Cyber-Physical Threat Intelligence for Critical Infrastructures Security

A Guide to Integrated Cyber-Physical Protection of Modern Critical Infrastructures

John Soldatos, James Philpot
and Gabriele Giunta
(Editors)

now
the essence of knowledge

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

ISBN: 978-1-68083-686-8
E-ISBN: 978-1-68083-687-5
DOI: 10.1561/9781680836875

Suggested citation: John Soldatos, James Philpot and Gabriele Giunta (eds.). (2020). *Cyber-Physical Threat Intelligence for Critical Infrastructures Security*. Boston–Delft: Now Publishers

Table of Contents

Part III Securing Critical Infrastructures of the Energy Sector

Foreword

As outlined in the new Security Union Strategy, protection and resilience of critical infrastructures remains among the top priorities of the European Union.

European critical infrastructure sectors find themselves in the midst of rapid digitization that is accelerated by the growth of technologies like cyber-physical systems (CPS), the Internet of Things (IoT) and artificial intelligence (AI). Besides this, critical infrastructure operators today are confronted with different types of risks in both the cyber- and physical domain. Notable attacks like the "Wannacry" ransomware and the "Mirai" botnet cyber-attacks, which affected critical infrastructure operations across different Member States and in multiple sectors, remind us of the risks that we face. This situation calls for innovative security concepts that take us beyond conventional policies that have been addressing either the physical or cyber-security domain. Currently, the ongoing COVID-19 pandemic shows us the vital role that digital critical infrastructure play in keeping different sectors like telecommunications, finance, energy, and health care running in the time of crisis.

The European Commission (EC) is supporting the Member States to protect and ensure the resilience of critical infrastructures. It has adopted an integrated framework based on both strong physical and cyber-security measures. Key pillars of this framework include the Directive on security of network and information systems NIS Directive), the Directive on protecting European Critical Infrastructures, the General Data Protection Regulation (GDPR), and the Cybersecurity Act Regulation. Furthermore, under the Commission's new digital strategy, additional actions are being considered. The EC is emphasizing the consistency and complementarity of these and other ongoing initiatives, including the revision of the EUs overall approach to critical infrastructure protection and resilience, notably the European Programme for Critical Infrastructure Protection (EPCIP). A package of measures will be put forward during the fall of 2020.

Besides policy development, the EC supports research and innovation projects under Horizon 2020 looking for innovative approaches to the protection and resilience in different sectors. Calls for proposals have been developed and financed jointly between the ECs Directorates-General for Migration and Home Affairs (HOME) and Communications Networks, Content and Technology (CNECT). This cooperation reflects our commitment to an integrated cyber-physical approach, reflected also in the projects' outcomes.

It is important to underline that research plays a vital strategic role for security policy in the EU. In this respect, the EC has encouraged and supported the clustering between projects, as a means of boosting their cooperation. As such, we welcome the creation of the European Cluster for Securing Critical infrastructures (ECSCI), which seeks to bring the many projects working to improve critical infrastructure protection and resilience together.

Based on results that have been achieved in EU-funded projects, this book describes innovative approaches to enhancing the protection of critical infrastructures. It also presents approaches that reduce fragmentation in security operations and improve the implementation of existing European regulation. The book provides insights of relevance to policy makers, researchers and practitioners who are working to ensure the functioning of digitally-enabled critical infrastructures that our societies rely on.

Brussels, Belgium, June 2020

Monique Pariat
Director General
Migration and Home Affairs
European Commission

Roberto Viola
Director General
Communications Networks, Content and Technology
European Commission

Preface

At the dawn of the fourth industrial revolution, governments and enterprises are increasingly deploying Cyber-physical Systems (CPS) as part of their critical infrastructures. CPS systems blur the boundaries between the physical and digital worlds and enable digital control of physical processes in sectors like healthcare, finance, energy, and industry. CPS systems are a core element of the popular Internet of Things (IoT) paradigm, which has a transformational impact on the critical infrastructures that support the functioning of our societies and economies. Based on CPS and IoT systems, critical infrastructures operators leverage large amounts of field data in order to optimize business processes and decisions associated with the operation of their infrastructures. Furthermore, the rapid digitalization of critical infrastructures facilitates their interconnection and the seamless exchange of information across different stakeholders and value chains.

Along with these benefits, the expanded deployment of CPS and IoT technologies within critical infrastructures introduces various cybersecurity challenges, which add up to conventional physical security issues. This is evident in some of the recent large-scale security incidents against critical infrastructures, which include attacks against both cyber and physical assets. In several cases, adversaries exploit vulnerabilities in the digital parts of the infrastructures in order to attack their physical parts and vice versa. Therefore, critical infrastructures security must be implemented based on a holistic, integrated approach that protects cyber and physical assets at the same time. This is increasingly acknowledged by critical infrastructures operators and supported by recent regulatory efforts as well. As a prominent example, in Europe, the NIS Directive (EU 2016/1148) underlines the importance of cybersecurity for critical market operators and instructs EU Member states to supervise cybersecurity for the critical infrastructures of key sectors (e.g., telecommunications, finance, energy, healthcare, transport) in their country.

Overall, critical infrastructures security is currently redefined in order to address cyber and physical aspects in an integrated way. Cyber and physical security functions are no longer "siloed," but rather combined in the scope of integrated security policies. This integration introduces new requirements such as the need to model security knowledge in an unified way, the need to address cascading effects between the two different types of attacks (i.e. cyber and physical attacks), as well as the need to integrate solutions for cybersecurity and physical security within commonly used security platforms. Likewise, integrated platforms for critical infrastructures security must provide functionalities for preventing, detecting, and responding to security incidents in a proactive and cost-effective manner. Moreover, they should provide the means for sharing information across security stakeholders [e.g. Security Teams, First Responders, Computer Emergency Response Teams (CERT)] to facilitate their effective collaboration.

The above-listed requirements are common to the critical infrastructures of the different sectors of the economy (e.g., finance, healthcare, energy, and transport). Nevertheless, there are also sector-specific security requirements, stemming from the different devices, control processes, and business operations of the various sectors. Moreover, installations in different sectors are interconnected in different ways and are subject to diverse sets of cascading effects.

In this context, the present book is dedicated to presenting novel solutions for integrated security of critical infrastructures, with emphasis on solutions in four sectors, namely finance, healthcare, energy, and communications. The book presents various technologies and building blocks for integrated security, along with sector-specific solutions for each one of the above four sectors. The presented security technologies cover a wide range of functionalities such as security knowledge modeling, risk assessment, information sharing for stakeholders' collaboration, Security Information and Event Management (SIEM), CPS systems and IoT devices security, regulatory compliance auditing solutions, and more.

The book is structured in five parts. The first four parts are dedicated to presenting solutions for each one of the above listed four sectors, i.e., finance, healthcare, energy, and communications, respectively. The fifth part comprises sector agnostic solutions including technologies and best practices that are applicable to critical infrastructures regardless of their sector. Specifically:

The first part of the book is titled: "**Securing Critical Infrastructures of the Financial Sector**" and consists of the following chapters:

- Chapter 1 "**Security Challenges for the Critical Infrastructures of the Financial Sector**" introduces the main challenges that are associated with physical security and cybersecurity for the critical infrastructures of the financial sector. The chapter presents recent security incidents against financial institutions as a main motivation behind integrated security. Moreover, it also outlines the main building blocks of integrated security solutions for the financial sector.

- Chapter 2 "**A Reference Architecture for Securing Infrastructures in the Finance Sector**" presents a Reference Architecture (RA) for integrated (i.e., cyber and physical) security in the financial sector. The chapter illustrates the main building blocks of the RA, as well as their technical specifications and the interfaces between them. Furthermore, it illustrates how the RA aligns to recommendations of security agencies and standards development organizations.
- Chapter 3 "**FINSTIX: A Security Knowledge Base for the Finance Sector**" introduces the importance of security knowledge modeling for securing critical infrastructures. It also presents relevant knowledge modeling standards and vulnerability knowledge bases. Accordingly, the chapter details the FINSTIX model, which has been introduced in the scope of the European Commission (EC) funded FINSEC project towards modeling security knowledge in the finance sector. The chapter presents the use of FINSTIX in different scenarios and use cases.
- Chapter 4 "**Artificial Intelligence Gateway for Cyber-physical Security in Critical Infrastructure and Finance**" presents an intelligent system for Cyber and Physical Security of critical infrastructures. The system collects and processes information from different security systems installed within a site or a perimeter (e.g., intrusion detection, access control, fire detection, technical alarms, etc.), including data and events from a video surveillance system. Accordingly, it processes this information towards identifying threats and events before they become critical. The module can, therefore, analyze security events and situations, allow the integration of cybersecurity concepts, and support the monitoring of risks associated with critical installations.
- Chapter 5 "**Information Sharing and Stakeholders' Collaboration for Stronger Security in Financial Sector Supply Chains: A Blockchain Approach**" describes a novel blockchain-based solution for sharing security data in the scope of the financial sector supply chains. It explains the rationale behind adopting and using a decentralized blockchain-based solution for information sharing, including its merits and advantages over conventional centralized solutions. A relevant prototype implementation based on a permissioned blockchain is also presented.

- Chapter 6 “**Automated Assistance to the Security Assessment of APIs for Financial Services**” introduces the need for security assessment and automated synthesis of mitigation measures towards securing Open Baking services in the context of the new Payment Services Directive (PSD2). Accordingly, it provides an overview of innovative approaches to addressing these challenges, based on: (i)Automated identification and mitigation of security misconfigurations underlying Transport Layer Security (TLS) sessions; and (ii) Automated pen-testing and synthesis of mitigations for both business (e.g. payments) and security (e.g. authentication or authorization) functionalities provided by the APIs.
- Chapter 7 “**Adaptive and Intelligent Data Collection for Security of Critical Financial Infrastructures and Services**” describes an adaptive and intelligent data collection and monitoring system. The system provides intelligent, resilient, automated, efficient, secure, and timely collection and analysis of security-related data towards protecting cyber-physical infrastructures and services of the financial sector. It also enables security teams to gain insights on how: (i) the nature and quality of collected data affects the efficiency and accuracy of methods of attack detection and defense; (ii) the detection capability can be improved by correlating wide-ranging data sources and predictive analytics; (iii) the rate of the data collection at the various monitoring probes is tuned by managing the appropriate levels and types of intelligence and adaptability of security monitoring; (iv) the optimization of bandwidth and storage of security information can be achieved by rendering adaptiveness and intelligence and by integrating adaptive strategies and rules, and (v) increased automation can be achieved through a feedback loop of collection, detection, and prevention.

The second part of the book is titled: “**Securing Critical Infrastructures of the Health Sector**” and consists of the following chapters:

- Chapter 8 “**Security Challenges for the Critical Infrastructures of the Healthcare Sector**” introduces the main challenges that are associated with physical and cyber security for the critical infrastructures of the healthcare sector. It presents recent security incidents against healthcare institutions, such as the WannaCry ransomware. It also outlines the rationale for new integrated security system in the healthcare sector, including systems that support risk assessments and various threat scenarios.
- Chapter 9 “**Security Systems in the Healthcare Sector**” presents how a video management system, an access control system, a fire detection system,

SCADA, ICS, and smart building sensors, as well as a cybersecurity protection system can work together to increase the security and resilience of healthcare systems and processes.

- Chapter 10 "**Integrated Cyber-physical Security Approach for Healthcare Sector**" explains how an integrated cyber-physical security system will be developed and why this is the optimal solution for the future security of healthcare systems. It also describes the architecture of the security platform of the EC-funded SAFECARE project, along with its main building blocks and the interfaces between them.
- Chapter 11 "**Vulnerability and Incident Propagation in Cyber-physical Systems**" provides a focused view on how to handle incidents and their propagation from an asset's point of view in a healthcare environment. It also presents an overview of relevant work conducted within the SAFECARE project.
- Chapter 12 "**Innovative Toolkit to Assess and Mitigate Cyber Threats in the Healthcare Sector**" presents a novel Cyber Security Toolkit for healthcare systems. Specifically, the toolkit provides the IT personnel at hospitals and care centers with highly comprehensive visual analytics that depict the cybersecurity situation on a near real-time basis, in an intuitive and user-friendly way. One of the merits of the presented Toolkit is that it is fully adaptable to the individual user's profile characteristics.

The third part of the book is titled: "**Securing Critical Infrastructures of the Energy Sector**" and includes the following chapters:

- Chapter 13 "**Security Challenges for the Critical Infrastructures of the Energy Sector**" discusses the major security challenges, along with the methods adopted to identify and describe new threats against Critical Energy infrastructures (CEI). It also outlines methods for evaluating and assessing risks associated with these threats.
- Chapter 14 "**Securing CEI By-Design**" focuses on the CEI Security "by-design" solutions implemented in the EC-funded DEFENDER project. These solutions include Double Virtualization (Resilience), Fault-Location and Restoration (Self-healing), as well as Cryptography and Blockchain (Data Protection) solutions.
- Chapter 15 "**Securing CEI By-Innovation**" describes the CEI Security "by-innovation" solutions developed in the DEFENDER project, including CEI Incidents Detection & Mitigation and the CEI Security Control Centre solutions.

The fourth part of the book is titled: "**Securing Critical Infrastructures of the Communications Sector**" and consists of the following chapters:

- Chapter 16 "**Security and Resilience Challenges for the Critical Infrastructures of the Communications Sector**" introduces the main challenges for the critical infrastructures in the communication sector. Specifically, it reviews the current threats that arise as a result of the inter-connection of cyber and physical systems. It also presents security strategies that exploit the features and dual (i.e., cyber and physical) nature of the critical infrastructures of the communication sector.
- Chapter 17 "**Resilience Enhancement and Risk Control Platform for Communication Infrastructure Operators**" introduces the platform of the EC-funded RESISTO project, as an innovative solution for communication Critical Infrastructures (CIs) situation awareness and enhanced resilience. The platform integrates two control loops both running on top of the Communication Infrastructure and interlinked with each other, namely a Long Term Control Loop (LTCL) and a Short Term Control Loop (STCL), which are both presented in the chapter. Moreover, the chapter illustrates the cyber/-physical threat detectors of the project, which take advantage of state-of-the art technologies such as Machine Learning, IoT security, airborne threat detection, and audio-video analytics.
- Chapter 18 "**Managed Security on 5G Communication Networks: The Software Defined Security Paradigm**" is devoted to the description of the interaction between the new communication system (i.e., the 5G framework) and the emerging security paradigm, known as Software Defined Security (SDS). SDS is considered as a new security model that is applied for the management of communication networks, in which security aspects are implemented, controlled, and managed at software level. The main objective is the decoupling of the control part and the operational part of a security system by exploiting the virtualization of security techniques. In addition to discussing the SDS paradigm, the chapter illustrates its application to IoT and 5G networks.

The fifth part of the book is titled: "**Sector Agnostic Issues in Critical Infrastructures Protection**" and consists of the following chapters:

- Chapter 19 "**Detection of Innovative Low-rate Denial of Service Attacks Against Critical Infrastructures**" analyzes the functioning of low-rate Denial of Service (DoS) attacks targeting network services, with emphasis on web protocols. It presents novel intrusion detection methods to identify last-generation threats that make use of low bandwidth to target network services.

- Chapter 20 "**Resilience Analysis and Quantification for Critical Infrastructures**" presents resilience analysis methods developed and performed in the scope of the EC-funded RESISTO project, notably methods that follow an enhanced risk and resilience management process based on the ISO-31000 standard. The chapter discusses the main inputs needed for the resilience quantification, which are gathered at separate steps of the management process. It also provides details about how the resilience quantification is performed, including information about the network simulation tool used.
- Chapter 21 "**CISIApro Critical Infrastructures Modeling Technique for an Effective Decision-making Support"** presents CISIApro 2.0, an agent-based simulator, that assesses the consequences of negative events on interconnected infrastructures. The output of CISIApro 2.0 is the set of possible devices and services that are affected by an event. This output is exploited by Decision Support Systems that mimic the operator decision processes.
- Chapter 22 "**Modern Innovative Detectors of Physical Threats for Critical Infrastructures**" starts with an overview of the current situation in Critical Infrastructures in terms of detecting physical threats, attacks, and hazards. Accordingly, it introduces novel threat detection techniques that cover a wider range of threats.
- Chapter 23 "**The Ethical Aspects of Critical Infrastructure Protection**" introduces the ethical challenges of critical infrastructure protection, including ethical issues associated with both cyber and physical security. It also outlines applicable rules and regulations, along with tools and techniques for ensuring compliance to them.

Most of the sector-specific solutions that are described in the book have been developed and validated in the scope of four EC co-funded projects on Critical Infrastructure Protection (CIP), which focus on the four sectors addressed in the book. These projects include:

- The H2020 FINSEC project, which focuses on integrated security solutions for the infrastructures of the financial sector, such as infrastructures that support ATM networks, payment networks, trading/investment, and wealth management infrastructures and more.
- The H2020 SAFECARE project, which provides solutions that improve physical and cybersecurity of healthcare infrastructures in a seamless and cost-effective way. It integrates advanced technologies from the physical and

cybersecurity spheres, in order to deliver high-quality, innovative, and cost-effective security solutions in healthcare settings.

- The H2020 DEFENDER project, which provides integrated (i.e., cyber and physical) security solutions for Critical Energy Infrastructures (CEI). It addresses distributed, complex, and large-scale Cyber-physical Systems that need proactive protection and fast restoration to mitigate physical or cyber incidents or attacks. Moreover, it provides the means to address the even more challenging combined cyber-physical attacks against CEIs.
- The H2020 RESISTO project, which develops innovative solutions for Communication Critical Infrastructures, including solutions for holistic situation awareness and enhanced resilience. The project is developed based on an Integrated Risk and Resilience analysis management and improvement process availing all resilience cycle phases (i.e., prepare, prevent, detect, absorb, etc.) and technical resilience capabilities (i.e., sense, model, infer, act, adopt).

Also, one of the chapters of the book has been contributed by the H2020 SPHINX project, which focuses on cybersecurity protection for Healthcare IT infrastructures. Specifically, it provides solutions for cyber protection, data privacy, and integrity, while proactively assessing and mitigating cybersecurity threats. It also evaluates and verifies new medical devices and services, while offering certification and near real-time vulnerability assessment services in the Healthcare IT ecosystem.

The target audience of the book includes:

- Researchers in the area of security and, more specifically, in cyber and/or physical security for critical infrastructures protection, who wish to be updated about latest and emerging security solutions.
- Critical Infrastructures owners and operators, with an interest in adopting, deploying, and fully leveraging next-generation security solutions for their infrastructures, including solutions for securing Cyber-physical systems and processes.
- Practitioners and providers of security solutions, which are interested in the implementation of use cases for the protection of their cyber and/or physical assets.
- Managers wishing to understand modern, integrated security approaches for industrial systems and critical infrastructures in their sectors, with emphasis on the finance, healthcare, energy, and communication sectors.

The book is made available as an Open Access publication, which could make it broadly and freely available to the Critical Infrastructure Protection and Security communities. We would like to thank now Publishers for the opportunity and their collaboration in making this happen. Most importantly, we take the chance to thank the contributing projects for their valuable inputs and contributions. Special thanks to Prof. Federica Battisti (Assistant professor at Università degli Studi Roma) for her supporting in the collection and editing of the chapters contributed to the book by the RESISTO project. Finally, we would also like to acknowledge funding and support from the European Commission as part of the H2020 FINSEC, SAFECARE, DEFENDER, RESISTO, and SPHINX projects, which made this Open Access publication possible.

March 2020
John Soldatos
James Philpot
Gabriele Giunta

Glossary

Symbols

5G - *Fifth generation of mobile phone systems*. 310, 319, 320

A

AAA - *Authentication, Authorization, and Accounting*. 21, 22, 25, 83, 84

AAC - *Audio Analytics Component*. 322, 406

AHP - *Analytical Hierarchical Process*. 265, 266, 268

AI - *Artificial Intelligence*. xv, 2, 9, 10, 12, 15, 23, 27, 32, 112, 228, 415–418, 420–425, 427

AIS - *Account Information Service*. 95, 96

AISP - *Account Information Service Providers*. 95

AIT - *Athens Information Technology*. 434

AMI - *Advanced Metering Infrastructure*. 251

ANSI - *American National Standards Institute*. 195

AP - *Access Points*. 322

API - *Application Programming Interfaces*. xx, 5, 6, 11, 20–26, 83, 84, 94, 95, 98–101, 114, 120, 121, 126, 133, 134, 330, 332, 333

App - *Mobile Application*. 218

APT - *Advanced Persistent Threats*. 241, 300, 304, 305

ARM - *Advanced RISC Machine*. 255

ASA - *Adaptive Security Analysis*. 134

ASM - *Application Service Management*. 239

ASPSP - *Account Servicing Payment Service Providers*. 95, 96

ATM - *Automated Teller Machine*. 42, 61–63, 65–67, 71, 72, 419

ATT&CK - *Adversarial Tactics, Techniques and Common Knowledge*. 315

B

B2B - *Business to Business*. 394

BFP - . 239

BMS - *Building Management Systems*. 182, 183, 186

BMSTDS - *BM Threat Detection System*. 186

BPM - *Business Process Model*. 319

BPMN - *Business Process Model and Notation*. 320

BSS - *Business Support Systems*. 331

BSSID - *Basic Service Set Identifiers*. 411

BTMS - *Building Threat Monitoring System*. 183–185

BYOD - *Bring Your Own Devices*. 216, 218, 219

C

CA - *Certification Authority*. 97

CaESAR - *Cascading Effects Simulation in urban Areas to assess and increase Resilience*. 365, 366, 378–380

CAGR - *Compound Annual Growth Rate*. 3, 4

CAPEC - *Common Attack Pattern Enumeration and Classification*. 9, 35, 67

CART - *Classification and Regression Tree*. 127

CCTV - *Closed Circuit Television*. 4, 8, 10, 117, 399, 400, 402, 406, 419

CEI - *Critical Energy Infrastructures*. xxi, xxiv, 228–232, 235, 245, 251, 252, 270, 271, 275–278, 280, 286–291

CEP - *Complex Event Processing*. 317

CERT - *Computer Emergency Response Team*. 6, 10, 43

CI - *Critical Infrastructures*. xxii, 179, 194–196, 297, 305, 310–312, 314, 316, 319, 325, 374, 378, 398, 399, 402–405, 407, 409–411, 413

CIIP - *Critical Information Infrastructure Protection*. 147–149

CIM - *Computer-integrated Manufacturing*. 195

CIP - *Critical Infrastructure Protection*. xxiii, 147, 149, 179, 180

CIRCL - *Computer Incident Response Center Luxemborg*. 323

CISIA - *Critical Infrastructure Simulation by Interdependent Agents*. 314, 319

CISIApro 2.0 - *Critical Infrastructure Simulation by Interdependent Agents pro 2.0*. 385, 387, 390–395

CJEU - *Court of Justice of the European Union*. 425

CN - *Communication Networks*. 329, 330

CNECT - *Communications Networks, Content and Technology*. xvi

CNN - *Convolutional Neural Network*. 73, 171, 172

CPE - *Common Platform Enumeration*. 9

CPS - *Cyber-Physical System*. xv, xvii, xviii, 54–63, 65, 73, 251, 296, 297, 299, 303, 304, 306, 415–418, 420, 421, 423, 425, 427

CPTI - *Cyber and Physical Intelligence*. 9, 26, 27, 29, 35, 42, 43, 49, 68

CPU - *Central Processing Unit*. 111, 117, 118, 128, 135, 337

CRUD - *Create-Read-Update-Delete*. 22, 25

CSIRT - *Computer Security Incident Response Teams*. 6, 43, 189

CSRF - *Cross-Site Request Forger*. 100

CT - *Computed Tomography*. 153

CTI - *Cyber Threat Intelligence*. 9, 36–38, 40, 41, 45, 67, 68

CTMS - *Cyber Threat Monitoring System*. 183, 186

CVE - *Common Vulnerabilities and Exposures*. 9, 35, 45, 46, 323

CVSS - *Common Vulnerability Scoring System collection*. 78

CWE - *Common Weakness Enumeration*. 9

D

DB - *Database*. 22–24

DC - *Data Collectors*. 26

DDoS - *Distributed Denial of Service*. 334, 336–339, 341, 375–377, 394

DER - . 267

DevOps - *Development and Operations*. 7, 12

DevSecOps - . 7, 8

DHCP - *Dynamic Host Configuration Protocol*. 333

DHS - *Department of Homeland Security*. 386

DL - *Deep Learning*. 106, 112

DLT - *Distributed Ledger Technology*. 417

DM - *Decision Maker*. 281–285

DMTF - *Distributed Management Task Force)*. 195

DNN - *Deep Neural Network*. 73

DNS - *Domain Name System*. 333, 350

DoS - *Denial of Service*. 122–124, 128, 135, 136, 302, 346–357, 359, 361

DPIA - *Data Protection Impact Assessment*. 426

DRUNK - *Dynamic Recon of UNKnown networks*. 323

DSO - *Distribution System Operator*. 239, 241, 279

DSS - *Decision Support System*. 290

DST - *Dempster-Shafer Theory*. 63

DV - *Double Virtulization*. 248, 250–260, 270, 271

DVA&M - *DV Administration & Management*. 252–260

DXL - *Data eXchange Layer*. 183, 187

E

EA - *Enterprise Architecture*. 195

EBIOS - *Expression des Besoins et Identification des Objectifs de Sécurité*. 142, 154, 155, 162

EC - *European Commission*. xv, xix, xxi–xxiii, 221, 435

ECI - *European Critical Infrastructures*. 147, 149, 227

ECSCI - *European Cluster for Securing Critical Infrastructures*. xvi

EDXL - *Emergency Data Exchange Language*. 187

EGE - *European Group on Ethics*. 425

eHealth - *Electronic Health*. 206, 209, 215–217, 221

EHR - *Electronic Health Record*. 219, 221

EI - *Economic Impact*. 377

eIDAS - *Electronic Identification, Authentication and Trust Services*. 95–97

EIT - *European Institute of Technology*. 101

ELES - . 239

EMR - *Electronic Medical Record*. 145

EMS - *Element Management System*. 331

ENISA - *European Union Agency for Network and Information Security*. 6, 14, 17–19, 30, 95, 194, 215, 301, 305

EOS - *European Organization for Security*. 435

EPCIP - *European Programme for Critical Infrastructure Protection*. xv

EPL - *Event Processing Language*. 318

ERN - *European Reference Networks*. 221

ETSI - *European Telecommunications Standards Institute*. 306, 331–333, 336

EU - *European Union*. xv, xvii, 6, 16, 30, 45, 210, 215, 216, 219–221, 326, 417, 418, 420–424, 426

EUTRAN - . 334

EWC - *Emergency Warning Communication*. 320

EWCF - *Emergency Warning Communication Function*. 319, 320

F

FAPI - *Financial-grade API*. 99

FDA - *Fault Detection Algorithm*. 259

FDO - *FINSTIX Domain Objects*. 41, 42, 45, 46

FI - *Financial Institutions*. 18

FINSEC - *FINancial SECurity*. 53, 54, 67, 68, 74, 104–106, 113–122, 124–126, 130, 132–134, 136, 137, 361

FINSTIX - *FINancial Structured Threat Information eXpression*. 119

FO - *Fiber Optic*. 374

FQ - *Frequency of Occurrence*. 377

FRAT - *Flexible Resilience Analysis Template*. 371, 376–378, 380

FS-ISAC - *Financial Services Information Sharing and Analysis Center*. 5, 11, 77

FTP - *File Transfer Protocol*. 350

G

GAN - *Generative Adversarial Networks*. 112

GCS - *ground control station*. 407

GDPR - *General Data Protection Regulation*. xv, 6, 11, 189, 190, 219–221, 425

GEN - . 241

GP - *General Practitioner*. 180

GPRS - *General Packet Radio Service*. 334

GPS - *Global Positioning System*. 298, 320, 407, 419

GPU - *Graphics Processing Unit*. 322

GSM - *Global System for Mobile*. 334

GSMA - *Global System Mobile Association*. 336

GUI - *Graphical User Interface*. 21, 79

H

H2020 - *Horizon 2020*. 314, 342, 361

HAMS - *Hospital Availability Management System*. 183, 187

HID - *Human Interface Devices*. 117

HIDS - *Host Intrusion Detection Systems*. 119

HILP - *High Impact but occur with Low Probability*. 261, 262, 269

HIPAA - *Health Insurance Portability and Accountability Act*. 336

HIS - *Hospital Information System*. 143, 153

HITL - *Human In The Loop*. 287, 288, 290, 291

HL7 - *Health Level Seven*. 216

HMI - *Human Machine Interface*. 24, 172, 173

HSS - . 333

HTTP - *Hyper Text Transfer protocol*. 78, 85, 255, 350, 353, 354, 356, 358–361

HTTPS - *Hypertext Transfer Protocol Secure*. 97, 255, 260

HVAC - *Heating, Ventilation and Air Conditioning*. 177

HW - *HardWare*. 315, 316

I

I2SP - *Incident Information Sharing Platform*. 230, 238

IaaS - *Infrastructure-as-a-Service*. 207

ICCS - *Institute of Communication and Computer Systems*. 408

ICMP - . 348

ICS - *Industrial Control System*. xxi, 167, 172

ICT - *Information and Communication Technology*. 2, 28, 29, 208, 209, 234, 252, 270, 416, 421

IDE - *Integrated Development Environment*. 127

IDF - *Incident Detection Framework*. 276, 279, 280, 285

IDS - *Intrusion Detection Systems*. 4, 111, 122, 133, 278, 352–354, 359, 387

IDSA - . 420

IEC - *International Electrotechnical Commission*. 208, 236–238, 336

IED - *Intelligent Electronic Device*. 172

IEEE - *Institute of Electrical and Electronics Engineers*. 336

IIM - *Input-Output Inoperability Model*. 388

IIoT - *Industrial Internet of Things*. 386

IISF - *Industrial Internet Security Framework*. 8

IM - *Improvement Measures*. 371, 376, 377, 381

IMF - *International Monetary Fund*. 3, 276, 287, 289–291

IMS - *Integrated Management System*. 36, 335

IMSI - *International Mobile Subscriber Identity*. 322, 411, 412

IoC - *Indicators of Compromise*. 37, 122, 134

IoMT - *Internet of Medical Things*. 217, 219

IoS - *Internet of Services*. 55

IoT - *Internet of Things*. xv, xvii, xviii, xxii, 15, 31, 32, 54, 55, 172, 173, 207, 209, 210, 217, 219, 223, 255, 296, 311, 320, 322, 335, 410, 415–418, 420, 421, 423, 425, 427, 429, 434

IP - *Internet Protocol camera*. 167

IP - *Internet Protocol*. 110, 111, 128, 134, 216, 333, 335

IP LAN - . 403

IPDSM - *Impact Propagation and Decision Support Model*. 183, 187

IPFS - *InterPlanetary File System*. 288, 290

IPM - *Impact Propagation Management*. 197, 199–201

IPMS - *IP Multimedia Subsystem*. 335

IPZS - *Istituto Poligrafico e Zecca dello Stato*. 101

IR - *Infrared*. 402, 407, 409

IS3 - *Intelligent Systems and Social Software*. 435

ISAC - *Information Sharing and Analysis Center*. 77

ISG - *Industry Specification Group*. 336

ISO - *International Organization for Standardization*. 195, 208, 236–238, 312, 347, 355

ISS - *Information Systems Security*. 154

IT - *Information Technology*. 63, 143, 144, 146, 150, 151, 154, 182, 183, 186, 207, 210–215, 217–219, 221

ITDS - *IT Threat Detection System*. 186

ITU - *International Telecommunication Union*. 298

IVD - . 145

IVR - *In-vitro Regulation*. 145

IVRE - *Instrument de Veille sur Réseaux Extérieurs*. 323

J

JIT - *Joint Investigation Team*. 45

JSON - *JavaScript Object Notation*. 9, 26, 68, 95, 100

K

KB - *Knowledge Base*. 80, 312, 325

KNN - *K-nearest neighbors*. 107, 114, 127–131

KPI - *Key Performance Indicators*. 280, 339

L

LDoS - *Low-rate DoS*. 350

LEA - *Law Enforcement Agencies*. 288

LIS - *Laboratory information system*. 143, 153

LRDoS - . 350

LTCL - *Long-term Control Loop*. xxii, 312, 314, 315, 325, 370

LTE - . 329, 335, 407, 411, 412

M

MANO - *Management & Orchestration*. 331, 332, 341

MAS - *Mobile Alerting System*. 183, 185, 187

MCDA - *Multi-criteria Decision Aid*. 281–283, 292

MCDM - *Multi-Criteria Decision Making*. 262, 264, 265, 267–269, 271

MDD - *Medical Device Directive*. 145

MDR - *Medical Devices Regulation*. 145

mHealth - *Mobile Health*. 206, 209, 217–219, 221

MHR - *Mixed Holistic Reductionist*. 387–391

MIP - *Mixed Integer Programming*. 261

MITRE - . 35, 37, 45

ML - *Machine Learning*. 9, 10, 109, 115, 317, 318

MLE - *Machine Learning Engine*. 121

MLP - *Multilayer Perceptron*. 127, 129, 130

MME - . 333

MOEA - *Multi-Objective Evolutionary Algorithm*. 261

MPLS - *Multi-Protocol Label Switching*. 333, 334, 394

MQ - . 84

MQTT - *Message Queuing Telemetry Transport*. 135, 136

MS - *Member State*. 189, 216, 221

MSA - *Microservices Architecture*. 23

N

NESCOR - *National Electric Sector Cybersecurity Organization Resource*. 234

NFC - *Near Field Communication*. 419

NFV - *Network Function Virtualisation*. 320, 329, 331, 332, 335–338

NFVI - *Network Function Virtualisation Infrastructure*. 331, 332

NFVO - *Network Function Virtualisation Orchestrator*. 331, 332

NG RAN - *NG Radio Access Network*. 334

NGO - . 435

NGSA-II - *Non-dominated Sorting Genetic Algorithm II*. 342

NHS - *National Health Service*. 149, 180

NIC - *Network Interface Controllers*. 118

NIDS - *Network Intrusion Detection Systems*. 119

NIS - *Network Information Systems*. xvii, 6, 16, 18, 30, 147–149, 188–190, 240

NIST - *National Institute of Standards and Technology*. 208, 212, 232, 311, 336

NOC - *Network Operation Centre*. 403

NS - *Network Services*. 329–331

NTF - *New Technologies Formation*. 423

O

O&M - *Operation and Maintenance*. 319

OAuth - *Open Authorization Protocol*. 96, 98–100

OES - *Operators of Essential Services*. 148, 188, 189

OIDC - *OpenID Connect*. 99, 100

OLT - *Optical Line Termination*. 394

ONAP - *Open Network Automation Platform*. 336

OODA - *Observe-Orient-Decide-Act*. 253, 254

OSCE - *Organization for Security and Co-operation in Europe*. 228

OSI - *Open Systems Interconnection*. 111, 347, 355

OSINT - *Open-Source Intelligence*. 317, 323

OSS - *Operations Support Systems*. 320, 331

OT - *Operational Technologies*. 186

OTX - *Open Threat Exchange*. 122, 134

OVAL - . 35

OWL - . 198, 201

P

P2P - *Peer-to-Peer*. 117, 119

PAC - *Personal Access Codes*. 400

PACS - *Picture Archive and Communications Systems*. 143, 150, 153, 159, 161

PC - *Personal Computers*. 27, 29, 67, 435

PCAP - *Packet Capture*. 355, 360

PDC - *Phasor Data Concentrators*. 251, 258, 259

PDE - *Pattern Detection Engine*. 121

PENTEST - *Penetration Testing*. 315

PET - *Privacy-enhancing Technologies*. 426

PGW - . 333

PHA - *Preliminary Hazard Analysis*. 375

PHI - *Protected Health Information*. 173

PHP - . 350

PHR - *Personal Health Records*. 219, 221

PII - *Personal Identifiable Information*. 173

PIIS - . 95

PIN - *Personal Identification Numbers*. 400

PIS - *Payment Inititation Services*. 95, 96, 153

PISP - *Payment Initiation Services Providers*. 95

PKCE - *Proof Key for Code Exchange*. 100

PLC - *Programmable Logic Controller*. 172, 173, 186

PMU - *Phasor Measurement Unit*. 239, 251, 253, 258, 259

PoD - *Ping of Death*. 348

PPP - *Public Private Partnerships*. 77

Pr(EA) - *Probability of success of each Elementary Actions*. 157

PSD - *Revised Directive on Payment Services*. 5, 95, 96

PSIM - *Physical Security Information Management*. 53, 54, 316

PSP - *Payments Services Providers*. 5

PTZ - *Pan, Tilt, and Zoom*. 402, 406

PV - *Priority Vector*. 266

Q

QC - *Qualified Certificates*. 95, 96

QoE - *Quality of Experience*. 299

QoS - *Quality of Services*. 297, 299, 316, 319, 341, 374, 393

QTS - *Qualified Trust Services*. 95

QWAC - *Qualified Website Authentication Certificates*. 95

R

R&D - *Research & Development*. 434, 435

RA - *Reference Architecture*. 8, 14–21, 23, 25–32, 106, 113, 115

RAN - *Radio Access Network*. 322, 333, 335, 336, 411, 412

RB-SRA - *Rule-based Service Restoration Algorithm*. 262, 263, 267–269

RBF - *Radial Basis Function*. 127

RCS - *Radar Cross Section*. 408, 409

RDM - *Reference Data Model*. 22, 23, 26

ReEIF - *Refined eHealth European Interoperability Framework*. 221

RESISTO - *RESIlience Enhancement and Risk Control Platform for Communication infraSTructure Operators*. 310–312, 314, 315, 317–323, 325, 326, 342, 365, 366, 370–372, 374, 375, 378, 380, 381

REST - *REpresentational State Transfer*. 21–23, 95, 101

REX - *Request-to-exit*. 169, 170

RF - *Radio Frequency*. 322, 408

RFC - . 329

RFID - *Radio Frequency Identification*. 31, 419

RI - *Resilience Indicator*. 312, 314, 323, 325, 380

RMON - *Remote Networking Monitoring*. 257

RNN - *Recurrent Neural Network*. 112

RTS - *Regulatory Technical Standard*. 95, 96

RTU - *Remote Telemetry Unit*. 172

S

SaaS - *Software-as-a-Service*. 207

SBA - . 334

SBID - *Statistical Based Intrusion Detection*. 353

SC - *System Components*. 372–376

SCA - *Strong Customer Authentication*. 96

SCADA - *Supervisory Control And Data Acquisition*. xxi, 167, 172, 173, 186, 386, 387

SCC - *Security Control Center*. 276, 285, 287, 289–291

sCPS - *SMART Cyber-Physical System*. 73

SDA - *Slow Dos Attacks*. 122, 349–351, 353

SDN - *Software Defined Networking*. 320, 329, 330, 333, 335, 337, 338

SDO - *STIX Domain Objects*. 37–39, 126

SDR - *Software Defined Radio*. 412

SDS - *Software Defined Security*. xxii, 320, 338, 340, 341

SE - *State Estimation*. 263, 264

SECaaS - *Security as a Service*. 19, 21, 30–32, 83

SEPA - *Single Euro Payments Area*. 10

SF - *System Function*. 373–376

SI - *Social Impact*. 377

SIEM - *Security Information and Event Management*. xviii, 4, 8, 11, 18, 20, 25, 30, 31, 36, 117, 119, 121, 122, 133, 134

SKB - *Security Knowledge Base*. 9, 35, 39, 40, 44, 45, 49

SLA - *Service Level Agreement*. 17, 19, 341, 342

SME - *Small Medium Enterprises*. 281

SMS - *Security Management Systems*. 402, 403

SMTP - *Simple Mail Transfer Protocol*. 135

SOA - *Service Oriented Architecture*. 253

SOAP - *Simple Object Access Protocol*. 350

SOC - *Security Operation Centre*. 186, 317

SPHINX - *A Universal Cyber Security Toolkit for Health-Care Industry*. 210–223

SPoF - *Single Point of Failure*. 260

SQL - *Structured Query Language*. 23, 100, 303, 318

SR - . 248, 250, 261, 262, 269, 271

SRO - *STIX Relationship Objects*. 37–39, 126

SSL - *Secure Sockets Layer*. 78, 176

STCL - *Short Term Control Loop*. xxii, 312, 314–316, 325

STIX - *Structured Threat Information eXpression*. 9, 19, 22, 26, 27, 35–37, 39–42, 45, 49, 67, 78, 125–127

SVM - *Support Vector Machine*. 114, 127–131

SVR - *Support Vector Regression*. 107

SW - *SoftWare*. 315

SWIFT - *Society for Worldwide Interbank Financial Telecommunication*. 3, 4, 7, 10

SYN - . 348, 351

T

TAXII - *Trusted Automated eXchange of Indicator Information*. 78

TC - . 435

TCP - *Transmission Control Protocol*. 347, 348, 356–359, 403

TIP - *Threat Intelligence Platform*. 37

TIUS - *Threat Intelligence Update Service*. 134

TLC - *Telecommunication*. 329

TLS - *Transport Layer Security*. 95–98, 100

TOPSIS - *Technique for Order of Preference by Similarity to Ideal Solution*. 266, 267, 271

TPP - *Third-party Providers*. 95, 96, 99

TRAS - *Threat Response and Alert System*. 183

TSO - *Transmission System Operator*. 239, 241

TTP - *Trusted Third Party*. 38, 78

U

UAV - *Unmanned Aerial Vehicles*. 316, 321, 406–409

UCO - *Unified Cybersecurity Ontology*. 37

UML - . 198

UMTS - *Universal Mobile Telecommunications System*. 334

URL - *Uniform Resource Locator*. 394

USB - *Universal Serial Bus*. 304

V

VAC - *Video Analytics Component*. 322, 406

VIM - *Virtualized Infrastructure Manager*. 331–333

VLAN - *Virtual Local Area Network*. 334

VM - *Virtual Machine*. 332, 336

VMS - *Video Management System*. 53, 54, 167–171

VNF - *Virtual Network Function*. 331, 332

VNF-FG - *Virtual Network Function Forwarding Graph*. 331

VNFM - *Virtual Network Function Manager*. 331, 332

VoLTE - *Voice over LTE*. 335

VPN - *Virtual Private Network*. 176

W

WAMPAC - *Wide-Area Monitoring, Protection and Control*. 251, 252, 258

WHO - *World Health Organization*. 149, 151

WLAN - *Wireless Local Area Network*. 173, 410, 411

WLS - *Weighted Least Square*. 263

X

XAI - *Explainable Artificial Intelligence*. 112

XL-SIEM - *Cross-Layer SIEM*. 119, 121, 124, 133, 134

XML - *eXtensible Markup Language*. 9, 95

XS2A - *Access to Account*. 95, 96

Part I

Securing Critical Infrastructures of the Financial Sector

DOI: 10.1561/9781680836875.ch1

Chapter 1

Security Challenges for the Critical Infrastructures of the Financial Sector

By Ernesto Troiano, Maurizio Ferraris and John Soldatos

This chapter is an introduction to the first part of the book, which deals with security technologies for the infrastructures of the financial sector. It motivates the need for strong security based on recent security incidents that affected financial institutions. Accordingly, it presents some of the main security challenges for the financial sector, where is also highlights the need for cyber-physical threat intelligence. Furthermore, the chapter presents state-of-the-art technologies that can help confronting the presented challenges. Some of the presented technologies are elaborated in subsequent chapters of the first part of the book.

1.1 Introduction

In the era of globalization, the financial sector comprises some of the most critical infrastructures that underpin our societies and the global economy. In recent years, the critical infrastructures of the financial sector have become more digitalized and interconnected than ever before. Advances in leading edge ICT technologies like Big Data, Internet of Things (IoT), Artificial Intelligence (AI), and blockchains, coupled with a wave of Financial Technology (FinTech) innovations, has resulted in an explosion of the number of financial transactions. Furthermore, the critical assets of financial institutions are no longer only physical (e.g., bank branches, buildings, ATM machines, computer rooms), but rather comprise many different types of cyber assets (e.g., computers, networks, IoT devices) as well.

The increased digitization and sophistication of the critical infrastructures of the financial sector has also raised the importance of cybersecurity in the financial sector. Nevertheless, despite significant investments in cybersecurity, recent large-scale incidents demonstrate that financial organizations remain vulnerable against cyberattacks. As a prominent example, the fraudulent SWIFT (Society for Worldwide Interbank Financial Telecommunication) transactions cyberattack back in February 2016 resulted in $81 million being stolen from the Bangladesh Central Bank. Likewise, the famous "WannaCry" ransomware attacked financial institutions and had a significant adverse impact on Russian and Ukrainian banks. Another major attack took place in 2017, when a data breach at Equifax created a turmoil in the global markets and affected more than 140 million consumers.

In addition to these major incidents, smaller scale attacks against financial institutions happen daily. While most of them are confronted, there are still many cases where these attacks affect the operations of banks and financial institutions, as well as their customers. For instance, back in February 2019, Metro bank was named as a victim of a cyberattack that targeted the codes sent via text messages to customers, as part of the transactions' verification process. A small number of customers of the bank were potentially affected, while the bank reported the issue to relevant security authorities [1]. During the same month, the Bank of Valletta had to shut down all its operations after hackers broke into its systems and moved €13 million into foreign accounts. Specially, the bank shut down all the bank's functions, including branches, ATMs, mobile banking, as well as email services and the website of the bank [2].

In general, the financial sector suffers from security attacks (notably cybersecurity attacks) more than other sectors. During 2016, financial services customers suffered over 60% more cyberattacks than customers in any other sector, while cyberattacks against financial services firms increased by over 70% in 2017. Moreover, a 2018 analysis from the IMF (International Monetary Fund) estimated that emerging cyberattacks could put at risk a significant percentage of the financial institutions' profits, which ranges from 9% to even 50% in worst-case scenarios [3].

In response to the rising number of attacks against financial institutions and their cyber assets, financial sector organizations are allocating more money and effort in increasing their cyber resilience. According to Netscribes, the global cybersecurity market for in financial services is expected to expand at a CAGR (Compound Annual Growth Rate) of 9.81%, leading to a global revenue of USD 42.66 billion by 2023. Other studies reflect a similar estimation, e.g., a Compound Annual

Growth Rate (CAGR) of 10.2% during 2018–2023 and a cybersecurity market growth from USD 152.71 billion in 2018 to USD 248.26 billion by 2023 [4].

1.2 Financial Sector Security Challenges

Through their security investments, financial organizations are striving to confront the challenges described in the following paragraphs. The importance of these challenges has been demonstrated during some of the above-listed security incidents.

1.2.1 Limited Integration Between Physical Security and Cybersecurity

Even though the critical infrastructures of the financial sector comprise both physical and cyber assets, physical security and cybersecurity are still handled in isolation from one another. Specifically, cybersecurity and physical security processes in financial organizations remain “siloed” and fragmented. The latter fragmentation concerns both the technical and the organizational levels, i.e., physical and cybersecurity are handled by different security technologies and different security teams. For instance, physical security systems such as CCTV (Closed Circuit Television) systems, intelligent visual surveillance, security lighting, alarms, access control systems, and biometric authentication are not integrated with cybersecurity platforms like SIEM (Security Information and Event Management) and IDS (Intrusion Detection Systems). Likewise, processes like vulnerability assessment, threat analysis, risk mitigation, and response activities are carried out separately by physical security officers and cybersecurity teams.

This “siloed” nature of systems and process leads to several inefficiencies, including:

- Inefficient security measures that consider the state of the cyber or the physical assets alone, instead of considering the global security context. There are specific types of security attacks (e.g., ATM Network attacks), where security processes like risk assessment and mitigation should consider the status of both types of assets.
- Inability to cope with combined cyber/physical attacks, which are set to proliferate in the years to come. For example, a physical security attack (e.g., unauthorized access to a device or data center) is nowadays one of the best ways to gain access to internal resources and launch a cybersecurity attack as an insider. Indeed, the recent cyberattack against the Bangladesh Central Bank exploited access to physical assets of the bank like SWIFT computing devices.

- Increased costs as several processes are duplicated and overlapping. In this context, an integrated approach to security could help financial organizations streamline their cyber and physical security resources and processes, towards achieving greater efficiencies at a lower cost.

1.2.2 Poor Stakeholders' Collaboration in Securing Financial Services

In an era where financial infrastructures are more connected than ever before, their vulnerabilities are likely to impact other infrastructures and systems in the financial chain, having cascading effects. In this context, stakeholders' collaboration can be a key towards identifying and alleviating issues in a timely manner. However, collaboration is currently limited to exchanging data as required by relevant security regulations and do not extend to join security processes like (collaborative) risk assessment and mitigation.

Information sharing between stakeholders of the financial supply chain is a first and prerequisite step to their collaboration in security issues. In the financial sector, the Financial Services Information Sharing and Analysis Center (FS-ISAC) has been established, as an industry forum for sharing data about critical cybersecurity threats in the financial services industry. FS-ISAC provides its members with access to threat reports with tactical, operational, and strategic levels of analysis for a greater understanding of the tools, methods, and actors targeting the sector. This allows them to better mitigate risk.

Information sharing (e.g., as implemented by FS-ISAC) is a foundation for collaboration in security processes like joint risk scoring for assets and services that are part of the financial services supply chain. Such IT-supported collaborative workflows have been demonstrated in many sectors, including the financial sector. Nevertheless, there are still trust barriers to information sharing and collaboration, especially when data must be shared across private enterprises. Recent advances in IT technologies like blockchain and cloud computing could facilitate the sharing of information and the implementation of collaborative security functionalities.

1.2.3 Compliance to Stringent Regulatory Requirements and Directives

Financial institutions are nowadays faced with a need of complying with a host of regulations, which has a severe impact on their security strategies. For example:

- The Second Payment Services Directive (PSD2): Compliance to the 2nd Payment Services Directive (PSD) demands for banks to be able to interact with multiple Payments Services Providers (PSPs) in the scope of an API-based

Open Banking approach. This raises more cybersecurity concerns and asks for strong security measures like pentesting and vulnerability assessment on the APIs.

- The General Data Privacy Regulation (GDPR): As of May 2018, financial organizations have to comply with the General Data Privacy Regulation (GDPR), which asks for stricter and effective security measures for all assets where personal data are managed and exchanged. Note that GDPR foresees significant penalties for cases of non-compliance, which is one of the reasons why financial organizations are heavily investing in security systems and measures that boost their compliance.
- The Network Information Systems (NIS) Directive [i.e., Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016] [5]: The NIS Directive prescribes security measures for the resilience of the IT systems and networks that support Europe's critical infrastructures, including infrastructures in the financial sector. The prescribed measures include the establishment of risk-driven security polices, as well as the collaboration between security teams (including CERTs (Computer Emergency Response Teams) and CSIRTs (Computer Security Incident Response Teams) at national and international level. The directive defines entities in the Financial services as 2 of the 7 critical sectors and called the member states upon actions to protect and guarantee the availability of their services. Financial organizations are therefore investing in the implementation of the NIS Directive's mandates.
- The EU legislative framework for electronic communications (EU Directive 2009/140/EC) was reformed in 2009 and Article 13a introduced into the Framework directive (Directive 2002/21/EC as amended by Directive 2009/140/EC). Article 13a concerns security and integrity of electronic communications networks and services. The first part of Article 13a requires that providers of networks and services manage security risks and take appropriate security measures to guarantee the security (paragraph 1) and integrity (paragraph 2) of these networks and services. The second part of Article 13a (paragraph 3) requires providers to report significant security breaches and losses of integrity to competent national authorities, who should report about these security incidents to ENISA and the European Commission annually.

1.2.4 The Need for Continuous Monitoring of Transaction and Limited Automation

Financial organizations are nowadays required to secure their infrastructures in a fast moving and volatile environment, which is characterized by a proliferating number of threats and vulnerabilities that are likely to emerge and affect critical infrastructures. Hackers and adversaries are continually taking advantage of leading-edge technologies in order to exploit the rising number of vulnerabilities of the physical and cyber assets of the critical infrastructures. Therefore, it is not practical, and in several cases not possible, to manually carry out all security and protection tasks such as detection, monitoring, patching, reporting, and security policy enforcement activities.

In this context, one of the main challenges faced by the security officers of financial organizations is the poor automation of security functions. To confront this challenge, there is a need for solutions that offer immediate mitigation actions, as well as (semi)automated enforcement of security policies. To this end, financial organizations can take advantage of recent advances in technologies like Artificial Intelligence, Machine Learning and automated orchestration of security functions.

The lack of significant automation is also a setback to fulfilling one of the main security requirements of the financial institutions, which is the ability to monitor transactions without interruptions, i.e. on a 24×7 basis. This is challenging as it requires significant amounts of human resources, including cybersecurity experts and members of security teams. However, it is an essential requirement given that adversarial attacks can happen at any time during the day. Some of the recent attacks against the SWIFT system might have been avoided should a close 24×7 monitoring of transactions and security events was in place.

1.2.5 Lack of Flexibility in Coping with a Proliferating and Dynamic Number of Threats

In addition to automation, security officers of financial organizations are very keen on being flexible when dealing with the proliferating number of threats, including the emergence of several new cyber threats every year. Hence, security departments must be able to deploy new security functions (such as patches or protection policies) very frequently, e.g., daily or even several times per day. In this direction, financial organizations could benefit from latest developments in software engineering practices and methodologies such as the DevOps (Development and Operations) paradigm. Recent research initiatives are exploring the use of DevOps in security systems engineering, which is sometimes called DevSecOps.

1.2.6 Digital Culture and Education

The human factor plays a significant role in alleviating cybersecurity attacks. Proper digital culture and education can provide a sound basis for complying with the mandates of security policies, while avoiding mistakes that could open backdoors to malicious parties. Nevertheless, there is currently a proclaimed gap in digital knowledge in general and specifically in cybersecurity. This holds true for physical security teams as well. Hence, the cybersecurity knowledge gap hinders the implementation of integrated security strategies, while being a setback to the cyber resilience of modern financial institutions.

1.3 Solution Guidelines

With these challenges in mind, the following paragraphs provide solution guidelines and recommendations about securing modern financial organizations. The presented solutions are empowered by advanced security technologies and include the technologies presented in subsequent chapters of this first part of the book.

1.3.1 Structuring and Developing Integrated Security Systems

For over a decade, financial organizations have been deploying and using systems that process and analyze digital information towards implementing cyber defense strategies. Prominent examples include network monitoring and analysis probes, SIEM systems, vulnerability scanners, and more. However, these systems cannot adequately support the definition and implementation of integrated security policies, i.e., policies addressing cyber and physical aspects at the same time. Therefore, there is a need for designing and implementing more integrated systems that will be able to combine cybersecurity aspects with information about physical security, such as information derived from CCTV (Closed Circuit Television) cameras, access control systems, biometric systems, and more.

The design and implementation of integrated security policies requires rethinking of the architecture of the various security platforms, to a direction that considers physical information and devices. Thus, there is a need for new security architectures. The latter can take advantage of the recent advances in Industry 4.0 and the Industrial IoT, including relevant reference architectures such as the Industrial Internet Security Framework (IISF) of the Industrial Internet Consortium [6]. In this context, Chapter 2 introduces the Reference Architecture (RA) of the FINSEC project, which is destined to facilitate the development of data-driven security systems for the financial sector, including systems that address the cyber-physical nature of modern cyber physical infrastructures. As outlined in Chapter 2, the FINSEC RA is implemented based on modern microservices-based approach and can be used to support DevSecOps methodologies in building software systems for security.

1.3.2 Integrated Security Knowledge Modeling

Integrated (i.e., cyber and physical) security systems must deal with data for both cyber and physical threats. Likewise, they should capture and maintain knowledge about both cyberattacks and physical attacks, including combined cyber/physical attacks. Thus, there is need for extending existing security models and format, with constructs that enable them to represent integrated security knowledge. State-of-the-art knowledge bases for cybersecurity consolidate several sources of knowledge for Cyber Threat Intelligence (CTI), such as:

- CPE (Common Platform Enumeration), which is a structured naming scheme for IT software, systems, and packages.
- CWE (Common Weakness Enumeration), which lists common software's vulnerabilities.
- CAPEC (Common Attack Pattern Enumeration and Classification), which lists common attack patterns on software and their taxonomy.
- CVE (Common Vulnerabilities and Exposures), which lists all publicly known cybersecurity vulnerabilities and exposures.

Furthermore, they can also collect and store external CTI data sources through available documents in various formats like JSON (JavaScript Object Notation) and XML (eXtensible Markup Language). There are several knowledge bases available, including commercial SKBs (Security Knowledge Bases) from major security vendors and SKBs from standards development bodies [e.g., the OWASP (Open Web Application Security Project) Security Knowledge Framework]. Nevertheless, these knowledge bases do not include security knowledge for physical assets, which limits their ability to support integrated (i.e., cyber/physical) security.

Hence, there is a need for enhancing knowledge bases and formats for representing cyber-threat intelligence, with information about physical assets and security, towards Cyber-physical Threat Intelligence (CPTI). In-line with this requirement, Chapter 3 introduces FINSTIX, a STIX (Structured Threat Information Expression) based format, for supporting integrated security modeling for critical infrastructures in the financial sector.

1.3.3 Automation and Flexibility

To increase the automation of security processes, financial organizations are nowadays offered with the opportunity of leveraging Machine Learning (ML) and Artificial Intelligence (AI) on large volumes of security data. Specifically, financial institutions are currently collecting large amounts of cybersecurity and physical security related information through many different sensors and probes. This information, if analyzed properly, could provide insights about possible security incidents. Moreover, it can also facilitate the extraction of hidden attack patterns, beyond the ones already known and registered within security knowledge bases. Also, it is possible to employ predictive analytics towards identifying and anticipating security threats before their materialize. This can greatly boost the preparedness of security teams like CERTs.

AI and ML algorithms can boost not only the intelligence and proactiveness of the security processes, but also their automation as well. Specifically, they can automate security and surveillance processes through obviating manual surveillance and tracking of security information streams (e.g., from CCTV systems). Furthermore, they can boost the continuous, 24×7, monitoring of financial systems and transactions, through lowering the human resources needed for the surveillance tasks.

Two of the following chapters introduce data-driven, AI-based solutions for security and surveillance. Chapter 4 presents an AI-based gateway that can combine cyber and physical surveillance in financial environments. The gateway offers a range of intelligence and performance features, which are detailed in the chapter. Also, Chapter 7 presents a novel system for collecting security data from different probes, which incorporates security intelligence (e.g., awareness about security events) towards adapting the rate, the scope, and the context of the data collection.

1.3.4 Information Sharing and Collaboration Across the Financial Services Supply Chain

As already outlined, financial institutions are nowadays digital interconnected as part of different value chains and purposes. SWIFT and SEPA (Single Euro Payments Area) transactions are, for example, carried out across interconnected institutions. As another example, various financial enterprises are interconnected in the scope of trading and stock exchange transactions. Interconnected enterprises are vulnerable to attacks that originate from attacks against other stakeholders in the value chains where they participate. Specifically, financial organizations should not only consider the status of their assets and infrastructures. Rather, they should keep an eye on the status of interconnected infrastructures as well. A potential vulnerability in a connected infrastructure can influence other stakeholders in the supply chain.

Moreover, to address supply chains security, stakeholders had better collaborate in their security processes. As a prominent example, enterprises could engage in collaborative assessments of the risk factors that are associated with their assets. Such processes can be empowered by the automated and seamless sharing of information across stakeholders of the supply chain. Currently, financial organizations share such information as part of regulatory mandates and in the scope of their participation in initiatives like the Financial Services Information Sharing and Analysis Center (FS-ISAC). Nevertheless, the level of security information sharing is still quite low. Lack of trust is one of the reasons that make organizations reluctant to share security information. In recent years, distributed ledger technologies (i.e., blockchain technologies) are explored as a means of sharing information across financial organizations in a decentralized and trustworthy way. Chapter 5 presents a relevant approach, where data shared through a blockchain is used to facilitate collaborative risk assessment.

1.3.5 Regulatory Compliance Technologies

To confront the challenges of regulatory compliance, financial organizations need technologies that facilitate the implementation of relevant technical measures. As a prominent example, data anonymization and data encryption can be used to facilitate adherence to GDPR principles. Likewise, SIEM systems can be used to collect and analyze information about access, transfer, and use of data in an organization, towards identifying potential data breaches. In this context, Chapter 6 presents a suite of security tools for PSD2 compliance. These include, for example, pentesting tools for Open Banking APIs (Application Programming Interfaces), which are destined to identify vulnerabilities of these APIs prior to their use in PSD2 compliant applications.

1.3.6 Security-by-Design and Privacy-by-Design

Beyond regulatory compliance, financial organizations need to adopt new principles regarding the design and implementation of their applications. Specifically, they are expected to adhere to the security-by-design and privacy-by-design principles. The latter should become the preferred path of the software design and development cycle for financial organizations like banks. Likewise, traditional serialized development approaches should be updated towards more flexible and responsive approaches that involve the design and implementation of security controls early in the application development life cycle. Note that privacy-by-design is referenced in the text of the GDPR regulation, and hence, it can serve as a basis for achieving GDPR compliance as well.

1.3.7 Security Education and Training

Financial organizations should heavily invest in security education and training with a twofold objective: First to close the knowledge gap about cybersecurity issues, and second towards engaging the organization's personnel in IT security, regardless of their background and security knowledge. Such measures will help ensuring that employees are no longer one of the weakest links in the security value chain. Along with investments in training and education, financial organizations should be investing in IT security awareness campaigns. In this direction, the FINSEC project is contributing to training and awareness raising based on various trainings and presentations that are available through the market platform of the project, i.e., finsecurity.eu.

1.4 Conclusions

The critical infrastructures of the financial sector are increasing in size, complexity, and sophistication, while at the same time comprising both cyber and physical elements. At the same time, financial organizations are obliged to comply with many and complex regulations and directives about security, privacy, and data protection. As a result, financial enterprises must deal with increased security vulnerabilities and threats in a rapidly evolving regulatory environment. To this end, they are increasing their investments in cybersecurity and its intersection with physical security. Despite the rising investments, they remain vulnerable to security and privacy threats, as evident in several notorious incidents that have occurred during the last couple of years.

In order to properly secure the critical infrastructures for the financial sector, there is a need for new integrated approaches that addresses physical and cybersecurity together rather than dealing with them in a "siloed" fashion. To this end, financial organizations should benefit from the capabilities of emerging technologies like Big Data and AI analytics for security monitoring and automation, while at the same time leveraging the flexibility of the DevOps paradigm that provides opportunity for frequent changes to security measures and policies (e.g., patching on a daily basis). Likewise, integrated approaches to security knowledge modeling and information sharing can be employed. Following chapters of the first part of the book will illustrate novel technologies for cyber-physical threat intelligence, which address several of the security challenges that are currently faced by financial organizations.

Acknowledgments

Part of this work has been carried out in the scope of the FINSEC project (contract number 786727), which is co-funded by the European Commission in the scope of its H2020 program. The authors acknowledge valuable help and contributions from all partners of the project.

References

[1] Natasha Bernal, "Metro Bank hit by cyber-attack used to empty customer accounts," The Telegraph, February, 2019, available at: https://www.telegraph.co.uk/technology/2019/02/01/metro-bank-hit-cyber-attack-used-empty-customer-accounts/

[2] "BOV goes dark after hackers go after €13 m," Time of Valletta, February 2019, available at: https://timesofmalta.com/articles/view/bank-of-valletta-goes-dark-after-detecting-cyber-attack.701896

[3] Antoine Bouveret, "Cyber Risk for the Financial Sector: A Framework for Quantitative Assessment," International Monetary Fund (IMF) Paper, July 2018.

[4] "Cyber security in Financial Services Market: Market players, Market Research, Growth During, to 2018–2023," Marketwatch Press Release, September 2019, available at: https://www.marketwatch.com/press-release/cyber-security-in-financial-services-market-market-players-market-research-growth-during-to-2018-2023-2019-09-17

[5] European Parliament and Council. Directive (EU) 2016/1148, measures for a high common level of security of network and information systems across the Union. 2016.

[6] The Industrial Internet Security Framework Technical Report, available at: https://www.iiconsortium.org/IISF.htm (Accessed February 2020).

DOI: 10.1561/9781680836875.ch2

Chapter 2

A Reference Architecture for Securing Infrastructures in the Finance Sector

By Ernesto Troiano, John Soldatos, Ariana Polyviou, Alessandro Mamelli and Ilesh Dattani

While the critical infrastructures of financial institutions encapsulate both physical and cyber assets, their security is regarded in isolation from one another making financial institutions vulnerable to attacks against their physical and cyber assets. Drawing on an analysis of the needs of financial institutions, this Chapter proposes the FINSEC Reference Architecture (RA). The FINSEC RA aims at addressing the cyber-physical integration, the collaboration of stakeholders in the financial sector, the regulations and recommendations, the low level of automation at probes and platform/services level, and the lack of flexibility in front of a wide range of threats. The proposed architecture accounts for the regulations and recommendations applicable to the financial security sector and considers relevant recommendations from standards development organizations such as ENISA. The Chapter reflects on the methodology adopted for building the RA and presents the schema of the RA including its structure tiers and building blocks corresponding to the different functionalities of FINSEC platform.

2.1 Securing Finance Sector Infrastructures: The Challenges

Responding to the demands of the global economy, the financial sector's critical infrastructures have become more digitalized and interconnected than ever before. And yet, the wide deployment of Big Data, Internet of Things (IoT), Artificial Intelligence (AI), blockchains, mobile Apps, Cloud services, and web infrastructures, coupled with a wave of financial technology (FinTech) innovations, has resulted in an explosion of the number of financial transactions. Thus, the critical assets of financial institutions are no longer only physical (e.g., bank branches, buildings, ATM dispenser), but comprise many different cyber assets (e.g., computers, networks, IoT devices) as well. As discussed extensively in the chapter that introduces the security challenges of the financial sector despite the existing cybersecurity measures, recent cyberattacks demonstrate that financial organizations remain vulnerable. For this reason, financial institutions are currently investing in extending their cyber-resilience capacity.

Despite the fact that critical infrastructures of the financial sector comprise both physical and cyber assets, their security is addressed in isolation from one another. This includes different technologies and security teams handling security attacks of the physical and cyber assets. As a result, these practices lead to inefficient security measures, increased cost, and limited ability to effectively address attacks that combine both the physical and cyber assets. Enhancing the collaboration of all relevant stakeholders is necessary for identifying and alleviating issues in a timely manner. However, rather than collaborating extensively on risk assessment and mitigation, their current collaboration is rather limited to the exchange of data as imposed by security regulations.

Another major challenge for the security of financial institutions is the automation of security functions. Hackers and adversaries are continually taking advantage of leading-edge technologies in order to exploit the rising number of vulnerabilities of the physical and cyber assets. While the level of automation is low, it is impossible for security officers to execute all security and protection tasks such as detection, monitoring, patching, reporting, and security policy enforcement activities. This challenge urges the need for solutions that offer immediate mitigation actions, as well as (semi)automated enforcement of security policies. Recent advances such as Artificial Intelligence, Machine Learning, and automated orchestration of security functions can provide solutions in this respect.

This Chapter proposes a reference architecture for securing critical infrastructures of the financial sector, FINSEC RA, while accounting for the emerging need to address physical and cyber asset security combined. The proposed architecture incorporates recent technological advances to offer an integrated security solution for the critical infrastructures of financial institutions.

2.2 A Reference Architecture for Securing Critical Infrastructures of the Financial Sector

2.2.1 FINSEC RA: Background and Rationale

In order to address these challenges, vendors and integrators of security solutions need security middleware libraries and blueprints for the development, deployment, and operation of security systems that address the limitations of existing platforms in terms of supporting integrated (cyber/physical) security, boosting regulatory compliance, increasing automation, as well as ensuring flexibility and speed in deploying security functions. In this direction, a security Reference Architecture (RA) offers a synthesis of best practices based on past experiences and relevant blueprints for security solutions. A Reference Architecture (RA) can also serve as a conceptual framework for building security systems faster, while minimizing development, deployment, and operational risks. Furthermore, an RA serves as a device for communicating security contexts and solutions requirements across interested stakeholders. It therefore provides a common context and vocabulary, along with a repository of patterns for use by interested stakeholders. As such it facilitates teamwork in developing, deploying, and operating security systems for the financial sector.

The initial proposal for a new NIS Directive and the discussions documentation [1] states that an "insufficient level of protection against network and information security incidents, risks and threats across the EU [..., may undermine, ed.] the proper functioning of the Internal market." This statement is particularly relevant in the finance sector, where a failure of critical IT infrastructure can lead to major damages to financial markets with deep economic consequences. The H2020 FINSEC project is intended to support the need for better protection and resilience of this critical infrastructure.

FINSEC is a joint effort of security experts and financial organizations towards providing integrated (i.e., cyber/physical) solutions for the critical infrastructures of the finance sector. One of the main results of the project is a Reference Architecture (RA) for the development, deployment, and operation of integrated solutions in the finance sector. The RA is motivated by the need to apply innovative patterns to the development and deployment of security systems for the critical infrastructures of the sector. As such it's a foundation for the solutions that the project provides to different financial organizations including banks, payments organizations, and FinTech enterprises.

The development of FINSEC RA has considered concepts and building blocks from some well-known and accepted generic RA, such as the RA of the Industrial Internet Consortium and its Industrial Internet Security Framework (IISF) [2]. In this way, the FINSEC RA leverages experiences from established communities, while at the same time being in-line with the evolution of security concepts that have emerged and/or evolved in these communities.

2.2.2 FINSEC RA: Accounting for ENISA's Recommendations

The specification of the FINSEC RA has also considered recommendations of standards development organizations for security in the finance sector. In particular, we have accounted for the recommendations provided by ENISA in terms of:

- Network and information security in the Financial Sector [3].
- The use of Cloud Computing in the Finance Sector [4].

With respect to FINSEC, the above-listed recommendations present some limitations for 2019–20 state of the art, as they are focused on circa 2014–15 issues of cybersecurity, rather than current integrated (cyber/physical) security. Nevertheless, they are motivated by the updated (circa 2019) finance sector regulatory landscape and deals with two of the most important elements of contemporary finance sector infrastructures, namely networks and cloud computing infrastructures. Note also that the above-listed reports and recommendations do not focus on technical measures only (that can be addressed in a technical architecture), but extend to policy and organizational recommendations.

ENISA's report on the security of cloud computing infrastructures for the financial sector [4] outlines the challenges of managing governance and compliance risk, while providing tools and techniques for negotiating SLAs, especially in cases where smaller institutions are involved as customers of the cloud services. Furthermore, the report outlines the importance of improving security and privacy certification. Relevant recommendations for confronting these challenges are included in the report and addressed to financial institutions, cloud services providers, and regulators. In brief, the report includes the following recommendations, which we have built on in the RA:

- Extending the national good practices and standards in the areas of Cloud governance and risk management.
- Defining practices and standards for incident information sharing.

- Defining minimum security requirements for adoption of Cloud computing in FIs.
- Provision of transparency and assurance from cloud providers to financial institutions.
- Better informing both regulators and financial institutions about the security risks and opportunities connected to the use of cloud computing.
- Continuing effort towards harmonizing the legal and regulatory environment within the European Union.

Similarly, ENISA's report on network and information security [3] provides recommendations for strengthening the security of networking and communication infrastructures, including:

- Consolidating scattered NIS obligations in supervisory guidelines.
- Establishing guidelines on how NIS supervision practices in the Finance sector apply by extension to their supply chain, including Cloud providers that operate financial services.
- Establishing guidelines which summarize the key conditions for the adoption of Cloud-based applications or services in the Finance sector.
- Organizing regular and voluntary NIS stress tests in the Finance sector towards identifying possible black swan risks and uncovering to the greatest extent possible unknown risks.

Table 2.1 outlines how FINSEC RA aligns to some of the above-listed recommendations, notably through relevant technical measures:

2.2.3 Logical Design

2.2.3.1 Tiered Approach

The main goal of the RA is to alleviate the currently "siloed" landscape of physical and cybersecurity through enabling financial organizations to deploy integrated security solutions. The latter are characterized by the seamless flow of security information for both cyber and physical assets to the security department and teams of the organization. Hence, FINSEC RA does not focus on the physical security and the IT departments only, but rather addresses the needs of the top level management of organizations, notably in terms of managers [e.g., CSO (Chief Security Officer) or CEO (Chief Executive Officer)] that are in charge of the resilience of the organization.

Solutions that adhere to the RA will leverage security monitoring probes available in the organizations, including existing cybersecurity applications (e.g., SIEM

Table 2.1. RA alignment according to ENISA recommendations.

ENISA Recommendations	FINSEC RA Alignment
Defining practices and standards for incident information sharing	FINSEC and the FINSEC RA come with a specification of a data model for exchanging security and threat intelligence information in the Finance Sector, i.e., the STIX-derived FINSTIX model. FINSTIX facilitates the modeling and sharing of security incidents as well. Moreover, the FINSEC RA includes a module dedicated to security information sharing (including sharing of incidents in the supply chain).
Boosting SLA management in the Financial Supply Chain	The information sharing module of the FINSEC RA facilitates transparent data sharing in the financial supply chain, which can boost SLA management. One of the FINSEC pilots involves provision of security services to smaller financial organizations that take advantage of cloud-based financial services. Moreover, the FINSEC RA is designed as a microservices-based architecture with data-driven security services in the cloud that can support a cloud-based SECaaS (Security as a Service) paradigm. It can also provide the means for sharing SECaaS-related information for managing the respective SLAs.
Address Risks through Stress Testing	The core security services of the FINSEC RA include risk assessment, vulnerability assessment, and pentesting. These can be used by financial organizations in order to perform security stress tests in their infrastructures in-line with ENISA's recommendations for the larger system players in the sector (i.e., financial institutions).
Support for Transparency and Assurance	FINSEC RA promotes a transparent, decentralized information sharing approach, which is realized in the project based on a blockchain infrastructure. In this way, FINSEC contributes to transparency, while at the same time providing data-driven tools (e.g., analytics) for the implementation of regulatory compliance and assurance services.

systems, antivirus applications, log scanning probes) and available physical security systems [e.g., a PSIM (Physical Security Information Management System), a CCTV (Closed Circuit Television System), biometric access control systems]. These probes will provide security data that will drive security functionalities such as risk assessments, management of alerts, and compliance auditing functions.

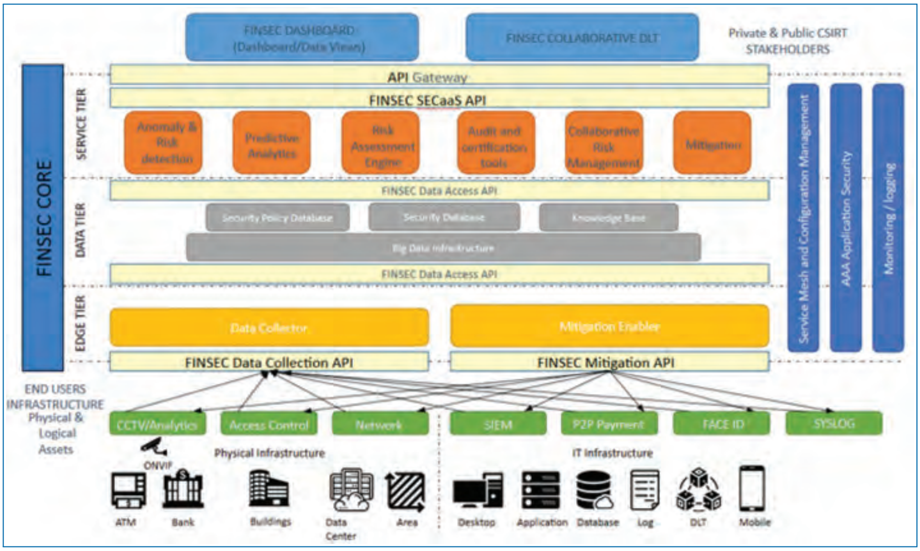


Figure 2.1. Data-driven API.

The RA is structured in tiered approach, yet it also includes cross-cutting elements that do not belong to a single, but rather to multiple tiers. Nevertheless, inline with the previously presented principles, the architecture is modular, and from a physical perspective, it enables every module to communicate with any other like a modern microservices architecture. The Figure 2.1 illustrates the main modules and tiers of the RA.

With reference to figure, the modules of the RA are structured based on the following tiers:

- **Field Tier:** The Field Tier is the lower level of the RA and includes the probes and their APIs, whose role is extracting raw data from the physical and logical assets to be protected against threats. For example, CCTV analytics and SIEM are involved in this layer to give useful information about potential attacks to the upper tiers.
- **Edge Tier:** The Edge Tier contains the Actuation Enabler and a Data Collection module, which is needed to filter the needed information during their flow towards the upper levels. The Actuation Enabler is responsible to allow some actions to be done from the upper layers onto the probes, such as the shutdown of a server in case of threat or the close of an automatic door of a protected room.
- **Data Tier:** The Data Tier is the logical layer where information is stored and is organized into three different storage infrastructures, providing consisting data access API to all other modules.

- **Service Tier:** The Service Tier is where the kernel applications of the FINSEC and the security toolbox are running, able to be used by the external world.
- **Presentation and Communication Tier:** The Presentation and Communication tier offers interface to rest of the world (e.g., consumers of the security services that adhere to the RA). This tier provides dashboards that monitor data and assets, along with the FINSEC Collaborative Module that supports sharing of security information with other financial organizations regardless of whether the latter are running systems compliant with the RA or not.

The FINSEC Core platform, delimited by the blue bar in the picture, comprises three tiers, namely the edge, data, and service tiers. It also specifies the two main interfaces that are used to support the interactions of the data-driven platform with other systems and applications:

- The northbound API towards higher level applications (e.g., end-user applications), called FINSEC SECaaS API to align to the DoA and the core concept of FINSEC. It represents a consistent and unified view of the individual APIs exposed by the service tier high-level services that represent the "major intelligence" of the platform. The FINSEC SECaaS API is exposed by the API Gateway, which is the single-entry point to the system for external clients. Among other capabilities, the API Gateway provides and supports Authentication, Authorization, and Accounting (AAA) services, which conceptually are part of the two vertical modules on the right of the figure (Application Security and Monitoring/logging).
- The southbound API interface, consisting of an EVENT API and PROBE API, allows communication between the Edge Tier and physical and cybersecurity probes.

The FINSEC SECaaS API is leveraged and invoked by external (north end) Business Client Applications (upper side of the figure). They are outside of the FINSEC core platform and interact with it only through the FINSEC SECaaS REST API. Some typical examples of business client applications include:

- **The FINSEC Dashboard application,** which is a (WEB) GUI used by the profiled end users of the platform. Note that in addition to the FINSEC dashboard application, additional dashboards can be implemented using the above-listed APIs.
- **The FINSEC Collaboration application,** which enables the collaboration of platforms and applications across the financial services supply chain (e.g., security data sharing). Likewise, additional applications for supply chain

security can be implemented by leveraging these APIs, such as applications for collaborative risk scoring and assessment.
- **Other Third-party Applications** that exploit the data and security capabilities of FINSEC.

2.2.3.2 The Service Tier

The Service Tier defines the high-level services that represent the "major intelligence" of the platform. The Service Tier services communicate with each other in three possible ways as follows:

- **SYNCHRONOUS,** through their REST API (in this case, being the services internal to the platform, it is not necessary to use AAA).
- **ASYNCHRONOUS,** via an MQ bus, yet in this case, queues and messages formats must be defined.
- **ASYNCHRONOUS,** through the DB (Database) Infrastructure.

2.2.3.3 The Data Tier

The Data Tier provides an infrastructure to serve data that follow the FINSEC Reference Data Model (RDM), which extends the STIX standards and is illustrated in latter paragraph. It provides access in read/write via a Data Access API, exposed by an ad hoc service of the platform (Data Manager). This module exposes convenient data access and manipulation functions to clients, is responsible for ensuring validation of input data against the data model and abstracts away the actual underlying DB engine(s), which can be changed without affecting upper-level services.

A possible alternative option, which enables the avoidance of an intermediate data access layer, is to use the CRUD (Create, Retrieve, Update, Delete) REST API already exposed by the DB engine (if available, depending on the DB engine chosen for the implementation) and rely on DB validation rules for ensuring consistency and validation of data with reference to the FINSEC RDM. The latter can be used by the modules of the Service tier to communicate with each other, in-line with the third of the above-listed approaches for communication.

In addition to data conforming to the common data model, the DB infrastructure may contain additional ad hoc data stores for private data reserved to the individual Service tier modules, useful for enabling their own internal logic. The concept in this case is that the individual service could still have a private DB schema for its own settings/local data, e.g., for processing with its own algorithms, and then proceed to publish data on the common DB schema (via the Data Access API) following the FINSEC data model only once it has identified events useful for the common intelligence of the system, as previously mentioned.

The Data Tier provides the fundamental service for and will be based on:

- A Data Base suited to manage the non-structured Threat Information made by events, incidents, logs, etc. It can be either a non-traditional DB (e.g., NoSQL, memSQL) or a conventional SQL relational DB.
- A Big Data Infrastructure to manage the large amount of data to be processed and distributed according to the requirements of the client modules, typically those ones of the Service tier needing BD/AI capabilities to perform their business logic.
- The Security Knowledge Base, which is used to automatically resolve observed data streams into known threats, vulnerabilities, and attacks encoded in the database.

2.2.3.4 The Service Tier

The Edge Tier communicates with the infrastructure (IT/Physical) through the southbound API interface. This API consists of the union of two distinct APIs:

- **Event API,** which is implemented to receive events in push and/or pull modes, and it is invoked by the probes.
- **Probe API,** which is implemented by probes to receive commands from FINSEC. In this case, probes can operate as actuators as well.

Overall, probes send events formulated based on the FINSEC RDM, in all cases when they want to publish data on the Data Tier (e.g., the DB infrastructure and possibly ingestion in the Big Data Infrastructure).

2.2.4 Main Services and Building Blocks of the RA

Each of the modules of the RA is thought as a black box with proper interfaces executing specific functions. Moreover, each module can be implemented as a software manageable and independently deployable service, i.e., respecting the microservices architecture (MSA) paradigm and communicating with standard interfaces (i.e., REST API). The list of the modules and services of the Reference Architecture that are depicted in the above figure are as follows:

- **FINSEC Dashboard:** Web application that presents events, threats, incidents, logs, etc., in a User Graphical Interface. The application will be web-based and will interact with the other microservices to gather information to be present to the dashboard graphically and intuitively.
- **FINSEC Collaborative Module:** Service application for collaborative security information sharing and Threat Intelligence. The application will have a

microservices interface that will provide APIs for exchange information about threats and mitigations. Exchanged data is based on FINSEC Data Model.

- **API Gateway:** API Gateway is a fully managed service that provides to other services ways to create, publish, maintain, monitor, and secure APIs at any scale.
- **Actuation:** Application that offers API to other services to operate on the physical and logical infrastructure sending commands to the physical or logical components.
- **Anomaly & Risk detection prioritization:** Application for anomaly and risk analysis. It consumes current data sources (logs, incidents, etc.) and produces incidents and alarms. Application consumes DATA Access API and push threat information using API of other services (e.g. Dashboard, Collaborative Module, etc.).
- **Predictive Analytics:** Application that will analyze risk and threats from current data sources (logs, incidents) and predicts threats and patterns of threats. Application uses DATA Access API and push threat information using API of other services (e.g., Dashboard, Collaborative Module, etc.).
- **Risk Assessment Engine:** Application for Real-time assessment of security risks, including business interpretation. It analyzes current model of assets associated with business risks levels stored as data model in DB and produces a risk assessment analysis. The Models are produced by the Audit and certification tool. Application uses DATA Access API push threat information using API of other services (e.g., Dashboard, Collaborative Risk Management modules, etc.).
- **Audit and certification tool:** Web Application with HMI to produce a data model representation of assets of the infrastructure. Application will be basically a Data Entry application plus Import from other data sources. Application produces reports displayable and exportable (e.g., pdf).
- **Collaborative Risk Management:** Application for collaborative risk analysis and management in the financial supply chain. It can be implemented based on either centralized (e.g., a centralized database accessed by all stakeholders) or decentralized approaches (e.g., a distributed ledger approach). The module provides an API for other services to push threat information.
- **MQ BUS:** This is an asynchronous Message Passing Application. It provides Push/Pull APIs for basic message passing.
- **Security Database:** A NoSQL application for storing data according to the FINSEC Data Model.
- **Knowledge Base:** A NoSQL application for storing data according to the FINSEC Data Model. It incorporates knowledge from various sources (including vulnerability databases) and used to automate the resolution of

threats and vulnerabilities as part of security functionalities like risk management. It provides a CRUD (Create, Retrieve, Update, Delete) API for storing Knowledge Base documents.

- **Big Data Infrastructure:** It is a distributed File System Application. It provides API for scaling data across multiple servers.
- **Data Collection:** Application module that provides API to EDGE services like CCTV or SIEM for pushing data (events, logs, etc.) to the Security Database. The application also performs normalization and prioritization to the information supplied by the EDGE applications.
- **Actuation Enabler:** Application module that provides API to the ACTUATOR service pushing action to the Logical and Physical infrastructure (e.g., shutdown of a server or close a door of a data center). The application performs abstraction and normalization to adapt to different EDGE components.
- **CCTV/Analytics:** Any Video Surveillance application can be integrated will use FINSEC Event API to push information to the FINSEC core and provides FINSEC Probe API to interact with the EDGE components.
- **SIEM:** Any Security Information and Event Management (SIEM) can be integrated as long as will use FINSEC Event API to push information to the FINSEC core and provides FINSEC Probe API to interact with the EDGE components.
- **Logical Probe:** Field-level logical sensor/actuator to give data on the status of the assets such as logs and actuation commands on logical assets (e.g., shutting down a server to protect it). Note that any probe can be integrated as long as will use FINSEC Event API to push information to the FINSEC core and provides FINSEC Probe API to interact with the EDGE components.
- **Pentest Tool:** Service application acting on the field to extract information on the status of the assets. The Pentest is an assessment of the capacity of an asset to react to a penetration attempt by an actor, through the simulation of an attack. Service provides APIs will give information about the status of the asset and the attack typology under simulation.
- **Service Mesh and Configuration Management:** Service application that provides API to discover services within the infrastructure; moreover, their configuration will be done via this building block. Its APIs will need to give the user the current configuration for each service within the RA.
- **AAA Application Security:** Service application that provides basic API for Authentication, Authorization, and Accounting users and other services within FINSEC platform. It is part of the "vertical" building blocks, not belonging to any specific tier.

- **Monitoring/Logging "Diagnostic" Module:** It provides API for storing and retrieving logs from other services (e.g., connection/disconnection, search for services, use of certain services, getting warning and alarms, actuation on cyber or physical assets, etc.).

All services must provide standard API to start/stop/shut down/monitor/status of the application.

2.2.5 FINSTIX: The FINSEC Reference Data Model

The data-driven operations and data flows specified in the FINSEC RA hinge on the adoption and use of common data semantics for the full set of data/information that are exchanged between the modules of the architecture. To this end, the FINSEC RA is accompanied by a Reference Data Model (RDM) specification, which specifies the format and the semantics of the security data that flow across the modules of the architecture. The RDM is based on the second version of the STIX^{TM} (Structured Threat Information Expression) (STIX2) [5], which is one of the most prominent standards for sharing threat intelligent information. In particular, the FINSEC RDM, which is conveniently called FINSTIX, is an extension of STIX2 into the physical and logical domain. FINSTIX has been developed based on the following principles that facilitate its implementation and integration with solutions that adhere to the FINSEC RA:

- The FINSTIX Data Model basic object is a sequence of key values that can be passed as JSON (JavaScript Object Notation).
- The FINSTIX Data Model general object is an aggregate of more objects and relations still expressed in JSON.
- FINSTIX includes information relevant and specific to the financial sector, including common threats and vulnerabilities faced by financial organizations.
- FINSTIX defines other objects and relations to STIX2 to cope with the correlation of physical and logical data, as a means of supported cyber and physical security integration.

