders have about the devices in the network and services provided by the GHOST network	No	Perception factor, to be quantified
Intangible/ Logical	Intangible assets handled by the GHOST network such as user data, forms of consent, blacklisted IP addresses, software integrity, etc.	Yes	Information/data and services generated/available in a smart home
Legal and Regulatory	Potential sanctions and/or penalties that might result from a breach	Yes	Data protection regulations, service contracts and legal obligations

The calculation model for RLs is defined as follows and is based on the balance table (**Table 2**):

$$RL_1 = T \times (V_1 \times A) \quad RL_2 = T \times (V_2 - AC_1) \quad RL_3 = T \times C \quad RL_4 = T \times (AC_1 + AC_2)$$

Where  $T$  = Time period,  $V_1$  = Value created by taking an action,  $A$  = Risk reduction as a result of action taken,  $V_2$  = Value created by not taking an action,  $AC_1$  = Additional internal cost,  $C$  = Cost associated with an action,  $AC_2$  = Additional external cost. Steps determining the RL in relation to an action taken:

- 1. If the action is completed, then go to step 2 else go to step 3.
- 2. If  $RL_1 > RL_3$ , then  $RL = RL_3$  else  $RL = RL_1$
- 3. If  $RL_2 > RL_4$ , then  $RL = RL_4$  else  $RL = RL_2$

## 5. Demonstration and evaluation

We use a scenario based approach, a common practice in Design Science Research Method (Samuel-Ojo et al, 2010) for ongoing work, to demonstrate and evaluate the application of the proposed RAM in the given scenario.

**Example scenario – A to B communication:** Internal IoT device A (**Table 4**) is sending data to malicious entity B (malware.com). B is already blocked by GHOST firewall (e.g. iptables).

**Table 4:** Device exposure vectors

Device	Exposure	Data
IP static camera	Wi-Fi connection, Motion detection, Remote control, Night vision, Video & sound capturing, Face recognition	System status, Configuration data, Video frames, Credentials, Facial profiles

Possible GHOST actions to take on this suspicious situation are listed in **Table 5**.

**Table 5:** Action and consequence correlation

Action	Positive Consequences	Negative Consequences
Block outgoing communication from device A to B	Controlled traffic, Avoiding privacy infringement of data sent to malware.com, Avoiding ransomware attack	Partial service disruption, User discomfort as no alert is received
Block all outgoing communication from device A	Controlled traffic, Avoiding ransomware attack	Full service disruption, Exposure to theft
Allow outgoing communication from device A to B	Continuous monitoring of sick (elderly) person, Physical security monitoring	Remote control by unauthorised party, Privacy violation, Involvement in DDoS, Potential danger in extreme scenario, GDPR regulatory fine, Ransomware

**Application of the Proposed Model:** The proposed RAM is applied to the above-mentioned scenario, and few assumptions are made for the data used in the calculations below to demonstrate the positive and negative values of doing or not doing the required action.

$RL_1$ : Positive Value – Activity Done

Let us assume that by removing the device from the network, we gain a positive value of EUR 5000 (from the positive consequences as listed in outlined scenario). Time period under consideration is 1 day. Risk reduction for the GHOST network in the given home is 90%.

Hence,  $T = 1$ ,  $V1 = 5000$ ,  $A = 90\%$ . Therefore,  $RL1 = 1 \times (5000 \times 0.9) = 4500$ .

*RL<sub>2</sub>: Positive Value – Activity Not Done*

Let us assume that by not removing the device from the network, we gain a positive value of EUR 3000 (from the positive consequences as listed in outlined scenario). Further, there is an additional cost associated with the unwanted data flow between A to B, which we assume as EUR 1000.

Hence,  $T = 1$ ,  $V2 = 3000$ ,  $AC1 = 1000$ . Therefore,  $RL2 = 1 \times (3000 - 1000) = 2000$ .

*RL<sub>3</sub>: Negative Value – Activity Done*

Let us assume that the negative consequences are critical in nature and by applying a method like Cyber Value-at-Risk (CVaR) for the above consequences as listed in outlined scenario, we get an estimated cost (negative consequence) of EUR 8000.

Hence,  $T = 1$ ,  $C = -8000$ . Therefore,  $RL3 = 1 \times (-8000) = -8000$ .

*RL<sub>4</sub>: Negative Value – Activity Not Done*

Since the device is not removed, the associated external cost is estimated by using a method like Single Loss Expectancy (SLE) for the above-mentioned negative consequences as listed in outlined. Let us assume that by applying SLE we get EUR 10000.

Hence,  $T = 1$ ,  $AC1 = 1000$ ,  $AC2 = -10000$ . Therefore,  $RL4 = 1 \times (1000 + (-10000)) = -9000$ .

Based on the output values at the respective risk levels for the given scenario, the user can take an appropriate risk management decision whether or not to take the underlying action.

## **6. Conclusion and future work**

The RAM presented in this paper is currently an ongoing research and development effort and is at the heart of the GHOST solution for RA. Deployed at the network traffic capture level, the incoming data is constantly monitored and fed into several distinct analysers. The resulting output is a set (zero or more) of risk related properties. Further grouped into identified risks, they serve as a base for the exposure value calculation. Various RLs at multiple stages of data processing are evaluated and monitored to ensure permitted RLs of current activity at each case, practically determining the required action to be taken. Experimental evaluation of the risk boundaries is enabling further fine-tuning of the calculation model to achieve automatic risks assessment. It is envisioned to perform several iterations of the model values refinement through the data obtained during the trials. Furthermore, a process on effective allocation and association of the mitigation actions should be identified. The current prototype relies on the hard-coded set of the actions extracted from the set of predefined attack scenarios.

## **Acknowledgements**

This work is partially funded by the European Union's Horizon 2020 Research and Innovation Programme through GHOST project under Grant Agreement No. 740923.

## **References**

- Almohri, H. et al. (2017) 'On Threat Modeling and Mitigation of Medical Cyber-Physical Systems', *Proceedings - 2017 IEEE 2nd International Conference on Connected Health: Applications, Systems and Engineering Technologies, CHASE 2017*, pp. 114–119. doi: 10.1109/CHASE.2017.69.
- Bertino, E. and Islam, N. (2017) 'Botnets and Internet of Things Security', *Computer*, 50(2), pp. 76–79. doi: 10.1109/MC.2017.62.

- Collen, A. et al. (2018) 'GHOST - Safe-guarding home IoT environments with personalised real-time risk control', in *Communications in Computer and Information Science*, pp. 68–78. doi: 10.1007/978-3-319-95189-8\_7.
- ISO/IEC TR 27016:2014 *Information technology -- Security techniques -- Information security management -- Organizational economics* ISO/IEC (2014). Geneva, CH.
- Kitchin, R. and Dodge, M. (2017) 'The (In)Security of Smart Cities: Vulnerabilities, Risks, Mitigation, and Prevention', *Journal of Urban Technology*. Taylor & Francis, 0(0), pp. 1–19. doi: 10.1080/10630732.2017.1408002.
- Martins, G. et al. (2015) 'Towards a systematic threat modeling approach for cyber-physical systems', *Resilience Week (RWS)*, 2015, pp. 1–6. doi: 10.1109/RWEEK.2015.7287428.
- Nari, S. and Ghorbani, A. A. (2013) 'Automated malware classification based on network behavior', in *2013 International Conference on Computing, Networking and Communications (ICNC)*. IEEE, pp. 642–647. doi: 10.1109/ICCNC.2013.6504162.
- Rao, A. et al. (2018) 'Probabilistic Threat Detection for Risk Management in Cyber-physical Medical Systems', *IEEE Software*, 35(1), pp. 38–43. doi: 10.1109/MS.2017.4541031.
- Saad, S. et al. (2011) 'Detecting P2P botnets through network behavior analysis and machine learning', in *2011 Ninth Annual International Conference on Privacy, Security and Trust*. IEEE, pp. 174–180. doi: 10.1109/PST.2011.5971980.
- Samuel-Ojo, O. et al. (2010) 'Meta-analysis of design science research within the IS community: Trends, patterns, and outcomes', in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer, pp. 124–138. doi: 10.1007/978-3-642-13335-0\_9.
- Schiefer, M. (2015) 'Smart Home Definition and Security Threats', *Proceedings - 9th International Conference on IT Security Incident Management and IT Forensics, IMF 2015*, pp. 114–118. doi: 10.1109/IMF.2015.17.
- Sivaraman, V., Habibi Gharakheili, H. and Fernandes, C. (2017) *Inside job: Security and privacy threats for smart-home IoT devices*, Australian Communications Consumer Action Network. Sydney. Available at: <https://www.runnersworld.com/running-gear/inside-job>.
- Zhao, D. et al. (2013) 'Botnet detection based on traffic behavior analysis and flow intervals', *Computers & Security*. Elsevier Ltd, 39(PARTA), pp. 2–16. doi: 10.1016/j.cose.2013.04.007.