In the scope of solutions that comply with the FINSEC RA, probes generate events and observed data according to the FINSTIX Data Model. Likewise, Data Collectors (DC) have the function to gather data from probes normalizing, sanitizing, prioritizing, and storing CPTI into the Data Layer. In other words, a DC knows the syntax-semantic and add or subtract further information to the FINSTIX objects passing through. Moreover, security knowledge [e.g., as of part of the Knowledge Base (KB) of the FINSEC RA] is represented with FINSTIX objects as well. Also, any analytics algorithms [including predictive analytics based on machine learning and Artificial Intelligence (AI) techniques] in security applications use events, observed data, and the Knowledge base and Asset Models to produce Cyber-physical Threat Intelligence.

FINSTIX includes the STIX Domain Objects (SDO) already defined by STIX2, including Identity, Observed Data, Indicator, Intrusion Set, Vulnerability, Tool, Attack Pattern, Campaign, Malware, Threat Actor, Course of Action, and Report. Nevertheless, FINSEC specifies several extensions to STIX2, notably extensions that address security use cases of the financial sector. These extensions are specified in terms of custom objects like:

- **Organization** that comprises information about a financial organization.
- **Asset** encoding information about an organization's valuable infrastructure such as PC, server rooms, ATMs, applications, and everything else inside the organization that is considered crucial.
- **Area of Interest,** a logical/physical area inside an asset such as the screen/keyboard of an ATM or an indoor area (server room).
- **Service** which signifies a collection of assets forming a publicly exposed service.
- **Probe** that is used to support the security monitoring infrastructure. A Probe usually monitors one or more areas of Interest.
- **Probe Configuration** that provides data sent to a probe in order to configure details such as the area under monitoring or the bit rate of the monitoring process.
- **Event** including information of something that happened or is happening.
- **Person** which extends the STIX Identity objects and is used to describe people involved in the events created by the probes.
- **Risk,** the calculated risk for a specific asset or service.
- **Risk Configuration** which provides information needed to optimize the risk assessment (e.g., triggers and other useful options).
- **Regulation,** an object used to depict a regulation violation. The regulation violation information can be communicated to Regulatory authorities and other Organizations.
- **CPTI** which is the principal object that collects and provides threat information. One or more CPTI objects are used to generate the output of the threat intelligence process, i.e., a report about ongoing or possible future attacks on one or more assets belonging to the infrastructure.

A detailed presentation of the FINSTIX specification is out of the scope of this whitepaper. Interested readers shall contact the FINSEC Project coordinator.

2.3 Security Use Cases for Financial Institutions

The FINSEC RA, along with the FINSTIX specification, enables the implementation of a wide range of security use cases for financial institutions. Some prominent examples follow and also illustrated in Figure 2.2.

2.3.1 SWIFT Network Attacks

The Society for Worldwide Interbank Financial Telecommunication (SWIFT) provides a network that enables financial institutions to send and receive information about financial transactions in a secure, standardized, and reliable environment. It is a messaging network, which facilitates the secure transmission of information and instructions based on a standardized system of codes. The SWIFT network is one of the most critical infrastructures of the financial sector, as it enables many financial transactions with very high monetary value. Any disruption to the operation of this network can have significant socio-economic implications (including significant financial losses), as evident in a number of recent attacks against it [6].

The operation of SWIFT network is based on a supply chain of relevant stakeholders, while entailing both cyber (e.g., networks, computers) and physical (e.g., SWIFT devices, SWIFT transactions rooms). Therefore, integrated (cyber/physical) approaches to securing the SWIFT network, along with stakeholders' collaboration in the supply chain, can increase its resilience. In-line with the FINSEC RA, probes can be used to collect security information about cyber and physical assets, as a means of identifying risks and non-obvious abnormalities. For example, a FINSEC compliant system can correlate information about attempts for unauthorized access to physical spaces or devices with information about vulnerabilities of the SWIFT ICT infrastructure. In this way, strong protection from insider threats can be provided, along with resilience against combined cyber/physical attacks [e.g., cases of an intruder (or insider) who exploits cyber vulnerabilities in order to give malicious SWIFT commands from the inside]. Likewise, the collaborative modules of the FINSEC RA can enable financial organizations that participate jointly in SWIFT transactions to share threat information and accordingly to use the shared information towards jointly scoring risks associated with SWIFT-related assets.

This scenario deals with the monitoring of events in a SWIFT system, where a SysLog events probe monitors the network traffic, and the upper layers of FINSEC platform analyze the risk level according to the login attempts and the related timestamps.

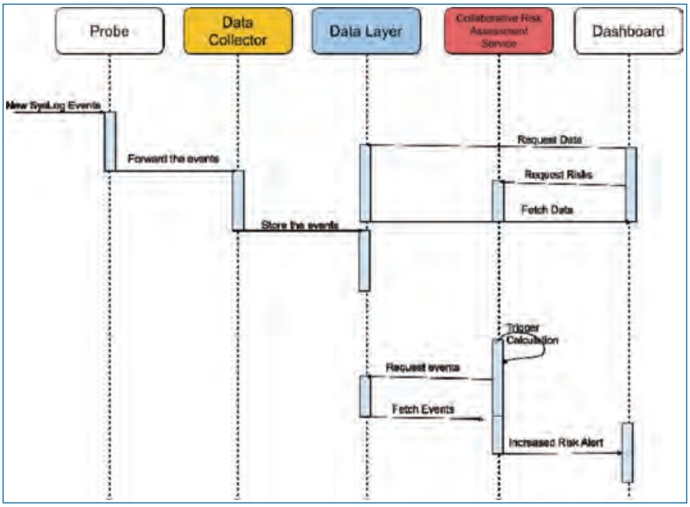


Figure 2.2. Pilot scenarios #1 – #2 data flow.

2.3.2 ATM Network Protection

The network of ATM (Automatic Teller Machines) is another prominent example of financial sector infrastructure that includes both cyber and physical elements. A single ATM includes a PC, a vault, and a printer, which are interconnected. Furthermore, ATM machines are themselves networked via ICT infrastructure. The integrated security capabilities of the FINSEC RA can be exploited in order to correlate cyber and physical security events that can indicate abnormal situations in the use of one or more ATM machines. To this end, appropriate probes like CCTV cameras and sensors that can provide information on the physical status of the ATM's objects are needed. Based on such probes and the analysis of their information, it is possible to extract and correlate a wide array of events and notifications such as a person entering or being the ATM area, detection of a use of a valid card, detection of whether the ATM case is open (e.g., based on vibration sensors), interaction between people in the ATM area (e.g., when two or more people are very close), people fighting, people leaving the ATM, and more. The correlation of such events can enable the detection and timely sharing of CPTI information between relevant security stakeholders such as the security officers of a bank, their IT department, law enforcement agencies, and more.

2.3.3 Regulatory Compliance

FINSEC RA can also enable the development and deployment of solutions that boost Data Privacy Compliance, as a means of boosting financial organizations' compliance with relevant directives and regulations such as the General Data

Protection Regulation (GDPR) for organizations that operate in Europe or collaborate with European financial institutions. This is very important for financial organizations, as they typically handle large amounts of sensitive consumer data. In particular, the FINSEC RA can boost the implementation of SIEM and other probes that:

- Record all events associated with handling of personal data, as a means of providing a complete and reliable audit trail for such data.
- Implementing and deploying advanced analytics algorithms over FINSTIX as a means of quickly detecting data breaches.
- Providing additional analytics tools for analyzing those data breaches and finding their root causes, along with relevant (i.e., responsible or liable) actors.
- Monitoring, logging, and analyzing changes to credentials and security groups, notably groups that handle personal data.
- Auditing and verifying security controls to ensure that user data is treated appropriately and in-line with GDPR principles.

Overall, the FINSEC RA forms a basis for the development of compliance auditing services for all operations that access and/or process private data.

Moreover, the NIS Directive [Directive (EU) 2016/1148] [1] advocates for a well-defined governance process, improved risk management, and management of the overall supply chain. In this direction, the FINSEC RA specifies:

- Audit and certification services to support improved governance.
- Risk assessment, anomaly, and risk detection to better support risk management.
- A supply chain collaboration concept that describes how FINSEC integrates security and risk management across the supply chain. In particular, the Edge Tier and SECaaS services of the architecture, together with the FINSEC Collaborative Module and Dashboard, provide integration/interaction with the supply chain.

In ENISA's work [4] on network and information security in the finance sector indicated also the need for risk transparency for the immediate operational circle in order to better manage the risks posed by the supply chain, which reinforces FINSEC's RA relevant for regulatory compliance.

2.3.4 Insider Threats

As briefly indicated in the scope of the SWIFT network protection use cases, insider threats can be a very big headache for financial organizations as they can be very hard to detect. This is because insiders can appear as legitimate users. Solutions compliant to the FINSEC RA can leverage SIEM-like functionalities in order to detect and understand insider threats based on recording and analysis of insiders' behavior. In practice, this can be implemented as follows:

- Detecting cases where users move across multiple systems within the intranet of the financial organization.
- Identifying cases where users' privileges and authorizations change, thus enabling users to access different systems and possibly gain additional authorizations.
- Detect "strange" and unusual behaviors, such as cases where users access systems during unusual days or times.
- Correlating events that do not have obvious links between them, such as changes in the quota or authorizations of specific groups of users and cybersecurity vulnerabilities of financial infrastructures (e.g., SWIFT/SEPA infrastructures).

2.3.5 IoT Devices Security

The Internet of Things (IoT) paradigm enables financial organizations to leverage data from the real world in designing and delivering their services [e.g., Point of Sales (POS) devices and RFID devices]. These devices add new points of vulnerability, given that the users of these devices may not take appropriate measures for their security. The FINSEC RA can enable the implementation of systems that can enable the security such devices through monitoring and analyzing their data flows, while at the same time activating pentesting and vulnerability assessment functionalities. In particular, the FINSEC RA can enable the implementation of analytics applications that generate alerts whenever unusual flows or patterns of data are detected. Such alerts can be visualized on appropriate dashboards and/or shared with security teams like CERT/CSIRTs.

2.3.6 Managed Security

The FINSEC RA promotes the implementation of security solutions based on modern cloud-based microservices architectures. As such it also provides the means for implementing cloud-based Security as a Service (SECaaS) applications. The latter are very important for financial organizations that lack the financial capacity and/or the technical knowhow to develop, deploy, and operate on-premise solutions. As a prominent example, SMEs (Small Medium Enterprises) dealing with algorithms trading or payments do not typically have organized security departments and teams. Thus, they would rather dispose with a managed security paradigm like SECaaS. The FINSEC RA can enable these organizations to access services like pentesting, risk management, and vulnerability assessments as a service (i.e. through a service provider) as soon as they can providing security data based on appropriate probes. The SECaaS model can provide them with flexibility as well, since they can request and access additional reports around compliance and privacy on demand, i.e., where and when needed. Overall, the FINSEC RA provides the means for implementing a wide range of managed security use cases based on the SECaaS paradigm, as a means of maximizing flexibility and obviating the need for significant capital investments on security infrastructures.

2.4 Conclusions

In order to properly secure the critical infrastructures for the financial sector, there is a need for a new integrated approach that addresses physical and cybersecurity together rather than having them treated by dedicated systems and processes. Likewise, financial organizations should benefit from the capabilities of emerging technologies like Big Data and AI analytics for security monitoring and automation, while at the same time leveraging the flexibility of the DevOps paradigm that provides opportunity for frequent changes to security measures and policies (e.g., patching on a daily basis). In response to these requirements, this Chapter has introduced a Reference Architecture developed in the frame of FINSEC project, as a blueprint for implementing, deploying, and operating integrated (cyber/physical) security systems.

Drawing on an in-depth analysis of the current needs of financial organizations on integrated security that that addresses physical and cybersecurity together rather than having them treated by dedicated systems and processes. The FINSEC RA is a modular architecture that adopts modern principles of microservices architectures and DevOps methodologies. It is a data-driven architecture that relies on the collection, analysis, and sharing of security information, as a means of identifying vulnerabilities and threats, but also as a means of instigating relevant remedial issues and actions.

Based on the FINSEC RA, a wide range of security use cases can be implemented and deployed. We have discussed some sample and very prominent use cases concerning attacks against the SWIFT network, protection of the ATM network, confronting insiders' threats, boosting compliance to GDPR and other data protection regulation, securing IoT devices, as well as implementing managed security based on the SECaaS paradigm.

References

[1] European Parliament and Council. Directive (EU) 2016/1148, measures for a high common level of security of network and information systems across the Union. 2016.

[2] Martin, R., Schrecker, S., Soroush, H., Molina, J., LeBlanc, JP., Hirsch, F., Buchheit, M., Ginter, A., Banavara, H., Eswarahally, S., Raman, K., King, A., Zhang, Q., MacKay, P. and Witten, B. Industrial Internet Security Framework Technical Report. 2016.

[3] Network and Information Security in the Finance Sector Regulatory landscape and Industry priorities. European Union Agency for Network and Information Security (ENISA). 2014.

[4] Secure Use of Cloud Computing in the Finance Sector. European Union Agency for Network and Information Security (ENISA). 2015.

[5] Structured Threat Information Expression (STIX). Accessed: 1 February 2020. Available at: https://oasis-open.github.io/cti-documentation/

[6] Bergin, T. and Layne, N. Special Report: Cyber thieves exploit banks' faith in SWIFT transfer network. Reuters. 2016. Accessed: 1 February 2020. Available at: https://www.reuters.com/article/us-cyber-heist-swift-specialreport-idUSKCN0YB0DD

DOI: 10.1561/9781680836875.ch3

Chapter 3

FINSTIX: A Security Data Model for the Financial Sector

By Giorgia Gazzarata, Ernesto Troiano, Enrico Cambiaso, Ivan Vaccari, Ariana Polyviou, Alessio Merlo and Luca Verderame

The numerous recent cybersecurity incidents in the financial sector prompt for the emerging need to elevate the security of the physical and cyber assets available for financial institutions. To this end, the FINSEC project aims at designing and building a reference architecture (i.e., the FINSEC Architecture) that offers integrated security services designed for financial institutions. This Chapter describes the Security Knowledge Base and the approach of the FINSEC project for sharing the knowledge of Cyber-Physical Threat Intelligence of financial infrastructures, based on a novel modeling language called FINSTIX.

FINSTIX enables the description and storage of cyber-physical security incidents and simplifies their exchange and processing among the different reasoning modules composing the FINSEC Architecture.

3.1 Introduction

In the last few years, the number of cybersecurity incidents against financial institutions is growing. As extensively addressed in the chapter that introduces the security challenges of the financial sector, the growing sophistication of recent technological innovations, the complex processes involving multiple organizations, and the fact that services are becoming more digitized and interconnected raise the cybersecurity risks for financial institutions. Financial institutions address cyber and physical security attacks usually in isolation. This often leads to inaccurate vulnerability assessment and risk analysis and, in general, limited security guarantees.

As a result, financial institutions must increase their robustness and develop integrated approaches for addressing physical and cyberattacks. Reflecting on this need, the FINSEC project aims to design and build a reference architecture for the integration of both physical and cybersecurity threats of financial institutions, thus proposing an integrated framework for predictive and collaborative security.

In particular, as described in the chapter that introduces the security challenges of the financial sector, the FINSEC Architecture defined in the project fosters the interactions between different financial institutions and third parties to enable the discovery and detection of sophisticated cyber-physical threats.

Within the FINSEC Platform, the Security Knowledge Base (SKB) is the cluster of Cyber-physical Threat Intelligence (CPTI) information. The SKB serves all the modules of the architecture with the information needed for their predictive and detection tasks.

FINSTIX is the data model employed by the SKB to represent every piece of information that transits in the FINSEC Architecture. Due to the FINSTIX data model, the SKB is able to represent the relationships among different cyber and physical threats, as well as new attack patterns involving the financial infrastructures discovered by the Service Tier of the FINSEC Platform.

The knowledge is collected from different publicly available threat intelligence sources, including Common Vulnerabilities and Exposures (CVE) databases, MITRE, Common Attack Pattern Enumeration and Classification (CAPEC) patterns, and existing OVAL specifications. Finally, the SKB enables the visualization of all the information on the vulnerabilities affecting the infrastructure assets, thanks to an ad hoc dashboard.

The FINSEC Consortium can boast the FINSTIX data model as a key innovation introduced by the FINSEC project, thereby addressing the current limitations in the mainstream approaches of threat intelligence and data modeling of cyber-physical threats.

In detail, the Chapter is structured as follows:

- Section 3.2 presents the background on Cyber Threat Intelligence, STIX, and the knowledge base technology.
- Section 3.3 introduces the Security Knowledge Base and the FINSTIX Data Model.

- Section 3.4 shows the interaction between the Security Knowledge Base and the FINSEC Dashboard and between the Security Knowledge Base and the Collaborative Risk Management module.
- Section 3.5 concludes the Chapter.

3.2 Background

3.2.1 Cyber Threat Intelligence

Cyber Threat Intelligence (CTI) is a holistic approach to the automated sharing of threat intelligence [1], and it is nowadays considered one of the most promising strategies in the cybersecurity topic [2]. In the CTI context, [3] proposes a classification and distinction among existing threat intelligence types. Similarly, [4] and [5] summarize and compare the most prevalent information-sharing models. Along the same lines, [6] and [7] propose a survey of the current platforms and formats available for threat information sharing. There are different CTI formats available, like OpenIOC [8], Trusted Automated eXchange of Indicator Information (TAXII) [2, 4], Real-Time Internetwork Defense (RID) [4, 9], Incident Object Description Exchange Format (IODEF) [1, 10].

Among the available CTI formats, Structured Threat Information Expression (STIX) [11] is considered the most commonly used CTI standard [12], although CyBOX and TAXII are considered good alternative solutions [6]. STIX provides a modular format that can also efficiently incorporate other standards [1]. STIX is adopted in different contexts of different nature. In this context, [13] adopts STIX as an input format for analyzing data for machine learning algorithms to increase new threat detection ability and responsiveness. Instead, [14] presents an innovative approach to automatically generate Cyber-threat Intelligence data as STIX documents, starting from raw threat data. [15] adopts STIX to share threats and security information in IoT contexts, while [16] and [17] make use of a blockchain-based system to share CTI data using STIX format. [18] proposes an industrial adoption of STIX to exchange information between Integrated Management System (IMS) and Security Information and Event Management systems (SIEM). [19] presents an alternative use of STIX, to describe the actual state of the reference system, instead of exchange attack information. [20] presents a collaborative platform to share cyber threat information using STIX by focusing on anonymity exploitation. [21] makes use of STIX for threat information inputs, combining it with other similar information sources to develop a collaborative cognitive system, able to detect threats by combining different collaborative agents, covering both host and network information. [22] combines STIX concepts with Markov chains ones, for cyber threats modeling, while [23] proposes a cyber threat protection solution based on a Threat

Intelligence Platform (TIP), based on both STIX and TAXII. Instead, [24] proposes MANTIS, a threat intelligence platform that makes use of different standards for threat data correlation, accomplished through a novel similarity algorithm. [25] proposes CyTIME, a framework that integrates CTI data like STIX under a global JSON format and automatically generates network security rules from the incorporated data seamlessly. Another innovative framework is proposed in [26], making use of STIX to exchange information about detected incidents, generated alerts, and applied mitigations. [27] introduces STIXGEN, a framework based on STIX able to generate error-free structured data.

Although it is widely used, the STIX format presents different limitations: [28] analyzes STIX by detailing the advantages and limitations of the format. Indeed, STIX is considered very complex to implement [4] and lacks support to reasoning [29]. In virtue of this, different extensions of STIX are proposed. For instance, UCO (Unified Cybersecurity Ontology) is a semantic-based alternative of STIX [29]. Also, [30] proposes some extensions of STIX, while [31] extends it to support the inclusion of relevant attack details on sophisticated attacks through the description of complex patterns. Similarly, while [32] extends STIX to support network and security events, [33] proposes a STIX extension to integrate and support additional cyber threats. Such extension is used in ChainSmith, a system able to extract Indicators of Compromise (IoC) by analyzing technical articles and industry reports.

The proposed work represents an extension of STIX in the fintech context. FINSTIX includes both cyber and physical security threats and enables the description of organization assets and how they are interconnected.

3.2.2 Structured Threat Information eXpression (STIX)

The OASIS Structured Threat Information eXpression (STIX) standard [11] is a language and serialization format that facilitates cyber threat intelligence (CTI) [34]. STIX has been designed with a focus on four different use cases [35] that include: (I) Analyzing Cyber Threats; (II) Specifying Indicator Patterns for Cyber Threats; (III) Managing Cyber Threat Response Activities; (III) Sharing Cyber Threat Information.

Although STIX was initially designed with the trademark of the MITRE Corporation, aiming to foster both the development of STIX and promote its adoption, it has been transitioned to OASIS. STIX defines two different types of objects, namely the STIX Domain Objects (SDOs) and the STIX Relationship Objects (SROs).

SDOs represent the concepts commonly used in CTI. The SROs represent the relationships between the SDOs. SDOs and SROs are listed and briefly described in Tables 3.1 and 3.2, respectively.

Table 3.1. STIX Domain Objects (SDOs) [11].

Object	Name	Description
Attack Pattern	Attack Pattern	A type of Tactics, Techniques, and Procedures (TTP) that describes ways threat actors attempt to compromise targets.
Campaign	Campaign	A grouping of adversarial behaviors that describes a set of malicious activities or attacks that occur over a period of time against a specific set of targets.
Course of Action	Course of Action	An action taken to either prevent an attack or respond to an attack.
Identity	Identity	Individuals, organizations, or groups, as well as classes of individuals, organizations, or groups.
Indicator	Indicator	Contains a pattern that can be used to detect suspicious or malicious cyber activity.
Intrusion Set	Intrusion Set	A grouped set of adversarial behaviors and resources with common properties believed to be orchestrated by a single threat actor.
Malware	Malware	A type of TTP, also known as malicious code and malicious software, used to compromise the confidentiality, integrity, or availability of a victim's data or system.
Observed Data	Observed Data	Conveys information observed on a system or network (e.g., an IP address).
Report	Report	Collections of threat intelligence focused on one or more topics, such as a description of a threat actor, malware, or attack technique, including contextual details.
Threat Actor	Threat Actor	Individuals, groups, or organizations believed to be operating with malicious intent.
Tool	Tool	Legitimate software that can be used by threat actors to perform attacks.
Vulnerability	Vulnerability	A mistake in software that can be directly used by a hacker to gain access to a system or network.

Table 3.2. STIX Relationship Objects (SROs) [11].

Object	Name	Description
Relationship	Relationship	Used to link two SDOs and to describe how they are related to each other.
Sighting	Sighting	Denotes the belief that an element of CTI was seen (e.g., indicator, malware).

Analyzing SDOs and SROs from a graphical representation viewpoint, the SDOs can be considered as the nodes and the SROs can be considered as the edges of a directed graph, as shown in the examples in [36].

In STIX 2, SDOs and SROs are represented in JSON as described in the specification of STIX [37]. The STIX standard enables customization in two different ways: Custom Properties and Custom Objects. The first regards the addition of properties not defined by the specification of existing SDOs, while the second one regards the creation of brand-new objects. Regardless of the means used to customize STIX, certain requirements need to be fulfilled in order to ensure the compatibility with the standard, as described in [38].

3.2.3 Knowledge Base Technology

Knowledge base technology is used to collect, organize, share, and retrieve complex structured and unstructured information representing facts and assertions about the world. Usually, the content of a knowledge base comes from several contributors who are well versed in the subject. Differently from a simple database, a knowledge base does not consist only of tables with numbers, strings, dates, etc., but also contains objects with pointers to other objects that, in turn, have additional pointers [39]. The ideal representation for a knowledge base is an object model, which is often called ontology.

There are two major types of knowledge bases: human-readable and machine-readable. A human-readable knowledge base enables the users to access and use the knowledge. Its content can consist of documents, manuals, troubleshooting information, and frequently answered questions. It can be interactive and can lead the users to the solutions to their problems, relying on the information provided by expert users to guide the process. A human-readable knowledge base is contrasted with a machine-readable knowledge base, which stores knowledge in system-readable forms and is limited in interactivity [40].

3.3 The FINSEC Security Knowledge Base

The FINSEC Security Knowledge Base (SKB) is one of the modules contained in the Data Tier of the FINSEC Reference Architecture. The Security Knowledge Base aims to collect information coming from different sources of Cyber Threat Intelligence. In particular, the added value of the Security Knowledge Base compared to the existing ones, is the definition of the relationships between different assets and their interactions as part of the critical infrastructures of the financial sector. This feature will enable the services of the Service Tier to consume the information contained in the Knowledge Base for producing new Cyber and Physical Threat Intelligence, namely new data about vulnerabilities, threats, risks, and attacks affecting the cyber and physical infrastructure of the financial organizations. Obviously, the Service Tier will feed the Security Knowledge Base with this new information.

The (SKB) is a mixture of human and machine-readable knowledge base. On the one hand, the knowledge should be consumed by the FINSEC Platform services to produce new knowledge. On the other hand, the content of the knowledge base should be consultable by users through a visual interface. To suit the FINSEC needs, the content of the Security Knowledge Base should satisfy two essential requirements:

(1) It should be structured in order to enable automatic processing;
(2) It should include information on the infrastructure and the organization assets, for enabling the FINSEC Platform to perform Cyber and Physical Threat Intelligence.

Below, Sections 3.3.1 and 3.3.2 explain the used ontology and the architecture of the FINSEC Security Knowledge Base, respectively. Instead, Section 3.3.3 presents a simple visual interface that allows to visualize the content of the Security Knowledge Base.

3.3.1 The Security Knowledge Base Data Model

To realize a knowledge base, it is essential to design the appropriate data model, which is the format used to represent the information contained in the knowledge base. In the design of a data model, one approach is to define a completely new set of objects coping with the business requirements of the considered use cases. However, this approach incorporates the risk of missing other relevant cases. Alternatively, another approach is to employ an existing standard (or a mix of standards) and then extend it such that missing components can be added. In our approach, we adopted the latter option and confronted it with the business cases of the project to assure consistency. Among the existing CTI formats, STIX is considered the most commonly used CTI standard [2]. Due to its modularity, it can also easily incorporate other standards [1]. Also, STIX enables easy customization and extension. For these reasons, the STIX was employed as the basis of the FINSEC Data Model. The result is the FINSTIX Data Model, which extends STIX 2.0 combining information coming from both physical and logical worlds. To this end, FINSTIX is the data model used not only by the FINSEC Security Knowledge Base, but also by the probes and all the other modules of the FINSEC Platform to communicate with each other. In particular, the design of

FINSTIX accounted for the needs of each module. To the best of our knowledge, FINSTIX is the first data model integrating CTI with the physical world, and thus, this could be considered as an innovative contribution of this book Chapter.

3.3.1.1 From STIX to FINSTIX

Due to its expressiveness, flexibility and extensibility, STIX can be considered as one of the most famous industrial standards used to represent and share CTI. However, it encapsulates two weaknesses. First, it does not provide an accurate representation of the financial institution's infrastructure. Second, it does not envisage physical systems, but it is rather limited to the cyber ones. Due to these limitations, STIX was extended.

There are two possible kinds of extensions of the STIX standard: the first regards the definition of custom parameters into STIX Domain Objects already defined by the standard itself; the second consists in the definition of brand-new custom objects. Within the FINSEC project, both approaches have been used. "STIX^{TM} Version 2.0. Part 1: STIX Core Concepts" [22] contains the rules to follow to extend STIX preserving compliance with the standard.

The resulting data model has been named FINSTIX (from FINSEC-STIX), which includes all the domain objects defined in STIX. The objects included in the FINSTIX Data Model take the name of FINSTIX Domain Objects (FDOs). Every FDO is a collection of key-value pairs that represent data or relations in JSON format. The FINSEC extension to STIX has been driven by the FINSEC project use cases, which led to the definition of the custom objects listed and briefly described in Table 3.3.

All the custom objects introduced in FINSTIX contain the following mandatory elements:

- *type*: type of FDO.
- *id*: univocal identifier of the FDO. It must have the form "<type>--uuid", where <type> is the FDO type.
- *name*: name of the FDO.
- *description*: description of the FDO.
- *subtype*: subtype of the FDO. Its value depends on the FDO type.
- *domain*: it can be "Cyber" or "Physical." It is used to distinguish between cyber or physical domain.
- *datatype*: it can be "Model" or "Instance." It is used to distinguish a model from an object created at run-time.
- *x_organization*: id of the referenced organization. This key is used for multi-tenant applications, in other words, to protect organizations from disclosure

Table 3.3. FINSTIX Domain Objects (FDOs).

Name	Description
Organization	Financial organization.
Asset	Organizations' valuable infrastructure. PCs, server rooms, ATMs, applications, and everything inside an organization that is crucial.
Area of Interest	Logical/physical area, for example, an indoor area (server room).
Service	A collection of assets forming a publicly exposed service, for example, a web application.
Probe	Object used to support monitoring infrastructure. A Probe usually monitors one or more areas of interest.
Probe Configuration	Data sent to a probe to configure details such as the area under monitoring or the bit rate of the monitoring process.
Event	Information on something happened/happening.
Collected data	A group of observed data collected by the network probe.
Agent	Person involved in the events created by the probes.
Risk	The calculated risk for a specific asset or service. The upper levels of FINSEC calculate it in real time.
Risk Configuration	Parameter specification to optimize the risk assessment process. It defines the triggers and other useful options.
Regulation	An object used to depict a regulation violation. FINSEC must deal with this kind of issue, even if different from attacks. The regulation violation information will be sent to regulatory authorities and other organizations.
Vulnerability score	Rating used to provide a score to a vulnerability.
Cyber-Physical Threat Intelligence	Data set fed and enriched by threat information as soon as they are gathered from the probes and processed by the Predictive Analytics module. One or more CPTI objects are used to generate the output of the intelligence process, which is a report about ongoing or possible future attacks on one or more assets belonging to the infrastructure.

of their data. x_organization is also a custom property in the FDOs already defined in STIX.

- *reference*: id of the referenced object. It usually refers to the object above in the hierarchy. It is used to create trees of objects.

datatype is an important and mandatory field, whose value can be "Model" or "Instance." In the first case, the objects represent a model that can be used as a basis

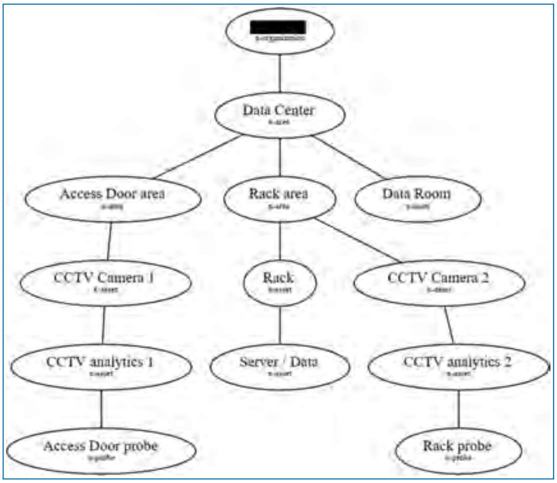


Figure 3.1. Hierarchical representation of an organization infrastructure.

for the analytics to recognize future malicious events; in the second case, instead, the objects are generated by the probes at run-time. Because of this distinction, it becomes clear that:

- The analytics core services match the events with known models and find correspondences (malicious events/attacks), meaning that from the events they produce threat intelligence instances.
- The predictive analytics module produces new models as a result of events analysis.

Every custom object presents a reference to a parent object (through the key *reference*), which enables to create a hierarchy of objects. For example, Figure 3.1 shows a graph representing an organization and its infrastructure. The graph has been created automatically starting from test data that have been pushed into the Data Tier in FINSTIX format.

Events and observed data produced by the organization probes are pushed into the FINSEC Platform, which correlates and aggregates information also gathered from asset models and external Cyber Threat Intelligence through the machine learning analytics and prediction algorithms. The result of this process is the Cyber-Physical Threat Intelligence (CPTI), which integrates important information coming from both the cyber and the physical world. The CPTI produced in the FinTech sector is the added-value information produced by the FINSEC platform that could be exchanged (in-out) between financial organizations and security organizations (CERT/CSIRT like). The integration between cyber and physical security aspects introduced by FINSTIX is an innovation attributable to the FINSEC Project.

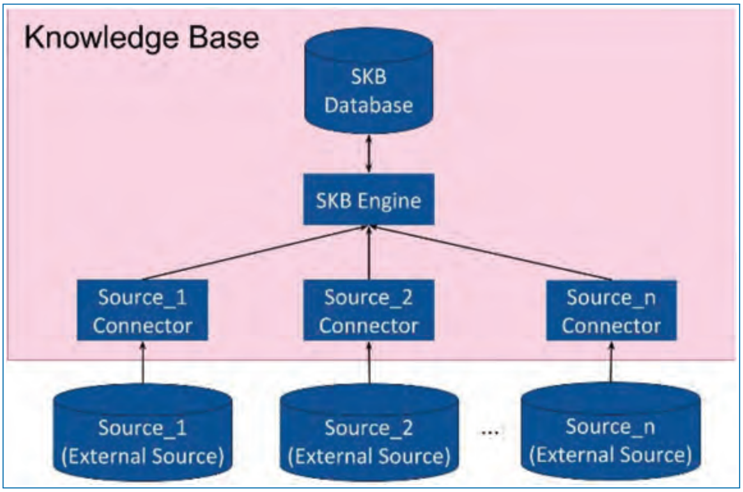


Figure 3.2. The Security Knowledge Base architecture.

3.3.2 The Security Knowledge Base Architecture

This section presents the architecture of the FINSEC Security Knowledge Base, which is depicted in Figure 3.2. The SKB consists of:

- The SKB Database, which actually stores the knowledge.
- The SKB Engine, which manages the operations on the database. It exposes REST API to interact with the other modules.
- The SKB Connectors (one for each external source), which translate the information coming from external threat intelligence sources into the data model to promote homogeneity and integrity among the FINSEC services.

The content of the FINSEC Security Knowledge Base is stored into an instance of MongoDB. As discussed in Section 3.3.1, the information is stored as FINSTIX Domain Objects to enable semantic interoperability among all the modules in the platform and between the FINSEC Platform and the external modules (such as the FINSEC Probes and the FINSEC Dashboard).

The Security Knowledge Base Engine is a microservice developed in Python-Flask. It is a simple module handling the operations on the database. Its functionalities are exposed through a REST API. The principal methods are described hereafter:

- The methods insertkb and insertkb_many enable to insert, respectively, one or more FINSTIX Domain Objects into the Knowledge Base. The input comes from the SKB Connectors or from the Service Tier.

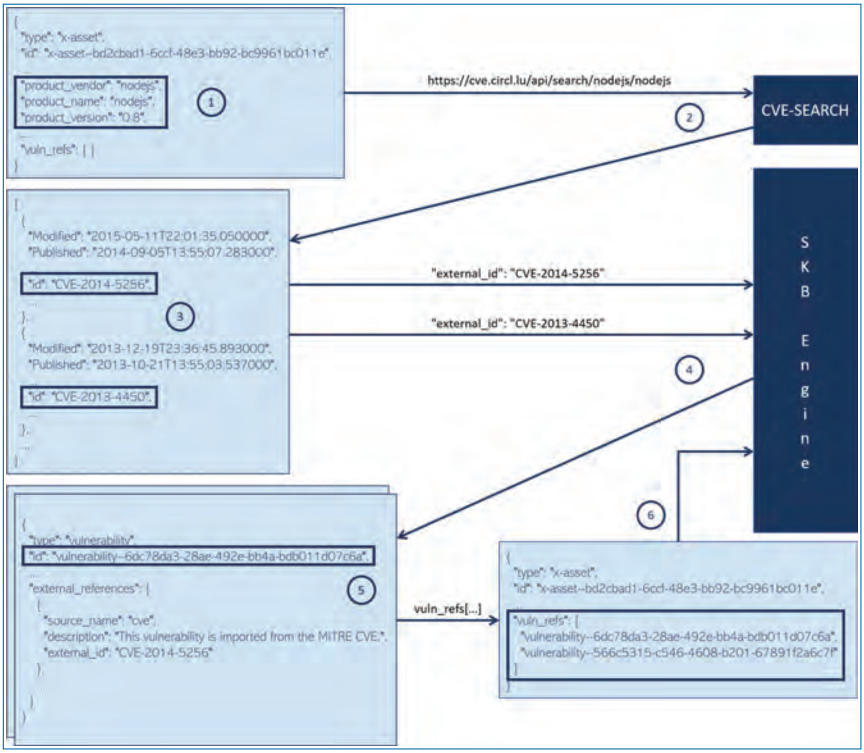
- The methods getkb and retrievekb enable to retrieve, respectively, one or more FINSTIX Domain Objects.

At the time of writing, the FINSEC Security Knowledge Base contains FDOs coming from two different external sources of threat intelligence: MITRE Adversarial Tactics, Techniques, and Common Knowledge (ATT&CK) and MITRE Common Vulnerabilities and Exposures (CVE). In the future, other sources of CTI will be imported.

Those "other sources" may include cooperation with financial-sector stakeholders not currently involved in FINSEC, to include their experiences of interoperability issues, both across the multiple security knowledge bases described above, and more widely (e.g., potentially including metadata and ontologies used in joint actions by police and anti-fraud agencies to produce and share their intelligence analysis reports). An EU-wide example is the collaboration of Eurojust and Europol, which have set up a Joint Investigation Team (JIT) to counter organized international criminal attacks on cyber and physical weak points in banking organizations across Europe. The JIT used Europol's Malware Analysis System to collate and interpret threat intelligence, including terabytes of evidence of banking sector frauds, which led to the conviction of dozens of skilled criminals who had widely shared their methods of defeating the security measures introduced by banks. The JIT's use of a single system side-stepped the risk, with multiple systems of incompatibilities in methods of collecting and appraising data. Contrast with the recent reports of such problems in the UK police system for citizens to report possible large-scale cybercrime, the UK lost many citizen reports and had to process many others by hand (e.g., [41]).

Since the semantics used to represent the information depends on the source of CTI, the knowledge base needs a connector for each external source of information. Each connector is responsible for the translation of the information coming from the source into the FINSTIX Data Model. Since ATT&CK content is provided in STIX, it does not need any transformations to comply with FINSTIX. On the contrary, the CVE Connector has to translate the vulnerability imported from CVE in FINSTIX: for each CVE vulnerability, it creates the vulnerability and the vulnerability score FDOs, then, as shown in Figure 3.3, it takes different steps to create the relationships among the vulnerability and the affected assets.

After the translation to FINSTIX, the connectors interact with the SKB Engine through the API to insert the new knowledge into the SKB Database. At the time of writing, the connectors do not implement any protocol to collect CTI; however, standards such as the OASIS Trusted Automated Exchange of Intelligence Information (TAXII) will be considered in the second phase of the FINSEC project [42].



```
{
 "type": "x-asset",
 "id": "x-asset--bd2cbad1-6ccf-48e3-bb92-bc9961bc011e",
 ...
 "product_vendor": "nodejs",
 "product_name": "nodejs",
 "product_version": "0.8",
 ...
 "vuln_refs": [ ]
}
```

```
[
 {
  "Modified": "2015-05-11T22:01:35.050000",
  "Published": "2014-09-05T13:55:07.283000",
  ...
  "id": "CVE-2014-5256",
  ...
 },
 {
  "Modified": "2013-12-19T23:36:45.893000",
  "Published": "2013-10-21T13:55:03.537000",
  ...
  "id": "CVE-2013-4450",
  ...
 },
 ...
]
```

```
{
 "type": "vulnerability",
 "id": "vulnerability--6dc78da3-28ae-492e-bb4a-bdb011d07c6a",
 ...
 "external_references": [
  {
   "source_name": "cve",
   "description": "This vulnerability is imported from the MITRE CVE.",
   "external_id": "CVE-2014-5256"
  }
  ...
 ]
}
```

```
{
 "type": "x-asset",
 "id": "x-asset--bd2cbad1-6ccf-48e3-bb92-bc9961bc011e",
 ...
 "vuln_refs": [
  "vulnerability--6dc78da3-28ae-492e-bb4a-bdb011d07c6a",
  "vulnerability--566c5315-c546-4608-b201-67891f2a6c7f"
 ]
}
```

1
https://cve.circl.lu/api/search/nodejs/nodejs
2
CVE-SEARCH
3
"external_id": "CVE-2014-5256"
"external_id": "CVE-2013-4450"
4
SKB Engine
5
vuln_refs[...]
6

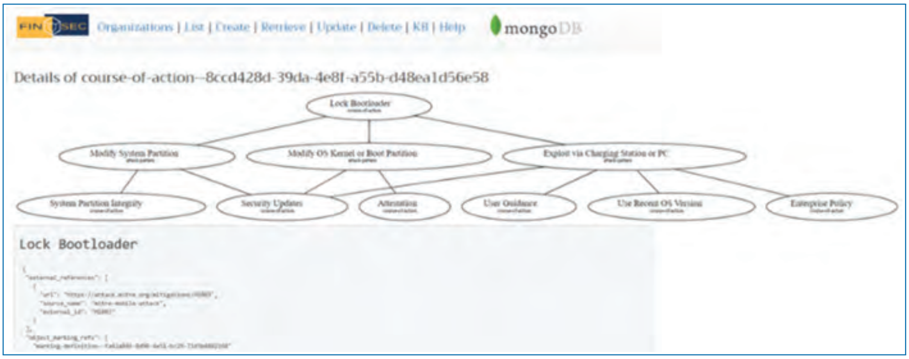
Figure 3.3. CVE connector.

3.3.3 The Security Knowledge Base Visual Interface

Other than the API, the content of the Security Knowledge Base is browsable through a primary visual interface, which is available in the KB page of the Data Layer application [43]. In this first page, the user can see a list of all the FINSTIX objects contained in the Knowledge Base. Clicking on the id of a specific object, the user can see further details of the object itself, as shown in Figure 3.4. In particular, the user can visualize a graph showing the relationships between the selected FDO and the other FDOs. The FINSTIX object data fields are then displayed under the graph. In the future, the visual interface will be improved, for example, supporting a sophisticated search of the FDOs through filters.

3.4 Interaction with the Dashboard and the Collaborative Risk Management Module

The FINSEC Dashboard allows a user to access some pieces of information related to the organization's assets and retrieved by the Security Knowledge Base.


FIN SEC
Organizations | List | Create | Retrieve | Update | Delete | KB | Help
mongoDB
Lock Bootloader
Modify System Partition
Modify OS Kernel or Boot Partition
Exploit via Charging Station or PC
System Partition Integrity
Security Updates
Attestation
User Guidance
Use Recent OS Version
Enterprise Policy

Details of course-of-action--8ccd428d-39da-4e8f-a55b-d48ea1d56e58

Lock Bootloader

```
{
 "external_references": [
  {
   "url": "https://attack.mitre.org/mitigations/M1003",
   "source_name": "mitre-mobile-attack",
   "external_id": "M1003"
  }
 ],
 "object_marking_refs": [
  "marking-definition--fa42a846-8d90-4e51-bc29-71d5b4802168"
```

Figure 3.4. Relational graph and object details.

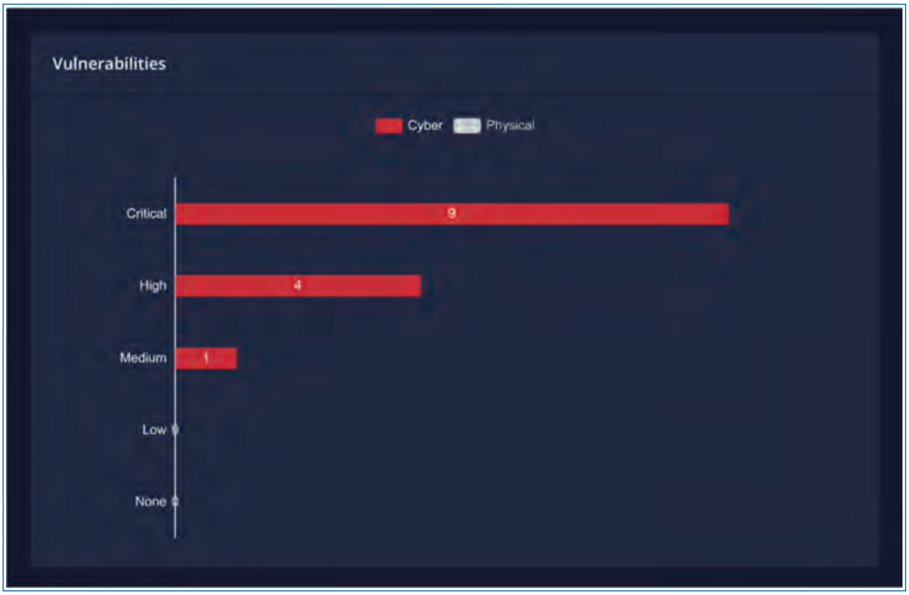
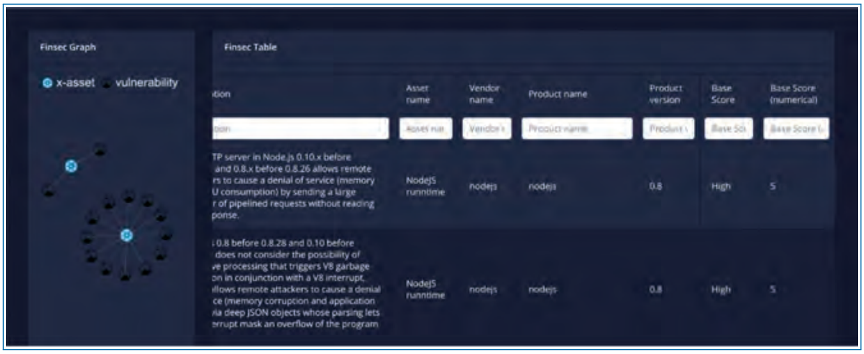


Figure 3.5. Criticality of the vulnerabilities.

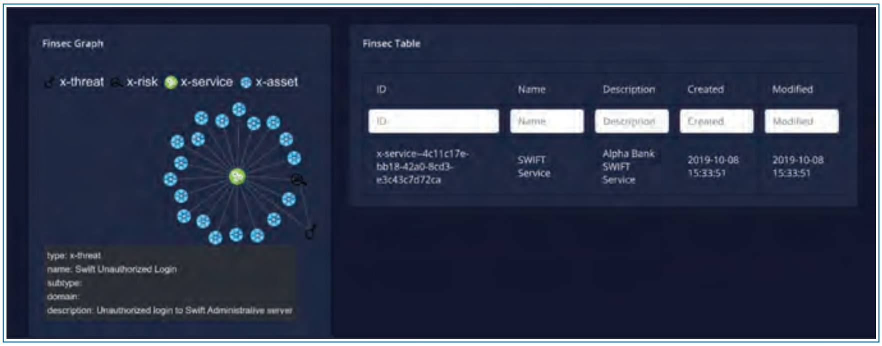
For example, in the main page, the user can see the graph in Figure 3.5, which enables to get a grasp on the criticality of the vulnerabilities affecting the assets of the infrastructure for the organization.

Figure 3.6 illustrates part of the Vulnerability page. From left to right, the user can see two graphs representing two assets with the affecting vulnerabilities; on the right, a table shows details on the vulnerabilities affecting the assets. In particular, each row presents the description of the vulnerability, vendor, product name and version of the affected asset, and the vulnerability score.



tion	Asset name	Vendor name	Product name	Product version	Base Score	Base Score (numerical)
TP server in Node.js 0.10.x before and 0.8.x before 0.8.26 allows remote rs to cause a denial of service (memory U consumption) by sending a large r of pipelined requests without reading ponse.	NodeJS runntime	nodejs	nodejs	0.8	High	5
0.8 before 0.8.28 and 0.10 before does not consider the possibility of ve processing that triggers V8 garbage on in conjunction with a V8 interrupt, llows remote attackers to cause a denial ce (memory corruption and application ia deep JSON objects whose parsing lets errupt mask an overflow of the program	NodeJS runntime	nodejs	nodejs	0.8	High	5

Figure 3.6. Vulnerabilities affecting the infrastructure.



type: x-threat
name: Swift Unauthorized Login
subtype:
domain:
description: Unauthorized login to Swift Administrative server

ID	Name	Description	Created	Modified
x-service--4c11c17e-bb18-42a0-8cd3-e3c43c7d72ca	SWIFT Service	Alpha Bank SWIFT Service	2019-10-08 15:33:51	2019-10-08 15:33:51

Figure 3.7. Risk associated to the service.

In the MVP, the information contained in the Security Knowledge Base is consumed by the Collaborative Risk Management module to calculate the risk associated with the infrastructure services. The Collaborative Risk Management module:

1. Retrieves the vulnerabilities and the related scores affecting the assets that compose the service.
2. Calculates the individual asset risk for each asset composing the service, based on the affecting vulnerabilities, the impact and the threat level for the asset itself.
3. Calculates the service risk starting from the assets' individual risks.

The user can see information on the organization services in the Service page of the Dashboard. In particular, Figure 3.7 shows a graph representing a service and the assets that compose it. As shown in the graph, there is a risk associated to the service, which is calculated by the Collaborative Risk Management module.

3.5 Conclusion

This Chapter presented the FINSEC Security Knowledge Base (SKB), a cluster of CPTI information. The SKB serves all the modules of the architecture with the information needed for their predictive and detection tasks. The Chapter also introduced FINSTIX, which is the data model used not only in the SKB, but also in the entire FINSEC Platform to enable the interaction among the different modules of the platform and between the platform and external modules (such as the FINSEC Probes and the FINSEC Dashboard). FINSTIX extends the STIX 2.0 standard by including custom objects and parameters tailored to the financial organization. To the best of our knowledge, FINSTIX is the first data model that bridges the cyber world and the physical world, thus enabling the extension from Cyber Threat Intelligence to Cyber-physical Threat Intelligence.

References

[1] E. W. Burger, M. D. Goodman, P. Kampanakis and K. A. Zhu, "Taxonomy model for cyber threat intelligence information exchange technologies," in *ACM Workshop on Information Sharing & Collaborative Security*, 2014.

[2] N. Gong, "Barriers to Adopting Interoperability Standards for Cyber Threat Intelligence Sharing: An Exploratory Study," in *Science and Information Conference*, 2018.

[3] W. Tounsi and H. Rais, "A survey on technical threat intelligence in the age of sophisticated cyber attacks," *Computers & Security*, vol. 72, pp. 212–233, 2018.

[4] P. Kampanakis, "Security automation and threat information-sharing options," *IEEE Security & Privacy*, vol. 12, no. 5, pp. 42–51, 2014.

[5] M. Liu, Z. Xue, X. He and J. Chen, "Cyberthreat-Intelligence Information Sharing: Enhancing Collaborative Security," *IEEE Consumer Electronics Magazine*, vol. 8, no. 3, pp. 17–22, 2019.

[6] M. Lutf, "Threat Intelligence Sharing: A Survey."

[7] A. Rattan, N. Kaur, S. Chamotra and S. Bhushan, "Attack Data Usability and Challenges in its Capturing and Sharing."

[8] D. Jaeger, M. Ussath, F. Cheng and C. Meinel, "Multi-step attack pattern detection on normalized event logs," in *2015 IEEE 2nd International Conference on Cyber Security and Cloud Computing*, 2015.

[9] H. Hazeyama, Y. Kadobayashi, D. Miyamoto and M. Oe, "An autonomous architecture for inter-domain traceback across the borders of network operation," in *11th IEEE Symposium on Computers and Communications (ISCC'06)*, 2006.

[10] F. Martinelli, O. Osliak and A. Saracino, "Towards General Scheme for Data Sharing Agreements Empowering Privacy-Preserving Data Analysis of Structured CTI," *Computer Security*, pp. 192–212, 2018.

[11] M. STIX. [Online]. Available: https://oasis-open.github.io/cti-documentation/stix/intro.

[12] D. Shackleford, "Who's using Cyberthreat Intelligence and how?," *SANS Institute*, 2015.

[13] K. Kim, J. H. An and J. Yoo, "A design of IL-CyTIS for automated cyber threat detection," in *2018 International Conference on Information Networking (ICOIN)*, 2018.

[14] F. Sadique, S. Cheung, I. Vakilinia, S. Badsha and S. Sengupta, "Automated structured threat information expression (STIX) document generation with privacy preservation," in *2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON) (IEEE UEMCON 2018)*, 2018.

[15] E. Ko, T. Kim and H. Kim, "Management platform of threats information in IoT environment," *Journal of Ambient Intelligence and Humanized Computing*, vol. 9, no. 4, pp. 1167–1176, 2018.

[16] J. Li and Z. Xue, "Distributed Threat Intelligence Sharing System: A New Sight of P2P Botnet Detection," in *2019 2nd International Conference on Computer Applications & Information Security (ICCAIS)*, 2019.

[17] V. Chia, P. Hartel, Q. Hum, S. Ma, G. Piliouras, D. Reijsbergen, M. Van Staalduinen and P. Szalachowski, "Rethinking blockchain security: Position paper," in *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, 2018.

[18] S. Abe, Y. Uchida, M. Hori, Y. Hiraoka and S. Horata, "Cyber Threat Information Sharing System for Industrial Control System (ICS)," in *2018 57th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, 2018.

[19] L. Leichtnam, E. Totel, N. Prigent and L. Mè, "STARLORD: Linked security data exploration in a 3D graph," in *2017 IEEE Symposium on Visualization for Cyber Security (VizSec)*, 2017.

[20] T. D. Wagner, E. Palomar, K. Mahbub and A. E. Abdallah, "Towards an Anonymity Supported Platform for Shared Cyber Threat Intelligence," in *International Conference on Risks and Security of Internet and Systems*, 2017.

[21] S. N. Narayanan, A. Ganesan, K. Joshi, T. Oates, A. Joshi and T. Finin, "Early detection of cybersecurity threats using collaborative cognition," in *2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC)*, 2018.

[22] R. Gore, J. Padilla and S. Diallo, “Markov Chain modeling of cyber threats,” *The Journal of Defense Modeling and Simulation*, vol. 14, no. 3, pp. 233–244, 2017.

[23] R. J. Ginn and I. Ionescu, “Cyber Threat Analysis,” 2017.

[24] H. Gascon, B. Grobauer, T. Schreck, L. Rist, D. Arp and K. Rieck, “Mining attributed graphs for threat intelligence,” in *Proceedings of the Seventh ACM on Conference on Data and Application Security and Privacy*, 2017.

[25] E. Kim, K. Kim, D. Shin, B. Jin and H. Kim, “CyTIME: Cyber Threat Intelligence ManagEment framework for automatically generating security rules,” in *Proceedings of the 13th International Conference on Future Internet Technologies*, 2018.

[26] A. M. Zarca, J. B. Bernabe, R. Trapero, D. Rivera, J. Villalobos, A. Skarmeta, S. Bianchi, A. Zafeiropoulos and P. Gouvas, “Security Management Architecture for NFV/SDN-aware IoT Systems,” *IEEE Internet of Things Journal*, 2019.

[27] Z. Iqbal, Z. Anwar and R. Mumtaz, “STIXGEN-A Novel Framework for Automatic Generation of Structured Cyber Threat Information,” in *2018 International Conference on Frontiers of Information Technology (FIT)*, 2018.

[28] A. Aviad and K. Wkecel, “Cyber Treat Intelligence Modeling,” in *International Conference on Business Information Systems*, 2019.

[29] Z. Syed, A. Padia, T. Finin, L. Mathews and A. Joshi, “UCO: A unified cybersecurity ontology,” in *Workshops at the Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[30] F. Fransen, A. Smulders and R. Kerkdijk, “Cyber security information exchange to gain insight into the effects of cyber threats and incidents,” *e & i Elektrotechnik und Informationstechnik*, vol. 132, no. 2, pp. 106–112, 2015.

[31] M. Ussath, D. Jaeger, F. Cheng and C. Meinel, “Pushing the limits of cyber threat intelligence: extending STIX to support complex patterns,” *Information Technology: New Generations*, pp. 213–225, 2016.

[32] M. Steinke and W. Hommel, “A data model for federated network and security management information exchange in inter-organizational it service infrastructures,” in *NOMS 2018-2018 IEEE/IFIP Network Operations and Management Symposium*, 2018.

[33] Z. Zhu and T. Dumitras, “Chainsmith: Automatically learning the semantics of malicious campaigns by mining threat intelligence reports,” in *2018 IEEE European Symposium on Security and Privacy (EuroS&P)*, 2018.

[34] S. Barnum, “Standardizing cyber threat intelligence information with the Structured Threat Information eXpression (STIX),” *Mitre Corporation*, vol. 11, pp. 1–22, 2012.

[35] “STIX Use Cases,” [Online]. Available: http://stixproject.github.io/usecases/.

[36] "STIX examples," [Online]. Available: https://oasis-open.github.io/cti-documentation/stix/examples.

[37] "STIX Version 2.0. Part 2: STIX Objects," [Online]. Available: http://docs.oasis-open.org/cti/stix/v2.0/stix-v2.0-part2-stix-objects.html.

[38] "STIX Version 2.0: STIX Core Concepts," [Online]. Available: http://docs.oasis-open.org/cti/stix/v2.0/stix-v2.0-part1-stix-core.html.

[39] "Wikipedia," [Online]. Available: https://en.wikipedia.org/wiki/Knowledge_base.

[40] "ATALASSIAN," [Online]. Available: https://www.atlassian.com/itsm/knowledge-management/what-is-a-knowledge-base.

[41] "The Guardian," [Online]. Available: https://www.theguardian.com/uk-news/2019/oct/24/police-database-flagged-9000-cybercrime-reports-as-security-risk?CMP=Share_iOSApp_Other.

[42] "OASIS TAXII," [Online]. Available: https://oasis-open.github.io/cti-documentation/taxii/intro.

[43] "Data layer," [Online]. Available: https://data-layer.dev.finsec-project.eu/dashboard/listkb.

[44] "STIX scenario," [Online]. Available: https://oasis-open.github.io/cti-documentation/examples/threat-actor-leveraging-attack-patterns-and-malware.

DOI: 10.1561/9781680836875.ch4

Chapter 4

Artificial Intelligence Gateway for Cyber-physical Security in Critical Infrastructure and Finance

By Marian Ghenescu, Serban Carata, Roxana Mihaescu and Sabin Floares

FINSEC Cyber-physical Security Gateway is an intelligent system that takes over information from different security systems installed within a site or a perimeter (intrusion detection, access control, fire detection, technical alarms, etc.) and, using the data extracted from the images provided by the video surveillance system, identifies threats and events before they become critical. The module is used to analyze security events and situations, allows the integration of cybersecurity concepts, as well as the monitoring of risks associated with critical installations for business processes: technical alarms, maintenance, internal procedures, and even commercial applications. Once the events are documented, they can be transformed into alarms on FINSEC Platform via FINSTIX Data Model or on other PSIM (Physical Security Information Management) and/or VMS (Video Management System) platforms, as the information may represent the best support in making decisions regarding threats either on-site or remotely.

4.1 Overview

FINSEC Cyber-physical Security Gateway is an intelligent system that takes over information from different security systems installed within a site or a perimeter (intrusion detection, access control, fire detection, technical alarms, etc.) and, using the data extracted from the images provided by the video surveillance system, identifies threats and events before they become critical. The module is used to analyze security events and situations, allows the integration of cybersecurity concepts, as well as the monitoring of risks associated with critical installations for business processes: technical alarms, maintenance, internal procedures, and even commercial applications. Once the events are documented, they can be transformed into alarms on FINSEC Platform via FINSTIX Data Model or on other PSIM (Physical Security Information Management) and/or VMS (Video Management System) platforms, as the information may represent the best support in making decisions regarding threats either on-site or remotely.

The system is designed to analyze security events and situations, allows the integration of cybersecurity concepts, as well as the monitoring of risks associated with critical installations for business processes: technical alarms, maintenance, internal procedures, and even commercial applications. Such system integrates specific functions that configure an internal security layer used to prevent cyberattacks on the connected systems. The entire hardware and application-based architecture allows different standard and protocols for sensors as well as different data models for uplink connection assuring, by design, the compatibility and integration with systems that use modern IoT (Internet of Things)-type technologies.

The overall system is based on a multi-sensor and multiple-technology detection, where each subsystem can be a standalone one but the processed information (physical events, cyber detections, associated data, other relevant sensing) is conveyed to the Cyber-physical Security System for local data fusion. The main objectives of the Cyber-physical Security System are:

- Properly detect, collect, and locally fuse the information from the local detection to achieve high detection probability;
- Verify the detection events so as to reduce the false alarm rate;
- Provide relevant information to the global data fusion and Global Platform;
- Include a mitigate layer for reactions to central commands, including predictive actions based on global knowledge and threat risk level.

4.2 Classical vs Smart Security Controllers

The appearance of the concept named “Cyber-physical Systems” (CPS) was determined in the first place by the need for evolution of the computer science and technology. The CPSs perform an embedding process of all their components and offer a close interaction between the cyber and physical components. This is the main improvement over the conventional systems, where computing systems and communications are added to the physical processes, by keeping the identity of every component.

The development of CPS was possible due to the existence of global networks as "Internet of Things" (IoT) and "Internet of Services" (IoS), along with complex networks of sensors. The interconnection between systems embedded in management system architectures and multiple sensor networks had led to the emergence of this new class of systems, named "Systems of systems". CPSs will transform the way it interact with the physical world, just like the Internet has transformed the way we interact with others.

By integrating all of the components, CPS ensures new capabilities as higher efficiency, lower costs, and the ability to interconnect in within the framework of complex structures. Therefore, these systems play an important role in the development of future systems [1].

4.2.1 Classical Cyber-physical Systems

The design and implementation of the CPSs was possible due to the emergence of sensor networks and network-embedded systems. These systems represent a complex new class, which can collect a large amount of real-time data from the sensor network, in order to take the most suitable decision. One of the complications that arise during the design of such a system is that the physical world is not entirely predictable. For this reason, the system must be robust to changing environmental conditions.

The ultimate goal of CPS is to use infrastructure cybernetics like detection, computing, communication, and hardware/software, in order to intelligently monitor (from physical to cybernetic) and control (from cybernetic to physical) the surrounding world.

The CPSs integrate sensors, execution elements, physical processes, electronic, and software devices in such a way that they allow the acquisition of strong interconnected systems, with real capabilities. The performances and the complexity of the systems are easy to highlight when two or more domains are interconnected. Due to the complex systems that have to interact continuously, it is necessary to design a precise architecture for the whole CPS.

Building blocks of the CPS have appeared since the last century. The blocks and concepts used in the development of this system are presented in Figure 4.1. One of those blocks is the embedded software system. These include the specialized processors both by construction (design) and by a specific software. This part of the system generates optimized solutions in order to ensure the control of a system.

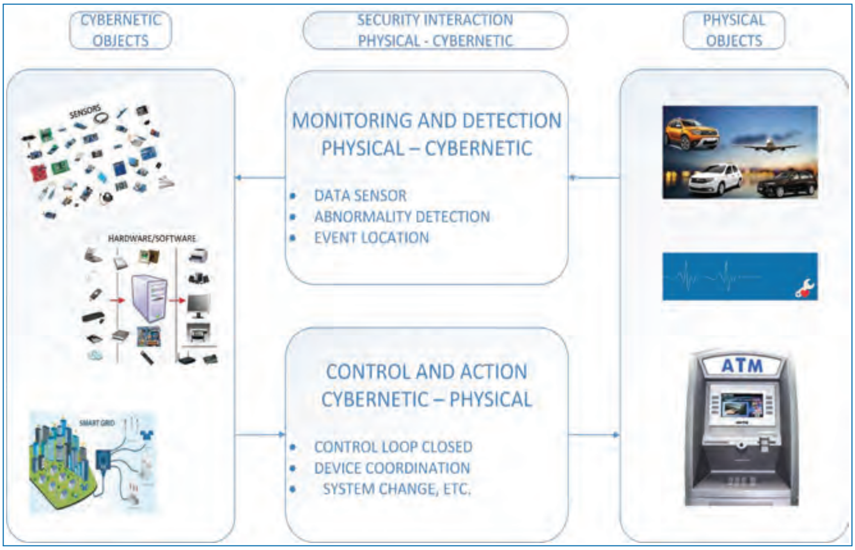


Figure 4.1. Cyber-physical system—examples and concepts.

But there are at least two problems regarding this block. The first is that it represents a closed solution, which implies a complete separation of the other systems. It is not intended for networking. The second problem relates to the fact that the operating safety has been tested and is guaranteed for a certain hardware and software implementation. If it is necessary to replace the microprocessor with a newer one, then the whole system will have to go through all the tests from the beginning.

The second block is the microsystem, which integrates in the same component the input transducer (sensor) or the output transducer (actuator), and the processing part of the electrical signal. This block ensures the interaction between CPS and physical environment. The accelerated development of the transducers that provide the interface with the physical world was one of the progress factors that lead to the new industrial revolution—Industry 4.0.

In addition to these two blocks, the human factor, the experts in the field, is incorporated in the decision process. Also, the client is included in the system, which allows the manufacture of personalized products.

It is more challenging to design a CPS, compared to a purely cyber system or a physical system. In this case, both the behavior of computational components and the physical environment must be taken into account, in order to obtain a unified framework. The design differences between software and mechanical engineering increase the difficulty level of implementing a CPS. As can be seen in Figure 4.2, during last years it was developed a design architecture consisting of five layers [2].

The first layer is the Smart Connection Level, which involves collecting data from the physical environment. In order to design a CPS, the connected systems

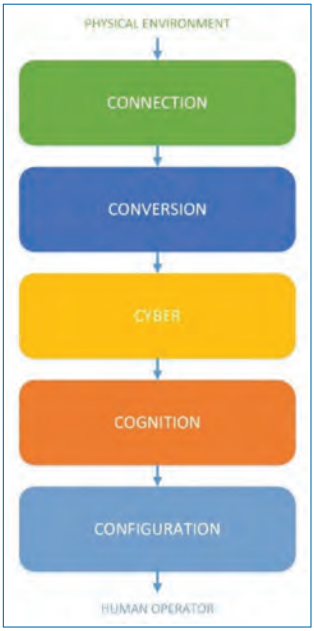


Figure 4.2. Levels of 5C CSP architecture.

and their components must acquire the data as accurately as possible. At this level, there are used different types of sensors, which can collect a large variety of data. Afterwards, the Conversion stage follows. At this point, the data collected from different resources is converted into relevant information, which can be used in the real-world application.

The next stage is the Cyber Level, which is one of the most important step in developing a proper design for CPS. After the information is extracted from all the interconnected machines, it must be compared with other similar machines, in order to discover details regarding system variation and life prediction. In this way, it can be created a base for every machine through the system, to know as much details about how a specific machine behaves in time.

At this point, the human operator has all the information needed so he can make the decisions. In the Cognition Level, the decisions must be prioritized and optimized, in order to determine the priority of tasks. In the last stage, the Configuration Level, the user is in charge of maintenance, by supervising the system and giving feedback to the physical part. The system can be reconfigured depend on the priority of the tasks and the risk criteria. In this way, it will be designed a self-adaptive and self-configured system [2].

So, during the five levels presented above, the CPS retrieves data from the sensor network and decides if an irregular action has taken place. If the alarm sensors detect an event which is not normally part of the environment, they send signals to the alarm panel. For the system to be fully operational, the sensors must be strategically placed in order to monitor all the areas that require security alerts [3].

The presence of an unusual event is signaled to the human operator, by generating an alarm. The performance of CPS is influenced by the way it detects those events that need immediate attention and also by the number of false alarms it generates. Also, the system depends on the human component. People are responsible for the proper functioning of the CPS. They must monitor the system and the generated alarms, in order to decide if any intervention is necessary to ensure that the technology is applied properly.

One of the main challenges, which comes from the interconnection of several systems, is to ensure the security of a CPS. Because of the existence of distinct components, a failure in one of them could lead to a whole system failure [4]. If taken separately, the components may not represent a real threat during an attack, when they are interconnected it could lead to serious consequences. The concern about real-time operation can interfere with the security of the system.

The system and the human operator interact in a spontaneous manner, which represents a real challenge for designing the system. The CPSs are described as "Systems of systems" as a result of interconnection of the embedded systems linked to the physical environment through sensor networks and execution elements. Those systems have partial autonomy and adaptive ability, and also an advanced cooperation between system and human component.

Everyday life is becoming increasingly dependent on such systems. For this reason, the ability to adapt to such intelligent systems need to be continuously improved through education and training. The tendency to increase the computational intelligence, the degree of automation, and control of some complex processes requires the rethinking of the human operator role and training for new skills and actions. In addition, the large amount of data that is retrieved from the physical environment leads to the generation of many alarms, which are becoming increasingly difficult to be managed by a human operator.

For those reasons, the classical CPS need further improvements, in order to keep up with the evolution of the physical environment.

4.2.2 Need of Improvement

The correct functionality of the CPSs depends on the correctness and accuracy of the sensor network. The data acquired from the sensors can be influenced by many external environmental factors. Furthermore, the sensors can be affected by faults and uncertainty, which can have a strong negative effect on the decision-making process. This will either lead to multiple false alarms or missed unusual events [5].

Other major design challenges of the CPS are the disconnection between abstraction layers, the lack of precise synchronization, or the inadequate consistency.

For CPS development through major progresses and by integrating advanced technologies is required a deeper understanding of the integration of real-time processing, with embedded wireless network which works with a wide range of sensors. For a complex and complete operation of the CPS is required an efficient detection and control of the physical systems, through a robust software architecture, focused on system hierarchies, protocols, and analytical procedures.

Furthermore, the CPSs should use concepts of new and highly intelligent programming and advanced hardware design. The mechanisms used to interact with sensors are not fully represented by existing programming languages. The CPSs must be concurrent, because the surrounding world is concurrent. So, an improvement is the usage of abstractions which can lead to an intuitive modeling of the real world [1].

CPSs must overcome all challenges and events that may take place in the future, in order to ensure security, safety, and predictability. These systems must operate in real time all of the operation needed and take into account data delays captured by the sensors.

4.2.3 SMART Cyber-physical Systems

Due to the continuous increase of the interconnection between the physical and the virtual world, and the development of increasingly sophisticated and complex algorithms, became necessary the emergence of a new generation of cyber-physical systems.

The Smart CPS represents a modern CPS system, which is able to integrate a larger number of physical components and, respectively, computation components. This new system can control the items from their environment in a more intelligent manner, in order to achieve a higher degree of efficiency [6].

The main difference between the classical CPS and the smart Cyber-physical Systems is related to how they process the data retrieved from the network of sensors. Sensors provide data taken from a physical object, which is further used by the device to perform a function.

Regarding the classical systems, the data generates alarms directly, without going through a processing level. This explains the occurrence of many false alarms or the loss of real alarms. Thus, an evolution of the system was essential, which is way were introduced multiple stages of data processing. In the case of smart CPS, after the data is collected by sensors, a number of algorithms are developed and simulated again until a correct action is selected and the response is sent to the action device. Thus, after the phases of data processing, aggregation, and fusion, the alarms received by the user will be fewer and more accurate.

The number of false alarms decreases, being generated only the alarms that represent a real risk. This is one of the major improvements made by smart CPSs,

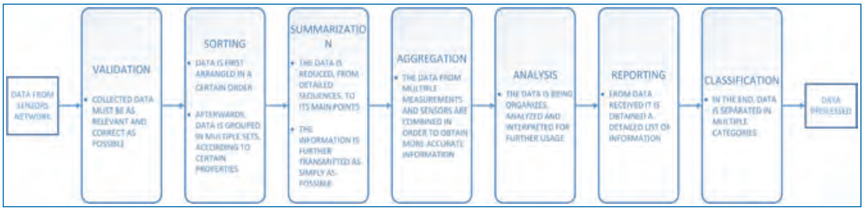

DATA FROM SENSORS NETWORK
VALIDATION
• COLLECTED DATA MUST BE AS RELEVANT AND CORRECT AS POSSIBLE
SORTING
• DATA IS FIRST ARRANGED IN A CERTAIN ORDER
• AFTERWARDS, DATA IS GROUPED IN MULTIPLE SETS, ACCORDING TO CERTAIN PROPERTIES
SUMMARIZATIO N
• THE DATA IS REDUCED, FROM DETAILED SEQUENCES, TO ITS MAIN POINTS
• THE INFORMATION IS FURTHER TRANSMITTED AS SIMPLY AS POSSIBLE
AGGREGATION
• THE DATA FROM MULTIPLE MEASUREMENTS AND SENSORS ARE COMBINED IN ORDER TO OBTAIN MORE ACCURATE INFORMATION
ANALYSIS
• THE DATA IS BEING ORGANIZES, ANALYZED AND INTERPRETED FOR FURTHER USAGE
REPORTING
• FROM DATA RECEIVED IT IS OBTAINED A DETAILED LIST OF INFORMATION
CLASSIFICATION
• IN THE END, DATA IS SEPARATED IN MULTIPLE CATEGORIES
DATA PROCESSED

Figure 4.3. Data processing flow.

which greatly help the human operator, while also increasing the efficiency of the alarm system. The processing phase of the data acquired by the sensors network consists of several techniques and processes presented in the diagram in Figure 4.3. The data collected by the sensors network must go through several processing steps in order to generate the information that finally reaches to the human operator.

A first step is data validation. The data must be as accurate as possible so that the final system does not generate false alarms. This is the phase in which the data is cleaned and verified, so that only the correct, consistent, and valid data will pass further through the system. The accuracy and the correctness of the final system depend on this first step.

Next step is the sorting of data. Previously obtained data is systematically arranged in a logical order and grouped into several sets according to certain properties. The relevance of the data is very important during processing phase. The existence of irrelevant data leads to a decreased quality of the information generated in the end.

The data is further reduced from detailed sequences, so that in the next step the information being represented as simple as possible. The aggregation step is one of the most important phases of data processing. Using various aggregation techniques, the data can be combined from several measurements, from various sensors in the network. When the data are aggregated, the observation groups are replaced by a statistic based on these observations. By combining several measurements, both the redundant information and the wrong information from the system can be eliminated. In this way, the smart CPS system will generate in the end fewer false alarms and more correct ones. And so, the accuracy and correctness of a smart CPS is increased compared to the classical systems.

The last three steps analyze the aggregated data and interpret them in order to generate detailed lists describing the information reached in this point. The last step, the classification of information, groups the data into several categories. In essence, the role of data classification is to use known variables in order to predict unknown or future variables.

So, data processing is the way in which all the data received from the sensors is converted into useful and correct information. This process is implement through a computer that receives at input rows of data and generates at the output the information used later by the human operator.

The smartness of this system is, for the most part, implemented in software. The smart CPS combine data from various sources and apply intelligent algorithms that process information from the real world. The interactions that occur at the physical level can change the behavior in the virtual world. This connection is exploited by the smart CPS system for continuous improvements of the processes, which is reflected in a higher level of adaptation to the physical environment, and also of optimization of the system. Thus, a more diverse range of applications and services can be developed.

The components of the smart CPS interact with each other in a robust and decentralized manner. These components must also maintain a high degree of autonomy. For this reason, there are several challenges during the implementation.

One of the challenges is to design intelligent physical infrastructures for communicating between physical objects and virtual world. Another major challenge is the processing of data collected by the sensors. It must be ensured the data stream processing along with data analytics, and the implementation of newer machine learning techniques, through development of self-adaptive software. Moreover, being a system with an increased complexity, and also having integrated a human component, it must be taken into consideration the social and behavioral problems that arise.

One other major problem that has arisen with the increase of the capabilities of smart CPS system is the need to increase the security. In order to increase efficiency, those systems have complex functionalities and advanced algorithms, which also increase the likelihood of a potential attack. This has a negative effect on the security system [6].

4.2.4 Applications of Cyber-physical Systems

The CPSs provide unique features, which is why they are used more and more often in the surrounding world. These systems have a very wide range of applicability, starting from automated machines to the medical field. In addition, these are very common in the field of security. Whether it is the supervision of personal houses or areas with higher risks, such as an Automated Teller Machine (ATM), the CPS systems are increasingly widespread.

One of the first areas in which those systems have been used is the field of transport. Modern automated vehicles represent the typical example of a CPS. These cars provide a number of features, as obstacle detection system, deceleration system in case of obstacle detection, automatic driving, and continuous surveillance of the energy consumption of the car, etc. Furthermore, those systems are used also in the aerospace field, the requirements being much more demanding in this case.

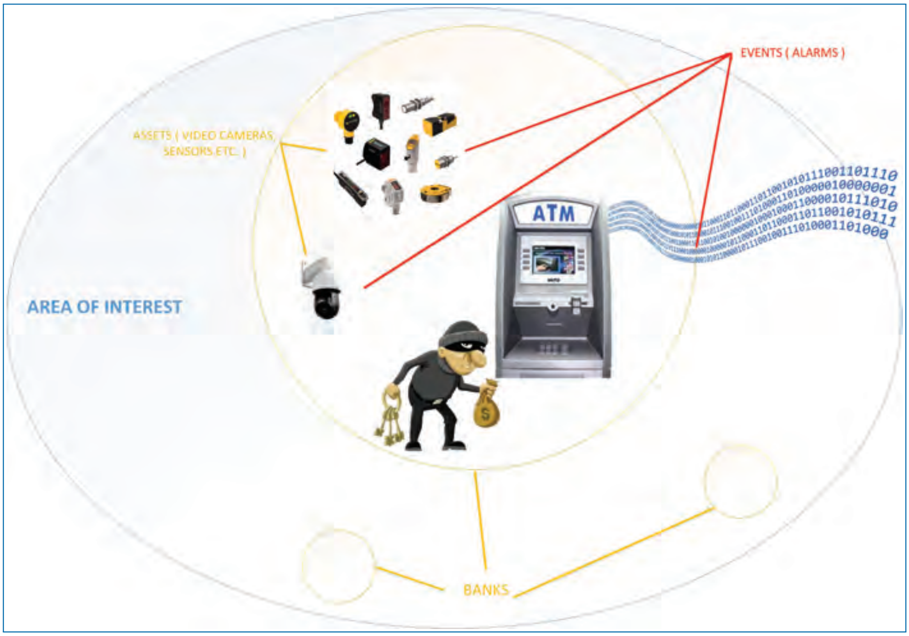


Figure 4.4. Applicability of CPSs—ATM's security.

Another area in which the CPS systems gain ground it the medical field. These make it possible to remotely monitor an elderly or a sick person. For this purpose, the sensors must be installed in the room or even in the clothing. Afterwards, the sensors will trigger alarms in case of any danger.

In recent years, the CPSs have begun to be used also in the field of environmental protection. These can be used in order to monitor the degree of pollution and to alert in case of danger situations such as floods, fires, earthquakes, etc. Besides that, the climate change can be closely monitored, together with its effects on the planetary scale.

Nowadays, the biggest spread of CPS systems is in the security field. These are commonly used to supervise various buildings or areas. The most frequent is currently the case where those systems are installed for the surveillance of personal houses. The use of CPS turns the house in a safer and more secure place, and, in addition, it can provide many other features designed for a "smart" house. Furthermore, another main role of the CPSs systems is the surveillance of several organizations, like banks. Such situation is presented in Figure 4.4.

This is the moment when it becomes obvious the importance of the evolution of CPS systems, from classical to smart systems. In situations related to security and surveillance of perimeters or buildings with high risk of attacks, a more efficient and accurate system is required. The main reason is that generating a large number of false alarms will tire the human operator, which could subsequently lead to the loss of other real alarms. The probability of not taking into account a real attack from inside the perimeter increases with the number of false alarms generated by the system.

On the other side may also arise a situation when a real alarm is not generated by the system. In this case, the ATM security is compromised. For this reason, the most appropriate system to monitor and protect an ATM, or other area prone to attacks, is the smart CPS system. This should lead to increased security in the financial-banking field, and, on a larger scale, to increased national security.

4.3 Data Fusion (Physical and Cyber Events)

4.3.1 Theoretical Aspects

Data fusion is the process of combining data from multiple sources (sensors, cameras, etc.) and processes (video analytics algorithms, cyber security agents) in order to obtain more relevant and useful information at the end.

The most common data fusion usage is in geospatial applications due to the necessity of correlating targets in space and time (the sensors do not report the information simultaneously).

Through the activities that an attacker can carry out in the pre-stages of an attack, as well as during the intrusion itself, they can use tactics specific to the military space, misleading detection systems, resulting in monitoring data being uncertain and confused.

In addition, the transposition of more and more human activities in the virtual space, combined with the dynamics of technological changes, will determine an increased complexity of IT architectures and their associated management processes. According to the principle of the incompatibility between precision and complexity, which manifests strongly in human monitored systems, it is expected that, in terms of security management, this complexity is translated as the availability of a large mass of data and information, but which will have an increasingly high content of imperfections.

The imperfection of the data must be incorporated into the systems that are trying to provide the most accurate modeling of reality. However, this is difficult to achieve with the use of current solutions offered by information management systems.

One of the mathematical models that allows working under uncertain conditions is known by the name of evidence-based reasoning theory (Dempster-Shafer theory—DST) [7, 8]. The premise of the theory was that the ignorance of an agent towards one statement must not cause the probability to be evenly divided between the value of truth and of false, as assumed in classical probabilistic reasoning. Even more, if there is the possibility of some mutually exclusive alternative, and the agent can only set the probabilities for some of them, according to the classical probabilistic reasoning, the remaining probabilities must be distributed in a certain way between the other alternatives.

Definition:

$$\Theta = \{\theta_1, \ldots, \theta_n\} \tag{4.1}$$

is called the fusion problem framework, and θ_i, where $i = 1, \ldots, n$, represents the set of hypotheses.

Shafer Model (M0(Θ)) assumes that $\theta_i (i = 1, \ldots, n)$ are precisely identified such that to ensure the exclusivity and completeness of the hypotheses. If Θ is open (condition exhaustiveness is not met), an element θ_{n+1} can be added closing so that it works with a closed case $\{\theta_1, \ldots, \theta_n, \theta_{n+1}\}$. Thus, without losing from generality, it will be considered that Equation (4.1) defines a closed discernment framework.

In the initial TDS, subsets are constructed as sentences, where the sentences of interest are shaped as:

$$P_\theta(A) = \textit{the truth value of } \theta \textit{ is in a subset } A \textit{ of } \Theta \tag{4.2}$$

Given the isomorphism between $P_\theta(A)$ and A, for simplicity and consistency with the terminology adopted in other theories, a representation based on sets is used in the definitions that follow.

Definition:
The power set

$$2^\Theta = (\Theta, \cup) \tag{4.3}$$

represents the set consisting of all subsets of Θ created based on the following rules:

- Ø, $\{\theta_1, \ldots, \theta_n\} \in 2\Theta$.
- If A, B ∈ 2Θ, then A ∪ B ∈ 2Θ.
- 2Θ does not contain any other element except those obtained using the first two rules.

For $\Theta = \{\theta_1, \theta_2, \theta_3\}$, we will obtain:

$$2^\Theta = \{\emptyset, \{\theta_1\}, \{\theta_2\}, \{\theta_3\}, \{\theta_1 \cup \theta_2\}, \{\theta_2 \cup \theta_3\}, \{\theta_1 \cup \theta_3\}, \{\theta_1 \cup \theta_2 \cup \theta_3\}\} \tag{4.4}$$

having the property $|2^\Theta| = 8$.

Definition:
It is called the basic trust table (simply called the table function); the function is:

$$m(\cdot): 2^{\Theta} \rightarrow [0, 1] \tag{4.5}$$

associated with a body of records B as follows:

$$m(\emptyset) = 0 \tag{4.6}$$

$$\sum_{A \in 2\Theta} m(A) = 1 \tag{4.7}$$

where the value m(A) is called the generalized mass of basic confidence of A.

4.3.2 Necessity of Data Fusion for Cyber-physical Security

Commonly cybersecurity and physical security are two completely different aspects and are treated by different security companies with different intervention protocols (Figure 4.5).

This can be a serious security vulnerability due to the fact that in many cases, the two types of attacks are correlated. The complexity of the attacks is also increasing and the technologies and techniques are getting harder to detect and combat. One typical example of a correlated attack is when a cyberattack is perform to disconnect an ATM from the back network and then a physical attack is conducted with divides that can perform unlimited withdraws.

There are several motives and flows in the current systems that permit these types of attacks to function.

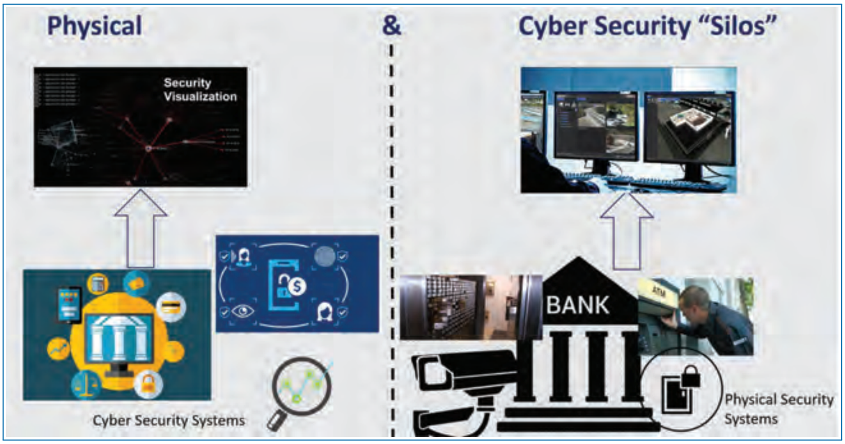


Figure 4.5. Current approach to CPS.

The first and most evident is the fact that the cybersecurity firms don't have intervention teams for the field, and when such an intervention is required, they rely on human intervention to notify the physical security team. This approach is inefficient and could be easily resolved if the end point could send certain events to both teams.

The second aspect that permits these attacks is the fact that the end point, in this case let's say an ATM, does not have any decision power. The most simple example of local decisions that could enhance security is the decision to put the ATM on lockdown in the case that it losses communications with the central servers. These approaches were prohibitive in the past due to the costs, size, and power consumption of systems that could make these aspects possible. Fortunately, currently embedded systems are extremely affordable and power efficient and have then sufficient computational power to handle the required tasks.

4.3.3 Implementation

As mentioned, NanoPCs such as the Raspberry Pi platform are more than powerful enough to operate a Data Acquisition module, a Data Fusion module, and a Decision and Distribution module, without requiring special consideration to space and power needs.

Such modules are being developed and launched to the market (Figure 4.6) at the time of writing this, but this can be considered as being quite late. Just in the UK, in 2019 the total ATM fraud damages were around 100 million Pounds [9], so if such solutions were implemented earlier, and assuming that they would prevent

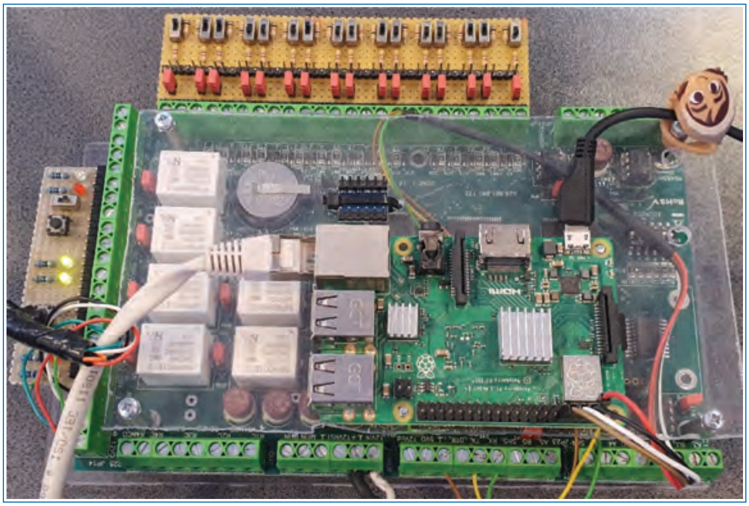


Figure 4.6. Intrusion detection and access control board.

only 10% of the thefts, the savings would be at around 10 million Pounds just in the UK.

With such a module connected between the sensors, the ATM PC and the central server could monitor all activity and determine if something is out of the ordinary, especially due to the fact that it has information about the whole state of the system.