# Applicability of Resilience Metrics in the Context of Telecommunications Services

Tarja Rusi

University of Jyväskylä, Finland

[Tarja.a.rusi@student.jyu.fi](mailto:Tarja.a.rusi@student.jyu.fi)

**Abstract:** This study evaluates the applicability of WEF resilience metrics framework as a tool for business owners in the telecommunications industry. Contact center and customer service solutions are critical communication services in the modern digital society. Digital society is highly dependent on customer service, be it digital, self-service or traditional phone service. Communication services and platforms are also an attractive target for nation-state cyber troops and criminals. Moreover, some of the cyber-attacks are specifically targeted to harm, steal and destruct critical customer data in the systems. These systems are also becoming more complex having more vulnerable components. Also, the service processes behind the services are becoming more complex with partners locating in another country. It is argued that risk analysis, risk management and business continuity management approaches are not sufficient to capture the complexities of uncontrolled, unexpected events. These approaches suit well in the foreseeable and calculable situations. They are not, however, sufficient to focus on increasing risks in the modern cyber-physical world having non-foreseeable, non-calculable stress situations. Thus, resilience must be built into the systems and it must have ownership, attention, proper ways to measure it and an interest to implement relevant metrics framework as a regular review. This study reviews briefly the frameworks of NIST, Linkov, and WEF and uses WEF to evaluate its applicability into real-life situation in a telecommunications service. The research question is how WEF framework can be applicable and possibly be improved as a tool of continuous review for business owners of communication platforms. The study uses case study method. The subject of the research is a particular case, a telecommunications company and a contact center solution. The aim is that the tested and further developed framework is general enough to be utilized by other service owners and platforms as well.

**Keywords:** resilience, resilience metrics, critical infrastructure protection, telecommunications

---

## 1. Background

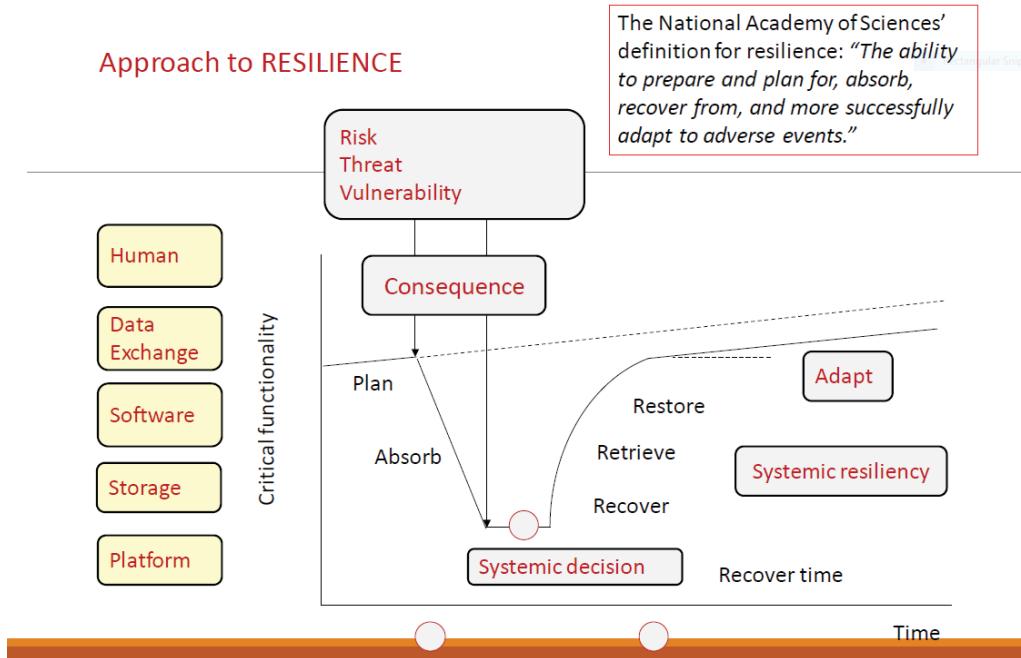
Digital society is highly dependent on customer service, be it digital, self-service or traditional phone service. Communication services and platforms are an attractive target for nation-state cyber troops as well as criminals. Moreover, some of the cyber attacks are specifically targeted to harm, steal and destruct critical customer data in the systems. These systems are also becoming more complex having more and more vulnerable components. Also, the service processes behind the services are becoming more complex with partners locating in another country. From the service owner perspective, there is a need to implement tools which help service owners get reliable and up-to-date information on the resilience status of services. It is challenging to have a needed overview due to the complexity and interdependency of the critical communication platforms. Thus, it is argued (ie. Collier et al., 2014; Erol et al, 2010; Rajamäki, 2017) that both risk analysis, risk management and business continuity management approaches are not sufficient to capture the complexities of uncontrolled, unexpected events. These approaches suit well in the foreseeable and calculable situations. These approaches are not, however, sufficient to focus on increasing risks in the modern cyber-physical world (the concept of cyber-physical: Song, Fink and Jeschke, 2017; National Science Foundation, cyber-physical program, 2018; Hu, Lu et al., 2016) having non-foreseeable and non-calculable stress situations. Resilience must be built into the systems and it must have ownership, attention, proper ways to measure it and an interest to implement relevant metrics framework as a regular review.

## 2. The concept of resilience and resilience metrics

Resilience refers to “the action or an act of rebounding or springing back; rebound, recoil”. Secondly, resilience can be understood as “elasticity; the power of resuming an original shape or position after compression, bending etc.” Thirdly, it can have a meaning of “the quality or fact of being able to recover quickly or easily from, or resist being affected by, a misfortune, shock, illness etc.; robustness, adaptability. (Oxford English Dictionary). Encyclopaedia Britannica, on the other hand, defines resilience as “the ability of a strained body to recover to its original size and shape after being compressed, bent, or stretched”. Secondly, it suggests that resilience can also be “the ability to recover from or adjust easily to misfortune or change”. (Encyclopaedia Britannica)

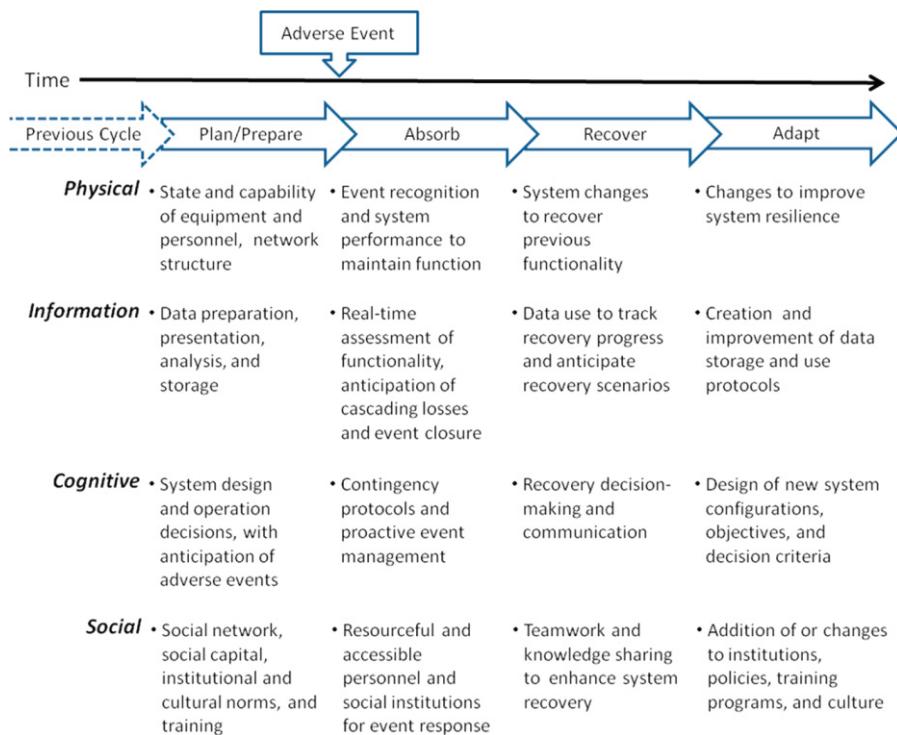
ENISA (2011) suggests that resilience metrics should be quantifiable, repeatable, and comparable to allow for viable and accurate comparison of different measurements. Secondly, they should also possess some non-

technical business characteristics. Thirdly, they should be easily obtainable, relevant to the business mission, and work toward the continuous improvement of resilience. Linkov et al. (2013) have presented a multi-level resilience matrix having the layers of physical, information, cognitive and social. They have also drafted a resilience-phasing model and integrated the layers with the event phases. The operational domains of physical, information, cognitive and social are derived from Network-Centric Warfare (NCF) doctrine developed by Alberts



**Figure 1:** Resilience (Rajamäki, 2017) Original: National Academy of Sciences (2012)

The National Academy of Sciences (2012) definition of resilience stems from the fact that it is necessary for any organization to have the ability to prepare and plan for, absorb, recover from, and more successfully adapt to adverse events.



**Figure 2:** Resilience matrix by Linkov, I. et al. (2013)

The National Institute of Standards and Technology (NIST) have published their framework for improving critical infrastructure cybersecurity (2014).



**Figure 3:** NIST Cybersecurity framework (National Institute of Standards and Technology, 2014)

World Economic Forum (WEF, 2016) has further developed the Linkov and NIST frameworks to better address the forum's core principles. WEF framework utilizes the NAS components of Linkov et al. framework except for the "Detect" from NIST framework. The domains of physical, information, cognitive and social refer exclusively to the Network-Centric Warfare.

	<i>Plan &amp; Prepare</i>	<i>Detect</i>	<i>Absorb</i>	<i>Recover from</i>	<i>Adapt to</i>
<i>Physical</i>	<ul style="list-style-type: none"> <li>(1) Implement controls/sensors for critical assets</li> <li>(2) Implement controls/sensors for critical services</li> <li>(3) Assessment of network structure and interconnection to system components and to the environment</li> <li>(4) Redundancy of critical physical infrastructure</li> <li>(5) Redundancy of data physically or logically</li> </ul>	<ul style="list-style-type: none"> <li>(1) Monitor the physical environment to detect potential cybersecurity events</li> <li>(2) Monitor personnel activity to detect potential cybersecurity events</li> </ul>	<ul style="list-style-type: none"> <li>(1) Signal the compromise of assets or services</li> <li>(2) Use redundant assets to continue service</li> <li>(3) Dedicate cyber resources to defend against attack</li> </ul>	<ul style="list-style-type: none"> <li>(1) Investigate and repair malfunctioning controls or sensors</li> <li>(2) Assess service/asset damage</li> <li>(3) Assess distance to functional recovery</li> <li>(4) Safely dispose of irreparable assets</li> </ul>	<ul style="list-style-type: none"> <li>(1) Review asset and service configuration in response to recent event</li> <li>(2) Phase out obsolete assets and introduce new assets</li> </ul>

	separated from the network (6) Protect data-in-transit				
--	---	--	--	--	--

### Tarja Rusi

<b>Information</b>	<ul style="list-style-type: none"> <li>(1) Inventory physical devices, systems, software platforms, and applications within the organization</li> <li>(2) Map organizational communication and data flows</li> <li>(3) Catalog external information systems</li> <li>(4) Categorize assets and services based on sensitivity or resilience requirements</li> <li>(5) Documentation of certifications, qualifications and pedigree of critical hardware and/or software providers</li> <li>(6) Prepare plans for storage and containment of classified or sensitive information</li> <li>(7) Identify external system dependencies</li> <li>(8) Identify internal system</li> </ul>	<ul style="list-style-type: none"> <li>(1) Detect malicious code</li> <li>(2) Detect unauthorized mobile code</li> <li>(3) Monitor external service provider activity to detect potential cybersecurity events</li> </ul>	<ul style="list-style-type: none"> <li>(1) Observe sensors for critical services and assets</li> <li>(2) Effectively and efficiently transmit relevant data to responsible stakeholders/ decision makers</li> <li>(3) Document, implement, and review audit/log records in accordance with policy</li> </ul>	<ul style="list-style-type: none"> <li>(1) Log events and sensors during event</li> <li>(2) Review and compare systems before and after the event</li> </ul>	<ul style="list-style-type: none"> <li>(1) Document incident's impact and cause</li> <li>(2) Document time between problem and discovery/discovery and recovery</li> <li>(3) Anticipate future system states post-recovery</li> <li>(4) Document point of entry (attack)</li> <li>(5) Categorize incidents consistent with response plans</li> <li>(6) Continuously improve protection processes</li> </ul>
--------------------	--	---	--	--	---

	dependencies				
<b>Cognitive</b>	<ul style="list-style-type: none"> <li>(1) Anticipate and plan for system states and events</li> <li>(2) Understand performance trade-offs of organizational goals</li> <li>(3) Scenario-based cyber wargaming</li> <li>(4) Include cybersecurity in human resources practices</li> <li>(5) Test response and recovery plans</li> </ul>	<ul style="list-style-type: none"> <li>(1) Analyze detected events to understand attack targets and methods</li> <li>(2) Aggregate and correlate event data from multiple sources and sensors</li> <li>(3) Determine impact of events</li> <li>(4) Establish incident alert thresholds</li> </ul>	<ul style="list-style-type: none"> <li>(1) Use a decision making protocol or aid to determine when event can be considered "contained"</li> <li>(2) Determine if mission can continue</li> <li>(3) Focus effort on identified critical assets and services</li> <li>(4) Utilize applicable plans for system state when available</li> </ul>	<ul style="list-style-type: none"> <li>(1) Review critical points of physical and information failure in order to make informed decisions</li> <li>(2) Establish decision making protocols or aids to select recovery options</li> </ul>	<ul style="list-style-type: none"> <li>(1) Review management response and decision making processes</li> <li>(2) Determine motive of event (attack)</li> <li>(3) Mitigate newly identified vulnerabilities or document as accepted risks</li> <li>(4) Understand the impact of incidents</li> </ul>
<b>Social</b>	<ul style="list-style-type: none"> <li>(1) Identify and coordinate with external entities that may influence or be influenced by internal cyber-attacks (establish point of contact)</li> <li>(2) Educate/train employees about resilience and organization's resilience plan</li> <li>(3) Manage identities and credentials for authorized devices and users</li> <li>(4) Manage and protect physical and remote access to assets</li> </ul>	<ul style="list-style-type: none"> <li>(1) Define roles and responsibilities for detection to ensure accountability</li> <li>(2) Communicate event detection information to appropriate parties</li> <li>(3) Continuously improve detection processes</li> </ul>	<ul style="list-style-type: none"> <li>(1) Locate and contact identified experts and resilience responsible personnel</li> <li>(2) Protect communications and control networks</li> <li>(3) Share effectiveness of protection technologies with appropriate parties</li> </ul>	<ul style="list-style-type: none"> <li>(1) Manage public relations and repair reputation after events</li> <li>(2) Communicate recovery activities to internal stakeholders and executive/management teams</li> <li>(3) Determine liability for the organization</li> </ul>	<ul style="list-style-type: none"> <li>(1) Evaluate employees response to event in order to determine preparedness and communications effectiveness</li> <li>(2) Assign employees to critical areas that were previously overlooked</li> <li>(3) Stay informed about latest threats and state of (the art protection methods/share with organization</li> <li>(4) Voluntarily share information with external stakeholders to achieve broader</li> </ul>

(5) Prepare/establish resilience communications (6) Establish a cyber-aware culture (7) Understand and manage legal and regulatory requirements regarding cybersecurity, including privacy and civil liberties obligations				cybersecurity situational awareness
--	--	--	--	-------------------------------------

**Figure 4:** WEF resilience metrics framework (World Economic Framework, 2016)

Linkov et al. (2013, 2014); Eisenberg et al. (2014) have further refined the resilience matrix to suit especially for cyber systems. The below table illustrates the core areas in the cyber resilience matrix.

**Table 1** The cyber resilience matrix

Plan and prepare for	Absorb	Recover from	Adapt to
<b>Physical</b>			
(1) Implement controls/sensors for critical assets [S22, M18, 20]  (2) Implement controls/sensors for critical services [M18, 20]  (3) Assessment of network structure and interconnection to system components and to the environment  (4) Redundancy of critical physical infrastructure  (5) Redundancy of data physically or logically separated from the network [M24]	(1) Signal the compromise of assets or services [M18, 20]  (2) Use redundant assets to continue service [M18, 20]  (3) Dedicate cyber resources to defend against attack [M16]  	(1) Investigate and repair malfunctioning controls or sensors [M17]  (2) Assess service/asset damage  (3) Assess distance to functional recovery  (4) Safely dispose of irreparable assets	(1) Review asset and service configuration in response to recent event [M17]  (2) Phase out obsolete assets and introduce new assets [M17]
<b>Information</b>			
(1) Categorize assets and services based on sensitivity or resilience requirements [S63]  (2) Documentation of certifications, qualifications and pedigree of critical hardware and/or software providers  (3) Prepare plans for storage and containment of classified or sensitive information  (4) Identify external system dependencies (i.e., Internet providers, electricity, water) [S31]  (5) Identify internal system dependencies [S63]	(1) Observe sensors for critical services and assets [M22]  (2) Effectively and efficiently transmit relevant data to responsible stakeholders/decision makers  	(1) Log events and sensors during event [M17, 22]  (2) Review and compare systems before and after the event [M17]  	(1) Document incident's impact and cause [M17]  (2) Document time between problem and discovery/discovery and recovery [S41]  (3) Anticipate future system states post-recovery  (4) Document point of entry (attack)
<b>Cognitive</b>			
(1) Anticipate and plan for system states and events [M18]  (2) Understand performance trade-offs of organizational goals  (3) Scenario-based cyber wargaming	(1) Use a decision making protocol or aid to determine when event can be considered "contained"  (2) The ability to evaluate performance impact to determine if mission can continue  (3) Focus effort on identified critical assets and services [M20]  (4) Utilize applicable plans for system state when available [M20]	(1) Review critical points of physical and information failure in order to make informed decisions [M17]  (2) Establish decision making protocols or aids to select recovery options  	(1) Review management response and decision making processes  (2) Determine motive of event (attack)