4.4 Integration Layer FINSTIX Data Model Implementation

In the FINSEC project, a solution has been developed and demonstrated for an integrated, intelligent, collaborative, and predictive approach to the security of critical infrastructures in the financial sector. To this end, a proper data model is crucial to provide an integrated representation of physical and cyber assets and their relationships, to operate on data and to define the scope of the prediction algorithms.

In the design of a data model, two different approaches can be adopted: the first one comprises the definition of the model from scratch, covering all the business requirements of the considered use cases. This approach has a number of advantages, most relevant here being the fact that the data model would be perfectly adapted for the task and would be likely of a smaller size. Even though these advantages are substantial, the fact that the model would not be in a known industry standard is enough motive to dismiss this approach and continue to the second option.

The second option comprises the expansion (i.e., particularization, detailing) of an existing standard with the objects individualized by the use cases and missing in the standard. Thus, the FINSEC project pursued the second solution, resulting in the FINSEC-FINSTIX data model. FINSTIX extends the Structured Threat Information eXpression (STIX) 2 [10] standard combining information coming from both physical and logical worlds (thus supporting defenses against both cyber and physical threats).

STIX has been chosen because it already defines concepts important for the CTI (such as Observed Data, Vulnerability, Attack Pattern, Malware, Course of Action), while enabling an easy extension through the addition of custom parameters to already existing STIX objects and/or the creation of brand-new custom objects. In addition, STIX allows easy references to other external sources of intelligence (such as CAPEC).

The FINSEC extension to STIX2 has been driven by the FINSEC Project use cases, which led to the inclusion of information relevant to the financial sector, enabling the correlation of physical and logical data.

The whole FINSEC Platform can be conceived as an "intelligent engine" capable of transforming events and observed data from the physical and digital world (physical-cyber infrastructure) into Threat Intelligence. The information produced will be referred to Cyber and Physical Threat Intelligence (CPTI). In the same way that Cyber Threat Intelligence (CTI) is valuable information exchanged in the Cyber Security Domain, the CPTI produced in the FinTech sector is the added-value information produced by the platform which could be exchanged (in-out) between Financial Organizations and Security Organizations (CERT/CSIRT-like).

As an overview, the FINSTIX data model describes in detail the following aspects:

- The location and the characteristics of surveyed site
- The asset that is protected and its complete description
- The sensors that feed info into the system and they're afferent description
- The event description

The data model is in the JSON format. Up next, we will present a few examples of this data model so things become clearer. The examples are disposed from the highest level to the lowest. Before the complete JSON example section, we will provide a graph schematic (Figure 4.7) that will make the whole process more intuitive to understand.

Example 1: Organization.

```
{
  "type": "x-organization",
  "subtype": "Financial",
  "datatype": "Model",
  "domain": "Physical",
  "id": "x-organization--d02ca029-9afb-46de-af03-893c6cb4d79c",
  "created": "2019-06-17T16:38:49.668788",
  "modified": "2019-06-17T16:38:49.668788",
  "name": "Wirecard Co, Roma",
  "description": "Service Provider of ATM network",
  "x_organization": "x-organization--d02ca029-9afb-46de-af03-893c6cb4d79c",
  "reference": [ "x-stray--d0da5445-c970-41cd-b24a-8200123cda9b" ],
  "contacts": {
    "tel": "+30 9293489719",
    "email": "security@wirecard.com"
  }
}
```

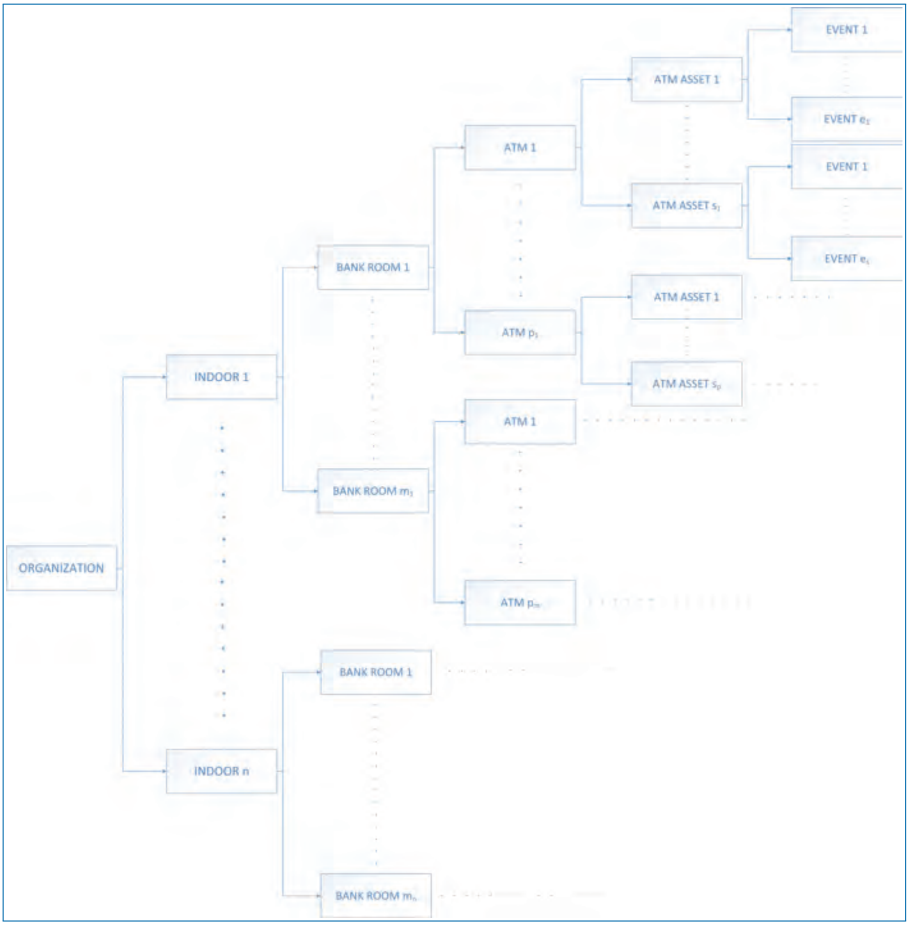


Figure 4.7. General architecture of the communication protocol.

Example 2: Indoor.

```
{
  "type": "x-area",
  "subtype": "area",
  "datatype": "Model",
  "domain": "Physical",
  "id": "x-area--660bb851-2194-4f90-8147-7ba34ef498fc",
  "created": "2019-06-17T16:38:50.167099",
  "modified": "2019-06-17T16:38:50.167099",
  "name": "Indoor",
  "description": "area inside the ATM",
  "x_organization": "x-organization--d02ca029-9afb-46de-af03-893c6cb4d79c",
  "reference": [ "x-organization--d02ca029-9afb-46de-af03-893c6cb4d79c" ],
  "coordinates": [
    44.463191,
    18.602537
  ],
  "area": [
    [
      -64.73,
      32.31
    ],
    [
      -80.19,
      25.76

  ],
    [
      -66.09,
      18.43
    ],
    [
      -64.73,
      32.31
    ]
  ],
  "asset_ref": []
}
```

Example 3: Bank Room.

```
{
  "type": "x-asset",
  "subtype": "main",
  "datatype": "Model",
  "domain": "Physical",
  "id": "x-asset--1dc5b93d-7f74-42a6-a428-508bd90dde50",
  "created": "2019-06-17T16:38:50.926607",
  "modified": "2019-06-17T16:38:50.926607",
  "name": "Room of a bank Branch",
  "description": "Bank Room with ATMs, cash desks and other assets",
  "x_organization": "x-organization--d02ca029-9afb-46de-af03-893c6cb4d79c",
  "reference": [ "x-area--572bc02a-cf16-44e0-bac2-64f003cf622e" ],
  "coordinates": [
    39.2301,
    -4.6262
  ]
}
```

Example 4: ATM Area.

```
{
 "type": "x-area",
 "subtype": "subarea",
 "datatype": "Model",
 "domain": "Physical",
 "id": "x-area--572bc02a-cf16-44e0-bac2-64f003cf622e",
 "created": "2019-06-17T16:38:50.344781",
 "modified": "2019-06-17T16:38:50.344781",
 "name": "ATM area",
 "description": "area of the ATM",
 "x_organization": "x-organization--d02ca029-9afb-46de-af03-893c6cb4d79c",
 "reference": [ "x-area--660bb851-2194-4f90-8147-7ba34ef498fc" ],
 "coordinates": [
  44.463191,
  18.602537
 ],
 "area": [
  [
    3.55,
    51.08
  ],
  [
    4.36,
    50.73
  ],
  [
    4.84,
    50.85
  ],
  [
    4.45,
    51.3
  ],
  [
    3.55,
    51.08
  ]
 ],
"asset_ref": []
}
```

Example 5: ATM Asset.

```
{
 "type": "x-asset",
 "subtype": "sub",
 "datatype": "Model",
  "domain": "Physical/Cyber",
  "id": "x-asset--65f16e13-a566-49b6-8f11-498eddf3c2b0",
  "created": "2019-06-17T16:38:51.054243",
  "modified": "2019-06-17T16:38:51.054243",
  "name": "ATM",
  "description": "ATM #534 inside the building",
  "x_organization": "x-organization--d02ca029-9afb-46de-af03-893c6cb4d79c",
  "reference": [ "x-asset--1dc5b93d-7f74-42a6-a428-508bd90dde50" ]
}
```

Example 6: ATM Probe.

```
{
  "type": "x-probe",
  "subtype": "LOG",
  "datatype": "Model",
  "domain": "Physical",
  "id": "x-probe--a68726cc-9e5b-44c2-90b5-776ca028eba1",

  "created": "2019-06-17T16:38:52.413222",
  "modified": "2019-06-17T16:38:52.413222",
  "name": "Access Probe",
  "description": "Permit legitimate access, forward sensor information",
  "x_organization": "x-organization--d02ca029-9afb-46de-af03-893c6cb4d79c",
  "reference": [ "x-asset--65f16e13-a566-49b6-8f11-498eddf3c2b0" ],
  "event_refs": [
    "x-event--e40bb60f-8b61-11e9-9fd4-ef9927e22a73",
    "x-event--e40bb60e-8b61-11e9-9fd4-ef9927e22a73"
  ]
}
```

Example 7: Event Intrusion.

```
{
  "datatype": "Model",
  "description": "Intrusion Event",
  "domain": "Physical",
  "id": "x-event--e40bb60f-8b61-11e9-9fd4-ef9927e22a73",
  "name": "Intrusion Event",
  "reference": [
            "x-probe--a68726cc-9e5b-44c2-90b5-776ca028eba1"
  ],
  "subtype": "room",
  "type": "x-event",
  "x_organization": "x-organization--d02ca029-9afb-46de-af03-893c6cb4d79c"
}
```

As it can be seen, the majority of the data model is human readable and the fields that are not are explained in the chapter detailing the protocol in more detail, here we are detailing the implementation.

When the connection is established to the end point, all the packages in the examples must be sent to the server in order to validate the connection. Afterwards, only the Event data is required, due to the fact that the server stores the end point description.

4.5 Conclusion

It is clear that standard security approaches are no longer viable for the modern world. The being said, the latest technologies available for artificial intelligence (a.i. Deep Neural Network—DNN, Convolutional Neural Network—CNN) have yet to be proven as secure and reliable as needed for financial institutions and applications. Another disadvantage of modern artificial intelligence approaches is the fact that they require a great deal of computational power, as such they consume a lot of electrical energy, thing that must be taken into account.

Taking this into account, we consider that the best approach for these cases is a combination of statistical validation, logical trees, and human monitoring. Such a system will yield the best results with the minimum of false alarms, without missing any relevant events, and notify all the interested parties. This fact will greatly reduce the time of intervention and drastically increase the system efficiency.

One of the latest technologies is the concept named "Cyber-physical Systems" (CPS). This system was determined in the first place by the need for evolution of the computer science and for increasing the security in fields like financial-banking. The main purpose of CPSs is to use infrastructure cybernetics like detection, computing, communication, and hardware/software, in order to intelligently monitor (from physical to cybernetic) and control (from cybernetic to physical) the surrounding world. Furthermore, the latest CPSs systems, called Smart Cyber-physical Systems (sCPSs), lead to an increased efficiency by processing the data collected by the sensors network (data aggregation, data fusion). Using such systems, it is possible to generate fewer false alarms and also to increase the final accuracy by not missing the real alarms. This is the reason why in areas like banks or other organizations, with an increased risk of cyberattacks, the CPSs represent the most suitable and secure system that can be used.

Data fusion is one of the most important steps in determining the validity and accuracy of the information provided by the sensors. There are many types of algorithms that are suitable for these applications. The thing they all have in common is that they all consider the data collected by the sensor being imperfect, until they can validate or invalidate the information. Data fusion represents the main reason why the system will not generate any false alarms in the end, and also why it assures that any relevant information will not be missed. And so, the human operator who must verify all the alarms and validate them will no longer have to deal with so many false alarms, and the efficiency of the system will be increased.

The last step important for an intelligent gateway is communication between all the systems and subsystems that compose the solution. In this chapter, we have presented examples of the FINSTIX protocol that are relevant for access control and alarm scenarios. The examples provided are complete, from the highest level (Monitoring Server) to the lowest level (sensors).

Acknowledgments

This work is partially supported by the by FINSEC project (contract number 786727, which is co-funded by the European Commission in the scope of its H2020 program), and also by SPIA-VA (Romanian PNIII-Solutii, UEFISCDI 2SOL/2017). The authors acknowledge valuable help and contributions from all projects' partners.

References

[1] S. Khaitan, J.D. McCalley, "Design Techniques and Applications of Cyber Physical Systems A Survey", IEEE Systems Journal, 2015.

[2] B. Bagheri, S. Yang, H.A. Kao, and J. Lee. "Cyber-physical Systems Architecture for Self-Aware Machines in Industry 4.0 Environment," IFAC-PapersOnLine, 2015, 48(3), 1622–1627.

[3] J. Kingsley-Hefty, "Physical Security Strategy and Process Playbook."

[4] M. Anand *et al.*, "Security challenges in next generation cyber physical systems," Beyond SCADA: Networked Embedded Control for Cyber Physical Systems, 2006.

[5] C. Alippi, "Intelligence for Embedded Systems," Springer, 2014.

[6] T. Bures *et al.*, "Software Engineering for Smart Cyber-Physical Systems: Challenges and Promising Solutions," ACM SIGSOFT Software Engineering Notes, June 2017.

[7] M. Beynon, B. Curry and P. Morgan, The Dempster–Shafer theory of evidence: an alternative approach to multicriteria decision modelling. Omega, 2000, 28(1), 37–50.

[8] T. Inagaki – Interdependence between safety-control policy and multiple-sensor schemes via Dempster-Shafer theory, IEEE Trans. on Reliability, vol. 40, no. 2, pp. 182–188, 1991.

[9] [https://www.ukfinance.org.uk/system/files/Fraud%20The%20Facts%20019%20-%20FINAL%20ONLINE.pdf]

[10] https://oasis-open.github.io/cti-documentation/

DOI: 10.1561/9781680836875.ch5

Chapter 5

Information Sharing and Stakeholders' Collaboration for Stronger Security in Financial Sector Supply Chains: A Blockchain Approach

By Ioannis Karagiannis, Konstantinos Mavrogiannis, John Soldatos, Dimitris Drakoulis, Ernesto Troiano and Ariana Polyviou

Security incidents in the finance sector highlight the need for sharing security information across financial institutions, as a means of mitigating risks and boosting the early preparedness against attacks. To address this issue and enhance the security and trust in the information sharing process, a blockchain-based solution for sharing security information in a decentralized way can be employed. Our earlier research work has reflected on this approach and proposed a reference architecture that incorporated a blockchain-based sharing of security information for critical infrastructures of the finance sector. In this Chapter, we extend this reference architecture by enhancing its collaborative risk assessment approach and a security knowledge database. We then employ an example to provide a demo of the dashboard that has been implemented.

5.1 Introduction

In recent years, we have witnessed a steady rise of cybersecurity incidents against infrastructures of the financial sector, such as phishing, ransomware, and DDoS (Distributed Denial of Service) attacks. These incidents include notorious attacks, which have resulted in significant economic damage, while decreasing trust in financial institutions and questioning their social value. As discussed extensively in the Chapter of this book which introduces the security challenges of the financial sector, the critical infrastructures of financial institutions are vulnerable. Some of the reasons of their vulnerability is the integration between physical and cybersecurity and the connectivity between the different systems and infrastructures. First, there is currently limited integration between physical and cybersecurity. This is because data-driven systems for the security of the finance sector are mainly addressing cybersecurity and ignore physical security systems. As a result, vulnerability assessment, threat analysis, risk mitigation, and response activities are fragmented. However, holistic approaches could assist financial institutions in better addressing security incidents involving both cyber and physical assets of their critical infrastructures. Second, as financial infrastructures are more connected than ever before, attacks are likely to impact other infrastructures and systems in the financial chain [1]. Thus, stakeholder collaboration could largely contribute identifying and alleviating such issues more effectively.

The exchange of security information across collaborating stakeholders of the financial services value chain can be a foundation for security collaboration in the relevant supply chain. In the scope of an integrated security approach, information for both cyber and physical security should be exchanged. This Chapter draws on [2] to extend the proposed blockchain-based system for collaborative security in the finance sector that includes an enhanced collaborative risk assessment approach and the incorporation of a security knowledge database.

5.2 Related Work

Collaboration is considered as one of the key activities in a plethora of European national cybersecurity strategies. Collaboration refers to the enhancement of cybersecurity at different levels so as to encapsulate threats sharing, risk assessment, and awareness raising. This is also reflected in the establishment of formal structures such as Information Sharing and Analysis Centers (ISAC) and Public Private Partnerships (PPP) [3]. In the finance sector, the Financial Services Information Sharing and Analysis Center (FS-ISAC) [4] was also established, as an industry forum for sharing data about critical cybersecurity threats in the financial services industry. ISAC centers support information sharing across stakeholders and assist the related collaborative workflows, such as those implemented in other sectors of the economy (e.g., the maritime [5] and transport sectors [6]). Collaborative security and information sharing options have been proposed in the literature (e.g., [7]), in order to support and complement conventional risk assessment techniques (e.g., [8, 9]). The rationale of information sharing is to trigger security processes like risk assessment and threat analysis, based on information received from other parties that join the collaborative security infrastructures. Trustworthiness and the security of the information sharing process are the two main obstacles in leveraging collaboration. This is because the use of a centralized database for sharing data involves disadvantages such as the requirement for a trusted third party (TTP) that will assume the ownership and will guarantee the integrity of the shared information. Additionally, it is susceptible to security attacks, which can compromise the shared data.

Financial organizations are overall reluctant to share information and thus avoid to share any information that lies beyond their compliance with regulations. Thus, a decentralized approach could provide solutions for addressing this issue. In particular, the use of blockchain technology could enable financial organizations to share information in a shared distributed ledger in a secure and decentralized way and hence in this way provides distributed trust. Alternative technologies that could be employed include STIX (Structured Threat Information Expression) [10], a protocol developed by OASIS to model cyber threat intelligence. TAXII (Trusted Automated Exchange of Intelligence Information) [11] refers to the application-layer protocol developed by OASIS to exchange STIX data. TAXII runs on top of HTTP and can provide secure connections over SSL (Secure Sockets Layer) if needed. But, TAXII is mainly a communication protocol, and thus, it does not provide storing capabilities. Hence, although it supports both publish-subscribe and client-server topologies, compared to blockchain, it lacks the guaranteed degree of confidentiality. Along the same lines, the alternative of pure P2P networks [12] could not provide a viable solution for sharing financial data. This is because the lack of solid authorization techniques could lead in information compromise, bad connections could possibly produce big network latency, while malicious files or messages can be easily implanted and consumed by other peers. For these reasons, information sharing is nowadays one of the most prominent blockchain use cases in the financial sector [13]. Existing literature provides a thorough analysis on the benefits arising by the use of blockchain technology in the financial sector [14, 15].

5.3 Collaborative Risk Assessment

In the proposed architecture, risks are calculated using specific metrics. These metrics include the vulnerability level, the impact level, and the threat level. Both vulnerability and impact levels were derived from the CVSS scores of the assets' vulnerabilities detected. The threat level is now a result of events occurring inside the organization and historical information. As a result, the calculations are more precise and are based on the current state of each organization.

Risk configuration is an object that follows the FINSTIX format (see Chapter 3) defined to make the risk calculation process easily adaptable to the needs of each organization allowing them to easily edit calculation triggers, add or remove events from the calculation scope, and, in general terms, enable customization. Essentially, through this object an officer can map events to threats and define trigger thresholds for the risk calculation.

5.3.1 Services

The first step to initialize a risk calculation suite is the creation of a Service. Services are stored in the FINSEC data-tier; hence, the communication with it is critical. In the current platform state, the data tier is protected using basic authentication. To protect the credentials, the username and password are provided as environment variables during the container initialization.

The form creation involves the asset selection as well as the vulnerability definition for each asset. The latter is now leveraged by the introduction of the Security Knowledge Base.

Important information related to a service includes:

- Name—which identifies the service along with the id;
- Description—which provides extra information for the security officers;
- Criticality—which defines the level of importance of the service. This information is important because mitigation actions are sometimes urgent and should be handled immediately;
- Subtype—which identifies the level of exposure (e.g., if the service is part of a supply chain, the subtype value will be "public");
- Service references—which lists the dependency of the current service to other services, either inside or outside the borders of the organization.

5.3.2 Threats

While Services provide the ability to group assets inside the organization, it could be impossible to calculate a risk on them without the detection of threats that may target the service. Likewise, a list of events should be defined. These events affect the level of the threat in real time. Threats are associated with the Service using the risk configuration object. Threat objects must be stored in the Security Knowledge Base. Therefore, the form endpoint of the Collaborative Risk Assessment GUI

(Graphical User Interface) will send a POST request to the deployed FINSEC KB (Knowledge Base).

The key properties of a Threat are:

- Name—identification of the threat;
- Description—details of the threat;
- Domain—cyber or physical;
- Subtype—related to the subtype. Example may be "natural disaster" in case of "physical subtype";
- Impact description—What may happen if the threat if realized;
- Likelihood.

5.3.3 Events

As mentioned before, events play a significant role in the risk calculation process. First, a security officer needs to define event models and then map them to a predefined threat. For instance, an "invalid login attempt" is related to a "SWIFT compromise threat." Consequently, when a probe produces an instance of this model, the Collaborative Risk Assessment platform detects it, and if the trigger value is reached for this specific event, the overall risk of the related threat is re-calculated.

Event details must include the following values:

- Name—identifies the event;
- Description—provides more information about the event;
- Domain—cyber or physical;
- Subtype—main or sub (in case the event is of subtype sub, it means that it is dependent of another parent event);
- Probe reference – defines the probe that produced the event;
- Coordinates—only for event instances;
- Observed references—provide the whole observation (may be pointing to an observable like IP address, binary file, etc.).

5.3.4 Triggers

A key consideration is the conditions that trigger the calculation process. In our approach, the calculation can be triggered in three ways:

- Manually;
- Vulnerabilities of the assets involved have changed;
- Event Instances reach a specified threshold.

The threshold is defined during the risk configuration by the security officer. It is an integer value which currently refers to the detections per day. Thus, when set to the number 3, the risk computation will run after the third detection of the specific event. The same event model may be associated with other threats, with a lighter or more sensitive bound. The threshold value is stored inside the Collaborative Risk Assessment platform's local storage (internally).

5.3.5 Risk Calculations

Figure 5.1 presents a high-level overview of the risk calculation process. For the service to function properly, certain preconditions need to apply. These include the service definition, the threat to event mapping, and the probe to be up and running.

As soon as a probe produces a new event, it is forwarded through the data collector to the FINSEC data layer. The Collaboration Service is connected to the data layer and is "listening" for event instances. After the event detection, the Collaborative Risk Assessment Engine:

- Examines all the Services of the organization;
- For each service, it checks the corresponding risk configuration;
- If the risk configuration does not define a relation of the current service to the event detected, the process is terminated;
- If the risk configuration defines a relation of the current service to the event detected, The Collaborative Risk Assessment Platform fetches the threats related to the event instance as well as all the vulnerabilities of the service (through its assets);
- The vulnerability, impact, and threat levels are calculated internally;
- A new FINSTIX risk object is created and sent to the data layer;
- The object is also displayed in the Dashboard;

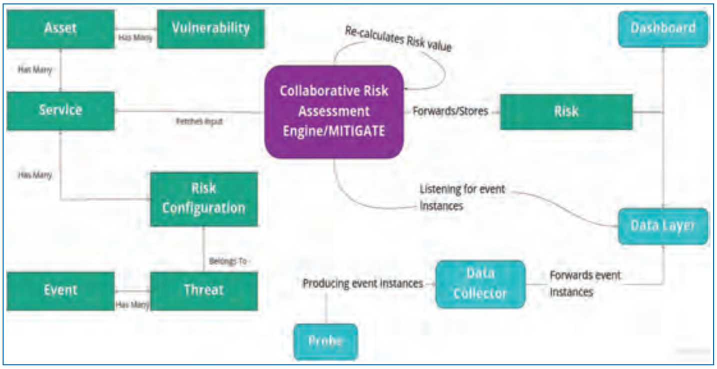


Figure 5.1. Collaborative risk assessment inputs/outputs.

- The logged in security officer checks the new risk calculation details;
- The officer can either approve or decline sharing the object with other stakeholders.

Note that the Collaborative Risk Assessment Engine is developed and customized based on the risk assessment platform of the H2020 MITIGATE project[1].

5.4 Information Sharing Architecture

5.4.1 FINSEC Platform Overview

Aiming to elevate security collaboration in the financial services supply chain, this Chapter extends the proposed information sharing architecture included in [2]. The proposed architecture (Figure 5.2) regards wider platform for financial infrastructures security developed in the frame of the FINSEC H2020 research project. The implementation of the FINSEC platform is based on a state-of-the-art microservices architecture. The platform encapsulates a Big Data system for security analytics, which provides the means for collecting security-related information from physical and cybersecurity systems. The platform can be viewed as a n-tier architecture, with a lower layer (i.e., the edge layer) that interfaces with the actual physical and logical infrastructures. Moreover, it includes several cross-cutting services, which are not confined to providing support to a single tier, but rather support functionalities that may reside in any of the layers of the architecture.

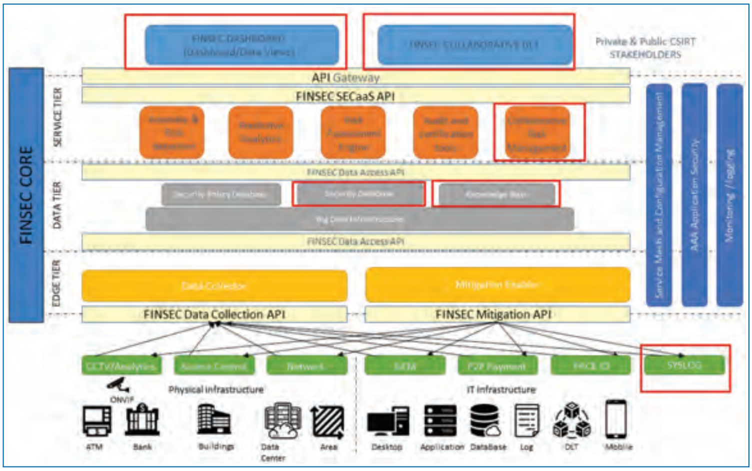


Figure 5.2. Main tiers of the FINSEC platform architecture.

[1] https://cordis.europa.eu/project/id/653212

The main tiers of the architecture enable the implementation of the previously presented building blocks and are as follows: (i) The Field Tier is the lower level and includes the probes and their APIs, whose role is extracting raw data from the physical and logical assets to be protected against threats; (ii) The Edge Tier contains the Actuation Enabler and a Data Collection module, which is needed to filter information as it flows towards the upper levels; (iii) The Data Tier is the logical layer where information is stored and organized into three different storage infrastructures, providing consisting data access APIs to all other modules; (iv) The Service Tier is where the kernel applications and the security toolbox will be running (i.e., the security kernel of the platform), able to be used by external applications via proper APIs; (v) The Business Client Applications tier is the layer where end users and business applications may actually get benefits from the platform capabilities. The FINSEC dashboard enables the end users to visually monitor in real time the data and assets managed by the platform, while the (Supply Chain) collaboration module enables the sharing of information with other instances of the platform, including instances deployed in different business organizations.

The core platform encapsulates three tiers: the Edge, Data, and Service tiers, which interact with the external environment with two main interfaces, northbound API and southbound API. (i) The northbound API towards higher level applications (e.g., end-user/business applications) called SECaaS (Security as a Service) API. It represents a consistent and unified view of the individual APIs exposed by the service tier high-level services that represent the "major intelligence" of the platform. The SECaaS API is exposed, and the API Gateway, which is the single-entry point to the system for external clients. Among other capabilities, the API Gateway provides and supports Authentication, Authorization, and Accounting (AAA) services, which conceptually are part of the two cross-cutting vertical modules on the right of the figure (Application Security and Monitoring/logging). (ii) The southbound API interface, consisting of an "Event API" and a "Probe API", allows communication between the Edge Tier and physical and cybersecurity probes.

The SECaaS API is leveraged and invoked by external (north end) Business Client Applications (upper side of the figure). They are outside of the core platform and interact with it only through the SECaaS REST API. Typical examples of business client applications include: (i) The Dashboard application, a web-based GUI used by the profiled end users of the platform; (ii) The Collaboration application, which enables the collaboration of multiple platform instances (data sharing etc.); (iii) Third parties' applications that exploit the capabilities of the platform, such as risk assessment and regulatory compliance applications. The Collaboration application is illustrated in following paragraphs, as it is based on the sharing of data in a blockchain infrastructure.

The Service Tier defines the high-level services that represent the "major intelligence" of the platform. The Service Tier services communicate with each other in three (3) possible ways: (i) Synchronous communications through their REST APIs. In this case, being the services internal to the platform, it is not necessary to use AAA functionalities; (ii) Asynchronous communications via an MQ bus; (iii) Asynchronous communications through the Database Infrastructure.

The collaborative module refers to a FINSEC service aims to provide a collaboration platform on top of a blockchain ledger. The module is deployed as a FINSEC service and provides endpoints to produce and consume FINSTIX messages across organizations. It was originally built to support the Ethereum blockchain; however, efforts are in progress for supporting Hyperledger Fabric. The Open API provided is not expected to change drastically, so the already available endpoints are used to push/pull messages from the blockchain. New capabilities, trust model definition and so on will not pose further issues, and the integration will be seamless. The integration with the collaborative module was rather simple. Instead of the MITIGATE UI, now the information sharing functionality is embedded inside the FINSEC Dashboard.

The security knowledge base essentially utilizes external sources of attacks and vulnerabilities. The most popular of which are NIST NVD and ATT&CK. In case a new asset is stored inside the data tier, it is automatically associated (based on product name and version) with all its known vulnerabilities. This fact eliminates the need of manually importing cyber vulnerabilities for each new asset. Only physical vulnerabilities should now be imported by a security officer. Figure 5.3 presents the Security Knowledge Base architecture.

Additionally, the introduction of the security KB ensures that the vulnerabilities are up to date and updated when necessary. Integrating with the KB required the utilization of its endpoints to persist and fetch information related to threats

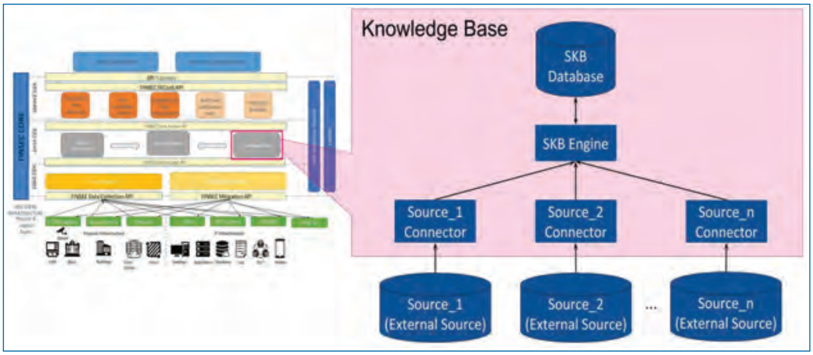


Figure 5.3. Security knowledge base—external sources.

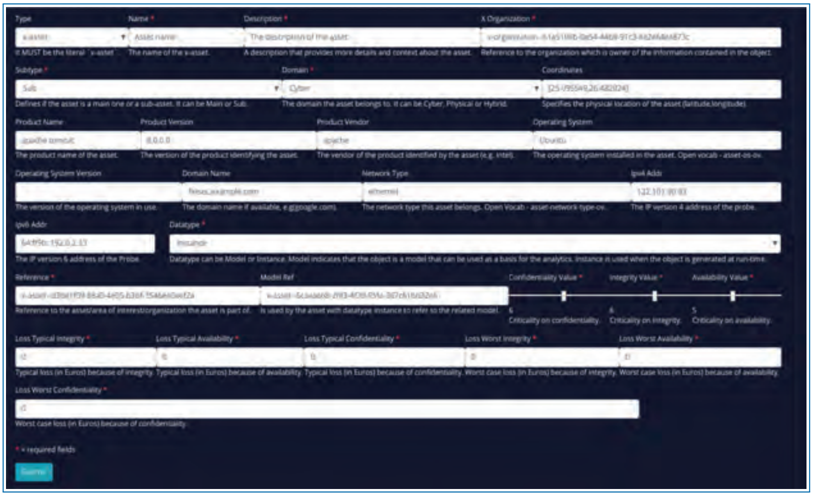
and vulnerabilities. The communication was RESTful, and the authentication was achieved using basic authentication[2] just as the data-layer case.

5.5 Implementation

5.5.1 CRUD Operations—User Interface

Collaborative Security Tools are encapsulated in the FINSEC Dashboard. Thus, all the forms needed are generated through the JSON schemas defined as a FINSTIX domain object. As a result, form validation coupled with form inputs needed for each object are provided for assets, threats, vulnerabilities, services, events and services. Association of domain objects lies on the security officer drag and drop actions, while notifications are still provided to the end user. The efforts were basically to update the FINSTIX schemas, align the Angular versions, code refactoring, so the forms can be automatically generated and other code adjustments on the Dashboard end to enable the full MITIGATE frontend operations.

Figure 5.4 presents the new form layout embedded in the FINSEC Dashboard. Both the validation errors and the input fields are auto-generated from a FINSTIX schema. Figure 5.5 illustrates the association functionality which is achieved with a dual filterable list box. Finally, 5 displays the sharing prompt as realized in the Dashboard.


Type: x-asset ▾
It MUST be the literal `x-asset`.

Name *: Asset name
The name of the x-asset.

Description *: The description of the asset
A description that provides more details and context about the asset.

X Organization *: x-organization--[illegible]
Reference to the organization which is owner of the information contained in the object.

Subtype *: Sub ▾
Defines if the asset is a main one or a sub-asset. It can be Main or Sub.

Domain *: Cyber ▾
The domain the asset belongs to. It can be Cyber, Physical or Hybrid.

Coordinates: [25.195549,26.482024]
Specifies the physical location of the asset (latitude,longitude).

Product Name: apache tomcat
The product name of the asset.

Product Version: 8.0.0.0
The version of the product identifying the asset.

Product Vendor: apache
The vendor of the product identified by the asset (e.g. Intel).

Operating System: Ubuntu
The operating system installed in the asset. Open vocab - asset-os-ov.

Operating System Version:
The version of the operating system in use.

Domain Name: [illegible]
The domain name if available, e.g(google.com).

Network Type: ethernet
The network type this asset belongs. Open Vocab - asset-network-type-ov.

Ipv4 Addr: 122.101.90.83
The IP version 4 address of the probe.

Ipv6 Addr: 64:ff9b::192.0.2.33
The IP version 6 address of the Probe.

Datatype *: Instance ▾
Datatype can be Model or Instance. Model indicates that the object is a model that can be used as a basis for the analytics. Instance is used when the object is generated at run-time.

Reference *: x-asset--[illegible]
Reference to the asset/area of interest/organization the asset is part of.

Model Ref: x-asset--[illegible]
Is used by the asset with datatype instance to refer to the related model.

Confidentiality Value *: 6
Criticality on confidentiality.

Integrity Value *: 6
Criticality on integrity.

Availability Value *: 5
Criticality on availability.

Loss Typical Integrity *: [illegible]
Typical loss (in Euros) because of integrity.

Loss Typical Availability *: [illegible]
Typical loss (in Euros) because of availability.

Loss Typical Confidentiality *: [illegible]
Typical loss (in Euros) because of confidentiality.

Loss Worst Integrity *: [illegible]
Worst case loss (in Euros) because of integrity.

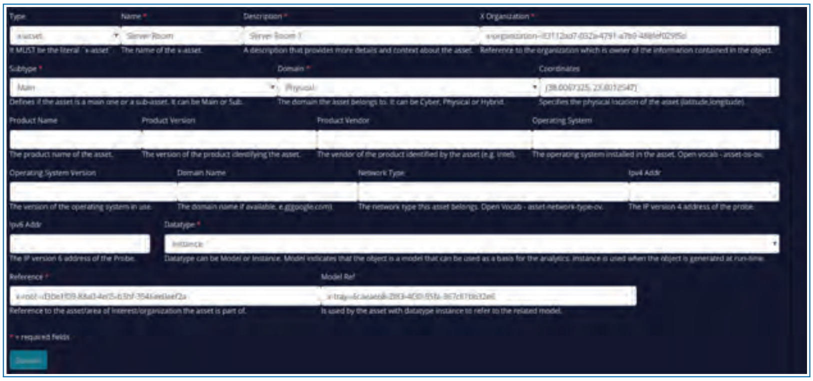
Loss Worst Availability *: [illegible]
Worst case loss (in Euros) because of availability.

Loss Worst Confidentiality *: [illegible]
Worst case loss (in Euros) because of confidentiality.

* = required fields

Figure 5.4. Form layout—dashboard integration.

[2] Basic Access Authentication requires a client to provide a username and a password during the HTTP message exchange. https://en.wikipedia.org/wiki/Basic_access_authentication



Type: x-asset
It MUST be the literal `x-asset`.

Name *: Server Room
The name of the x-asset.

Description *: Server Room 1
A description that provides more details and context about the asset.

X Organization *: x-organization--83112ad7-032a-4791-a7b9-488ef029f5a
Reference to the organization which is owner of the information contained in the object.

Subtype *: Main
Defines if the asset is a main one or a sub-asset. It can be Main or Sub.

Domain *: Physical
The domain the asset belongs to. It can be Cyber, Physical or Hybrid.

Coordinates: [38.0067325, 23.8012547]
Specifies the physical location of the asset (latitude,longitude).

Product Name:
The product name of the asset.

Product Version:
The version of the product identifying the asset.

Product Vendor:
The vendor of the product identified by the asset (e.g. Intel).

Operating System:
The operating system installed in the asset. Open vocab - asset-os-ov.

Operating System Version:
The version of the operating system in use.

Domain Name:
The domain name if available, e.g.google.com).

Network Type:
The network type this asset belongs. Open Vocab - asset-network-type-ov.

Ipv4 Addr:
The IP version 4 address of the probe.

Ipv6 Addr:
The IP version 6 address of the Probe.

Datatype *: Instance
Datatype can be Model or Instance. Model indicates that the object is a model that can be used as a basis for the analytics. Instance is used when the object is generated at run-time.

Reference *: x-root--d3be1f09-88a0-4e05-b3bf-3546ae8aef2a
Reference to the asset/area of interest/organization the asset is part of.

Model Ref: x-tray--6caeae68-2f83-4030-9384-867c87bb32e6
Is used by the asset with datatype instance to refer to the related model.

* = required fields

Figure 5.5. Server room asset creation.

5.6 Demonstrator

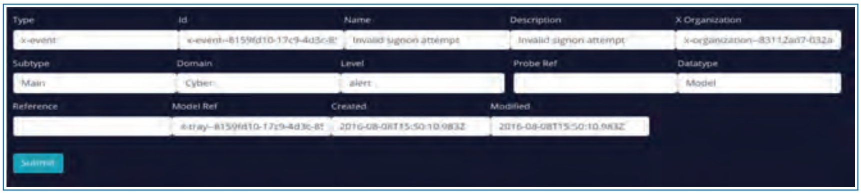
Drawing on an example of the behavior of a logged in security officer, in this Section we provide a demo of the proposed approach.

5.6.1 Initialization

As a first step, the security officer logs in and navigates to the Assets page. By clicking the button "Add New," the tool displays a form which must be filled and submitted to generate the new Asset. Figure 5.5 illustrates the generation of the first Asset detected.

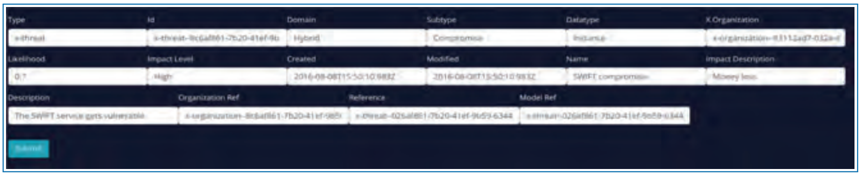
Next up, the security officer navigates to the Events page and creates the event models which will be considered for the risk calculations of the current demonstrator (Figure 5.6).

Soon after the event model definitions, the security officer must introduce a Threat. The operation is illustrated in Figure 5.7. Additionally, Figure 5.8 sketches the mapping of the Threat created with the appropriate event models. This step is crucial for the dynamic risk calculations. Note that threats are stored inside the Security Knowledge Base.



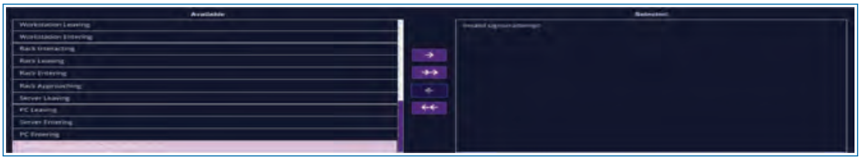
Type: x-event
Id: x-event--8159fd10-17c9-4d3c-8
Name: Invalid signon attempt
Description: Invalid signon attempt
X Organization: x-organization--83112ad7-032a
Subtype: Main
Domain: Cyber
Level: alert
Probe Ref:
Datatype: Model
Reference:
Model Ref: x-tray--8159fd10-17c9-4d3c-8
Created: 2016-08-08T15:50:10.983Z
Modified: 2016-08-08T15:50:10.983Z

Figure 5.6. Invalid Signon event model creation.



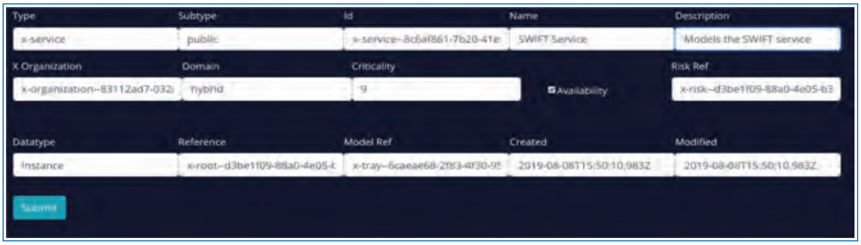
Type	Id	Domain	Subtype	Datatype	X Organization
x-threat	x-threat--8c6af861-7b20-41ef-9b	Hybrid	Compromise	Instance	x-organization--83112ad7-032a-4
Likelihood	Impact Level	Created	Modified	Name	Impact Description
0.7	High	2016-08-08T15:50:10.983Z	2016-08-08T15:50:10.983Z	SWIFT compromise	Money loss
Description	Organization Ref	Reference	Model Ref		
The SWIFT service gets vulnerable	x-organization--8c6af861-7b20-41ef-989	x-threat--026af861-7b20-41ef-9c59-6344	x-stream--026af861-7b20-41ef-9e59-6344		

Figure 5.7. SWIFT compromise threat creation.



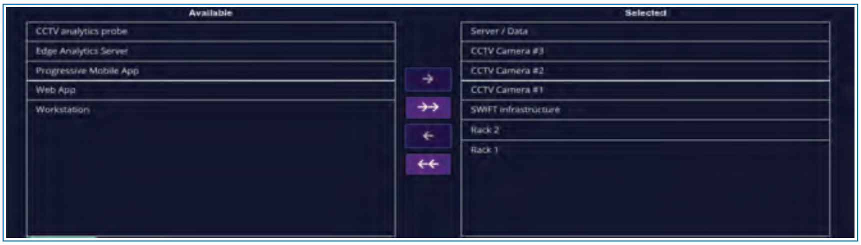
Available	Selected
Workstation Leaving	Invalid signon attempt
Workstation Entering	
Rack Interacting	
Rack Leaving	
Rack Entering	
Rack Approaching	
Server Leaving	
PC Leaving	
Server Entering	
PC Entering	

Figure 5.8. SWIFT compromise threat mapping to relevant events.



Type	Subtype	Id	Name	Description
x-service	public	x-service--8c6af861-7b20-41e	SWIFT Service	Models the SWIFT service
X Organization	Domain	Criticality		Risk Ref
x-organization--83112ad7-032	hybrid	9	☑ Availability	x-risk--d3be1f09-88a0-4e05-b3
Datatype	Reference	Model Ref	Created	Modified
Instance	x-root--d3be1f09-88a0-4e05-b	x-tray--6caeae68-2f83-4f30-95	2019-08-08T15:50:10.983Z	2019-08-08T15:50:10.983Z

Figure 5.9. SWIFT service creation.



Available	Selected
CCTV analytics probe	Server / Data
Edge Analytics Server	CCTV Camera #3
Progressive Mobile App	CCTV Camera #2
Web App	CCTV Camera #1
Workstation	SWIFT infrastructure
	Rack 2
	Rack 1

Figure 5.10. SWIFT service asset attachment.

At this stage, the security officer is ready to create the SWIFT service. The pure service information is initially provided. Consequently, the threats are associated with the service, and finally, a risk calculation object is being filled in to define risk triggering conditions. Figures 5.9 and 5.10 illustrate the steps followed.

Inputs/Outputs

All the FINSTIX objects created via the FINSEC Dashboard. These objects will serve as input for the MITIGATE tool. The objects cover both the use cases defined in the SWIFT Service pilot and include the Assets detected, the Event models, the

Threats identified, the Service, and finally the Risk Configuration Object. Additionally, the MITIGATE will listen for Probe events, and thus, these events are also considered input for the Collaborative Risk Assessment Service. Vulnerabilities detected for every asset are used in the risk calculations. They are the building blocks for calculating the Vulnerability and Impact metrics. Using the aforementioned inputs, the MITIGATE platform will produce a risk object which will be available for sharing with other stakeholders. The risk object essentially constitutes the output of the Collaborative Risk Assessment Service.

Demonstrator

As soon as all the necessary input is provided by the Security Officer, the vulnerability constitution is available in the FINSEC Dashboard home page. Figure 5.11 illustrates the vulnerabilities for the SWIFT service pilot, categorized by their domain (cyber/physical).

Figure 5.12 displays the auto-imported vulnerabilities from the Security Knowledge Base, while Figures 5.13 and 5.14 compose a proof that the vulnerabilities detected for the NodeJS server are also defined in the external source (CVE).

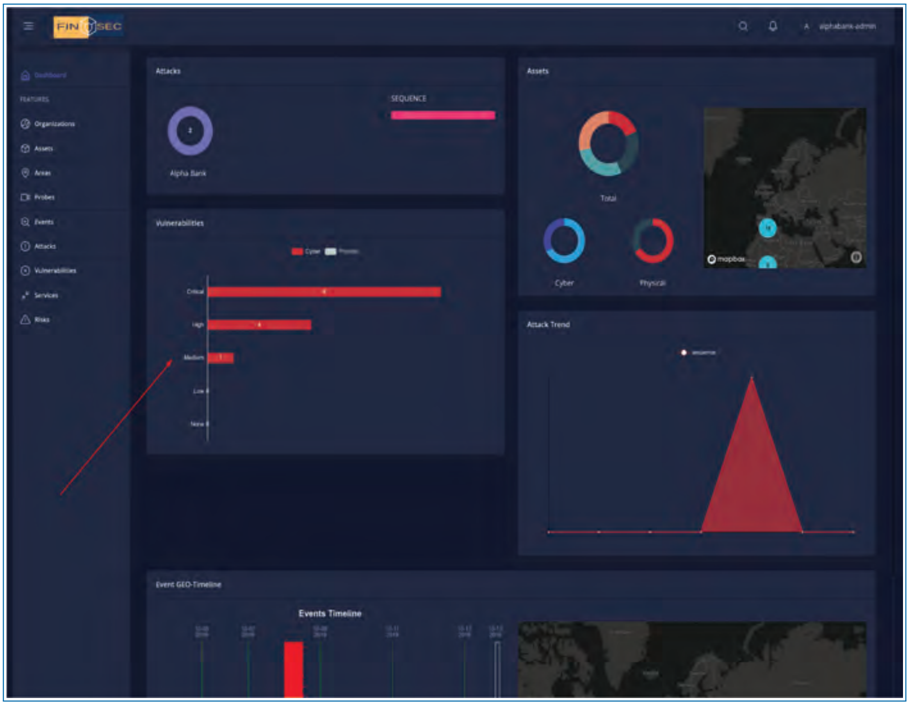
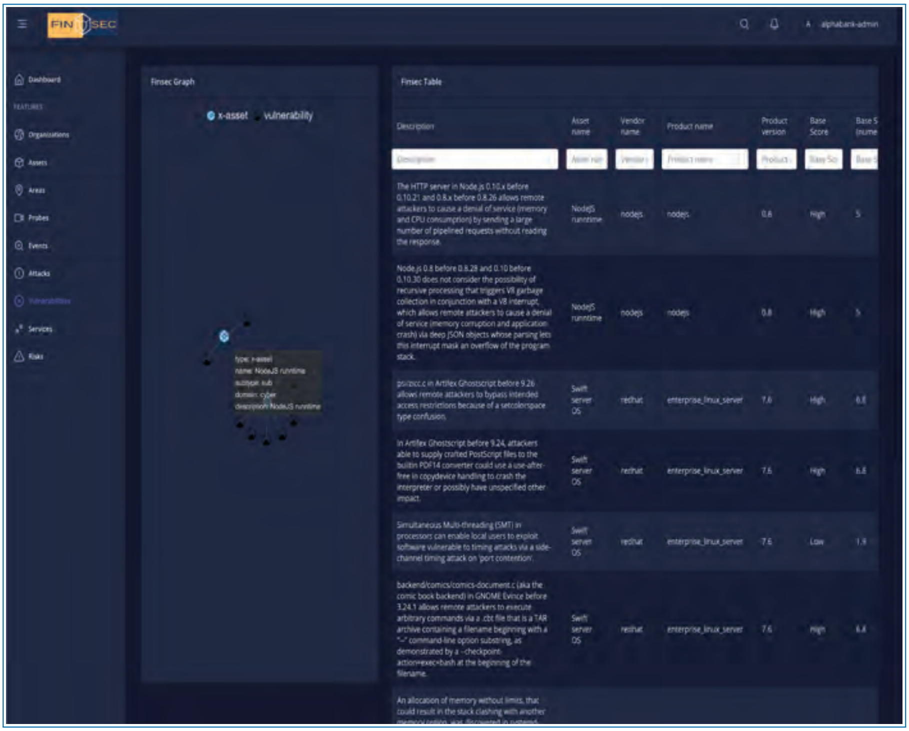


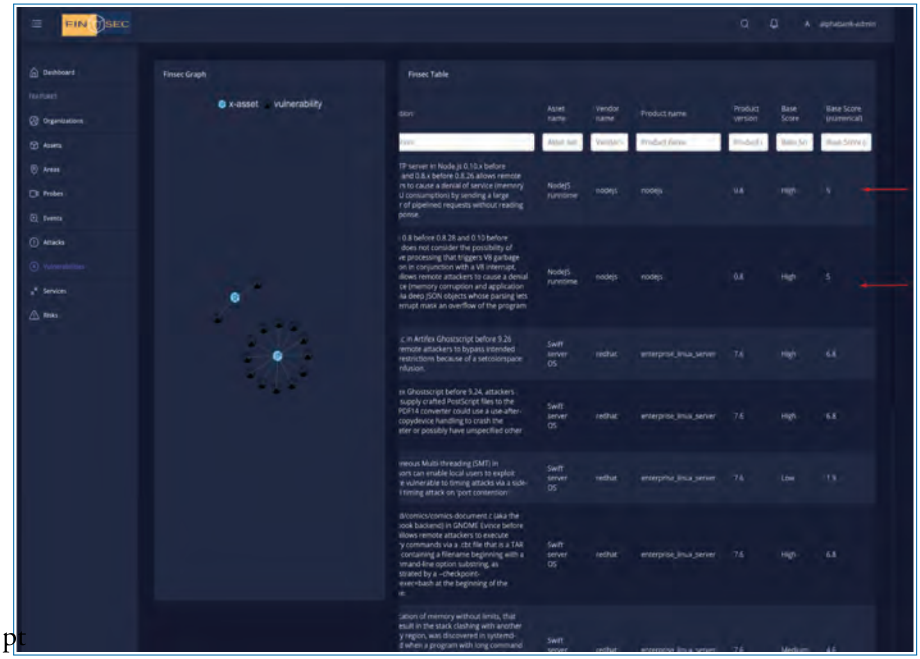
Figure 5.11. FINSEC Dashboard homepage—vulnerability categorization.



Finsec Table

Description	Asset name	Vendor name	Product name	Product version	Base Score	Base S (nume
The HTTP server in Node.js 0.10.x before 0.10.21 and 0.8.x before 0.8.26 allows remote attackers to cause a denial of service (memory and CPU consumption) by sending a large number of pipelined requests without reading the response.	NodeJS runntime	nodejs	nodejs	0.6	High	5
Node.js 0.8 before 0.8.28 and 0.10 before 0.10.30 does not consider the possibility of recursive processing that triggers V8 garbage collection in conjunction with a V8 interrupt, which allows remote attackers to cause a denial of service (memory corruption and application crash) via deep JSON objects whose parsing lets this interrupt mask an overflow of the program stack.	NodeJS runntime	nodejs	nodejs	0.8	High	5
psi/zicc.c in Artifex Ghostscript before 9.26 allows remote attackers to bypass intended access restrictions because of a setcolorspace type confusion.	Swift server OS	redhat	enterprise_linux_server	7.6	High	6.8
In Artifex Ghostscript before 9.24, attackers able to supply crafted PostScript files to the builtin PDF14 converter could use a use-after-free in copydevice handling to crash the interpreter or possibly have unspecified other impact.	Swift server OS	redhat	enterprise_linux_server	7.6	High	6.8
Simultaneous Multi-threading (SMT) in processors can enable local users to exploit software vulnerable to timing attacks via a side-channel timing attack on 'port contention'.	Swift server OS	redhat	enterprise_linux_server	7.6	Low	1.9
backend/comics/comics-document.c (aka the comic book backend) in GNOME Evince before 3.24.1 allows remote attackers to execute arbitrary commands via a .cbt file that is a TAR archive containing a filename beginning with a "--" command-line option substring, as demonstrated by a --checkpoint-action=exec=bash at the beginning of the filename.	Swift server OS	redhat	enterprise_linux_server	7.6	High	6.8
An allocation of memory without limits, that could result in the stack clashing with another memory region, was discovered in systemd-						

Figure 5.12. Vulnerabilities—auto-imported from the security knowledge base.

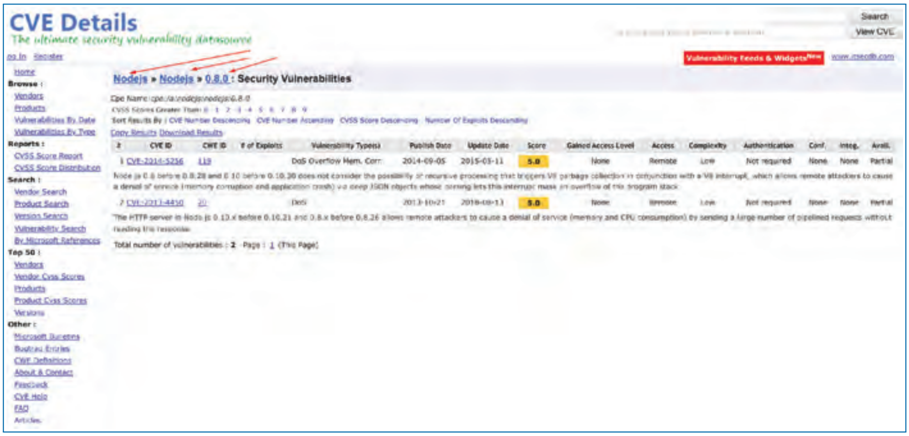


Finsec Table

tion	Asset name	Vendor name	Product name	Product version	Base Score	Base Score (numerical)
TP server in Node.js 0.10.x before and 0.8.x before 0.8.26 allows remote rs to cause a denial of service (memory U consumption) by sending a large r of pipelined requests without reading ponse.	NodeJS runntime	nodejs	nodejs	0.8	High	5
0.8 before 0.8.28 and 0.10 before does not consider the possibility of ve processing that triggers V8 garbage on in conjunction with a V8 interrupt, llows remote attackers to cause a denial ce (memory corruption and application ia deep JSON objects whose parsing lets errupt mask an overflow of the program	NodeJS runntime	nodejs	nodejs	0.8	High	5
.c in Artifex Ghostscript before 9.26 remote attackers to bypass intended restrictions because of a setcolorspace nfusion.	Swift server OS	redhat	enterprise_linux_server	7.6	High	6.8
ex Ghostscript before 9.24, attackers supply crafted PostScript files to the PDF14 converter could use a use-after-copydevice handling to crash the eter or possibly have unspecified other	Swift server OS	redhat	enterprise_linux_server	7.6	High	6.8
neous Multi-threading (SMT) in sors can enable local users to exploit e vulnerable to timing attacks via a side-l timing attack on 'port contention'.	Swift server OS	redhat	enterprise_linux_server	7.6	Low	1.9
d/comics/comics-document.c (aka the book backend) in GNOME Evince before llows remote attackers to execute y commands via a .cbt file that is a TAR containing a filename beginning with a mmand-line option substring, as strated by a --checkpoint-exec=bash at the beginning of the ie.	Swift server OS	redhat	enterprise_linux_server	7.6	High	6.8
cation of memory without limits, that esult in the stack clashing with another y region, was discovered in systemd-d when a program with long command	Swift server	redhat	enterprise_linux_server	7.6	Medium	4.6

pt

Figure 5.13. NodeJS vulnerabilities detected.



Nodejs » Nodejs » 0.8.0 : Security Vulnerabilities

Cpe Name: cpe:/a:nodejs:nodejs:0.8.0

CVSS Scores Greater Than: 0 1 2 3 4 5 6 7 8 9

Sort Results By : CVE Number Descending CVE Number Ascending CVSS Score Descending Number Of Exploits Descending

Copy Results Download Results

#	CVE ID	CWE ID	# of Exploits	Vulnerability Type(s)	Publish Date	Update Date	Score	Gained Access Level	Access	Complexity	Authentication	Conf.	Integ.	Avail.
1	CVE-2014-5256	119		DoS Overflow Mem. Corr.	2014-09-05	2015-03-11	5.0	None	Remote	Low	Not required	None	None	Partial

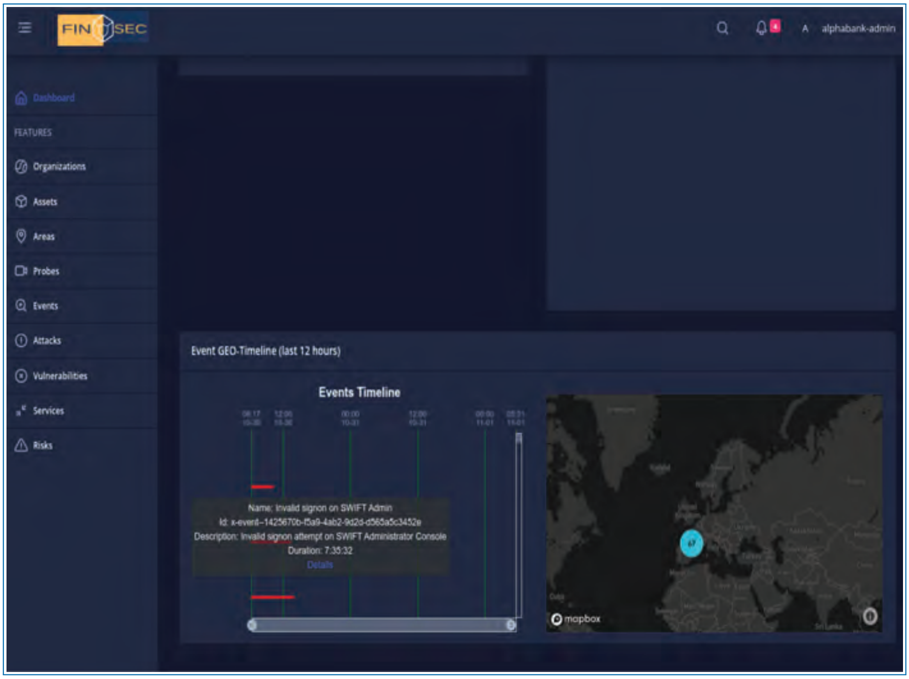
Node.js 0.8 before 0.8.28 and 0.10 before 0.10.30 does not consider the possibility of recursive processing that triggers V8 garbage collection in conjunction with a V8 interrupt, which allows remote attackers to cause a denial of service (memory corruption and application crash) via deep JSON objects whose parsing lets this interrupt mask an overflow of the program stack.

#	CVE ID	CWE ID	# of Exploits	Vulnerability Type(s)	Publish Date	Update Date	Score	Gained Access Level	Access	Complexity	Authentication	Conf.	Integ.	Avail.
2	CVE-2013-4450	20		DoS	2013-10-21	2018-08-13	5.0	None	Remote	Low	Not required	None	None	Partial

The HTTP server in Node.js 0.10.x before 0.10.21 and 0.8.x before 0.8.26 allows remote attackers to cause a denial of service (memory and CPU consumption) by sending a large number of pipelined requests without reading the response.

Total number of vulnerabilities : 2 Page : 1 (This Page)

Figure 5.14. CVE vulnerabilities cross check.

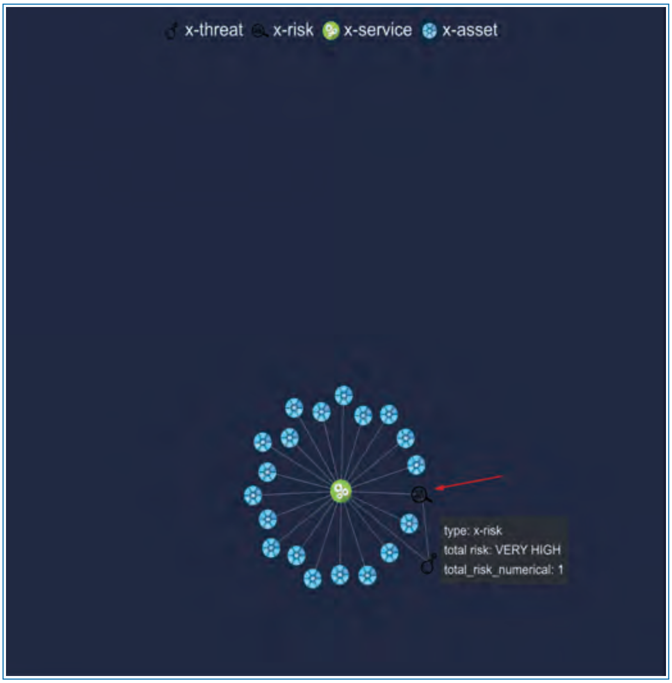


Name: Invalid signon on SWIFT Admin
Id: x-event--1425670b-f5a9-4ab2-9d2d-d565a5c3452e
Description: Invalid signon attempt on SWIFT Administrator Console
Duration: 7:35:32
Details

Figure 5.15. Probe events detected.

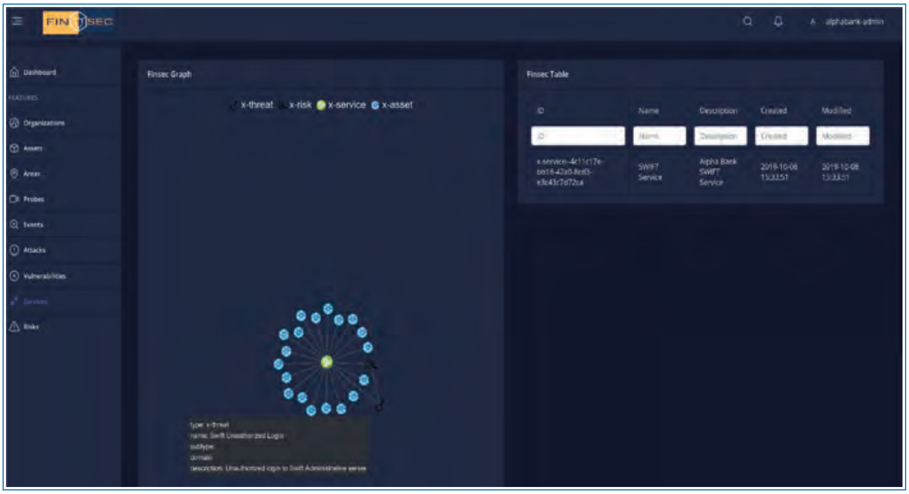
Figure 5.15 illustrates the Probe events detected. For the specific SWIFT service scenario, they are both the "Invalid Signon Attempt" and the "Submission of SWIFT messages outside working hours."

One notification is displayed on the upper right corner as soon as a risk value is changed. The risk value calculated for the SWIFT service and especially the SWIFT



type: x-risk
total risk: VERY HIGH
total_risk_numerical: 1

Figure 5.16. Risk results—graphical representation of the service generated.



type: x-threat
name: Swift Unauthorized Login
subtype:
domain:
description: Unauthorized login to Swift Administrative server

ID	Name	Description	Created	Modified
ID	Name	Description	Created	Modified
x-service--4c11c17e-bb18-42a0-8ed3-e3c43c7d72ca	SWIFT Service	Alpha Bank SWIFT Service	2019-10-08 15:33:51	2019-10-08 15:33:51

Figure 5.17. Threat identified for the SWIFT service.

Service Compromise Threat due to “Invalid Signon Attempt” events produced by the Syslog Probe is provided in Figure 5.16.

SWIFT Service details are illustrated in Figure 5.17. Both a table detail view and a relation graph are available.

The incidents related to the Compromise of the SWIFT service were successfully detected for both use cases without providing false positives.

5.7 Conclusions

This Chapter extended the approach introduced in [2] for sharing security information across financial organizations, towards enabling collaborative security in the financial services supply chain. In particular, it described a blockchain infrastructure, as a means of leveraging the advantages of auditability, security, and distributed trust offered by distributed ledger technologies. The blockchain infrastructure is appropriately integrated to a wider platform for financial services security, which is destined to protect both cyber and physical assets. In particular, this Chapter has introduced the extended collaborative risk assessment functionalities of the platform as well as the platform's security knowledge base. Then, drawing an example of the behavior of a logged in security officer, it has demonstrated the functionality of the user interface and dashboard.

Acknowledgments

Part of this work has been carried out in the scope of the FINSEC project (contract number 786727), which is co-funded by the European Commission in the scope of its H2020 program. The authors acknowledge valuable help and contributions from all partners of the project.

References

[1] Rinaldi, S.M., Peerenboom, J.P., and Kelly, T.K., Identifying, Understanding and Analyzing Critical Infrastructure Interdependencies, IEEE Control Systems Magazine, 11–25 (2001).

[2] Karagiannis I., Mavrogiannis K., Soldatos J., Drakoulis D., Troiano E., and Polyviou A. Blockchain Based Sharing of Security Information for Critical Infrastructures of the Finance Sector. In: Fournaris A. *et al.* (eds.) Computer Security. IOSEC 2019, MSTEC 2019, FINSEC 2019. Lecture Notes in Computer Science, vol. 11981. Springer, Cham, (2020).

[3] Maurer, T., Levite, A., and Perkovich, G. Toward a global norm against manipulating the integrity of financial data. Economics Discussion Papers, 38, Kiel Institute for the World Economy, Kiel, Germany (2017).

[4] Hussain, K. and Prieto, E. Big Data in the Finance and Insurance Sectors. In: Cavanillas J., Curry E., Wahlster W. (eds.) New Horizons for a Data-Driven Economy. Springer, Cham. (2016).

[5] European Union Agency for Network and Information Security. Safeguarding the Global Financial System by Reducing Cyber-Risk, Heraklion, Greece (2016).

[6] Financial Services Information Sharing and Analysis Center, https://www.fsisac.com/, last accessed 2019/07/09.

[7] Ntouskas, T. and Polemi, N. A Secure, Collaborative Environment for the Security Management of Port Information Systems. 5th International Conference on the Internet and Web Applications and Services, pp. 374–379, IEEE Press, Barcelona, Spain, (2010).

[8] Theoharidou, M., Kandias, M., and Gritzalis, D. Securing Transportation-Critical Infrastructures: Trends and Perspectives. 7th IEEE International Conference in Global Security, Safety and Sustainability, pp. 171–178, Springer, Greece, (2011).

[9] Kampanakis, P. Security Automation and Threat Information-Sharing Options. IEEE Security & Privacy, 12(5), 42–51, (2014).

[10] European Network and Information Security Agency. Inventory of Risk Management/Risk Assessment Methods. rm-inv.enisa.europa.eu/rm_ra_methods.html, last accessed 2019/07/09.

[11] Ekelhart, A., Neubauer, T., and Fenz, S. Automated Risk and Utility Management, In: 6th International Conference on Information Technology: New Generations, IEEE Computer Society, 393–398, Las Vegas, NV, USA (2009).

[12] Jordan, B., Piazza, R., and Wunder, B. Stix Core Concepts. https://docs.oasis-open.org/cti/stix/v2.0/stix-v2.0-part1-stix-core.html, last accessed 2019/07/09.

[13] Wunder, J., Davidson, M., and Jordan, B. TAXII^{TM} Version 2.0. Retrieved from https://docs.google.com/document/d/1Jv9ICjUNZrOnwUXtenB1QcnBLO35RnjQcJLsa1mGSkI/edit, last accessed 2019/07/09.

[14] Parameswaran, M., Susarla, A., and Whinston, A.B. P2P Networking: An Information-Sharing Alternative, Computer, 34(7), 31–38, 2001.

[15] Bosco, F., Croce, V., and Raveduto G. Blockchain Technology for Financial Services Facilitation in RES Investments. In: 4th International Forum on Research and Technology for Society and Industry (RTSI), pp. 1–5, IEEE Italy Section, Palermo, Italy, (2018).

DOI: 10.1561/9781680836875.ch6

Chapter 6

Automated Assistance to the Security Assessment of API for Financial Services

By Andrea Bisegna, Roberto Carbone, Mariano Ceccato, Salvatore Manfredi, Silvio Ranise, Giada Sciarretta, Alessandro Tomasi and Emanuele Viglianisi

This chapter presents the challenges related to the security assessment and the automated synthesis of mitigation measures of APIs for financial services. The focus is on the APIs supporting the implementation of the new Payment Services Directive [PSD2]. It also gives an overview of an innovative approach to address these challenges by (i) the automated identification and mitigation of security misconfigurations underlying sessions based on Transport Layer Security [TLS], which is ubiquitously used to build a foundation layer of security; and (ii) the automated penetration testing and synthesis of mitigations for the functionalities provided by APIs built on top of it, both business (e.g., payments) and security (e.g., authentication or authorization). The main novelty of the proposed approach lies in the tight integration of identification and mitigation phases by means of actionable measures that allow users to significantly strengthen the security posture of the entire API ecosystem.

The Regulatory Landscape

The Electronic Identification, Authentication and Trust Services [eIDAS] Regulation is the keystone regulation that defines requirements granting legal validity throughout the internal market to electronic transactions, equivalent to previous paper-based documents. To that end, it regulates Qualified Certificates (QC), electronic seals and signatures, and trust service providers. Security guidelines for the appropriate use of QCs have been published by ENISA QTS [ENISA QTS].

The Revised Directive on Payment Services [PSD2] is intended to protect and promote competition in the internal market by mandating that Account Servicing Payment Service Providers (ASPSP)—most likely traditional banking institutions—open their services to Third-party Providers (TPP) of Services including account information (AISP) and payment initiation (PISP) providers.

The Regulatory Technical Standard [RTS] defines requirements on the use of QCs for website authentication and electronic seals for communication among TPPs and ASPSPs. Guidance on the use of QCs is included in [EBA-OP-2018-7]. The [ETSI TS 119 495] standard defines how to implement the requirements of the RTS for use of QCs to meet the regulatory requirements of PSD2. For instance, it defines the requirements for Qualified Website Authentication Certificates (QWACs), and it clarifies specifically that a QWAC "should be used to establish a secure TLS channel to protect the communication (in the transport layer) from potential attackers on the network."

Open Banking API Security Recommendations

Under PSD2, banks are to provide an interface for third parties to access account information and perform operations (e.g., payments) on behalf of the account holder. The regulation does not specify technical solutions.

The Berlin Group standards and harmonization initiative proposes several possible approaches in its detailed "Access to Account (XS2A) Framework," including XML/JSON data model and associated messaging, as well as OpenAPI files to assist developers with implementation. At its core, XS2A provides a detailed description of REST API and their usage for the purposes of authentication of involved parties and authorization to access Service resources, such as Account Information (AIS), Payment Initiation (PIS), and Confirmation of Funds (PIIS).

The security of these APIs is based on both the transport and application layers. The first core technology explicitly identified by the guidelines is the Transport Layer Security [TLS] protocol: in particular, "the communication between the TPP and the ASPSP is always secured by using a TLS-connection using TLS version 1.2

or higher."[1] [XS2A-IG]. Additionally, [XS2A-IG] requires mutual authentication of TPP and ASPSP using eIDAS- and RTS-compliant QCs, which must include all the roles for which the TPP is authorized.

On the application layer, the core technology for authorization is the Open Authorization Protocol [OAuth2], in particular the "Authorisation Code Grant" flow is mandated for PIS and AIS. While other options are available and discussed below, OAuth is seen as preferable.

Strong Customer Authentication in XS2A

Strong Customer Authentication (SCA) is one of the main requirements set out by PSD2 (article 97) and RTS (Chapter III). The ASPSP must determine how to enforce SCA on a per-transaction basis, in compliance with those requirements.

In the XS2A framework, TPPs have three broad categories of options to allow compliance with SCA requirements:

1. Redirection—of users to their account holders and back to the TPP—using an authentication solution based on, e.g., OAuth 2, such as [OIDC];
2. Decoupling, in which the communication between user and account holder proceeds on an entirely separate channel; and
3. Embedding, in which the TPP has to embed the PSP's entire SCA flow in their own app.

Approach 3 involves a deep level of integration with every single account holder, which is much more work than the other options and requires an extremely high level of trust between the parties as it requires the sharing of user credentials. Approach 2 is more lightweight and scalable but incurs a higher risk of hanging business processes as the TPP must wait for notification of a completed operation on a separate channel. Option 1 is clearly seen as preferable.

Approach	Redirect (OAuth 2)	Decoupled	Embedded
SCA	Directly between user and PSP		Entirely at XS2A interface
Third-party Provider	Does not need detailed information about the individual steps of SCA	No impact on the user/provider interface	Needs SCA details for the user, e.g., displays challenge
Example	Standard interface, e.g., "scope" attribute of authentication request is linked to payment initiation or consent resource	Push notification with payment transaction details to dedicated mobile app or via any other application or device, independent of online banking front-end	Users enter username and password through their browser and are shown a QR code to be scanned

[1] We note that TLS 1.2 is now officially marked as obsolete; TLS 1.3 is the current standard.

Automated Analysis of TLS

Transport Layer Security [TLS] consists of a set of cryptographic protocols designed to provide secure communications over a network. The popularity of TLS has encouraged attackers to find vulnerabilities and develop exploits. The variety of known attacks is the result of (i) maintaining backward compatibility and (ii) evolving use-case scenarios in which TLS is deployed.

One cannot "just deploy" TLS. Setting up a TLS server requires some amount of configuration, including:

- Choosing a set of cipher(s);
- Choosing the versions of TLS to be offered;
- Setting a certificate issued by a trustworthy CA;
- Coping with implementation issues (e.g., vulnerable libraries).

Several tools have been developed to help administrators deploy secure TLS instances. While such tools are quite effective in automatically finding vulnerabilities and issuing warnings about possible attacks, the burden of finding adequate mitigation measures is left to administrators who must first collect information about the identified problem and related fixes. Typically, such information is distributed in several sources ranging from scientific papers to blog posts. Even disregarding the effort to collect enough material to enact a mitigation, administrators should have enough skills to understand the often subtle details and turn the information in a concrete strategy to fix the problem. Additionally, each tool has varying degrees of coverage and does not specify mitigations for the issues identified. In other words, there is a problem in making the tools' reports actionable.

To address these issues, we developed TLS Assistant [MRS19], an open source tool that combines state-of-the-art TLS analyzers with a report system that shows the full set of viable attacks and suggests appropriate mitigations. The tool's architecture is summarized in Figure 6.1. Its goal is to assist an administrator in securing TLS configurations by:

- Detecting TLS and HTTPS misconfigurations;
- Providing
  - ○ A brief attack description;
  - ○ A mitigation description;
  - ○ Mitigation code snippets (for Apache and nginx web server).

We successfully tested the use of TLSAssistant in the deployment of an eIDAS solution based on the new Italian identity cards before its submission for eIDAS notification, discovering that the first release was prone to Lucky 13 [AFP13]

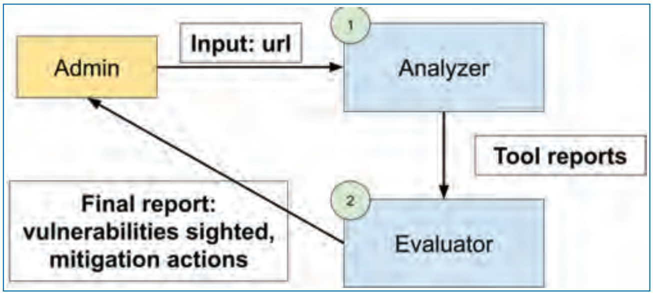


Figure 6.1. TLSAssistant workflow.

and 3SHAKE [BDLFPS14]. The server-side vulnerabilities issues were promptly patched, and the report was judged to be both easy to read and complete.

RESTful API Security Testing

API security issues can have a serious impact on all the applications that depend on them. Indeed, not only is there a growing business for API management [GMQAPI19] but there is a dedicated [OWASP API] top 10 security issue list, of which we highlight "API2:2019 Broken User Authentication" and "API7:2019 Security Misconfiguration." For example, the Harbor enterprise docker container management service was found to expose a "POST /api/users" registration API in which new users could self-register and inject a "HasAdminRole=true" attribute, thereby mounting an escalation of privilege attack remotely on any service exposing this API—see [CVE-2019-16097].

Specifically in the financial sector, a report by TrendMicro [HMcAM19] highlights challenges arising from the new paradigm, for instance, due to the different trust model underpinning the open banking framework. Among several issues, the basic building block of authorization protocols is still a work in progress.

While OAuth 2.0 is arguably the de facto standard for authorization protocols, it is a family of profiles tailored to specific use cases and scenarios. The higher security requirements inherent to the financial sector and the intrinsic novelty of exposing banking APIs to third parties have prompted the establishment of a working group for a dedicated Financial-grade API profile [FAPI], designed to harden OAuth under more adversarial circumstances—for instance, by assuming that sensitive tokens can be leaked by the user's browser or operating system, as is the case for many man-in-the-middle attacks, and allowing for the possibility that API endpoints may be misconfigured. Several mitigations have been proposed, for instance, requiring the use of mutual TLS between third parties and account providers; nevertheless, researchers in [FHK19] found that the expected security properties did not appear to hold in all cases, for instance, allowing malicious actors to force an honest TPP to perform write-like operations (e.g., payment authorizations) from the attacker's device on an honest user's account.

We note that the use of OAuth on its own for authentication is considered improper; the OpenID Connect [OIDC] protocol builds an authentication layer on top of OAuth, and indeed, this is used in FAPI.

Automated Black-box Testing of RESTful APIs

We developed a synthesis of functional and security black-box tests, to appear in [VDC20]. It allows the automatic generation of test cases for RESTful API against errors and vulnerabilities. Indeed, errors can be indicators of potential vulnerabilities that may be exploited to mount attacks.

The tool's architecture is summarized in Figure 6.2. It takes as input an OpenAPI specification, containing all the necessary information to reach the API and the description of the endpoints. The first module generates an Operation Dependency Graph that, together with the Swagger specification, is given as input parameter to the Nominal Tester module in order to test the API's nominal behavior. The Nominal Tester outputs the nominal test cases and a set of structured reports that are given as input to both Error Tester and Security Tester. The former tests the

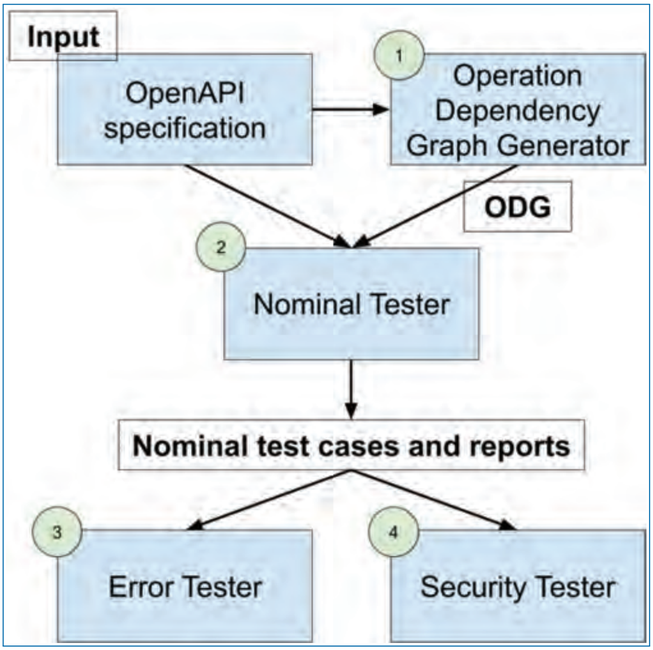


Figure 6.2. Black-box tool workflow.

correct error handling in case of malformed requests, for example, missing required parameters. The latter tests the API against common security vulnerabilities issues, such as SQL injection.

The execution scenarios generated by Nominal Tester, Error Tester, and Security Tester are run in the RESTful-API-under-test and its responses are monitored to spot the presence of programming mistakes, errors, and vulnerabilities. A set of *oracles* are defined to this aim, which check responses across multiple dimensions, such as error status code, data consistency with the OpenAPI specification, syntax and well-formed output data, traces of injection vulnerabilities.

Interesting execution scenarios generated by nominal, error and security testers are output as a set of test cases, consisting of JSON description of steps and java code using swagger codegen, to document and reproduce the issues.

OAuth/OIDC Testing

We also developed a tool for automated OAuth/OIDC penetration testing as a plug-in for the Burp Suite, designed to be integrated in our security training and pen-testing environment Micro-ID-Gym [BCMOPR19]. Our plug-in performs both passive and active tests over the traffic generated during an OIDC flow.

Passive tests do not interfere with the flow itself but analyze the recorded traffic, checking, for instance, standard compliance—whether exchanged messages conform to specifications—and Cross-Site Request Forgery (CSRF) protection—e.g., by correct implementation of Proof Key for Code Exchange (PKCE). Active tests verify the behavior of the endpoints when subject to unexpected, modified, or removed input parameters during the OAuth flow.

The plug-in is built on top of Burp Proxy, a tool which allows testers to intercept all requests and responses and leverages the selenium-webdriver browser automation library. The input is a recorded test track, used as a guide for a selenium instance. The track contains the instructions to guide the selenium driver through an OAuth/OIDC flow. The track can be played back so that a tester may observe whether the browser, controlled by the selenium driver, is performing as expected. The tool is designed to pinpoint the step of the flow in which incorrect behavior has been sighted, and courses of action to mitigate against it are to be integrated.

Summary

Our proposed approach to TLS and API security is one that integrates the generation of actionable intelligence and offers concrete courses of action for the mitigation of vulnerabilities. Our ongoing work includes the integration of TLSAssistant in the FinSec platform, the identification of compliance impacts of identified vulnerabilities, and models for continuous risk assessment. In future work, we aim at extending API testing with new penetration testing functionalities, bundle them to build a set of cooperating security services, and integrate the resulting component in a suitable platform.

Acknowledgments

Black-box and white-box security tests for REST API were developed as part of Teîchos, an EIT Digital Finance project.

TLSAssistant was developed in a joint lab with IPZS and is currently being enhanced and integrated in FINSEC, a H2020 Critical Infrastructure Innovation Action project (Contract Number: 786727), which is co-funded by the European Commission in the scope of its H2020 program.

References

[AFP13] N. J. Al Fardan and K. G. Paterson: "Lucky Thirteen: Breaking the TLS and DTLS Record Protocols." 2013 IEEE Symposium on Security and Privacy, Berkeley, CA, 2013, pp. 526–540, doi: 10.1109/SP.2013.42.

[BDLFPS14] K. Bhargavan, A. Delignat-Lavaud, C. Fournet, A. Pironti and P. Strub: "Triple Handshakes and Cookie Cutters: Breaking and Fixing Authentication over TLS". IEEE Symposium on Security and Privacy 2014: 98–113.

[BCMOPR19] A. Bisegna, R. Carbone, I. Martini, V. Odorizzi, G. Pellizzari and S. Ranise: "Micro-Id-Gym: Identity Management Workouts with Container-Based Microservices." IJISC 8 (1), pp. 45–50, 2019.06.28.

[CVE-2019-16097] NIST National Vulnerability Database: Common Vulnerabilities and Exposures #2019-16097. URL: https://nvd.nist.gov/vuln/detail/CVE-2019-16097

[EBA-OP-2018-7] "Opinion of the European Banking Authority on the use of eIDAS certificates under the RTS on SCA and CSC." URL: https://eba.europa.eu/file/58802/

[eIDAS] "Regulation (EU) No. 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC." URL: http://data.europa.eu/eli/reg/2014/910/oj

[ENISA QTS] "ENISA studies on qualified trust services." URL: https://www.enisa.europa.eu/topics/trust-services/qualified-trust-services

[ETSI TS 119 495] "Electronic Signatures and Infrastructures (ESI); Sector Specific Requirements; Qualified Certificate Profiles and TSP Policy Requirements under the payment services Directive (EU) 2015/2366". V1.4.1, November 2019. URL: https://www.etsi.org/standards-search#page=1&search=TS119495

[FAPI] OpenID Financial-grade API (FAPI) Working Group. URL: https://openid.net/wg/fapi/

[FHK19] D. Fett, P. Hosseyni and R. Kuesters: "An Extensive Formal Security Analysis of the OpenID Financial-Grade API." Proceedings of S&P 2019, pp. 1054–1072. doi: 10.1109/SP.2019.00067.