Social	(1) Identify and coordinate with external entities that may influence or be influenced by internal cyber attacks (establish point of contact)	(1) Locate and contact identified experts and resilience responsible personnel [S40]	(1) Follow resilience communications plan	(1) Evaluate employees response to event in order to determine preparedness and communications effectiveness [S18, 19]
	(2) Educate/train employees about resilience and organization's resilience plan [M17, S29]		(2) Determine liability for the organization	(2) Assign employees to critical areas that were previously overlooked [S22]
	(3) Delegate all assets and services to particular employees [S15, 22]			(3) Stay informed about latest threats and state of the art protection methods/share with organization
	(4) Prepare/establish resilience communications [S17]			
	(5) Establish a cyber-aware culture			

Rows represent the four domains taken from Network-Centric Operations doctrine (Alberts 2002); columns represent the four lifecycle stages of resilient systems defined by the US National Academy of Science (2012). Each cell represents metrics of resilience for each domain and lifecycle stage based on Allen and Curtis (2011) (Reference S#) and Bodeau and Graubart (2011) (Reference M#)

**Figure 5:** Cyber resilience matrix (Linkov et al., 2013)

Both Linkov et al. and WEF framework have many similar elements. However, WEF framework takes a step further from Linkov adding a “detect” element from NIST framework. Moreover, WEF framework is more business-process oriented having more elements in the cognitive and social dimensions. This is very important in self-learning organizational cultures. NIST framework also lacks the domains derived from network-centric warfare concept and is more requirement-oriented and risk-oriented as opposed to process-oriented approach of WEF. It should be notified, however, that all the studied frameworks have a similar systems-based view on cyber attack. They all seem to lack the more modern side of cyber attacking: attacks connected to hybrid warfare. In hybrid influencing a foreign state combines a variety of activities. These activities can take place in the political, economic, military, civil or information domains. They are conducted using a wide range of means and designed to remain below the threshold of detection and attribution. (HybridCoE, 2019).

In conclusion, this study uses WEF framework to test its applicability in the communications services to be used by business owners. For the purposes of this case study, WEF framework fulfills the need to have a business-process oriented approach with more elements in the social and cognitive areas. It will also be necessary to study how hybrid influencing should be taken into account in the framework.

### 3. Research problem

The task of this study is to test the WEF resilience framework in the critical communications platform service. The framework should follow the principles set by ENISA to be applicable in the real-life setting. The underlying purpose is to use empirical case data to determine the applicability and needed improvements for the framework.

The research questions are:

- How can WEF framework of cyber resilience be applicable and possibly be improved as a tool of continuous review for business owners of communication platforms?
- How should hybrid influencing be taken into account in the framework?

The research follows the principles of a case study research (Byrne & Ragin, 2009; Gomm & Hammersley, 2000; Travers, 2001; Farquhar, 2012; 2018). Yin (2009) defines the case study method as an empirical inquiry that investigates a contemporary phenomenon in depth and within its real-life context. The strategy of data analysis is relying on explanation building.

### 4. Conclusions

The study is important because it will have implications for both theory and practice. Firstly, there are not many studies where resilience metrics is tested in a real-world setting. Secondly, at least some of the metrics are originally made from American perspective (f.ex. Linkov), more cross-country data is needed to test the matrix constructs and their suitability. Thirdly, Linkov model does not adequately address the privacy requirements generated by EU-GDPR. Fourth, real data from critical infrastructure service provider point is scarce. Fifth, there is a need to find useful and applicable resilience metrics for business owners facilitating continuous and up-to-date status understanding of the service resilience. Sixth, the study aims at providing more understanding on the specifics of telecommunications services. Thus, there is a need for real-life evaluation and testing of applicability and meaningfulness of established measures and constructs of resilience metrics.

## References

- Alberts, D. (2002). Information age transformation, getting to a 21st century military. *DOD Command and Control Research Program*. Retrieved from <http://www.dtic.mil/get-tr-doc/pdf?AD=ADA457904>
- Byrne, D. S., & Ragin, C. C. (Eds.). (2009). *The SAGE handbook of case-based methods*. Sage, London.
- Collier, Z. A., DiMase, D., Walters, S., Tehranipoor, M. M., Lambert, J. H., & Linkov, I. (2014). Cybersecurity standards: Managing risk and creating resilience. *Computer*, 47(9), 70-76.
- Cyber-physical systems. In <https://ptolemy.berkeley.edu/projects/cps/>
- Eisenberg, D. A., Linkov, I., Park, J., Bates, M. E., Fox-Lent, C., & Seager, T. P. (2014). Resilience metrics: Lessons from military doctrines. *Solutions*, Vol.5, issue 5, 76-87.
- Encyclopaedia Britannica (2018). <http://academic.eb.com.ezproxy.jyu.fi/levels/collegiate/search/dictionary?query=resilience&includeLevelThree=1&page=1>. Accessed 19.2.2018.
- ENISA (2011). Measurement Frameworks and Metrics for Resilient Networks and Services: Challenges and Recommendations. <https://www.enisa.europa.eu/topics/critical-information-infrastructures-and-services/internet-infrastructure/metric>
- Erol, O., Henry, D., Sauser, B., & Mansouri, M. (2010). Perspectives on measuring enterprise resilience. Paper presented at the 587-592.
- Farquhar, J. D. (2012). *Case study research for business*. Sage, London.
- Farquhar, J. D. (2018). *Case study research for business*. Sage, London.
- Gomm, R., & Hammersley, M. (Eds.). (2000). *Case study method : Key issues, key texts*. Sage, London.
- Hu, F., Lu, Y., Vasilakos, A. V., Hao, Q., Ma, R., Patil, Y., . . . Xiong, N. N. (2016). Robust Cyber–Physical systems: Concept, models, and implementation. *Future Generation Computer Systems*, 56, 449-475.
- HybridCoE (2019). The European Center of Excellence for Countering Hybrid Threats. <https://www.hybridcoe.fi/what-is-hybridcoe/>
- Linkov, I., Bridges, T., Creutzig, F., Decker, J., Fox-lent, C., Kröger, W., . . . Thiel-clemen, T. (2014). Changing the resilience paradigm. *Nature Climate Change*, 4(6), 407.
- Linkov, I., Eisenberg, D., Plourde, K., Seager, T., Allen, J., & Kott, A. (2013). Resilience metrics for cyber systems. *Environment Systems and Decisions*, 33(4), 471-476.
- Linkov, I. et al. (2013). Measurable resilience for actionable policy. *Environmental Science & Technology*, 47 (18), pp 10108–10110.
- National Academy of Sciences. (U.S.). (2012). Disaster resilience: A national imperative. Washington, D.C.: National Academies Press. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=867788&site=ehost-live>
- National Institute of Standards and Technology. (2014). Framework for Improving Critical Infrastructure Cybersecurity. Retrieved from <http://www.nist.gov/cyberframework/upload/cybersecurity-framework-021214.pdf>
- National Institute of Standards and Technology (2018). Framework for Improving Critical Infrastructure Cybersecurity. Revised version. Retrieved from <https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf>
- National Science Foundation (2018). [https://www.nsf.gov/news/special\\_reports/cyber-physical/](https://www.nsf.gov/news/special_reports/cyber-physical/)
- Oxford English Dictionary. (2018). <http://www.oed.com.ezproxy.jyu.fi/view/Entry/163619?redirectedFrom=resilience#eid>. Accessed 19.2.2018.
- Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed ed.). Sage, CA.
- Rajamäki, J. (2017). Resilience of cyber-physical systems. Lecture at the University of Jyväskylä, Finland, ITKST65. 2017.
- Song, H., Fink, G., & Jeschke, S. (2017). Security and privacy in cyber-physical systems. Foundations, principles, and applications. eBook. 1<sup>st</sup> ed. Wiley, UK.
- Travers, M. (2001). *Qualitative research through case studies*. Sage, CA.
- World Economic Forum (2016). A framework for assessing cyber resilience. A report for the World Economic Forum. Prepared by Keys, Chhajer, Liu and Horner.
- Yin, R. K. (2009). *Case study research: Design and methods* (4th ed.). Sage, LA.

# Mitigating Return Oriented Programming

Lee Speakman<sup>1</sup>, Thaddeus Eze<sup>1</sup>, David Baker<sup>2</sup> and Samuel Wairimu<sup>1</sup>

<sup>1</sup>Computer Science Department, University of Chester, Thornton Science Park, Chester, Cheshire, UK

<sup>2</sup>Pennaf Housing Group, UK

[l.speakman@chester.ac.uk](mailto:l.speakman@chester.ac.uk)

[t.eze@chester.ac.uk](mailto:t.eze@chester.ac.uk)

[David.Baker@pennaf.co.uk](mailto:David.Baker@pennaf.co.uk)

[1720190@chester.ac.uk](mailto:1720190@chester.ac.uk)

**Abstract:** Code-reuse attack techniques, such as Return Oriented Programming (ROP), pose a significant threat to modern day systems as they are able to circumvent both traditional and more modern protection mechanisms such as antivirus, antimalware, Address Space Layout Randomisation (ASLR) and W⊕X/Data Execution Prevention (DEP). IT companies are actively researching ways in which ROP attacks can be mitigated, emphasising the importance of research in this area. Various defence mechanisms have been designed and developed to attempt to prevent ROP attacks, however, vulnerabilities still exist, and some attacks are still able to bypass these. This paper proposes a solution – ROPMit – that successfully mitigates ROP attacks without the caveats of other current research. ROPMit is a collection of base techniques that detects function boundaries and randomises at the function level the memory layout to mitigate against ROP, even when an info-leak is present, to reveal the address of part of the code section. ROPMit is implemented and tested on Linux 32bit binaries compiled with gcc. Testing is done on a binary with an info-leak and buffer overflow vulnerability on the call stack. A ROP attack attempts to call gadgets in the binary but is blocked by ROPMit with high likelihood. The likelihood of blocking an attack is proportional to the factorial of the number of functions present in the binary.

**Keywords:** code injection, ROP, ROPMit, security, ASLR, software protection

---

## 1. Introduction

Code injection attacks against the call stack of a target process have resulted in the development of countermeasures. These countermeasures include the use of canaries to check if an overflow has occurred, ASLR which affects the location of sections in memory at program load time, and W⊕X (Writable XOR eXecutable), specifically in the form of the NX (Non-eXecutable) / DEP bits on the memory page to render the page unable to have machine code instructions executable if delivered to the CPU from that page. These are all considered the minimum standard for running programs in modern computer systems, and are to be found on all common operating systems and in most programs.

Code injection attacks could no longer succeed with these countermeasures employed. ASLR causes a program to have different sections loaded into memory at a different address each time it is run. ASLR initially applied to the location of the call stack, so that after code injection, a ‘return’ into the injected code on the call stack might not succeed due to the absolute address not being known. Canaries attempted to detect and catch overflows before they can subvert the program. And NX/DEP on the stack prevented execution of machine code anyway.

Each of these countermeasures have their weaknesses. ASLR is only able to randomise to a certain degree. In 32bit memory address spaces only a few bits of randomisation may be possible, while in 64bit memory address spaces much more protection likelihood can be given. Canaries on the stack also attempt to detect stack corruption, but at what point in the flow of the program the check is made is vital to whether it is too late, i.e. the exploitation may have already happened before the canary checks are made. Furthermore, the availability of an information leak (info-leak) can be used to replace a canary value with itself, rendering it inert. NX/DEP still prevent execution of injected machine code. However, ROP exploits have been used by attackers to deactivate the security enforced by the NX/DEP. According to research, Carlini *et al.* (2015), Davi & Sadeghi (2015) and Schwartz *et al.* (2011), DEP/W⊕X was only effective against malicious code injection since attackers could no longer execute a malicious code that has been written through a data variable section to a RW-marked memory page, but not against code-reuse attacks like ROP.

ROP was generated as a counter-countermeasure to achieve the equivalent of code injection without injection of actual code. This works by injecting a sequence of return address which are continually popped off the stack and ‘returned’ to, to execute code of the attacker’s choice that is already in an executable part of memory. These

snippets of machine instructions, finalised with a ‘ret’ (return) instruction, are called **gadgets**. ASLR can still present a countermeasure to ROP; with Position Independent Code (PIC), the code (“.text”) section can also be loaded at a randomised address, along with the libraries, and also data and heap. This presents a challenge to an attacker to gain the right ‘return’ addresses of gadgets in executable memory.

The use of an Info-Leak can be used to bypass ASLR. If information can be gathered by an attacker on the position or likely position of any section of executable code, then all code will simply be at a relative and predictable offset to this leaked position, therefore rendering ASLR inert. A ROP-attack can then inject a series of return addresses with the known gadget address and other necessary data to be popped off the stack into registers. A further weakness of ASLR is that the randomisation is performed once at load time and thereafter the program will continue with the same randomised offsets for the duration of the running program. Depending on the program, this could be anything from a fraction of a second to years. Our proposal seeks to create a solution that randomises code and the addresses of code read into memory, in order to provide sufficient entropy to prevent ROP based attacks, without requiring access to source code, disassembly information or incurring a significant performance overhead.

## **2. Related work**

The current literature on mitigating ROP falls into 3 areas – Compilation, Dynamic Checking, and Randomisation. Koo et al. (2018) introduced a solution called compiler-assisted code randomisation (CCR), an approach that depends on a compiler-rewriter combination to assist in code randomisation on end-user systems. The concept of compiler-rewriter combination was discussed by Larsen et al. (2014) as a potentially interesting solution that would randomise a code thus offering a high entropy. The compiler method of mitigation is dependent upon altering the code of the program at the point it is compiled and according to Larsen et al. (2014), the randomisation entropy achieved by compiler-based approach was proved to be very high. While CCR is able to randomise the functions and basic block, studies have shown that compiler-based randomisations are not desirable as they require access to source code, custom compilers and need modifications to existing software distribution mechanisms (Larsen et al., 2014), Onarlioglu et. al. (2010) and Bletsch et al. (2011). Compiler based solutions are not very portable for end users.

Dynamic checking is the analysis of a program during execution. ROPdefender by Davi et al. (2011) is an example. Before executing an instruction, the instruction is checked to determine the instruction type, if the instruction type is a call, the return address is added to the shadow stack, if the instruction is a return, the address at the top of the program stack is compared with the instruction on top of the shadow stack, if these instructions do not match then there must have been corruption and execution is halted. The recent work by Intel (2017) highlights the relevance of research into this area. Dang et al. (2015) describes a parallel shadow stack that incurs a lower but still significant runtime performance overhead of 3.5%.

Gupta et al. (2016), Hiser et al. (2012) and Pappas et al. (2012) all propose methods of mitigating ROP attacks by randomising the address space in memory. With randomisation, in order to create a ROP attack, the attacker must be able to locate the necessary instructions to formulate gadgets. But by randomising the address space it is harder to locate or predict where the necessary instructions will be located in memory during runtime. Once the address space has been randomised, memory patching takes place; this is where the addresses to functions or instructions that have moved are updated to reflect the new location within memory. Randomisation mitigation methods have a number of drawbacks, including:

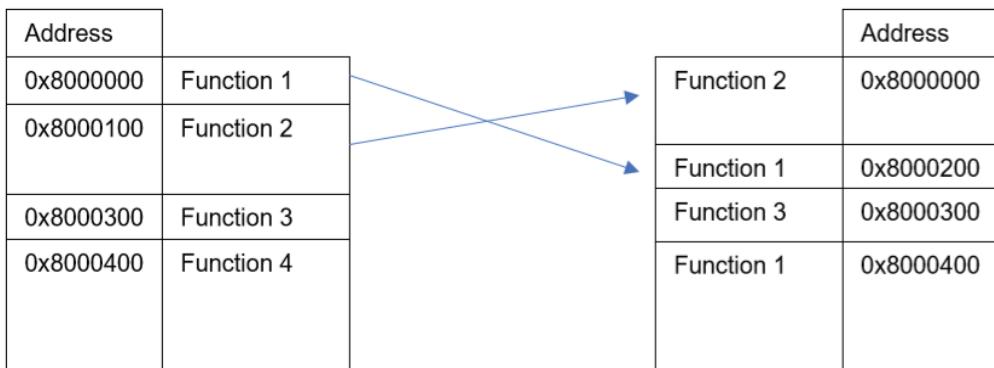
- Needs access to symbol information or able to restore symbol information (Gupta et. al., 2016).
- Need disassembler information (Gupta et. al., 2016, and Hiser et al., 2012).
- Use of 3rd party tools or programs, such as IDA Pro or objdump (Pappas et al., 2012, and Gupta et al., 2016).
- Additional space required (Pappas et al., 2012).
- Overhead during runtime (Hiser et al., 2012).

## **3. ROPMit collection**

Here ROPMit - ROP Mitigation – is proposed as a collection of base techniques to address the aforementioned known-offset problem (via info-leak) within a section and the issues associated with long-term processes running in memory and keeping the same initially randomised offsets. ROPMit is positioned to successfully thwart ROP

attacks without the major caveats of previous studies. ROPMit uses a very low disruptive method of randomisation which is based on swapping adjacent functions. The reasoning behind this is that there are no instructions between the functions that will require moving about. In addition, the address space taken up by the two functions will be the same after being swapped meaning that those instructions before or after the functions will also not require moving

ROPMit takes the approach of randomising the functions within a binary at load time or with a pre-load algorithm before being passed to the standard loader. This does not require knowledge of or information generated from the source code, it does not need additional complexity in the binary executable file, and can act on any binary so long as the code is compiled as PIC and is therefore relocatable by the standard ASLR. The binary is scanned to detect function boundaries. Once identified, a random decision can be made as to whether it will be swapped with the adjacent following function. Consider a simplified view of a binary with four functions, as shown in Figure 1. Here is shown Function 1 being swapped with Function 2.