[GMQAPI19] Gartner "Magic Quadrant for Full Life Cycle API Management" 2019. URL: https://www.gartner.com/doc/reprints?id=1-1OGPZC68&ct=190905&st=sb

[HMcAM19] F. Hacquebord, R. McArdle, F. Mercês and D. Sancho: "Ready or Not for PSD2: The Risks of Open Banking." TrendMicro, 2019. URL: https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/the-risks-of-open-banking-are-banks-and-their-customers-ready-for-psd2

[MRS19] S. Manfredi, S. Ranise and G. Sciarretta: "Lost in TLS? No More! Assisted Deployment of Secure TLS Configurations." In: DBSec 2019: Data and Applications Security and Privacy XXXIII pp. 201–220. LNCS 11559. doi: 10.1007/978-3-030-22479-0_11. URL: https://stfbk.github.io/tools/TLSAssistant

[OAuth2] "The OAuth 2.0 Authorization Framework." IETF proposed standard. URL: https://tools.ietf.org/html/rfc6749

[OIDC] "OpenID Connect". URL: https://openid.net/connect/

[OWASP API] OWASP foundation Top 10 API security issue list. URL: https://owasp.org/www-project-api-security/

[PSD2] "Directive (EU) 2015/2366 of the European Parliament and of the Council of 25 November 2015 on payment services in the internal market, amending Directives 2002/65/EC, 2009/110/EC and 2013/36/EU and Regulation (EU) No. 1093/2010, and repealing Directive 2007/64/EC." URL: http://data.europa.eu/eli/dir/2015/2366/oj

[RTS] "Commission Delegated Regulation (EU) 2018/389 of 27 November 2017 supplementing Directive (EU) 2015/2366 of the European Parliament and of the Council with regard to regulatory technical standards for strong customer authentication and common and secure open standards of communication (Text with EEA relevance)." URL: http://data.europa.eu/eli/reg_del/2018/389/oj

[TLS] "The Transport Layer Security (TLS) Protocol". IETF proposed standard. URL: https://tools.ietf.org/html/rfc8446 (v1.3), https://tools.ietf.org/html/rfc5246 (v1.2 – obsolete).

[VDC20] E. Viglianisi, M. Dallago, M. Ceccato: "RESTTESTGEN: Automated Black-Box Testing of RESTful APIs." Accepted to appear in ICST 2020 Research Papers.

[XS2A-OR] NextGenPSD2 Access to Account Interoperability Framework – Operational Rules V1.3 2018.12.21. URL: https://www.berlin-group.org/nextgenpsd2-downloads

[XS2A-IG] NextGenPSD2 Access to Account Interoperability Framework – Implementation Guidelines V1.3.4 2019.07.05. URL: https://www.berlin-group.org/nextgenpsd2-downloads

DOI: 10.1561/9781680836875.ch7

Chapter 7

Adaptive and Intelligent Data Collection and Analytics for Securing Critical Financial Infrastructure

By Habtamu Abie, Svetlana Boudko, Omri Soceanu, Lev Greenberg, Aidan Shribman, Beatriz Gallego-Nicasio, Enrico Cambiaso, Ivan Vaccari and Maurizio Aiello

This chapter presents the FINSEC adaptive and intelligent data collection and analytics system for securing critical financial infrastructure. It enhances the intelligent, resilient, automated, efficient, secure, and timely manner the collection and analysis of security-related data for securing cyber-physical financial infrastructure and services. Making security data collection and analysis intelligent and capable of quickly spotting, learning from, and addressing zero-day threats is essential to economizing of resources and accessing the right information at the right time. This is achieved through the configuration of configurable collection probes and the adaptation of different collection strategies. The chapter further addresses how, inter alia, (i) the nature and quality of collected data affects the efficiency and accuracy of methods of attack detection and defense, (ii) the detection capability can be improved by correlating wide-ranging data sources and predictive analytics, (iii) the rate of the data collection at the various monitoring probes is tuned by managing the appropriate levels and types of intelligence and adaptability of security

monitoring, (iv) the optimization of bandwidth and storage of security information can be achieved by rendering adaptiveness and intelligence and by integrating smart security probes and a set of adaptive strategies and rules, and (v) the increased automation is achieved through a feedback loop of collection, detection, and prevention that allows the early detection and prevention of security compromises and consistently makes security analysis more effective.

7.1 Introduction

Cyber-physical attacks are growing rapidly and posing a substantial risk to the stability of the overall financial sector. Attacks are increasing in number, scope, and sophistication, making it difficult to predict their total impact. The nature and frequency of cyber risks have changed rapidly in the directions not anticipated before, and more risk-managers are becoming aware of the value of engaging with Fintech R&D to keep track of new types of attack surfaces and risk management options, such as FINSEC is addressing. Leading security researchers are coming to the same conclusions (e.g., the state-of-the-art 2019 Cyber Risk Outlook report [1]). In this chapter, we look further ahead than this Cambridge report, blending risk policy, risk technology, and risk-management best practice. Our findings include:

- It is essential to track and maintain the security of critical financial infrastructure and services through the collection and analysis of security-related data in an intelligent, resilient, efficient, secure, and timely manner.
- Making security data collection and analysis intelligent and capable of quickly spotting, learning from, and addressing zero-day threats is essential to economizing resources and accessing the right information at the right time through the configuration of data collection probes and the adaptation of different collection strategies.
- The nature and quality of collected data affects the efficiency and accuracy of methods of attack detection and defense.
- The detection and defense capability can be greatly improved by correlating wide-ranging data sources and by predictive analytics.

Adversaries may attack financial services, damage infrastructure, manipulate critical information, therefore causing serious financial loses. Considering the risks of a large-scale financial network system, it is important to calculate not only the risks of separate nodes but also the risks from connections. Furthermore, adaptive attackers will adapt their strategies to the current security situation and to newly deployed countermeasures. Such emerging attacks can become very sophisticated and can be coordinated, persistent, collaborative, or cooperative with specialized attack expertise. Therefore, there is a need to implement adaptive and intelligent data collection and analytics to cope with a constant update of attack vectors.

The amount of data collected by financial organizations to maintain security is growing every day, and these huge amounts of data can no longer be stored efficiently or processed in real time. Therefore, due to a high dimensionality of data collected from cyber-physical systems, a constant growth of data due to improvements and exposure to new vulnerabilities, and a constant update of attack vectors, Deep Learning (DL)-based security models are essential for adaptability and extendibility with the data drift, continuous discovery of new system threats, and vulnerabilities [2]. In this chapter, we present a model for developing an adaptive and intelligent data collection and analytics that adapts the collection rate and storage state configuration to the analytical systems, threats detected by those systems over time, and economizing the cost of collection and storage resources.

The rest of the chapter is organized as follows: Section 7.2 briefly reviews related work. Section 7.3 sets the scene by describing data collection and analytics. Section 7.4 presents the architecture of adaptive and intelligent data collection and analytics and its implementation in the overall FINSEC Reference Architecture (RA) highlighting its peculiar characteristics. Section 7.5 presents the adaptive data collection strategies which are used to economize use of resources and optimize bandwidth and collection rate. Section 7.6 describes the implementation of different modules and the validation of the predictive analytics algorithms for intelligent processing. Finally, Section 7.7 concludes the chapter.

7.2 Related Work

Adaptive data collection refers to the collection of security-related data to improve collection efficiency, ensure collection accuracy, reduce the amount of collected data to minimize the effect of data collection, and automate the data collection by adjusting to different environmental contexts and situations. Several authors address adaptive data collection in different settings.

Lin *et al.* [3] present the design and implementation of an adaptive security-related data collector based on network context in heterogeneous networks, and they used adaptive sampling algorithm to reduce the amount of collected data. The authors argue that sampling methods to collect data and the collection frequency need to be determined according to specific conditions. For instance, if the data variation is large, the collection interval should be reduced, so as to reflect the variation trend of data; and if the data variation is small, the collection interval can be increased, so as to reduce the amount of data collected while ensuring the accuracy of data collection. They propose an adaptive collection frequency adjustment strategy based on predicted variation ratio. They argue that regression algorithms can be used for prediction, such as linear regression, support vector regression (SVR), logistic regression, KNN regression, etc. They further argue that data variation can also be represented by calculating the ratio of predicted accuracy, which is the ratio of the predicted value of the data to the real value of the data. When the predicted value of the data is close to the real value, it indicates that the data variation is small, and when the predicted value of the data is very different from the real value, the data variation is large.

Habib *et al.* [4] investigated self-adaptive data collection and fusion for body sensor networks. Their approach uses an early warning score system to optimize data transmission and estimates in real time the sensing frequency, and uses a data fusion model using a decision matrix and fuzzy set theory. Their adaptive sampling algorithm adapts the sampling rates of sensors to the vital sign dynamic evolution. An adaptive data collection protocol was proposed in [5], which collects periodically sensor readings and prolongs the lifetime of a periodic sensor network. Authors' sampling rate adaptation is based on the similarity between periods of cycles using Euclidean distance measure to adapt its rate of sampling according to the dynamic modification of the monitored environment. An efficient adaptive sampling approach based on the dependence of conditional variance on measurements varies over time as proposed in [6], which adapts sampling rates to the physical changing dynamics and minimizes over-sampling, and improves resource efficiency of the overall network system. An adaptive sampling approach for energy-efficient periodic data collection in sensor networks is proposed [7]. The approach provides each sensor node the ability to identify redundancy between collected data over time, by using similarity functions and allowing adaptive sampling rate.

Ji and Ni [8] present an adaptive data collection method based on the network data correlation and variation routines. Their method selects the data collection in association with network data variation and adjusts collection frequency based on the ratio of the data variation amplitude. It can adjust data collection according to network load to reduce the burden on network bandwidth and processing resources. The frequency adjustment strategy can reduce data collection times when the data vary gently and increase data collection times when the data vary dramatically. Tang and Xu [9] investigate data collection strategies in lifetime-constrained wireless sensor networks. Their objective is to maximize the accuracy of data collected. They developed adaptive update strategies for both individual and aggregate data collections.

Lin *et al.* [10] highlight the challenges posed in collecting security-related data, which indicates relevance to security, safety, privacy, and trust, in the big data era. Their examples of making data collection difficult are due to its 5Vs (volume, variety, value, velocity, and veracity) characteristics and further the

5G networks' characteristics of being heterogeneous, supporting device-to-device, machine-to-machine and other communication technologies, and different networks such as Internet, Mobile Ad hoc Networks, mobile cellular networks and wireless sensor networks. Security-related data fundamentally affects the efficiency and accuracy of attack detection and defense methods. Jing *et al.* [11] survey existing studies about security-related data collection and analytics for measuring the Internet security. They argue that for measuring the security of the internet and detecting the Internet attacks, collecting different categories of data and employing methods of data analytics are essential. A number of surveys of data collection approaches exist [3, 10–15], addressing different settings.

As demonstrated above, there exist many adaptive data collection methods using different strategies. However, few of them are aimed at adaptive multi-layer data collection applying artificial intelligence and deep learning. This chapter addresses adaptive and intelligent multi-layer data collection through the correlation of wide-ranging data sources and predictive analytics to improve the detection capability, the improvement of the quality of collected data that affects the efficiency and accuracy of methods of attack detection and defense, the rendering of adaptiveness and intelligence, and the integration of smart security probes and a set of adaptive strategies and rules. It also addresses the different means for physical and cybersecurity as means of tuning the rate of the data collection at the various monitoring probes.

7.3 Data Collection and Analytics

7.3.1 Requirements

7.3.1.1 Data collection requirements

Before going through data collection in a physical system, one may verify a set of requirements aspects that are identified and summarized below, but more details can also be found in [15]:

- **Efficiency:** On one hand, the collected data should be compact, the unnecessary data that are useless in attack detection should not be collected. On the other hand, the needed data should be collected in a real-time and high-speed manner to decrease the time delay of attack detection.
- **Privacy:** In the data collection process, the sensitive information of some particular data should be protected.
- **Resource consumption:** The consumption of resources including power, memory, and network bandwidth in the process of data collection and data communication should be well considered.
- **Adaptability and Intelligence:** The data collection process should be adaptable to the context of the physical and cyber-world, as well as to the

security context. In particular, the rate of information acquisition/collections, along with the type of data collected, should be adaptable to changing security contexts. Adaptability should be performed in an intelligent way, i.e., towards optimizing the amount of information available for the security task at hand, while ensuring availability of the proper information.

- **Configurability:** To support adaptability and configurability in data collection, the data collection systems to the used in the project (e.g., probes) must be configurable.
- **Automation:** To automate the data collection and adaptation by adjusting to different environmental contexts and situations. Machine Learning (ML) techniques are helpful for implementing automatic adaptable solutions capable of adjusting to new situations and timely reacting in the face of threats and anomalies [16].

The authors [10] specify 13 functional requirements and 5 security requirements, and 9 functional objectives and 6 security objectives, and the relationship between these.

7.3.1.2 Quality attributes for data analytics

The authors in [17] present a systematic review aimed at identifying the most frequently reported quality attributes and architectural tactics for big data security analytic systems. Their findings are twofold: (i) identification of most frequently reported quality attributes and the justification for their significance for big data cybersecurity analytic systems; and (ii) identification and codification of architectural tactics for addressing the quality attributes that are commonly associated with big data cybersecurity analytic systems. The identified tactics include six performance tactics, four accuracy tactics, two scalability tactics, three reliability tactics, and one security and usability tactic each.

- ***Performance*** is a measure of how quickly a system responds to user inputs or other events.
- ***Accuracy*** is a measure to which a system provides the right results with the necessary degree of precision.
- ***Scalability*** is a measure of how easily a system can grow to handle more user requests, transactions, servers, or other extensions.
- ***Reliability*** is a measure of how long a system runs before experiencing a failure.
- ***Usability*** is a measure of how easy it is for people to learn, remember, and use a system.

- ***Interoperability*** is a measure of how easily a system can interconnect and exchange data with other systems or components.
- ***Adaptability*** is a measure of how easily a system adapts itself to different specified environments using only its own functionality.
- ***Modifiability*** is a measure of how easy it is to maintain, change, enhance, and restructure a system.
- ***Generality*** is a measure of the range of attacks covered by a security analytic system.
- ***Privacy assurance*** is the measure of the ability of a system to carry out its business according to defined privacy policies to help users trust the system.
- ***Security*** is the measure of how well a system protects itself and its data from unauthorized access.
- ***Stealthiness*** is the measure of the ability of a security analytic system to function without being detected by an attacker.

7.3.2 Data Sources

Data sources from which security event data are collected include, but are not limited to, network traffic data, firewall logs, web logs, system logs, router access logs, database access logs, and application logs, system statistics, etc. [11].

7.3.3 Data Collection Categories

The following categories of data collection can be distinguished [11].

Packet-level data: A packet consists of a packet header and a packet payload. They are generated when using protocols like TCP, UDP, ICMP, etc. Based on this definition, a classification of these data for detecting DDoS and Worm attacks can be as: Source/Destination IP address, Source/Destination port, Time to live, Timestamp, Packet payload, Packet size, and Number of packets.

Flow-level data: In high-speed networks with rates up to hundreds of Gigabit per second, collection of packet-level data requires expensive hardware. Thus, flow-level data was introduced and can be considered as a stream of packets. The flow-level data is classified into Flow count, Flow type, Flow size, Flow direction, Flow duration, and Flow rate.

Connection-level data: A connection is defined as the aggregated traffic between two IP addresses from the perspective of a specific network. A connection will contain many flows. Thus, a difference between a connection and a flow is the flow does not have size restriction, that is to say, the flow is generated even if a single packet has been exchanged. But, a connection is generated by at least two packets. The connection level data can be divided into the following types: Connection size, Connection duration, Connection count, and Connection type.

Host-level data: This data is collected from a host. This data provide comprehensive knowledge of system events as it records host activities, changes, resource consumption, etc. These changes are widely used in Host-based IDS. We mention in the following two commonly used types of host-level data in attack detection: CPU and Memory usage and Operation log.

7.3.4 Security Probes

Security probes are created to capture and assess the overall security of servers, networks, databases, etc. and to generate events when they find problems, and have the following abilities:

Topology probes: Probes that have the ability to capture network topology, interface, bridge, namespace attributes. Examples include ethtool (a utility for Linux kernel-based operating system for displaying and modifying some parameters of network interface controllers and their device drivers), Network system simulation software (this includes Software-Defined Network or similar software to simulate the real network functions; An example is the Open vSwitch Database management protocol), Simple Network Management Protocol (an Internet Standard protocol for collecting and organizing information about managed devices on IP networks and for modifying that information to change device behavior), Telnet (protocol to provide a bidirectional interactive text-oriented communication facility, which can be used to connect to network equipment and extract management data), Network Interface Filtering Card (a hardware-based probe that can be remotely configured to focus in more detail on selected traffic and/or filter out malicious forms), etc.

Flow probes: Probes that have ability to follow a flow along a path in the topology. Examples include sFlow (sampled flow) (an industry standard for packet export at Layer 2 of the OSI model), Data Plane Development Kit (a set of data plane libraries and network interface controller drivers for fast packet processing, currently managed as an open-source project under the Linux Foundation), libpcap (commonly used packet capture library, which also defines the de facto external format for packets), sCap (a more efficient implementation of the standard libpcap, using shared memory and so-called subzero packet copy), Internet Protocol Flow Information Export (a protocol for exporting Internet Protocol flow information from routers, probes and other devices), NetFlow (a feature of Cisco routers that provides the ability to collect IP network traffic as it enters or exits an interface), Flowmon probe (a hardware-based probe that uses IPFIX protocol), etc.

7.3.5 Predictive Security Analytics for Adaptive Data Collection

Predictive analytics are used to predict security attacks, threats, and anomalies. Based on the predicted security events, mitigation measures can be triggered, for example, to adapt the data collection rate, close a door, etc. It requires constant monitoring, capturing, and processing large amounts of various data. These data is often redundant. Thus, the storing and processing resources are used unnecessary, and the same prediction results can be achieved with significantly less data. It is, therefore, important to develop lightweight predictive data analytics that can give earlier indications about possible cyberattacks based on less data amount and processing. This will allow reducing the amount of collected and processed data while maintaining the required level of threat detection. We need to select the algorithms that give best prediction results and can, therefore, function as a base for the predictive analytics. For this purpose, several machine learning, deep learning (DL), artificial intelligence (AI) algorithms are to be selected and tested using different datasets available online.

DL or deep neural networks are especially relevant for scenarios where massive datasets are collected. One of the principal DL features is the ability of a DL model to adapt to the behavior of systems to previously unseen scenarios in cybersecurity, thus ensuring generalization of the models [2], which is one of the key goals of AI. Trust and explainability are two other important features to ensure trustworthiness of AI-based cyber systems. Recent research in Explainable AI (XAI) successfully showed how deep neural-network-based intrusion detection systems can help in improving user trust [18]. Adversarial learning offers an approach to increase our understanding of these models. Adversarial learning exploits how a DL system can be "fooled" to wrong conclusions. This knowledge strengthens the system against incorrect intrusion detection decisions. Hence, trust of the system is increased and explainability is improved [18].

Berman *et al.* [19] survey DL methods for cybersecurity applications covering a broad array of attack types including malware, spam, insider threats, network intrusions, false data injection, and malicious domain names used by botnets. They discussed the DL architecture and training process for popular and emerging methods ranging from RNNs to GANs and their application to a wide variety of these cybersecurity attack types.

In this Project, we among other things use AI-based (i.e., deep learning mechanisms) predictive analytics that enable us the identification of complex attack patterns.

7.4 Adaptive and Intelligent Data Collection and Analytics

This section first presents the architecture overview and its implementation in the overall FINSEC Reference Architecture (RA) followed by the descriptions of the various features and services. The security of critical financial infrastructure and services must be tracked and maintained through the collection and analysis of security-related data in an intelligent, efficient, secure, and timely manner. Making security data collection and analysis intelligent and capable of quickly spotting, learning from, and addressing zero-day threats is essential to economizing of resources and accessing the right information at the right time through the configuration of configurable data collection probes and the adaptation of different collection strategies.

The nature and quality of collected data affects the efficiency and accuracy of methods of attack detection and defense. The detection capability can thus greatly be improved by correlating wide-ranging data sources and by predictive analytics. Managing appropriate levels and types of intelligence and adaptability of security monitoring is achieved through different means for adaptive data collection and predictive analytics. This is important for physical and cybersecurity as a means of tuning the rate of the data collection at the various monitoring probes. The cyber and physical data need to be correlated taking the latency of communication into account.

7.4.1 Adaptive and Intelligent Data Collection and Analytics Architecture

Figure 7.1 shows the architecture of the multi-layer adaptive and intelligent data collection and analytics, which extends the classical data collection and analytics process that includes data collection, data parse, data analysis, and data processing. The approach makes this process adaptive by introducing feedback control loop and letting the data collection depends on the result of the last data processed. Adaptability refers to how a collection mechanism can adjust to different environmental contexts and situations.

In Figure 7.1, the process modules include **Monitor** (data collector), **Analyser** (data parser & analyser), **Adapter** (data processor), and **Multi-layer Probes** (Implemented FINSEC Probes). The arrow between modules is data flow and control direction.

The FINSEC project integrates smart security probes and a set of adaptive strategies for the multi-layer data collection functionality, which includes rendering adaptiveness and intelligence, optimizing bandwidth and storage of security

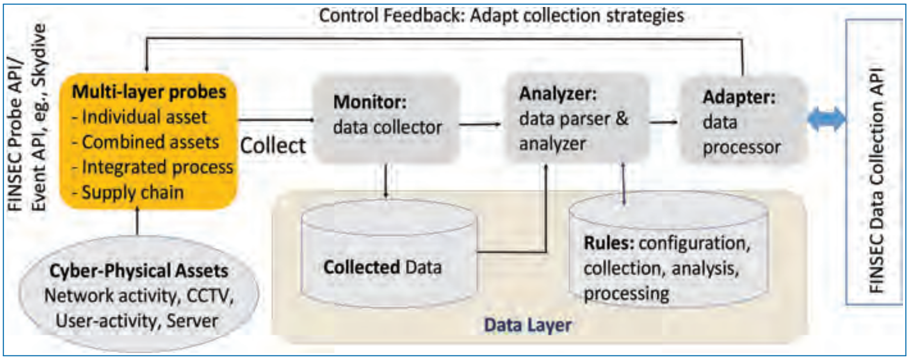


Figure 7.1. FINSEC adaptive and intelligent data collection and analytics architecture.

information, and boosting the intelligence of the probes. Security data analytics methods are integrated in the process at appropriate level-specific analytics. While predictive/regression algorithms such as linear regression, support vector machine (SVM), logistic regression, KNN regression, and Random Forest (classification & regression), K-nearest neighbors, and Decision Tree have been evaluated for the lightweight analysis of adaptive strategies with promising accuracy results of 93%–99%, deep learning mechanisms are under evaluation for the identification of complex risk and attack patterns. These will be described in a later section. A set of rules (both static and adaptive) will be defined for data processing and analysis, configuration, collection, and adaptation.

In the **Multi-layer Probes**, the FINSEC Data Collection API is called by the actual implemented probes, e.g., skydive, to collect data from cyber and physical assets at different levels (individual asset, combined assets, integrated process, and supply chain).

The **Monitor** collects the data using the FINSEC Data Collection API and stores it in the DB at the Data Layer. It analyzes and summarizes the probe data from some probe types and integrates the probes and the Data Layer. The Monitor notifies the Analyser module of collected data.

The **Analyser** module, such as anomaly detection and predictive analytics, analyzes the data and converts the standard data to service data (threats, anomalies, attacks, etc.). Further, it passes the service data to the Adapter module.

The **Adapter** module disposes the service data depending on its value such that it adapts collection strategies and controls the probes through the FINSEC Mitigation API and sends notification to external modules such as alarms and/or data visualization tool or database.

The combinations of Deep Learning algorithms and statistical approaches are utilized to deliver intelligence on anomalies and attacks with the sort of speed to maximize the value of that intelligence. This allows to (i) enhance components of the FINSEC toolbox with more data and predictive security capabilities; (ii) train predictive models running different iterations of different algorithms; (iii) use different models on the same set of data, determine the one that best fits; (iv) establish predictive models to be used for wider use in the financial sector; (v) correlate cyber-physical events and detect cross domain anomalies through pattern detection engine; and (vi) learn typical behavior of the system and detect anomalies through machine learning engine.

The issue of false positives will be addressed to ensure reliability and accuracy. For the quality attribute performance, accuracy, and security & privacy [17, 20, 21], different measures are taken. Performance can be met through ML algorithm optimization, feature selection and extraction, data cutoff, etc. Accuracy can also be improved through alert correlation, combining signature-based and anomaly-based detection, etc. The Security and privacy of the collected and analyzed data is protected through encryption and cross-cutting security services of the FINSEC platform such as authenticity and integrity protection.

7.4.2 Implementation in the FINSEC Reference Architecture

The FINSEC Reference Architecture (RA) provides capability to foster new, intelligent, collaborative and more dynamic approaches to detect, prevent, and mitigate integrated (cyber & physical) security incidents, intelligent monitoring, and data collection of security-related information (the topic of this section); predictive analytics over the collected data; triggering of preventive and mitigation measures in advance of the occurrence of the attack; and allowing all stakeholders to collaborate in vulnerability assessment, risk analysis, threat identification, threat mitigation, and compliance.

Figure 7.2 depicts the implementation of the adaptive and intelligent data collection and analytics architecture with the process modules in the overall FINSEC RA, closing the adaptive loop of **Monitor**, **Analyse**, and **Adapt/Configure** through a feedback control loop. The **Monitor** module maps to the **Data Collection** module in the FINSEC RA, the **Analyser** module maps to the **Predictive Analytics**, **Anomaly Detection,** and **Risk Assessment** services in the FINSEC RA, and the **Adapter** module maps to the **Mitigation** service and **Mitigation Enabler** in the FINSEC RA.

Having data collected with flexible granularity on one hand and with high redundancy on the other allows the correlation of information between locations and layers and the use of various algorithms to produce insights. In this way, increased automation and optimization of bandwidth and storage of security information is achieved using the adaptive collection strategies such as security threats, content variation, collection/sampling rate, bandwidth variation/communication

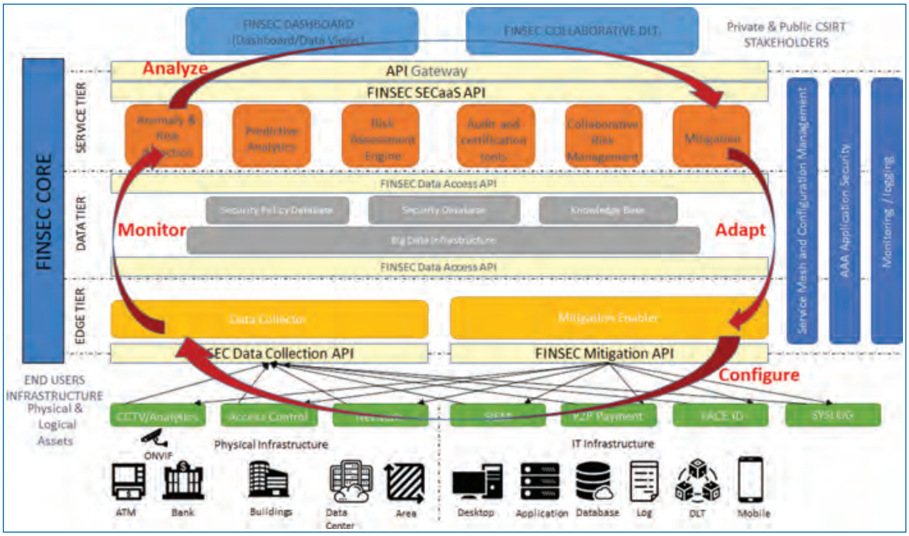


Figure 7.2. Implementation of the adaptive and intelligent data collection and analytics architecture in the FINSEC reference architecture.

dynamics, application needs, context changes, and storage needs. This automation can also be controlled through the FINSEC Dashboard user interface allowing the human in the loop.

7.4.3 Automation Through Predictive Analytics

As mentioned above, the increased automation and optimization of bandwidth and storage of security information is achieved using adaptive data collection strategies such as security threats, content variation, collection rate, bandwidth variation, communication dynamics, application needs, context changes, and storage needs. This, in turn, is achieved through predictive analytics and is achieved at different levels [16]:

- Automation of the data collection, which is inherently automatic in capturing and recording of data for later processing and analysis;
- Automation of the data pre-processing, normalization, and preparation to feed the inputs of the system;
- Automation of the analysis, training, and learning from the collected data, and the detection process;
- Automation of the mitigation process for taking mitigating actions to avoid escalation of the detected anomaly, intrusions, attacks through either passive reaction such as raising alarms or stopping of the system or active reaction such as avoiding system failure.

The FINSEC solution adds another level of automation by tying these automation levels to an overall adaptive and automation level through the feedback control loop (monitor, analyze, and adapt) increasing the automation of monitoring, analyzing, and adapting to the environmental context. This automation and adaptive nature of the FINSEC data collection and analytics allows us to meet quality attributes requirements described in [17, 20, 21] by adjusting its collection mechanism to different environmental contexts and situations, which are termed adaptive data collection strategies and will be described in Section 7.5.

7.4.4 FINSEC Multi-layer Security Probes

The FINSEC probes implemented for data collection and analytics are CCTV probe, Access Control probe, Network Skydive probe, SIEM probe, P2P Payment probe, FaceID probe, and Syslog/App Login as shown in Figure 7.2. This section describes these in brief.

7.4.4.1 CCTV probe

The CCTV probe monitors CCTV, analyzes movements, and detects physical events that may cause threats. The analytics service produces events coming from observations of physical interactions by CCTV.

7.4.4.2 Access control probe

The Access Control probe correlates cyber-physical events by checking the access to a secured area by both the use of a badge and a fingerprint and the state change signaled by movement sensors, vibration sensors, gas sensors, and temperature sensors. Data access events indicate legitimate authentication through HID (Human Interface Devices) readers and fingerprint readers.

7.4.4.3 Network Skydive probe

Skydive is an open source real-time network topology and protocols analyser. It provides real-time insights on network activity which can be used for anomaly detection. It provides agents that act as data collectors, employing efficient mechanisms to control the granularity of data collected and collection intrusiveness, which ensure minimal CPU, memory and network overheads on the monitored system. These mechanisms allow for extra flexibility in capturing network topology and network flow data, as compared to other existing tools. The challenge is to efficiently collect data with minimal disturbance to the production workloads. This includes memory and CPU but also the network itself that is shared in some level between the monitoring and data acquisition tooling and the production workloads. In addition to the common methods, sFlow, netFlow, pcap, etc., a modern advanced networking infrastructure for Host level capturing known as bpf and eBPF is utilized. Those capturing methods make use of Linux Kernel and outperform legacy methods in a wide range of scenarios. With ebpf/bpf capturing, it is possible on one hand to limit and slice the networking data captured to some defined value, and even to change dynamically the capture to fit to on-going security demands and on the other hand allow much more efficient capturing that required significantly less CPU and Memory. All this optimization is achieved through configuring and re-configuring of the frequency of data collection based on different adaptive strategies. This is achieved using the probe configuration data model.

The Skydive probe is composed of Skydive Agents that collect topological information (the Hosts, Switches and NICs (Network Interface Controllers) in the system) and flow information (the L3 traffic streams; using powerful protocols analyzers to understand the traffic). This information is reported by the Skydive Agents to a Skydive Analyzer which aggregates the information at the cluster level and stores it in a time-series database. Figure 7.3 provides a multi-layer Skydive probe architecture.

The Skydive Analyzer exposes the real-time Flow information via a WebSocket which enables construction of Export pipelines. It processes these flows (transforming, encoding, compressing, and storing) and thus facilitates the construction of analytical tools that consume Skydive flow information.

The FINSEC Skydive Adapter (also implemented in Python) pushes network data as observed data to the data collector layer by performing the following steps:

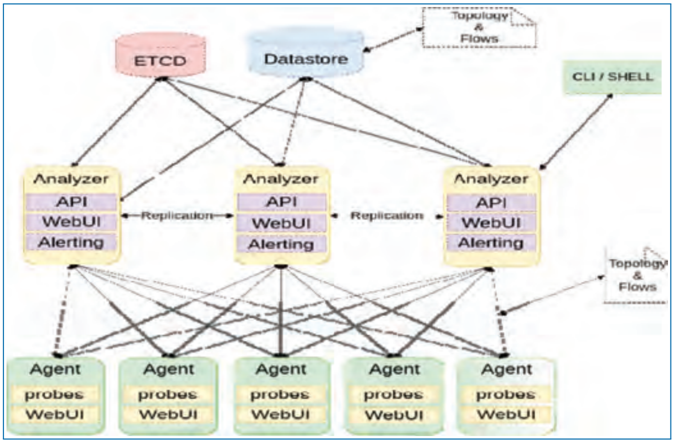


Figure 7.3. Multi-layer Skydive probe architecture.

- Classify flows according to traffic type (internal, ingress, egress, unknown)
- Reformat flows to FINSTIX (FINSEC Data Model)
- Submit flows to data-collector layer

7.4.4.4 SIEM probe

Security Information and Event Management (SIEM) systems have been used in IT since long ago to guarantee security in computer transactions and technological environments. SIEMs collect information about the monitored IT system by using agents deployed close to the infrastructure elements. This information is encapsulated in the form of events, stored and correlated to identify anomalous behaviors, discover possible threats, and detect security incidents. This way the SIEM offers a security administrator a view of the security status and of the activity that is going on in the monitored system.

In FINSEC, the SIEM probe is based on the XL-SIEM (Cross-Layer SIEM) tool developed by Atos [22], which produces alarms by correlating events received from different sources to offer extended information to other components. The event sources are typically application logs and sensors such as HIDS (Host Intrusion Detection Systems), NIDS (Network Intrusion Detection Systems), and AntiVirus.

7.4.4.5 P2P Payment probe

The P2P Payment Probe includes the following three modules that contribute the following features to the FINSEC platform: The P2P Pay module monitors and collects data of peer-to-peer payments sent on Blockchain infrastructure by end users via their commercial banks; The Block chain module monitors and collects Blockchain infrastructure parameters useful for anomaly detection on payments sent on Blockchain and Blockchain itself; and The Actuation module provides a web service interface to send specified events and commands to P2P Payment probe.

7.4.4.6 FaceID probe

The FaceID probe is two factors identification probe that combines physical level (face recognition) and credential entering to authenticate users.

7.4.4.7 Syslog/App Login

Syslog Probe analyzes the logs generated by the internal Bank monitoring infrastructure. It is installed inside the Bank premises in a virtual machine with access restrictions to users and software that can be added.

The responsibilities of the Syslog probe are to send initial information to the data collector with the FINSTIX x-assets, x-probes, x-probe-configurations, to monitor a local database which stores in near real time all the syslog events provided by the Bank's internal monitoring infrastructure, to filter and analyze records received from the Syslog, and to generate corresponding x-event and observed-data FINSTIX objects based on a set of rules; and the events generated are related to a predefined threat providing the collaborative risk module the ability to perform risk calculations.

7.5 Adaptive Data Collection Strategies

The FINSEC data collection strategies are based on security threats, content variation, collection rate, bandwidth variation, application needs, communication dynamics, and environmental context change which all are addressed in the ensuing sections.

7.5.1 Content Variation and Security Threats

To adapt the collection rate to content variations, FINSEC will implement the adaptive sampling rate algorithm that is defined and presented in [7]. The algorithm uses a score for sets similarity, which is defined in this study. The algorithm computes the similarity between datasets collected during successive slots of monitoring. Further, the amount of the redundant data is determined based on the similarity score; thus, the size of the data sent for further processing is reduced.

To adapt the collection rate to security threats, the predictive analytics analyzes collected data and predicts security attack, threat, or anomaly. Then, predictive analytics initiates mitigation measure, in this case adaptive data collection strategy via the FINSEC mitigation service. The FINSEC mitigation service instructs the FINSEC Mitigation Enabler to adapt the collection rate. The FINSEC Mitigation Enabler instructs the Field tier probe to re-configure collection rate and the Field tier probe re-configures its collection rate and pushes data accordingly, thus adapting the rate of data collection based on the security context.

7.5.2 Anomaly Detection Driven Data Collection

Figure 7.4 shows a generic anomaly driven adaptive data acquisition approach proposed for the FINSEC platform. It is composed of three components: (1) **Mitigation rules** defined using FINSTIX and stored in the Data layer. These rules will define what events or attacks should trigger probe activations. **Mitigation service** will apply these rules to decide when and which Probe Mitigation API should be called; (2) **Probes Mitigation API** exposed by the probes to control what operations should be performed by probes for the mitigation; and (3) Analytics and probes produce event and attack **mitigation triggers** to trigger mitigation rules.

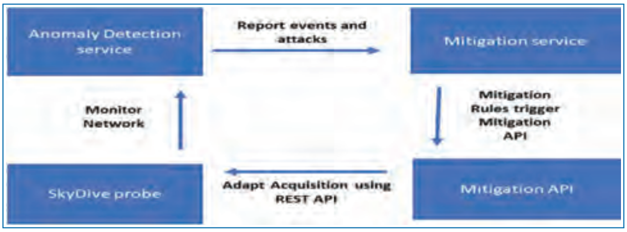

Anomaly Detection service
Report events and attacks
Mitigation service
Monitor Network
Mitigation Rules trigger Mitigation API
SkyDive probe
Adapt Acquisition using REST API
Mitigation API

Figure 7.4. Anomaly driven adaptive data acquisition.

For Skydive probe the above components become: Mitigation rules specify which Skydive probe Actuation APIs should be called for anomalies detected on network data (Network events) or cyber-physical attacks as reported by Anomaly Detection service; Skydive's probe exposes an API to control what types of the netflows should be acquired; and Anomaly Detection service reports network anomalies and cyber-physical events to the Data Layer to trigger adaptive rules (e.g., start acquiring internal traffic).

The adaptive anomaly detection comprises Pattern Detection Engine (PDE), which correlates cyber-physical events and detects cross domain anomalies, and Machine Learning Engine (MLE), which learns typical behavior of the system and detects anomalies on Netflows. The online adaptive training updates models with the most recent observations and gradually "forgets" old behaviors. The Big data Spark-based process aggregates events over time periods and anomaly scores based on the deviation of the observed behavior from the learned models. The platform is modular that can be easily extended with new feature extractors, models, scorers, and pattern detection components.

The adaptive strategies for anomaly driven data collection include more historical data, physical measurement, change of acquisition, and outlier-driven rate of acquisition. The adaptive approach consists of adaptive rules defined using FIN-STIX and stored in the Data layer; Adaptive service applies these rules to decide when and which Probe Activation API should be called, and Probes Activation API exposed by the probes to control how the data acquisition should be adapted.

7.5.3 Enhanced Security Analysis

The Atos XL-SIEM probe has been extended in FINSEC to support adaptive security data collection and this way, enhancing SIEM's security analytics capabilities. With this purpose, a new functional component has been designed, the SIEM Probe Analysis module, which is aimed to be deployed in the FINSEC platform. This module is in charge of analyzing the information received through the FINSEC Data Collector, from the SIEM Probe, and invokes the XL-SIEM Mitigation API to take the necessary adaptive actions. Through this, the SIEM probe can reconfigure itself and the different sensors involved in the data collection, deployed at the target IT infrastructure of the organization, and thus adapts to a new cybersecurity context.

The SIEM Probe Analysis module analyzes FINSTIX data available in the FINSEC platform, together with other relevant Threat Intelligence retrieved from external sources. Two different strategies for FINSTIX data analysis are used:

- **Detection of noisy events to adapt the quantity of events received from the SIEM probe.** This is implemented by creating filtering rules in the SIEM, on-demand, to mute some specific kind of events. This improves the data collection rate in the SIEM probe by lowering down the frequency of periodic non-relevant events. Events are still collected in the SIEM but not reported to the FINSEC platform;
- **Exploitation of IoCs (indicator of compromises) to improve or extend SIEM capabilities to detect security incidents and thus enhance the quality of events received from the SIEM.** The SIEM Probe Analysis module will retrieve IoCs from external sources [e.g., OTX (Open Threat Exchange)], related to events or attacks reported to the FINSEC Platform. IoCs related to suspicious activity already detected in the FINSEC Platform contain valuable and high-quality information that, for instance, an IDS can use to improve or extend their detection capabilities.

7.5.4 Application-driven Innovative Attacks

In the context of the detection algorithms investigated, the focus is on the detection of application layer attacks. Threats like Slow DoS Attacks (SDA) [23], tunneling, and covert channels [24] belong to this category. In the anomaly based intrusion detection topic, after appropriate training on allowed scenarios, a characterization of legitimate conditions is accomplished and used for detection. Particularly, the aim of the algorithm is to monitor and analyze run-time traffic (through on-line or off-line techniques), hence flag as legitimate or anomalous the analyzed traffic.

In order to analyze a potentially anomalous situation, a capture of network traffic is needed to extrapolate predefined representative features able to characterize the considered scenario. If we consider, for instance, Slow DoS Attacks [23], such features may be related to the Delta parameters, extrapolated from network traffic and representing timings used during single connections lives [25]. By using such approach, by considering each Delta parameter, a proper threshold is defined as a consequence of the initial training [26]. The legitimate traffic is characterized to be included under the defined threshold, with a given confident interval. When

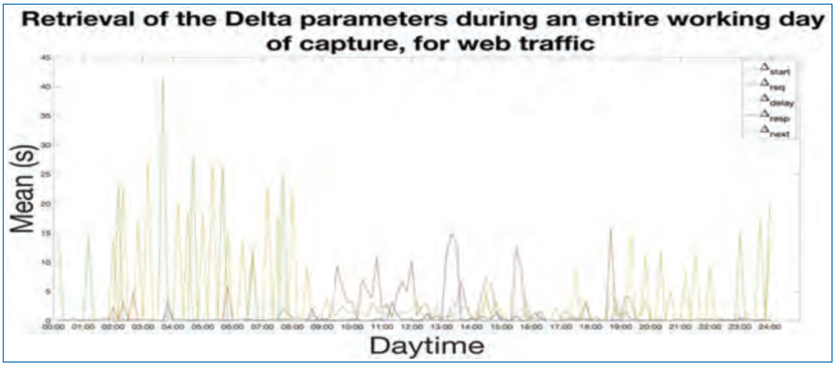


Figure 7.5. Overview of delta parameters dynamic behavior over time, for HTTP traffic between 00:00 and 24:00.

processing run-time traffic, each connection related to a Delta parameter exceeding the defined thresholds will be flagged as anomalous, hence, potentially legitimate.

Although such approach is potentially able to identify run-time threats, in particularly advanced scenarios, a malicious user may attempt to elude the detection system, by modifying the attack to make it behave like a legitimate condition. In this case, if the detection system is not able to refine the calculated Delta thresholds in real time, hence making the Delta threshold assume some sort of "dynamic" behavior, detection may fail, hence expose the system to the attack without triggering any detection.

Figure 7.5 reports the means of the Delta parameters over an entire day, from 00:00 to 24:00, when computing them on a network composed of around 50 nodes, in office environment, for HTTP network traffic.

As can be seen, their values are not static over time, but, instead, they assume some sort of "dynamic" behavior depending on the day time. Such behavior may depend, for instance, on scheduled backup activities executed overnight or on users browsing during office hours. Because of this, a first detection approach is based a dynamic adaptation of the Delta thresholds depending on the time of the day the (potentially anomalous) traffic is captured. By adapting the thresholds through such approach, it is possible to improve the detection of unknown threats, by contextualizing the detection algorithm on the time of the day considered. An extension of this approach may also monitor an entire week of traffic, to also extend the concept to non-working days like Saturday and Sunday, even though the run-time thresholds update activities.

By considering adaptive approaches, in conjunction with the approach described above, it is possible to dynamically enable and disable the network analysis process with a function of the network status. For instance, considering protection from slow DoS threats, it is possible to enable such analysis only when critical conditions are measured. Hence, considering attacks targeting network services, it is possible to adaptively monitor traffic only when the service load exceeds a predefined threshold. This means that in case a network service is under loaded, or a partial DoS [27] is executed, protection may not be enabled, also in view of the application a green approach to cybersecurity [28]. Similarly, adaptive data collection and consequent analysis may therefore be enabled only for the features that characterize specific categories of attacks.

By considering a network platform like FINSEC, the detection algorithm may be represented as the execution of the following steps:

1. The network probe captures information from live traffic.
2. The data collector receives captured information for collection/storage.
3. The data monitor component extrapolates features from collected data.
4. The data analyzer component identifies anomalies/threats
5. The data adapter component re-configures the detection system, involving steps 3 and 4.

By adopting this approach is possible to build an adaptive detection system able to identify cyber threats.

7.6 Implementation and Validation

This section provides a prototype implementation of the adaptive and intelligent security monitoring infrastructure for the FINSEC project and its validation with the anomaly detection example.

In this first phase, a prototype implementation of the adaptive and intelligent security monitoring infrastructure is provided, which covers predictive analytics describing the most relevant approaches to analyze the collected data and detect attack patterns. In addition, the security threat and collection rate strategies are implemented. Various alternative adaptive strategies are also defined: (i) application layer adaptive collection strategies (Request start duration, Request duration, Request management duration, Response duration, and Next request start duration), (ii) adaptive techniques for data acquisition for anomaly detection (More historical data, Physical measurement, Change of acquisition, and Rate of acquisition), and (iii) Adaptive data collection for enhanced security analysis (Data Collection manager for reconfiguring the infrastructure of XL-SIEM agents, Threat intelligence update service, and Adaptive security module, which analyzes the events and alarms generated). The combination of these three architectural elements implements a feedback loop of collection, detection, and prevention that allows for early detection of security compromises and consistently makes security analysis more effective.

7.6.1 Data Collector and Mitigation Enabler

The Data Collector (Monitor) conveys information from the probes to the Data Layer, and it may also perform additional functions for each probe. For the Skydive probe, it summarizes, at a regular interval, all "observed-data" objects seen during this interval and sends this summary to the Data Layer. The summary is created as an "x-collected-data" object, whose structure is fully described in The FINSEC Data Model (FINSTIX). It includes a list of IDs of the summarized objects, a sequence number and a time range bracketing the first and the last observed object. The Data Collector has three endpoints for the Skydive probe, supporting respectively ingress, egress, and internal traffic. Each of these traffic types is treated separately by the Data Collector, so that separate summaries are created for each traffic type, with separate sequence numbering.

In the prototype implementation, the Data Collector receives STIX objects of type "observed-data" from the probes. The Data Collector stores these objects in the Data Layer.

In the case of the Skydive probe in particular, the Data Collector also performs a summarization service of the "observed-data" objects received. Each "observed-data" object contains a set of "x-skydive-flow" objects representing native Skydive flow objects. At a regular, configurable interval (which is 10 minutes by default), the Data Collector sends an "x-collected-data" object to the Data Layer. This object contains a summary of all the "observed-data" objects received from the Skydive probe within the last interval. A separate series of 'x-collected-data' objects is created for every combination of network flow type (ingress, egress, and internal) and organization ID, and every object contains a sequence number within that series. These "x-collected-data" objects are intended to inform the analyzer that new data are available in the Data Layer.

7.6.1.1 Interface to Skydive

Skydive is a real-time network topology and protocol analyzer that can be used to capture network topology and data flows. The Skydive architecture consists of two types of software: agents and analyzers. The purpose of an agent is to collect topology and flow data various types of probes. Thus, an agent needs to be deployed on each computer to be monitored. The purpose of an analyzer is to consolidate the information collected from a set of agents. Only one analyzer is needed, although there may be more than one if redundancy is required.

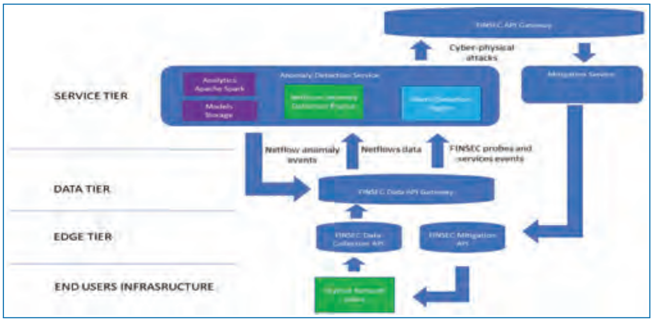


Figure 7.6. Anomaly detection with Skydive probe.

Each analyzer offers two types of interfaces for accessing its functions. The first is a graphical user interface for interactive use of management and monitoring functions. The second is an API that can be integrated with applications. This API is based on the JSON format for exchanging data and the Gremlin language for executing queries on the topology graph.

Figure 7.6 depicts the end-to-end data flow from Skydive probe to Anomaly Detection service. Here are the main steps of the dataflow:

1. The netflow collected by Skydive Network probe are pushed to FINSEC Data Layer through FINSEC Data Collection API as "observed-data" objects.
2. Data Collection service periodically produces "x-collected-data" object that references the "observed-data" objects.
3. Network Anomaly Detection Engine analyzes new "observed-data" objects and reports anomalies as to FINSEC Data Layer as "x-event" object.
4. Alerts Detection Engine correlates reported events according to "x-attack" models and report "x-attack" instances to FINSEC API Gateway
5. Mitigation Service (not implemented yes) will analyze produced "x-events" and "x-attacks" to activate adaptive Mitigation API of Skydive Network probe.

7.6.1.2 STIX and customizations

Structured Threat Information Expression (STIX^{TM}) is a JSON-based language for expressing cyber threat and observable information. A STIX [29] description consists of a set of STIX Domain Objects (SDOs) and a set of STIX Relationship Objects (SROs). The SROs describe relations between the SDOs, forming a graph. In addition to these types of objects, there are also STIX Cyber Observables, which are used by various SDOs to provide additional context to the data that they characterize.

The STIX language can be customized and remain compatible with STIX, as long as certain syntactic rules are observed. In the case of the Data Collector, two custom STIX object types, "x-skydive-flow" and "x-collected-data," were introduced. This was necessary, since Skydive delivers very detailed information on flows in its own JSON-based format, which is incompatible with STIX. Each of these flow descriptions is converted into an "x-skydive-flow" STIX object containing the same structure and the same properties as the Skydive flow object, except that the properties are converted to be compatible with STIX syntax, and "type" and "extensions" properties are added. The "x-collected-data" object type is used to summarize the aforementioned objects.

7.6.2 Predictive Analytics

The general goals of predictive analytics models are to reduce false-positive rates and to deal with a large amount of data for training and prediction, imbalanced datasets, a large number of features, and categorical and continuous features [30]. Random Forest models outperform in achieving these goals due to their advantages of low training time complexity, fast prediction, resilience to deal with imbalanced datasets, embedded feature selection method and intrinsic metrics to rank features by importance, and for their ability to deal natively with categorical and continuous features [30].

To evaluate approaches for the adaptation of the data collection strategies and intelligent processing, we have studied and tested predictive analytics based on machine learning algorithms. At this stage, the following machine learning algorithms have been selected for the predictive analytics toolkit: Support Vector Machine (SVM) using the RBF (Radial Basis Function) kernel method, K-nearest neighbors (KNN), Decision Tree using the Classification and Regression Tree (CART) algorithm, Random Forest, and Multilayer Perceptron (MLP). These algorithms are often applied to solve classification problems. We used the scikit-learn package, Python 3. The PyCharm Integrated Development Environment (IDE) was used for coding. pPckle files have been generated for each model and saved. Furthermore, we explored the possibility of using deep learning algorithms and tested a multi-layer perceptron neural network with 3 layers (on the CICIDS 2017 (Intrusion Detection Evaluation Dataset) dataset mentioned below).

The toolkit has been tested using the datasets KDDCup-99 [31], CICIDS 2017 [32], and UNSW-NB15 [33], which are described below.

The KDDCup99 is a relatively old dataset that was used for "The Third International Knowledge Discovery and Data Mining Tools Competition." The competition's task was to build a predictive network intrusion detector model capable of distinguishing between attacks and normal network traffic. This database contains a standard set of data to be audited, which includes a wide variety of intrusions simulated in a military network environment. All features provided with this dataset have been applied.

The CIC IDS 2017 dataset was created by the Canadian Institute for Cybersecurity. It contains benign traffic and the most up-to-date common attacks. According to the authors, the network traffic analysis was performed using CICFlowMeter with labeled flows based on the time stamp, source, and destination IPs, source and destination ports, protocols and attack (CSV files). Generating realistic background traffic was prioritized. The authors used their B-Profile system (Sharafaldin *et al.* [34]) to profile the abstract behavior of human interactions and generate naturalistic benign background traffic. The dataset is built upon the abstract behavior of 25 users based on the HTTP, HTTPS, FTP, SSH, and email protocols. The CIC IDS 2017 dataset has over 2.83 M examples (2.27 M benign and 557,646 malicious ones) in contrast to KDDCup-99 dataset with 148,517 flows including 77,054 benign and 71,463 malicious ones. For prediction, we used all provided features.

The UNSW-NB15 dataset was created as an IoT dataset in the Cyber Range Lab of the Australian Centre for Cyber Security (ACCS). The authors aimed to generate a hybrid of real modern normal activities and synthetic contemporary attack behaviors. According to the authors, the raw network packets of the UNSW-NB 15 dataset was created by the IXIA PerfectStorm tool. This dataset has nine types of attacks: Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode, and Worms. The authors used the Argus and Bro-IDS tools to generate totally 49 features with the class label. All features that were included in the datasets have been used.

The test results for these three datasets are depicted in Tables 7.1–7.4, respectively. The tests were done using a laptop with Intel(R) Core(TM) i5-5300 U CPU, 2.30 GHz, RAM 16.0 GB, and 64-bit OP.

Table 7.1. Classification results for KDDCup-99 dataset.

	Training Set	Training Time	Testing Set	Testing Time	Accuracy
SVM	3428901	3721.03	1469529	648.56	0.922812634431
KNN	3428901	88.5	1469529	24.94	0.986108253728
Decision Tree	3428901	26.394	1469529	4.74	0.9983703703703
RF (100 estimators)	3428901	712.58	1469529	43.645	0.999948282784
RF (300 estimators)	3428901	2095.318	1469529	120.91274	0.99994896327
RF (500 estimators)	3428901	3424.962	1469529	229.4986	0.999949643763

Table 7.2. Classification results for CIC IDS 2017 dataset, the training set contains 3428901 records, and the testing set contains 1469529.

Algorithm	Training Time	Testing Time	Accuracy	Precision	Recall Score
SVM	2599.515625	226.453125	0.9352371389758432	0.9359399980539957	0.9319887947439143
KNN	3.28125	10.875	0.9977408304293899	0.997765750563484	0.9976316312033666
Decision Tree	2.546875	0.28125	0.9997932786013821	0.9997894281071482	0.9997894281071482
RF (100)	20.140625	0.6875	0.9998375760439431	0.9998122482419608	0.9998569235972009
RF (300)	60.59375	1.546875	0.9998375760439431	0.9998122482419608	0.9998569235972009
RF (500)	98.78125	2.296875	0.999822810229756	0.9997992315023175	0.9998398488439311
Light Gradient Boosting	69.625	1.25	0.9960132301695116	0.9964189831191492	0.9954752405759268
Logistic Regression	27.046875	0.359375	0.943860374461	0.9530305105666308	0.9359238242332494
	Training Set	Training Time	Testing Set	Testing Time	Accuracy
SVM	158021	2921.30	67724	267.26	0.935237138976
KNN	158021	14.32	67724	4.43	0.9977408
Decision Tree	158021	3.3072	67724	0.3744	0.9997637
RF (100 estimators)	158021	32.853	67724	0.9672	0.999948282784
RF (300 estimators)	158021	93.132596	67724	2.19961	0.999837576
RF (500 estimators)	158021	143.73932	67724	3.43202	0.9998228102
MLP (3 layers, 256 initial nodes, 4 epoch)	158021	41.21 (~10 sec. per epoch)	67724	3.11	0.9903431575217058 – I epoch 0.9986858425373575 – II epoch 0.9988335006792275 – III epoch 0.9990402220778454 – IV epoch

Table 7.3. Classification results for UNSW-NB15 dataset.

	Training Set	Training Time	Testing Set	Testing Time	Accuracy
SVM	82332	201.4	17534	37.2	0.720092009961
KNN	82332	14.32	17534	4.43	0.874246715967
Decision Tree	82332	0.1248	17534	0.0468	0.9474364579967
RF (100 estimators)	82332	1.1856	17534	0.234	0.958576507043
RF (300 estimators)	82332	3.2916	17534	0.5928	0.9625738272
RF (500 estimators)	82332	5.2884	17534	0.9984	0.9689476329

This preliminary study has shown that the SVM method has performed inadequately for training/testing time. It has also achieved lower accuracy for the UNSW-NB15 Dataset. We concluded that this method could be excluded from further work stage. The random forest method performs well while requiring slightly more time for training than the decision tree method; the deep learning MLP also performs well, with training time between random forest and the decision tree algorithms, and validation accuracy comparable with both, especially from the second epoch (after which the validation accuracy does not improve much).

In this preliminary study, we have used the datasets that were available online. All features supplied with these datasets have been applied. In the next stage, we plan to define how to select a feature set that produces acceptable results with predefined accuracy while reducing the volume of the collected and stored data. Further, we need to develop methods for predictive analytics that operate on real-time data collections and investigate new efficient predictive algorithms based on deep learning techniques. We, therefore, need to investigate how combining various deep learning approaches can improve the quality of the attack detection.

7.6.3 Anomaly Detector Service

7.6.3.1 Architecture overview

The Anomaly Detection service is composed of External and Internal Anomaly Detection services as depicted in Figure 7.7. The Internal Anomaly Detection service is part of the FINSEC infrastructure, and the External Anomaly Detection service is running outside of the FINSEC infrastructure on the IBM cloud. The External Anomaly Detection service is composed of two analytic engines: Network Anomaly Detection engine and Attack Detection engine.

Other related FINSEC components are the Dashboard, Data Layer, Data Collector, and the Skydive probe. Figure 7.8 shows the data flow between the

Table 7.4. Classification results for UNSW-NB15 dataset, the training set contains 490001 records, and the testing set contains 210000 records.

Algorithm	Training Time	Testing Time	Accuracy	Precision	Recall Score
SVM	2766.109375	346.53125	0.9687191965752544	0.7843639878854836	0.5002233948306303
KNN	17.9375	27.65625	0.9848238817910391	0.9268482036258789	0.802971164004187
Decision Tree	5.03125	0.515625	0.9981571516326113	0.984694679159947	0.9849088498677336
RF (100)	113.265625	3.28125	0.9986047685487212	0.9854760420933983	0.9916943506145978
RF (300)	329.59375	7.921875	0.9985809591382898	0.9849070673636564	0.9919029984302696
RF (500)	542.265625	12.828125	0.9985619116099447	0.9846162034317678	0.9918931670869641
Light Gradient Boosting	144.984375	2.296875	0.9687144346931681	0.48435721734658405	0.5
Logistic Regression	11.828125	0.578125	0.9685287212918033	0.7170305523649205	0.5185364976312337

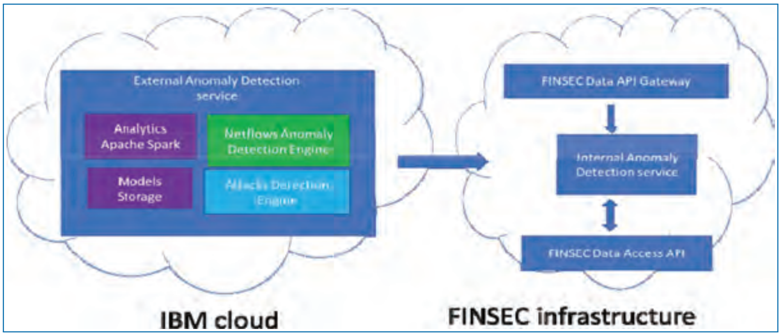


Figure 7.7. External and internal anomaly detection services.

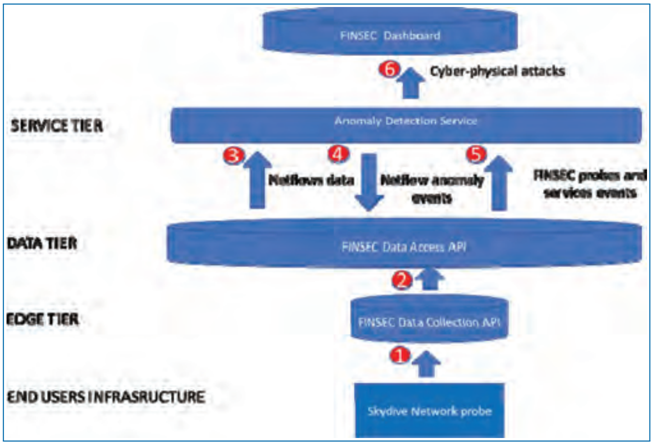


Figure 7.8. Anomaly detection data flow.

above-mentioned components and the Anomaly Detection service, starting from the probe data acquisition and culminating in attack detection and reporting to the dashboard. Described below are the main steps of the data flow:

1. The Netflow data is acquired by the Skydive Network probe and pushed into the Data Collector.
2. The Data Collector aggregates the data and pushes it to the Data Layer.
3. The Netflow data from the Data Layer is processed by the Netflow Anomaly Detection Engine of the Anomaly Detection Service.
4. The Netflow anomaly events detected in the previous step are reported to the Data Layer.
5. Netflow anomaly events along with events produced by other services are analyzed by the Attack Detection Engine.
6. The Cyber-physical attacks that are detected in the previous step are exposed to the FINSEC Dashboard.

7.6.4 SIEM Probe Analysis

As previously introduced in Section 5.3, the Atos XL-SIEM technology has been extended with a new module, the SIEM Probe Analysis module, that supports the implementation of adaptive data collection strategies with the ultimate goal of improving the quality of security events collected and controlling the data collection rate. This module, deployed as a service in the FINSEC platform, works in combination with other services and modules of the XL-SIEM probe running in the field. Figure 7.9 depicts all the elements that compose the XL-SIEM probe adaptive infrastructure and illustrates their intended deployment. The figure also shows the interaction of these elements with other services and components of the FINSEC platform, such as the Data Collector.

On the left hand side of Figure 7.9, Monitored Infrastructure is the target infrastructure under surveillance. This infrastructure is composed by different logical and physical assets such as laptops, servers, routers, printers, and the local area network. These elements are monitored by different typical security sensors or probes such as Host-based Intrusion Detection System (IDS), Network-based Intrusion Detection System (NIDS), or Antivirus (AV), all of them under the control of one or more XL-SIEM agents. XL-SIEM agents are in direct communication with the XL-SIEM probe to send security events or retrieve monitoring configuration updates. XL-SIEM Probe represents the core of the XL-SIEM technology. The Data Collection Manager module, the Data Collection Rules database, and the Configuration Update Service, which will manage the configuration of the remote monitoring components, deployed at the Monitored Infrastructure. This configuration can be updated as a result of an invocation of a specific adaptive action through the XL-SIEM Mitigation API. This API is used by the XL-SIEM to allow modifying the

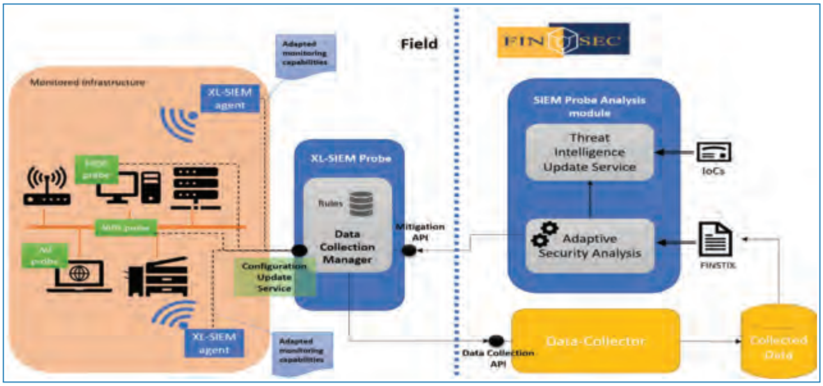


Figure 7.9. Overview of the XL-SIEM probe adaptive infrastructure.

configuration of the XL-SIEM, particularly the configuration of data collection process.

On the right hand side of Figure 7.9, the modules and services are represented which are running under the umbrella of the FINSEC platform. This is the case of the SIEM Probe Analysis module and the Data Collector with the corresponding database to store the collected data. As part of the SIEM Probe Analysis module, the Adaptive Security Analysis (ASA) service is in charge of, first, analyzing the information received in the data-collector from the SIEM Probe and, second, taking decisions on which adaptive strategy to invoke through the XL-SIEM Mitigation API. The Threat Intelligence Update Service (TIUS) supports the ASA service and is responsible for retrieving additional high-quality information about certain security events under analysis. This additional information can be obtained from another FINSEC source of security intelligence, such as the Knowledge Base, or from external sources of IoCs, e.g., Open Threat Exchange (OTX) [35].

Each FINSTIX instance received from the XL-SIEM probe at the Data-Collector is processed at the ASA to extract candidate IoCs from the list of attributes, e.g., URLs, IPs, domains, malware hashes, etc. If the IoC is a public IP, ASA uses the TIUS to consult in the OTX service and returns a list of related "pulses." The TIUS can subscribe the XL-SIEM probe to pulses in order to automatically get new relevant IoCs. The subscription is done only if it is a trusted pulse, i.e., if the number of subscriptions that this pulse already have is above a threshold. Through this process, also known as IoCs Expander, the ASA component can, for example, dynamically update the NIDS (e.g., Suricata [36]) with new rules retrieved from the official NIDS update service (Emerging Threats [37], in the case of Suricata). This way, the XL-SIEM probe adapts to collect additional relevant security events from the monitored infrastructure.

On the other hand, the decision of the ASA after the analysis of the FINSTIX collected data could be to reduce the quantity of events received from the XL-SIEM probe for various reasons, e.g., because the information about a specific IP address is considered not relevant (i.e., it is in a whitelist) or the probe can be instructed to send FINSTIX events wrapping XL-SIEM alarms (high-level correlated data) instead of XL-SIEM events (low level security information). The XL-SIEM probe can be instructed to mute a particular type of event through the invocation of the corresponding method of the Mitigation API. This results in one or more filtering rules created in the XL-SIEM probe. These rules do not prevent the XL-SIEM to generate the event and its corresponding FINSTIX instance but will not send it to the FINSEC Data Collector. This way, the muted events can be recovered upon request at a later point in time if necessary. Filtering rules can be retrieved and removed too, by using the corresponding methods of the Mitigation API.

7.6.5 Innovative Attacks

If we consider the last generation threats, it is important to consider that they may expose characteristics that make them improve their efficacy, compared to old-style threats. If we consider, for instance, the Slow DoS Attacks [23], the focus of our work, compared to old-style flooding threats, the quality of the attack is in this case enhanced, in terms of effects on the system and requirements to the attacker. This is due to the fact that during the execution of an "innovative attack" like the Slow DoS, almost all the packets composing the communication between the attacker and the victim contribute and are important for the success of the attack itself. This means that there is less waste of packets, from the attacker's perspective, compared to old-style flooding attacks, whose approach is to send a huge amount of packets to the victim to attempt to saturate its resources, in case of a slow DoS, a smarter approach is adopted. In virtue of this, reduced attack resources (CPU, memories, bandwidth, etc.) are required.

Considering innovative attacks we have investigated, it is important to consider that the Slow DoS category we have investigated is able to target application layer protocols based on TCP. Known attacks [23, 27] are found in literature for protocols like HTTP, HTTPS, or SMTP. Nevertheless, it is important to consider that the same concept can be adapted and ported to affect different protocols as well. In this case, it may be required to adapt the attack to make it able to target the considered protocol. If we consider, for instance, the MQTT protocol [38], widely used in the machine-to-machine (M2M) context, it may be required to send specific commands like CONNECT (with consequent reception of CONNECT+ACK) messages to perpetrate a long request DoS attack [23]. Preliminary tests executed [39] against a real MQTT service supporting secure communications are shown in Figure 7.10.

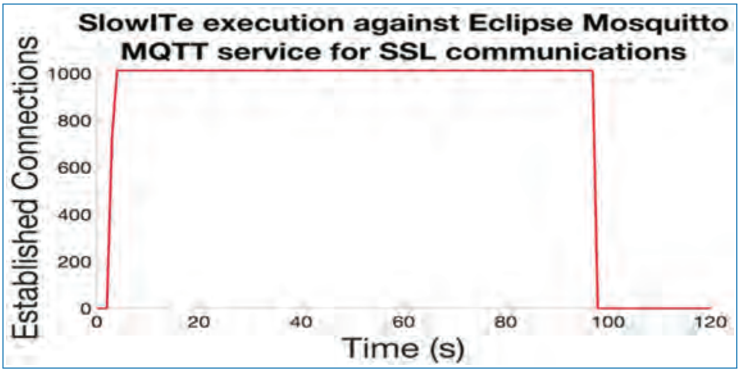


Figure 7.10. Results of tests of the SlowITe slow DoS attack against MQTT service [29].

Figure 7.10 shows that the attack is successful, even on encrypted communications, and it is able to initiate a large number of connections. Tests also reported that the denial of service is reached on the server just after the establishment of 1012 connections that are closed around 90 seconds after their establishment. Although such number of connections may appear high, in this case, since the application layer daemon is targeted, compared to the number of TCP connections, a network host is able to manage (in the order of tens of thousands), such number is considered low. In addition, it is important to consider that no application layer packets are exchanged after the establishment of a connection. Hence, required bandwidth is extremely low. Indeed, we measured that around 340 Kbps were used required for the attack.

In the cyber-security topic, it is therefore important to consider that innovative attacks may create serious damage to the network and its components. Therefore, it is extremely crucial to deploy appropriate monitoring and protection methods and, at the same time, investigate the cyber-security field to acquire knowledge on emerging threats.

Concerning detection from attacks that target specific protocols like MQTT, it is important to consider that efficient detection is still an open issue in research [39], since legitimate clients exploiting such protocols may be characterized by long times of inactivity. This can be also found on SSH protocol, for instance, where connected users may not exchange (at the application layer) any data with the server, even for hours, without experiencing any connection closure. In virtue of this, it is particularly important to investigate the topic and to adapt slow DoS detection algorithms to such kind of “silent” protocols.

7.7 Conclusions

This chapter presented the FINSEC adaptive and intelligent data collection and analytics system for securing critical financial infrastructure. Making the data collection intelligent, resilient, automated, efficient, secure, and timely is essential to economizing of resources, accessing the right information at the right time, and quickly spotting, learning from, and addressing zero-day threats. This is achieved through the configuration of configurable collection probes and the adaptation of different collection strategies. The chapter further addresses how, inter alia, (i) the nature and quality of collected data affects the efficiency and accuracy of methods of attack detection and defense, (ii) the detection capability can be improved by correlating wide-ranging data sources and predictive analytics, (iii) the rate of the data collection at the various monitoring probes is tuned by managing the appropriate levels and types of intelligence and adaptability of security monitoring, (iv) the optimization of bandwidth and storage of security information can be achieved by rendering adaptiveness and intelligence and by integrating smart security probes and a set of adaptive strategies and rules, and (v) the increased automation is achieved through a feedback loop of collection, detection, and prevention that allows the early detection and prevention of security compromises and consistently makes security analysis more effective.

The chapter also presented the adaptive data collection strategies, implementation of the different components of the system, and validation of the predictive analytics algorithms for intelligent processing using publicly available and widely used datasets with promising results. In our future work, we plan to validate the efficiency of all components in real-life use-case scenarios of the FINSEC project.

Acknowledgments

Part of this work has been carried out in the scope of the FINSEC project (contract number 786727), which is co-funded by the European Commission in the scope of its H2020 program. The authors gratefully acknowledge the contributions of the funding agency and of all the project partners.

References

[1] Cyber Risk Outlook, Cambridge Centre for Risk Studies, May 2019, https://www.jbs.cam.ac.uk/fileadmin/user_upload/research/centres/risk/downloads/crs-cyber-risk-outlook-2019.pdf

[2] C. Wickramasinghe, D. Marino, K. Amarasinghe, and M. Manic, "Generalization of Deep Learning For Cyber-Physical System Security: A Survey" in Proc. 44th Annual Conference of the IEEE Industrial Electronics Society, IECON 2018, Washington DC, USA, Oct. 21–23, 2018. PDF, doi: 10.1109/IECON.2018.8591773

[3] H. Lin, Z. Yan, and Y. Fu, Adaptive security-related data collection with context awareness, Journal of Network and Computer Applications, Volume 126, 2019, Pages 88–103, ISSN 1084-8045, https://doi.org/10.1016/j.jnca.2018.11.002

[4] C. Habib, A. Makhoul, R. Darazi, and C. Salim, "Self-Adaptive Data Collection and Fusion for Health Monitoring Based on Body Sensor Networks," in IEEE Transactions on Industrial Informatics, vol. 12, no. 6, pp. 2342–2352, Dec. 2016. doi: 10.1109/TII.2016.2575800

[5] A. Al-Qurabat and A. Idrees (2017). Adaptive Data Collection protocol for Extending Lifetime of Periodic Sensor Networks. Qalaai Zanist Scientific Journal, 2. doi: 10.25212/lfu.qzj.2.2.11.

[6] D. Laiymani and A. Makhoul, "Adaptive data collection approach for periodic sensor networks," 2013 9th International Wireless Communications and Mobile Computing Conference (IWCMC), Sardinia, 2013, pp. 1448–1453. doi: 10.1109/IWCMC.2013.6583769

[7] H. Harb, A. Makhoul, A. Jaber, R. Tawil, and O. Bazzi, (2016). Adaptive data collection approach based on sets similarity function for saving energy in periodic sensor networks. Int. J. Inf. Technol. Manage. 15, 4 (January 2016), 346–363. doi: https://doi.org/10.1504/IJITM.2016.079603

[8] Z. Ji, Z. Kuang and H. Ni, "A Novel Two-Dimension Adaptive Data Collection Method for Network Management," 2009 WRI International Conference on Communications and Mobile Computing, Yunnan, 2009, pp. 237–241. doi: 10.1109/CMC.2009.10

[9] X. Tang and J. Xu, "Adaptive Data Collection Strategies for Lifetime-Constrained Wireless Sensor Networks," in IEEE Transactions on Parallel and Distributed Systems, vol. 19, no. 6, pp. 721–734, June 2008. doi: 10.1109/TPDS.2008.27

[10] H.Q. Lin, Z. Yan, Y. Chen, and L.F. Zhang, "A Survey on Network Security-Related Data Collection Technologies," IEEE Access, vol. 6, issue 1, pp. 18345–8365, Dec. 2018. doi: 10.1109/ACCESS.2018.2817921 (IF: 3.224)

[11] X.Y. Jing, Z. Yan, and W. Pedrycz, "Security Data Collection and Data Analytics in the Internet: A Survey," IEEE Communications Surveys and Tutorials, 2018. doi: 10.1109/COMST.2018.2863942 (IF: 20.23)

[12] R. Erbacher (2008). Steps for Improving Data Comprehension for Digital Security and Forensics. Proceedings of the 2008 International Conference on Security and Management, SAM 2008. 318–326.

[13] D.H. Zhou, Z. Yan, Y.L. Fu, and Z. Yao, "A Survey on Network Data Collection," Journal of Network and Computer Applications, 2018. doi: 10.1016/j.jnca.2018.05.004 (IF: 3.5)

[14] G. Liu, Z. Yan, and W. Pedryczc, "Data Collection for Attack Detection and Security Measurement in Mobile Ad Hoc Networks: A Survey," Journal of Network and Computer Applications, vol. 105, pp. 105–122, March 2018. doi: https://doi.org/10.1016/j.jnca.2018.01.004 (IF: 3.500)

[15] L.M. He, Z. Yan, and M. Atiquzzaman, "LTE/LTE-A Network Security Data Collection and Analysis for Security Measurement: A Survey", IEEE Access, vol. 6, issue 1, pp. 4220–4242, 2018. doi: 10.1109/ACCESS.2018.2792534

[16] L. Cazorla, C. Alcaraz, and J. Lopez (2013). Towards Automatic Critical Infrastructure Protection through Machine Learning. In: Luiijf E., Hartel P. (eds.) Critical Information Infrastructures Security. CRITIS 2013. Lecture Notes in Computer Science, vol. 8328. Springer, Cham.

[17] F. Ullah and M.A. Babar, “An Architecture-Driven Adaptation Approach for Big Data Cyber Security Analytics,” 2019 IEEE International Conference on Software Architecture (ICSA), Hamburg, Germany, 2019, pp. 41–50. doi: 10.1109/ICSA.2019.00013

[18] C. Rieger and M. Manic, On Critical Infrastructures, Their Security and Resilience – Trends and Vision, 2018, https://arxiv.org/pdf/1812.02710.pdf

[19] D. S. Berman, A. L. Buczak, J. S. Chavis and C. L. Corbett, A Survey of Deep Learning Methods for Cyber Security, Information 2019, 10, 122.

[20] F. Ullah and M.A. Babar, “QuickAdapt: Scalable Adaptation for Big Data Cyber Security Analytics,” 2019 24th International Conference on Engineering of Complex Computer Systems (ICECCS), Guangzhou, China, 2019, pp. 81–86. doi: 10.1109/ICECCS.2019.00016

[21] F. Ullah and M.A. Babar (2019). “Architectural Tactics for Big Data Cybersecurity Analytic Systems: A Review,” Journal of Systems and Software, vol. 151, May 2019, Pages 81–118.

[22] G. Gonzalez-Granadillo, S. Gonzalez-Zarzosa and M. Faiella, “Towards an Enhanced Security Data Analytic Platform”. 15th International Conference on Security and Cryptography (SECRYPT), 2018.

[23] E. Cambiaso, G. Papaleo, G. Chiola, and M. Aiello, “Slow dos attacks: definition and categorisation,” International Journal of Trust Management in Computing and Communications, vol. 1, no. 3–4, pp. 300–319, 2013.

[24] M. Aiello, M. Mongelli, E. Cambiaso, and G. Papaleo, (2016). Profiling DNS tunneling attacks with PCA and mutual information. Logic Journal of the IGPL, 24(6), 957–970.

[25] E. Cambiaso, G. Papaleo, G. Chiola, and M. Aiello, “A network traffic representation model for detecting application layer attacks,” International Journal of Computing and Digital Systems, vol. 5, no. 01, 2016.

[26] M. Aiello, E. Cambiaso, S. Scaglione, and G. Papaleo, “Asimilarity based approach for application dos attacks detection,” in 2013 IEEE Symposium on Computers and Communications (ISCC), pp. 000430–000435, IEEE, 2013.

[27] E. Cambiaso, G. Papaleo, G. Chiola, and M. Aiello, “Designing and modeling the slow next dos attack,” in Computational Intelligence in Security for Information Systems Conference, pp. 249–259, Springer, 2015.

[28] L. Caviglione, M. Gaggero, E. Cambiaso, and M. Aiello (2017). Measuring the energy consumption of cyber security. IEEE Comm. Magazine, 55(7), 58–63.

[29] Introduction to STIX, https://oasis-open.github.io/cti-documentation/stix/intro.html

[30] P.A.A. Resende and A.C. Drummond, "A Survey of Random Forest Based Methods for Intrusion Detection Systems", ACM Computing Surveys (CSUR), vol. 51, no. 3, pp. 48, 2018

[31] KDD Cup 1999 Data, http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[32] CICIDS 2017 (Intrusion Detection Evaluation Dataset), https://www.unb.ca/cic/datasets/ids-2017.html

[33] The UNSW-NB15 Dataset Description, https://www.unsw.adfa.edu.au/unsw-canberra-cyber/cybersecurity/ADFA-NB15-Datasets/

[34] I. Sharafaldin, A. Gharib, A.H. Lashkari and A.A. Ghorbani, Towards a Reliable Intrusion Detection Benchmark Dataset, Journal of Software Networking, 2017, 177–200. doi: 10.13052/jsn2445-9739.2017.009

[35] [OTX] Open Threat Exchange database, https://otx.alienvault.com/

[36] Suricata website, https://suricata-ids.org/

[37] Emerging Threats rule server, https://rules.emergingthreats.net/

[38] R. Light (2017). Mosquito: server and client implementation of the MQTT protocol. Journal of Open Source Software, 2(13), 265.

[39] I. Vaccari, M. Aiello, and E. Cambiaso (2020). SlowITe, a novel denial of service attack affecting MQTT. 2nd Workshop on Attackers and Cyber-Crime Operations.

Part II

Securing Critical Infrastructures of the Health Sector

DOI: 10.1561/9781680836875.ch8

Chapter 8

Security Challenges for the Critical Infrastructures of the Healthcare Sector

By Eva Maia, Isabel Praça, Vasiliki Mantzana, Ilias Gkotsis, Paolo Petrucci, Elisabetta Biasin, Erik Kamenjasevic and Nadira Lammari

Healthcare organizations are an easy target for cybercrime due to their critical and vulnerable infrastructure. Increasing digitalization has led to the emergence of several security challenges. It is crucial to identify these critical challenges, not only from a technical point of view but also from a legal and management perspective. Recognition of the threats that may arise is also important to be able to fight cybercrime. Not just physical and/or cyber threats are relevant but also the combination of both. It is important to understand how they can impact and destabilize health services, and how they are being used by attackers to achieve their aims. This chapter provides a brief introduction to the critical challenges in the healthcare sector and a list of recent security incidents. Five main groups of threats and a critical assets categorization are also presented. Finally, the EBIOS methodology is introduced and used to describe two relevant cyber-physical scenarios of threat.

8.1 Introduction

Over the last decade, cybercrime has been the greatest threat to every sector in the world. Due to its critical and vulnerable infrastructure, the health sector is an easy target for hackers. Moreover, healthcare organizations are highly trusted entities that hold valuable and personal information, meaning that exploiting its vulnerabilities brings huge potential financial and political gain.

Several security challenges emerge from the needs of the healthcare sector. It is important to ensure the security of data without impacting the availability of the healthcare services, as they are crucial to human life. The increasing interconnection of physical and cyber assets of the hospital brings new threats that should be considered to ensure patient safety. Also, legal requirements like GDPR in Europe need to be taken into account to ensure patient data protection and compliance with the regulations.

Being aware of security incidents that have occurred is very important for understanding the risks that healthcare facilities can face. It is also important to know which critical assets are present and their impact on the availability of systems. Only then is it possible to identify and design scenarios that can help recognize threats that a security solution for a healthcare facility must cover. These scenarios should exploit combined physical and cyber threats in the context of cascading attacks, since they are the most complex and interesting threats to cope with.

8.2 Challenges in Healthcare Sector

Nowadays, healthcare structures are equipped with common perimeter precautions and active and predictive cybersecurity solutions. With these cybersecurity systems, the possibility of successfully carrying out an attack on critical assets (for example, on the main IT systems, HIS hospital information system, PACS picture archiving and communication system, LIS laboratory information systems, and other vertical software like for ER-ED) remains very low.

There is no other possibility of breaking the perimeter defenses, even using a physical attack, since it is necessary to connect directly to servers and networks sections that are not accessible from the outside. In this case, access with violence, theft, or other fraudulent access should only be seen as complementary action of a cyberattack.

It is known that hospital or healthcare structures do not work properly without IT systems, in particular the PACS and the LIS, without which it is very difficult to work with radiological images and laboratory tests, making diagnosis and therefore treatment of patients difficult, or extremely slow. This could be considered an attack damage "multiplier effect."

Therefore, it is essential to face the risk of attacks on hospital IT systems in order to decrease the functioning capacity of hospitals and to absorb patients in the emergency department, in the event of a terrorist attack and consequent maxi-inflow of wounded people. For that, threats can no longer be analyzed solely as physical or cyber. It is critical to develop an integrated approach in order to fight against such combination of threats.

8.2.1 The Protection of Critical Assets—the Point of View of Healthcare Structures Management

The management of Healthcare structures are used to facing complex challenges, such as the typical complexities of the healthcare sector, and a number of internal and external emergencies that may occur and have actually occurred; but the challenge of cybersecurity and physical security is something that in most European hospitals and Healthcare structures there is not yet full knowledge of and is not yet being considered; or perhaps, better expressed, that we are only now beginning to consider as an emerging problem, but are still lacking widespread and shared solutions.

Certainly, the IT sectors of Healthcare structures have, in recent years, had to face a number of malware attack campaigns (the most famous being Wannacry, NotPetya and CryptoLocker) that are not specifically directed to a particular type of structure.

For some hospitals, the damage was greater than expected (for countless causes, such as the diffusion of computer clients of different management and origin), but this had the benefit of putting the structures and management on alert, and considering the problem as a possible threat, like any other.

Only in recent years (mainly in USA and Asia) have attacks specifically targeting healthcare facilities been reported (like orangeworm, kwampirs, medjack). This confirmed a certainty: *Hospitals* are no stranger to malware and ransomware cyber*attacks.*

In some cases, vulnerabilities of medical device systems have been exploited; medical devices, something that was not considered a possibility, likely for cultural reasons, coupled with the fact that the medical device suppliers themselves did not consider an attack possible and were not prepared to deal with the possibility. Indeed, it is necessary to consider the particular market of medical devices:

- Productions in small series, sometimes very small series (for example, in Radiotherapy);

- Highly complex and innovative systems and therefore high costs for research and development;
- Complex sector regulations (MDD, MDR, IVD, IVR), with the need to certify every different model;
- The consequent difficulty of keeping operating systems and antivirus updated.

In recent years, there have been reported numerous local health structures affected by massive ransomware attacks, with the consequence of the total blockade of some departments, such as the emergency room and hospitalizations (!), with the sole exclusion of the "most critical" patients not diverted elsewhere (in danger of life, in other words, negotiable without the help of a computer system). A criminal attack with the explicit request for cash ransom, an operation organized on a larger scale than the typical ransomware already widespread at the level of individual personal computers, more organized as entire networks and computer servers are affected, making entire hospital systems unavailable.

Of course, we do not know the full consequences of the attacks, only what was reported to the press—in some cases, it has been reported that the very few infected computers have been reformatted, and restored, without significant loss of data; in others, the administration admitted that it preferred to pay the ransom after several days. But many operators are convinced that the cases disclosed are only a part of those actually verified and never spread for obvious reasons of bad publicity.

The latest attacks that have been reported in the news took place in October 2019 (USA and Australia). Again, with reference to the world of the United States, the analysis lead experts to believe that hackers are increasingly concentrated on the portals, patients that are increasingly popular, as they are connected with EMR/EHR (Electronic Medical/Health Record). At the same time in Europe, in recent years (at least after the serious attacks on crowded and critical structures like railways, undergrounds, airports), all critical structures are expected to be prepared to be hit by attacks. Until now, attacks using explosives on hospital are documented only in East and Middle East (Egypt, Afghanistan, Pakistan).

In short, there is a need for European structures to be prepared for the worst, in order to deal systematically with these threats, simply because of the obvious consideration that these threats will, sooner rather than later, hit the old continent as well. Without forgetting the considerable latency due to finding suitable solutions and the time necessary to spread them in the structures.

- To understand the reference context (and consequent difficulties and facilities), it is important to consider the particular situation of typical European healthcare structure: Entrances and access control—unlike public offices or other public buildings, no hospital or healthcare structure has the possibility to restrict access to one or two "single point of entry," nor is the commissioning of check points at a visitor control desk, with the control and filing of identity documents, possible.

The reasons can be many and among these certainly is the fact of not having so far hypothesized the need to protect these buildings from specific attacks (a cultural aspect that is certainly erroneous). We cannot ignore also:

  - The considerable inflows—many thousands of people for some European hospital districts;
  - The dimensions (hectares) of hospital enclosures, within the urban fabric context, sometimes in historical contexts and historical buildings;
  - The simultaneous presence of different organizations—such as universities, with consequent additional inflows of students and various attending people.

- Critical assets—hospital structures are characterized by the presence of a very large number of critical assets, probably of a small size when compared to industrial plants, but with the contemporaneity of a huge number of different types of implants and different specific safety systems (idem—when compared to industrial plants). For example: cryogenic systems, RX systems, handling radioactive isotopes, big magnetic field systems, gas tanks, hyperbaric systems and so on, And with the greatest difficulties deriving from the co-presence of large number of people: patients, visitors, students (the largest being some 10–20 thousand people);
- Separate management of IT assets (IT department) and medical devices assets (clinical engineering department), for cultural and historical reasons; this separation was culturally motivated in the last century for the absence of networked medical devices systems, at least those few who were computerized. Nowadays, the opposite happens: very few medical devices are not computerized and not connected to the IT network. Without denying the specific skills of the two staff (IT and CE), there is a strong need for coordination in management for cybersecurity aspects;
- Emergency Plans—all the hospital structures have a well-established habit of confidence and have long established various emergency plans, maxi-influx of patients, evacuation of patients, etc; the staff is therefore trained for even disastrous events and can therefore also face the consequences caused by attacks;
- Provision of video surveillance systems—due to the difficulties of inserting access controls, many hospital structures are equipped with several video surveillance systems, videocameras and a videoserver, mainly for crime prevention purposes (only with video-recording).