**Figure 1:** Function relocation by adjacent function swapping

Adjacent functions are swapped so that the sum, at the time the swap takes place, takes up the same space in memory and does not shift the other remaining functions in location. This is iterated over the whole code section. Multiple parses proportional to the number of functions can completely randomise the location of the functions within the memory address space, though a limit can be put on the number of passes to preserve load time or once sufficient entropy is reached. As more functions are present, the number of possible combinations increase to the factorial.

#### 4. Protection capability

After the randomisation of the function locations within the binary, any info leak of the location of any one function within the running program does not reveal to an attacker all the potential ROP gadget locations within the whole program as the binary no longer have relative-addresses guarantees within the code section. This renders ROP significantly less likely to succeed, especially where the function count is great. After the functions have moved, anything that calls these functions will have to have the call addresses patched to the new location, additionally any calls or offsets within the functions will also need their addresses patching. This is known as Memory Patching – a method of updating the addresses of instructions or functions that have changed during the randomisation process (Figure 2).

The patching algorithm works by going through each of the instructions until a call instruction is encountered. Upon detecting a call instruction, the next 5 bytes which make up the offset are read. The target address of an offset is calculated by looking at the next address and adding it to the offset. The new offset is calculated based upon where the function has moved to, the old offset is then overwritten with the new offset. If the next instruction is an add to the EAX register (the code must then be PLT offset), the PLT offset is read, the new offset is calculated and then this new offset is written over the current PLT offset

#### 5. Implementation

The implementation was trialled on Linux with 32bit code generated from gcc – the GNU compiler collection. Function boundaries can be detected predictably in all trials leading to the preparation of function swaps. In addition to swapping functions, addresses of function calls must be updated along with jumps, as the location of the target code may have moved. This address patching was performed per swap-iteration; if code was moved, then the address patching was done in the same iteration.

```

FOR EACH INSTRUCTION
    IF instruction is a call
        Next Address = current instruction + 5
        Read next 5 bytes and store in Current Offset
        Target Address = Next Address + Current Offset
        New Offset = Target Address – Next Address + difference from old position
        Replace Current Offset with New Offset
    IF Next Instruction is Add to EAX register
        Next Address = current instruction + 5
        Read next 5 bytes and store in Current Offset
        Target Address = Next Address + Current Offset
        New Offset = Target Address – Next Address + difference from old position
        Replace Current Offset with New Offset
    END IF
END IF
END FOR

```

**Figure 2:** Pseudocode for Memory Patching calls within swapped functions.

A variety of function types were generated resulting in a variety of machine code generated with calls and jumps, including conditional jumps. Furthermore, additional patching to the Procedure Lookup Table (PLT) and Global Offset Table (GOT) was done to allow functions to use shared code. This required patching of the EAX register offset in order to make sure the PLT/GOT-dependent code continued to work. A number of optimisations can be performed that are too complex to cover in this short paper. Optimisation will be important for scalability – where the number of functions increases significantly. A high level explanation of the proposed randomisation is presented in Figure 3.

<pre> int main (int argc, char *argv[]) {     if(argc != 2) {         printf("USAGE: ropmit.out &lt;filename&gt; \n");         return 0;     }     //ensures there are at least 2 arguments      int fd = open(argv[1], O_RDWR);     struct stat statbuf;     fstat(fd, &amp;statbuf); //the argument, passed file, is mapped into memory     char *fbase = mmap(NULL, statbuf.st_size, PROT_READ   PROT_WRITE, MAP_SHARED, fd, 0); </pre>	<pre>     int n = 0;     int findRes = 0;     while(findRes != -1) {         findRes = findFunctions(fbase,statbuf.st_size,findRes,0);         n++; //at end of data, the number of swappable               functions found is returned     }     printf("Number of swapable functions found: %d\n",n); </pre>
<pre>     prepare_random_num_generator();     int x;     int y;     int randNumber = 0; //a count of the number of functions swapped is     int swapCount = 0; kept and incremented if random number is 1     findRes = 0;     for(x=0;x&lt;(n-1);x++) {         for(y=0;y&lt;n-x-1;y++) {             randNumber = rand() % 2 + 1;             if(randNumber == 1){                 swapCount++;             }             findRes = findFunctions(fbase,statbuf.st_size,findRes,randNumber);         }     } </pre>	

**Figure 3:** High level overview of ROPMit implementation

## 6. Testing and results

The tested code consisted of binaries with stack buffer overflow vulnerabilities in them. The code base was big enough to have gadgets made out of them. The stack was marked NX (Non-eXecutable) and a ROP attack was constructed based upon the attack by Reece (2013). The results show that the effect of randomisation was able to thwart the attacks with very high likelihood proportional on average to the factorial of the number of functions.

## 7. Discussion and conclusions

ROPMit is a work-in-progress and is still under development as an exploratory research project. It is relatively simple in its implementation and currently has limited optimisation. Even though shown to be effective for any size of program under test conditions, the authors have identified potential weaknesses and limitations to this approach with research ongoing to overcoming these weaknesses and limitations. One of these is the ability to identify function boundaries; this is currently limited to one signature for the function prologue and one for the function epilogue. This detects the vast majority of functions successfully, but many programs with high code optimisation often have more difficult to detect function boundaries. These must be mapped and understood to see if they can work with ROPMit without much additional complexity. Another is the challenge of more exotic code calls, such as via a trampoline. ROPMit deals with the standard PLT & GOT as mentioned above, but doesn't yet deal with application-specific code calls via redirection. The presence and nature of these in target software must be mapped and understood to see if they can be handled by ROPMit. A third is on the scalability of ROPMit to larger target programs and whether these can be fully handled by optimisations to the implementation or the overarching approach.

ROPMit currently swaps adjacent functions with multiple passes over the whole program until sufficient randomisation has occurred. This can be optimised to less than  $O(n^2)$ , though the nature of the scaling with optimisations has not yet been studied. It may be found that compile-time generated or other supporting metadata is required to make ROPMit effective, though this would be counter to the simplicity of ROPMit.

## 8. Future work

In addition to address the points given in the previous section, further related research is planned. One particular interesting area is applying this randomisation technique at run-time which would help to solve the weakness of ASLR only being applied at load time and therefore overcome its limitation in very long lived processes. This requires more radical thinking and a number of challenges and obstacles stand in the way to run-time sub-section randomisation. Furthermore, it would be useful to study whether it would be effective to apply to object oriented programming techniques and particularly the binaries generated from a C++ compiler, including the variety of ways code can be executed including on dynamic objects with dynamic dispatch.

Another proposal that we are working on is a randomisation-based hybrid solution which aims to shuffle the memory layout at different instances, thus hindering ROP attacks. The main concept behind this idea is to randomise the memory layout at the level of function block and basic block granularity thus offering a high entropy which prevents brute-force and ROP attacks.

## References

- Bletsch, T., Jiang, X., & Freeh, V. (2011). *Mitigating code-reuse attacks with control-flow locking*. ACSAC '11 Proceedings of the 27th Annual Computer Security Applications Conference. Pages 353-362, doi:10.1145/2076732.2076783
- Carlini, N., Barresi, A., Payer, M., Wagner, D., & Gross, T. R. (2015). *Control-Flow Bending: On the Effectiveness of Control-Flow Integrity*. SEC'15 Proceedings of the 24th USENIX Conference on Security Symposium. Pages 161-176
- Dang, T. H. Y., Maniatis, P., Wagner, D. (2015). *The Performance Cost of Shadow Stacks and Stack Canaries*. Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security. doi:10.1145/2714576.2714635
- Davi, L., & Sadeghi, A. R. (2015). *Building Secure Defenses Against Code-Reuse Attacks*. Springer International Publishing.
- Davi, L., Sadeghi, A., & Winandy, M. (2011). *ROPdefender: A detection tool to defend against return-oriented programming attacks*. Published in AsiaCCS 2011. DOI:10.1145/1966913.1966920
- Gupta, A., Habibi, J., Kirkpatrick, M. S., & Bertino, E. (2015). *Marlin: Mitigating code reuse attacks using code randomization*. IEEE Transactions on Dependable and Secure Computing, 12(3), 326-337. doi:10.1109/TDSC.2014.2345384
- Hiser, J., Nguyen-Tuong, A., Co, M., Hall, M., & Davidson, J. W. (2012). *ILR: Where'd my gadgets go?* 2012 IEEE Symposium on Security and Privacy. doi:10.1109/SP.2012.39