8.2.2 The Protection of Critical Infrastructures for the Healthcare Sector in Europe: Legal Challenges

8.2.2.1 The EU legal framework for the protection of critical infrastructures

The legal framework concerning security and the protection of critical infrastructures in Europe is characterized by its complexity. This is due to the presence of heterogeneous laws differing in scope (applicable at national or EU level) and matter (ranging from civil protection law, security laws, privacy laws, etc.) applicable to the subject matter. Moreover, the protection of critical infrastructures encompasses two parallel aspects, the physical and the cyber. Each aspect corresponds to what is commonly referred to as "Critical Infrastructure Protection" (CIP) and "Critical Information Infrastructure Protection" (CIIP) (1). This parallel is also evident in the EU legislation concerned with the topic[1] [1]. CIP and CIIP are regulated by respective directives (legislative acts setting out only a goal that all MS must achieve via national laws). The most important piece of legislation concerning CIP is the ECI Directive [2] dealing with the 'European Critical Infrastructures' (ECI). The most relevant legislation dealing with CIIP is the NIS Directive [3] the aim of which is to set up measures for a high common level of security of network and information systems across the Union.[2]

8.2.2.2 Challenges originating from the EU legal framework

With regard to the CIP, the status of protection of national healthcare critical infrastructure results to be "disparate" [4] among the MS, which is due to the regulation of security by national laws. As a consequence, some MS (e.g., the Netherlands [5]) do not explicitly mention "healthcare" as a sector worthy of protection under national CIP legislations.[3]

[1] The outline of the legislative developments of CIP and CIIP legislation in Europe falls outside the purposes of the present article. For an overview of the main pieces of legislation and policy-making instruments in the EU, see A. Kasper, A. Antonov, "Towards Conceptualizing EU Cybersecurity Law" (2018).

[2] The ECI Directive has been approved in 2008 and, although devoted to CIP, it applies only to CI that fall under the definition of 'European Critical Infrastructures. Member States' national Critical Infrastructures fall outside the scope of the ECI Directive. The ECI Directive remains, however, a key reference within the EU CIP framework as it provides meaningful legal definitions on CI (such as, the definition of Critical Infrastructure, under art. 2). Furthermore, the ECI Directive does not consider the healthcare sector as worthy of being protected, while, the NIS Directive considers the healthcare sector as falling within the scope of the legislation.

[3] France is an example of a Member State that has included the healthcare within CIP legislation. The French Defence Code ("Code de la Défense") considers critical infrastructures as the ones that are vital for the maintenance of the social and economic progress. It considers 12 sectors for critical infrastructures and includes healthcare.

With regard to the CIIP, the NIS Directive represents an important step towards reaching a common level of cyber resilience across the EU as it has set, among others, security and notification requirements for operators of essential services (OES) (i.e., "healthcare providers" for the healthcare sector) [6]. Nonetheless, many challenges—concerning the implementation and the interpretation of the law by the EU MS—await to be addressed. For example, many MS have not respected the deadline (9 May 2018) for the adoption of national laws implementing the NIS Directive [7]. This has implied uncertainty for many stakeholders willing to put in place the necessary measures foreseen by the EU law and national law. Furthermore, in order to identify the OES (such as hospitals, clinics, etc.) [8]. MS have adopted methodologies that have proven to be heterogeneous. [9] For instance, some MS have identified a very high number of OES (for instance, Finland) [9], whereas others have identified less.[4] Such difference in numbers may have a negative impact on the coherent application of the NIS Directive within the Union, with possible consequences for the whole internal market and the effective handling of cyber-dependencies [9]. Moreover, the Directive states that OES have to notify incidents "having a significant impact to the continuity of the essential services they provide" [10]. Since the purpose of the Directive is to provide a level of minimum harmonization [11], the body of the text does not specify what "significant impact" means—leaving MS to provide their own definition. This may consequently lead to fragmentation among operators across Europe who will have to follow their respective national approaches with regard to incident notification.[5] Similarly, the Directive does not granularly define the security measures that OES must adopt "to prevent and minimize the impact of incidents affecting the security of the network and information

[4] To give an example, according to the data provided by the EC Report [8], Finland has identified 10.897 OES for all NISD sectors—due to the high number of OES identified for the healthcare sectors (see [9], p. 27, footnote 8). This number appears to be very high, considering that the sum of all OES identified by all the other MS for all the NISD sectors is 4.925. To give a comparative example with another MS, Italy has identified 553 OES for all NISD sectors [9]. Furthermore, according to the preliminary documentation available, Italy has identified 326 OES for the healthcare sector—see Presidenza del Consiglio dei Ministri. Intesa ai sensi dell'articolo 4 del decreto legislative 18 maggio 2018, n. 65, recante attuazione della direttiva (UE) 2016/1148 del Parlamento Europeo e del Consiglio, 6 luglio 2016: misure per un livello comune elevato di sicurezza delle reti e dei sistemi informativi dell'Unione, tra il Governo e le Regioni e Province autonome di Trento e Bolzano, sullo schema di decreto del Ministero della salute (version of 7 November 2018, available at: www.statoregioni.it).

[5] While this problem remains, for the sake of completeness it is also true that the European Commission is putting in place also coordinative efforts to tackle this kind of issues. As an example, see the European Commission guidelines on Incident Reporting, which have been drafted within the framework of a Cooperation Group composed by Member States' experts. European Commission, Reference document on Incident Notification for Operators of Essential Services. Circumstances of notification, CG Publication (February 2018). To be noted that the document is not binding.

systems" [12]. This may bring further fragmentation among healthcare operators in Europe.

Although the challenges mentioned above might appear copious, it should also be stressed that during recent years, the EU has put in place several legislative measures to increase the level of CIP in Europe [13]. While there is still enough room for improvement, legislative instruments such as the ECI Directive and the NIS Directive have served as a catalyst in many Member States to pave the way for real change in the institutional and regulatory landscape of critical infrastructures. Further non-binding guidance at an EU level and the already established coordinative mechanisms between Member States (most importantly the recently established NIS Cooperation Group [14]) could be beneficial to achieve a higher degree of coherence for CIP, and especially CIIP, in Europe.

8.3 Recent Security Incidents

According to the World Health Organization (WHO) definition "Hospitals complement and amplify the effectiveness of many parts of the health system, providing continuous availability of services for acute and complex conditions" [15]. They are an essential element to health systems as they support care coordination and integration and play a key role in supporting other healthcare providers, such as primary health care, community outreach and home-based services. For these reasons, cyber and physical attacks against hospitals, patients, healthcare workers, and facilities have been on the rise worldwide [16].

More specifically, in terms of cyberattacks, it has been reported that 81% of 223 healthcare organizations surveyed and >110 million patients in the US had their data compromised in 2015, with only 50% of the providers thinking that they could protect themselves from cyberattacks [17]. In addition, between 2009 and 2018, there have been 2.546 healthcare data breaches that resulted in theft/exposure of 189,945,874 records [18]. In the healthcare sector, hacking and malware (including ransomware) are the leading attack type of health data breaches [19]. These data breaches result in large financial losses, but also in loss of reputation and reduced patient safety.

Several cyberattacks in the healthcare sector have been reported and some examples of such incidents are presented below:

- 2017 WannaCry attack infected more than 300,000 computers across the world demanding that users pay bitcoin ransoms. The WannaCry cyberattack targeted the UK's National Health Service (NHS). By exploiting a Windows vulnerability, the hackers managed to infect at least 16 health centers and

200,000 computers, which led to the cancellation of nearly 20,000 appointments and paralyzed more than 1,200 pieces of diagnostic equipment [20]. Moreover, according to US media, the Presbyterian Medical Centre shut down for 10 days until it paid a $17,000 ransom [21].

- Medical Device Hijack (Medjack) is another known attack that injects malware into unprotected medical devices to move laterally across the hospital network [22]. Between the first detection of Medjack in 2015 and now, there have been many variations of the attack with several hospitals' medical devices, including X-ray equipment, Picture Archive and Communications Systems (PACS), and Blood Gas Analyzers (BGA), etc., having been attacked. The attacker establishes a backdoor within the medical device, and almost any form of manipulation of the unencrypted data stored and flowing through the device is possible.
- It was reported in the press that in January 2019 hackers performed a ransom attack in a heart specialist clinic in Melbourne, where the hackers hit patient files [23]. As a result, staff was unable to access some patient files for more than three weeks. The Clinic could have mitigated the impact if data was properly and fully backed and if they were investing consistently in IT security.
- A billing company based in the USA, which operates the online payment system used by a network of 44 hospitals in the USA, discovered that some of its databases that contained 2,652,537 patients' records had been compromised in 2018. Upon discovery of the breach, access to data was terminated and forensic specialists were hired to review the incident, secure affected databases, and improve security controls (HIPAA, 2018).
- In 2019, it was revealed that a billing services vendor American Medical Collection Agency was hacked for eight months between August 1, 2018, and March 30, 2019. Since the breach was revealed, at least six covered entities have come forward to report their patient data was compromised by the hack. So far, up to 25 million patients from were affected [24].
- 128,400 records were affected by a sophisticated phishing incident that happened at New York oncology and hematology clinic. More specifically, fourteen employee email accounts clicked on phishing emails, which exposed health information in the email accounts. The clinic hired forensic specialists to assess the breach and types of data affected. Moreover, improvements to data security following the incident included active monitoring of affected systems, regular password resets, additional employee training, and new email protocols [19].

On a similar line, in the USA, the U.S. Department of Health and Human Services has developed a breach portal, the aim of which is to gather information on healthcare sector physical and cyber breaches. According to this portal, in 2019, 407 entities have been attacked and 40.267.487,00 individuals have been affected. In addition, there have been identified four main types of breaches hacking/IT incident (61%), improper disposal (1%), loss (3%), theft (8%), and unauthorized access/disclosure (28%), with the hacking/IT incidents affecting a total of 35.381.048,00 individuals and the unauthorized access/disclosure affecting 4.551.487,00 individuals [25].

In addition, physical attacks deprive people of urgently needed care, endanger healthcare providers, and undermine health systems. The WHO created the Attacks on Health Care initiative to systematically collect evidence on attacks on healthcare, to advocate for the end of such attacks, and to promote best practices for safeguarding healthcare from attacks [26]. The initiative is global, but its main geographic focus is at the country level. According to this initiative, in 2018 the healthcare sector (19 countries) was attacked 388 times and this caused 322 deaths and 425 injuries. The attacks were mainly bombings (51%), shootings (14%), threats of violence (9%), etc. Several physical attacks, such as violence against physicians (including hostage taking), fires, shootings, bombings against infrastructures, have been reported all around the world and some examples of such incidents are described below:

- While a nurse was examining a female patient, the accompanying Roma (gypsies) group attacked her and injured her face. The incident happened at the Salamina Island Health Center, Greece (POEDIN, 2018). Similar incidents have been reported to other countries, such as Cyprus [27], Louisiana [28], Kolkata [29], Australia [30], etc.
- A UK A&E registrar was held hostage when she had gone to check on a young patient, who was having a mental health episode after taking drugs. Unfortunately, the patient had managed to hide a pair of scissors, which she pulled out before backing the doctor into a corner. The police were eventually called and restrained the patient [31].
- A woman opened fire at a flat opposite a Catholic Hospital and then inside the hospital in the south-western town of Lorrach in Baden-Wuerttemberg, Germany, killing at least three, including one child, and wounding several patients before police shot her dead [32].
- A gunman killed six patients in a hospital waiting room in the Czech city of Ostrava and drove off. Police launched two helicopters to search for him, once they had obtained pictures of the suspect from security cameras. When one of the helicopters was flying over the car, the man shot himself in the head and later died of his injuries [33]. It has been reported that shooting rates in hospitals, increased from 9 per year from 2000 to 2005 to 17 per year

from 2006 to 2011, according to a study published in 2012 in the Annals of Emergency Medicine [34].

Physical or cyber incidents like the above could affect the healthcare services provision and could cause overwhelming pressure, such as loss of infrastructure or a massive patient surge. Hospitals not only provide care services but they are also the last resort for disaster victims seeking care and represent an icon of social security, connectivity, and community trust [35]. Thus, in this context, it is fundamental for a hospital to remain resilient, maintain the level of provided care, and be able to scale up its service delivery in any given emergency situation.

8.4 Threat and Risk Analysis

Threats are actions that can negatively impact valuable resources of an organization. Typically, threats exploit vulnerabilities of the system, i.e., take advantage of some weaknesses in the system to trigger an undesired outcome, as damage or loss of an asset.

To guarantee the safety of the systems, it is very important to determine the possible root causes of threats. According to ENISA [36], we can identify five main groups of threats faced by healthcare organizations:

- **Malicious actions** that are deliberate acts performed by an internal or external person or organization to destroy or steal data or sabotage the system. Malware (e.g., virus, ransomware), hijacking, social engineering, medical device tampering, device and data theft are examples of malicious actions;
- **Human errors** that are related with misconfiguration or improper use of devices and information systems, and incorrect execution of processes;
- **System failures;**
- **Supply chain failures** that are responsibility of third-party suppliers, for example, power suppliers, medical device manufacturers, etc.;
- **Natural phenomena.**

The person or entity who is responsible for conducting these threats (threat actor) can also be classified according to its role:

- **Insider threat actor:** this category is composed of the hospital staff (physicians, nurses, administrative staff, etc.);
- **Malicious patients and guests;**
- **Remote attackers:** actors who are not physically in the hospital;
- **Other causes:** such as environmental or accidental equipment failure.

Table 8.1. Asset categories.

Category	Example
Specialist personnel	Employees, Persons with special functions, etc.
Buildings and Facilities	Main and ancillary buildings, Technical buildings, Power and climate regulation systems, temperature sensors, medical gas supply, room operation, automated door lock system, etc.
Identification Systems	Tags, bracelets, badges, biometric scanners, CCTV (video surveillance), RFID services, etc.
Networked Medical Devices	Mobile devices (e.g., glucose measuring devices), wearable external devices (e.g., portable insulin pumps), implantable devices (e.g., cardiac pacemakers), stationary devices [e.g., computed tomography (CT) scanners], support devices (e.g., assistive robots), etc.
Networking Equipment	Transmission media, network interface cards, network devices (e.g., hubs, switches, routers, etc.), telephone system, etc.
Interconnected Clinical Information Systems	Hospital information system (HIS), Laboratory information system (LIS), Pharmacy information system (PIS), Picture archiving and communication system (PACS), blood bank system, etc.
Mobile Client Devices	Mobile clients (e.g., laptops, tablets, smartphones), mobile applications for smartphones and tablets, alarm, and emergency communication applications for mobile devices, etc.
Remote Care System Assets	Medical equipment for tele-monitoring and tele-diagnosis, medical equipment for distribution of drugs and telehealth equipment (cameras, sensors, telehealth computer system for patients to register their physiological measurements themselves, etc.)
Data and records	Clinical and administrative patient data, financial, organizational and other hospital data, staff data, vendor details, tracking logs, etc.
Operating resources	Medicinal products, medical consumables, Laundry supply, Sterile supply, Food supply, etc.

Organizations have a wide range of entities, the assets, which are essential for their operation. Thus, it is crucial to identify the critical assets in the hospital to ensure the patients' safety. Table 8.1 presents the list of critical asset categories.

A cyber-physical attack scenario is a combination of threats, vulnerabilities, and assets. In the next section, we will describe nine different relevant attack scenarios against critical health infrastructures.

8.5 Scenarios of Threat

The definition of the cyber-physical scenarios of threat, to be clearer, should follow a methodology. Several methodologies for risk assessment exist, such as ISO 31000:2018, IEC 31010: 2019, ISO27005, etc. EBIOS methodology is compliant with the standards, as ISO3100, ISO/IEC 27001, ISO/IEC 15408, etc., and it is commonly used to describe the scenarios of threat. EBIOS [37] is a French acronym meaning Expression of Needs and Identification of Security Objectives (Expression des Besoins et Identification des Objectifs de Sécurité) and was developed by the French Central Information Systems Security Division. EBIOS is used to assess and treat risks related to information systems security (ISS). It can also be used to communicate this information within the organization and to partners and therefore assists in the ISS risk management process since it is compliant with major IT security standards. EBIOS can be employed in different fields (using the appropriate techniques and knowledge bases) [38], even if it was initially designed for information security. To apply EBIOS in a specific field, it is generally sufficient to adapt the terminology and exploit the techniques and the knowledge bases specific to that field concerned if the knowledge does not seem to be applicable or understood (primary assets, considered criteria, potential impacts, etc.).

EBIOS uses a progressive risk management approach (see Figure 8.1): it starts in the major missions of the object under study (highest level) and goes to the business functions and techniques (lowest level), studying possible risk scenarios [39]. It aims to obtain a synergy between compliance and scenarios, positioning these two complementary concepts in the best way, i.e., where they bring the highest value. The compliance approach is used to determine the security base of the scenarios,

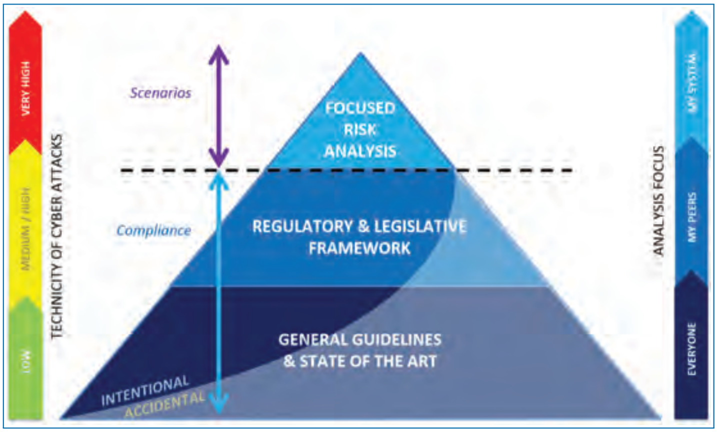


Figure 8.1. Digital risk management pyramid.

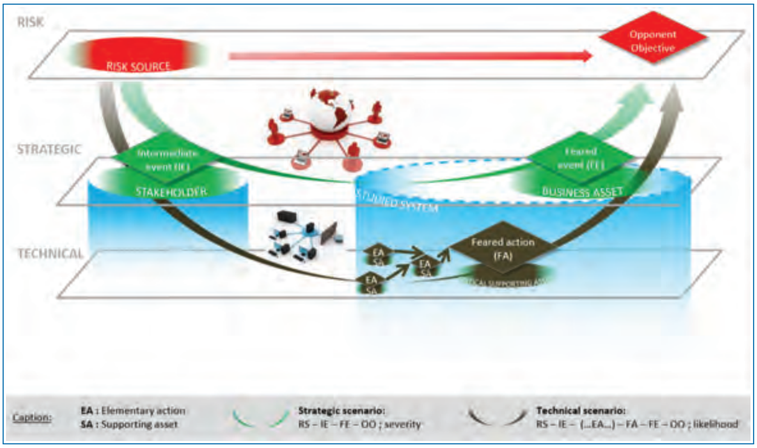


Figure 8.2. Cyber risk scenario.

particularly to develop targeted or sophisticated scenarios. This assumes that accidental and environmental risks are treated a priori by the compliance approach. Thus, scenario risk assessment focuses on intentional threats.

The EBIOS method consists of five iterative workshops (Figure 8.2):

- **Scope and Security Baseline:** the first workshop aims to identify the scope of the study, the workshop participants and the time frame. During this workshop, essential and support assets and business values should be listed. Threat events and their impact should be identified at this stage. The security baseline should also be defined. This first workshop follows the compliance approach: it corresponds to the first two stages of the digital risk management pyramid.
- **Risk Sources:** in the second workshop, the risk sources and their high-level objectives should be identified and characterized.
- **Strategic Scenarios:** in this workshop, it is possible to have a clear vision of the ecosystem, which allows to build high-level scenarios of threat. They represent the paths of attack that a risk source can take to achieve its objective. These scenarios are conceived taking into account the ecosystem and the business values of the object, and they are evaluated in terms of severity. At the end of the workshop, it is already possible to define security measures on the ecosystem.
- **Technical Scenarios:** the purpose of this workshop is to build scenarios containing the technical procedures that can be used by the risk sources to carry out the strategic scenarios. This workshop adopts a similar approach of the previous one but focuses on the critical assets. Then, the likelihood of the technical scenarios should be evaluated.

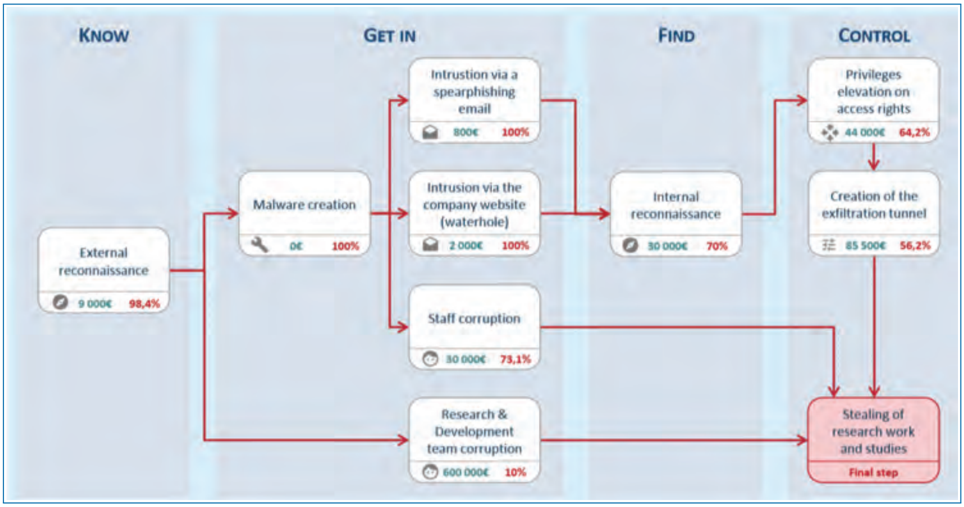

Know
Get in
Find
Control
Intrustion via a spearphishing email
800€ 100%
Privileges elevation on access rights
44 000€ 64,2%
Malware creation
0€ 100%
Intrusion via the company website (waterhole)
2 000€ 100%
Internal reconnaissance
30 000€ 70%
Creation of the exfiltration tunnel
85 500€ 56,2%
External reconnaissance
9 000€ 98,4%
Staff corruption
30 000€ 73,1%
Research & Development team corruption
600 000€ 10%
Stealing of research work and studies
Final step

Figure 8.3. Example of a technical scenario description using attack graphs.

- **Risk Treatment:** in the last workshop, all the risks studied in the previous workshops are considered to define the risk treatment strategy. Then, a set of safety measures are defined and included in a continuous improvement plan. In this workshop the residual risks are also summarized and the risk monitoring framework is defined.

Therefore, we can summarize the construction of a cyber risk scenario as described in Figure 8.3.

A technical scenario can be represented in the form of an attack graph to visualize the operational modes planned by the attacker to achieve its objective. An example of an operational scenario is given in Figure 8.3.

The proposed model consists of 4 phases:

- **KNOW:** set of targeting, reconnaissance, and external discovery activities conducted by the attacker to prepare his attack and to increase his chances of success (ecosystem mapping, information on key people and systems, search and evaluation of vulnerabilities, etc.). Such information shall be collected according to the determination and resources of the attacker: intelligence, economic intelligence, exploitation of socio-professional networks, direct approaches, specialized meetings for information inaccessible in open source, etc.
- **GET IN:** all activities carried out by the attacker to digitally or physically introduce either directly and frontally into the target information system or in its ecosystem for a rebound attack. The intrusion is usually carried out

through "border" goods that serve as the entry points due to their exposure, for example, user post connected to the Internet, maintenance tablet of a provider, TV-maintained printer, etc.

- **FIND:** internal recognition of networks and systems, localization, elevation, and persistence, which allows the attacker to locate the desired data and material. During this phase, the attacker usually seeks to remain discreet and erase his traces.
- **CONTROL:** all the data and media activities found in the previous stage. For example, in the case of sabotage, this phase includes the activation of the active load, for example, ransom; in the case of an espionage operation aimed at ex-filtering emails, it may be necessary to establish and maintain discrete capacity for data collection and exfiltration.

After the definition of the technical scenario, it is important to evaluate its overall likelihood, which reflects its probability of success or feasibility. To begin, the elemental likelihood of each action in the scenario should be assessed. This can be estimated by the judgment of an expert or using metrics. Then, the overall likelihood of the scenario is evaluated from the elementary likelihoods.

Three different approaches can be considered to rate the likelihood of the operational scenarios:

- **Express method:** direct quotation of the likelihood of the scenario;
- **Standard method:** rating of the "probability of success" of each elemental action of the scenario, from the point of view of the attacker;
- **Advanced method:** in addition to the "probability of success," rating of the "technical difficulty" of each elementary action of the scenario, from the point of view of the attacker.

We will consider the standard method. The following scale will be used to determine the probability of success of each elementary actions (Pr(EA)) [40]:

- **4 – Almost certain:** Probability of near-certainty >90%;
- **3 – Very High:** Very high probability of success >60%;
- **2 – Significant:** Probability of significant success >20%;
- **1 – Low:** Success probability low <20%;
- **0 – Very Low:** Success probability very low <3%.

The overall likelihood score of the scenario can be evaluated using the following rule:

$$Index\_\mathrm{Pr}(EA_n) = Min\{Pr(EA_n),\ Max(Index\_Pr(EA_{n-1}))\}$$

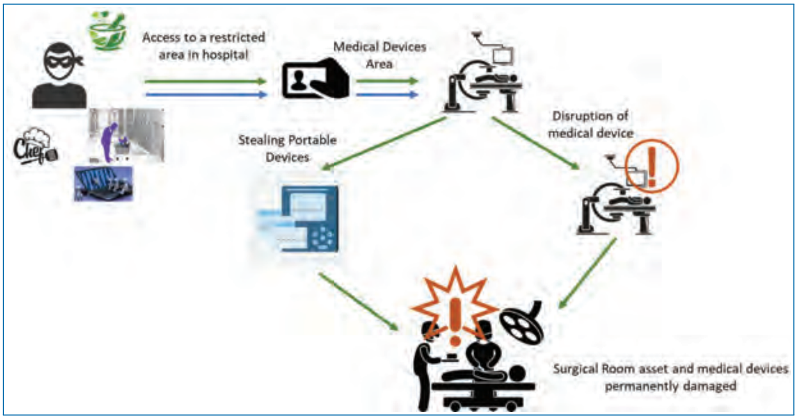

Access to a restricted area in hospital
Medical Devices Area
Stealing Portable Devices
Disruption of medical device
Surgical Room asset and medical devices permanently damaged

Figure 8.4. Medical devices attack: sketch.

The idea is to evaluate step by step an intermediate cumulative probability index from the elementary action “EA_n” of a node n and the cumulative indices of the previous node $n-1$. The overall probability of success index (final step) is obtained by taking the highest intermediate cumulative probability index among the procedures that lead to the final step. It corresponds to the mode(s) operating(s) whose chance of success seems the highest.

8.5.1 Scenario Example 1: Cyber-physical Attack to on Medical Devices

Medical devices are an important asset in healthcare infrastructure. They improve the quality of life of the patients, but they are also a source of threat due to the increasing connectivity to all parts of the hospital. Several attacks on medical devices have been reported during the last years. For example, in 2018, security researchers demonstrated that they have founded security weaknesses in Medtronic pacemakers that leaves the life-saving device vulnerable to hackers and puts patients at risk.[6] Figure 8.4 shows an example of medical devices attack.

An attacker, in order to influence treatment outcome or for financial gain, can obtain physical or remote access to medical device and use reverse engineering to identify a vulnerability and exploits it. Then, the attacker can take advantage of the exploit of the medical device to alter its software and/or cause a disruption in health systems, which can potentially harm patients and/or staff (see Figure 8.5).

Thus, in this case, after the vulnerability scanning, the medical device system is changed or a denial of service is launched to interrupt the health system (Figure 8.6).

[6] https://www.cnbc.com/2018/08/17/security-researchers-say-they-can-hack-medtronic-pacemakers.html

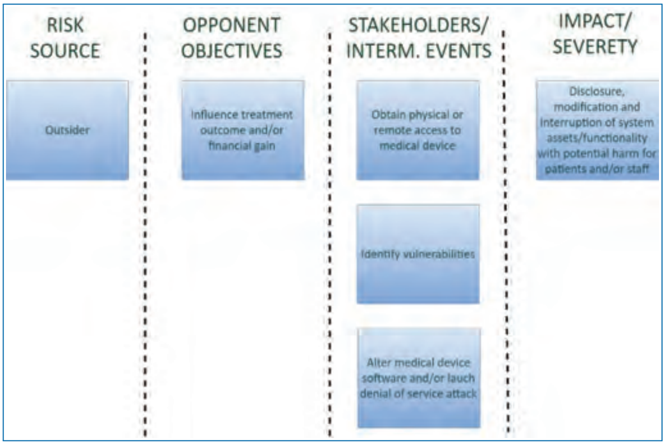

RISK SOURCE
OPPONENT OBJECTIVES
STAKEHOLDERS/ INTERM. EVENTS
IMPACT/ SEVERETY
Outsider
Influence treatment outcome and/or financial gain
Obtain physical or remote access to medical device
Identify vulnerabilities
Alter medical device software and/or lauch denial of service attack
Disclosure, modification and interruption of system assets/functionality with potential harm for patients and/or staff

Figure 8.5. Medical devices attack: strategic scenario.

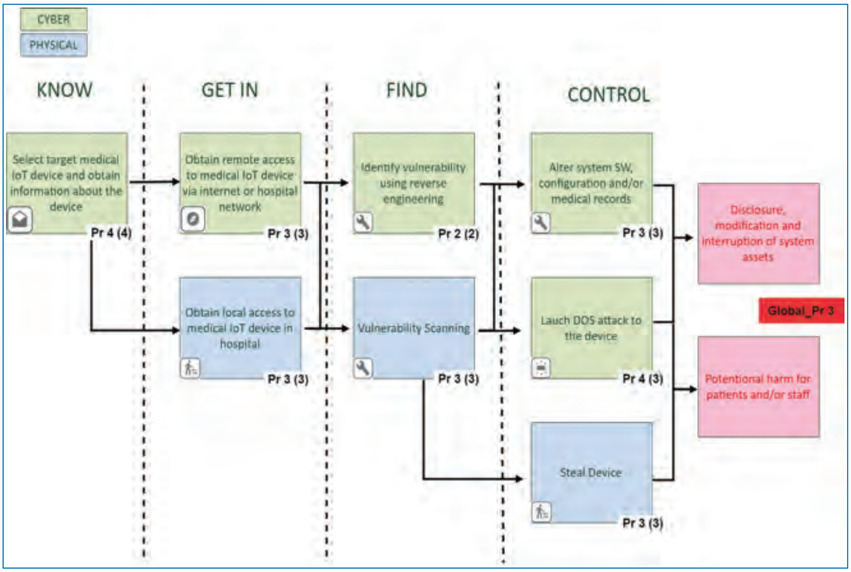

CYBER
PHYSICAL
KNOW
GET IN
FIND
CONTROL
Select target medical IoT device and obtain information about the device
Pr 4 (4)
Obtain remote access to medical IoT device via internet or hospital network
Pr 3 (3)
Obtain local access to medical IoT device in hospital
Pr 3 (3)
Identify vulnerability using reverse engineering
Pr 2 (2)
Vulnerability Scanning
Pr 3 (3)
Alter system SW, configuration and/or medical records
Pr 3 (3)
Lauch DOS attack to the device
Pr 4 (3)
Steal Device
Pr 3 (3)
Disclosure, modification and interruption of system assets
Global_Pr 3
Potentional harm for patients and/or staff

Figure 8.6. Medical devices attack: technical scenario.

The attacker can also steal the device. This will cause a disclosure, modification, and/or disruption of the medical device which will impact patient and/or staff safety. This attack has a very high probability of success (Pr 3).

Several assets are compromised in this type of scenario, for example: Buildings and Facilities, e.g., technical room; Identification Systems, e.g., badge (physical), credentials (cyber); Networked Medical Devices, e.g., Wearable Medical IoT; Networking Equipment, e.g., Router; Interconnected Clinical Information Systems, e.g., PACS; and Data and records, e.g., patient data.

Some practices can be considered to minimize the impact of this attack, which are:

- Establish and maintain communication with vendors security teams;
- Implement access controls for vendor support staff;
- Implement security operations practices for devices;
- Develop and implement security measures for a devices network.

8.5.2 Scenario Example 2: Cyber-physical Attack to Cause a Hardware Fault

As any critical infrastructure, the interruption of services in healthcare facilities has a huge impact on patients. Attackers can take advantage of this feature of hospitals for financial gain, or for attention or other motives. A sketch of an attack that can cause a hardware fault is represented in Figure 8.7.

The usual aims of this kind of attack are extortion, sabotage, or even intimidation. Therefore, the attacker is only concerned with finding a way to cause system unavailability, damaging the system permanently (or not) and without worrying whether the patient will be at risk or not (Figure 8.8).

The attacker can use social engineering to obtain information about the hospital infrastructure. With this knowledge, he/she could exploit the system's vulnerability, gain administrator privileges, and cause a hardware failure (Figure 8.9). The unavailability of the healthcare system can cause death or serious injury to patients, because the assistance services cannot work properly with hardware failures.

Some of the affected assets in this scenario are: Networked Medical Devices, e.g., medical devices that communicate with central system; Networking Equipment, e.g., externally accessible server; Interconnected Clinical Information Systems,

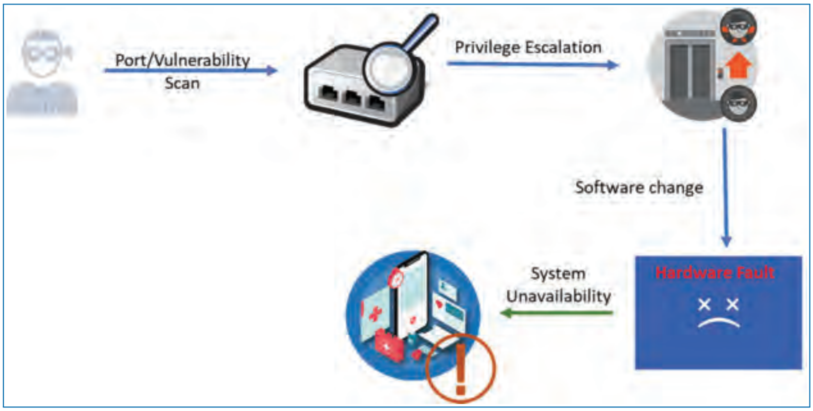


Figure 8.7. Hardware fault attack: sketch.

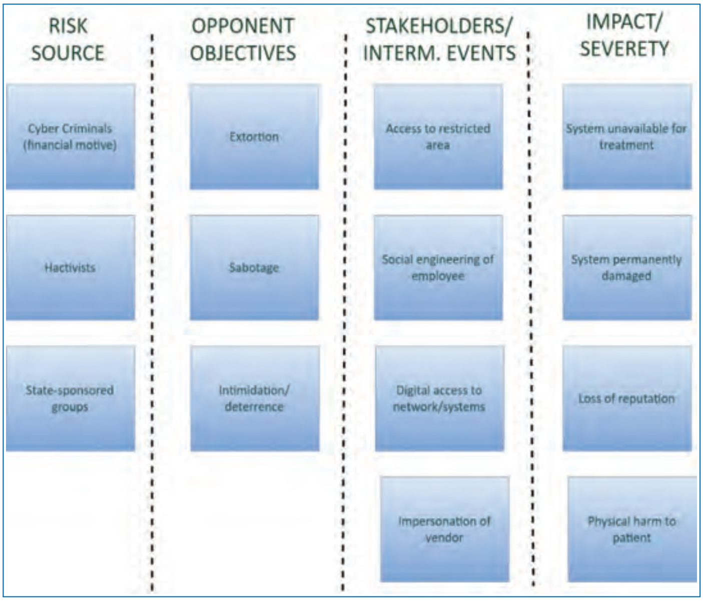

RISK SOURCE
OPPONENT OBJECTIVES
STAKEHOLDERS/ INTERM. EVENTS
IMPACT/ SEVERETY
Cyber Criminals (financial motive)
Extortion
Access to restricted area
System unavailable for treatment
Hactivists
Sabotage
Social engineering of employee
System permanently damaged
State-sponsored groups
Intimidation/ deterrence
Digital access to network/systems
Loss of reputation
Impersonation of vendor
Physical harm to patient

Figure 8.8. Hardware fault attack: strategic scenario.

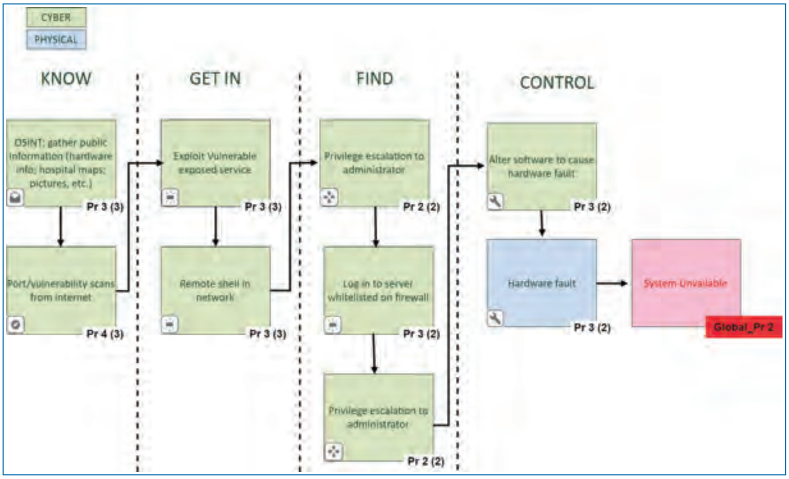

CYBER
PHYSICAL
KNOW
GET IN
FIND
CONTROL
OSINT: gather public information (hardware info; hospital maps; pictures, etc.)
Pr 3 (3)
Port/vulnerability scans from internet
Pr 4 (3)
Exploit Vulnerable exposed service
Pr 3 (3)
Remote shell in network
Pr 3 (3)
Privilege escalation to administrator
Pr 2 (2)
Log in to server whitelisted on firewall
Pr 3 (2)
Privilege escalation to administrator
Pr 2 (2)
Alter software to cause hardware fault
Pr 3 (2)
Hardware fault
Pr 3 (2)
System Unvailable
Global_Pr 2

Figure 8.9. Hardware fault attack: technical scenario.

e.g., PACS; Mobile Client Devices, e.g., mobile applications for smartphones and tablets; Remote Care System Assets, e.g., medical equipment for tele-monitoring; Data and records, e.g., health records.

It is important to note this attack could be, at least partially, mitigated if:

- An appropriate intrusion detection system had been deployed to detect early the attack;

- Exists an endpoint security system that prevents the connection of unknown devices;
- The staff had been trained to understand the threat, recognize suspicious emails, and never open email attachments from unknown senders;
- Privileged access management tools to report access to critical infrastructures had been deployed;
- Devices have been patched after the patches have been validated and distributed by medical device manufacturer;
- A restricted and rigid access controls policy for clinical and vendor support staff (including remote access and monitoring of vendor access) had been implemented.

8.6 Conclusion

Healthcare organizations are a fruitful target for crime. The increasing integration of cyber and physical systems and connected devices in its environment brings new challenges to these organizations, especially from a security perspective. To combat the threats that emerge from this healthcare technology era, hospitals need to implement cyber and physical controls, reducing the risks that can cause harm to people, property, and environment.

In this chapter, we have presented the main security challenges in the healthcare environment, not only from a structure management point of view but also from a legal perspective. A survey about the recent security incidents was performed in order to understand the type of vulnerabilities exploited by the attackers in the health sector. Inspired by this research, five main groups of threats and a critical assets categorization were defined. Finally, using EBIOS methodology, that is also briefly described, two combined cyber and physical scenarios of threat are described. All this information should clarify and alert the reader to the security issues faced by healthcare facilities in this smart hospital's era.

Acknowledgments

This work has received funding from European Union's H2020 research and innovation programme under SAFECARE Project, grant agreement no. 787002.

References

[1] Dunn, M. Understanding Critical Information Infrastructures. [book auth.] M. Dunn and V. Mauer. International CIIP Handbook vol. 2. 2006.

[2] (Directive), Council Directive 2008/114/EC of 8 December 2008 on the identification and designation of European critical infrastructures and the assessment of the need to improve their protection [2008] L 345/75 (ECI.

[3] (Directive), Directive (EU) 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 [2016] L 194/2 (NIS).

[4] THREATS. An Analysis of Critical Infrastructure Protection Measures Implemented within the European Union: Identifying which European Member States includes the Health Sector as part of Critical National Infrastructure and which facets of Health Infrastructes are. 2014.

[5] Biasin, E., *et al.* SAFECARE Deliverable 3.9 – Analysis of ethics, privacy, and confidentiality constraints. 2018.

[6] NIS Directive, Annex II.

[7] Howard, Casey. 20 EU Member States haven't implemented the NIS Directive. *itgovernance.eu.* [Online] 22 May 2018. [Cited: 3 January 2020.] https://www.itgovernance.eu/blog/en/20-eu-member-states-havent-implemented-the-nis-directive.

[8] Art 5 NISD.

[9] Commission, European. *Report from the Commission to the European Parliament and the Council Assessing the Consistency of the Approaches Taken by Member States in the Identification of Operators of Essential Services in Accordance with Article 23(1) of Directive 2016/1148/EU on.* 2019.

[10] NIS Directive, art 14(3).

[11] NIS Directive, art 3.

[12] NIS Directive, art 14(2).

[13] Kasper, A. and Antonov, A. Towards conceptualising EU cybersecurity law. 2018.

[14] NISD, art 11.

[15] World Health Organisation. [Online] 2019. [Cited: 15 09 2019.] https://www.who.int/hospitals/en/.

[16] International Committee of the Red Cross (ICRC). Health Care in Danger: Making the Case. *International Committee of the Red Cross (ICRC).* [Online] 2011. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=2ahUKEwi7i_XttvjlAhWXxcQBHbGNDpoQFjABegQIAxAC&url=https%3A%2F%2Fshop.icrc.org%2Ficrc%2Fpdf%2Fview%2Fid%2F2033&usg=AOvVaw2mlBEe7xYmGywa0Gea9edg.

[17] KPMG. Health care and cyber security: increasing threats require increased capabilities. [Online] 2015. https://assets.kpmg/content/dam/kpmg/pdf/2015/09/cyber-health-care-survey-kpmg-2015.pdf.

[18] HIPAA. HIPAA Journal. *Healthcare Data Breach Statistics.* [Online] 2018. https://www.hipaajournal.com/healthcare-data-breach-statistics/.

[19] —. HIPAA Journal. *Largest Healthcare Data Breaches of 2018.* [Online] 2018. https://www.hipaajournal.com/largest-healthcare-data-breaches-of-2018/.

[20] Perlroth, N.; Sanger, D.E. The New York Times. [Online] 2017. https://www.nytimes.com/2017/05/12/world/europe/uk-national-health-service-cyberattack.html.

[21] Stormshield. Top 5 cyberattacks against the health care industry. [Online] 2019. https://www.stormshield.com/top-5-cyberattacks-against-the-health-care-industry/.

[22] Hei X., Du X. Conclusion and Future Directions. In: Security for Wireless Implantable Medical Devices. *SpringerBriefs in Computer Science.* 2013.

[23] Martin L. The Guardian. [Online] 2019. https://www.theguardian.com/technology/2019/feb/21/hackers-scramble-patient-files-in-melbourne-heart-clinic-cyber-attack.

[24] Healthitsecurity. The 10 Biggest Healthcare Data Breaches of 2019, So Far. [Online] 2019. https://healthitsecurity.com/news/the-10-biggest-healthcare-data-breaches-of-2019-so-far.

[25] U.S. Department of Health and Human Services. Breach Portal. [Online] 2019. https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf.

[26] World Health Organisation. Health Systems. [Online] 2019. [Cited: 01 09 2019.] http://www.euro.who.int/en/health-topics/Health-systems/pages/health-systems.

[27] OFFSITE. OFFSITE. 27 χρονοςεπιτέ ϑηκε σε γιατρούς στο Νοσοκομείο Λάρνακας. [Online] 2017. https://www.offsite.com.cy/articles/eidiseis/topika/231890-27hronos-epitethike-se-giatroys-sto-nosokomeio-larnakas.

[28] Nurse. Nurse Dies After Being Attacked By Mental Health Patient – Manslaughter Charges. [Online] 2019. https://nurse.org/articles/nurse-attacked-by-patient-dies-manslaughter/.

[29] Times of India. Kolkata doctor beaten up after Garden Reach child's death. [Online] https://timesofindia.indiatimes.com/city/kolkata/doctor-beaten-up-after-garden-reach-childs-death/articleshow/69656632.cms.

[30] ZeroToleranceWorldwide. Patient charged with attempted murder after firing police officer's gun in Canberra Hospital. [Online] 2019. http://zerotoleranceworldwide.com/2018/07/19/patient-charged-with-attempted-murder-after-firing-police-officers-gun-in-canberra-hospital/.

[31] The Guardian. Violence in the NHS: staff face routine assault and intimidation. [Online] 2019. [Cited: 11 12 2019.] https://www.theguardian.com/society/2019/sep/04/violence-nhs-staff-face-routine-assault-intimidation.

[32] The Times. Woman kills 3 in hospital shooting spree. [Online] 2010. https://www.thetimes.co.uk/article/woman-kills-3-in-hospital-shooting-spree-n9q9nbhws9b.

[33] BBC. Czech shooting: Gunman kills six at hospital in Ostrava. [Online] 2019. https://www.bbc.com/news/world-europe-50725840.

[34] Kelen, Gabor D. *et al.* Hospital-Based Shootings in the United States: 2000 to 2011. 2012, vol. 60, 6, pp. 790–798.

[35] World Health Organization. Hospital Safety Index: Guide for Evaluators. [Online] 2015. [Cited: 10 09 2019.] https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=2ahUKEwi_tdq7n_kAhVE2aQKHVZfBbkQFjAAegQIAxAC&url=https%3A%2F%2Fwww.who.int%2Fhac%2Ftechguidance%2Fhospital_safety_index_evaluators.pdf&usg=AOvVaw3Jb3x3xUgBh-IK84EtnKD8.

[36] ENISA. Smart Hospitals: Security and Resilience for Smart Health Service and Infrastructures. [Online] 2016. [Cited: 23 April 2019.] https://www.enisa.europa.eu/publications/cyber-security-and-resilience-for-smart-hospitals.

[37] —. EBIOS Risk Manager: Guide Method. [Online] 2018. [Cited: 23 April 2019.] https://www.enisa.europa.eu/topics/threat-risk-management/risk-management/current-risk/risk-management-inventory/rm-ra-methods/m_ebios.html.

[38] EBIOS, Club. EBIOS: the risk management toolbox. [Online] 09 05, 2018. [Cited: January 23, 2020.] https://club-ebios.org/site/wp-content/uploads/productions/EBIOS-GenericApproach-2018-09-05-Approved.pdf.

[39] ANSSI. EBIOS risk manager. [Online] November 2019. [Cited: January 23, 2020.] https://www.ssi.gouv.fr/uploads/2019/11/anssi-guide-ebios_risk_manager-en-v1.0.pdf.

[40] ANSSI. EBIOS Risk Manager, Le Supplément. [Online] January 31, 2019. https://www.ssi.gouv.fr/uploads/2018/10/fiches-methodes-ebios_projet.pdf.

DOI: 10.1561/9781680836875.ch9

Chapter 9

Security Systems in the Healthcare Sector

By Mathias Normann, George Suciu, Vasiliki Mantzana, Ilias Gkotsis, Mari-Anais Sachian, Gabriel Petrescu, Hussain Ijaz and Barry Norton

To efficiently protect the healthcare sector is a major task. Healthcare is a highly specialized sector where the physical facilities are a mix of publicly-available, semi-private, private, and areas housing critical infrastructure. The healthcare sector spans many types of buildings, some open to the public with zones restricted only to staff and some completely restricted to the public. Some of the buildings will have patients all the time, some only during daytime, some will have important and expensive medical devices, and some will contain servers. The buildings can have physical medical records, computers used by the staff can have access to personal data, and the servers that store huge amounts of personal data, critical information, and critical software for the healthcare sector. Besides the normal software used for an office building, the healthcare sector's internal software is used to handle the medical data of patients and to run and control medical equipment and machines.

To protect all this, the healthcare sector keeps expanding the security systems within their facilities, so it is common to have several security systems to handle different security areas, as no system covers everything. This chapter will present an overview of how a video management system, an access control system, a fire detection system, SCADA, ICS, and smart building sensors, as well as a Cybersecurity protection system works to make the healthcare sector a more secure environment.

9.1 Video Management Systems

With regard to physical security, camera surveillance is the most common way of getting an overview of what is happening. In years past, when analogue cameras were the only type available, large hardware set-ups were needed to record the video feeds from the cameras and store them on physical tapes. With such set-ups, one could not combine cameras from different manufacturers, as they only worked with their own management systems. In the 1990s the Internet Protocol camera (IP) was invented. IP-cameras can be connected to the internal network using ethernet cables, instead of being directly hooked up to specific hardware, as analog cameras must. It is easy to view video feeds, given further standardization in this sector, from an IP that is connected to the internal network, and this does not require a large hardware set-up. These days the video can even be viewed using a Web browser, or some of the commonly available media players, by accessing the IP using these networking protocols, which have now spread from the Internet to almost every corporate, and indeed home, computer network. The switch to IP-enabled digital cameras came to revolutionize the world of Video Management Systems (VMSs), as it is now a lot easier and cheaper to install new cameras, and VMSs can manage video from cameras across many manufacturers.

In the healthcare sector large installations of IPs are common, given the profusion of assets to protect—both people, medical devices and data—so to sufficiently protect everything, installations may require hundreds or thousands of cameras. For a human to keep track of all these cameras manually, using a browser or video players is completely impossible, and this is where a VMS comes into the picture, to help manage all cameras, to record and store their feeds when relevant, and to help retrieve or show live the most relevant segments of video streams to the operator.

In general, a modern VMS is constructed to have a central Video Management Server with the responsibility of handling all cameras, the data they create, enabling reactions to the data and enabling the user to interact with it. The general architecture of a VMS can be viewed in Figure 9.1, and the parts of the VMS is described in detail below:

- **Storing** the configurations and information of all registered cameras in the Device Management Database;

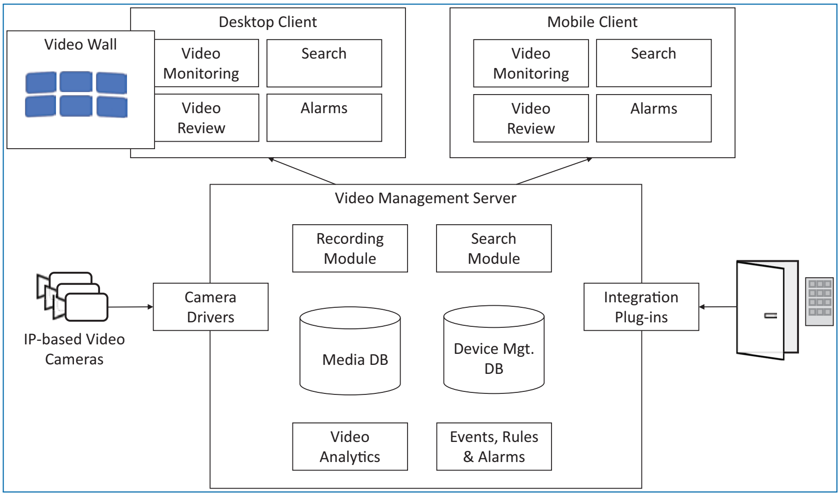


Figure 9.1. VMS architecture.

- **Recording** and **storing** the video streams from all cameras registered with the VMS into the Media Database;
- Enable **searching** for cameras and relevant video streams stored in the Media Database;
- Enable **Video Analytics** to be run on video streams, such as motion detection which can be used to only store video frames in the Media Database when there is motion in the view or to prioritize which video stream the human should watch;
- Handling of **events** happening within the system, upon which **rules** can be defined to react to conditions that can happen within the system and through which **alarms** can be raised by the system for the human operator to consider.

As mentioned above, cameras are manufactured by different companies, and even though the mission of the ONVIF standard (ONVIF, 2020) is to "provide and promote standardized interfaces for effective interoperability of IP-based physical security products," camera manufacturers are not obliged to use this standard, and even if they do, cameras can allow the configuration of features that the standard does not cover. Therefore, for a camera to work with a given VMS, in general a driver has to be developed for the camera, or for that series of cameras. Most VMSs today have drivers for at least the most common cameras in use.

In order for the operator to be able to view the live and recorded video streams, to review the alarms and to search, VMSs often provide both a desktop client and a mobile client. The desktop client can either be native to its host operating system or developed using Web interfaces to be used in the browser. Desktop clients can also enable display in a video wall, wherein the application displays across several windows, with one or more video streams in each window, where each window is mapped to an entire monitor within a bank of such.

Alongside VMSs, there are a number of other software-based systems used in physical security, such as access control of doors within the buildings covered. In recent years, there has been a demand from users to integrate these systems together with the VMS. This integration enables a better understanding of what happens at the installation site and better alarm handling.

9.2 Integrations with Video-based Security Systems

Many security systems in an installation are managed by a single central system, which is used to set up the devices and configure the system to run without further human interaction. Take, for example, an access control system: after the system has been set up to allow the right people through the right doors, the system will run without human interaction to the management system, unless something goes wrong or some access rights have to be changed. This section describes the integration of some of such building security systems and sensors that can help improve the security in the healthcare sector.

9.2.1 Access Control Systems

An access control system restricts access to areas within a building, by having one or more door controllers connected to locking mechanisms on important doors, together with card/PIN readers and request-to-exit (REX) systems, as illustrated in Figure 9.2 and described below.

Door controller The central system that receives input from the lock, reader, and REX, stores or forwards the events, as well as applying defined rules, and thereby communicating: to the lock whether it should open and, to the reader what state to display.

Lock Inside the door, and potentially the door frame, is a mechanism by which the door can be held in a locked state. One important characteristic of this mechanism is that the locking can be “fail safe”, i.e., unlocked if the power is removed, or “fail secure”, i.e., locked if the power is removed.

Reader Any system that gives permission to pass through an access control point if the correct credentials are provided, such as card readers, PIN pads, and finger print readers.

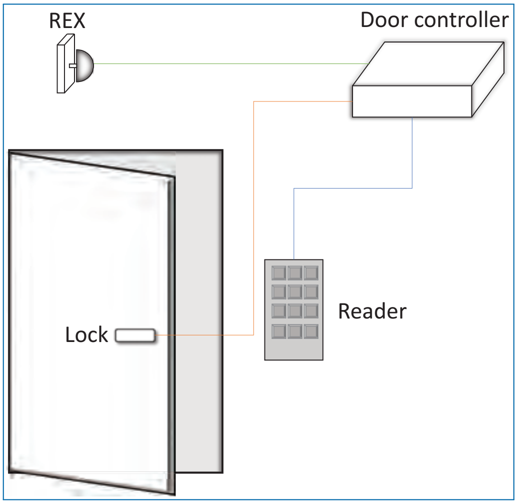


Figure 9.2. Access control system.

REX Any system that gives permission to pass through an access control point without providing credentials, such as a push button, an asymmetric door handle mechanism, or a motion sensor.

In an office building, restricting door access, and sometimes lift access, to employees and invited guests only is often a simple and easily-defined problem. This access regime can be effected by installing an access control point at every entrance of the outer perimeter of the building and only allowing the public into the reception area, until invited further in.

In the healthcare sector, however, defining those areas where the public may enter can be a much more complicated problem. There might be a lot of areas that are restricted to staff only, but that are still physically accessible to the public, as it would disrupt the hospital workflow too much to have too many access control points. To ensure adequate surveillance of such restricted areas, more intelligent solutions are required in addition to basic access control, such as integration with the video-based security systems.

When an access control point is used, if the access control system is integrated with a VMS, then the VMS can be informed of authentication and door open/-close events, in order to capture the relevant video feed covering the access control point and associate this video with the access control events produced. It is a feature of such integration that the video preserved and associated with the access control event precedes the triggering event by several seconds, requiring buffering. By knowing the video feeds covering the access control point, the VMS may further apply analytics to determine and classify malicious uses. One example is to automatically detect “tailgating,” by determining if more than one person passes through the access control point, while access through the access controller only grants one authorization.

9.2.2 Fire Detection System

The fire detection system in a hospital is vital for the safety of the personnel and patients. Fire is not uncommon in a hospital environment, either malicious or accidental, and false alarms are even more common.

The main components in such a system are the sensors and the control panel which supervises the whole building. In providing physical security to the healthcare sector, there is often a larger monitoring system, which allows operators to inspect and detect various attacks which could affect equipment and put the personnel of the hospital in harm’s way. In installing a fire detection system, it is necessary that each room should have specialist sensors to detect both heat and smoke. Each fire incident detected will directly be forwarded as an alert to the control panel, and so on to the people which are in charge of the monitoring, can act fast, and call the firefighters.

In case of fires, integrated physical security systems can be used to diagnose causal or contributory factors such as:

1. The bad wiring of a power socket;
2. Inflammable substances left unattended in certain operations rooms;
3. An arson started by a malicious intended person;
4. A bombing attack;
5. Mishandling of electrical equipment.

As regards the use of video in fire detection, lately various convolutional neural networks (CNN)-based methods have been applied in specific environments with reasonable accuracy and execution time. However, those approaches failed to detect fire in uncertain environments, for instance, those having excessive smoke, fog, fire, and snow. Furthermore, achieving efficiency with reduced running time and model size is quite challenging for resource-constrained devices, such as edge-based analytics, i.e., within cameras, motivating the centralized approach of VMS-based video analytics.

A CNN-based method, illustrated in Figure 9.3, can thereby be used for fire detection in videos of health facility. The approach can be extended for the extraction of detailed contextual information from fire scenes such as an object on fire, burning degree, and fire growth rate, etc. Furthermore, a hybrid system can be developed by integrating smoke detection methods with the current work for

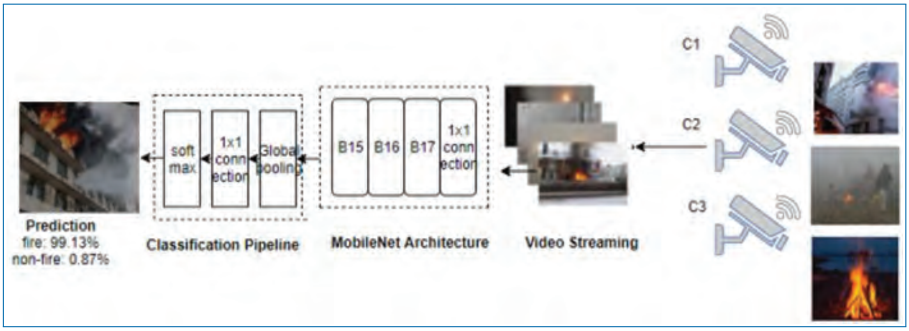


Figure 9.3. Efficient deep CNN for fire detection in video captured in uncertain environment C1, C2, and C3.

intelligent management of fire disasters. Finally, such an approach can be combined with industrial systems, 5G IoT, traffic, and robotics for more safe automation, traveling, more vibrant, and trustworthy experience (Muhammad *et al.*, 2018).

9.2.3 SCADA, ICS, and Smart Building Sensors

SCADA is a control systems architecture, based in both software and hardware, that has many benefits for the industry. SCADA's features include to process real-time data, record specification into a log document, control mechanical procedures, and connect further devices. The architecture consists of several interconnected elements, each with a different purpose and design, varying from a Remote Telemetry Unit (RTU) that interacts with the physical environment to an Human Machine Interface (HMI) that connects with the users (Rodofile *et al.*, 2017). An Industrial Control System (ICS) is one of the various kinds of control systems used to monitor industrial processes. Depending on the size, it can be made up of several controllers or a complex network of interactive control systems. These systems obtain data from remote sensors that monitor and measure process variables that will be compared with set-points. Hence, SCADA and ICS infrastructure capture data relevant to security issues which can affect the well-being of the personal and patients in a hospital environment.

A SCADA system can be used to connect to sensors and actuators which are in charge of collecting various parameters from devices on the field. The signal sent by SCADA devices is stored in an analogue format, and it is converted by a RTU, a Programmable Logic Controller (PLC), or a Intelligent Electronic Device (IED). After this process is done, the converted data is sent via a communication channel to the respective SCADA presentation and the control unit, whereas the sent data is analyzed, and each operation is sent back to each sensor (Mobolarinwa, 2017). The communication between devices and the SCADA host can be classified as dial-up, satellite, telephone, radio, and Wireless Local Area Network (WLAN). Within a SCADA system exists four layers, such as a collection, conversion, communication, and control layer. Each of these layers can be used as an attack entry point into the system, because protocols such as WLAN do not have authenticity and encryption from the manufacturing phase. Subsequently, the data sent to end-devices can be intercepted by an attacker and also operational errors can lead to a vulnerability in a HMI, and this can be exploited by the malicious intended attacker.

A recent approach called "Tactile Internet" involves gathering multiple technologies by permitting intelligence through mobile edge computing and data transmission over a 5G network, though time will tell whether this approach gains traction in the healthcare space.

Major classes of security vulnerabilities (Mobolarinwa, 2017) in the Industrial SCADA IoT Infrastructure are:

- Human Machine Interface (HMI) vulnerabilities: Hard-coded Credentials, Poor Input Field Validation, Poor Authentication and Authorization, Zero-day Exploits
- PLC vulnerabilities
- Social Engineering
- Inadequate Physical Security
- SCADA Protocol Vulnerabilities
- Connection with the Corporate Network

9.3 Cyber-security in Healthcare Contexts

As introduced previously, the healthcare sector faces unprecedented risks and compounding regulatory compliance requirements. It is usual that healthcare organizations have many assets that are essential for their operation and should be protected. Assets that can be attacked include the facilities and buildings themselves, data, interconnected clinical information systems, mobile devices, networking equipment, identification systems, networked medical devices, and remote care systems, with the two most critical hospital's assets being the interconnected clinical information systems and networked medical devices (Independent Security Evaluators, 2016). Patient records contain valuable information, such as Personal Identifiable Information (PII) and Protected Health Information (PHI), that can be the most attractive information for attackers. Healthcare organizations and their assets suffer from vulnerabilities that can be technical (application & OS, control gaps and design flaws, unpatched devices, unprotected networks, weak credentials, lack of cyber threat prevention and detection, lack of smart sensors, remote access policies, lack of employee training and awareness, etc.) or organizational and social (behavior of users, human errors, etc.).

These vulnerabilities can be exploited in different ways by attackers that use different types of malicious actions (e.g., virus, ransomware, hijack). The probability of these attacks can increase as healthcare organizations suffer also from system failures (e.g., software, hardware and network failure, inadequate firmware); human errors (users systems' misuse, unauthorized access, absence of audits and logs, etc.); and natural phenomena. Attackers have different goals, as they might wish to cause damage, obtain a ransom, cause the interruption of service, or collect data to prepare future attacks.

As such, health infrastructure is identified as a significant potential target of cyberattacks, which highlights the need to enhance protection from them. In order for healthcare organizations to prevent, or at least reduce, unauthorized access, use, disruption, deletion, and corruption, to respond effectively, quickly, and efficiently, and to minimize the impact of attacks to their networks and systems, it is important to take organizational and technical measures, such as those nominated below.

With regard to organizational measures that will enhance cybersecurity in healthcare organizations, it has been widely claimed that it is important for healthcare organizations to assess cyber risks. Cyber risk assessments are used to identify, estimate, and prioritize risk to organizational operations, organizational assets, individuals, other organizations, and national concerns, resulting from the operation and use of information systems (NIST, 2019b).

In addition, healthcare organizations should develop and incorporate both generic and case-specific laws, standards, plans, and policies that outline cybersecurity measures and crisis management procedures, such as the NIS directive (EU, 2016) and ISO 27001 (ISO, 2019), security procedures application in order to protect the venue and other sensitive, critical, or valuable assets and areas (e.g., computer room, central servers, clinical information systems, and electronic healthcare records) from attacks.

Since the human factor is one of the major security threats in the health domain, it is important that personnel are aware of the basic cybersecurity-related issues and their skills—both technical and behavioral—are improved (ECSO, 2018). Moreover, healthcare staff (including researchers, administrators, front desk workers, medics, transcriptionists, handlers of medical claims to IT, and technical staffs) should be properly trained on cybersecurity protection and crisis management issues, standards, plans, and protocols (Martin *et al.*, 2017). In doing this, stakeholders that find themselves affected by, or actively seek involvement in crisis management processes, can manage and cooperate effectively and in timely fashion on security planning, preparedness, response, recovery, and impact mitigation. With regard to technical measures, it has been reported that healthcare organizations should adopt and implement different practices that will enhance data, systems, devices, and networks security, such as the following, according to ISO (2018) and NIST (2019a):

Authentication ensures the validity of the claimed identities of the entities participating in communication (e.g., person, sensors, service or application) and provides assurance that an entity is not attempting an unauthorized replay of a previous communication.

Access control (authorization)—much like physical access control, described above—guarantees that only individuals, as well as software and IT infrastructure, can only gain access to, and perform operations on, stored information and flows that they are authorized for. Unlike physical access control, different access levels can be granted to systems, devices, and networks.

Availability describes a security dimension that ensures there is no denial of authorized access to network elements, stored information, information flows, services, and applications due to events impacting the network.

Reliability has been defined as the ability of the system to perform its functions for a period of time. This is a high-level security requirement and to be achieved different mechanisms should be implemented (e.g., availability, communication security), as described in the respective sections above.

Non-reputation refers to the ability to prevent an individual or entity from denying having performed a particular action related to data, by making available proof of various network-related actions (such as proof of obligation, intent, or commitment; proof of data origin; proof of ownership; proof of resource use).

Data confidentiality ensures that the data content cannot be understood by unauthorized entities.

Data integrity is a security dimension that ensures the correctness or accuracy of data. Data should be protected against unauthorized modification, deletion, creation, and replication and provide an indication of these unauthorized activities.

Backup is the process of backing up the operational state, architecture and stored data of database software.

Tracing systems should log access and errors to the collected and stored data (e.g. time, date, users' accessing the system, fails, wrong password).

Log files are automatically produced files, recording events, messages from certain software and operating systems.

Communication security is the security dimension that ensures that information flows only between the authorized end points; i.e., the information is not diverted or intercepted as it flows between these end points. To obtain communication security, mechanisms such as encryption through Secure Sockets Layer (SSL), Virtual Private Networks (VPNs), timestamps, auditing and restricting access per-user-group should be implemented.

To secure networked devices and assets in the healthcare, it has been reported that: (a) inventories should be created and maintained, as they can ensure a sound understanding of the systems and their components, support configuration, and automated remediation management processes (Independent Security Evaluators, 2016); and (b) software should be regularly patched and updated.

In addition, the network can be protected through the implementation of a firewall and thereby segmentation and segregation techniques. Moreover, monitoring mechanisms should be employed, so as to support: (a) network protection from attacks, e.g., Intrusion Prevention Systems that detect threats over the network by examining communications and scanning ports for anomalies and can execute a real-time response to stop an immediate threat, detection of attacks, i.e., Intrusion Detection Systems that monitor systems, network traffic, data, and files access, etc. and detect attacks; and (c) response to attacks (Intrusion Response systems that choose the necessary action to take to respond to attacks and ensure the security of networks and computational system.

Finally, a security-by-design approach would complete the above countermeasures, focusing on the cybersecurity concerns with respect to new devices or systems that need to be planned and implemented from the start of the procurement, design, development, and maintenance phases.

9.4 Conclusion

This chapter has presented how security systems can be applied within the healthcare sector. The SAFECARE[1] project[2] is working to provide an integrated solution for both physical and cybersecurity in the healthcare domain. In terms of physical security, video surveillance, access control, fire detection, and building management

[1] The SAFECARE project has received funding from the European Union's H2020 research and innovation programme under Grant Agreement no. 787002.

[2] https://www.safecare-project.eu/

sensors are combined with video analytics and novel rule support across the various modalities of input data from all of these systems. In this way, integrated systems are made capable of signaling security incidents via intrusion detection; fire detection; detection of attacks on building management systems, such as power and heating, ventilation and air conditioning (HVAC); and suspicious behavior detection. Further, this approach to security sits alongside state-of-the-art cybersecurity provisions and, for both, SAFECARE provides sophisticated analyses of impact propagation, as described in a later chapter.

Acknowledgments

The SAFECARE project has received funding from the European Union's H2020 research and innovation programme under Grant Agreement no. 787002.

References

ECSO. 2018. *Healthcare Sector Report – Cyber security for the healthcare sector*. European Cyber Security Organisation (ECSO), Rue Montoyer, 10, 1000 Brussels Belgium.

EU. 2016. *Directive (EU) 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*. URL: https://eur-lex.europa.eu/eli/dir/2016/1148/oj (accessed on 12/10/2019).

Independent Security Evaluators. 2016. *Securing Hospitals: A research study and blueprint*. URL: https://www.securityevaluators.com/wp-content/uploads/2017/07/securing_hospitals.pdf.

ISO. 2018. *ISO 22300:2018-Security and resilience—Vocabulary*. URL: https://www.iso.org/obp/ui/#iso:std:iso:22300:ed-2:v1:en (accessed on 12/10/2019).

ISO. 2019. *ISO/IEC 27001:2013 [ISO/IEC 27001:2013] Information technology—Security techniques—Information security management systems—Requirements*. URL: https://www.iso.org/standard/54534.html (accessed on 12/10/2019).

Martin G., Martin P., Hankin C., Darzi A., and Kinross J. 2017. "Cybersecurity and healthcare: how safe are we?" *BMJ* 358(j3179).

Mobolarinwa T. Balogun. 2017. *A Comparative analysis of heathcare system IoT and Industrial SCADA IoT for Cyberterrorism*. URL: https://pdfs.semanticscholar.org/75db/5aae5318ee60d43db4b2fcc46aadbdbfe1be.pdf (accessed on 01/27/2020).

Muhammad, K., R. Hamza, J. Ahmad, J. Lloret, H. Wang, and S. Baik. 2018. *Secure Surveillance Framework for IoT systems using Probabilistic Image Encryption*. URL: https://www.researchgate.net/publication/322408159_Secure_Surveillance_Framework_for_IoT_systems_using_Probabilistic_Image_Encryption.

NIST. 2019a. *NIST – Digital Identity Guidelines*. URL: https://pages.nist.gov/800-63-3/sp800-63-3.html#def-and-acr (accessed on 12/10/2019).

NIST. 2019b. *NIST glossary*. URL: https://csrc.nist.gov/glossary/term/RA (accessed on 12/10/2019).

ONVIF (2020). *Open Network Video Interface Forum*. URL: https://www.onvif.org/ (accessed on 01/14/2020).

Rodofile, Nicholas R. and Radke, Kenneth and Foo, Ernest. 2017. *Framework for SCADA Cyber-Attack Dataset Creation*. URL: https://doi.org/10.1145/3014812.3014883.

DOI: 10.1561/9781680836875.ch10

Chapter 10

Integrated Cyber-Physical Security Approach for Healthcare Sector

By Fabrizio Bertone, Francesco Lubrano, Marco Gavelli, Olivier Terzo, Elisabetta Biasin, Erik Kamenjasevic, Samantha Dauguet-Demailly, David Lancelin, Silvia Andernello, Francesco Tresso, Luca Viarengo and George Suciu

10.1 Introduction

Modern societies are strongly dependent on the continuous function of Critical Infrastructures (CI) that ensure the supply of crucial goods and services such as power, Information and Communication Technologies, or drinking water. Critical Infrastructures are essential for the maintenance of vital societal functions, such as health, safety, security, economic or social well-being of people, etcetera. These aspects are also relevant with regard to the healthcare sector, where any interruption, damage, or unavailability of healthcare services may provoke economic and non-economic damages for individuals, organizations, States, and society as a whole.

The healthcare sector is among the most critical sectors in Critical Infrastructure Protection (CIP). Healthcare services considered "critical" are, for instance, emergency healthcare; hospital care (inpatient & outpatient); the supply of pharmaceuticals, vaccines, blood, medical supplies; and infection/epidemic control, to name but a few. The disruption of one of these critical healthcare services could imply several damages for society. This happened, for instance, after the Wannacry ransomware attack on the National Healthcare Services (NHS), in the UK (Ghafur *et al.*, 2010). According to NHS England, the ransomware affected at least 80 out of 236 trusts across England, because they were infected by the ransomware or turned off their devices or systems as a precaution. Furthermore, 603 primary care and other NHS organizations were infected, including 595 GP practices. Thousands of appointments and operations were cancelled, and in five areas, patients had to travel further to reach accident and emergency departments.

Having recognized the increasing role of CIP, the EU legislator and majority of the EU Member States have adopted national strategies to increase the level of protection of critical infrastructures in the EU. Concerning CIP, in the last few years, many of the Member States adopted national CIP strategies and consider healthcare as one of the sectors requiring protection.

Nevertheless, to protect Critical Infrastructures such as a hospital is a huge and very complex task that requires particular attention and knowledge of defense and prevention strategies, as well as of vulnerabilities and potential attacks that may occur.

Critical Infrastructures such as hospitals are constantly threatened by different kinds of potential attackers with different resources available. Some could be simply motivated by visibility. Others could be driven by profit gain (Sultan *et al.*, 2018; Tonutti, 2016). The last few decades have also seen an increase of state-sponsored attacks, which can be motivated by espionage, retaliation, intimidation or as a stealth way to create disruption in case of escalating conflicts (Geers *et al.*, 2013).

Dangerous threats are also presented by increasing terrorist activities in recent years. While, traditionally, terrorist attacks have mostly targeted the physical world, cyberattacks are getting more and more popular, for financing purposes, to collect intelligence information or to cause disruption. When expertise is not available internally, other people can be persuaded to do the job without knowing the real objective (Mitnick and Simon, 2010).

Nowadays, physical and cyber systems are more and more interconnected, in some cases being so integrated to be indivisible. From a security point of view, this greatly enhances the attack surface and the possibility for remote actors to reach their goals. Moreover, physical intrusions are still possible and currently used by criminals.

Let us consider the motivation for a generic attacker to access confidential information stored in local servers inside a hospital. This could be achieved by various means and exploiting different levels of physical and cyber intrusions.

A more "traditional" attack would be to enter the hospital's premises and collect the desired information manually. An evolved remote attacker would rather look for a chain of vulnerabilities in the "cyber" services exposed by the hospital and get the data from a completely different country. In the middle, there are many different mixes of physical and cyber steps that can allow the accomplishment of the same goal. Social engineering techniques could be used to trick hospital staff in order to gain access to the systems (Medlin *et al.*, 2008). Alternatively, a USB drive infected with malware could be given as a gift to a doctor during a conference (De Falco, 2012; Cluley, 2010).

It is therefore clear that physical and cyber threats should be considered, analyzed, and treated together, as cyber-physical threats. An integrated approach that considers both physical and cyber worlds is therefore required.

This chapter presents the description of threats, potential incidents, and issues regarding the protection of the critical infrastructures like hospitals, and it presents the first results of the SAFECARE project, describing the internal architecture of the whole system.

10.2 Safecare Approach

The idea behind the SAFECARE project is to respond to the growing demand for an integrated cyber-physical security solution for Critical Infrastructures, in particular hospitals. The challenge is to bring together the most advanced technologies from the physical and cyber security spheres, to achieve a global optimum for systemic security and for the management of combined cyber and physical threats and incidents, their interconnections and potential cascading effects. Indeed, the main objective of the SAFECARE project is to increase the protection and resilience of Healthcare facilities and services, allowing for a better response in case of emergencies. This is done using a holistic approach that considers the physical and cyber worlds in a single integrated system.

In addition, innovative services enable first responders and other relevant actors to get real-time updated information about the availability of healthcare services. This is useful both in case of incidents inside the facility itself, so that the information about unavailable services is easily accessible, and in case of large-scale emergencies such as earthquakes that can require the involvement of many facilities and an efficient routing of patients.

Within the SAFECARE project, a global architecture schema that includes all the physical, cyber, and cyber-physical module has been designed. This architecture can be broken down into 3 parts, as shown in Figure 10.1:

- Physical security solutions
- Cybersecurity solutions
- Integrated cyber-physical security solutions

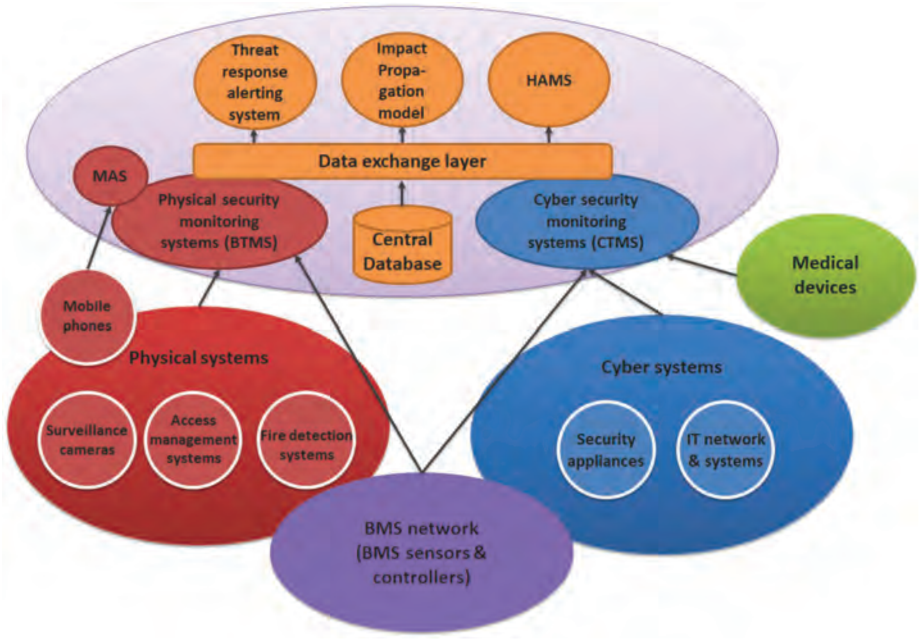

Threat response alerting system
Impact Propa-gation model
HAMS
Data exchange layer
MAS
Physical security monitoring systems (BTMS)
Central Database
Cyber security monitoring systems (CTMS)
Medical devices
Mobile phones
Physical systems
Surveillance cameras
Access management systems
Fire detection systems
Cyber systems
Security appliances
IT network & systems
BMS network (BMS sensors & controllers)

Figure 10.1. SAFECARE global architecture.

The physical and cybersecurity solutions consist of smart modules and efficient integrated technologies to, respectively, improve physical security and cybersecurity.

More specifically, physical security solutions embed integrated intelligent video monitoring and interconnect building monitoring systems as well as management systems. Meanwhile, cybersecurity solutions consist of cyber monitoring systems as well as threat detection systems related to Information Technology (IT), Building Management Systems (BMS), and e-health systems.

Both physical and cybersecurity monitoring tools are interconnected thanks to the integrated cyber-physical security solutions. They consist of intelligent modules whose role is to integrate different data sources and better take into account the combination of physical and cybersecurity threats.

In order to fulfill their role, each solution ensemble is composed of several dedicated systems.

The physical security solutions rely on:

- The Suspicious Behavior Detection System;
- The Intrusion and Fire Detection System;
- The Data Collection System;
- The Mobile Alerting System;
- The Building Threat Monitoring System.

The Building Threat Monitoring System (BTMS) is the aggregation point for physical security and makes the link between physical systems and the rest of the architecture. The Mobile Alerting System (MAS) is intended for local security operators, providing them with a quick way to report physical security events and visualize contextual information.

The cybersecurity solutions rely on:

- The IT Threat Detection System;
- The BMS Threat Detection System;
- The Advanced File Analysis System;
- The E-health Devices Security Analytics;
- The Cyber Threat Monitoring System.

Just like the BTMS, the Cyber Threat Monitoring System (CTMS) connects the cyber systems to the rest of the architecture and is the central point for cybersecurity.

Finally, the integrated cyber-physical security solutions rely on:

- The Data eXchange Layer;
- The Impact Propagation and Decision Support Model;
- The Threat Response and Alert System;
- The Hospital Availability Management System;
- The E-health Security Risk Management Model.

The Data eXchange Layer (DXL) enables communication between all of the SAFECARE subsystems. It works in pairs with the central database which stores static and dynamic data characterizing the whole system. The Impact Propagation and Decision Support Model (IPDSM), the Threat Response and Alert System (TRAS), along with the Hospital Availability Management System (HAMS) are three decision-making modules. They respectively enable inferring cascading impacts of physical and/or cybersecurity incidents, alerting internal and external practitioners (or any other appropriate defined response) upon reception of an "impact" message from the IPDSM, and providing information about health services availability.

10.2.1 Intercommunication Layer and Central Database

In order to cope with such a combined approach for the management of cyber and physical security, the SAFECARE project implements a central database and a common data exchange layer to connect the different modules of the SAFECARE platform.

Dentral Database

The SAFECARE Central Database is a single, unique repository that stores multiple types of data needed for the other modules in the platform. In particular, two different categories of data have been identified and modeled in the database:

- Static data, all the information related to assets, facilities, buildings, and services inside the hospital. Furthermore, this category includes interconnections and relations among the assets.
- Dynamic data, all the information that is generated by SAFECARE modules, such as incidents, impacts, and all the other responses/messages. Relations among incidents, impacts, etc. are also represented in the database and can be used for further analysis.

Data eXchange Layer

The Data eXchange Layer constitutes the core of the communication layer in the SAFECARE architecture. It allows all the other modules to communicate with each other in near real time and provides relevant interfaces to extract data stored in the database. Five types of dynamic-data messages are defined:

- Incident: message generated by the monitoring tools; it reports information related to the incident, it is validated by human operators, and it triggers decision-making modules.
- Impact: reports the potential impacts after an incident occurs allowing prevention of potential cascading effects.
- Threat response: provides a predefined reaction plan to mitigate the effects of incidents and improve time to response.
- Notification: exchange the communication between Threat Response and Alerting System and Mobile Alerting System.
- Availability: reports the updated availability of assets involved in the incident.

10.2.2 Cyber and Physical Security Solutions

In order to detect possible incidents, some monitoring systems are required. This section describes the set of tools that are integrated in SAFECARE for this kind of job, logically subdivided between physical and cybersecurity.

Building Threat Monitoring System

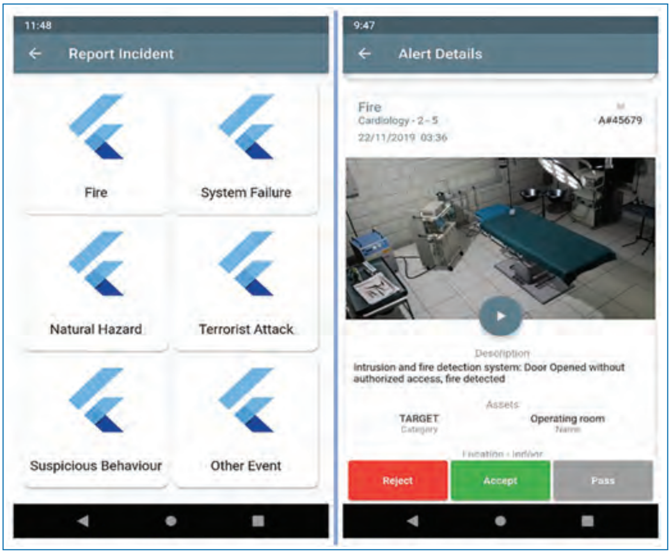
The Building Threat Monitoring System is the module in charge of monitoring the physical assets. BTMS is an event-based server that tracks physical events coming from different subsystems, such as: the Suspicious Behavior Detection System, that analyzes the video surveillance detecting irregular movements or behavior such as loitering or tailgating; the Intrusion and Fire Detection System, that is connected to the access control system and to the fire alarm system of the hospital; the Data Collection System, that collects data from many different type of sensors and controllers; and the Mobile Alerting System. The BTMS is the central point for communicating physical incidents, which are alerts that have been judged to require a security response by operators in charge.

Finally, the BTMS is also responsible for receiving and relaying the incident handling responses elaborated by the Impact Propagation and Decision Support Model.

Mobile Alerting System

Smartphones and tablets are powerful network-connected devices, constantly available and low cost; therefore, they are perfect tools for widespread use by human operators.

Through the MAS, coupled with the mobile app specifically developed in SAFECARE, a building security officer via a smartphone has the ability to quickly report specific categories of security threats or alerts (system failure, natural hazard, terrorist attack, etc.), as depicted in Figure 10.2(left). On the other side, automatic alerts generated by detection systems can be validated or cancelled by the operator as can be seen in Figure 10.2(right), where a false fire alarm is shown.



Intrusion and fire detection system: Door Opened without authorized access, fire detected

Figure 10.2. Incident reporting (left) and alert validation (right).

Cyber Threat Monitoring System

The objective of the Cyber Threat Monitoring System is to collect and centralize security events from the cyber threat detection systems, organize the information, and provide user-friendly interfaces to SOC[1] operators so that they can visualize the threats and have an overview of the potentially impacted assets.

The CTMS receives security events from the following systems: the IT Threat Detection System (ITDS) that monitors the IT network and receives log messages from the different components in order to detect threats targeting the IT infrastructures; the BMS Threat Detection System (BMSTDS) that analyzes the Operational Technologies (OT) protocols used in building automation systems (such as SCADA systems and PLC controllers); the advanced file analysis system that performs in-depth analysis of files extracted by the ITDS or the BMSTDS, thus allowing malware detection; the e-health devices security analytics that monitors medical devices by collecting their log messages and rely on an e-health security risk management model to identify any related risk.

Rules are implemented within the CTMS to automatically generate alerts from the received security events. The CTMS is the entry point for SOC operators to monitor in real time all incoming alerts regarding cyber threats as it centralizes them. After a first analysis phase, the SOC operators must confirm the alerts as either incidents or false-positive alerts. From there, the CTMS enables tracking of incidents and coordination of incident responses.

Finally, the CTMS receives potential impacts, which are computed from both physical and cyber incidents by the Impact Propagation and Decision Support Model, in order to provide SOC operators with a clear understanding of potential impacted assets and services.

10.2.3 Integrated Cyber-physical Security Solutions and Decision Support

This section provides a brief description of the SAFECARE subsystems that handle incidents that generate potential impact and cascading effects, alerting relevant recipients following predefined reaction plans and providing updated information related to the hospital status.

Impact Propagation and Decision Support Model

The ability to simulate the propagation of impacts caused by incidents and to mitigate risk is the cornerstone of the SAFECARE project. The module in

[1] Security Operation Centre.

charge of these functionalities is the Impact Propagation and Decision Support Model.

The objectives of the IPDSM are:

- Combine physical and cyber incidents that occur on assets
- Infer cascading effects as impacts that could potentially affect the same or related assets
- Alert other modules about the potential impacts and severity.

In order to reason about incidents and their potential impacts, the IPDSM needs to know detailed information about physical and cyber assets and their relations. This information is collected in a custom ontology defined for the project. Following incidents, the IPDSM simulates a set of potential impacts on directly or indirectly involved assets. This is done by employing a set of rules derived by domain knowledge.

Threat Response and Alert System

The Threat Response and Alert System is a specific module devoted to alerting relevant recipients by providing information about incidents, potential impacts and sharing the predefined reaction plan, according to incident type and severity. It is activated by an "impact" message received from the IPDSM through the DXL. Once triggered, the module runs the corresponding predefined response plan and alerts internal and external practitioners via different media (SMS, emails, phone calls,…) and possibly also by using the MAS.