- Intel. (June 2017). *Control-Flow Enforcement Technology Preview*. Retrieved from  
<https://software.intel.com/sites/default/files/managed/4d/2a/control-flow-enforcement-technology-preview.pdf>
- Koo, H., Chen, Y., Lu, L., Kemerlis, V. P., & Polychronakis, M. (2018). *Compiler-assisted Code Randomization*. 2018 IEEE Symposium on Security and Privacy (SP)
- Larsen, P., Homescu, A., Brunthaler, S., & Franz, M. (2014). *SoK: Automated software diversity*. SP '14 Proceedings of the 2014 IEEE Symposium on Security and Privacy. Pages 276-291
- Onarlioglu, K., Bilge, L., Lanzi, A., Balzarotti, D., & Kirda, E. (2010). *G-free: Defeating return-oriented programming through gadget-less binaries*. ACSAC 2010, Annual Computer Security Applications Conference.  
doi:10.1145/1920261.1920269
- Pappas, V., Polychronakis, M., & Keromytis, A. D. (2012). *Smashing the gadgets: Hindering return-oriented programming using in-place code randomization*. SP '12 Proceedings of the 2012 IEEE Symposium on Security and Privacy. Pages 601-615. doi:10.1109/SP.2012.41
- Reece, A. (2013). Introduction to Return Oriented Programming. Retrieved from  
<http://codearcana.com/posts/2013/05/28/introduction-to-return-oriented-programming-rop.html>
- Schwartz, E. J., Avgerinos, T., & Brumley, D. (2011). *Q: Exploit Hardening Made Easy*. Retrieved from  
[https://www.academia.edu/2842524/Q\\_Exploit\\_hardening\\_made\\_easy](https://www.academia.edu/2842524/Q_Exploit_hardening_made_easy)

# US-Russian Relations in Cybersecurity: The Constructivist Dimension

Ilona Stadnik

School of International Relations, Saint-Petersburg State University, Russia

[Ilona.st94@gmail.com](mailto:Ilona.st94@gmail.com)

**Abstract:** This case study of US-Russian cyber relations seeks to answer two critical questions. Firstly, can one country change the discourse of cybersecurity in another country even without committing malign activity? Secondly, whether shifts in the state of discourse will lead to changes in foreign policy toward cybernorms? The answers will help to fill gaps in the constructivist literature on cybersecurity, providing theoretical ground for the concept of cyberpower through adding to its materialist understanding. Also, this research will contribute to discourse studies explaining how a change in the conception of what is secure in cyberspace has led to cyber policy change. The case study is focused on the alleged Russian interference in the US presidential elections in 2016, the shift in domestic American cybersecurity discourse towards infosecurity, and US-Russian competition for international cybernorms.

**Keywords:** cybersecurity, infosecurity, cybernorms, constructivism, realism, securitization, US-Russian relations

---

## 1. Introduction

The lack of constructivist perspective on cyber conflicts between nations leaves a significant gap in modern academic scholarship. Constructivism and liberalism nominally have more to say about security in the digital age because of the diversity of actors and a wide range of topics including information society and networked economies (Ericksson and Giacomello, 2006). Yet, the realist paradigm has been dominating political literature on cybersecurity and warfare for the last two decades, by focusing on strategic studies and the military dimension.

In a realist perspective, a state develops its cybersecurity policy to achieve national interests. The underlying premises of cyber policy were thus transferred from classical and neorealist works about the struggle for power (Morgenthau, 1948) and balance of power (Waltz, 1979), or maximization of power to ensure survival in anarchy (Mearsheimer, 2001). Reardon and Choucri (2012) noted this transfer and identified the tendencies in academic and policy literature.

In the past decade, a variety of works have described the essence of cyberconflict with evolving narratives. Firstly, there were two competing views on the probability of a cyberwar – while Clarke and Knake (2012) claimed that cyberwar is a very real and pressing threat to national security, Rid (2013) argued that cyber war does not represent true violence in the Clausewitzian sense and is unlikely to be in the future. Segal (2016) seemed to support this view and emphasized that cyberattacks pose less of a threat of bodily harm but more to infrastructures such as financial institutions, power grids, and networks. Secondly, authors covered specific features of cyberconflict dynamics using terminology of strategic studies, to mention a few of them. Cyber offense was claimed by Libicki (2009) more cost-effective than cyber defense. Gratzke (2013) suggested the security dilemma for cyberspace has a reverse effect, - arguing that cyberweapons loose their capability after their usage because exploited vulnerabilities in adversarial networks are patched and secured. Then Gartzke and Lindsay (2014) put forward the idea of cross domain deterrence including cyber. Nye (2017) developed the cyber deterrence concept and highlighted its main components: punishment and denial (coercive options), and entanglement and norms (restraining options). Coercion in cyberspace was studied and theorized by Borghard and Lonergan (2017), Sharp (2017). Valeriano and Maness (2015) introduced the theory of cyber restraint saying that adversaries are unlikely to engage in cyber conflict because of normative restrictions, the ease of proliferation of cyber weapons, and other unknown risks.

The constructivist approach to cyber conflict has been undereappreciated. The Copenhagen school and securitization theory got its second birth with studies on new cyber threats and new referent objects by Dunn-Cavelty (2008, 2013) and Hansen and Nissenbaum (2009). Paletta et al. (2015) drew attention to the media coverage of an unseen cyber arms race. Craig and Valeriano (2016) demonstrated a relationship between build-up of cyber capabilities and mutual perceptions of threat and competition between states in a select number of cases. Also, they defined militarization of cyberspace by a particular discourse expressed in new military organizations, cyber-military doctrines, cybersecurity budgets.

However, constructivist view on developments in cyber policy doesn't confine itself to empirical and methodological issues of research as it may seem first. In contrast, it can explain developments in cybersecurity policy through changes in discourse, in other words, conceptions of what is secure mean in cyberspace. Recent history of US-Russian relations in cybersecurity provides a fresh set of facts for analysis and opens new perspectives for theoretical research. The power of discourse has influenced formulation of foreign policy. Since discourse emanates from domestic actors and processes it is necessary to track the evolution of cybersecurity discourse in both countries. Globally, there are two prevailing discourses for security in cyberspace: cybersecurity and infosecurity. The former deals predominantly with the technical dimension of network security while the latter incorporates issues of content regulation – how information affects national security and social order in addition to network security. As a result, there is a tacit distribution of countries belonging to the two global discourses: the more a country is authoritarian the stronger is information security discourse and vice versa. Such a split became entrenched during the work of the first UN group of governmental experts (UN GGE) in 2004 that examined the existing and potential threats from cyberspace and possible cooperative measures to address them. Group members, the US and Russia in particular, couldn't agree whether to address issues of information content, or only network infrastructure. Thereafter, UN GGE groups began to use the neutral wording of "ICT use" to facilitate consensus in their reports.

A critical change in US cybersecurity discourse subsequent to the alleged information operations and hacks associated with the 2016 presidential elections politically attributed to Russia serve as a key case study for this study. The second research question is whether this discourse change will lead to changes in American foreign policy towards international cybernorms to an infosecurity based approach.

## **2. US discourse and milestones for changes**

The understanding of cybersecurity in the US has gone through several stages of evolution. Whereas at the end of the 1990s attention was focused mainly on the internal security of networks, with an emphasis on the technical aspects of security, the spread of ICT provided for human and political dimension of cyber threats. The second half of the 2000s marked the emergence of an international dimension in cybersecurity policy, because cyberspace became truly global, and for the safe and cost-effective use of the Internet it was necessary to build an international system for ensuring cybersecurity. The discourse of cyber threats has also changed in key strategic American documents – a more diverse typology of threats, the growth of potentially dangerous actors and “naming and shaming” of adversarial countries for their malicious activity in cyberspace. Finally, the hacker attack on the Democratic National Committee preceding the presidential elections in the United States in 2016, as well as the scandal with Facebook and Cambridge Analytica, led to a change in the discourse of cyber threats. For the first time, the American political establishment talked about violation of US sovereignty through information propaganda in social networks. The leitmotif of the congressional hearings of NSA and FBI directors about DNS hack was the recognition of the fact that the cyberattack “violated American sovereignty” by the theft and publication of H. Clinton’s confidential correspondence that ultimately influenced “the most sacred act of democracy” - elections (McFaul, 2017).

“Hacked” elections became a milestone for changes in the US cybersecurity discourse. The new National Cyber Strategy signed by President D. Trump in 2018 cemented the change in discourse to infosecurity. In the introduction, the strategy lists, among other threats, cyber tools that adversaries use to “sow discord in our democratic process.” Moreover, the document has a separate section devoted to malign cyber influence and information operations. It claims that the US will “counter the flood of online malign influence and information campaigns and non-state propaganda and disinformation <...> and prevent the use of digital platforms for malign foreign influence operations while respecting civil rights and liberties”. Thus, infosecurity has been communicated in the discourse on the highest official level.

Interestingly, private sector followed the new strategy immediately. Facebook launched its “war room” to address the specific elections-related issues on its social media platforms. Personnel of the war room had fight with voter suppression efforts and civic-related misinformation during the US mid-term elections in the 2018 fall (Constine, 2018). The company is planning to continue this tactic and expand the work of war rooms around the globe.

### **3. What is cyberpower?**

Constructivism can provide new explanations for the question of power in cyberspace. Power is a key concept for political realism, and it already has several interpretations for the cyber domain from a realist materialistic perspective. Nye (2011) defined cyberpower as “the ability to obtain preferred outcomes through use of the electronically interconnected information resources of the cyber domain”. Valeriano and Maness (2015) measured cyber power as cyber capabilities – resources, manpower and money to support cyber operations. In other words, cyberpower is the sum of cyber offense, cyber defense, and cyber dependence. Segal (2016) based his criteria of cyberpower on the presence of: (1) large and technologically advanced economy to produce hardware, software, and services; (2) public institutions that channel innovations by the private sector; (3) adventurous and rapacious military and intelligence services; (4) an attractive story to tell about cyberspace and cybernorms.