Hospital Availability Management System

The Hospital Availability Management System service aims to improve the resilience of health services and the communication of availability information among hospital staff and first responders. The HAMS is an integral part of the incident management process in SAFECARE. Based on incidents that are received from monitoring modules, it updates the availability of assets involved, considering the incident nature and the asset category. Once the impacts are reported, HAMS can examine them, updating the availability of assets that are involved (even indirectly) in the incident. Furthermore, HAMS provides a web interface with which users can check the status of the hospital and eventually manually update resources/availability status. Finally, HAMS provides an interface to export hospital status/information compliant with the EDXL-HAVE standard.[2]

[2] http://docs.oasis-open.org/emergency/edxl-have/v2.0/edxl-have-v2.0.html

10.3 Ensuring Security, Privacy, and Data Protection within the EU Legal Requirements

Security and confidentiality are key factors when it comes to privacy and data protection. In that regard, healthcare infrastructures process on a daily basis personal health-related data of vulnerable individuals (i.e., patients), due to the nature of the services they provide. These kinds of activities are likely to result in a high risk, especially when they are performed on a large scale. If a healthcare infrastructure falls victim to an attack and a security incident occurs, appropriate steps should be taken. To do so, it is important to follow procedures determined by the relevant legal frameworks on incident reporting and notification.

The paragraphs that follow provide a brief overview of the applicable EU security, privacy, and data protection legal requirements that may be considered when dealing with reporting and notification of incidents.

10.3.1 Security of Networks and Information Systems

Network and information system security is a matter that has been regulated at European level in 2016 with the NIS Directive. This legislative instrument has provided a minimum set of rules ("harmonization") with the aim of achieving a common level of security resilience across the European Union. Every Member State has to transpose the Directive via national legislation. The NIS Directive requires entities providing services considered "essential" (i.e., "Operators of Essential Services" or "OES"—e.g., healthcare providers such as hospitals and private clinics) to ensure the security of their network and information systems and to adopt a risk-based approach.[3] OES must put in place technical and organizational measures appropriate to the risk posed to their networks and information systems. Among these, OES/healthcare providers should enact measures aimed at preventing, detecting, and handling incidents[4] and at mitigating their impact.

(Security) Incident: prevention, detection and notification under the NIS Directive

The NIS Directive has established the duty for operators to notify, without undue delay, to the competent authorities or Computer Security Incident Response Teams

[3] See art. 5(2) NIS Directive for the definition and criteria of identification of OES, which have to be identified by the Member States.

[4] Incidents are defined by the NIS Directive as "any event having an actual adverse effect on the security of network and information systems" (art. 4(1)(7)).

(CSIRTs) incidents having a significant impact[5] to the continuity of the essential services they provide. This requirement implies that OES/healthcare providers must set up measures to detect incidents as they have to be prepared to gather key information on incidents to be notified to the competent authorities. Furthermore, OES/healthcare providers should notify incidents *as soon as they can.* As cybersecurity incidents are dynamic and the situation can change rapidly, operators should first send an immediate alert notification to the national competent authority and/or CSIRTs in order to allow them, for instance, to offer support concerning the handling of the incidents or to assess the potential impact for essential services, individuals, society, economy, etc. An incident notification may happen via different means, such as a phone call, a plain email, a web service, an online paper. Procedures regarding modalities of incident reporting and the information that has to be provided (which may concern the nature of the incident, the impact of the incident, operational information such as time or status, etc.) may vary between Member States, as they must be determined by each MS individually.

10.3.2 Security of Personal Data

Integration of security architecture in hospital's infrastructure in order to prevent security incidents from happening entails the application of EU privacy and data protection laws (i.e., GDPR). Unlike the NIS Directive, the GDPR is directly applicable in all EU Member States. Healthcare organizations, healthcare professionals, and healthcare staff are bound by the requirements of the GDPR. The Regulation requires all persons and legal entities (e.g., healthcare providers) acting as controllers[6] to abide by the key principles of data protection laws and shall be responsible for and be able to demonstrate compliance with the law. With regard to security, the integrity and confidentiality principles require from the healthcare providers to process personal data securely. This shall include protection against unauthorized or unlawful processing and against accidental loss, destruction, or damage; and it must imply the use of appropriate technical and organizational measures according to the risk inherent to the processing.

[5] Parameters to determine the "significance" of the impact of an incident are listed under art. 14 NIS Directive, i.e.: (a) the number of users affected by the disruption of the essential service; (b) the duration of the incident; (c) the geographical spread with regard to the area affected by the incident. The NIS Coordination Group provided further guidance in this regard. See: Reference document on Incident Notification for Operators of Essential Services (February 2018); Guidelines on notification of Operators of Essential Services incidents (May 2018).

[6] The 'controller' is "the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law" (art 4(1)(7) GDPR).

Personal data breaches: prevention, detection and notification under the GDPR

The GDPR also requires healthcare infrastructures, when acting as controllers, to notify the supervisory authority in case a personal data breach[7] occurs. The concept of "personal data breach" is close to the NIS Directive concept of "incident" analyzed above. However, it differs significantly in scope: the former concerns personal data only, whereas the latter concerns any kind of security incidents. In other words, every personal data breach is a security incident, but not every security incident is necessarily a personal data breach.

As the Article 29 Data Protection Working Party puts it, a key element of any data security policy is being able, where possible, to prevent a breach and, where it nevertheless occurs, to react to it in a timely manner. Personal data breaches must be notified to competent authorities without undue delay, and no later than 72 hours after becoming aware of it. The notification should describe the nature of the personal data breach including the categories and approximate number of data subjects concerned as well as the categories and approximate number of personal data records concerned. Data breaches must also be notified to data subjects when the breach is likely to result in a high risk for their rights and freedoms.

10.3.3 Relevance of the NIS and Privacy and Data Protection Requirements within the SAFECARE Framework

The solutions presented within the SAFECARE architecture are aimed at establishing monitoring mechanisms and internal incident detection mechanisms. By monitoring and preventing incidents, these solutions may thus represent a security measure with which healthcare providers may manage the risks posed to security of their network and information systems, including to the risk posed to the processing of patients' data concerning health. By doing so, healthcare providers may be facilitated in their process of compliance with prevention, detection, and notification requirements set by the NIS Directive and the GDPR.

10.4 Conclusions

The threats that target critical infrastructures, in particular the healthcare sector, are multiple and manifold. The actors that can act against critical infrastructure, their

[7] A personal data beach consists in "a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to personal data" (art. 4(1)(12) GDPR).

motivations, and means of operating can also be varied and involve either physical and cyber assets, or a combination of them. These reasons explain the motivation behind the need for an integrated cyber-physical security solution.

The SAFECARE project conceived an integrated cyber-physical security approach and designed an architecture that combines together different monitoring and management tools, each considering a specific aspect of the global solution. Assets, vulnerabilities, threats, incidents, and impacts are all considered together with their dependencies, forming a shared intelligence that greatly enhances the value of each single piece of data. This approach allows us to extract much more information and uncover possible menaces previously unseen.

Finally, an important aspect to consider while implementing an organization's security plan is the compliance with relevant legislation. For this reason, security, privacy, and data protection requirements have been analyzed from a legal point of view, giving a brief overview of the relevant legislation concerning SAFECARE framework.

Acknowledgments

The SAFECARE project has received funding from the European Union's H2020 research and innovation programme under Grant Agreement no. 787002.

References

Cluley, G. (2010, May 21). *IBM distributes USB malware cocktail at AusCERT security conference*. Retrieved from Naked Security: https://nakedsecurity.sophos.com/2010/05/21/ibm-distributes-usb-malware-cocktail-auscert-security-conference/

De Falco, M. (2012). *Stuxnet Facts Report: A Technical and Strategic Analysis*. Tallin: NATO Cooperative Cyber Defense Centre of Excellence.

Geers, K., Kindlund, D., Moran, N., and Rachwald, R. (2013). *WORLD WAR C: Understanding Nation-State Motives Behind Today's Advanced Cyber Attacks*. FireEye.

Ghafur, S., Kristensen, S., Honeyford, K., Martin, G., Darzi, A., and Aylin, P. (2019). A retrospective impact analysis of the WannaCry cyberattack on the NHS. *NPJ Digital Medicine*.

Medlin, B. D., Cazier, J. A., and Foulk, D. P. (2008). Analyzing the Vulnerability of U.S. Hospitals to Social Engineering Attacks: How Many of Your Employees Would Share Their Password? *International Journal of Information Security and Privacy (IJISP)*, 71–83.

Mitnick, K. D. and Simon, W. L. (2005). When Terrorists Come Calling. In K. D. Mitnick, and W. L. Simon, *The Art of Intrusion: The Real Stories Behind the Exploits of Hackers, Intruders and Deceivers* (pp. 23–47). John Wiley & Sons.

Sultan, H., Khalique, A., Tanweer, S., and Alam, S. I. (2018). A Survey on Ransomeware: Evolution, Growth, and Impact. *International Journal of Advanced Research in Computer Science*.

Tonutti, S. (2016, September 16). *Genetic Data Theft as a new type of Biocrime: legal and social aspects of genetic privacy*. Retrieved from Privacy Genetica: http://www.privacygenetica.it/2016/09/genetic-data-theft-as-a-new-type-of-biocrime-legal-and-social-aspects-of-genetic-privacy/

DOI: 10.1561/9781680836875.ch11

Chapter 11

Vulnerability and Incident Propagation in Cyber-physical Systems

By Faten Atigui, Fayçal Hamdi, Nadira Lammari and Samira Si-said Cherfi

11.1 Introduction

Hospitals are cyber-physical systems that are vulnerable by nature to a multitude of attacks that can occur at their communication, networking, and physical entry points. Such cyber-physical attacks can have detrimental effects on their operation and the safety of their patients. Thus, to properly secure these systems, it is of utmost importance to: (i) understand their underlying assets with related vulnerabilities and associated threats, (ii) quantify their effects, and (iii) prevent the potential impacts of these attacks. This implies addressing a challenging objective of understanding the tight relationships between the asset's characteristics and the propagation of attack's effects to better prevent the impacts and consequences of incidents. Such an approach needs a detailed knowledge of intrinsic and contextual assets properties. However, hospitals host a variety of medical and IT assets with very different characteristics. The next section reports on the state of the art of assets and assets interdependencies modeling as well as on incidents propagation approaches.

11.2 Related Work

This section presents existing work on impact propagation of incidents and the methods used to assess the severity of incidents and risks.

11.2.1 Characterization of Dependencies Between and Within Critical Infrastructures

Although the terms "dependency" and "inter-dependency" are commonly used interchangeably, some research work distinguish them. The consensual distinction is this of Rinaldi *et al.*, 2001. The authors define a dependency as a relationship between two infrastructures in a single direction, whereas inter-dependency is bidirectional (implicitly multi-directional) with two (implicitly more) infrastructures influencing each other. This definition is also shared by Stapelberg (2008). A more precise definition of the dependency concept is given by Schmitz *et al.* (2007). The European Union Agency for Cybersecurity (ENISA) proposes to consider dependencies within critical infrastructures (CIs) and dependencies between CIs. These kinds of dependencies are qualified as upstream, internal, or downstream dependencies in Petit *et al.* (2015). An upstream dependency expresses the fact that the products or services provided to one infrastructure by another external infrastructure are necessary to support its operations and functions. Downstream dependencies are the consequences to a critical infrastructure's consumers or recipients from the degradation of the resources provided by a critical infrastructure. Internal dependencies represent the internal links among the assets constituting a critical infrastructure. Therefore, upstream and downstream dependencies are between CIs, whereas internal ones are within CIs. Several works have focused on the characterization of dependencies between CIs. Zimmerman (2008) distinguishes spatial dependencies from functional ones. Rinaldi *et al.*, 2001 and Schmitz *et al.* (2007) propose a categorization of dependencies into physical, cyber, geographic, and logical ones. Dudenhoeffer *et al.* (2006) and Clemente (2013) consider physical, informational, geo-spatial, policy/procedural, and societal dependency. For reasoning purposes, Adetoye *et al.* (2011) propose another taxonomy of dependencies. They suggest considering five types of dependencies: generic, indirect, inter, co, and redundant dependency.

11.2.2 Models Serving the Incidents' Impact Propagation

In addition to the existing inter-dependencies between infrastructures, to deal with the impact of cascading effects that a disruption of an asset may have on the internal and external context of a critical infrastructure, one must also have, for each asset, a clear knowledge of the kinds of threats that could affect this asset, its vulnerabilities, and its relation to the other assets. These three aspects have been the subject of several research studies. From a threat perspective, to our knowledge, there is no research work that provides a high-level ontology of threats for CI. However, let us note that the European Commission reported a generic classification of threats for CI in which natural hazards are distinguished from non-malicious man-made hazards and malicious man-made hazards (Theocharidou and Giannopoulos, 2015). The HITRUST alliance have also published a threat taxonomy where at the top level logical, physical, and organizational threats are distinguished (HITRUST, 2019). Other works concentrate on specific threats. In ENISA (2016a) the most common threats affecting ICS/SCADA systems are shown. The top 10 threats affecting these systems have been published by CTED and UNOCT (2018). In the context of physical security risk assessments, Liu *et al.* (2012) propose a list of threats from terrorism. We can also find in "Common Criteria" and ANSSI portals security protection profiles for some software and physical equipment of CI where threats affecting these components are listed. In the context of the healthcare sector, ENISA (2016b) provided an overview of the cyber threats faced by smart hospitals. Taxonomies of threats for healthcare infrastructures are also proposed by Almohri *et al.* (2017) and Agrafiotis *et al.* (2018). Regarding the links between assets, we can consider research works that give much attention to the hierarchical links between assets (Silva and Jacob, 2018; Brocke *et al.*, 2014; Jakobson, 2011; Tong and Ban, 2014; Breier and Schindler, 2014). They model an infrastructure into levels to which the assets belong. The contributions differ in terms of kind and number of layers. The representation models used are also varied, ranging from simple oriented graphs to light ontologies. To define models that consider the hierarchical dependency between assets while emphasizing the links between assets within the hierarchical layers, one can rely on Enterprise Architecture (EA) modeling languages and standards or methodological guides existing in the industrial world. These tools are not specifically dedicated to critical infrastructures. As an example, we can mention ArchiMate 2.1, an open and independent EA modeling language within TOGAF Framework 9.2. We can also mention the CIM standard produced by DMTF (formerly known as the Distributed Management Task Force) that is internationally recognized by ANSI (American National Standards Institute) and ISO (International Organization for Standardization). There also exists several security risk analysis methodologies that give descriptions of assets, most of which are based on standards. These descriptions are very often informal and sometimes accompanied by catalogues. This is the case of EBIOS RM (EBIOS, 2019) and MAGERIT 3.0 (Amutio *et al.*, 2014) methodologies.

11.2.3 Incidents Propagation

Several approaches have dealt with the incidents propagation issue. We can classify these approaches into three categories: empirical, agent-based, and network-based approaches.

Empirical approaches analyze asset's interdependencies according to experts' opinions and past incidents traces. The underlying assumption is that it is difficult to identify assets' interdependencies in normal situations. Thus, analyzing the incidents helps rising intangible relationships among assets under extreme situations such as disasters, failures, or attacks. Laefer *et al.* (2006) defined accuracy, comprehensibility, timeliness, and accessibility of data as key characteristics to store, analyze, query, and visualize critical incident. This data could then be analyzed to mine records of frequent failure patterns as presented in Chou and Tseng (2010). To highlight the relationship between interdependencies and incident propagation, Mendonça and Wallace (2006) studied the 9/11 World Trade Center attacks and their impact on critical infrastructures and their services. The study showed that 20% of reported disruptions involved interdependency. Considering CI as "systems of systems" may improve response to incidents. Kotzanikolaou *et al.* (2013) combine common-cause and cascading events to assess the potential risk caused by complex situations. They considered the cumulative dependency risk of cascading chains.

Agent-based approaches consider a CIS as a complex adaptive system that could be analyzed as a complex phenomenon emerging from individual and autonomous agents. This kind of approaches captures all types of interdependencies among CIS by event simulations. It also provides scenario-based what-if analysis and the effectiveness assessment of different control strategies. Barrett *et al.* (2010) investigated cascading effects in three closely coupled systems: cellular networks, transportation networks and phone call networks. They studied the interaction between these systems and the challenges raised by their co-evolution and reaction to incidents. Gómez *et al.* (2014) proposed a method for clustering a network into agents called decision units. This method deals with the complexity by exploring relationships between agents' local decisions and their impact at the global level.

Network-based approaches represent the connected infrastructures interdependencies as a graph to show paths for incidents propagation. Shah and Babiceanu (2015) propose to evaluate the resilience of a system under attacks. The infrastructures are modeled using networks of interdependent processes. Based on this model, the authors provide simulations to predict the network behavior to face different attacks.

11.3 Impact Propagation and Decision Support Model

This part describes the impact propagation model and decision support model solution that includes the specification of the IPM ontology (SafecareOnto) and the IPM rules.

11.3.1 SAFECARE Ontology

The Safecare ontology, called SafecareOnto, describes both cyber and physical assets, their vulnerabilities, and their interdependence, as well as the risks and threats. It is the cornerstone of the knowledge graph used by the Impact Propagation and Decision Support Model module to infer the propagation of impacts over cyber and physical assets. In the following sections, we will describe the construction process of this ontology and its modular structure.

Overview of the ontology building process

For the determination of the approach to build SafecareOnto described in Figure 11.1, we have been inspired by NeOn methodological framework Suárez-Figueroa *et al.* (2012).

In the first phase, we provided information about the scope of the ontology (its purpose, the language to be used during its implementation, the target users for which it is intended, its requirements expressed under competency questions).

In the second phase, we started by studying the available resources (ontological and non-ontological) favoring the elaboration of SafecareOnto. The lack of ontological resources that perfectly meet our requirements led us to choose the option of building a first draft of the ontology from portions of non-ontological resources through an abstraction process. The objective is to identify a core of

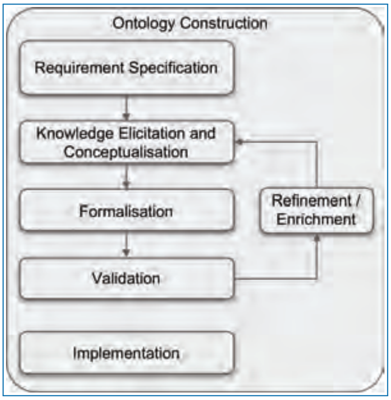


Figure 11.1. Construction process of SafecareOnto.

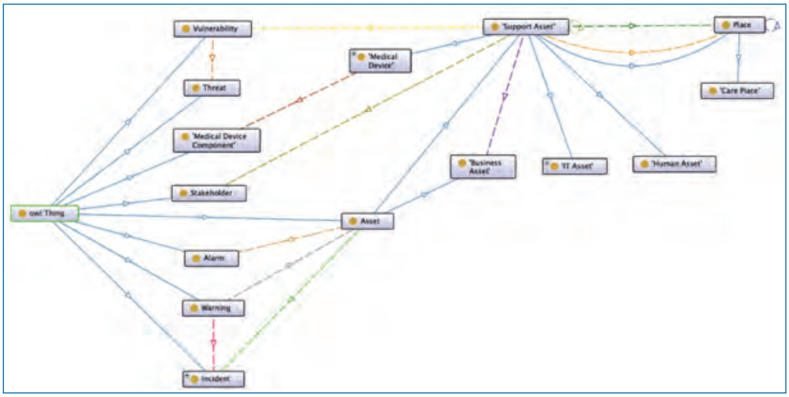


Figure 11.2. Excerpt of the SafecareOnto.

basic concepts and relationships that must be part of our ontology. As an example of non-ontological resources, we can mention the description of EBIOS RM methodology (EBIOS, 2019) and the description of medical devices of the MITRE (Connolly *et al.*, 2019). The conceptualization activity consisted of summarizing, organizing, and structuring the required knowledge into a meaningful model. In our case, for representing knowledge modeling, we opted for the UML class diagram. The benefits of such a model for ontology conceptualization have been acknowledged in several studies. One of its main advantages is that it is widely used. Furthermore, users are likely to be more familiar with a class diagram representation of the ontology (since it is a semi-formal model) than with OWL which representation is purely textual. Thus, it is more relevant for the verification of the ontology scope.

The resulting conceptual model (the first draft of SafecareOnto) has been translated, during the formalization phase, into a formal model using OWL2 that offers a highly expressive language and inference capabilities. Figure 11.2 represents an excerpt of the SafecareOnto.

The last phase consists of evaluating SafecareOnto regarding the ability of the impact propagation module to deal with the threat scenarios defined in the SAFECARE project. The validation step will lead to a refinement and enrichment of the ontology.

SafecareOnto, a modular ontology

The impact propagation and decision support model relies on both structural information about the assets and their intrinsic properties and structural relationship and on knowledge about the incidents that they suffered from. It also holds knowledge about how to infer and propagate impacts. This second knowledge evolves continuously and is more dynamic than the structural knowledge. For example,

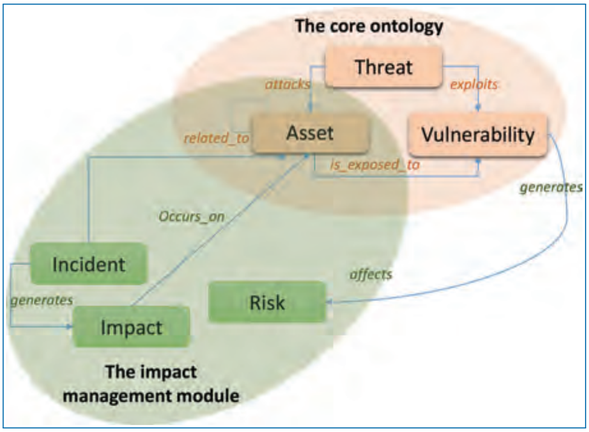


Figure 11.3. The modular structure of SafecareOnto.

the software of a medical asset could be updated to correct a known vulnerability. This kind of operation is less dynamic and more predictable that the occurrence of incidents.

To cope with the static and dynamic knowledge and to confer more stability to the IPM module, we have adopted a modular vision of the ontology. At a high level of abstraction, we could view the whole picture as depicted in Figure 11.3.

The core ontology captures essentially the static and is centered essentially on three concepts that are Asset, Vulnerability, and Threat.

An asset is any "thing" that has value. Within the SAFECARE projects assets could be business assets such as personal data about patients and personnel or the patients themselves or support assets such as medical or IT devices or medical staff. Assets are related to other assets through several kinds of relationships. A vulnerability is any weakness of an asset that could be used to generate a threat. A vulnerability assesses the protection of an asset against attacks. A threat could be accidental or malicious. As an example for "a radiology room" could have as vulnerability "likely to be subject to unauthorised access" and a "patient report" could have as vulnerability "lack of encryption."

A threat is the operationalization or a materialization of a vulnerability. An asset could be exposed to several vulnerabilities that are known or that could emerge after incident occurrence. The information about vulnerabilities is updated consequently to regular maintenance operations or after incident analysis. "Unauthorized access" or "personal data disclosure" are examples of threats. The more we know about the threats that relate to an asset, the more efficient its protection can be and the better we can react when incidents occur. These basic concepts are further refined and characterized. An excerpt is formalized in the next section. This formalization is done in such a way that it can easily be extended to meet emerging requirements.

The impact management module is an extension to the core ontology that relies on the previous concepts. It allows defining the concepts that are essential to the computation of impact propagation and provide indicators to help decide about the suitable countermeasures to face attacks consequences. It relies on concepts such as Incident, Risk, and Impact.

An incident, according to NIST (Stouffer *et al.*, 2011), is "an occurrence that actually or potentially jeopardizes the confidentiality, integrity, or availability of a system or the information the system processes, stores, or transmits or that constitutes a violation or imminent threat of violation of security policies, security procedures, or acceptable use policies." An incident could be an attack against one or several assets by exploiting vulnerabilities. In SAFECARE, we handle both physical and cyber incidents. We also have to assess the severity of an incident to better compute its propagation. An incident could be the expression of a known risk or completely unexpected. Indeed, a risk is the probability that a threat will exploit a vulnerability. When an incident occurs, it is likely to have impacts on assets. An impact needs to be qualified and/or quantified to efficiently help decide about the mitigation plans.

11.3.2 IPM Rules Specification

There are several approaches for impact propagation management such as agent-based and graph-based approaches that are mainly structure oriented. However, from our investigations, it appears that an added value that the project may produce is to combine cyber and physical incidents and to take into account the variety of interdependencies to provide a semantic oriented approach based on semantic web technologies. A first solution is consequently based on the exploitation of the ontologies' expressiveness expanded by the usage of inference rules. Indeed, the idea of the IPM module is to use axioms describing the concept and properties of SafecareOnto as well as a set of rules to deal with different threat scenarios. The creation of these rules follows the steps below (cf. Figure 11.4):

- *Knowledge elicitation*: in this phase, threat scenarios are analyzed and discussed with domain experts to identify, on the one hand, all the assets that could be impacted in each scenario, and on the other hand, the relationships between assets that lead to the propagation of impacts. Moreover, all the situations of a given scenario are analyzed to see if it is possible to generalize common parts. The objective is to avoid redundant rules.
- *Formalization*: in this phase, the concepts and properties of SafecareOnto that can be used to write rules are identified. A rule-engine (e.g., SWRL, JENA) is then used to implement these rules in the form of premises and conclusions.

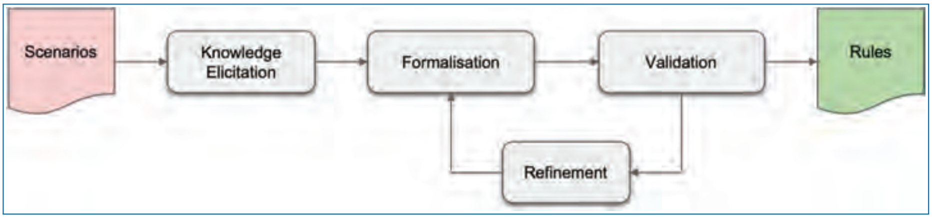


Figure 11.4. IPM rules construction process.

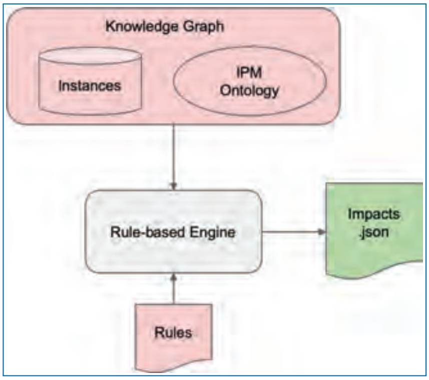


Figure 11.5. Architecture of the IPM prototype.

As existing rule-engine are often equipped with semantic reasoners, the implemented rules can be applied to automatically infer impact propagation.

- *Validation and refining*: in this phase, implemented rules are tested on different scenarios, and inferred impacts on different assets are evaluated by domain experts. At the end of the validation, IPM rules could be refined to better meet the expected results.

A first version of a prototype that simulates impacts propagation was implemented on a near-real scenario (cf. Figure 11.5). Based on the knowledge graph and on IPM rules, a reasoner is used to infer impacts propagation on assets. In this prototype, the IPM rules were expressed in terms of OWL concepts (classes, properties, individuals) using the JENA rule engine. Each rule is composed of a list of body terms (premises), a list of head terms (conclusions).

The following example presents a JENA rule that propagates warnings in case of assets located in the same places:

```
(?asset ipm:hasLocation ?place), (?warning ipm:attachedTo ?place),
(?warning ipm:hasCause ?incident), makeSkolem(?new_warning, ?warning) =>
 (?new_warning rdf:type isid:Warning), (?new_warning ipm:hasCause ?incident),
 (?new_warning ipm:attachedTo ?asset), (?asset ipm:hasWarning ?new_warning)]
```

The premise of this rule instantiates all the assets having a place, the warnings triggered in this place and the incidents causing these warnings. The conclusion attaches warnings to all assets located in the same place. An application of this rule may be a fire detection incident in a server room that could affect all the materials inside this room.

11.4 Conclusion

This chapter presents a focused view on how to handle incidents and their propagation from an assets point of view in a healthcare environment. It presents an overview of work conducted within the Safecare project. The state of the art shows that dealing with incidents and their propagation requires a detailed knowledge on assets, their context and an as precise as possible vision of the historical data about the assets, their real time state and the incidents that impacted them. From our experience within the Safecare project, it appears that collecting such data is not an easy task. It requires an additional effort from health actors, whose priority is care, although they are aware that safety is also a major issue. Consequently, we could not adopt one of the existing approaches for incidents propagation as they rely on either detailed traces, in case of empirical approaches, or a quasi-complete structure knowledge of systems as required by network-based approaches. The proposed solution is semantics based. It relies on an evolving knowledge captured by a modular ontology. The propagation is managed by rules that exploit assets states, incidents, and domain knowledge which all evolve continuously.

Acknowledgments

This work is part of SAFECARE project. This project has received funding as part of the “Secure societies – Protecting freedom and security of Europe and its citizens”, challenge of the Horizon 2020 Research and Innovation programme of the European Union, under grant agreement 787002.

References

Adetoye, A. O., M. Goldsmith, and S. Creese. 2011. “Analysis of dependencies in critical infrastructures.” In: *International Workshop on Critical Information Infrastructures Security*. pp. 18–29.

Agrafiotis, I., J. R. Nurse, M. Goldsmith, S. Creese, and D. Upton. 2018. “A taxonomy of cyber-harms: Defining the impacts of cyber-attacks and understanding how they propagate.” *Journal of Cybersecurity*. 4(1): tyy006.

Almohri, H., L. Cheng, D. Yao, and H. Alemzadeh. 2017. "On threat modeling and mitigation of medical cyber-physical systems." In: *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. IEEE. pp. 114–119.

Amutio, M., J. Candau, and J. Mañas. 2014. "Magerit-version 3, methodology for information systems risk analysis and management, book I-the method." *Ministerio de Administraciones Públicas*.

Barrett, C., R. Beckman, K. Channakeshava, F. Huang, V. A. Kumar, A. Marathe, M. V. Marathe, and G. Pei. 2010. "Cascading failures in multiple infrastructures: From transportation to communication network." In: *2010 5th International Conference on Critical Infrastructure (CRIS)*. IEEE. 1–8.

Breier, J. and F. Schindler. 2014. "Assets dependencies model in information security risk management." In: *Information and Communication Technology-EurAsia Conference*. Springer. 405–412.

Brocke, J. vom, A. M. Braccini, C. Sonnenberg, and P. Spagnoletti. 2014. "Living IT infrastructures—an ontology-based approach to aligning IT infrastructure capacity and business needs." *International Journal of Accounting Information Systems*. 15(3): 246–274.

Chou, C.-C. and S.-M. Tseng. 2010. "Collection and analysis of critical infrastructure interdependency relationships." *Journal of Computing in Civil Engineering*. 24(6): 539–547.

Clemente, D. 2013. *Cyber Security and Global Interdependence: What is Critical?* Chatham House, Royal Institute of International Affairs.

Connolly, J., S. Christey, R. Daldos, M. Zuk, and M. Chase. 2019. "Medical Device Cybersecurity Regional Incident Preparedness and Response Playbook. October 2018." MITRE.

CTED and UNOCT. 2018. "The protection of critical infrastructures against terrorist attacks: compendium of good practices." United Nations.

Dudenhoeffer, D. D., M. R. Permann, and M. Manic. 2006. "CIMS: A framework for infrastructure interdependency modeling and analysis." In: *Proceedings of the 38th conference on Winter simulation*. Winter Simulation Conference. 478–485.

EBIOS. 2019. "EBIOS Risk Manager—The method." The French National Cybersecurity Agency (ANSSI).

ENISA. 2016a. "Communication network dependencies for ICS/SCADA Systems." European Union Agency For Network and Information Security.

ENISA. 2016b. "Cyber security and resilience for Smart Hospitals." European Union Agency For Network and Information Security.

Gómez, C., M. Sánchez-Silva, and L. Dueñas-Osorio. 2014. "An applied complex systems framework for risk-based decision-making in infrastructure engineering." *Structural Safety*. 50: 66–77.

HITRUST. 2019. "The HITRUST Threat Catalogue." Health Information Trust Alliance.

Jakobson, G. 2011. "Mission cyber security situation assessment using impact dependency graphs." In: *14th International Conference on Information Fusion*. IEEE. 1–8.

Kotzanikolaou, P., M. Theoharidou, and D. Gritzalis. 2013. "Cascading effects of common-cause failures in critical infrastructures." In: *International Conference on Critical Infrastructure Protection*. Springer. 171–182.

Laefer, D. F., A. Koss, and A. Pradhan. 2006. "The need for baseline data characteristics for GIS-based disaster management systems." *Journal of Urban Planning and Development*. 132(3): 115–119.

Liu, C., C.-K. Tan, Y.-S. Fang, and T.-S. Lok. 2012. "The security risk assessment methodology." *Procedia Engineering*. 43: 600–609.

Mendonça, D. and W. A. Wallace. 2006. "Impacts of the 2001 world trade center attack on New York City critical infrastructures." *Journal of Infrastructure Systems*. 12(4): 260–270.

Petit, F., D. Verner, D. Brannegan, W. Buehring, D. Dickinson, K. Guziel, R. Haffenden, J. Phillips, and J. Peerenboom. 2015. "Analysis of critical infrastructure dependencies and interdependencies." *Tech. Rep.* Argonne National Lab. (ANL), Argonne, IL (United States).

Rinaldi, S. M., J. P. Peerenboom, and T. K. Kelly. 2001. "Identifying, understanding, and analyzing critical infrastructure interdependencies." *IEEE Control Systems Magazine*. 21(6): 11–25.

Schmitz, W., F. Flentge, H. Dellwing, and C. Schwaegerl. 2007. "The integrated risk reduction of information-based infrastructure systems, interdependency taxonomy and interdependency approaches." *IRRIS Project*. (027568): 82.

Shah, S. S. and R. F. Babiceanu. 2015. "Resilience modeling and analysis of interdependent infrastructure systems." In: *2015 Systems and Information Engineering Design Symposium*. IEEE. 154–158.

Silva, F. and P. Jacob. 2018. "Mission-Centric Risk Assessment to Improve Cyber Situational Awareness." In: *Proceedings of the 13th International Conference on Availability, Reliability and Security*. ACM. 56.

Stapelberg, R. F. 2008. "Infrastructure systems interdependencies and risk informed decision making (RIDM): impact scenario analysis of infrastructure risks induced by natural, technological and intentional hazards." *Journal of Systemics, Cybernetics and Informatics*. 6(5): 21–27.

Stouffer, K., J. Falco, and K. Scarfone. 2011. "Guide to industrial control systems (ICS) security." *NIST Special Publication*. 800(82): 16–16.

Suárez-Figueroa, M. C., A. Gómez-Pérez, and M. Fernández-López. 2012. "The NeOn methodology for ontology engineering." In: *Ontology Engineering in a Networked World*. Springer. 9–34.

Theocharidou, M. and G. Giannopoulos. 2015. "Risk assessment methodologies for critical infrastructure protection. Part II: A new approach." *Tech. Report EUR 27332 EN*.

Tong, X. and X. Ban. 2014. "A hierarchical information system risk evaluation method based on asset dependence chain." *International Journal of Security and Its Applications*. 8(6): 81–88.

Zimmerman, R. 2008. "Understanding the implications of critical infrastructure interdependencies for water." *Wiley Handbook of Science and Technology for Homeland Security*. 1–25.

DOI: 10.1561/9781680836875.ch12

Chapter 12

Innovative Toolkit to Assess and Mitigate Cyber Threats in the Healthcare Sector

By Marco Manso, Bárbara Guerra, George Doukas and Vasiliki Moumtzi

Cybersecurity is an increasingly critical aspect of healthcare information technology infrastructure. Nowadays, the rapid digitization of healthcare delivery, from electronic health records and telehealth (eHealth services) to mobile health (mHealth) and network-enabled medical devices, introduces risks related to cybersecurity vulnerabilities that are particularly worrisome because cyberattacks in a healthcare setting may result in the exposure of highly sensitive personal information, cause disruptions in clinical care, or affect the safety of patients, for example, by compromising the integrity of data or impairing medical device functionality. The threat is real and growing in tandem with the pace of the healthcare industry digitization [1]. Yet, cybersecurity capacities currently remain behind the pressing needs, lagging the robust pace of adoption of digital networks by threat actors. This discontinuity places the multitrillion-euro healthcare sector at risk of even more significant cyberattacks. A new generation of cybersecurity tools, specifically designed for the healthcare domain, takes on the challenge of surpassing that discontinuity and setting higher standards on cybersecurity for healthcare organizations. The proposed architecture combines a smart and robust security awareness layer, equipped with a wide range of tools that build a personalized data security management platform. The combined use of state-of-the-art technologies to effectively prevent, respond, and recover from cyberattacks, while managing to raise awareness and provide timely actionable information, is a promising compound for enhancing cybersecurity within the healthcare IT ecosystem.

12.1 Introduction

Over the past ten years, adversaries around the world have been constantly using more sophisticated methods to attack organizations' digital surfaces. There used to be a clear distinction between the inside and outside of an organization, and infrastructures had clearly defined boundaries. However, with the rise of mobile computing and cloud services, that endpoint has expanded and there is no clear, easily protected line that can keep data secured. Now that the Internet of Things (IoT) has also been adopted and has entered the mainstream, the perimeter and number of vulnerabilities are set to expand yet again.

There are few industries that need strong cybersecurity as much as the healthcare industry. Healthcare technology is prevalent around the world and creates huge potential to improve clinical outcomes and transform care delivery, but there are also increasing concerns relating to the security of healthcare data and devices. With the healthcare sector's rapid adoption of digital systems and the spending in technology growing, so does the sector's cybersecurity attack surface. Today, healthcare networks not only include hospitals, clinics, and doctor's offices but also start to accommodate Internet-based medical consulting with remote healthcare providers or patients, multi-cloud Infrastructure-as-a-Service (IaaS) and Software-as-a-Service (SaaS) environments and connected medical devices both inside hospitals and deployed at the patients' homes.

In the past five years, healthcare has been plagued by a myriad of cyber threats, with weaponized ransomware, misconfigured cloud storage buckets, and phishing emails dominating [2]. And, indeed, the healthcare sector experiences twice the number of cyberattacks as other industries [3]. A global study by NetDiligence of cyber insurance claims in 2017 found that healthcare accounted for 18% of breaches across all sectors and that 63% of healthcare breaches were caused by criminal or malicious activity [4]. Healthcare data breaches grew in both size and frequency, with the largest breaches impacting as many as 80 million individuals, exposing highly sensitive information, from personally identifiable information to health insurance information and patients' medical histories [5]. Cybercriminals seem to display more creativity and adopt faster new technologies, as healthcare organizations strive to develop better awareness of cyber threats and adequate security strategies.

While almost every healthcare provider is committed to secure patient privacy, most of them are behind in terms of cybersecurity adoption and advancement. Healthcare providers are highly attractive targets, because they can provide access to a huge amount of valuable data. In order to secure the available information, when adversaries are relentlessly in pursuit of weaknesses, the only way is trying to be always one step ahead. In order to achieve this, sufficient levels of cybersecurity awareness must be ensured. However, healthcare providers are usually focused on upgrading their medical technology and investing on employing the best medical staff to ensure they save lives faster and provide better overall care. Therefore, this approach overlooks the importance of cybersecurity and its complementarity to core provided services.

Apart from the awareness, the existence of enforced consistent security standards and procedures seems to be a crucial factor that impacts the levels of cybersecurity. Standards such as the ISO/IEC 27000 family or the NIST Cybersecurity Framework encourage the implementation and adoption of clear and comprehensible structures that define procedures, techniques, and measures, highly important for both cyber and information security, especially when it comes to areas such as healthcare. Towards this end, several worldwide initiatives and agencies try to specify rules and necessary measures under a common cybersecurity strategy and, in parallel, introduce a common cybersecurity certification scheme. Most healthcare providers acknowledge that such guidelines can strengthen their cybersecurity, but they try to tackle the problem with simpler procedures.

Despite the current selection of defensive measures, healthcare providers should always take into serious consideration that the attacks tend to be more opportunistic and difficult to detect or predict. Threats are getting more potent because systems are more interconnected; and people, business, and government will have a greater reliance on Information and Communications Technologies (ICT) to function. Defenders should prepare to face self-propagating network-based threats that can even be disguised in encrypted traffic, or hidden threat actors in popular cloud services which cannot be dealt with traditional security tools, or even orchestrated attacks using several tools to induce confusion and make attacks undetected for long periods of time. Developing at this pace, cybercrime threatens to become even more devastating for all types of businesses in years to follow, and the only certainty is that in order to effectively address all emerging threats, every possible combination of available innovative technologies will be required.

12.2 The Challenge

Hospitals and care centers have adopted ICT to handle, store, and exchange large amounts of sensitive patient data from medical devices, but also wearables and IoT devices, like smart watches and sensors, which are not categorized as medical devices but still produce sensitive data of patient's health and wellness status.

Digital health service providers are improving their efforts to reduce inefficiencies, improve access, reduce costs, increase quality, and make medicine more personalized for patients [6]. In the meantime, patients use digital health to better manage and monitor their health and wellness-related activities [7]. This increasing market, which is becoming more open and accessible, raises the volume of potential vulnerability risks that may lead to cyber threats and incidents.

The operating system and medical devices' firmware of hospitals and care centers typically consists of legacy hardware or software without a systematic patch management procedure which leads to vulnerabilities that cyberattackers may exploit, in addition to the many vulnerabilities of the increasing healthcare ecosystem brought by IoT devices. As a result, hospitals and care centers are prime targets for cyber criminals, especially concerning data theft: a FortiGuard Labs study states that almost half of the top 10 threats in healthcare were triggered by botnets, some of which leveraged through compromised IoT medical devices.

Moreover, several security aspects such as the vulnerability assessment of existing and newly added devices and the real-time alerting and mitigation of security threats are not addressed, rendering those solutions unable to cope with state-of-the-art cybersecurity threats. The security posture of IoT providers should constantly adapt to the ever-evolving cybersecurity risks imposed by multiple sources. Turning to commercial solutions, the traditional penetration testing is still very much a point-in-time assessment, dealing with specific tasks, and thus, most solutions cannot manage new sophisticated advanced attacks. This may cause major disruptions of healthcare services and the loss of patient data for hospitals and care centers, since it is almost impossible for them to maintain and administer a widespread heterogeneous infrastructure across the region or even the nation, thus becoming prone to cyberattacks.

The healthcare ecosystem is therefore missing a tailored cybersecurity vulnerability assessment toolkit, able to proactively assess and mitigate cyber threats, either known or unknown, imposed by mobile and IoT-based devices and eHealth and mHealth services.

12.3 The Solution

This challenge has been identified in the EU-funded Research and Innovation Action SPHINX [8] that aims to develop a health-tailored universal Cyber Security Toolkit to enhance the cyber protection of the healthcare information technology (IT) ecosystem and ensure patient data privacy and integrity.

The vision is to build a transparent cybersecurity environment to be effortlessly tailored to the specifications (addressing infrastructure, devices, and software) of different hospitals and care centers. This environment offers the professionals in healthcare organizations the opportunity to identify cybersecurity risks when they occur and take decisions to safeguard their cybersecurity and privacy. As a result, the innovative Cyber Security Toolkit transforms today's cybersecurity practices in healthcare organizations, by offering reliable automated cybersecurity services and enhanced transparency. In this respect, the Toolkit adopts a strong human-centered design that:

- Alleviates healthcare IT personnel from the burden of unnecessary decisions by offering them extensive automated security.
- Prioritizes the human factor to the advantage of security, by exploiting security-friendly behavioral patterns of users (healthcare organization's professionals) and beneficiaries (patients).
- Adheres to applicable legislation in order to increase trust in case of a security or data protection breach, facilitating the overall recovery process.

The SPHINX Cyber Security Toolkit provides the IT personnel at hospitals and care centers with highly comprehensive visual analytics that depict the cybersecurity situation on near real-time basis, in an intuitive and user-friendly way. In addition, the Toolkit is fully adaptable to the individual user's profile characteristics.

Concerning IoT devices, manufacturers are also able to use the SPHINX Toolkit security services and certify that their devices respect the security and privacy needs of their beneficiaries and users. Concerning patients, these benefit greatly from the increased security of their IoT devices brought by the SPHINX Toolkit that, in an automated manner, monitors and controls potential cyber risks, known or unknown cyber threats and incidents. As a result, patients are easily prompted to customize personalized data privacy policies, distinguishing between work and home-related data among different IoT devices in use.

The SPHINX Cyber Security Toolkit concentrates data of both real and cyber situation awareness within the healthcare IT ecosystem to analyze potential, imminent, and forecasted cyber vulnerabilities. An automated zero-touch device and service verification solution is adapted to or embedded in existing infrastructures, facilitating the IT personnel's work through the selection of a number of available security services capable of enhancing the IT ecosystem's self-defense countermeasures and increase the effectiveness of cybersecurity services' automation. These innovative technologies amplify the cybersecurity for healthcare ecosystems, building a certified trust between IT personnel and the remaining healthcare professionals.

The Toolkit is built upon an interoperable architecture that concentrates and handles data from multiple devices and services, covering thus a wide range of use cases to be demonstrated in SPHINX's at advanced testbeds.

12.4 The Technical Architecture

SPHINX introduces a universal Cyber Security Toolkit for the Healthcare domain that enhances the cyber protection of the healthcare IT ecosystem and ensures the patients' data privacy and integrity. The SPHINX Toolkit offers an embedded, smart, and robust security awareness layer, able to identify modern and advanced cyber threats, enhanced with a personalized data security management tool.

The SPHINX architecture [9] has the capability to concentrate and handle the data of many devices or services, thus covering a wide range of use-case scenarios. The SPHINX users are kept informed at any time via highly comprehensive dashboards and visual analytics, while being able to interact with the services and functions of the proposed solution in an intuitive and user-friendly way. Furthermore, the SPHINX Toolkit provides automated intrusion prevention and data filtering algorithms, fully adaptable to the individual user's profile characteristics.

The SPHINX Toolkit is designed to facilitate the operation of the SPHINX cybersecurity tools in real-life conditions, allowing regular technology users (not limited to cybersecurity experts) to operate the system. Advanced data flow analysis is applied on a packet and session basis to build the context of communication. From this context, data are classified into user and device profiles, in line with appropriate categorization methodologies. User and device profiles are used for automated real-time risk assessment, based on evaluation, comparison, and matching with safe data flow patterns, utilizing a self-learning approach performed at application layer. Data analytics and visualization techniques are deployed to ensure enhanced user awareness and understanding of the security status, potential threats, risks, and associated impacts.

The SPHINX main building blocks are depicted in Figure 12.1. The figure illustrates the SPHINX Toolkit in a healthcare IT operational environment (involving users, workstations, servers, medical devices) in which the SPHINX tools are deployed as part of the SPHINX operational environment. The device verification and certification tool is isolated in SPHINX's sandboxed environment.

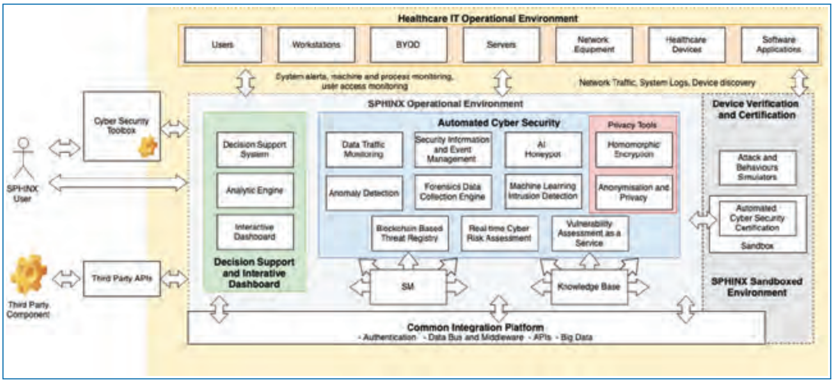

Healthcare IT Operational Environment
Users
Workstations
BYOD
Servers
Network Equipment
Healthcare Devices
Software Applications
System alerts, machine and process monitoring, user access monitoring
Network Traffic, System Logs, Device discovery
SPHINX Operational Environment
Cyber Security Toolbox
SPHINX User
Third Party APIs
Third Party Component
Decision Support System
Analytic Engine
Interactive Dashboard
Decision Support and Interative Dashboard
Automated Cyber Security
Data Traffic Monitoring
Security Information and Event Management
AI Honeypot
Anomaly Detection
Forensics Data Collection Engine
Machine Learning Intrusion Detection
Blockchain Based Threat Registry
Real time Cyber Risk Assessment
Vulnerability Assessment as a Service
Privacy Tools
Homomorphic Encryption
Anonymisation and Privacy
SM
Knowledge Base
Device Verification and Certification
Attack and Behaviours Simulators
Automated Cyber Security Certification
Sandbox
SPHINX Sandboxed Environment
Common Integration Platform
- Authentication - Data Bus and Middleware - APIs - Big Data

Figure 12.1. SPHINX platform high-level architecture.

At the core of the SPHINX solution lies the **Common Integration Platform**, providing the necessary integration framework and infrastructure for all SPHINX tools and services. It is built upon the basic concepts of virtualization and containerization, allowing each tool and service to be deployed independently and to concurrently support the aggregation of heterogeneous external services, making use of various data exchange protocols.

The **Device Verification and Certification** building block provides functionalities for the verification of the cybersecurity level of software applications and devices, including assessment of vulnerabilities. It is provided through a sandboxed environment and provides a safe and isolated testing environment where deployment and cybersecurity testing can be performed. It also delivers a certification report concerning the compliance of third-party software applications and devices with SPHINX standards.

The **Automated Cyber Security** building block is the smart and robust security awareness layer of SPHINX, encompassing a wide range of cutting-edge cyber protection technologies and innovative tools focusing on the detection of anomalous behavior, security event handling, intrusion detection, vulnerability, and risk assessment. The SPHINX Cyber Situational Awareness Framework combines every available information, generated by the SPHINX tools, to provide effective and efficient identification, investigation, mitigation, and reporting of realistic multi-dimensional attacks and advanced persistent threats within the healthcare IT ecosystem and its interconnected cyber assets.

The SPHINX Toolkit: contributing tools and their key functionalities

The SPHINX Toolkit displays advanced cybersecurity capabilities that embrace the identify, protect, detect, respond, and recover functions of the NIST Framework for Improving Critical Infrastructure Cybersecurity [10].

On the monitoring of the IT ecosystem and the building of a cyber situational awareness, several tools allow a beyond state-of-the-art cyber protection to the critical assets in the healthcare IT ecosystem. The *Data Traffic Monitoring and Anomaly Detection* tools track information regarding the connected devices in the network and the data they access, the transfer loads, and the transmitted packets' in order to detect suspicious programs' network traffic, unexpected events, and any other activity or observation that raise suspicion by differing significantly from the normal infrastructure/component/user behavior. The *Security Information and Event Management* tool implements a query interface where other components or users can distinguish between normal and abnormal operations. The tool's log management capabilities facilitate the collection, aggregation, retention, analysis, searching, and reporting of high volumes of computer-generated log messages that allow the end user to provide real-time analysis of security alerts generated by network hardware and applications. The *Vulnerability Assessment as a Service* tool dynamically assesses network entities against certain vulnerabilities and outputs a Common Vulnerability Scoring System score that reflects the level of security of that particular entity. The *Real-time Cyber Risk Assessment* tool deals with advanced and automated features to assess the level of risk associated with cybersecurity incidents, determine their probable consequences, and present warning levels and alerts to the users of the healthcare IT operational environment. The *Artificial Intelligence Honeypots* are part of the cyber defense arsenal and are used to prevent, detect, and respond to cyberattacks. Their value resides on luring the adversaries to attack them instead of the real production IT systems, by emulating services or even complete systems that may be considered targets from an adversary. In SPHINX, the Honeypot tool provides data dynamically to the Artificial Intelligence algorithms designed to detect anomalies. The *Machine Learning-empowered Intrusion Detection* tool operates in conjunction with Honeypots to gather attack information from intruders and supervised machine learning and/or deep learning algorithms for dynamic learning of both registered and unregistered data. Outperforming current solutions that are typically capable of coping with known threats, SPHINX makes a step forward developing an intelligent defensive system capable of either detecting existing threats or learning new uncategorized ones. The *Forensic Data Collection Engine* tool provides the basis required for supporting the processing and storage of data gathered from various sources into a unified structure in order to discover the relationships between devices and the related evidence and produce a timeline of cybersecurity incidents, including a map of affected devices and a meaningful chain of evidence.

On the exploitation of acquired knowledge to establish an enhanced cybersecurity awareness, two SPHINX tools come into play. The *Knowledge Base* aims to represent domain-specific knowledge in a form that can be used by both computers and humans to effectively operate on the knowledge acquired by SPHINX. Towards forming knowledge, it collects anonymized security intelligence and insights from external repository sources (autonomous agents search and mine reliable web resources), as well as from SPHINX tools. This information is translated into security rules and shared among the network by updating the respective advanced threats registries. The Knowledge Base gathers security incentives for a collective wisdom creation, as well as interconnects/integrates with third parties threat intelligence. On its turn, the *Blockchain Based Threats Registry* acts as a background infrastructure that safely stores different logs from different sources within the healthcare domain. It can also be used to store any kind of interesting information, such as critical logs or thread information. The main advantage of using Blockchain is to have a distributed ledger with unalterable information, synchronized between all parties.

A key aspect of the SPHINX Toolkit deals with privacy assurance and testing features. In this context, the *Homomorphic Encryption* tool serves as a backbone for the SPHINX Toolkit and ensures user data privacy and security by storing all sensitive data in an encrypted format. Also the *Anonymization and Privacy* tool assists in this endeavor, delivering a dataflow with high throughput for processing large text datasets in unstructured formats and performing user-defined transformations to clean, structure, anonymize, and/or encrypt. Further, the *Attack and Behavior Simulators* deliver a reliable ground for testing SPHINX tools: by providing routines/scripts of already documented cyberattacks, with known effects, outcomes, and consequences, the simulators allow for the operational capability of the SPHINX Toolkit to be tested.

Overall, the **Cyber Security Toolbox** enables SPHINX users to select the cybersecurity services that best match their needs, to use within the SPHINX ecosystem. It allows users to *plug* cybersecurity services into their existing connectivity services and configure/adapt them according to their security needs. In this context, the SPHINX *Decision Support and Interactive Dashboards* target a panoply of user-centered functionalities related with decision support, providing recommendations on the suitable courses of action following upcoming, ongoing, or forecasted cyberattacks or incidents. The decision-making process is supported by an analytic engine for the visualization of data in near real time that delivers a first insight into users' behaviors, as well as by customizable dashboards that interactively display and share trends, forecasts, and answers to business questions on the cybersecurity and protection levels of the healthcare IT ecosystem.

Finally, the SPHINX Toolkit also considers the interaction with third parties: the *Application Programming Interface for Third Parties* is a tool specifically designed to enable third-party solution providers to access and interact with the SPHINX Toolkit and its tools. Subject to authentication and using end-to-end encryption, it exposes advanced cybersecurity functionalities implemented by SPHINX, from device/application certification and verification to threat registry notifications and the detection of anomalies.

12.5 Application Scenarios for SPHINX Cybersecurity Tools

The innovative SPHINX cybersecurity tools are designed and developed to address a set of application scenarios specific to the healthcare domain, based on a report by the European Union Agency for Network and Information Security (ENISA) concerning security challenges and risks in eHealth [11]. In SPHINX, five application scenarios were defined focusing on the adoption of novel information and communication technologies by healthcare stakeholders, giving way to national eHealth strategies and a common EU eHealth policy, including healthcare data capture (secure collection of patient data from multiple sources), analysis (data processing and analytics to extract actionable information from captured healthcare data), and sharing (deployment of healthcare information networks that securely retrieve patient data from multiple sources and make it available to the patient and the responsible healthcare professional), in order to improve significantly the delivery of high-quality cost-efficient healthcare via informed decision-making.

12.5.1 Scenario 1: Digital Transformation in Healthcare

Healthcare is still new to digitization, with the vast majority of related investments on software and services in frontline clinical and administrative healthcare occurring in the last decade. Throughout the years, rendering administrative processes, clinical pathways, and patient data into digital realities has driven a focus on data standardization, integration, and security that holds together disparate system workflows. Adding new computers, servers, and devices and creating more dedicated networks has led to a panoply of different operating systems, applications, and databases that resulted in unique IT architectures and specialist cybersecurity needs. In the mix, outdated and legacy firmware with unaddressed bugs and known vulnerabilities compound the difficulty to maintain up-to-date security policies and systems, increasing the number of vulnerabilities or risks. Moreover, with increased digitization, new privacy regulations and more integration between different systems bring new risks and an increased burden of regulatory compliance.

Indeed, the current pace, scale, and complexity of technology adoption is putting healthcare organizations at a significant risk of multiplying its cyber vulnerabilities. When it comes to data as sensitive as private health information, the potential for an attack is surmountable for healthcare data has become one of the most desirable premium commodities for sale on black market sites. Not only do multiple sites require access to patient information across a spectrum of health facilities—such as local clinics, physician offices, hospitals, laboratories, and pharmacies—but the information also needs to be readily available to support open new healthcare services, such as allocation of medical practices, second opinion consultation services, comparison of diagnostic protocols, or participatory healthcare. Add to this the organizations' willingness to allow their employees to bring their own devices, and it is understandable how extremely challenging it is to implement network-wide security practices and data protection.

This application scenario illustrates several challenges in healthcare delivery that the SPHINX tools are designed to address:

- The digitized healthcare databases and services;
- The outdated (legacy) operating systems, applications, and databases;
- The integration of healthcare and patient data from multiple databases;
- The availability, integrity, and confidentiality of healthcare and patient data;
- The users' authentication and profile management;
- The integration of Bring Your Own Devices (BYOD), including professionals' and patients' tablets and smartphones, in healthcare organizations' networks.

The Digital Transformation in Healthcare is a common application scenario across Europe. For healthcare ecosystems to remain safe from cyber exploitation, cybersecurity strategies need to move beyond servers and desktops to reflect a world of interconnected networks, equipment, devices, and users.

12.5.2 Scenario 2: eHealth Services

EU Member States (MS) are working on an eHealth Digital Service Infrastructure under the aegis of the eHealth Network, the network of national authorities responsible for eHealth (2011/890/EU) [12]. In addition to Finland, Greece, Italy, Portugal, Spain, France, Denmark, Estonia, and Czech Republic, 18 MS are expected to exchange Electronic patient summaries and ePrescription by the end of 2021.

Healthcare organizations are gradually adopting new technologies to deliver nation-wide healthcare services online (eHealth), such as ePrescription/eDispensation, Electronic patient summary, eReferrals, and eBilling, that significantly facilitate the interaction of citizens and patients with healthcare organizations, as well as the daily work of thousands of healthcare professionals and employees. The added-value eHealth services are adopting widely used Internet-based technologies (e.g., IP and web services) and open standards (e.g., HL7) allowing access from commodity devices (e.g., mobile phones and web browsers) for users and services (intra- and extra-organization). Herein, organizations expose resources to external entities where security controls cannot be enforced.

As healthcare systems increasingly rely on web-enabled eHealth services and online transactions for care delivery, they also become more vulnerable to cyberattacks, requiring appropriate cybersecurity policies and solutions. An array of vulnerabilities is exposed and brings heightened concerns regarding privacy and security about third-parties' risks, inappropriate releases of sensitive and private information from healthcare records, and the systemic flows of information throughout healthcare organizations.

This application scenario addresses a set of challenges in healthcare delivery that the SPHINX tools are designed to address:

- Untrusted environments and devices;
- Web-based online healthcare services;
- Exposition of web services to external entities;
- The availability, integrity, and confidentiality of healthcare and patient data;
- The users' authentication and profile management.

With healthcare data breaches on the rise, healthcare organizations are committed to understand the perceived risks of eHealth services and the security and privacy measures patients expect, so they can begin to diagnose and overcome the barriers to adopting and embracing eHealth services.

12.5.3 Scenario 3: mHealth and Remote Patient Monitoring Platforms

Mobile health (mHealth) supports the delivery of healthcare via remote access medical devices, IoT-based health devices (the Internet of Medical Things or IoMT), and mobile applications that connect to healthcare IT systems through computer networks, empowering the sharing of health and well-being information, enabling the shifting of healthcare to a more preventative care outside of the hospital environment, giving rise to services such as telehealth (video appointments and consultation) and remote patient monitoring platforms, and delivering high-quality healthcare.

Experts estimate that there will be more than 64 billion IoT devices by 2025 [13], and a significant portion of these will be medical devices, from heart monitoring implants and pacemakers to infusion pumps, mobile medical workstations, in-home monitors and personal fitness devices or wearables. According to a study conducted by the McKinsey Global Institute, spending on the Healthcare IoT solutions will reach $1 trillion by 2025 [14]. Currently, 3 million patients worldwide are connected to a remote monitoring device that performs routine tests—such as checking glucose levels for patients with diabetes or checking blood pressure for patients receiving cardiac care—and sends personal medical data to their healthcare provider [15].

The use of personal health monitoring devices and smartphone applications (Apps) is also on the rise. Most of these devices are connected to patient remote monitoring Apps that focus on the collection of patient-generated health data from home, through devices and mobile health platforms that connect via the patient's home network or cellular network, to the primary care provider or care team. With mHealth tools and platforms, telehealth and remote patient monitoring platforms, healthcare organizations not only seize the potential to extend care management and coordination into the patient's home but also take the opportunity to deliver highly personalized, accessible, and on-time healthcare services; reduce the number of visits and hospitalizations; eliminate unnecessary waste; contain healthcare costs; and save lives. At the same time, the boundaries of cybersecurity are stretched, creating new, often insecure, entry points for hackers and rising data security and liability risks.

As healthcare systems become interconnected, especially as numerous wireless medical devices start connecting to web-enabled IT systems, they become increasingly vulnerable. Not only medical and health remote monitoring devices may be vulnerable to viruses and malware that can compromise the effectiveness of the devices (device failure or malfunction), patients' privacy, and the healthcare organization's IT ecosystem, but also the transmission of patient data enabled by those devices represents a risk of data breach if the information is not properly secure. In addition, the increasing use of BYOD is a potential issue as they may have developed networks and connectivity glitches and may very easily provide an on-ramp for attackers to healthcare networks. Moreover, they are prone to be lost or stolen, which could lead to identity theft and loss of privacy. Since these devices are outside the healthcare organization's control, there is also a lack of visibility and control over personal devices, as well as the absence of awareness of these devices' vulnerabilities that attackers could take advantage of. From a cybersecurity perspective, healthcare organizations need to rethink medical and health device management and consider all the variables this mobile technology introduces, compared to traditional workstations and laptops.

This application scenario especially illustrates the challenges in healthcare delivery that the SPHINX tools are designed to address:

- Untrusted environments and devices;
- Remote healthcare services (in-home care), such as telehealth consultations and remote patient monitoring platforms;

- Integration of IoT-enabled medical and health devices in the healthcare organizations' networks;
- Integration of patients' BYOD devices in the healthcare organizations' networks;
- The availability, integrity, and confidentiality of healthcare and patient data;
- The users' authentication and profile management.

With the development of the smart home concept, the IoMT and the advent of better mHealth technology, telehealth and remote patient monitoring platforms stand to become an accepted standard of high-quality healthcare delivery for the 21st century.

12.5.4 Scenario 4: Sharing and Exchange of Healthcare Information

Before the wide-scale adoption of Electronic Health Records/Personal Health Records (EHRs/PHRs), access to healthcare information entailed paper records, in-person requests to health information management offices, and the payment of fees. The increasing digitization of health records has improved access to health information, with healthcare professionals being able to easily access and view diagnosis, medication history, clinical decision support notes, lab results, imaging, treatment plans, and post-treatment monitoring. In this context, EHRs/PHRs act as pillars of point of care information systems, facilitating the sharing and exchange of health information among healthcare stakeholders, such as healthcare providers, pharmacies, insurance companies, and researchers. Currently, the ability of European citizens to access their electronic medical records across the EU varies from one country to another: such services are either operational (for example, in Luxembourg, Denmark, Finland, Estonia, France, Romania, and Portugal) or under development (for example, in Greece, Cyprus, and Italy). Thus, the European Commission is working to facilitate access across borders to healthcare data, namely to laboratory tests, medical discharge reports and images, and imaging reports, in full compliance with the General Data Protection Regulation (GDPR). Healthcare data interoperability and security are top priorities to ensure patient data protection and prevent data breaches.

Highly important for the EHRs/PHRs operations are interoperability standards and well-established integration profiles (adopted as EU standard specifications under the 1025/2012 EU regulation [16]), allowing the services to be provided to the appropriate users, across a variety of IT systems, diverse levels of sophistication and interoperable capabilities, a legal landscape of varying degrees, and various levels of privacy and rules, ensuring data availability, integrity, non-repudiation, resilience, and privacy. Healthcare organizations need to be knowledgeable of the

EU and national regulations and requirements with regard to healthcare interoperability, ensuring that they remain compliant to further healthcare data sharing and exchange so that the clinical or operational purpose and meaning of the data is preserved and unaltered. Data security is also a top interoperability priority. Ensuring privacy and the security of the health information throughout the entire data exchange process is a key component to building the trust required to realize the benefits of health information sharing and exchange. As such, access to data needs to be well defined and controlled (e.g., who can access, for how long), and performed operations (e.g., read, modify, delete) and support detailed auditing. It is also paramount to ensure data integrity throughout the complete workflow and data lifetime, clearly generating alerts if otherwise.

Along with improving health data security, it is important to consider patient preferences in how their data is handled, allowing them to understand how their information is used and how they could assert more control over which information is shared. Also, healthcare professionals should be aware of the security measures needed to protect their patient data.

This application scenario encapsulates specific challenges in healthcare delivery that the SPHINX tools are designed to address:

- Standardization and common data exchange formats, complying with EU and national regulations on interoperability;
- Availability, integrity, and confidentiality of patient records and healthcare information across the complete workflow and data lifetime;
- Detailed auditing on every data operation;
- The users' authentication and profile management.

The national push for healthcare interoperability continues to gain strength, as a common set of rules for trusted and secure exchange is established between networks across multiple jurisdictions, taking into account applicable legislation, including intellectual property rights, and supporting healthcare organizations in the process.

12.5.5 Scenario 5: Cross-border Healthcare Service Delivery

Cross-border healthcare has been introduced in the EU as required to secure universal quality of service delivered across the Member States, by allowing the flow of healthcare data across borders. Enabling citizens to securely access and share their healthcare data across borders is one of the priorities of the Communication on enabling the digital transformation of health and care in the Digital Single Market. Moreover, the GDPR underlines that citizens have the right to access their personal data and provides the legal framework for its protection, setting out directly applicable rules for the processing of the individuals' personal data, including their health data. And rules for facilitating the access to safe and high-quality cross-border healthcare are specifically provided for by the Directive on patients' rights in cross-border healthcare. Technical specifications for healthcare information exchange were defined, focusing on two sets of health data: Electronic patient summaries and ePrescription. The first exchanges took place between Estonia and Finland in January 2019 [17] and their example will be followed by another 22 EU MS by 2021. With the development and implementation of several EU-funded projects involving standardization and the exchange of healthcare data in Europe (projects epSOS, EXPAND, Antilope, and HITCH), the Refined eHealth European Interoperability Framework (ReEIF) is instrumental to the facilitation of EU-wide healthcare service delivery.

Currently, healthcare information on specific cases is exchanged among EU MS and Norway through the 24 thematic European Reference Networks (ERNs) that virtually connect 900 highly specialized healthcare units located in 300 hospitals and gather panels of clinicians to diagnose and treat suffering from rare, complex, and low prevalence diseases. Healthcare organizations refer patients to the relevant Network, with their consent and upholding existing national regulations, so citizens do not have a direct access to these networks. On the contrary, the digital transformation of healthcare, the creation of eHealth services, the leverage of mHealth and Remote Patient Monitoring platforms and the exchange and sharing of healthcare information, based on the cross-border interoperability of EHRs, PHRs, and ePrescription, is focused on the citizen. It will ensure that EU citizens can securely access and exchange their healthcare data wherever they are in the EU.

On February 6, 2019, the European Commission's Recommendation on a European Electronic Health Record exchange format (C(2019)800) [18] sets the framework to further develop a European EHR exchange format that will enable citizens to securely access and exchange their health data across borders in the EU. Further, it underlines the importance of ensuring data protection and security, in line with the GDPR, and full compliance with the cybersecurity framework. A joint coordination process involving the EU MS and the European Commission (EC) is envisaged to conduct this process, engaging relevant stakeholders, including healthcare professional organizations, national competence centers, industry actors, and patients' groups, as well as other EU and national authorities.

This application scenario encompasses broad challenges in healthcare delivery that the SPHINX tools are designed to address:

- A common vision for EU healthcare service delivery;
- The trusted chain of transactions that ensures data confidentiality;
- Authentication of all involved individuals and IT components (residing in different states);

- Availability, integrity, and confidentiality of healthcare and patient data;
- Standardization, interoperability, and common data exchange formats;
- Different national legislation frameworks on healthcare data.

Built on adequate technical expertise and open standards, the European electronic health record exchange format is set to become the future *de facto* standard for secure cross-border healthcare service delivery across Europe, taking into account full compliance with data protection legislation and ethical principles and abiding to a rigorous cybersecurity framework.

The five application scenarios identified in SPHINX enable the construction of the environment or context for the common identification of challenges, problems, needs, gaps, and opportunities and for the broaden debate of the SPHINX tools' added-value for the cybersecurity of healthcare organizations.

12.6 Conclusions

Building upon the exploitation of system and network vulnerabilities, the types of cyberattacks are rapidly increasing and constantly evolving. From individuals' personal information to confidential healthcare data, the field is vast, and the consequences can be devastating: impersonation, sensitive data fraudulent use, blackmail, ransom demand.

Healthcare providers should consider adopting new technologies to protect patient information and prevent their systems from being compromised. Artificial Intelligence to monitor shared networks, encryption techniques to further protect shared information, advanced systems for identifying in near real-time vulnerabilities and risks, and intuitive dashboards that provide rapid situational awareness are needed to promptly identify and respond to existing and new cyber threats. Importantly, SPHINX delivers actionable information related with cybersecurity to users within the Healthcare domain, contributing to increase the degree of cybersecurity awareness within the organization and put in place appropriate policies and practices. Like with any other transition, improving the levels of cybersecurity in Healthcare organizations is not going to be achieved instantly. It will be an ongoing process that shall require the commitment of all related stakeholders.

To fight existing and emerging cyber threats in the healthcare domain, a holistic approach to cybersecurity needs to be developed, enabling cybersecurity to become an integral part of patient safety. New legislation and regulations are in place to facilitate change, which applies to human behavior, technology, and processes. The SPHINX Toolkit allows healthcare organizations to understand and adapt to threats as they evolve while creating a layered security framework that promotes technological innovation to better protect patient data, minimize threats to patient health and safety, and ensure the privacy and confidentiality of sensitive information shared through mobile, cloud, or IoT-enabled environments. All users become aware of how cybersecurity works, the cyber vulnerabilities and threats that exist, and how they may be better managed. SPHINX technologies offer cutting-edge visualization with the risk radar method to ease awareness and decision-making in critical cases, including when human intervention is required.

Overall, SPHINX increases cybersecurity protection levels in the healthcare domain, tackling three main barriers of limited awareness and understanding on cybersecurity: (a) knowledge on cybersecurity issues and processes, (b) low usability of cybersecurity solutions at hospitals and care centers, and (c) the current vulnerabilities of cybersecurity solutions.

Acknowledgments

The present study was funded through the Project SPHINX “A Universal Cyber Security Toolkit for Health-Care Industry” of the European Union’s Horizon 2020 research and innovation program under the grant agreement No 826183.

References

[1] Use Cases Definition and Requirements Document v1. SPHINX Research and Innovation Action. Grant Agreement No. 826183, December 2019.

[2] Cyber Attacks: In the Healthcare Sector, Center for Internet Security, 2019.

[3] Ladi Adefala, Healthcare Experiences Twice the Number of Cyber Attacks As Other Industries, Fortinet, March 6th 2018, https://www.csoonline.com/article/3260191/healthcare-experiences-twice-the-number-of-cyber-attacks-as-other-industries.html.

[4] Cyber Claims Study, Net Diligence, Version 1.0, https://netdiligence.com/wp-content/uploads/2018/11/2018-NetDiligence-Claims-Study_Version-1.0.pdf, 2018.

[5] Nate Lord, Top 10 Biggest Healthcare Data Breaches of All Time, Digital Guardian, https://digitalguardian.com/blog/top-10-biggest-healthcare-data-breaches-all-time, 2018.

[6] D. Cutler, *et al.*, Reducing Administrative Costs and Improving the Health Care System. November 15, 2012. N Engl J Med 2012; 367:1875–1878. doi: 10.1056/NEJMp1209711

[7] Reimagining the FDA's Approach: Digital Health Innovation Action Plan https://www.fda.gov/MedicalDevices/DigitalHealth/default.html, 2019.

[8] SPHINX project, 'A Universal Cyber Security Toolkit for Health-Care Industry', has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 826183.

[9] SPHINX Architecture v2 Document SPHINX Research and Innovation Action. Grant Agreement No. 826183, 2020.

[10] National Institute of Standards and Technology (2018). Framework for Improving Critical Infrastructure Cybersecurity. Version 1.1. April 16, 2018. https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf.

[11] ENISA Security and Resilience in eHealth – Security Challenges and Risks. European Union Agency for Network and Information Security. 2015.

[12] 2011/890/EU: Commission Implementing Decision of 22 December 2011 providing the rules for the establishment, the management and the functioning of the network of national responsible authorities on eHealth. Official Journal of the European Union. Accessible at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32011D0890.

[13] The Internet of Things 2019. Peter Newman. Business Insider Intelligence. January 2019.

[14] Unlocking the potential of the Internet of Things. James Manyika, Michael Chui, Peter Bisson, Jonathan Woetzel, Richard Dobbs, Jacques Bughin, and Dan Aharon. McKinsey Global Institute, June 2015.

[15] 19 million will use remote patient monitoring by 2018. MEDCITY News. http://medcitynews.com/2014/06/biggest-market-remote-patient-monitoring/.

[16] Regulation (EU) No 1025/2012 of the European Parliament and of the Council of 25 October 2012 on European standardisation, amending Council Directives 89/686/EEC and 93/15/EEC and Directives 94/9/EC, 94/25/EC, 95/16/EC, 97/23/EC, 98/34/EC, 2004/22/EC, 2007/23/EC, 2009/23/EC and 2009/105/EC of the European Parliament and of the Council and repealing Council Decision 87/95/EEC and Decision No 1673/2006/EC of the European Parliament and of the Council Text with EEA relevance. Accessible at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32012R1025.

[17] Exchange of Electronic Health Records across the EU. European Commission. Last update: 6 February 2019. https://ec.europa.eu/digital-single-market/en/exchange-electronic-health-records-across-eu.

[18] C(2019)800. Commission Recommendation of 6.2.2019 on a European Electronic Health Record exchange format. European Commission. 2019. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=57253

Part III

Securing Critical Infrastructures of the Energy Sector

DOI: 10.1561/9781680836875.ch13

Chapter 13

Security Challenges for the Critical Infrastructures of the Energy Sector

By Dušan Gabrijelčič, Denis Čaleta, Theodore Zahariadis, Francesca Santori, Corrado De Santis and Teni Gasparini

13.1 Introduction

Security is of key importance for the development of an individual and the society. Particularly the means for and the forms of an organized provision of security have changed dramatically throughout history, influenced significantly by constantly new technologies and scientific evidence. The globalization of the world, and thus indirectly of security, poses serious dilemmas to the modern society about how to continue basing its development on the fundamental requirements related to the free movement of goods, services, and people, and, on the other hand, about how to keep threats at an acceptable risk level. The emergence of asymmetric threats to national and international security is based on completely different assumptions and perceptions which we were used in the past based on the static approach of managing conventional threats. The changing social conditions and tensions caused by the rapid technological development found particular social environments totally unprepared for confronting the new global security situation and, above all, the newly-emerging complex security threats. Dynamic changes and unexpected technological development have contributed to even greater complexity of this dimension. The fact that the modern society depends entirely on the functioning of technology makes this society even more vulnerable in terms of security. Energy and especial electricity are in this respect even more important for normal functioning of modern society based on technology. Moreover, it makes individual threats and risks related to the proper functioning of infrastructure even more uncontrollable. Certain infrastructure segments, especially sector of electrical power, are so important for the functioning of the society that their non-functioning or limited functioning could have serious consequences or cause serious trouble for that society [1].

Critical infrastructure and business-core applications can be attacked by means of many different vectors. Expanding on the previous analysis, it should be kept in mind that CI is, at an operative level, ordinary business with all the typical weaknesses that this implies.

Electrical grids offer a wide range of targets that can impart a great deal of damage on an entire system. Small-scale attacks can affect much greater systems because the entire grid is interconnected. Project participants warned that once one component is compromised, an entire system could be subject to a cascading failure, thus impacting far more than the initial target [2].

13.2 Energy Sector as a Critical Infrastructure

A critical infrastructure is often identified as that infrastructure whose incorrect functioning, even for a limited time period, may negatively affect the economy of individual subjects or groups, involving economic losses and/or even expose people and things to a safety and security risk [3]. Within the European Union, a Critical Infrastructure is defined as “an asset, system or part thereof located in member states which is essential for the maintenance of vital societal functions, health, safety, security, economic or social well-being of people, and the disruption or destruction of which would have a significant impact in a member state as a result of the failure to maintain those functions” [4]. While a European Critical Infrastructure (ECI) is defined as a “critical infrastructure located in Member States the disruption or destruction of which would have a significant impact on at least two Member States. The significance of the impact shall be assessed in terms of crosscutting criteria. This includes effects resulting from cross-sector dependencies on other types of infrastructure”. The designation of a critical infrastructure as an ECI is the result of a complex technical-political process, which arises from the potential impact that can be caused by a failure/destruction of an infrastructure in terms of sectoral and inter-sectoral relevance.

13.3 CEI Basic Concepts, From Threats to Risk

In this section, basic concepts addressed and discussed in the paper will be defined. A *threat* is core concept. Following from standards and attack perspective, as defined in Internet Security Glossary RFC4949 [5], threats are related to *threat agents* and *actions* against *system resources*, exposed by *vulnerabilities* and protected by *countermeasures*. Common Criteria [6] defines similar concepts, namely *threat actions* that give rise to *threats* which increase *risks* to *assets* protected by *countermeasures*. In the discussion below, both standards concepts will be used interchangeably.

A threat can be defined as "*Any circumstance or event with the potential to adversary impact an asset resulting in unauthorized disclosure, deception, disruption or usurpation as a threat consequence.*" A study of threats has been performed based on broad cyber oriented "ENISA Threat Taxonomy – A tool for structuring threat information" [7] and complementary, more physically oriented characterization of hazards and triggered events in CEI defined by OSCE [8]. Both has been used as a basis to study the threats that indirectly and directly affect the continuous functioning of CEI and are not limited only to the physical and cyber risks but cover the entire segment of risks.

An asset is "*An entity of a value for its owner.*" Smart Grid Coordination Group Smart Grid Architectural Model (SGAM) [9] defines the smart grid scope according to use cases studied in the group. The scope is defined in 3D cube model, namely (1) domain, from generation, transmission, distribution, distributed energy resources to customer; (2) zones, from process to market; and (3) interoperability layers, from components, communication, information, function to business layer. In all dimensions, various assets can be identified, like generators, transformers, poles, lines, communication links, information, processes, etc.

Threat agent, according to ENISA Threat Landscape report [10], is defined as follows: "*A threat agent is any person or thing that acts (or has the power to act) to cause, carry, transmit, or support a threat.*" The document exposes cyber-criminals, insiders, nation states, corporations, hacktivists, cyber-terrorists, and script kiddies as most visible agents. The taxonomy covers only cyber threats. Additional agents can be added to the selection, like nature and environment, hooligans, vandals, military, or even AI.

Threat agents exploit *vulnerabilities* to realize a threat. A vulnerability is defined [5] as "*A weakness in a system's design, implementation, or operation and management that could be exploited to violate the system's security policy.*" All systems have weaknesses, but they don't get always exploited. For example, in energy domain, a well-known weakness is separation in organizational silos; every silo implements technical and security solutions for itself; policies are often misaligned;

and organizations communicate in case of crisis mainly through unformal channels. The vulnerability exists but seems to be rarely exploited.

Risk combines vulnerabilities and threats with probabilities of loss: "*An expectation of loss expressed as the probability that a particular threat will exploit a particular vulnerability with a particular harmful result.*" Risk is important for threat analysis to be able to assess an impact of a threat or vulnerability.

13.4 CEI Threats Analysis and Evaluation Methodology

Threats to CEI systems are broad and varying, from very simple to complex and cascading as has been presented in introduction. A depth of threats space is further stretched due to a number of stakeholder domains, from generation, large-scale renewables, transmission to distribution. The space needs to be addressed holistically, allowing for continuous exploration and evaluation. Proposed methodology, evaluated through the DEFENDER project, is presented in Figure 13.1. The methodology interweaves studies, analysis, and evaluation with real stakeholder environments. In the figure, three flows are presented: analysis and evaluation flow denoted with full, requirements capturing; feedback flow denoted with dashed; and unknown threats flow denoted with dotted lines.

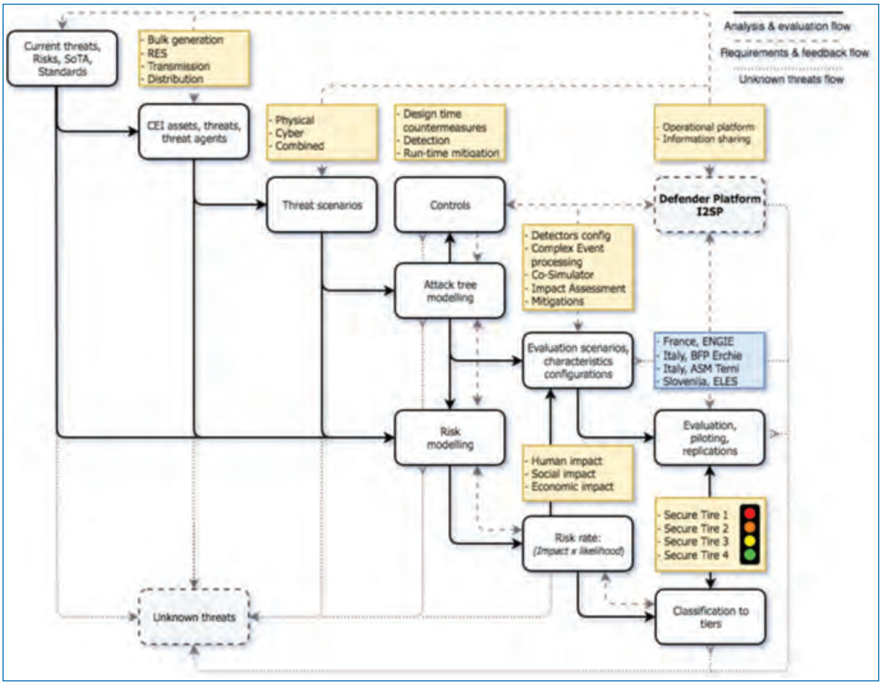


Figure 13.1. CEI threats methodology.

The analysis and evaluation flow starts with a review of current threats, state of the art, and status of standardization in the field. The current threats are of utmost importance since they indicate what kind of challenges needs to be addressed in threat mitigation. The threats examples are further summarized in Section 13.5. The flow continues with stakeholder domains investigation. CEI assets need to be determined, specific threats to the domains identified, and threat agents named and analyzed.

Current threats and domain-specific inventory are analyzed through threat scenarios. A threat scenario is a template allowing to collect threat-specific concerns, main assets involved, relevant vulnerabilities the attacker can exploit, consequences if the attack is successful, and initial detection and mitigation possibilities. Threat scenarios are a basis for attack tree modeling where the attacker steps are laid down and analyzed. The mitigation possibilities are further detailed in controls specification as design time countermeasures and run-time mitigations besides detection capabilities. Attack tree modeling and threat scenarios are further detailed in Section 13.6.

All steps so far form an input to a risk modeling phase as well evaluation phase of the methodology. The risk modeling aims at evaluation of risk rate per attack tree step and combined risk value of an asset looking from human, social, and economic point of view. The combined risk value is used to classify the assets into secure tiers. Risk modeling is further explained in Section 13.7.

In the evaluation phase, the threat scenarios and modeled attacks are transformed to evaluation scenarios. The evaluation scenarios adapt the threat scenarios to the pilots, define success and failure outcomes of the evaluation, and specify evaluation characteristics that will be evaluated during evaluation. Detailed configurations are prepared for DEFENDER Platform components, including with detectors, complex event processing, co-simulator, impact assessment, and mitigations configurations. Evaluation phase is continued with evaluation in pilots, in real environment, in trial sites related to stakeholder domains. Evaluation modeling is explained in brief in Section 13.8 and the piloting in Section 13.9.

All analysis and evaluation flow steps provide requirements for a DEFENDER Platform and Incident Information Sharing Platform (I2SP). On other hand, the development of the platforms needs to be in tight sync with the steps impacting the specification, feedback from the development phase in form of updates to attack modeling phase and evaluation specification phase is required. In particular, the attack trees and configurations of the components are tightly connected to the platform implementation, see (reference to other two articles).

Finally, all analysis and evaluation flow, and requirements and feedback flow steps need to be continuously evaluated to address a challenge of unknown threats. Threats need to be continuously managed in their lifecycle for the system and its administrators to be able to anticipate and be prepared for unknowns. Brief overview of unknown threats dimensions is presented in Section 13.10.

13.5 Challenging CEI Threats

Overview of current CEI threats can give insights what are top challenges for CEI protection. The overview is of crucial importance for selection of threat scenarios, assets, risks, and controls to be evaluated in the CEI threats analysis and evaluation process. The focus of the process is both physical and cyber threats and combination of them both. Asset and organization span are cross entire CEI sector. The threats are summarized and grouped below in few distinct and most challenging categories, namely physical threats, cyber threats, organizational threats, and technology threats. Any combination of such threats is possible, due to modernization and decentralization of the grid attack possibilities will be only larger.

Physical threats can be divided in two large groups, physical attacks and forces of nature. Physical attacks on the infrastructure are not rare at all [11]. Attacks are directed at all electrical grid segments and elements. Gun usage is often reported near power lines, substations get broke in and vandalized, switching fields are visited by people or animals, generation facilities like PV plants are targeted by thieves [12]. Targeted attacks on transformers can have huge and long-lasting impact on entire grid energy services provisioning as well on the environment.

Forces of nature threat the infrastructure, in particular the power lines. Strong wind, landslides, fires, and, especially, glazed ice can have strong and broad impact on both transmission and distribution services. For example, because of glazed ice in Slovenia in 2014 more than 15% of population was without electricity for several days [13]. The situation was close to disaster, a 400 KV line put in test operation just a month before was the only reason for a third of the nation not to be out of electricity for extended period of time [14]. The primary vulnerability of the energy infrastructure lies in its spread distribution all-over populated areas, exposed and long interconnections between grid segments and fragility of the infrastructure not designed to sustain all scope of nowadays threats.

Cybersecurity main concern is preservation of confidentiality, integrity, and availability properties of the CEI information system [15]. Confidentiality is related to personal or customer data, industrial processes, confidential information, financial data, or sensitive data used to build up cyberattacks. Social attacks, espionage, attentional or unintentional disclosure, malware exfiltration, or unauthorized network access are primary ways to realize threats to confidentiality.

Integrity loss of industrial control systems can have far-reaching consequences if the attack moves the industrial process beyond controllable levels, tampers the possibility of the system trigger emergency procedures, or modifies sensitive information related to, for example, to billing which leads to financial loss or fraud. Typical example of integrity loss incident is the Stuxnet [16]. Loss of availability refers to loss of provisioning of electricity services; all segments of the electricity grid need to be resilient in provisioning to be able to guarantee the service. Most eminent cyberattacks causing loss of availability in the domain were carried out in Ukraine in 2015. It is easy to generalize such attacks to other parts of the world [17].