Segal seemed to start conceptualizing cyberpower including not only the material but also using an idealist base. However, this is not enough for filling the gap in constructivist theory for cybersecurity. Cyber capacities are more difficult to be counted than nuclear warheads and missiles, yet some countries are still identified as cyberpowers. This means that the perception of a particular country comes from rumors and hard-proven intelligence about its offensive cyber capabilities, as well as from attacks and campaigns it had (allegedly) committed. The last part is a big puzzle for international cybersecurity, since there is no reliable attribution mechanism for cyber incidents, and international law needs to be developed to establish responsibility of states for acts of aggression committed in cyberspace. Thus, a new trend for political attribution of cyber incidents has emerged (Schulzke, 2018, Kaushik, 2018).

### **4. US-Russian fight for cybernorms**

The application of international law to cyberspace and its development for cyberconflict remains a cornerstone issue in US-Russian international relations. Back in 1998, Russia was the first to draw attention to the danger of conflicts with use of ICT for international peace and security and started to call for preventing military conflicts in cyberspace. As a result, while there were five UN GGE groups, only three were successful enough to produce consensus reports about the applicability of international law to cyberspace and to introduce norms for responsible state behavior. US-Russian collaboration was rather fruitful, until the allegations of cyberattacks and meddling in the 2016 American election. The failure of the last UN GGE in 2017 to reach the consensus on applicability of international humanitarian law, was partly due to political tensions between countries, but also because the format has exhausted itself. However, Russia claimed to continue this work by preparing a draft resolution for the 73<sup>rd</sup> UN session creating a new group based on an open-ended principle in contrast to previous 25 members selected by geographically equal representation. Interestingly, the US also introduced their own resolution seeking to continue the GGE format without any changes. Finally, the General Assembly voted and passed both resolutions by the end of 2018, so this opens up a competition between the two newly established groups of governmental experts on cybernorms. The voting records show the traditional international split between countries on the principle of adhering to the one of the dominating cybersecurity discourses.

### **5. Preliminary conclusions and further research**

Obviously, the trend for shifting cybersecurity discourse towards infosecurity in the US is backed by the release of new strategic documents by the new administration. However, the impact of perceived Russian influence on American discourse still needs to be proven. Without a doubt, meddling in the election process has triggered the changes, but the findings of the Mueller commission have yet to be released. Whether or not evidence of Russian interference is proven, its perception has already confirmed the hypothesis of the research. A country can change the discourse of another country by either conducting or being perceived to be conducting malign activity against it.

However, the second part of the research question still remains unanswered. Recent developments in the UN GGE process signals that the fight for cybernorms is continuing. While Russia is pushing an infosecurity agenda, the US tries to keep its cybersecurity policy on a separate track. Since the composition of working groups is as yet unknown, we have to watch whether there will be a joint collaboration between them and wait for the final reports due in 2020 and 2021.

Finally, there is an interesting game-changer for cybernorms. Private sector actors have started to promote their agenda for responsible behavior of states and non-state actors in cyberspace (Stadnik, 2018). The tech giants

have tried to consolidate the international community around basic cybersecurity principles while keeping the privacy and freedom of the Internet. However, as several global social media platforms were forced to impose additional restrictions in their content policies, they have moved closer to an infosecurity discourse themselves.

## References

- Borghard, E. D., Lonergan, S.W. (2017) «The Logic of Coercion in Cyberspace». *Security Studies* 26 (3): 452–81. <https://doi.org/10.1080/09636412.2017.1306396>.
- Clarke, R., Knake, R. (2012) *Cyber War: The Next Threat to National Security and What to Do About It*. Harper Collins.
- Constine, J. (2018) «Facebook Denies Report That Election War Room Was Disbanded». *TechCrunch* (blog). December 2018. <http://social.techcrunch.com/2018/11/26/facebook-war-room-rages-on/>.
- Craig, A., Valeriano, B. (2016) "Conceptualising Cyber Arms Races." IEEE Proceedings for CCDCOE CyberCon, 8th International Conference on Cyber Conflict: Cyber Power, 141–58.
- Dunn Cavelti, M. (2008) *Cyber-Security and Threat Politics: US Efforts to Secure the Information Age*. CSS Studies in Security and International Relations. London ; New York: Routledge.
- Dunn Cavelti, M. (2013) «From Cyber-Bombs to Political Fallout: Threat Representations with an Impact in the Cyber-Security Discourse». *International Studies Review* 15 (1): 105–22. <https://doi.org/10.1111/misr.12023>.
- Eriksson, J., Giacomello, G. (2006) «The Information Revolution, Security, and International Relations: (IR)Relevant Theory?» *International Political Science Review/ Revue Internationale de Science Politique* 27 (3): 221–44. <https://doi.org/10.1177/0192512106064462>.
- Gartzke, E. (2013) "The Myth of Cyberwar: Bringing War on the Internet Back Down to Earth". *International Security* 38(2): 41–73.
- Gartzke, E., Lindsay, J. (2014) «Cross-Domain Deterrence: Strategy in an Era of Complexity» [https://quote.ucsd.edu/deterrence/files/2014/12/EGLindsay\\_CDDOverview\\_20140715.pdf](https://quote.ucsd.edu/deterrence/files/2014/12/EGLindsay_CDDOverview_20140715.pdf)
- Hansen, L., Nissenbaum, H. (2009) «Digital Disaster, Cyber Security, and the Copenhagen School». *International Studies Quarterly* 53 (4): 1155–75. <https://doi.org/10.1111/j.1468-2478.2009.00572.x>.
- Kaushik, A. (2018) «Attribution in Cyberspace: Beyond the "Whodunnit"». GLOBSEC. May 2018. <https://www.globsec.org/wp-content/uploads/2018/05/GLOBSEC-cyber-attribution.pdf>.
- Libicki, M.C. (2009) *Cyberdeterrence and Cyberwar*. Rand Corporation.
- McFaul, M. (2017) «The real winner of the House Intelligence Committee hearing on Russia». *The Washington Post*. 23 March 2017. [https://www.washingtonpost.com/news/global-opinions/wp/2017/03/23/the-winner-of-the-house-intelligence-committee-hearing-on-russia-vladimir-putin/?utm\\_term=.fb3dfffa121d2](https://www.washingtonpost.com/news/global-opinions/wp/2017/03/23/the-winner-of-the-house-intelligence-committee-hearing-on-russia-vladimir-putin/?utm_term=.fb3dfffa121d2).
- Mearsheimer, J. J. (2001) *The tragedy of Great Power politics*. New York: Norton.
- Morgenthau, H. J. (1948) *Politics among Nations: The Struggle for Power and Peace*. New York: A.A. Knopf.
- Nye, J. (2011) «Cyber Power». *The Future of Power in the 21st Century*, Public Affairs Press.
- Nye, J. (2017) «Deterrence and Dissuasion in Cyberspace». *International Security* 41 (3): 44–71. [https://doi.org/10.1162/ISEC\\_a\\_00266](https://doi.org/10.1162/ISEC_a_00266).
- Paletta, D., Yadron, D., Valentino-Devries, J. (2015) «Cyberwar Ignites a New Arms Race». *The Wall Street Journal*, 11 October 2015.
- Reardon, R., Choucri, N. (2012) *The Role of Cyberspace in International Relations: A View of the Literature*. San Diego: MIT. [http://ecir.mit.edu/images/stories/Reardon%20and%20Choucri\\_ISA\\_2012.pdf](http://ecir.mit.edu/images/stories/Reardon%20and%20Choucri_ISA_2012.pdf).
- Rid, T. (2012) «Cyber War Will Not Take Place». *Journal of Strategic Studies* 35 (1): 5–32. <https://doi.org/10.1080/01402390.2011.608939>.
- Schulzke, M. (2018) «The Politics of Attributing Blame for Cyberattacks and the Costs of Uncertainty». *Perspectives on Politics* 16 (4): 954–68. <https://doi.org/10.1017/S153759271800110X>.
- Segal, A. (2016) *The Hacked World Order: How Nations Fight, Trade, Maneuver, and Manipulate in the Digital Age*. PublicAffairs.
- Sharp, T. (2017) «Theorizing cyber coercion: The 2014 North Korean operation against Sony». *Journal of Strategic Studies* 40 (7): 898–926. <https://doi.org/10.1080/01402390.2017.1307741>.
- Stadnik, I. (2018) «A New Cybersecurity Diplomacy: Are States Losing Ground in Norm-Making?» *Russian Council on International Affairs* (blog). <http://russiancouncil.ru/en/analytics-and-comments/analytics/a-new-cybersecurity-diplomacy-are-states-losing-ground-in-norm-making/>.
- Valeriano, B., и Maness, R. (2015) *Cyber War versus Cyber Realities: Cyber Conflict in the International System*. Oxford University Press.
- Waltz, K. N. (1979) *Theory of International Politics*. New York: Random House.