Electrical grid companies face many challenges that could be understood as *organizational threats*. The security services are provided by different parts of the organization, physical security is separated from cybersecurity, cybersecurity for a technical network is provided separately from business network. Information silos lead to inadequate security and risk management, unaligned and dependency weak inventory management, poor cooperation between departments and inefficient incident response procedures.

Modern technology advances expose novel *technology threats* to the CEI infrastructure beyond cyber. Drones are nowadays easily accessible off the shelf. They can carry a load of few kilos, fly fast, and at distance. Off-the-shelf drones were used for drug smuggling [18] and in Syrian war [19]. A distributed and overly exposed energy grid infrastructure is very hard to protect. Power lines, transformers, and switching fields could be an easy target with huge consequences.

13.6 Threat Modeling

Threat modeling starts with collection of threat scenarios. A threat scenario has a number of definitions in literature. The Canadian government Threat and risk assessment working guide has specified the threat scenario as: "*A threat scenario consists of one or more threat events, carried out by a threat agent, that could result in the compromise of an asset*" [20]. The definition hasn't changed much latter; NIST has added in its Guide for Conducting Risk Assessments [21] a notion of ordering of the events in time: "*A threat scenario is a set of discrete threat events, attributed to a specific threat source or multiple threat sources, ordered in time, that result in adverse effects.*"

The definitions tell already a lot. There is a need to know what are the *threat events*, *threat agents*, *threat sources,* and *assets* and how are the events of the scenario ordered in time. Based on the definitions, the following threat scenario template is proposed to collect the threats, tying the threat analysis to a pilot site:

- *Pilot site*: defines the assets of the pilot site and main configuration site details that can affect any threat scenario essential components. The information helps to focus the scenario towards the site and report on realistic environment where the scenario and consequent detection and mitigation can happen.

- *Main scenario concern*: lays out the main scenario concern, what will go really wrong if the scenario happens? The concern allows multiple variations of the scenario to happen. The concern helps to keep the focus on main issues the scenario addresses. If the concern is not clear, the scenario should be split in smaller, more focused parts.
- *Description*: specifies the steps of the scenario, how the scenario is implemented, refers to pilot site description and elements when relevant. Here the template asks for the events and possibly for the time order of the events, but the description serves more as an illustration of possible threat scenario realization.
- *Relevant vulnerabilities*: describes the vulnerabilities the scenario assumes to be used/misused. The vulnerabilities here are not detailed vulnerabilities as are reported by CERTS or similar organizations, but general vulnerabilities that can be further specified towards the detailed during the analysis.
- *Threat consequences*: the threat consequences of the scenario, strongly related to main concern. Multiple diverse consequences are possible. The consequences can have important impact on the further scenario evaluation and their categorization. More detailed, more precise the categorization can be.
- *Related scenarios*: for documentation purposes, primary for similar scenarios already discussed in research and scientific literature.

The threat scenarios give needed information for the next step, modeling the threats through attack trees.

13.6.1 Attack Trees

Threat and attack modeling have been widely used by security experts as means to identify how an attack agent or group is able to exploit the vulnerabilities of a system to compromise its security, attack trees being one of the most popular relevant techniques. Though attack trees have been used for years under various forms and names, it was not until 2000 [22] that they were clearly described as a systematic method to characterize system security against attacks of various types and manifestation [23]. An attack tree defines a collection of possible attacks, simply called *attack suite*, against a given system. Usually, attack trees represent the end result (e.g., compromise of the system security or survivability) as the root of the tree, whereas the ways that the attackers can cause such compromise are represented as lower-level tree nodes. In this sense, each attack tree presents the full range of actions that an attacker could take to compromise the system; each path through the attack tree represents a unique attack on the system, the set of all available attacks being called *attack surface*. Typically, a complete system comprises a large set of attack trees (called forest) that pertain to its operational features.

Modeling via attack trees has been mainly employed in cybersecurity contexts, primarily in the field of information and communication technology (ICT), most often to deliberately present the security of software systems and applications. However, attack trees have been also employed in the framework of modeling the cybersecurity of smart energy systems. Indicatively, NESCOR (the U.S. National Electric Sector Cybersecurity Organization Resource initiative) has performed substantial research on the cybersecurity failures for energy grids and has published scenarios, relevant impact analysis, as well as proposed mitigation actions [24, 25]. However, in contrast to the goals presented in this chapter, NESCOR only targets at cybersecurity, neglecting the cyber-physical aspects. Presented attack tree analysis is extended as well with detection possibilities, and evaluation and risk modeling phase.

13.6.2 Attack Tree Example

In Figure 13.2, an attack tree representing a cascading threat of cutting water supply through electrical grid is presented. The following notations are used in the tree: octagons present *attack tree nodes*; dashed lines, OR relationship; and full lines, AND relationship. Double octagons present common attack tree nodes as being composed from many nodes and relationships. For example, “Getting into substation” is based on a common node “Getting into building.” Triple octagons present other threat scenario, in our example “Damage power line or pole.” The attack tree indicates that the power supply can be cut if the water pumps are stopped and the water reservoir is drained. The water pump can stop working if the power supply is cut and the backup power supply has been disabled or is not available. The power supply service can be loss because of power line failure or it is maliciously controlled through powering substation. Powering substation can be attacked either physically or through cyber means.

13.6.3 Attack Tree Analysis Complementary Information

In analysis phase, each attack tree node is complemented with information describing means of specific attack node detection and mitigation. The detection part describes node-specific details of detection together with rough indication of information needed to implement the detection. The mitigation part describes design-time mitigations and run-time mitigations. The design-time mitigations are countermeasures that can prevent a tree node attack step by design, for example, by

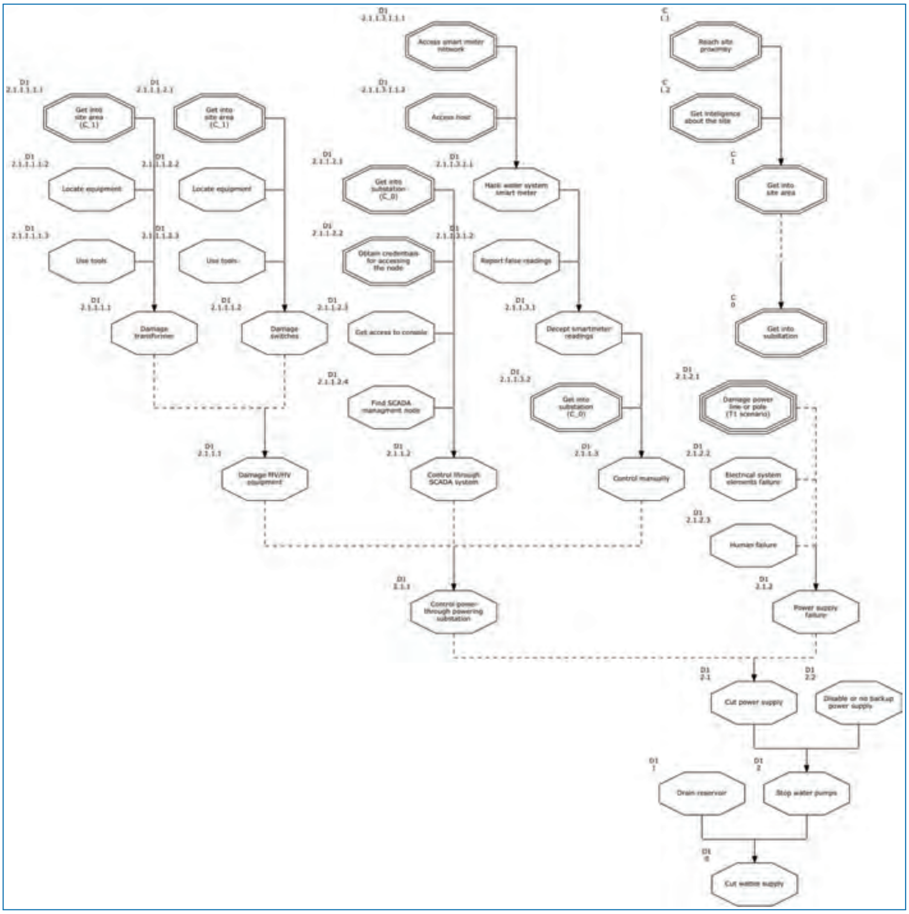


Figure 13.2. Attack tree representing a cascading threat of cutting water supply through electrical grid.

introducing a redundancy or additional fortifications in the system. The run-time mitigations are one that can be applied dynamically during the system operation.

The analysis is further extended with system response to threat realization as well to run-time mitigation triggering. In both cases, the consequences for the system and services the system provides are studied and noted.

13.7 Modeling Risks

The primary objective of risk modeling is to define a process for analyzing and assessing the attack trees in terms of their risk exposure, which is expressed as Risk Rate and ultimately to result in a CEI secure tier classification.

13.7.1 Risk Management Process

The Risk management process selected to be followed is based on ISO/IEC 27005 [26]. The process consists of the following steps: *Context Establishment*, where the scope and the boundaries of the effort are defined; *Risk Assessment*, where risks are identified, analyzed, and evaluated; *Risk Treatment*, where a risk treatment plan is prepared and any residual risks are identified; *Risk Acceptance*, where the risk treatment plan and related residual risks are accepted; *Risk Communication*, which involves the exchange and/or sharing of information about the identified risks with the relevant stakeholders; and *Risk Monitoring and Review*, where actions are taken to monitor the overall process and to review it for improvements.

13.7.2 Context Establishment

To define the impact of a threat scenario, one should consider a number of parameters: cost of equipment and maintenance, and cost of losses, namely material, reputation, availability, or confidentiality. For every site, examples will be given for piloting sites in DEFENDER project, the assets need to be identified, threats to the assets need to be analyzed, threats mapped to the assets, vulnerabilities of the assets listed, consequences of the threats identified, and existing controls recognized. All basic relationships are indicated in Figure 13.1.

13.7.3 Risk Estimation and Classification

For the estimation of the risk, as parameters the human, economic, and social impact and the risk likelihood has been used. Following the approach proposed by ANSSI [27], in the risk estimation and assessment process, a qualitative approach has been selected. The selection has assumed that the qualitative approach can be easily converted to a quantitative one, by assigning to each level a range of specific values.

Impact categories selected were from minimal to severe in five levels from 1 to 5. Risk likelihood was defined in five levels as well, from rare to almost certain, from 1% to 80% probability. Calculation of the likelihood and impact has been performed on each attack tree node as suggested by Edge [28].

For the attack tree, in case of OR relationship, it is necessary that at least one of the ancestor nodes is accomplished in order for the child node to be accomplished, too. In order to calculate the probability for the child node, we need to find out the probability of NOT accomplishing any of the parent nodes. If this probability is subsequently subtracted from 1, then the outcome is the probability of having the child node accomplished. As far as the impact is concerned, the impact of the child node is calculated as the maximum impact of any of the ancestors.

In case of an AND relationship, all of the ancestor nodes must be accomplished in order for the child node to be accomplished, too. This lead, logically, to the conclusion that the probability of the child node is calculated as the product of the probabilities of its ancestors. As far as the impact is concerned, the logic is similar to the logic used for the calculation of the probability in the case of an OR relationship, in combination with a normalization operator.

Combining the risk impact and the risk likelihood, by actually producing their product, we are able to determine the risk rating that refers to the level of the risk exposure for a particular threat scenario. There exist 25 combinations of risk impact and likelihood which give combined a risk rating matrix. The matrix can be divided into security tires, where the green tire represents most secure (low likelihood and impact) to red tire (high likelihood and impact) as is presented in Figure 13.1.

13.8 Modeling Evaluation

Evaluation is organized around two concepts: scenarios and characteristics. The evaluation scenarios are a simple concept allowing organizing the evaluations of controls and procedures developed for threat mitigation into manageable units. The evaluation characteristics allow to assess features and qualities of the solutions in a unified and standardized way.

The evaluation scenario presents a single testing unit which can be meaningfully evaluated on its own. It is intended that the single unit would allow testing of a single component, solution, or feature of the developed technologies. The scenarios can be interrelated to each other. Therefore, one successful scenario evaluation can require previous scenarios to be successful as well. Main components of a scenario are scenario steps, which define what is the scenario's basic purpose. For these steps, within the same evaluation scenario, the testing steps are proposed. The evaluation of testing steps results in success or failure state of the system. Each evaluation scenario is evaluated according to its scenario test success or failure, and evaluation characteristics as are specified below.

The standard way to evaluate a product provides a series of ISO/IEC standards known as SQuaRE (System and Software Quality Requirements and Evaluation). From five ISO/IEC standard divisions, the threat mitigation evaluation can mostly benefit from quality model, quality measurements, and quality evaluation division. The ISO/IEC 25010 [29] standard quality model is defined with a number of characteristics and sub-characteristics in product quality and quality of use categories. From the model, the following characteristics are most suitable for common, cross evaluation scenarios quality assessment: functional, performance, usability, reliability, security, and reusability from product quality perspective and efficiency and dissatisfaction from quality of use perspective. ISO/IEC 25023 on "Measurement of system and software product quality" [30] provides some great insights how to assess the product quality characteristics. The standard defines simple and applicable measurement functions for functional completeness based on proportions estimation, performance measures based on mean response time, response time adequacy, mean turnaround time, throughput, resources utilization, and capacity adequacy, etc.

Some other characteristics should be evaluated beyond ones defined in ISO/IEC 25000 family, for example:

- *Threat mitigation*: how well did the controls and procedures address the threat scenario that has been evaluated?
- *Risk mitigation*: how well did the controls and procedures address the risks originating in the threat scenario?
- *Ethic and privacy*: per threat scenario handling of ethical and privacy aspects needs to be evaluated.

13.9 Pilot Evaluation

The threats analysis and evaluation methodology has been evaluated in the DEFENDER project.[1] The project has developed a Defender Platform, an implementation of controls for cyber-physical security provisioning process. The platform is accompanied by an Incident Information Sharing Platform (I2SP) allowing to exchange attack-specific indicators among Defender Platform systems.

13.9.1 Defender Platform Configuration

The Defender Platform provides a complete system for data fusion, attack detection, situation awareness enrichment, optimized attack mitigation selection, and visualization and control. All the methodology steps described in previous sections have contributed to requirements of the platform. Attack trees, controls, detectors, and mitigations give the platform a skeleton, sensors, and actuators. The risk analysis as presented in Section 13.7 enable a co-simulator to estimate future states of the system according to the attack tree and to calculate impacts of attacker next steps as well of the system response as part of the situation enrichment. All the elements were bind together with a configuration of the Defender Platform and external

[1] DEFENDER project has received funding from the European Union's Horizon 2020 Innovation Programme under grant agreement No 740898, see the project home page for details: https://defender-project.eu/

components—detectors, controls, mitigations, like their geolocation, orientation, angles, timings, etc.

13.9.2 Pilots

The described methodology and the DEFENDER Platform have been evaluated in DEFENDER project pilots. The project has implemented four pilots in France, Italy, and Slovenia. The French pilot operated by ENGIE was in a Combigolfe power plant (France), focused on drone attacks, drone video detection and mitigation by jamming, protection of the perimeter by laser detection and automated drone inspection.

Italian trials were at ASM Terni (Italy), a Distribution System Operator (DSO), and at BFP large-scale renewable, wind farm in Erchie (Italy). In Terni, a Phasor Measurement Unit (PMU)-based fault detection and localization was experimented together with evaluation of the attack on a substation powering a water system of Terni, as is presented in Figure 13.2. In Erchie, a laser-based drone detection and protection of the windmills by a hunting drone was experimented, as well cyber-physical protection of the windmills and substation.

In Slovenia, the trials were led by ELES, a Transmission System Operator (TSO). The Slovenian trials have focused on electrical network fault location by utilizing collocated optical network properties and cross cyber-physical access control improvements.

13.9.3 Result Replication

For the project pilots, more than 20 threat scenarios were described and analyzed. The threat scenarios attack trees were complemented by 28 common attack trees that could be utilized across all the pilots and situations. Experience from the field has shown that it easy to merge and extend the attack trees. Configurations require some time to be adapted to new installation and deployment but allow adaptation and fine tuning. Replicating detectors and controls from pilot site to pilot site was possible and was even considered as a part of development process resulting in components and configurations improvements.

13.10 Unknown Threats

Unknown threats are threats previously unknown to the observed system. According to the definition of threats as is specified in Section 13.3, one can make two observations regarding the definition of the unknown: security needs to be defined and there are many views the unknown can be related to.

Addressing the unknown threats, a number of views have been explored. Most common is a *threat space*, when new threats continuously emerge due to newly discovered vulnerabilities. A *threat scenario space* allows defining and exploring new scenarios as well identifying possible missing ones. An *attack space* views the unknown from an attacker perspective and can give answers to where attack tools are coming from and how they are used for target purpose. A *domain space* lights the issues pertinent to domains like energy, water, finance, etc. Every domain opens new unknown possibilities; threats from one domain can be meaningfully transferred to other domains. A *technology space* introduces novel threats by just either emergence of new technology, for example, drones, automation systems, etc., or opening administrative boundaries of existing technologies. On the end, a *research space* could be used to identify gaps and possible alternative usage of existing techniques.

Methodology as presented in Figure 13.1 indicates the unknown threat spaces and a need to continuously follow the information in entire process of threat analysis and evaluation. In this way, the process becomes permanent and requires a lifecycle management of identification, detection, evaluation, and mitigation of unknown threats.

13.11 Energy Sector Challenges

In this section, major energy sector challenges will be summarized based on the experience of threat modeling and evaluation as were laid out in the previous sections. The energy sector challenges can be roughly divided into three groups: political, organizational, and techno-social. In cross-group dimension, one particular property of the sector can be identified—general disconnection or lack of cooperation—between diverse layers, segments, sectors, organizational units, organizations, stakeholders, etc., resulting in major sector weaknesses and vulnerabilities.

From the *political* point of view, more determination, regulation, and will to support development of sector cyber-physical threats resistance is needed. In Europe, the NIS directive has initiated a framework for addressing the cyber-security challenges. Operators of essential services have been identified and tasked with needed set of security services to be provided for secure operation. List of services, on the end per country specific, is often not comprehensive or complete. The reason for this could be that a comprehensive and complete list would be very demanding to implement in cost and effort needed. Smaller entities in the sector find any demands form regulation already very challenging.

While the NIS directive provides needed scope and vertical structure to the cybersecurity services facilitation, a need to strengthen cooperation of the entities in the field in depth is only partially covered by the directive. The actors, TSOs, DSOs, GEN, Renewables, Aggregators, etc., should cooperate closely at least in best-practice exchanges, attack information, and indicators of compromises exchange and common planning for incident response and consequences mitigation. The regulators and countries policy-makers should foresee and encourage such cooperation not only in the sector but in cross CI sectors as well. It has to be noted that physical threats are selectively covered by regulation, cooperation within sector and cross CI sectors is on discretionary basis only.

Organizational challenges are pertinent to the organizations in the sector. Organizations vary from large to small, so the level of cyber-physical security readiness is diverse as well. What seems to be still common to most of them are silos, dividing operation in technical and business information systems, corresponding two cybersecurity systems governance is disconnected from physical security, etc. Organizations don't have top security policies governing a general direction of security in organization, sometimes even silo policies are missing. The silos challenge overall risk management, dependencies between silos are weakly defined, risks do get biased towards more expensive physical equipment, while the services are what matters. For dynamic threat management systems, a lack of coherent inventory and service dependency management is of crucial concern.

Technological challenges are supercharged by diversity of equipment and standards in the sector, fast digitization, innovative use cases, and novel technologies. They lead to decentralization of the grid, introduction of off-the-shelf technologies, and activation of even the smallest actors, making the grid on one hand more vulnerable and on the other hand more resilient to failure. While the attack surface of the digitized grid has been significantly enlarged, the physical infrastructure has remained almost the same, including with long amortization time. The main vulnerability of physical equipment has remained the same: open exposure, e.g., transformers, and large geographic dispersion, as in case of power lines. While novel technologies development, e.g., drones, enable easier attack of the infrastructure, countermeasures are not yet fully ready for its protection.

Another face of the *techno-social* challenges are targeted attacks. The targeted attacks are most challenging to prevent and can have vast consequences. They often combine multiple types of threats, from social, physical to cyber. Techniques like social engineering, advanced persistent threats (APT), combination of cyber-physical attacks, cascading attacks, and destructive assets attack are so diverse that are hard to detect and mitigate with a single system. The attack life cycle can span from years in case of APT to seconds in case of destructive assets attack. Detection techniques face challenges of minimal data available for detection and very long time spans. On the other hand, the mitigations would need to be available at all geographically disperse locations to be able to be triggered in few seconds time span for meaning full action. The fact also introduces a requirement for full threat management automation which can be a challenge by itself.

13.12 Conclusions

Presented threat modeling and evaluation methodology, together with the process of its implementation, have shown a number of benefits for involved stakeholders. Through the implementation of the methodology in the pilots, pilot-specific threat scenarios have been developed to address the most feared concerns of the pilot owners. The methodology has been able to expose many of the organizational and technology challenges discussed in Section 13.11. Systematic work and improvements of the pilot owner systems has enabled the Defender Platform to successfully fuse sensors data, detect attacks, assess impact through co-simulation, propose mitigation actions, and start the countermeasures. Nevertheless, the methodology could only indicate and expose the challenges as a requirements, threat scenarios, attack trees, attack steps risks, and probabilities. In this way, it can fuel a techniological solution to address some of the challenges. The political and organizational challenges need to be addressed separately requiring all stakeholders' cooperation.

References

[1] Čaleta, D. (2011). A Comprehensive Approach to the Management of Risks Related to the Protection of Critical Infrastructure: Public-Private Partnership. In: Čaleta, D. & Paul Shemella (Eds.). Counter terrorism challenges regarding the process of Critical Infrastructure Protection (pp. 15–26). Ljubljana, Monterey: ICS, Centre for Civil Military Relations.

[2] Centre for European Policy Studies (CEPS). Protecting Critical Infrastructure in the EU. CEPS Task Force Report. 2010.

[3] Brunner, M. and Suter, E. M. International CIIP Handbook 2008/2009. Center for Security Studies, ETH Zurich, 2008.

[4] Directive of the council on the identification and designation of European critical infrastructures and the assessment of the need to improve their protection. COUNCIL DIRECTIVE 2008/114/EC, 2008.

[5] Shirey, R., "Internet Security Glossary, Version 2," FYI 36, RFC 4949, DOI 10.17487/RFC4949, August 2007. https://www.rfc-editor.org/info/rfc4949 [Online; accessed on 8th of February 2020].

[6] ISO/IEC 15408:2009, Information technology—Security techniques—Evaluation criteria for IT security—Part 1: Introduction and general model. Third edition 2009, corrected version 2014. 2009.

[7] ENISA. ENISA Threat Taxonomy. 2016. https://data.europa.eu/euodp/en/data/dataset/enisa-threat-taxonomy-1 [Online; accessed on 8th of February 2020].

[8] OSCE, Organization for Security and Co-operation in Europe. Protecting Electricity Networks from Natural Hazards, 2016. http://www.osce.org/secretariat/242651?download=true [Online; accessed on 8th of February 2020].

[9] CEN-CENELEC-ETSI Smart Grid Coordination Group. SG-CG/M490/F_Overview of SG-CG Methodologies Version 3.0. 2014.

[10] ENISA Threat Landscape Report 2016–2018. https://www.enisa.europa.eu/publications/enisa-threat-landscape-report-2018 [Online; accessed on 8th of February 2020].

[11] U.S. Department of Energy. Electric Grid Security and Resilience: Establishing a Baseline for Adversarial Threats. 2016.

[12] Solar panel theft on the rise in Germany. Financial Times JUNE 7, 2015 Chris Bryant in Frankfurt. For details see: https://www.ft.com/content/745382e0-0525-11e5-9627-00144feabdc0 [Online; accessed on 8th of February 2020].

[13] Troppauer W., Lovrenčić V., Gubeljak N., Nemeth B., Kovaè M., Gocsei G. and Krisper U. Advanced monitoring of icing and prevention against icing on overhead power lines. Proceedings of Int. Workshop on Atmospheric Icing of Structures IWAIS 2019 – Reykjavík. June 2019.

[14] Energy Agency. Report on energy sector in Slovenia 2014. 2014.

[15] Avi, A., Zienis, J.-C., Laprie, B., Randell, C. and Landwehr, C. (2007). Basic concepts and taxonomy of dependable and secure computing. IEEE Transactions on Dependable and Secure Computing – TDSC. 1.

[16] Kushner, D., 2013. The real story of Stuxnet. IEEE Spectrum, 3(50), pp. 48–53. 2013.

[17] Sullivan, J.E. and Kamensky, D. How cyber-attacks in Ukraine show the vulnerability of the US power grid. The Electricity Journal, 30(3), pp. 30–35. 2017.

[18] CNN online. Drone carrying drugs crashes south of U.S. border, http://edition.cnn.com/2015/01/22/world/drug-drone-crashes-us-mexico-border [Online; accessed on 8th of February 2020].

[19] Washington Post: Use of weaponized drones by ISIS spurs terrorism fears, https://www.washingtonpost.com/world/national-security/use-of-weaponized-drones-by-isis-spurs-terrorism-fears/2017/02/21/9d83d51e-f382-11e6-8d72-263470bf0401_story.html [Online; accessed on 8th of February 2020].

[20] Government of Canada, Communications Security Establishment. Threat and risk assessment working guide. 1999.

[21] NIST Special Publication 800-30, Revision 1. Guide for Conducting Risk Assessments. 2012.

[22] Schneier, B. Secrets & Lies: Digital Security in a Networked World: Wiley Computer Publishing Inc., pp. 90–91. 2000.

[23] Moore, A., Ellison, R. and Linger, R. "Attack modeling for information security and survivability," Pittsburgh, Pennsylvania, 2001.

[24] Electric Power Research Institute (EPRI). Integrating Electricity Subsector Failure Scenarios into a Risk Assessment Methodology, 2013.

[25] NESCOR. Analysis of Selected Electric Sector High Risk Failure Scenarios V. 3.0, 2015.

[26] ISO/IEC 27005:2011 – Information technology—Security techniques—Information security risk management (second edition), 2011.

[27] Classification Method and Key Measures, Cybersecurity for Industrial Control Systems, ANSSI. 2013.

[28] Edge, K., Raines, R., Grimaila, M., Baldwin, R., Bennington, R. and Reuter, C. The Use of Attack and Protection Trees to Analyze Security for an Online Banking System, presented at 40th Annual Hawaii International Conference on System Sciences (HICSS'07), 2007.

[29] ISO/IEC 25010, Systems and software engineering—Systems and software Quality Requirements and Evaluation (SQuaRE)—System and software quality models. 2011.

[30] ISO/IEC 25023, Systems and software engineering—Systems and software Quality Requirements and Evaluation (SQuaRE)—Measurement of system and software product quality. 2016.

DOI: 10.1561/9781680836875.ch14

Chapter 14

Securing CEI "By-Design"

By Nikolaus Wirtz, Alberto Dognini, Abhinav Sadu, Antonello Monti,
Guilherme Brito, Giovanni di Orio, Pedro Maló
and Cosmin-Septimiu Nechifor

14.1 Resilience and Vulnerability of CEI

Critical Energy Infrastructures (CEI) do not only consist of the power grid and electrical equipment, but also of the communication infrastructure, measurement devices, and control functionalities, which are networks in different domains and at different scale. Dependencies between these networks include power supply of communication infrastructure by the power grid and the dependence of components in the power grid on Control Centre commands. In combination, the networks and their interdependencies form a complex system where a small initial set of component failures has the potential to cause cascading failures leading to partial or complete breakdown of the system. Consequently, for a comprehensive analysis of the vulnerability and resilience of such a system, all the relevant domains and their interdependencies must be included and modeled.

For distributed systems with many components, such as the transmission or distribution grids, a graph-theoretic approach can be applied to perform an analysis of potential cascading in the system. A graph $G = \langle V, E \rangle$ is a mathematical object formed by a set of vertices V and a set of edges E, where each edge $e = \langle u, v \rangle$ is connecting two vertices u and v. We are using simple graphs to represent the network's topology and directed graphs to represent the dependency between components. Simple graphs do not allow self-loops (edges that are connected to the same vertex) and duplicate edges (multiple edges that are connected to the same vertices) and its edges are bidirectional. Self-loops are not required for the modeling, while duplicate edges could be used to represent redundant lines, cables, or links. However, redundancy can also be expressed as an attribute of an edge in a simple graph. In general, power grids and communication networks allow the flow of energy and data in both directions of a line or link, making a simple graph a suitable representation of the network. Directed graphs, however, as described in [41], consist of arcs as unidirectional edges, which point from one vertex to another. The dependency graph represents components as vertices of the graph and arcs as a dependency, pointing from the dependent component to the supporting component.

14.1.1 Example of use Case and Implementation

The system investigated in this sample use case is based on a generic distribution grid segment [20], which has been extended by a physical and a logical communication network and a measurement and control network as described in Figure 14.1.

The **power grid** consists of multiple loads at the LV level, supplied by four different substations in a radial configuration. Since the grid is meshed, different paths are available to supply the loads under normal operating conditions. This redundancy is utilized for network reconfiguration in case of a fault. For the vulnerability analysis, we simplified the network by aggregating the loads of each section; since in case of a fault in the grid, we expect either all the load nodes of a section to function or none of them.

The real physical communication equipment is not known, and we assume that an **optical communication network** is used to support measurement and control of the power grid. The optical cables are parallel to power lines, thus connecting the same vertices of the graphs; and there is a communication node at each load node of the power grid.

The **logical communication network** uses the optical network's infrastructure to exchange data between nodes. Each link of this network connects two logical

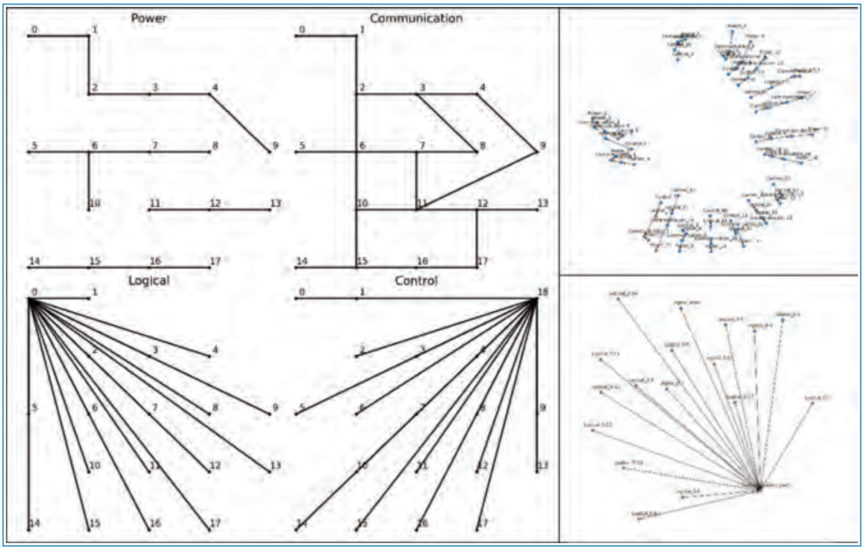


Figure 14.1. Network topologies and node and edge dependency graphs for the example use case.

communication nodes and requires a path in the optical network between the respective physical communication nodes.

The grid operation uses a **measurement and control network** with a central control node (the control center node), which gathers measurement data, processes it, and provides control functionality. This control center node is connected to each of the measurement and control nodes, representing the nodes in the power grid that are equipped with measurement devices and switching equipment that can be remote-controlled.

To model the **interdependencies**, we assume that each node of the power grid supplies the local communication node with power, which in turn supports the local node of the logical communication network. Each node in the logical communication network finally supports one or multiple nodes of the measurement and control network. All the dependencies between nodes are represented in a node dependency graph, which is shown in Figure 14.1 for the default configuration of the scenario. Every link in the logical communication network is dependent on the availability of a path between the respective nodes in the optical communication network. If there is no path available, the link fails, as data can no longer be exchanged.

In addition to the interdependencies, there are domain-specific **intra-dependencies** in each of the networks. For the power grid, we assume each subgraph must include at least one load node as energy consumer and one substation node as energy provider. If this is not the case, all nodes of the subgraph fail. For the optical and logical communication networks, we assume that isolate nodes fail, since they are no longer able to send or receive data and thus cannot fulfill their purpose. For the measurement and control network, we assume each subgraph must include at least one control center node as measurement data consumer and control command provider and one measurement node as measurement data provider and control command consumer.

In this example scenario, four different configurations of the system are evaluated and compared:

- A **default configuration,** where no by-design measures are implemented
- The **DV configuration,** where Double Virtualization is applied to virtualize part of the functionality of the measurement and control network and thus make it independent from the actual hardware. In this configuration, the control functionality is virtualized, and we assume that it can be hosted by any of the communication nodes at the power grid substations. This is represented in the model by adding dependencies, as "Control_18" is now depending on "Logical_5," "Logical_13" and "Logical_14" in addition to the dependence on "Logical_0." Consequently, additional logical communication links are added, to connect any of the logical communication nodes located at loads to any of the nodes located at substations.
- The **SR configuration,** where a network reconfiguration algorithm provides service restoration by design for the power grid, to resupply lost loads after a failure in the grid. This is represented in the model by adding edges to the power network, representing the lines with switches that are normally open and can be used to provide redundant paths. These additional lines are only used in the reconfiguration after a fault happened, therefore depending on a control command from the control center. In case the control command cannot be received, the switches cannot close, and the line will not be put in operation.
- The **DV_SR configuration,** where both by-design measures are applied.

To evaluate the performance of the complete system after a fault has occurred, we choose two performance indicators:

- The **number of supplied loads,** as a measure of the service level to energy consumers, is maintained.
- The **number of loads that are controllable,** as a measure of the reliability of the final configuration after the cascading sequence has ended. For loads that

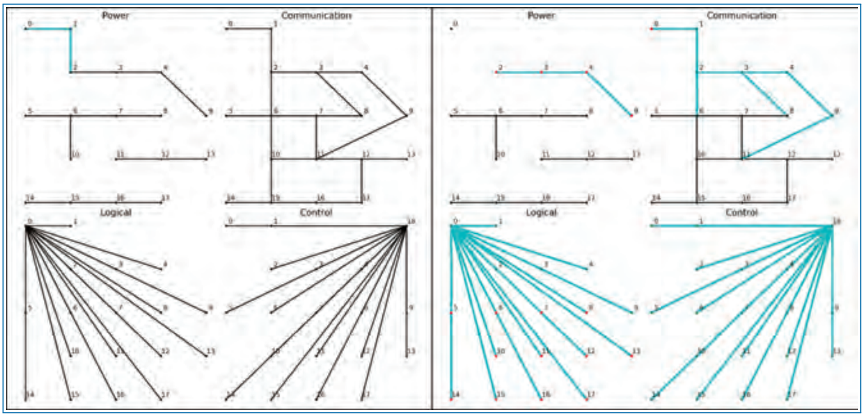


Figure 14.2. Initial and cascading failures of the example use case.

are not controllable, measurements are not available, and targeted reconfiguration is not possible. If the number of controllable loads is smaller than the number of supplied loads, the capability to react to load changes or additional failures is impaired.

14.1.2 Results and Conclusion

The results for a failure at load node 1 are explained in detail to show how the failure is cascading through the system. Initially, as shown in Figure 14.2, load node 1 is failing, causing the edges to the neighboring nodes to fail. Since nodes 2, 3, 4, and 9 are no longer connected to a substation, they are unsupplied and fail. The respective communication nodes are no longer supplied and fail, too, isolating physical communication node 0 and causing it to fail. The cascade proceeds to the logical and the measurement and control networks, causing in both cases the failure of some nodes (including the node that supports or provides control, respectively), then causing all links to fail, and finally also the isolated nodes. In the final state, a large part of the power grid is still supplied, but the controllability has been lost completely and the remaining grid can no longer be monitored.

For a comparison of the different by-design measures, the failure scenario described above has been repeated for each of the nodes in the power grid, representing a single fault happening in different parts of the grid. For each fault, the final state of all networks has been determined via the cascading analysis. The result is shown as boxplots in Figure 14.3, where the box is marking the upper and lower quartile and the orange line marking the median, while the whiskers mark the minimum and maximum.

Even in the default scenario, most of the loads remained supplied no matter where the initial fault happened. However, as presented in the example, the

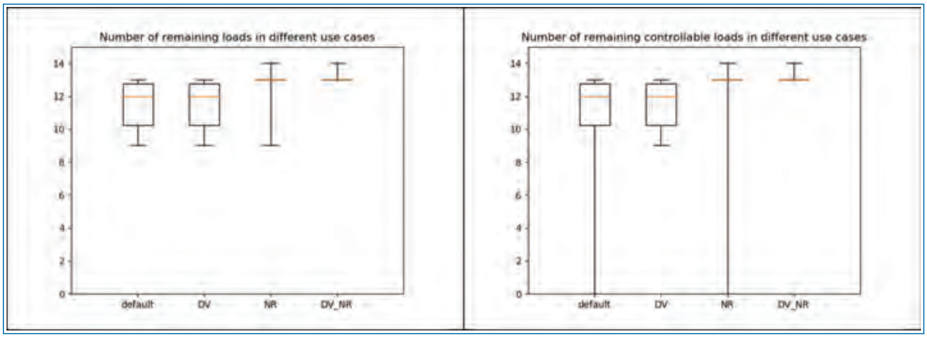


Figure 14.3. Assessment of cascading for different scenarios of the example use case.

capability to monitor and control the remaining part of the power grid is lost completely, if the substation hosting the control center is affected by the initial fault.

The DV configuration manages to solve this issue, enabling the other substations to provide redundancy for supporting the control center functionality. In effect, the remaining part of the power grid remains controllable due to the DV application. Yet, it must be noted that DV does not improve the number of supplied loads. Since a reconfiguration of the power grid is not considered here, it is likely that the initial fault causes additional load nodes of the subgraph to fail. Due to the radial topology of the grid, the closer the fault is to the substation, the more load nodes are failing.

The SR configuration greatly improves this situation and enables more supplied loads in the final state. Due to the meshed topology of the grid that can be utilized for reconfiguration, only the initially failed load is lost in the final state, if the initial fault occurs at a load node. If it occurs at a substation node, there may not be a load node failure at all. However, if the substation hosting the control center functionality fails, the ability to reconfigure the grid is lost and the final state of the system is as in the default configuration.

Only the combination of both by-design measures provides complete containment of a fault independent from its location. If a load node fails initially, only this node is failed in the final state. If a substation node fails initially, the power supply of all load nodes can be maintained. Finally, the capability of monitoring and control of the grid is secured.

The above results show how the impact on the power supply can be minimized effectively by applying by-design measures. While each of the investigated measures improved the resilience of the grid, the combination of both measures provided additional synergies and can avoid the worst case of a failure at the substation hosting the control center. Containing initial failures and reducing cascading to a minimum independent from the location of the initial failure is of increased importance in case of targeted attacks on the most critical components of the system, to stop attackers exploiting vulnerabilities of the system.

14.2 Double Virtualization

Recently, critical infrastructures have been evolving into more complex networks of Cyber-physical Systems (CPS), creating several challenges in the monitoring and controlling of these systems [24]. Double Virtualization (DV) is a specific strategy capable of addressing this, by providing a solution based on the cloud computing paradigm, while enabling a certain degree of decentralization.

For realizing DV, it works on two logical layers: the **Functional Layer** which abstracts the computational resources for management, control, and monitoring functionalities of an asset, and the **Data Layer,** where the logic and features of the deployed applications, such as connectivity and computational operation, are virtually represented. The former offers the remote connectivity, leveraging the control features of the device (while possibly including self-awareness features), while the latter encompasses, in the virtualization process, the set of applications running of the given devices (e.g., logic and configuration for acquiring data, pre-processing, and database query).

By keeping the Functional and the Data Layers decoupled from one another, but acquiring their virtualizations, the DV opens the path on installed devices in a given network to enable real-time reconfiguration and to control running applications and move them from one device to another. This is particularly useful to facilitate the monitoring and control of the Critical Energy Infrastructures (CEI) domain, which comprises a wide variety of dispersed and heterogeneous assets. As matter of fact, in modern power system, the challenge of these types of systems has moved from networking and hardware (such as connection protocols, CPU power and consumption, etc.) to how to connect this amount of different data sources into the specific demands of the hosting platforms and applications. In this context, virtualizations of physical assets—such as the ones offered by DV—and their delivery as services over the network ensure the separation of the functionalities from the specific runtime, protocols, and communication in order to construct highly dynamic, extensible, and flexible environments, as confirmed in [25–28].

The control and monitoring of CEIs deeply rely on the evolvement of the smart grid concept, which incorporates technologies to enhance and provide a better "awareness" of the grid state [29], such as *Advanced Metering Infrastructure* (AMI) or *Wide-Area Monitoring, Protection and Control* (WAMPAC) systems based on *Phasor Measurement Units* (PMU) and *Phasor Data Concentrators* (PDC), aiming at the provision of the guidelines for collecting, transport, and use of data generated on

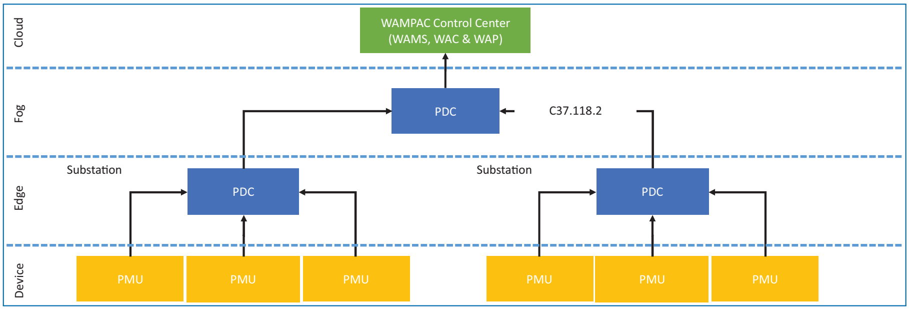


Figure 14.4. Typical WAMPAC system architecture.

the grid. However, these technologies heavily rely on Information & Communication Technologies (ICT), thus exposing the smart grid to a wide range of possible cyberattacks [30].

In this sense, the DV applied to the dedicated computational units of the WAMPAC architecture, as exemplified in Figure 14.4, represents an alternative solution to mitigate cyberattacks that can possibly jeopardize the complete smart grid. To this aim, the DV separates the logical control from the hosting computational hardware into another device and efforts on performing early detection of cyber-physical attacks while enabling mechanisms that provide a continuous operation of the CEI by reallocation of application logic into another asset.

14.2.1 Double Virtualization System Model

For accomplishing DV in a system, it is necessary to adopt the relevant set of assets—DV Assets—with the necessary logic, by either transforming the already existing and/or adding new devices. Additionally, DV demands the inclusion of control, monitoring, and management methodology of these DV Assets, which implies the addition of extra devices in the system—DV Administration & Management (DVA&M). Although in an ideal implementation, DVA&M should be also considered a DV Asset (whose running applications are exclusively for managing and monitoring); in the current implementation, this is not mandatory, since we focus primarily on the already existing devices of the WAMPAC system for demonstrating the concept. Considering all this, and taking the WAMPAC architecture as application model, the system’s transformation is depicted in Figure 14.5.

14.2.2 Double Virtualization Assets

In this type of component, the implemented mechanisms related to the DV provide the ability to gather the necessary information about its resources and functionalities: virtualize the resources/applications running (e.g., bash/Python scripts and

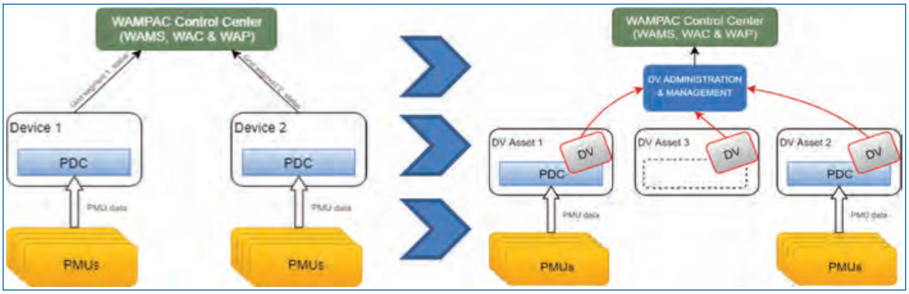


Figure 14.5. Simple grid-monitor model using PMU measurements and equivalent model with double virtualization integration.

their network connections configuration), in a structured data format that may be moved and interpreted on other devices; remote access points that enable the DVA&M devices to access over an internet connection for performing the DV monitoring/administration.

The DV Asset must then be provided, among others, with the following specifications and features:

- **Virtualization:** the DV logic included in the DV Assets must be capable of representing the device and its logical applications in a defined data format.
- **Connectivity:** server and client endpoints must be present, in order to interact with the DVA&M for sending and receiving information, as is the virtualization of the applications or control commands.
- **Monitoring:** the DV Asset must include the necessary services that enable its monitoring by the DVA&M. Furthermore, it may contain self-awareness features that track inner changes that may be also relevant and is able to forward them to the DVA&M.

14.2.3 Double Virtualization Administration and Management

The pivot point of development of the DVA&M component is the ability to execute the monitoring and administration of a set of DV Assets which are connected to it over the network. In this sense, taking in consideration the requirements of the overall system, the chosen approach envisions the use of diverse software patterns, as for example, Service Oriented Architecture (SOA) for addressing communication between devices and Service Orchestration in the optic of management of these same machines, and which was inspired in previously researches, such as presented in [31]. Moreover, the DVA&M is structured taking as base the Observe-Orient-Decide-Act (OODA) pattern, introduced by John Boyd and firstly drafted in [32], in order to achieve the desired logic. Assisted by the OODA loop, the

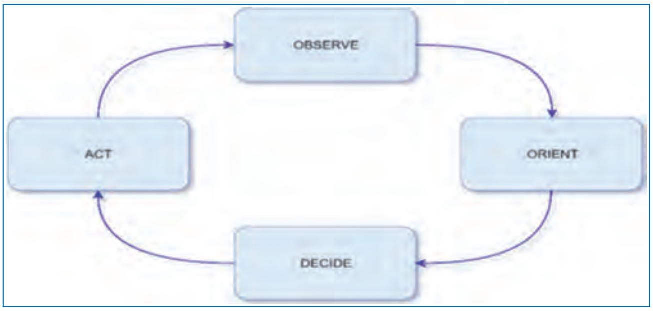


Figure 14.6. Simple OODA loop.

DVA&M entity gathers the abilities to lightly anticipate harmful situations through the continuous monitoring of the several assets behaviors, making use of incoming sensing information, and use it to perform decisions and actions to mitigate detected failures.

The four steps of the OODA loop (Figure 14.6) are described as:

- **Observe:** acquisition of information data, incoming from the detection modules of the system. This information about the DV Assets and can be obtained from internal logic or from an external source.
- **Orient:** in this stage, the received control data is provided with meaning, so that analyses mechanisms can be applied. For example, matching the data to the respective DV Asset and respective previous samples for tracking relevant changes.
- **Decide:** this is where the gathered data is analyzed with the provided algorithms for discovering or handling the detected failures, and furthermore to decide what is the next action. That is to say, if and in what terms the system shall react to the attack detection. This step also provides all output necessary for enforcing the reaction, such as is the case of a migration, where the virtualized logic of an attacked DV Asset needs to be moved into another one.
- **Act:** when this step is activated, it uses all the gathered information to trigger and complete all the mitigation process, while handling the involved DV Assets, in any way possible, through the established connections implemented specifically for administration purposes.

The DVA&M must then be implemented in accordance with the following specifications and features:

- **Database/Registry:** necessary for storage of the relevant information of the DV Assets, like the specifications of the device and the respective virtualizations.

- **Connectivity:** The DVA&M provides the necessary server and client endpoints in order to receive and send information to the other components (DV Assets, external detectors, …).
- **Monitoring:** the DVA&M hosts simple monitoring mechanisms (Acknowledge and Heartbeat/Watchdog techniques) dedicated to the connectivity status of DV Assets, yet it shall also be able to handle incoming information from external detectors and forward it into the decision algorithms.
- **Decision:** the DVA&M must be able to filter the incoming information from the multiple DV Assets and decide whether any action shall be activated, and in that case, handle all the consequent process
- **Mitigation:** the mechanisms to autonomously interact with the faulty DV Assets, while performing the necessary control commands and exchanging the necessary information.

14.3 Example use Case and Implementation

14.3.1 Technological Details

Node-Red framework has been selected as the development and deployment tool for the DV system. Node-Red offers a browser editor for development and deployment and runs over Node.js runtime environment, which stands as one of the prevailing software for development of applications under the Internet of Things (IoT) scope. Moreover, Node-Red applications are constructed on a flow-based semantics, by wiring nodes, and allow an easy creation and setup of computational resources that provide functions, APIs, and online services supported by a wide number of protocols usage. Node-Red also enables the creation and integration of custom nodes (provided by a highly active community), extending its potential for connectivity to, for example, legacy systems.

Another important highlight is the fact that Node.js is supported by a variety of operating systems and processor architectures, such as ARM processors used in single board computers like Raspberry Pi or Odroid. This leverages the cross-platform implementation and widens the number of possible resources to use.

With respect to the interoperability among the DV components in use, the created endpoints that are related to DV functionalities follow the REST pattern, while most of the inherent data is represented in JSON format.

Regarding the security mechanisms, several options are available, including the standard authorization schemes for HTTP. However, while adopting the use of certificates to enable HTTPS for encryption, client certificate authorization, which is built in the HTTPS handshake, was tested and subsequently included, while being optionally customized.

14.3.2 Internal Detectors

The implemented DVA&M comprises built-in detection methods oriented to evaluate connectivity status of the DV Assets. Even if not necessarily the lost connection of a device is caused by some sort of attack, either physical or cyber, it is still plausible to assume it. Moreover, to ensure the resilience of the system, the detection of such failure is used by the DV to trigger the migration of the faulty DV Asset into an available one, as a mitigation strategy.

For both Acknowledge and Heartbeat techniques, the DVA&M has a specified timeout, in which the DV Asset must report to the DVA&M that it is available. The difference is that in the Acknowledge technique, the DV&AM makes a request and the timeout refers to the response time, while in the Heartbeat/Watchdog technique, the DV Asset itself periodically sends acknowledge messages and the timeout is used within the Watchdog. It must be considered that, in both cases, the period of the acknowledge messages and the timeout value must be set so that there is no overlap within the sequence, making it susceptible to induce false failure detections.

14.3.3 External Detectors

Event detection represents the activity of detecting relevant events in (near) real-time from the stream of raw data observations. Most event detection systems are generic, where the user must deploy a set of processing rules at design time, which are used to push observations at run time. The result of the processing is delivered back to the application in form of events.

The event detection engines can be evaluated according to the following categories:

- Development platform, representing the programming language used for event detection applications development
- Event detection language, the operators which can be used to define event extraction rules
- Development model, representing the flexibility used for defining event detection patterns
- Advertised event rate
- Out of the box deployment possibilities
- Integration/compatibility with other technologies
- Licensing.

The network and the communication infrastructure represent an important commodity of an IT system, including the Smart Grid ones. For such a system, it is important to detect as early as possible any attempt of unauthorized access/usage of the network. The network monitoring module has the scope to detect any abnormal behavior of the network.

Regardless of the network assets (routers, switches), the monitoring can be done using the Remote Network Monitoring (RMON) Protocol [40]. The RMON protocol can be used to extract real-time information about the device, such as bandwidth or ports connected/disconnected. Depending of the device type, the processing logic can be embedded into the switch (if it has enough processing power) or a field device (like a Raspberry PI) can be located in the nearby area to execute this activity. In most of the cases, a centralized solution will overload the network. The list of switch operating parameters can include:

- Network utilization (per each port or overall);
- Error Rate;
- Port connectivity.

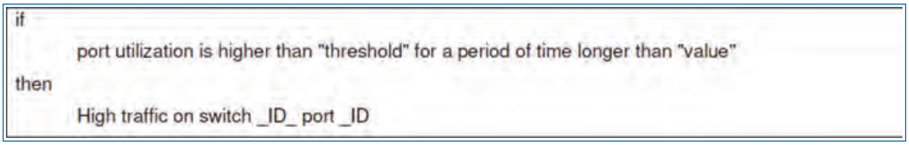
In general, the various components of a system generate log data which is used for monitoring the component status and for debugging. Depending on the architecture, each component can have its own log file or the system can have a centralized logging infrastructure. In most of the cases, when one component is affected by a perturbation, several components might report the abnormal behavior in their log file.

External detectors can be plugged into the DVA&M, by accessing specific endpoints (REST) created for the effect. In the DV case study, this was tested with the log data pattern matcher implemented by SIEMENS, offering the following features:

- Merge multiple log files considering the log event timestamp
- Define domain specific log data patterns (at design time)
- Apply the log data pattern on the streams of log events (at run time).

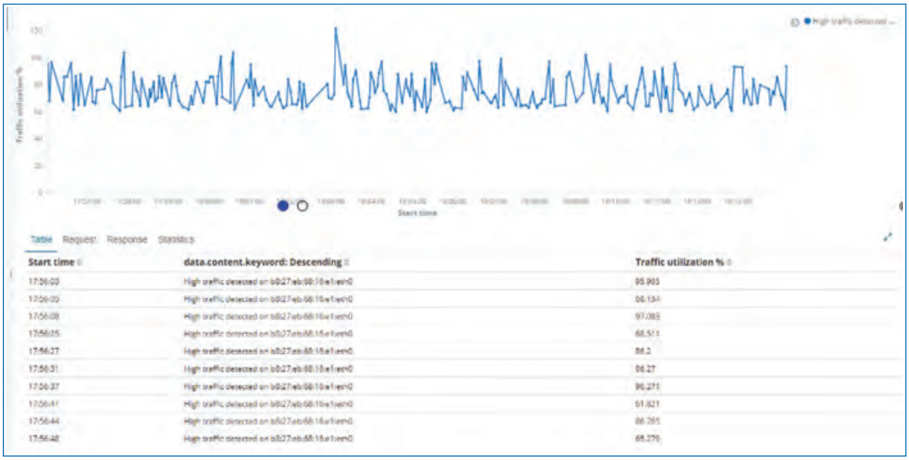
The following example is relevant for high traffic on device interface observation. The logic of observation pattern is depicted in Figure 14.7.

The observed behavior in case of an attack in the context of this use case is presented in Figure 14.8.



```
if
        port utilization is higher than "threshold" for a period of time longer than "value"
then
        High traffic on switch _ID_ port _ID
```

Figure 14.7. Observation pattern logic.



Start time	data.content.keyword: Descending	Traffic utilization %
17:56:03	High traffic detected on b8:27:eb:68:16:e1:eth0	95.965
17:56:05	High traffic detected on b8:27:eb:68:16:e1:eth0	88.134
17:56:08	High traffic detected on b8:27:eb:68:16:e1:eth0	97.083
17:56:25	High traffic detected on b8:27:eb:68:16:e1:eth0	68.511
17:56:27	High traffic detected on b8:27:eb:68:16:e1:eth0	88.2
17:56:31	High traffic detected on b8:27:eb:68:16:e1:eth0	88.27
17:56:37	High traffic detected on b8:27:eb:68:16:e1:eth0	96.271
17:56:41	High traffic detected on b8:27:eb:68:16:e1:eth0	61.821
17:56:44	High traffic detected on b8:27:eb:68:16:e1:eth0	86.765
17:56:48	High traffic detected on b8:27:eb:68:16:e1:eth0	65.270

Figure 14.8. Observed behavior of the log data pattern matcher in case of an attack.

The messages received from external detectors may lead the DVA&M to take some decision and possibly trigger some mitigation action, like trying to reconfigure some DV Asset or perform a migration. In the case of the log data pattern matcher, the approach is to evaluate if a determined number of warnings related to a given DV Asset is received during a time window, thus inducing the DVA&M to activate the mitigation process for the faulty device.

Also, by including this and other detectors that provide a wider panoply of parameters of the DV Assets (e.g., network interfaces traffic information, CPU loads, temperatures of the CPUs, response time of the APIs and services, etc.), the decision algorithms can evolve to more accurate results, like in the case of an occurring migration, where the DVA&M should decide which is the more adequate DV Asset to receive and start running a new set of applications.

14.3.4 Use Case

For demonstrating the DV functionality, the use-case scenario is based on the previously shown WAMPAC system, where PDCs are wired up to PMUs for collecting data. Besides this, a spare development of the PDC consists in hosting small preprocessing algorithms. These PDCs were adopted with the DV logic, and furthermore, its applications were virtualized in compliance with the DV specifications.

Finally, in order to gather the necessary results, a connection failure was induced in one PDC, by unplugging the ethernet cable. When doing this, the DVA&M is able to detect that the unplugged PDC is no longer responding, triggering the migration process, and consequently, the set of application is launched on another

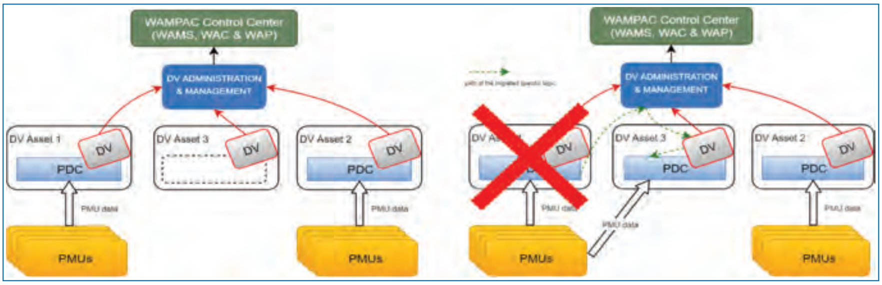


Figure 14.9. System state before and after the detection of lost connectivity of a DV asset and migration strategy applied as mitigation strategy.

PDC which was chosen during the decision algorithm of the DVA&M, as depicted in Figure 14.9.

The context of the use case is as described below.

The Fault Detection Algorithm (FDA) plays an important role in power grid observability and is also the first functionality of a self-healing grid. Depending upon the different grounding schemes of the different grids, the impacts of the fault currents in the grid varies [33]. Furthermore, with the availability of high accuracy and high frequency of measurements from PMUs, advanced FDA schemes, based on PMU data, are being designed [34, 35]. The FDA is deployed in a dedicated hardware that receives continuously the stream of PMU data corresponding to the voltages at different nodes and currents flowing through specific branches in the network. The PMU data is parsed into the FDA after proper protocol translation. The parsed PMU data is then processed by the FDA which detects the occurrence of faults based on the changes in the zero sequence components. Given that the reporting frequency of PMU can be as high as 50 frames per second or more for power networks with nominal frequency of 50 Hz, the fault inception moment can be captured with delay of 20 ms at maximum. Since the timestamp of an event is critical information for correct evaluation of fault location, it is of paramount importance to ensure the uninterrupted operation of FDA. With introduction of DV, the availability of FDA can be substantially increased when fault in communication network or cyberattack occurs, and therefore, the robustness of the fault detection scheme is ensured.

14.3.5 Conclusion

The DV model implemented in the use-case scenarios have, in general, fulfilled the stipulated outcomes in terms of functionality and proof of concept. More specifically, the defined mitigation strategy—migration—was achieved, once a connectivity failure was detected, by completing the transaction of the running application from the faulty device into the best suitable device.

It can also be stated that the main requirements were accomplished: the virtualization process of the assets functional and application layers; communication system for supporting the data transaction inherent to the DV, using REST endpoints; implementation and integration of, respectively, internal and external detectors and corresponding monitoring mechanisms; decision algorithms that can be shaped according to the monitoring parameters in use; and the processing/management of the mitigation actions by either DV Assets and DVA&M components of the system.

In this sense, the DV is a viable solution to augment the resilience of the system. However, taking in consideration that the DV application is still in an early stage, several items that shall be developed and/or improved in the future implementations have been already identified.

For instance, in such scenarios as the Fault Detection described in the previous section, where it is of such crucial importance to minimize the downtime caused by the network failure, some technical choices can be made towards this improvement, such as minimizing the routes of the network connections (number of intermediary routers, not using VPN, etc.), opt for a faster alternative to using client certificate authorization (which takes some time for validation during the HTTPS handshake) and also refining the detection, decision, and action processes of the DV itself.

Another improvement to be considered, for a more proactive solution, is to pre-setup DV Assets with one another's logic, for minimizing the amount of data to be passed to trigger the mitigation and, consequently, the time of the process. Of course, this comes at the cost of more storage and CPU load, but depending on the use case, it may be profitable.

Regarding the critical Single Point of Failure (SPoF) paradigm, the current DV system is not yet completely capable to solve this thematic. In a more close-up glance, it can be noticed that the SPoF was removed from the "functional" area of the system (where DV Assets co-exist); however, the introduction of DVA&M device results in a new SPoF. One possible solution that was put on the table is to create a cluster of DVA&M devices in the system and apply also the DV solution to them, with the respective nuances. For example, in order to avoid the hierarchical structure that induces SPoFs, one can adopt monitoring patterns such as the circular pattern. Furthermore, the Blockchain technology may directly offer a solution for decentralization, but from the performed investigation, we found that the requirements for implementing Blockchain (e.g., high-performance CPUs, big data storage) for very demanding time requirements, this may be a challenge and other instances of Distributed Ledger Technology may be required in these demanding scenarios.

14.4 Service Restoration

The basic functionality of a self-healing power grid is to restore the loads that were de-energized either due to natural disasters or targeted attacks on the grid. There are two kinds of events that create outages in the grid. One that occurs frequently but have lower magnitude of outage, like tripping of lines due to faults in the lines due to ageing of the cables. The other type of events are the ones that have High Impact but occur with Low Probability (HILP events). A HILP event introduces severe and rapidly changing circumstances that may have never been experienced before, causing multiple outages in the network and creating large de-energized sections [1].

A typical Service Restoration (SR) scheme for distribution grids, after successful fault detection and isolation, should be able to perform the following:

- Restore as much out-of-service customers as possible in a minimum time, by providing a sequence of operations to the switches. Preference in use of tele-controlled switches in re-powering process should be given, to reduce the restoration time.
- Consider the priority of the loads and restore the most crucial customers (hospitals, devices controlling the gas network pumps, cellular base stations, and other critical infrastructures) first.
- Preserve radiality of the grid with every switching operation prescribed in the sequence.
- Maintain the voltage of the grid as per the limits imposed in the grid codes of the specific country.
- Satisfy loading constraints of the lines and substation loading.

For the outages caused by the non-HILP events, different approaches are proposed in the literature for optimal selection of the tie switches (normally open switches, generally connecting different feeders or segments of the same feeder) to be closed. These can be classified into expert systems [3–5], Multi-Objective Evolutionary Algorithm (MOEA) [6–8], heuristic-based systems [9–12], meta-heuristics and mathematical programming based [2]. Though the MOEA methods and heuristic methods for SR are popular, they have longer running times and are sensitive to the accuracy of generating the feasible topologies, from which the optimal solution would be deduced. Furthermore, Mixed Integer Programming (MIP)-based methods have also been proposed [13–15]. The mathematical programming methods, especially the MIP based, prove to be computationally expensive.

Hence, they would not be suitable for real-time application with stringent time constraints. One of the major challenges in designing a service restoration scheme to cope with the HILP events is that, in addition to the aforementioned attributes, the SR scheme should also react to rapidly changing system condition. It should be able to consider, in real time, the uncertainties in the power generation and load demand so that possible network congestion is avoided while grid restoration. Furthermore, it should adapt to changes in grid topology as subsequent multiple faults may occur due to the propagation of the HILP events. Unlike the heuristic, meta-heuristic, and mathematical programming-based SR schemes, the rule-based algorithms have been found better suited for real-time applications. They can provide sub-optimal, interim solutions to cater to emergency situations [5], due to their lower computational complexity. Nevertheless, the Rule Based Service Restoration Algorithm (RB-SRA) should also incorporate distribution grid operator preferences in selecting the optimal service restoration option. This is vital as grid operators in different countries have to follow different operational norms and adhere to specific grid codes. Thus, for designing a universal service restoration, the RB-SRA has to be extended with a Multi-Criteria Decision Making (MCDM) algorithm. The MCDM approach facilitates optimal selection of solution from the set of possible solutions considering the dynamic preferences of the decision-maker [16–18]. In this case, the set of solutions are the choice of optimal restoration sequence considering the preferences of the network operator. It should be noted that the MCDM enabled RB-SRA would not be the most optimal service restoration (considering the cost of operation, power losses, etc.) but would be a best effort solution catering to dynamic outages caused by propagation of the HILP events.

14.4.1 MCDM-Enabled Rule-Based Service Restoration Algorithm (RB-SRA)

The MCDM-enabled RB-SRA has 4 major sequential steps, Namely:

(1) Identification of loads to be restored;
(2) Determination of alternative reconfigurable paths;
(3) Network security assessment with state estimation;
(4) MCDM-based selection of optimal restoration path.

Identification of loads to be restored

When multiple faults occur or the continuous load growth is not promptly combined with substation reinforcement, the reconnection of all de-energized loads cannot be achieved [36]. Hence, it is necessary to identify the most critical load and quickly restore it. The RB-SRA selects, among the de-energized nodes that are outside the fault-zone, the one with the highest priority index. This parameter is an independent characteristic of each load, assigned by the grid operator to indicate its criticality. If multiple loads have the same index, the algorithm selects the one consuming (or generating) the highest active power. The chosen node is taken as target for the restoration plan.

Determination of alternative reconfigurable paths

Once the load to be restored, named $\boldsymbol{b}$, has been identified as described above, the best reconfiguration topology to re-energize it has to be computed. Firstly, the proposed algorithm inspects all the $\boldsymbol{n}$ primary substations present in the network. For each substation, it determines the most suitable path towards the load by using Dijkstra's algorithm. Considering two nodes $\boldsymbol{a}$ and $\boldsymbol{b}$ in a weighted graph G, Dijkstra's algorithm calculates the shortest path that connects them, named $G'_{a,b}$. In this case, $\boldsymbol{a}$ is the selected substation, $\boldsymbol{b}$ is the load to be restored, and the weight of each edge is the series impedance of the line. Hence, considering $|\overline{Z}_{x,y}|$ as the magnitude of complex series impedance $|\overline{Z}_{x,y}| = R_{x,y} + jX_{x,y}$ between adjacent nodes $\boldsymbol{x}$ and $\boldsymbol{y}$ of the graph, the shortest path $G'_{a,b}$ has a total impedance Z_a^b such that $Z_a^b = \min \sum_{x,y \in G'_{a,b}} |\overline{Z}_{x,y}|$. With respect to total impedances of other paths that can connect $\boldsymbol{a}$ to $\boldsymbol{b}$, Z_a^b has the minimum value for graph $G'_{a,b}$. If the shortest path exists, it includes at least one bus tie which is currently open and, by closing it, allows to energize load $\boldsymbol{b}$ from substation $\boldsymbol{a}$. With the switches now closed and the path made electrically continuous, the whole network topology has changed, represented by the graph $G_{a,b}$ for which $G'_{a,b} \subset G_{a,b}$ ($G'_{a,b}$ is a subset of $G_{a,b}$). This procedure is repeated for each substation present in the grid. If multiple faults occur or the continuous load growth is not promptly combined with substation reinforcement, the reconnection of all de-energized loads could not be achieved [36].

Network security assessment with state estimation

Each network configuration proposed by the RB-SRA is to be checked versus line congestion and voltage security limits via State Estimation (SE) [37]. Many methods are available for SE but among them the Weighted Least Square (WLS) approach is most popular [37, 38]. It is based on the minimization of the square of the measurement residual vector. With input as the set of measurements, the uncertainty class of the measurement devices, network topology and its parameters, the WLS based SE is able to provide the estimate of the state of the grid that may be magnitude and angle of node voltage (for node voltage-based SE) or the magnitude and angle of the branch current (for branch current-based SE). Furthermore, from the estimated states all the power flows in the grid, loading of the network lines, and the power losses can be calculated.

For each proposed grid topology, represented with graph $G_{a,b}$, the following constraints should be respected:

- Radiality of the network: if a path exists between nodes a and b, the closing of the tie switches in this restoration scheme must maintain each substation electrically disconnected from the others.
- Voltage limits: at each node of the grid, the voltage magnitude must remain in the range of ±10% of the nominal value [39].
- Respect of loading limits: the current flowing in each edge must comply with the cable/conductor or substation transformer specification $|\overline{I}_{x,y}| \pm 3\mu_{|\overline{I}_{x,y}|} < I_{max_{x,y}}$.

Where $|\overline{I}_{x,y}|$ is the line current magnitude at the generic edge $(\boldsymbol{x}, \boldsymbol{y})$ and its uncertainty $\mu_{|\overline{I}_{x,y}|}$; $I_{max_{x,y}}$ indicates the continuous current carrying capacity of the line or the overcurrent limit of the transformer at the primary substation, and in emergency situation, a certain percentage of overloading is acceptable for limited amount of time. If the proposed configuration $G_{a,b}$ does not fulfill the requirements, it is discarded and the restoration of load $\boldsymbol{b}$ cannot be achieved by the substation $\boldsymbol{a}$.

MCDM-based selection of optimal restoration path

Once the set of secure reconfiguration paths has been determined, the optimal solution has to be identified. To do this, the proposed algorithm combines two criteria dependent upon settings predefined by the user, namely the power losses and the utilization of the lines, as described below.

- *Total power losses in the network* $(P_{x,y})$: The power loss $P_{x,y}$, between two generic nodes $\boldsymbol{x}$ and $\boldsymbol{y}$, is estimated by the following formula, using the estimated line-to-ground node voltages by the SE algorithm and the electrical lines are modeled as the π equivalent circuit.

$$P_{x,y} = 3\,\Re\left[\overline{V}_x\left(\overline{V}_x\frac{G_+ + jB_+}{2}\right) + \overline{V}_y\left(\overline{V}_y\frac{G_+ + jB_+}{2}\right) + \left((\overline{V}_x - \overline{V}_y)\left(\frac{(\overline{V}_x - \overline{V}_y)}{R_+ + jX_+}\right)\right)\right]$$

Where R_+, X_+, G_+, B_+ are the positive sequence line resistance, reactance, conductance, and susceptance, respectively. The power losses are added for each line of the network (edges of graph $G_{a,b}$) to obtain the total power loss P^a of the candidate topology restored through substation $\boldsymbol{a}$.

- *The utilization of electrical lines* ($\theta_{x,y}$): It is a measure of overloading of the lines in the grid. The different service restoration options can be ranked on the basis of their relative network loading. The higher the value of $\theta_{x,y}$, the better is the distribution of power flow in the specific network configuration

$$\theta_{x,y} = \frac{I_{max_{x,y}} - |\overline{I}_{x,y}|}{I_{max_{x,y}}} \tag{14.1}$$

Where $\boldsymbol{x}$, $\boldsymbol{y}$ are two nodes between which the current $\overline{I}_{x,y}$ flows and the line connecting the nodes $\boldsymbol{x}$ and $\boldsymbol{y}$ has the maximum current carrying capacity of $I_{max_{x,y}}$. For each network topology that is analyzed, three minimum values of $\theta_{x,y}$ are recorded in descending order. In the case of graph $G_{a,b}$, they are indicated as θ_1^a (the minimum one), θ_2^a & θ_3^a, for which θ_1^a is related to the electrical line having the current most close to its specific ampacity. The selection of the optimal solution requires the combination of these two aspects, summarized as the four criteria $P^a, \theta_1^a, \theta_2^a, \theta_3^a$. Then, using the MCDM technique, the optimal restoration path is selected. The MCDM technique is a two-step algorithm. In the first step depending upon the relative pairwise weight of the criteria, an absolute weight for each criterion is deduced. In the second step, these weights are used to determine the relative closeness of the available solution to the ideal solution. For the first step, the Analytical Hierarchical Process (AHP) is implemented to determine the absolute weights of criteria with the pairwise weights of the criteria. The pairwise weights are assumed to be provided by the network operator. The comparison matrix Γ is calculated using the pairwise weights provided by the operator [22].

$$\Gamma = \begin{bmatrix} 1 & \omega_{P\theta_1} & \omega_{P\theta_2} & \omega_{P3} \\ 1/\omega_{P\theta_1} & 1 & \omega_{\theta_1\theta_2} & \omega_{\theta_1\theta_3} \\ 1/\omega_{P\theta_2} & 1/\omega_{\theta_1\theta_2} & 1 & \omega_{\theta_2\theta_3} \\ 1/\omega_{P\theta_3} & 1/\omega_{\theta_1\theta_3} & 1/\omega_{\theta_2\theta_3} & 1 \end{bmatrix} \tag{14.2}$$

In which ω is the comparison value between the attributes indicated by the subscripts, which ranges from 1/9 (attribute of second subscript is extremely important with respect to the first one) to 9 (attribute of first subscript is extremely important with respect to second one) according to the AHP scale. A detailed possible AHP weights and their interpretation is provided in Table 14.1. The subscripts

Table 14.1. AHP weight interpretation.

Intensity of Importance (**$\omega_{P\theta}$**)	1	3	5	7	9	2, 4, 6, 8
Interpretation	P, θ are of equal importance	P is slightly more important than θ	P is strongly more important than θ	P is very strongly more important than θ	P is extremely more important than θ	Intermediate importance between two adjacent judgment

$P^a, \theta_1^a, \theta_2^a, \theta_3^a$ represent the power losses and the line utilization of the three most consumed lines, respectively. The priority vector is obtained, which ranks the four criteria and shows relative weights among them. The approximate calculation of the priority vector (PV) can be done as shown in equations:

$$PV_j = \frac{\sum_{l=1}^{m} \overline{p}_{jl}}{m} \quad \text{where } \overline{p}_{jk} = \frac{\Gamma(j,k)}{\sum_{l=1}^{m} \Gamma(l,k)} \text{ and } m\text{: number of criteria} \tag{14.3}$$

Then, these relative weights are combined with the power losses and line utilizations of each feasible solutions, indicated by the different values of a as reference substation, according to the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) [23]. A generic set of equations that govern the TOPSIS-based ranking of the alternatives are given below. For $\boldsymbol{t}$ alternatives and $\boldsymbol{m}$ number of criteria, the decision matrix $\boldsymbol{D}$ can be created as shown below in Equation (14.4):

$$D = \begin{bmatrix} d_{11} & \dots & d_{1m} \\ \vdots & \dots & \vdots \\ d_{t1} & \dots & d_{tm} \end{bmatrix} \tag{14.4}$$

The weighted normalized decision matrix integrating the Priority Vector calculated in Equation (14.3) can then be derived as in Equation (14.5):

$$v_{ij} = PV_j \, r_{ij} \quad \text{where } r_{ij} = \frac{d_{ij}}{\sqrt{\sum_{i=1}^{m} d_{ij}^2}} \tag{14.5}$$

The next step in the TOPSIS-based ranking is to calculate the positive ideal solution ($\boldsymbol{A}^+$) and the negative ideal solution ($\boldsymbol{A}^-$) where B is a set of Benefit criteria and C is a set of Cost criteria:

$$A^+ = \{v_1^+, \dots v_m^+\} \quad \text{where } v_j^+ = \{\max(v_{ij}) \text{ if } j \in B; \min(v_{ij}) \text{ if } j \in C\} \tag{14.6}$$

$$A^- = \{v_1^-, \dots v_m^-\} \quad \text{where } v_j^- = \{\min(v_{ij}) \text{ if } j \in B; \max(v_{ij}) \text{ if } j \in C\} \tag{14.7}$$

The purpose to calculate the $\boldsymbol{A}^+$ and $\boldsymbol{A}^-$ is to measure the distance of the alternatives from the positive ideal solution and negative ideal solution. The best alternative would be the one that is as close to the positive ideal solution and as far as from the negative ideal solution. In order to rank the alternatives, the relative closeness is then calculated as in Equation (14.8):

$$C_t^+ = \frac{S_t^+}{S_t^+ + S_t^-} \quad \text{where } S_t^+ = \sqrt{\sum_{i=1}^{m} (v_{ij} - v_j^+)^2} \quad \text{and}$$

$$S_t^- = \sqrt{\sum_{i=1}^{m} (v_{ij} - v_j^-)^2} \tag{14.8}$$

For the service restoration process, the TOPSIS allows us to compute the closeness of each candidate solution to the ideal one, which is composed by the minimum power loss P^a and maximum reserve line capacity $\theta_1^a, \theta_2^a, \theta_3^a$. Among the possible solutions, indicated by the different values of $\boldsymbol{a}$ as reference substation, the one having the highest closeness $C_a{}^+$ is chosen to reconnect the load $\boldsymbol{b}$. Then, a closing signal is sent to the open tie switch related to this configuration.

The selected optimal solution is, then, a trade-off between the minimization of power losses in the line and the avoidance of lines having the current close to its limit. The comparison parameters ω are defined by the grid operator before the service restoration is started; they are set depending on whether more importance is assigned to power losses or line utilization aspect. For example, in case of aged cables, one can place greater emphasis on line utilization (by decreasing $\omega_{P\theta_1}$, $\omega_{P\theta_2}$ and $\omega_{P\theta_3}$) in order to avoid the excessive worsening of line condition.

Test case

The test grid used to validate the MCDM-based RB-SRA is a medium voltage distribution grid at 13.8 kV with four primary substations. Figure 14.3 represents its single line diagram and includes the elements naming used below [20]. Its 37 nodes connect loads that range from 100 kW to 1 MW; the length of the electrical lines varies from 800 m to 2200 m. Nodes I2 and L1 host DERs of 200 kW and 250 kW, respectively. Complete data of the grid can be found in [21]. The black

Table 14.2. AHP pairwise priorities.

Pair of criteria	$\omega_{P\theta_1}$	$\omega_{P\theta_2}$	$\omega_{P\theta_3}$	$\omega_{\theta_1\theta_2}$	$\omega_{\theta_1\theta_3}$	$\omega_{\theta_2\theta_3}$
AHP pairwise weights	7	8	9	2	2	2

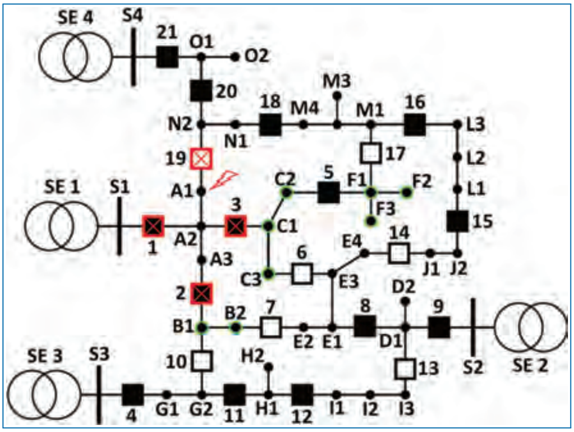


Figure 14.10. Location of the fault for test case.

squares indicate the normally closed switches, whereas the normally open ones are shown with white squares.

This scenario inspects the occurrence of a single fault and the presence of multiple loads having the same priority index, for which the nominal active power is considered to determine the restoration target. In this test case, the most important criterion of the service restoration is the minimization of the power losses, with marginal relevance of lines utilization in the decision process. The comparison parameters ω are reported in Table 14.2.

The single electrical fault occurs at node A1; the protection system opens the upstream circuit breaker indicated by number 1 and, in order to isolate the fault area, the downstream circuit breakers 2, 3, and 19 (which is already open). These four switches, indicated with red frames in Figure 14.10, remain in a tripped condition and cannot be reclosed until the fault has been repaired.

The MCDM-enabled RB-SRA receives the status of the tripped breakers, and it first identifies the faulty zone; hence, it excludes the nodes A1, A2, and A3 from the restoration process.

The de-energized loads are downstream of the switches 2 and 3; they are marked with green circles in Figure 14.8, whereas their priority indexes and nominal active power are reported in Table 14.3. Both the loads B2 and C2 have the highest priority index; since the nominal active power of B2 is higher, it is selected as target for the restoration process.

Table 14.3. De-energized Loads and their critical index.

Loads	B1	B2	C1	C2	C3	F1	F2	F3
Priority Index	3	1	4	1	2	4	3	4
Active Power [MW]	0.37	0.70	0.27	0.46	0.17	0.27	0.22	0.35

Table 14.4. Computational performance of MCDM-enabled RB-SRA.

Test Case	Test Case 1
Loads	B2 C2
Min (Seconds)	2.15 2.97
Average (Seconds)	2.29 3.16
Max (Seconds)	2.97 3.84

In the next step, the algorithm evaluates the possible reconfiguration paths associated to each substation. The substation SE 1 is not suitable, since the switches in the fault zone cannot be operated. Moreover, SE 4 is excluded too, because the radial topology cannot be maintained (SE 4 and SE 2 would be electrically connected). The power loss related to SE 2 is 4.5% larger (corresponding to 30.7 kW) than SE 3, making SE 3 solution the closest to the ideal one ($C^{+}_{SE2} = 0$ and $C^{+}_{SE3} = 1$). Hence, the closing command to switch 10 is sent. Once the database updates the switch status and the closing command is sent to field device, the SR algorithm restarts; the loads B1 and B2 are now energized by SE 3; hence, the algorithm evaluates the restoration of the loads in the branch downstream of switch 3, selecting the node C2 as first target. Only SE 2 or SE 4 could restore the selected load (and, consequently, all the nodes in the same branch) by maintaining the radial structure. The restoration is achieved by closing the switch 6, for which P^{SE2} is smaller of 36% than P^{SE4} making the closenesses to ideal solution $C^{+}_{SE2} = 0.82$ and $C^{+}_{SE4} = 0.18$. All the de-energized loads outside the fault zone are reconnected; then, the algorithm is concluded. The algorithm always checks for the real-time switch position data before it closes the tie switch to reenergize the loads, by doing so it detects if a subsequent fault had occurred that triggered other switches/circuit breaker position. If it detects a change, then it stops the current operation and reruns the complete MCDM-enabled RB-SRA for the new topology of the grid and lost load configuration. This functionality helps in handling multiple sequential failures introduced by the HILP events.

The performance of the MCDM-enabled RB-SRA is tabulated in Table 14.4. The restoration of grid for the same fault locations have been performed 100 times for a stochastic evaluation.

Conclusion

Self-healing or automated service restoration of the power grids is one of the important functionalities of resilient smart grids. With increasing frequency of occurrence of natural disasters and targeted cyber-physical attacks on the power grids, the automated service restoration becomes a vital functionality of the electrical energy infrastructure. Furthermore, with higher dependence on the Information and Communication Technology (ICT) infrastructure for the operation of the power grid, the automatic service restoration becomes a complex problem as the ICT infrastructure might also fail due to power network failures thus jeopardizing the automatic service restoration procedure. Therefore, in addition to the aforementioned algorithm for emergency service restoration, mechanisms should also be included to optimally restore the grid considering the availability of ICT infrastructure. The proposed methodology enables to restore the grid considering the criticality of the load and the preferences of the network operator for single and multiple faults. However, it should also be extended to also optimize the life of each switch or circuit breaker thus making sure that switches and breakers are not stressed by extremely high number of switching made during the restoration process over a period of time.

14.5 Conclusion of the Chapter

In the first part of this chapter, we have shown that applying by-design measures, the resilience of CEI can be increased. Utilizing additional redundancies and enabling the physical and communication and control networks to autonomously adapt in case of incidents enables the infrastructure to self-heal and to either minimize the impact or recover from it. While applying by-design measures individually can already improve the service level, combining different by-design measures (including different domains as power, communication and control) provides synergies since this approach recognizes the interdependencies between different domains of CEI.

Double Virtualization has been introduced as by-design measure to avoid single points of failure in the functional layer of grid monitoring and control, represented by centralized monitoring and control functionalities that are dependent on a specific device. Virtualizing these functionalities enables utilization of redundancies provided by the availability of various devices in the infrastructure that are able to host respective functionalities. DV has been applied to Fault Detection Algorithm as example use case to showcase the principle. Restrictions and recommendation for applying DV as well as the potential for future improvements of the principle have been given in the conclusions.

Service Restoration has been introduced as by-design measure to reconfigure the electrical grid after occurrence of one or multiple faults. It has been described how a multi-criteria approach can be implemented with the help of MCDM and TOPSIS, enabling distribution system operators to configure the reconfiguration strategy based on predefined priorities assigned to a set of criteria as restoration of power supply to critical loads as well as complying to voltage limitations of the grid. The principle has been demonstrated based on test cases covering faults in test grid that represents a part of a distribution system. Based on relevant criteria, the priorities for restoring the grid after faults can be refined and a suitable reconfiguration strategy can be derived.

Although the comparison of cascading effects in a test system indicates that resilience can be increased by applying by-design measures as DV and SR, requirements of a specific CEI must be considered before implementing the proposed or other by-design measures. Applicability of DV, for example, might depend on requirements of the functionality, which is to be virtualized. Applicability of SR might depend on the capability of the electrical equipment to perform flexible reconfiguration as well as regulatory constraints for system operation.

References

[1] M. Panteli and P. Mancarella. The grid: Stronger, bigger, smarter? Presenting a conceptual framework of power system resilience. IEEE Power and Energy Magazine, 13(3):58–66.

[2] C. Liu, S.J. Lee, and S.S. Venkata. An expert system operational aid for restoration and loss reduction of distribution systems. IEEE Transactions on Power Systems, 3(2):619–626, May 1988.

[3] C.-S. Chen, C.-H. Lin, and H.-Y. Tsai. A rule-based expert system with colored petri net models for distribution system service restoration. IEEE Transactions on Power Systems, 17(4):1073–1080, Nov. 2002.

[4] T. Ananthapadmanabha, A.D. Kulakarni, A.S.G. Rao, J.G. Char, K.R. Rao, and K. Parthasarathy. Knowledge-based methodology for intelligent sequence switching, fault identification and service restoration of distribution system. International Journal of Electrical Power & Energy Systems, 19(2):119–124, 1997.

[5] Y. Kumar, B. Das, and J. Sharma. Multiobjective, multiconstraint service restoration of electric power distribution system with priority customers. IEEE Transactions on Power Delivery, 23(1):261–270, Jan. 2008.

[6] R. Srinivasa Rao, S. V. L. Narasimham, M. Ramalinga Raju, and A. Srinivasa Rao. Optimal network reconfiguration of large-scale distribution system using harmony search algorithm. IEEE Transactions on Power Systems, 26(3):1080–1088, Aug. 2011.

[7] D.S. Sanches, J.B.A. London, A.C.B. Delbem, R.S. Prado, F.G. Guimaraes, O.M. Neto, and T.W. de Lima. Multiobjective evolutionary algorithm with a discrete differential mutation operator developed for service restoration in distribution systems. International Journal of Electrical Power & Energy Systems, 62:700–711, 2014.

[8] S. K. Goswami and S. K. Basu. A new algorithm for the reconfiguration of distribution feeders for loss minimization. IEEE Transactions on Power Delivery, 7(3):1484–1491, July 1992.

[9] S. Dimitrijevic and N. Rajakovic. Service restoration of distribution networks considering switching operation costs and actual status of the switching equipment. IEEE Transactions on Smart Grid, 6(3):1227–1232, May 2015.

[10] M. Gholami, J. Moshtagh, and L. Rashidi. Service restoration for unbalanced distribution networks using a combination two heuristic methods. International Journal of Electrical Power & Energy Systems, 67:222–229, 2015.

[11] M. E. Baran and F. F. Wu. Network reconfiguration in distribution systems for loss reduction and load balancing. IEEE Transactions on Power Delivery, 4(2):1401–1407, April 1989.

[12] C. Lee, C. Liu, S. Mehrotra, and Z. Bie. Robust distribution network reconfiguration. IEEE Transactions on Smart Grid, 6(2):836–842, March 2015.

[13] B. Chen, C. Chen, J. Wang, and K. L. Butler-Purry. Sequential service restoration for unbalanced distribution systems and microgrids. IEEE Transactions on Power Systems, 33(2):1507–1520, March 2018.

[14] T.T. Borges, S. Carneiro, P.A.N. Garcia, and J.L.R. Pereira. A new opf based distribution system restoration method. International Journal of Electrical Power & Energy Systems, 80:297–305, 2016.

[15] N.R.M. Fontenele, L.S. Melo, R.P.S. Leao, and R.F. Sampaio. Application of multi-objective evolutionary algorithms in automatic restoration of radial power distribution systems. In 2016 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS), pages 33–40, May 2016.

[16] A. Kumar, B. Sah, A. R. Singh, Y. Deng, X. He, P. Kumar, and R.C. Bansal. A review of multi criteria decision making (MCDM) towards sustainable renewable energy development, Renewable and Sustainable Energy Reviews, vol. 69, 2017.

[17] R. R. Yager, "Modeling prioritized multicriteria decision making," in IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 34, no. 6, pp. 2396–2404, Dec. 2004.

[18] P. Espie, G. W. Ault, G. M. Burt and J. R. McDonald, "Multiple criteria decision making techniques applied to electricity distribution system planning," in IEE Proceedings – Generation, Transmission and Distribution, vol. 150, no. 5, pp. 527–535, 15 Sept. 2003.

[19] A. Angioni, A. Kulmala, D.D. Giustina, M. Mirz, A. Mutanen, A. Dede', F. Ponci, L. Shengye, G. Massa, A. Repo S. Monti, (2017, April). Design and implementation of a substation automation unit. IEEE Transactions on Power Delivery, vol. 32, no. 2, pp. 1133–1142.

[20] N. R. M. Fontenele, L. S. Melo, R. P. S. Leao, and R. F. Sampaio. Application of multi-objective evolutionary algorithms in automatic restoration of radial power distribution systems. In 2016 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS), pages 33–40, May 2016.

[21] A. Dognini and A. Sadu. Networkdata_MV_FLISR_use_case.pdf. https://www.fein-aachen.org/en/projects/rbosr/ accessed on 11th September 2019.

[22] A. Jaiswal and R. B. Mishra. "Cloud Service Selection Using TOPSIS and Fuzzy TOPSIS with AHP and ANP." In: Proceedings of the 2017 International Conference on Machine Learning and Soft Computing. ICMLSC '17. Ho Chi Minh City, Vietnam: ACM, 2017, pp. 136–142. isbn: 978-1-4503-4828-7. doi: 10.1145/3036290.3036312. url: http://doi.acm.org/10.1145/3036290.3036312.

[23] G.H. Tzeng and J.J. Huang. *Multiple Attribute Decision Making: Methods and Applications*. A Chapman & Hall book. Taylor & Francis.

[24] Q. Zhu and T. Başar, "Robust and resilient control design for cyber-physical systems with an application to power systems," in 2011 50th IEEE Conference on Decision and Control and European Control Conference, 2011, pp. 4066–4071.

[25] A. Rocha *et al.*, "An agent based framework to support plug and produce," in 2014 12th IEEE International Conference on Industrial Informatics (INDIN), 2014, pp. 504–510.

[26] G. Di Orio, A. Rocha, L. Ribeiro, and J. Barata, "The PRIME Semantic Language: Plug and Produce in Standard- based Manufacturing Production Systems," presented at The International Conference on Flexible Automation and Intelligent Manufacturing (FAIM2015), Wolverhampton, UK, 23–26 June 2015, 2015.

[27] J. Soldatos, S. Gusmeroli, P. Malo, and G. Di Orio, "Internet of Things Applications in Future Manufacturing," in *Digitising Industry – Internet of Things Connecting the Physical, Digital and Virtual Worlds*, River Publishers, 2016.

[28] G. D. Orio, P. Maló, J. Barata, M. Albano, and L. L. Ferreira, "Towards a Framework for Interoperable and Interconnected CPS-populated Systems for Proactive Maintenance," in 2018 IEEE 16th International Conference on Industrial Informatics (INDIN), 2018, pp. 146–151.

[29] A. Ashok, A. Hahn, and M. Govindarasu, "Cyber-physical security of Wide-Area Monitoring, Protection and Control in a smart grid environment," *J. Adv. Res.*, vol. 5, no. 4, pp. 481–489, Jul. 2014.

[30] S. Majumder, A. Mathur, and A. Y. Javaid, "Cyber-Physical System Security Controls: A Review," in Cyber-Physical Systems: Architecture, Security and Application, S. Guo and D. Zeng, Eds. Cham: Springer International Publishing, 2019, pp. 187–240.

[31] G. Brito, G. Di Orio, and J. Barata, "Orchestrating loosely coupled and distributed components for product/process servitization," presented at the 2017 IEEE 15th International Conference on Industrial Informatics (INDIN), 2017, pp. 1199–1204.

[32] J. R. Boyd, "The essence of winning and losing," *Unpubl. Lect. Notes*, vol. 12, no. 23, pp. 123–125, 1996.

[33] A. Zidan *et al.*, "Fault Detection, Isolation, and Service Restoration in Distribution Systems: State-of-the-Art and Future Trends," in IEEE Transactions on Smart Grid, vol. 8, no. 5, pp. 2170–2185, Sept. 2017.

[34] S. Barman and B. K. S. Roy, "Detection and location of faults in large transmission networks using minimum number of phasor measurement units," in IET Generation, Transmission & Distribution, vol. 12, no. 8, pp. 1941–1950, 30 4 2018.

[35] M. A. Ebrahim, F. Wadie and M. A. Abd-Allah, "Integrated fault detection algorithm for transmission, distribution, and microgrid networks," in IET Energy Systems Integration, vol. 1, no. 2, pp. 104–113, 6 2019.

[36] James Northcote-Green and Robert Wilson. *Control and Automation of Electrical Power Distribution Systems*. CRC Press, 2007.

[37] A. Abur and A.G. Exposito. *Power System State Estimation: Theory and Implementation*. CRC Press, March 2004.

[38] R.F. Stengel. *Optimal Control and Estimation*. Dover Books on Mathematics. Dover Publications, 2012.

[39] EN 50160:2010. Voltage characteristics of electricity supplied by public electricity networks. Standard, CENELEC, July 2010.

[40] S. Waldbusser, C. Kalbfleisch, D. Romascanu, "Introduction to the Remote Monitoring (RMON) Family of MIB Modules," Network Working Group, Standard, https://tools.ietf.org/html/rfc3577

[41] van Steen, Maarten: Graph theory and complex networks. An introduction. 2010.

DOI: 10.1561/9781680836875.ch15

Chapter 15

Securing CEI "By-Innovation"

By Cédric Buron, Simon Fossier, Edward-Benedict Brodie of Brodie, Antonello Corsi, Artemis Voulkidis and Nikos Drosos

15.1 Introduction

In this chapter, CEI Security solutions for situation awareness are described. We show how the adoption of innovative processes and leading-edge technologies can improve evaluation of threats scenarios in an automated manner. We also show how it can lead to evaluate the occurrence of potential attacks and to apply proper countermeasures to mitigate them while not sacrificing operational task for security.

In DEFENDER, the *situational awareness* is an iterative process; it starts with the detection of *simple events* performed by the cyber-physical detectors at the sensor level. The latter are then correlated, according to a time/space dimension, to provide *complex events* that represent the CEI's *state of the environment*.

Subsequently, the complex events are compared with a model—the *attack tree model*—that formalizes the structure of well-known cyber-physical attacks. It allows to compute and simulate the probability on future complex events, based on the current knowledge—complex event that have already occurred—and the common knowledge—the attack trees. This step is performed by the DEFENDER Co-simulator. Its output, consisting of the set of the most likely attack subtrees, is then evaluated looking at the impact of each of these subtrees with regard to various dimensions of the situation, such as economical and societal, so as to have a complete view and a full understanding of what is occurring in the CEI assets (*situation comprehension*). Finally, after calculating a risk score for each attack subtree, the most appropriate *countermeasures* or *mitigations* (*i.e.,* measures able to stop the attack without severe damages) to the ongoing attack is determined and suggested to the end user through a human-friendly interface for the security decision-making. In DEFENDER, these tasks have been assigned to the CEI *Incident Detection Framework* (IDF), the CEI *Incident Mitigation Framework* (IMF) and the *Security Control Center* (SCC).

In order to formalize the notion of threat scenario, the DEFENDER consortium relies on a high-level description of the attacks—*attack trees*—that have already been presented in Section 12.4. This representation has the advantage to preserve genericity in the downstream analyses, which is quite important in a project like DEFENDER that must be applicable to any kind of CEI. In this chapter, we show how these trees can be used to deal with incoming attacks. It is first necessary to enrich the trees with attributes representing the link between the attack tree node and the CEI assets targeted by the threat. This process allows to quantify the impact of a malicious action based on the importance of the underlined asset. For example, it can represent the difference between a risky situation where a drone fly over a secondary electrical substation and a less risky one where the same drone flying over a general CEI perimeter zone.

In addition to the nodes to CEI assets association, attack trees can also be enriched by adding other elements among which: the response of the system after a successful attack; the consequences of the attack; alternative relationship between alternative root nodes, and/or system response. Before going through the best strategy evaluation process, it is worth describing the information flow through IDF, IMF, and SCC.

Figure 15.1 describes the processing path in and around the detection and mitigation tasks. In this description, the DEFENDER Co-simulator essentially plays the role of event probability estimator for the threat evaluation module—which propagates them as impacts and risks to the aggregation and mitigation modules. The mitigation module output is then pushed to the Security Control Center, where interactions with the end user take place. The next subsections of this chapter successively describe the detection task (Section 15.2), the mitigation task (Section 15.3), and the human-centric Control Center (Section 15.4).

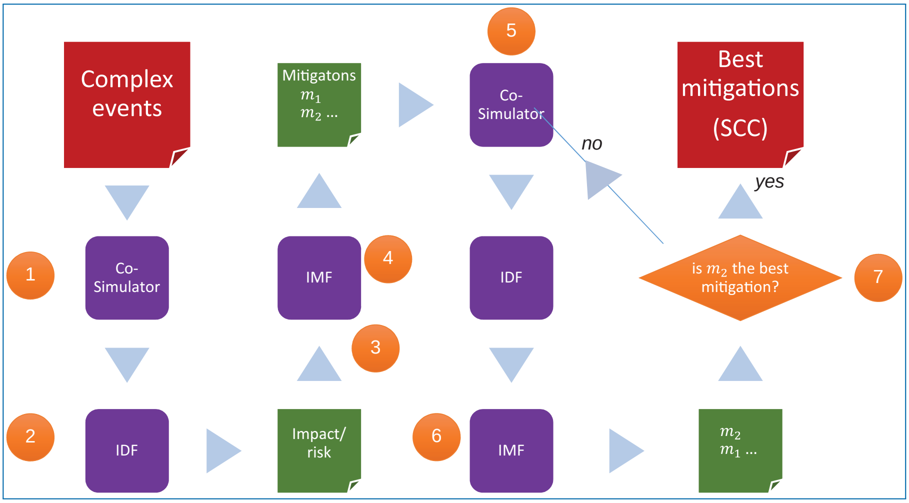


Figure 15.1. Information flow and processing in the situation awareness modules.

15.2 Detecting Ongoing Incidents

With the rapid growth and the increasing complexity of new digital objects, CEI are subject to more evolved cyber-physical threats. For this reason, identifying and preventing attack against critical infrastructure is getting more and more strategic, and it is difficult to ensure an adequate degree of protection from the multiple sources of attacks. In this situation, numerous procedures are put in place to at least lower the risk of an attack when it is not directly possible to mitigate immediately any attack on the infrastructure normal operation mode.

For this reason, we approached the problem about detection of ongoing threats, previously mentioned, with the modeling provided by the attack graph model developed in DEFENDER that illustrates possible multi-stage attacks in a CEI network, typically by presenting the logical causality relations among multiple exploitable vulnerabilities and infrastructure.

To understand the difficulty of this kind of detection, it is important to underline the concept of detection as a process that lasts in time, instead of detection as a punctual event.

In fact, as reported in Figure 15.2, if an attacker compromises the asset by gaining access to p_1, they have to fulfill at least one other condition (p_2 or p_3), in order to move towards the attack and to cause severe damage. Suppose we want to use this piece of information in real-time security analysis. When we suspect the p_1 node has been compromised, with how much confidence can we say that the information related to $p4$ have been compromised? The appropriate response is far less sure than

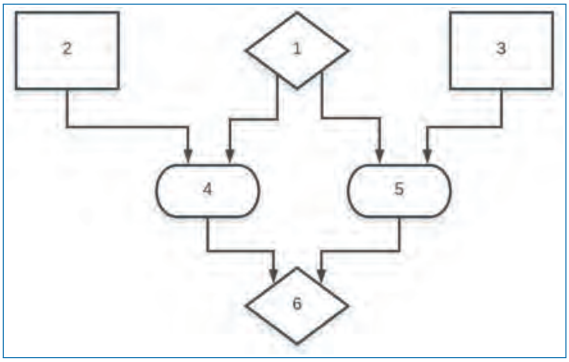


Figure 15.2. General attack tree with 6 nodes $\{p_1, p_2, p_3, p_4, p_5, p_6\}$.

the analysis of the attack tree can give. How might we know whether the attacker has decided to launch this assault? Regardless of whether he did as such, how might we realize that the assault will proceed by steps ahead in time?

We do not know the attacker’s choices once that one node has been violated; thus, there is the uncertainty from unknown attacker behaviors. The defender’s observations on potential attack activities are limited, at least with a finite amount of resources to be used in cybersecurity; and as a result, we have the uncertainty from false positives and false negatives of *intrusion detection system* (IDS) sensors.

For this reason, it is useful to equip the different nodes of the attack tree with a list of metrics that can allow us to rank, by risk, the different evolutions in time of the attack path once that a node has been detected.

The first information needed is then the probability that a node will be exploited across the tree in a certain direction, and this is the main functionality developed by the Co-Simulator component of the DEFENDER platform. The second information is the one provided by the *Incident Detection Framework*, and it consists in a given metric related to the domain under attack/protection. In the DEFENDER platform, we used the impact measure to provide to the mitigation part a list of possible threats ordered by means of risk, defined as the product of probability by impact.

Attack graphs produced by DEFENDER and related threat models allow us to specify the different ways that a CEI can be compromised. For a given attacker, in the attack tree, the root represents the goal of the attacker and the leaves depict initial entry points for the attacker that are necessary to achieve their goals. It is also useful to underline the fact that in DEFENDER, each node can be associated to a given asset (in the case of a smart grid, we can have a heat pump, a wind-turbine, a smart meter, or a perimeter fence that protect the CEI premises), and therefore, the impact calculated exploiting its vulnerabilities can change depending on the asset.

This DEFENDER IDF can be further detailed describing the different modules that are in charge of deploying the main functionalities for threat detection. The IDF is directly connected with the Co-simulator, and it receives from it attack pattern subtree with confidence interval (probability). The result of the incident processing is to assess the impact of each prognosed subtrees and evaluate the related risk based on the results of threat analysis available in literature. The IDF provides to the mitigation part a ranked list based on a set of predefined rules enabling the alert on incidents. The IDF's main elements are the following:

- **Monitor and analytic modules:** these two functions implement the base functionalities of a data analytics tool that allows for multi-criteria analysis and detection of spatio-temporal patterns playing a very important role for the threats identification and related countermeasure selection. In case the IDF is used to quantify the impact of possible time evolution of an attack tree, these functionalities are deployed in two modules that receive information from the Co-simulator and move it, as is, to the semantically enhanced threat module that is able to evaluate the impact of the forecasted part of the attack tree providing further analysis.
- **The Semantically Enhanced Threat Extraction Tool** is a component capable to receive from the monitor and analytic modules a list of attack path—portion of attack tree—with related probability and has to assess, for each of this attack vector, an impact based on the loss estimation and the related risk evaluation. This component will provide the risk evaluation of each forecast attack path to be shown in the threat identifier, leaving to the mitigation component the decision of the mitigation action to be taken. It is based also on historical information contained in the threat model and risk repository that in turn is populated with the detected threats associated with impact values. The loss estimation value assesses the losses when an attack actually occurs. For instance, a secondary electric substation is running many energy related infrastructure services for a Distribution System Operator (DSO). When it is unavailable for a day, the company will have a severe impact because the business services cannot be provided any more. At this point, estimating the loss may include the hourly revenue of business, loss of data, and implementing a mitigation plan. The estimation of loss in DEFENDER is quantified as one monetary value which should include all types of losses. The DEFENDER approach for evaluation of techniques can be based on theoretical assumption to produce loss estimates using theoretical aspects such as statistical distributions. The risk related to the evaluated impact then follows naturally from the product of the probability of an attack path with the cumulative impact calculated across the chosen attacker's path.

Table 15.1. Metrics for information propagation.

	AND	OR
Probability	$\prod_{i=1}^{n} prob_i$	$1 - \prod_{i=1}^{n} (1 - prob_i)$
Cost	$\sum_{i}^{n} cost_i$	$\frac{\sum_{i=1}^{n} prob_i X cost_i}{\sum_{i=1}^{n} impact}$
Impact	$\frac{10^n - \prod_{i=1}^{n} (10 - impact_i)}{10^{(n-1)}}$	$MAX_{I=1}^{n} impact_i$

Table 15.2. KPI libraries.

Name	Domain
Mean Time to Breach	Time
Mean Time to Recovery	Time
Mean Time to First Failure	Time
Number of trigger complex event	Time
Confidential impact	Impact
Integrity Impact	Impact
Availability impact	Impact

To perform this calculation, the metrics in Table 15.1 are used to propagate the information along the attack tree [1].

The KPI libraries in Table 15.2 contain the models of security metrics that allow the monitor of attack vectors and provide information for the characterization of the forecasted time evolution of the threats linked to the detected complex event. These metrics are based on the work done in [2] even though adapted to serve the CEI context.

The KPI libraries enable the "scorecard" functionality containing the set of general metrics and their definitions. The metrics are general characteristics that we deemed relevant to the IDF. The metrics have been divided into two classes: *Time* and *Impact*. The former allows evaluation based on the number of events that happen in time and the latter are fitted for allow the multidimensional impact evaluation.

15.3 Mitigating Ongoing Incidents

After the detection, the system should determine a suitable set of mitigations or countermeasures able to stop the attack. However, more criteria should be taken into account than simple efficiency—for instance, the countermeasure implementation cost. Humans are notoriously bad at making decisions with several criteria on a rational basis, even with expertise. *Multi-criteria Decision Aid* [3] (MCDA) is a framework designed to support a *decision-maker* (DM) in choosing rationally among several alternatives, when there is no obvious best choice that would satisfy all criteria and it boils down to user preference. MCDA relies on two elements: a *preference model*, which it represents how alternatives compare to each other, and a *process*, which defines how the DM applies their knowledge to the model.

The impact of each mitigation, both from material, immaterial, and human perspective, must be taken into account, and each set of mitigations must be evaluated to assess its adequacy. In this section, we provide the description of a tool well suited to evaluate automatically the sets of mitigations under consideration based on the knowledge of a *Subject Matter Expert* (SME) [4].

A classical way of solving the MCDA problem is to rely on MCDA trees [5]. MCDA trees are used to disaggregate and analyze the DM preferences along a set of criteria. In turn, these criteria can be disaggregated, leading to subcriteria, until reaching atomic criteria corresponding to measurements. The structure of the decision tree defines how the DM decomposes his decision. Each internal node of the tree corresponds to an aggregation function depending on the lower branches. There are plenty of aggregation functions. Some are very simple; the most common one is the *weighted sum*. This function is very simple but cannot take into account the complexity of some situations, where the criteria interact with each other. More complex functions can do so—it is the case of the *Choquet integral* [6]—but they also require more complex methods to be calibrated. The MCDA approach must therefore be divided in two parts: a disaggregation phase and an aggregation phase.

During the disaggregation phase, the DM is first interviewed in order to structure the information and build a decision tree; and for each criterion, partial utility functions are built. Then comes the calibration of the aggregation functions for each of the aggregating nodes. To do so, the DM is asked her preferences among several alternatives. The overall methodology to calibrate the partial utility functions and the aggregation functions is called MACBETH [7] and has proven reliable for various study cases [5, 8].

Finally, the second part of the process is the aggregation part, which consists in applying the model obtained in the first part to alternatives in order to score them.

The Choquet integral

The Choquet integral [9] is a much more complete function and can express interactions, such as a veto. It is also monotonous, which means that if a first alternative is better than a second one *on all criteria*, then the Choquet integral of the first alternative is higher than the second. The Choquet integral is a function built on so-called capacities, a set of functions with properties described in Definition 15.1.

Definition 15.1 (Capacity). A *capacity* (also called *fuzzy measure*) on $N = \{1, \dots, n\}$ is a set function $\mu : 2^N \rightarrow [0, 1]$ such that $\mu(\phi) = 0, \mu(N) = 1$ (boundary conditions) and $\forall A \subseteq B \subseteq N, \mu(A) = \mu(B)$ (monotonicity)

The Choquet integral is both symmetric and additive:

Definition 15.2 (Symmetric and additive capacities). A capacity is said to be additive if $\mu(A \cup B) = \mu(A) + \mu(B)$ for every pair (A, B) of disjoint coalitions.

A capacity is said to be symmetric if $\mu(A)$ depends only on $|A|$, the cardinality of A.

Finally, the Choquet integral itself is defined as:

Definition 15.3 (Choquet Integral.). The Choquet integral of $a = (a_1, \dots, a_n) \in R^n$ defined with regard to a capacity μ can be expressed as:

$$C_\mu(a_1, \dots, a_n) = \sum_{(i=1)}^{n} a_\sigma(i) \times [\mu(\{\sigma(i), \dots, \sigma(n)\}) - \mu(\{\sigma(i+1), \dots, \sigma(n)\})]$$

where σ is a permutation on N such that $a_{\sigma(1)} \leq a_{\sigma(2)} \leq \dots \leq a_{\sigma(n)}$, and $\mu(\{\sigma(i+1), \dots, \sigma(n)\}) = 0$ when $i = n$.

Disaggregation phase

The first phase of the MCDA process is to disaggregate the preferences of the DM, in order to produce the decision tree. This part can be divided into three:

- The structuring stage, used to determine the structure of the tree, i.e., how the criteria relate to each other;
- The calibration of the partial utility function, used to get partial utility functions from the universes on which the measures are made;
- The calibration of the aggregation function in the aggregation nodes.

Note that all the stages of disaggregation systematically involve the DM, who can express her expertise on the topic in her interview. Ultimately, the decision tree is composed of:

- The *universe nodes*: attributes from which the decision is made, and leaves of the MCDA tree. These values are not bounded by the model and depend on what is measured. It may be a qualitative or a quantitative value.

- The *utility nodes*, present directly above each universe node: associates a partial utility in [0, 1] to the measures of universe nodes.
- The *aggregation nodes*, used to aggregate criteria—these criteria being themselves either utility nodes or aggregation nodes. In particular, the root of a MCDA tree is an aggregation node which represents the overall utility of an alternative.

The second stage of the disaggregation phase deals with the calibration of the partial utility function. To do that, we can rely on the MACBETH approach [6, 7]. This approach can be divided into three steps: defining perfectly satisfactory $\mathbb{1}$ and inacceptable level $\mathbb{0}$ on the attribute universe, define key values in the universe, and finally define the difference between consecutive key values using an interval scale. When the utility function is monotonous, the totally satisfactory value in the attribute universe defines the point from which getting a "better" value will not increase the utility of the DM. More formally, let u be the partial utility function of an attribute for which the utility is increasing. Then: $\forall x \leq \mathbb{0},\ u(x) = 0$, and $\forall$ x $\geq \mathbb{1}$ $u(x) = 1$. The case of a decreasing partial utility function is similar.

It is then necessary to define several key values. They represent inflection points of the utility function. For the DM, it represents points between which she can define the intensity of the difference and the levels of difference between consecutive points—taking, e.g., the values *very weak, weak, mean, strong, very strong, extreme*.

Finally, the aggregation functions must be set. Here again, we use the MACBETH methodology to set the value of the coefficient associated to each subcriterion and the one associated to each interaction. First step: determine a set of option that the DM is able to order, which must be sufficient to determine all the coefficients. A good way to get such a set is to provide to the DM the options for which all the attributes are set to either 0 or 1, order these options, and ask the DM to provide information on the level of preference between the options.

Aggregation phase

The second phase of the algorithm is the aggregation phase. During this phase, the model designed in the disaggregation phase is applied to one or more alternatives. This alternative must be formalized on the aforementioned attributes. Note that the system can support and adapt to undefined values among the attributes. In this case, the user must define whether the undefined values correspond to non-applicable attributes or to missing values. Another element is computed during this phase. As the 2-additive Choquet integral is more complete than the weighted sum, it may also be more difficult to interpret what criteria have more influence on the result. Moreover, this value depends on the values of the criteria. During the aggregation phase, a value is computed that represents the sensitivity of the result with regard to each of the criteria. This sensitivity index indicates the change on an aggregation node when one criterion is modified, and the other ones are kept at the same value.

Feedback on DM interviews

In theory, the tree structure must reflect the concerns of the end users. However, it had proven complicated for an operator to directly provide the tree structure. In order to help them, some proxies can be used. For instance, providing a sample tree with several criteria/metrics and aggregations, then asking them to extend these using the criteria they may use in the eventuality of an attack. This can typically lead to them suggesting new nodes in the tree, but also the deletion of less relevant nodes.

Computation of mitigation impact

Mitigation relevance is typically computed using the notions of attack probability and attack impact; however, these values must be balanced with by the impact of the mitigation itself. This is related to the fact that, for an attack with a very low impact, there may be a mitigation that is very effective, but at the same time has a very high impact, possibly even worse than the initial attack. In this case, it would be better either to choose a mitigation with lower impact or even do nothing in some situations. The impact of the mitigation is computed considering:

- The human resources (i.e., engagement of security teams). This criterion is in turn divided into two subcriteria, physical security and cybersecurity. The corresponding metrics are computed in percentages. For trial sites without physical security team, this criterion is replaced by a police intervention;
- The material resources that correspond to the usage of tools either provided in the context of the project or owned by the end users themselves;
- The non-functional infrastructure, involving network infrastructure, access to the offices, etc. It is subdivided into two subcriteria: the criticality of the impacted non-functional infrastructure and the disruption time.
- The outage, also computed on two subcriteria: first, the amplitude of the outage, which can either be computed in terms of number of impacted customers or in terms of missed gains, in currency. Second, the duration of the outage importance of non-relevant criteria is set to 0, so that it has no influence on the computations.

Computation of the overall mitigation score

The root of the tree is the global utility ("grade") of the mitigation. It is this value that will be considered when deciding whether the mitigation should be proposed to the end user. In addition, with the attack risks and probabilities, the decision must take into account the mitigation impact (see above) and its *efficiency*.

This efficiency corresponds to the reduction of risk of the attack once the mitigation has been applied.

The whole computation therefore stems from a co-computation with the Co-simulator and the IDF, before sending the result to the SCC as represented on Figure 15.1:

1. Complex events are transmitted to the Co-Simulator. It computes attack probability and time for possible attacks paths;
2. The Incident Detection Framework receives the attack paths and computes the corresponding attack impacts;
3. The impacts and probability are sent to the Incident Mitigation Framework;
4. The Incident Mitigation Framework computes the relevant mitigations, making the supposition that the attack is blocked by the mitigation;
5. The most promising mitigations are sent back to the Co-Simulator so that it computes the new risk level if this mitigation was implemented. Note that the most probable attack may have changed: if the initially planned attack is blocked by the mitigation, the attackers may try another attack path;
6. The Incident Detection Framework computes the new impact;
7. The new impact is sent back to the Incident Mitigation Framework. If one of the computed mitigations is still optimal in the Incident Mitigation Framework, the process stops and the most promising mitigations that have been evaluated twice are sent to the Security Control Center. If not, the most promising mitigations not yet evaluated twice are sent back to the Co-Simulator.

   The process ends anyway if all the mitigations have been evaluated twice.

Explaining the model

When provided with the possible mitigations, the DM can have the following needs:

- *Interpretability*: what are the most important attributes on average?
- *Explicability*: why is the preference higher for this mitigation option than the other? why has the relevance of a given mitigation option significantly increased over last minutes?
- *Sensitivity Analysis*: what changes in the attributes would most increase the relevance?

An index that achieves this [10] is an extension of the Shapley and Owen values (defined in Cooperative Game Theory) on trees. Note that the use of the values has recently gained interest for interpretability in Classification [11, 12]. Formally, the goal is to construct an easily understandable indicator $I_i(x, y, T, U)$

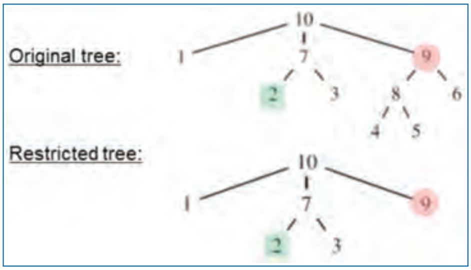


Figure 15.3. Example of a decision tree with transformation.

that measures the influence of factor i, when comparing two options x, y, in a preference model with criteria represented in a tree T, quantified by a utility model U.

We want to impose certain properties on the index (the indices refer to Figure 15.3):

- *Restricted Value*: $\boldsymbol{I_i}$ depends only on combinations of $\boldsymbol{x}$ and $\boldsymbol{y}$.
- *Consistency with Restricted Game*: $\boldsymbol{I_2}$ is equal for the original tree and for the derived subtree where **9** becomes a leaf.
- *Null Attribute*: If changing $\boldsymbol{x_i}$ to $\boldsymbol{y_i}$ never changes $\boldsymbol{U}$, then $\boldsymbol{I_i = 0}$
- *Additivity*: $\boldsymbol{I_i(U + U') = I_i(U) + I_i(U')}$
- *Restricted Equal Treatment*: All attributes are treated symmetrically.
- *Generalized Efficiency*: $\boldsymbol{I_{10} = U(y) - U(x)}$ and, e.g., $\boldsymbol{I_9 = I_6 + I_8}$.

These properties ensure a good "behavior" of the explanation, such as stability towards a slight change to the underlying attributes. Furthermore, note that we can demonstrate that there exists only one influence index fulfilling these properties.

This index computation has exponential complexity, like the Shapley value, making its computation intractable, even for small numbers of criteria. However, taking profit of symmetries among permutations can drastically speed-up the computation time. This idea was proposed by Owen for his value [13] and is extendable to any tree. First, the hierarchical structure of the tree makes some interactions between criteria null by design, which reduces the number of permutations required to compute the index. Then, using the Consistency with Restricted Game property allows to "trim" the tree dynamically when computing the index. The computational complexity is still exponential but on a much smaller tree.

15.4 Control Center Tools for Better Incident Monitoring

Assisting the operation of the CEI owners and operators, DEFENDER offers a rich graphical user interface served in the form of a control center web dashboard

(Security Control Center—SCC) that allows the users to get useful information on the security status of the managed CEI and quickly act to mitigate an identified, previously uncovered vulnerability, a security event or a validated security breach (attack). Notably, SCC acts as an integration point for the majority of the DEFENDER components and processes, including the IMF and the *Human In The Loop* (HITL) framework, detailed in the following paragraphs. In short, the key features of the SCC are listed below:

- Ability to visually overview of the CEI security state in real time as well as in historical terms;
- Ability to overview a countermeasure strategy as proposed by the IMF component, including the valuated risk mitigation impact as described in economic, human, and social terms and, if needed, execute it;
- Ability to register HITL targets in coordination with the HITL backend infrastructure and the blockchain, view their messages, and respond with further notices;
- Integrate with the DEFENDER Pan-European I2SP in order to send information over local CEI attacks and countermeasures and retrieve information regarding pan-European attacks against CEI. The DEFENDER Pan-European I2SP is a global monitoring entity responsible for analyzing attack patterns at European level and deducing coordinated or, possibly, cascaded attacks;
- Allow SCC operators to configure the CEI data sources, elements, and available mitigation mechanisms.

In the following, a quick presentation of the HITL concept and framework is given, followed by the SCC architecture and core interactions.

15.4.1 Building a Culture of Security: The Human in the Loop Framework

The final aim of DEFENDER is to build a culture of CEI security in a controllable manner, starting from small teams of first responders and, via the well-defined communication strategy of DEFENDER, eventually considering public masses. These teams of first responders will be gradually integrated into the DEFENDER CEI security framework in the form of virtual security sensors, effectively materializing crowd sensing and giving rise to the DEFENDER HITL concept. According to the latter, people living in the vicinity of CEI will be empowered to providing feedback (free or structured text, photographs, and video) from the CEI locations. This will be made possible via (a) trained employees implementing the best security policies tailored to CEI operational environment and (b) social networking and trusted information exchange between volunteers, while preserving by design the privacy and the security of the citizens involved in the process.

In the context of DEFENDER, HITL holds a special role since it constitutes an invaluable source of information that actively engages real human interaction with the DEFENDER system. Granted this remark, the security and privacy requirements are explicitly hardened; the real identity of the humans is absolutely protected, and all data exchanges are traceable to prevent intended or unintended misuses. At the same time, the information flows are bidirectional since the HITL targets are able to send plain text messages, images, and videos to the CEI operators and Law Enforcement Agencies (LEAs), whereas the latter are also able to either send commands as a response to the initial HITL targets reportings.

In the context of DEFENDER, the HITL targets are equipped with an Android messaging application that enables end-to-end encryption and that is tightly integrated with the Ethereum blockchain infrastructure so that:

- All message exchanges are kept encrypted at all times;
- The digital identity of the users is validated and managed by the BC infrastructure and is guaranteed to be decoupled by the physical one;
- All large files are stored in the IPFS distributed filesystem, encrypted;
- All information exchanges are permanent, immutable, and traceable;
- No individual's special categories of personal data are processed (GDPR art. 9);
- Messages containing text, images, and video collected through the HITL app will be deleted on individuals' devices immediately after sending them to the CI system.

To satisfy the stringent privacy requirements of the overall HITL concept, the DEFENDER HITL solution bases its functionality on the orchestrated operation of an Ethereum Blockchain to store basic information and perform user identity management, a distributed filesystem for storing larger files (text, photographs, and videos) as well as secured Application Programming Interfaces (APIs) to facilitate interactions at application level.

Notably, the DEFENDER HITL framework may be seen as both a monitoring and an actuation component, depending on the communication flow direction, Humans → CEI standing for the former and CEI → HITL corresponding to the second one. Figure 15.4, below, depicts the framework flow regarding the unidirectional human-to-CEI communication.

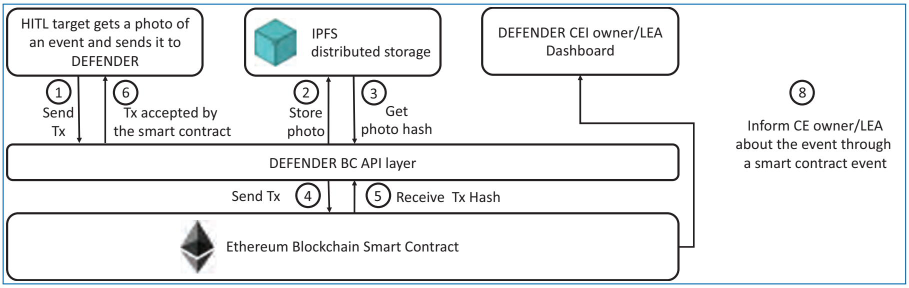

HITL target gets a photo of an event and sends it to DEFENDER
IPFS distributed storage
DEFENDER CEI owner/LEA Dashboard
1 Send Tx
6 Tx accepted by the smart contract
2 Store photo
3 Get photo hash
8 Inform CE owner/LEA about the event through a smart contract event
DEFENDER BC API layer
Send Tx 4
5 Receive Tx Hash
Ethereum Blockchain Smart Contract

Figure 15.4. Overview of the bidirectional human-to-CEI communication.

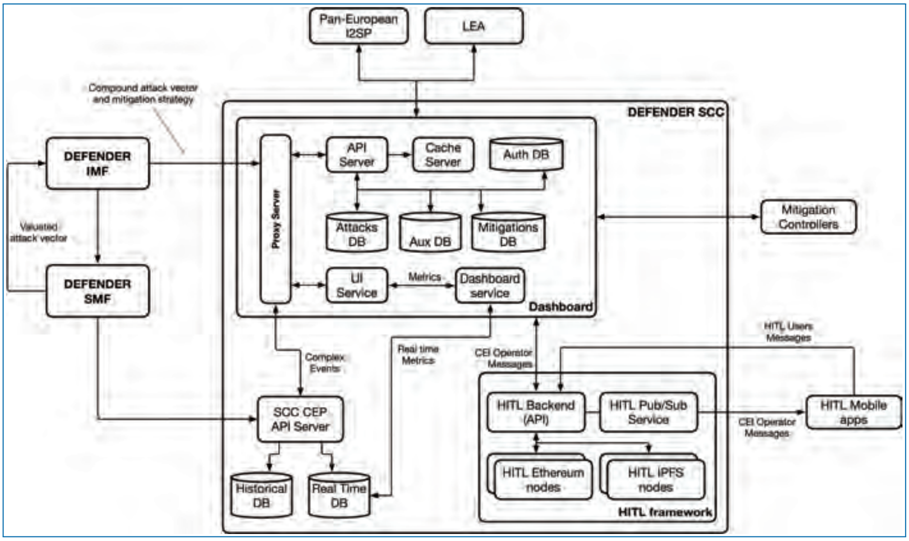


Figure 15.5. DEFENDER SCC high-level architecture and interactions.

15.4.2 Control Center Architecture and Interface with other Modules

An integration point for the vast majority of DEFENDER operations, the DEFENDER SCC is a complete system featuring multiple sub-components, orchestrated in a distributed manner, so that possibly failing functionalities do not affect the operational status of the rest of the sub-components; effectively, the DEFENDER SCC has been designed featuring a microservices-oriented architecture, as may be seen in Figure 15.5, below.

The following table overviews the SCC *dashboard-composing entities* along with their functionality.

From a process perspective, whenever a compound attack vector and mitigation strategy is being sent to the SCC core by the IMF, it gets passed via a reverse proxy server (mask the various functionalities of the SCC) to the SCC API server and later via the cache server. In this context, the SCC acts as server and the IMF as client. In turn, the API server notifies the UI service using a web socket service of the new attack and the mitigation strategy proposed by the IMF are presented to the CEI operator. Next, the CEI operator selects whether to activate the countermeasures or not. In any case, regardless of whether the CEI operator chooses to activate the countermeasure or not, an event is being sent to a DEFENDER Pan-European platform instance for further processing under a more general perspective.[1]

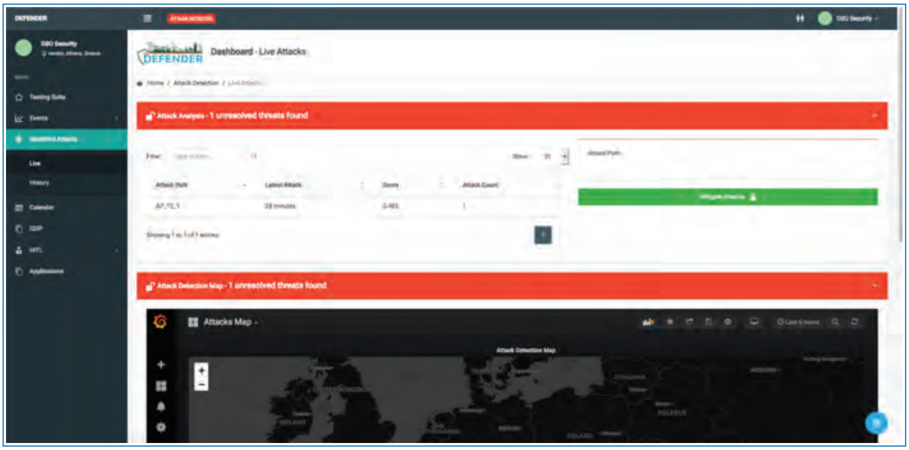
The following figures briefly present the functionalities identified in Table 15.3. In particular, Figure 15.6 depicts the SCC view when a new attack gets detected and mitigations are notified by the IMF, whereas Figure 15.7 highlights a view of the DEFENDER SCC HITL integration.

Finally, the DEFENDER SCC can act as a comprehensive toolbox offering *Decision Support Systems* (DSS) services to CEI operators. Coupled with the

Table 15.3. Core services of the DEFENDER SCC.

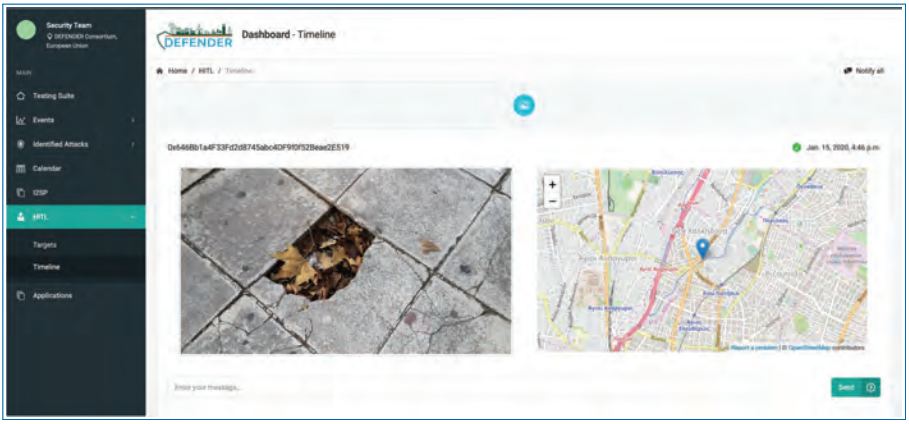
Component	**Cause**
Proxy Server	Consolidates the various services access to reduce the open web ports of the system and facilitate integration
API Server	The core API service handling REST requests from the various components of the system
Cache Server	In-memory database used for caching parts of the UI and API services for speeding up access to commonly used resources
UI Service	Core dashboard user interface-implementing component
Dashboard Service	Implements mapping and charting functionality for overviewing the security state of the CEI as well as for displaying the complex events received from the SMF F3 interface
SCC CEP API Server	The core API service handling the F3 requests, validating and pushing them to the underlying database structures
HITL Backend	The API exposed to the HITL Android devices, responsible for retrieving the HITL reports
HITL Ethereum nodes	The blockchain node(s) of the DEFENDER instance
HITL IPFS nodes	The distributed storage node(s) of the DEFENDER instance
HITL Pub/Sub	A publish/subscribe mechanism, enabling the CEI-to-human unidirectional communication flow

[1] In fact, this information is used to uncover attacks at pan-European level, or, possibly, attacks that could cascade to other infrastructures as well.



Attack Path	Latest Attack	Score	Attack Count
AP_72_1	28 minutes	0.483	1

Figure 15.6. DEFENDER SCC view highlighting an identified attack.



0x646Bb1a4F33Fd2d8745abc4DF9f0f52Beae2E519

Figure 15.7. Integration of SCC with HITL.

sophisticated mitigation strategy extraction services exposed by the IMF and the trusted communication flows offered by HITL, the SCC can radically change the way CEI operators perceive the security of their infrastructures, allowing them not only to better monitor them but also optimally secure them against a large variety of cyber-physical threats.

15.5 Conclusion

15.5.1 The Big Picture

Innovation is a key element in the protection of Critical Energy Infrastructures. Going from a model of an ongoing attack to the explanation of all the elements, including the incident detection and its mitigation, is absolutely necessary, especially given the criticality of the infrastructure dealt with. In this chapter, we presented a methodology to handle these incidents. The first step is to evaluate the ongoing attack, which relies on two key components: the use of data analytics to determine the possible time evolution of an attack within an attack tree; and the proposition of a semantically enhanced threat extraction tool, able to provide the risk evaluation of each forecast attack path, based on historical information of the threat model and risk repository. The second component of our approach is an innovative incident mitigation framework that is able, based on the detection module, to score mitigations against detected attack and provide explanation about the selected ones. This module is based on a solid mathematical framework, Choquet integral, and information provided by subject matter experts through a formal methodology. The result of these two elements is finally provided to the control center. We presented the capabilities of the latter: visualization of the situation in real time and historical terms, interaction with the user for countermeasures strategy, configuration and information—the whole operation being secured through blockchain technology.

15.5.2 Extending the Architecture to Other Settings

The architecture presented here is aimed at situations where the possible attack paths are known—but in some cases, it may not be the case. In particular, in a cybersecurity setting, the security loopholes may not be identified. In this case, building an attack tree as we did may not be possible. Some other methods can then be used to detect incoming events, for instance, machine learning. These attacks, and all the formalism related to them, are being investigated in other research projects, such as the H2020 PHOENIX (832989) project. Though the setting of that project is different, some of the elements presented in this chapter can be reused. It is, for instance, the case of the evaluation of the mitigation through an MCDA approach, as well as the technical elements of the Security Control Center, which make it possible for the operator to visualize the situation and the ongoing attack. Some requirements also stay identical, for instance, the need to provide explicability to the end users. Only minor changes are therefore necessary to utilize and adapt the elements of this chapter to this type of new settings.

References

[1] Kenneth S. Edge, George C. Dalton, Richard A. Raines and Robert F. Mills, “Using attack and protection trees to analyze threats and defenses to homeland security. In Military Communications Conference, 2006. MILCOM 2006. IEEE (pp. 1–7). IEEE,” *Military Communications Conference*, 2006.

[2] N. Idika and B. Bhargava, "Extending attack graph-based security metrics and aggregating their application," *IEEE Transactions on Dependable and Secure Computing*, pp. 75–85, 2012.

[3] J. Figueira, S. Greco and M. Ehrgott, Multiple criteria decision analysis: state of the art surveys, vol. 78, Springer Science & Business Media, 2005.

[4] M. D. Teter, "Applying subject matter expertise (SME) elicitation techniques to TRAC studies," 2014.

[5] C. Labreuche and F. Le Huédé, "MIRIAD: a tool suite for MCDA.," in *EUSFLAT Conf.*, 2005.

[6] C. Labreuche and M. Grabisch, "The Choquet integral for the aggregation of interval scales in multicriteria decision making," *Fuzzy Sets and Systems*, vol. 137, no. 1, pp. 11–26, 2003.

[7] C. A. Bana e Costa, J.-M. De Corte and J.-C. Vansnick, "On the mathematical foundations of MACBETH," in *Multiple criteria decision analysis*, Springer, 2016, pp. 421–463.

[8] D. Lafond, J.-F. Gagnon, S. Tremblay, N. Derbentseva and M. Lizotte, "Multi-criteria assessment of a whole-of-government planning methodology using MYRIAD," in *2015 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision*, 2015.

[9] G. Choquet, "Theory of capacities," in *Annales de l'institut Fourier*, 1954.

[10] C. Labreuche and S. Fossier, "Explaining Multi-Criteria Decision Aiding Models with an Extended Shapley Value.," in *IJCAI*, 2018.

[11] A. Datta, S. Sen and Y. Zick, "Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems," in *2016 IEEE Symposium on Security and Privacy (SP)*, 2016.

[12] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, 2017.

[13] G. Owen, "Values of games with a priori unions," in *Mathematical Economics and Game Theory*, Springer, 1977, pp. 76–88.

Part IV

Securing Critical Infrastructures of the Communications Sector

DOI: 10.1561/9781680836875.ch16

Chapter 16

Security and Resilience Challenges for the Critical Infrastructures of the Communications Sector

By Federica Battisti, Marco Carli, Federica Pascucci, Mirjam Fehling-Kaschek, Rodoula Makri, Maria Belesioti, Ioannis Chochliouros, Ioan Constantin and Xiao-Si Wang

This chapter introduces the main challenges for critical infrastructures in the communication sector. Specifically, the chapter will review the current threats that arise upon cyber and physical systems interconnection. At the same time, security strategies exploiting both the features (cyber and physical) of critical infrastructures will be introduced.

16.1 Introduction

The Internet of Things (IoT) (Lin *et al.*, 2017) revolution has brought the presence of the Internet in almost everything and has changed several aspects of our daily lives. The IoT technology, indeed, allows the massive introduction of Cyber-physical Systems (CPS) (Bordel *et al.*, 2017), that provide embedded intelligence, smart actuation, monitoring, and control to the peripheral nodes at the network edge. Relevant examples of CPS are represented by Critical Infrastructures (CIs), such as water distribution systems, smart grids, or telecommunication systems. CIs are of paramount importance in economy and social-being of citizens; therefore, they ask for protection against threats able to affect their operating level and, subsequently, the quality of our lives (Massel, 2018).

Threats for CIs are commonly divided into two classes (i.e., planned and unplanned), according to the possibility of forecasting them. Unplanned threats are represented by non-intentional human errors or natural disasters. In the last few years, the prevention of failure induced by extreme weather events is becoming more challenging, since climate changes make them more frequent and intense (Labelle *et al.*, 2008). Planned threats are mainly represented by cyberattacks. CIs have been the predominant target of several attacks that propagated due to domino effects.

Regardless the type of threat, the design and the development of resilience strategies are fundamental for protecting CIs. Specifically, CIs should be able to recover quickly from failures: they should be able to cope with either known or unknown threats according to the well-known paradigm detect, absorb, recover, and adapt (Sterbenz *et al.*, 2014). To this aim, CIs need to be equipped with detection tools to successfully identify a threat and reduce its impact. Moreover, a CI should react to the system performance degradation provoked by the threat guaranteeing a certain Quality of Service (QoS).

Among CIs, a key role is played by telecommunication networks: they are essential to support and maintain public and private services. Private businesses, government agencies, and other bodies rely on phone and Internet services provided by telecommunications networks to carry out daily operations. Telecommunication networks also supply services to health and social life. Since telecommunications are pivotal infrastructures, their protection requires more concern. This feature is becoming even more critical facing the 5G revolution. The extensive use of programmable platforms and exponential growth of connected devices require paradigms and tools to protect complex and flexible architectures.

Although since 2002 Universal Service Directive requires telecom companies to maintain the security and resilience of their networks (European Commission, 2002), there is no security and resilience standard for this CI. Commonly, the resiliency is addressed by using redundancy: most critical segments of the infrastructure are duplicated, and back-up power supplies are installed. Moreover, cyber and physical security issues are considered as independent, while recent events demonstrate that cyber-physical can affect the physical systems (Center for Strategic and International Studies, 2020; Computer Emergency Response Team-Coordination Center, 2020).

In this chapter, we address the security and resilience challenges for telecommunication infrastructure. To this aim, we provide an overview of the current and future structure of the telecommunication networks in Section 16.2; we classify the threats for telecommunication systems in Section 16.3 in order to understand the challenges in building a resilient system as detailed in Section 16.4. Finally, we draw some conclusive remarks in Section 16.5.

16.2 Current Telecommunication Infrastructures

Telecommunication networks exploit physical infrastructure for connecting users. They can be decomposed into two main components:

- The *core (backbone)* networks;
- The *access* networks.

The core network provides connectivity between sub-networks carrying a large amount of data. Core networks of different countries are implemented mainly by fiber infrastructure and the satellite links. Radio signals from satellite are used to connect remote communities, oil rigs, ships, and airplanes. The used radio frequency spectrum and the paths of their orbits are registered by the International Telecommunication Union (ITU). Telecommunication networks rely on information from global positioning system (GPS) satellites to synchronize with each other. Recently, several concerns have been raised about the cyber and physical security of both the undersea and the satellite links that carry a large number of global communications (Rishi Sunak, 2017): this aspect needs to be addressed when designing a security and resilience strategy. The access network is the component supplying the user with access to services. According to the type of access provided by telecommunication operators to users, traditionally networks have been further classified as:

- *Fixed-line networks*;
- *Mobile networks*.

The fixed-line network provides the connection to end customer by means of cables, through which a user can make phone calls, receive TV signals, or connect to the Internet. Its core network is composed of copper and fiber optic cables, having high bandwidth to connect switches and route communication. The access network is mainly composed by copper paired wires connecting the users; however, in the last few years, the use of fiber optic lines for the last mile is increased.

Mobile networks connect users to the network via wireless transmission technologies. Therefore, a mobile access network consists of base stations that communicate with the user handsets by using radio signals. Base stations provide access to the network over a limited area (i.e., the cell). The access network is connected to a backbone infrastructure composed of mobile switching centers using fixed-line (fiber optic cables) or radio links.

Fixed and mobile networks have mostly been developed separately. However, the rapid evolution of 5G mobile technologies leads to higher fiber demand, thus boosting the convergence of networks. Indeed, the new mobile technologies allow the development of a novel set of applications (Agiwal *et al.*, 2016), mainly focused on the fulfillment of user requirements. To this aim, the Quality of Experience (QoE) is replacing the QoS in the management of the networks. According to Qualinet (Brunnström *et al.*, 2013), QoE can be defined as: *"The degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user's personality and current state."* As can be noticed, the fulfillment of the QoE requirement is much more demanding than that of QoS. From the telecommunication point of view, the main challenge is the development of optimized self-organized networks able to timely provide services. Software Defined Network and cloud technologies may represent appropriate tools to allocate the available resources. However, the increasing demand for connectivity enables the antenna densification process, i.e., the deployment of cells covering a small area. Concerning the security issue, the novel architectures provide redundancy by design; however, novel CPS threats will emerge due to the increased number of attack surfaces.

16.3 Classification of Threats in Telecommunication Infrastructures

In Euchner *et al.* (2015), the requirements, that need to be taken into account when facing the challenge of securing a telecommunication system, are:

- The *parties involved*;
- The *assets* that need to be protected;
- The *threats* against which those assets must be protected;
- The *vulnerabilities* associated with the assets and the environment;
- The *overall risk* to the assets from those threats and vulnerabilities.

Concerning the parties involved, the main role is played by customers/ subscribers: they expect that the network is available and that the services offered are reachable, especially in emergency scenarios. Public authorities ask for security by directive and legislation to guarantee service availability, privacy protection, and fair competition. Network operators and services providers need to preserve their operating level and business so that to meet the demand of customers, business partners, and the requirements of public authorities.

The assets are represented by the personnel, the infrastructure (communication and computing devices, equipment, and facilities), the information and data, and the services provided.

A security threat is defined as a potential violation of security, that is: a possible danger that might exploit a vulnerability or weaknesses of the system to breach security and, therefore, to lead to risky impact. As stated previously, threats can be regarded either as unplanned/accidental or planned/intentional (Jones *et al.*, 2012). Moreover, they can be either active or passive; active threats significantly affect information and/or operation in the system, while the passive ones do not provide any change in the information and/or operation of the corresponding systems.

About the telecommunication security threats and related risks, different classification can be drawn according to different purposes. Therefore, in the literature, several threat classification schemes have been proposed, upon the basis of a variety of criteria. Understanding the potential threats is of paramount importance to deeply get insights on the security and resilient challenges for the telecommunication systems. To this aim, we consider the Recommendation ITU-T X.1205 (ITU-T, 2008) that provides a taxonomy of security threats from an organizational point of view, along with a discussion of the threats at the various layers of a network. Specifically, we consider three different types of threats, according to the part of the systems affected, that is:

- *Physical threats*;
- *Cyber threats*;
- *Cyber-physical threats*.

The physical threats affect the physical assets (i.e., communication and computing devices, equipment, and facilities), the cyber threats exploit vulnerabilities in the cyber space to harm the digital assets, and the cyber-physical threats exploit vulnerabilities in the cyber space to disrupt the physical assets. Moreover, we also analyze *Advanced Persistent Threat* (APT), that is considered the most demanding threat to detect and defend against to date.

In the following, these types of threats and APT for telecommunication infrastructure are reviewed to further understand the security challenges.

16.3.1 Physical Threats

Physical threats damage the physical infrastructure of telecommunication networks and can be either planned or unplanned (Jones *et al.*, 2012; Electronic Communications Resilience and Response Group, 2004).

The planned threats are related to intentional events, motivated by financial gain, internal sabotage, terrorism, and vandalism. Example of planned physical threats are damages to the transmission equipment (telecom pillars, antenna, buildings, etc.), by using weapons or drones, copper or fiber optics cables theft to interrupt network services, or signal jamming to disrupt wireless networks.

Unplanned threats can be roughly divided into two main categories. The first one is related to hardware and/or software failure due to unintentional human actions; the second one concerns natural hazards. The most common cause of telecommunication failure, as reported by ENISA (European Union Agency for Network and Information Security) every year since 2012 (ENISA, 2019), is associated with power breakdown: lack of fuel for backup generators, excavators shearing through cables, anchors damaging undersea cable are examples of this type of threat. It is worth noticing that power and telecommunication infrastructures rely on each other: on the one hand, the telecommunication infrastructures depend on continuous supply of power and, on the other hand, the electrical power industry depends on telecoms to run their extensive network of generators and grid distribution. Although the intentional damages can seriously harm telecommunication networks, according to ENISA reports the most prolonged disrupts are caused by natural hazards (i.e., weather events, seismic activity, fire, and explosions). Flooding, strong winds, lightning, cold weather, and heatwaves can affect telecommunication physical assets either directly or indirectly by damaging the power infrastructures. Furthermore, changes in the near-Earth space environment can influence the performances of the telecommunication systems.

Concerning the planned events, some prevention actions need to be set-up, by applying suitable frameworks and related measures. Concerning unplanned events, only mitigation strategies could be applied due to the unpredictable characteristics of the incidents. It is worth mentioning that due to climate change, natural hazards are becoming more frequent.

16.3.2 Cyber Threats

Cyber threats affect the telecommunication operation, software system, and services. They can be divided into intentional and accidental, like the physical ones.

The unplanned threats are represented by system and software failures. System failures occur when the performances of a telecommunication system are downgraded due to system errors. The challenge is to avoid the single points of failure by enhancing resilience strategies. However, not all parts of the network can be made redundant and, in these cases, the complementary restore and repair procedures need to be strengthened. The software failures are usually related to bugs in the algorithms that control the equipment. Although errors in software are acceptable for personal computers, a telecommunication network cannot bear crashes and delays in services. The most challenging software issue is represented by the systemic or common-mode failure; in this case, a software error in one network node causes the same fault to occur in other connected nodes, leading to a runaway failure of the whole network.

The planned cyber threats are related to hacking activities and attacks. They include both typical cyberattacks and specific ones, tailored to the specific infrastructure(s). The eavesdropping aims at breaching the system or service, to spoof the user identity, to disclose information (privacy breach or data leak), and to gain knowledge on the system. The man in the middle attack covertly intercepts the communication between two nodes, records the information, and even alters it. The denial of service (DoS) targets at making a resource unavailable to the users by temporarily or indefinitely disrupting services of a host connected to the Internet. A denial-of-service attack floods systems, servers, or networks with traffic to spill over resources and bandwidth; as a result, the system is unable to fulfill legitimate requests. Attackers can also use multiple compromised devices to launch this sort of attack. A semantic attack is devoted to change and disseminate correct and/or incorrect information to cover tracks of malicious activities. Attacks launched by malicious codes include the execution of viruses, worms, Trojan horses, and active Web scripts aiming at destroying or stealing information. They represent well-known computer security threats, since a computer virus is a program written to alter the way a computer operates, without the permission or knowledge of the user. A virus replicates and executes itself, usually doing damage to the computer in the process. Cyber threats most connected with infrastructure are related to vulnerabilities in system security procedures, hardware design, internal controls, or software code. They could be exploited to gain unauthorized access, to manipulate the integrity or to affect the availability of both classified or sensitive information and non-sensitive information of protocols, procedures, and equipment. One of the main weakness is represented by the legacy protocols: some protocols, indeed, are old and were designed without considering future security issues. Concerning equipment, backdoor attacks and device compromise are mighty threats. Regarding the first, it is set up by software development companies or hardware providers that leave a single point of failure in order to obtain access to a system or application found in production. The second one hits the devices used in telecommunication networks (e.g., home routers): once they are compromised, malicious attackers can anonymously access services. The zero-day exploit is a cyberattack that occurs on the same day a weakness is announced in software, just before a patch or a solution is provided. A Structured Query Language (SQL) injection targets SQL servers by introducing malicious code into vulnerable website and retrieving data and information. The most challenging attack, however, is represented by phishing, since it includes the human in the attack loop. Phishing is an example of social engineering techniques, based on email spoofing and instant messages. It is the fraudulent attempt to gain sensitive information (usernames, password, personal identification number, credit card numbers, etc.) by appearing as a request from reputable sources.

16.3.3 Cyber-physical Threats

Cyber-physical threats encompass attacks to information systems that have an impact on physical assets (cyber to physical threats) and *vice versa* (physical to cyber threats), i.e., physical threats that disrupt the information systems. The physical to cyber threats can be accidental, whereas the cyber to physical attacks are always intentional.

According to Paridari *et al.* (2018), physical threats are represented by both physical intrusion and attacks to sensors and actuators resulting in system failure. The physical intrusion refers to an intruder that circumvents the physical security of an infrastructure in order to harm the cyber domain. The sensor or actuator attacks refers to a physical damage that brings to a system fault. In Paridari *et al.* (2018) also, cyber to physical threats are considered. Specifically, the most common attacks are network disruption to harm physical assets and electronic jamming to deliberate cause losses in physical assets. For a telecommunication system, the first cyber to physical threats is related to remote action in the cyber domain that cause failure in providing services due to physical issue (i.e., power shutdown). The second cause denial of service due to traffic overloads and can cause damages in the interconnected critical infrastructures (e.g., control systems of power distribution networks).

The main approach adopted to prevent cyber-physical threats in a CPS is by controlling its vulnerability; however, it constitutes a challenge. CPS, indeed, are composed by heterogeneous building blocks. From a hardware perspective, they are composed by different components (i.e., sensors, actuators, controllers, physical structures, and embedded systems). CPS also include firmware, communication channel, proprietary, and commercial software for controlling and monitoring the systems. Every single component as the whole integrated system represents an attack surface. Therefore, a fundamental task is to get insights on the vulnerability risk in order to identify missing pieces, gaps, and weak links. Another challenge is related to privacy preserving issues in the CPS: in a CPS, indeed, it is difficult to identify, trace, and examine the attacks, which may originate from, move between, and target at multiple CPS components. An in-depth understanding of the vulnerabilities, threats, and attacks is essential to the development of defence mechanisms.

Telecommunication infrastructures are rarely regarded as a CPS, although the exponential growth in the development and deployment of networked systems has brought impacts to almost all aspects of daily life. It is worth noticing, however, that telecommunication systems provide and manage the communication channels of all the other critical infrastructures (e.g., power distribution systems, water distribution systems, transportation networks, etc.); therefore, they are tightly related and interconnected with CPS. Moreover, the facilities of a telecommunication system, as well as all the physical devices (i.e., antenna pillars, network control systems or wireless sensor networks) that the emerging 5G technologies foresee, make the telecommunication system itself a CPS (Hutchison and Sterbenz, 2018).

16.3.4 Advanced Persistent Threats

The most challenging threat to detect and defend against is considered the APT. An APT is a set of stealthy and continuous computer hacking processes, which gain unauthorized access to a computer network and remain undetected for an extended period. It is set up by group driven by political and/or economic motivations; the actors behind an APT have the capability and determination to achieve a specific target. An APT usually targets either private organizations, states or both, and requires a high degree of covertness over a long period of time.

As suggested by the name, it consists of three main components, namely advanced, persistent, and threat. The advanced component implies that sophisticated techniques are adopted: traditional espionage vectors, social engineering, human intelligence, and infiltration are used to gain access to a physical location to enable network attacks. Commonly, the main target is to place custom malicious code on one or multiple computers in order to accomplish a specific task. The persistent component implies that an external command and control system is continuously monitoring and extracting data from a specific target during the dwell time (i.e., the time an APT attack goes undetected). This provides to the attackers a significant amount of time to go through the attack cycle, propagate, and achieve the objective. The threat component involves human in orchestrating the attack (Alshamrani *et al.*, 2019).

APTs exploit Internet and/or infected media to breach the target system. Internet connections are used to send malicious payload via email attachments, peer-to-peer file sharing, or spear-phishing. Media infection may consist of infected Universal Serial Bus (USB) memory sticks, infected memory cards, or infected appliances. Furthermore, cyber threats (i.e., zero day attack, man in the middle, etc.) can be applied.

To date, every major business sector has recorded instances of attacks by advanced actors with specific goals seeking to steal, spy, or disrupt. The most famous APT for industrial control system is considered Stuxnet (Albright *et al.*, 2010), while the world's first global ransomware attack, Wannacry (Ghafur *et al.*, 2019), was shown to be based on code produced by a known APT.

ENISA's threat landscape report predicts that high-capability agents will specialize in the future on more off-the-shelf campaigns rather than custom techniques, so as to enhance stealthiness and further improve APT effectiveness (ENISA, 2019), showing that APTs exemplify the advanced cyber threat due to increasing frequency, importance, and complexity in countering.

16.4 Resilience in Telecommunication Systems

Resilience for a telecommunication system is defined as the capability of a network to prepare, prevent, protect, respond, and recover against a challenge by maintaining an acceptable quality of service (Thoma *et al.*, 2016). Resilience is regarded as a major requirement as well as a design objective for CIs; however, it is of paramount importance for Internet that is the *"critical infrastructure used by citizens, governments, and businesses"* [as described ENISA (2019)].

Resilience represents a cross-cutting edge between information and network security, fault-tolerance, dependability (Avižienis *et al.*, 2004), performability (Meyer, 1992), and network survivability (Ellison *et al.*, 1997; Sterbenz *et al.*, 2002). It is useful to underline that engineering resilience has a monetary cost: to this aim, it is critical to maximize the effectiveness of committed resources.

In telecommunication systems foreseen by the 5G architecture, the main resilience challenges are related to software-based networks. The radio access network allows to add/remove nodes by easily reconfiguring the network in an automated way. This capability enables the set-up of automatic redundant configuration, while introduces new security and resilience challenges, namely the risk of accepting a malicious node. Software-based networks, indeed, rely on centralized control that can represent a single point of failure. The key challenge is to make the control level resilient and secure in order to avoid the propagation of attack and failure from this level to the data and application ones.

In the literature, two mitigating strategies are considered: the cross-layer fault management and the learning dynamic resource dependencies. The cross-layer fault management aims at timely diagnosing faults and attacks in order to set up recovery strategies to guarantee a suitable level of service. To achieve this goal, proper metrics to detect and identify system malfunctioning need to be defined: they are represented by the key performance indicators that may give relevant insights on the behavior of the system. By learning dynamic resource dependencies, it is possible to build a run-time model for the software-based network that allows to track faults and alarms.

Another challenge that arises considering software-based networks is related to network slicing that enables the coverage of different use cases (NGMN Members, 2015) by mapping virtual resources into physical infrastructure. In this case, the network resilience depends on the resilience of the slicing service and of the physical infrastructure. A fault on the physical layer, indeed, propagates into virtual resources.

16.5 Conclusions

This chapter analyzes the security and resilience of telecommunication systems considering the challenges that an improved connectivity may induce in CPS. Furthermore, the telecommunication infrastructure is regarded itself as a CPS. We investigated the novel challenges and the security issues arising when the next generation of telecommunication systems is considered. The main concerns are related with the convergence between fixed and mobile networks, the exploitation of cyber threats to damage the physical layer, and the novel network technologies used by the 5G generation of mobile networks.

The challenges in security, however, can be considered also as opportunities. Resilient systems, indeed, can be easily set up by using the network functions virtualization, as foreseen by the ETSI (the European Telecommunications Standards Institute) (NFV ETSI Industry Specification Group, 2017). The future telecommunication systems, indeed, will be composed of Physical Network Functions that cannot be virtualized and Virtual Network Functions that run in commodity hardware. These two components will realize network services and /or application coordinated by an orchestrator, able to implement the appropriate policy. Here, the opportunity is to exploit the orchestrator also for security purposes. Finally, the networks virtualization and the software defined networking can be synergetically used to set up automatic network.

Acknowledgments

This chapter is partially supported by the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 786409 (RESISTO – RESIlience enhancement and risk control platform for communication infraSTructure Operators).

References

Agiwal, M., A. Roy, and N. Saxena (2016). “Next Generation 5G Wireless Networks: A Comprehensive Survey.” *IEEE Communications Surveys & Tutorials*. 18: 1617–1655. DOI: 10.1109/COMST.2016.2532458.

Albright, D., P. Brannan, and C. Walrond (2010). “Stuxnet malware and Natanz: Update of ISIS.” *Tech. Rep.* Institute for Science and International Security. URL: https://isis-online.org/uploads/isisreports/documents/stuxnet_update_15Feb2011.pdf (accessed on 02/21/2020).

Alshamrani, A., S. Myneni, A. Chowdhary, and D. Huang (2019). “A Survey on Advanced Persistent Threats: Techniques, Solutions, Challenges, and Research Opportunities.” *IEEE Communications Surveys & Tutorials*. 21(2): 1851–1877. DOI: 10.1109/COMST.2019.2891891.

Avižienis, A., J. C. Laprie, B. Randell, and C. Landwehr (2004). “Basic concepts and taxonomy of dependable and secure computing.” *IEEE Transactions on Dependable & Secure Computing*. 1(1): 11–33. DOI: 10.1109/TDSC.2004.2.

Bordel, B., R. Alcarria, T. Robles, and D. Martin (2017). “Cyber-physical systems: Extending pervasive sensing from control theory to the Internet of Things.” *Pervasive and Mobile Computing*. 40: 156–184. DOI: 10.1016/j.pmcj.2017.06.011.

Brunnström, K., S. Ariel Beker, K. de Moor, A. Dooms, S. Egger, M. N. Garcia, T. Hossfeld, S. Jumisko-Pyykkö, C. Keimel, and M. C. Larabi (2013). “Qualinet White Paper on Definitions of Quality of Experience.” *Tech. Rep.* Qualinet. URL: https://hal.archivesouvertes.fr/hal-00977812/document (accessed on 02/21/2020).

Center for Strategic and International Studies (2020). “Significant Cyber Incidents since 2006.” URL: https://csis-prod.s3.amazonaws.com/s3fs-public/200108_Significant_Cyber_Events_List.pdf?aj4_%5C-VlDq2hSan2U8O5mS29Iurq3G1QKa (accessed on 02/21/2020).

Computer Emergency Response Team-Coordination Center (2020). “CERT Vulnerability Notes Database.” URL: https://www.kb.cert.org/vuls/bypublished/desc/ (accessed on 02/21/2020).

Electronic Communications Resilience and Response Group (2004). “Recommendation ITU-T X.1205: Overview of cybersecurity.” URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/61335/telecommunications_sector_intro.pdf (accessed on 02/21/2020).

Ellison, R. J., D. A. Fisher, R. C. Linger, H. F. Lipson, T. Longstaff, and N. R. Mead (1997). "Survivable Network Systems: An Emerging Discipline." *Tech. Rep.* Software Engineering Institute, Carnegie Mellon Universit. URL: https://resources.sei.cmu.edu/asset_files/TechnicalReport/1998_005_001_16598.pdf (accessed on 02/21/2020).

ENISA (2019). "ENISA Threat Landscape Report 2018." URL: https://www.enisa.europa.eu/publications/enisa-threat-landscape-report-2018 (accessed on 02/21/2020).

Euchner, M., G. Sebek, H. Bertine, and M. Harrop (2015). "Security in Telecommunications and Information Technology: An overview of issues and the deployment of existing ITU-T Recommendations for secure telecommunications." *Tech. Rep.* ITU. URL: https://www.itu.int/dms_pub/itu-t/opb/hdb/T-HDB-SEC.05-2011-OAS-PDFE.pdf (accessed on 02/21/2020).

European Commission (2002). "Directive 2002/22/EC." URL: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2002:108:0051:0077:EN:PDF (accessed on 02/21/2020).

Ghafur, S., S. Kristensen, K. Honeyford, G. Martin, A. Darzin, and P. Aylin (2019). "A retrospective impact analysis of the WannaCry cyberattack on the NHS." *npj Digital Medicine*. 2(98). DOI: 10.1038/s41746-019-0161-6.

Hutchison, D. and J. P. G. Sterbenz (2018). "Architecture and design for resilient networked systems." *Computer Communications*. 131: 13–21. DOI: 10.1016/j.comcom.2018.07.028.

ITU-T (2008). "Recommendation ITU-T X.1205: Overview of cybersecurity." URL: https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-X.1205-200804-I!!PDF-E&type=items (accessed on 02/21/2020).

Jones, J., G. Carpenter, M. Kilger, and S. Bodmer (2012). *Reverse Deception: Organized Cyber Threat*. McGraw-Hill.

Labelle, L., R. Rodschat, T. Vetter, and K. Ludwig (2008). "ICTs for e-Environment." *Tech. Rep.* ITU. URL: https://www.itu.int/dms_pub/itu-d/opb/str/D-STR-ENV.CLIMATE-2008-PDF-E.pdf (accessed on 02/21/2020).

Lin, J., W. Yu, N. Zhang, X. Yang, H. Zhang, and W. Zhao (2017). "A Survey on Internet of Things: Architecture, Enabling Technologies, Security and Privacy, and Applications." *IEEE Internet of Things Journal*. 4(5): 1125–1142. DOI: 10.1109/JIOT.2017.2683200

Massel, L. (2018). "The Research Convergence of Critical Infrastructures, Quality of Life and Safety." In: *Proceedings of the Vth International Workshop Critical Infrastructures: Contingency Management, Intelligent, Agent-based, Cloud Computing and Cyber Security*. Atlantis Press. DOI: 10.2991/iwci-18.2018.24.

Meyer, J. F. (1992). "Performability: a retrospective and some pointers to the future." *Performance Evaluation*. 14(3): 139–156. DOI: 10.1016/0166-5316(92)90002-X.

NFV ETSI Industry Specification Group (2017). "Network Function Virtualisation (NFV); Reliability; Report on the resilience of NFVMANO critical capabilities." *Tech. Rep.* European Telecommunications Standards Institute. URL: https://www.etsi.org/deliver/etsi_gr/NFV-REL/001_099/007/01.01.02_60/gr_nfv-rel007v010102p.pdf (accessed on 02/21/2020).

NGMN Members (2015). "5G White Paper." *Tech. Rep.* NGMN Alliance. URL: https://www.ngmn.org/wp-content/uploads/NGMN_5G_White_Paper_V1_0.pdf (accessed on 02/21/2020).

Paridari, K., N. O'Mahony, E. D. Mady, R. Chabukswar, M. Boubekeur, and H. Sandberg (2018). "A Framework for Attack-Resilient Industrial Control Systems: Attack Detection and Controller Reconfiguration." *Proceedings of the IEEE*. 106(1): 113–128. DOI: 10.1109/JPROC.2017.2725482.

Rishi Sunak, M. P. (2017). "Undersea Cables Indispensable, insecure." *Tech. Rep.* Policy Exchange. URL: https://policyexchange.org.uk/wpcontent/uploads/2017/11/Undersea-Cables.pdf (accessed on 02/21/2020).

Sterbenz, J. P. G., D. Hutchison, E. K. Cetinkaya, A. Jabbar, J. P. Rohrer, M. Schöller, and P. Smith, "Redundancy, diversity, and connectivity to achieve multilevel network resilience, survivability, and disruption tolerances." *Telecommunication Systems*. 56: 17–31. DOI: 10.1007/s11235-013-9816-9.

Sterbenz, J. P. G., R. Krishnan, R. R. Hain, A. W. Jackson, D. Levin, R. Ramanathan, and J. Zao (2002). "Survivable Mobile Wireless Networks: Issues, Challenges, and Research Directions." In: *Proceedings of the 1st ACM Workshop on Wireless Security. WiSE '02*. Atlanta, GA, USA: Association for Computing Machinery. 31–40. DOI: 10.1145/570681.570685.

Thoma, K., B. Scharte, D. Hiller, and T. Leismann. (2016). "Resilience Engineering as Part of Security Research: Definitions, Concepts and Science Approaches." *European Journal for Security Research*. 1: 3—19. DOI: 10.1007/s41125-016-0002-4.

DOI: 10.1561/9781680836875.ch17

Chapter 17

RESISTO—RESIlience Enhancement and Risk Control Platform for Communication infraSTructure Operators

By Alberto Neri and Alessandro Neri

17.1 Introduction

Communications play a fundamental role in the economic and social well-being of the citizens and on operations of most of the Critical Infrastructures (CIs). Thus, they are a primary target for criminals having a multiplier effect on the power of attacks and providing enormous resonance and gains. Also extreme weather events and natural disasters represents a challenge due to their increase in frequency and intensity requiring smarter resilience of the Communication CIs, which are extremely vulnerable due to the ever-increasing complexity of the architecture also in light of the evolution towards 5G, the extensive use of programmable platforms, and exponential growth of connected devices. The fact that most enterprises still manage physical and cybersecurity independently represents a further challenge. RESISTO platform is an innovative solution for Communication CIs to increase situation awareness and enhance CIs resilience. An integrated Risk and Resilience analysis management and improvement process is in charge to identify threats and prevent impacts as well as RESISTO implements an innovative Decision Support System to protect communication infrastructures able to detect negative events, respond, and recover from physical, cyber, and combined cyber-physical threatening events. A suite of state-of-the-art cyber/physical threat detectors (Machine Learning based, IoT security, Airborne threat detection, holistic audio-video analytics) complete the platform. Through RESISTO, Communications Operators will be able to implement a set of recovery actions and countermeasures that significantly reduce the impact of negative events in terms of performance losses, social consequences, and cascading effects in particular by bouncing efficiently back to original and forward to operational states of operation. RESISTO adopts a unified approach to face physical as well as cyber threats as well as a double and integrated approach between offline and run-time activities applicable to different kinds of CIs.

17.2 RESISTO Architecture

The logical architecture of RESISTO integrates two control loops both running on top of the Communication Infrastructure and interlinked with each other (Figure 17.1) that implement the five core security functionalities introduced by the USA National Institute of Standards and Technology (NIST) in the "Framework for Improving Critical Infrastructure Cybersecurity," namely Identify, Protect, Detect, Respond, and Recover.

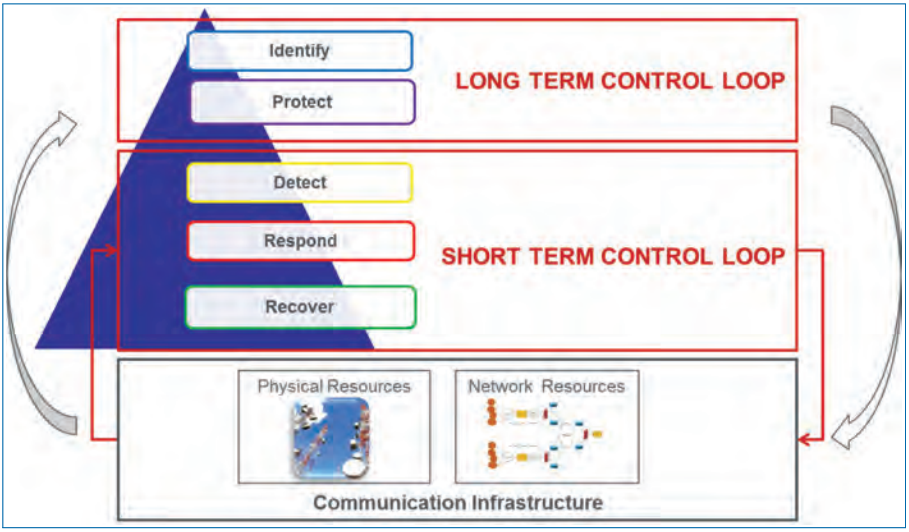


Figure 17.1. RESISTO logical architecture.

The **Long-term Control Loop** (LTCL) is an offline activity, following a well-defined methodology and supported by advanced tools, aimed to identify infrastructure vulnerabilities and cyber and physical security threats and, consequently, to define assets configuration and interventions in order to improve CI's resilience and robustness. For each loop cycle, a set of Resilience Indicators (RIs), relevant to critical threat event typologies, are estimated and stored in a Knowledge Base (KB) as described in detail in Section 17.5. An LTCL cycle is performed on a periodic basis or when particular events take place (new threats or discovery of previously undetected vulnerabilities). It is typically conducted annually, quarterly, or even monthly.

The **Short-term Control Loop** (STCL) is the runtime component of the platform. It promptly responds to detected cyber/physical attacks and events that may impact the operational life of the system. It enhances situation awareness and provides operators with a Decision Support System cockpit able to implement the best response to an identified adverse event with the aim of mitigating the event's effects and recovering standard operating conditions. While facing adverse cyber/physical events, some actual RIs values are measured and stored in the KB (see Section 17.5).

Moreover, LTCL and STCL are strongly interlinked with each other. In fact, comparison between target RIs estimated by the LTCL and their actual values measured by the STCL facing run-time threat events establishes a higher level global control loop able to continuously review and improve infrastructure resilience and methods.

17.3 The Long-term Control Loop

The RESISTO LTCL is in charge of the configuration of the Communication's Critical Infrastructure according to the security assessment.

While STCL provides tools for immediate reaction against attacks in real time, LTCL leads to the identification of criticalities and definition of long-term strategies. Therefore, it is conducted on a periodic basis as well as in case of specific events, e.g., detection of new types of vulnerabilities, expectation of new threats, or after significant CI changes that may impact on security.

The LTCL implementation is based on a sophisticated risk and resilience management process, aimed at identifying and evaluating risks and suggesting treatment and mitigation strategies, that extends the ISO 31000 standard.

Each cycle is structured into the following nine sequential steps (Figure 17.2):

1. **Context analysis:** it is devoted to the general description of the system, including societal, economic, legal, and ethical context, and includes the

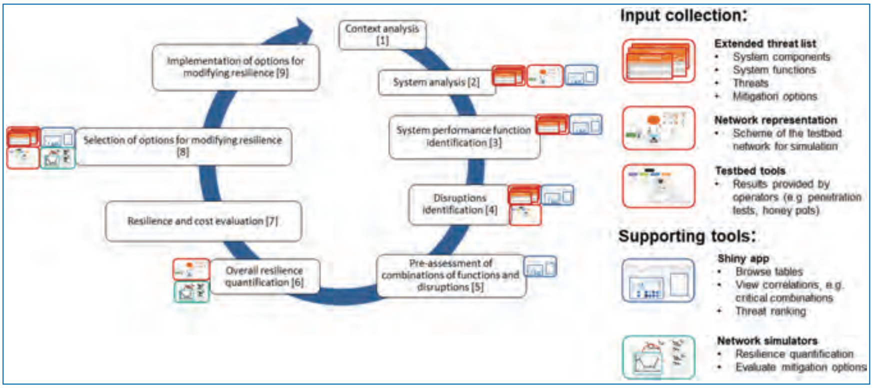

Context analysis [1]
System analysis [2]
System performance function identification [3]
Disruptions identification [4]
Pre-assessment of combinations of functions and disruptions [5]
Overall resilience quantification [6]
Resilience and cost evaluation [7]
Selection of options for modifying resilience [8]
Implementation of options for modifying resilience [9]
Input collection:
Extended threat list
• System components
• System functions
• Threats
• Mitigation options
Network representation
• Scheme of the testbed network for simulation
Testbed tools
• Results provided by operators (e.g. penetration tests, honey pots)
Supporting tools:
Shiny app
• Browse tables
• View correlations, e.g. critical combinations
• Threat ranking
Network simulators
• Resilience quantification
• Evaluate mitigation options

Figure 17.2. Risk and resilience management process.

identification of key stakeholders, resilience objectives, restrictions, and evaluation criteria.

2. **System analysis:** it is aimed at the analysis of the system environment and interfaces, including boundary definitions, static and dynamic analysis, and (graphical) modeling/representation.
3. **System performance function identification:** it is targeted to the definition of (non)performance (service) functions of the system, including qualitative and quantitative descriptions. The system (non)performance functions in combination should cover the expected system behavior and its assessment.
4. **Disruptions identification:** it identifies threats, hazards, and disruptions (classical risk events) that might affect system (non)performance, as well as potentially affected system functions, system layers, and resilience capabilities.
5. **Pre-assessment of combinations of functions and disruptions:** it analyses all combinations of system functions (step 3) and potential disruptions (step 4), in order to identify critical combinations which need to be further evaluated (in step 6). Step 5 is typically conducted analytically using a semi-quantitative approach. Step 5 and step 6 take account of all resilience cycle phases.
6. **Overall resilience quantification:** based on system modeling and simulation, it is aimed at determining resilience quantities, i.e., at quantification of the resilience of the system (non)performance functions regarding the identified threats based on the criticalities identified in the previous step 5. Step 6 covers advanced (overall) resilience quantification approaches.
7. **Resilience and cost evaluation:** it is devoted to the comparison of resilience performance, illustration of the performance loss, and evaluation of the acceptance level for all threats. Step 7 evaluates the results of steps 5 and 6.

8. **Selection of options for modifying resilience:** it selects the best options for resilience improvement based on a preselected decision-making method. Step 8 includes the re-execution of all previous steps that affect the resilience (semi) quantification to assess the resilience gain taking into account the planned improvement methods.
9. **Implementation of options for modifying resilience:** development and implementation of those options for improving resilience select at Step 8, based on domain-specific standards as far as possible and efficient methods corresponding to determined resilience levels for all subsystems.

In principle, the methodology employed in RESISTO is applicable to any kind of CI subjected to both physical and cyber threats. However, some elements and tools have been adapted or added in order to exploit specific aspects of the Communications domain.

Specific data concerning CI characteristics, potential threats, vulnerabilities, and their exploitation, and countermeasures required by LTCL are stored in a dedicated Data and Knowledge Base. However, a web application supporting fast and easy browsing through its content and further information inference, such as critical combinations of system functions and threats and threat ranking (Step 5), which serve as additional input to other steps of the resilience management process, has been realized. This application is based on the Shiny package of the free programming language R for statistical computing and allows a semi-quantitative assessment of critical risks. Tabular Excel templates for data import/export complement the tool.

Quantification of the resilience Matrix of the critical risks based on CI simulations is further supported. At present, two simulation tools have been integrated in RESISTO: the platform-integrated CISIApro simulator, also employed in real-time by the STCL, and the offline CaESAR simulator. CISIApro (Critical Infrastructure Simulation by Interdependent Agents) is a software engine, developed by the University of Roma Tre, able to calculate complex cascading effects, taking into account (inter)dependencies and faults propagation among the involved complex systems. CISIApro has been developed in the framework of the H2020 ATENA Project and in RESISTO has been updated to version 2.0, adding some important functionalities related with the modeling of telecommunication infrastructures.

In addition to the RIs computation, the simulators allow to assess and rank possible mitigation strategies. It should be noted that these simulations are not performed on an event basis, as a consequence of a detected attack, like in STLC, but rather on a periodic basis to identify weak points of the current setup of the infrastructure.

In Figure 17.2, the tools supporting the different steps of the risk and resilience management are indicated.

Among the methodologies and tools adopted in RESISTO in order to improve the knowledge about possible threats and their consequences (Figure 17.2, Step 4 from the Risk and resilience management process), we cite:

1. *Attack Trees*: Attack trees are a way to describe an attack. In the LTCL, they can be used to identify situations that the organization has no defense process in place. Each attack should have countermeasures in place.
2. *Honeypots*: A honeypot is a set of physical, HW, and/or SW modules simulating legitimate interactions with external users while they are instead separate entities not performing any real operation and/or handling real data. Post-processing of honeypot logs is a long time used technique to detect and analyze sophisticated intrusions while keeping the real system safe. Moreover, it provides the possibility to observe an attack over time, then enabling both: learning about new threats and assessment of known ones. Since it collects detailed historical information, the technique is particularly useful for the LTCL purposes.
3. *Penetration tests*: A penetration testing (PENTEST) is a combination of techniques that considers various issues of the systems and tests, analyses, and gives solutions. It is based on a structured procedure that performs penetration testing step-by-step. Undertaking a series of penetration tests helps test security arrangements and identifies improvements. When carried out and reported properly, a penetration test can give knowledge of nearly all technical security weaknesses and provides the information and support required to remove or reduce those vulnerabilities.
4. "*MITRE ATT&CK*": it consists in the constant monitoring of the information in the "MITRE ATT&CK" knowledge base on potential attacks in order to develop and update threat models for risk assessment of their networks.

17.4 The Short-term Control Loop

The RESISTO Short-term Control Loop (STCL) is a typical run-time control loop. It is in charge of detecting potential physical, cyber, and physical/cyber combined threat events that may impact on the operational life of the system and react promptly.

The STCL:

- Monitors the physical and cybersecurity status of the infrastructures, correlating physical and cyber domain events, and monitoring communication

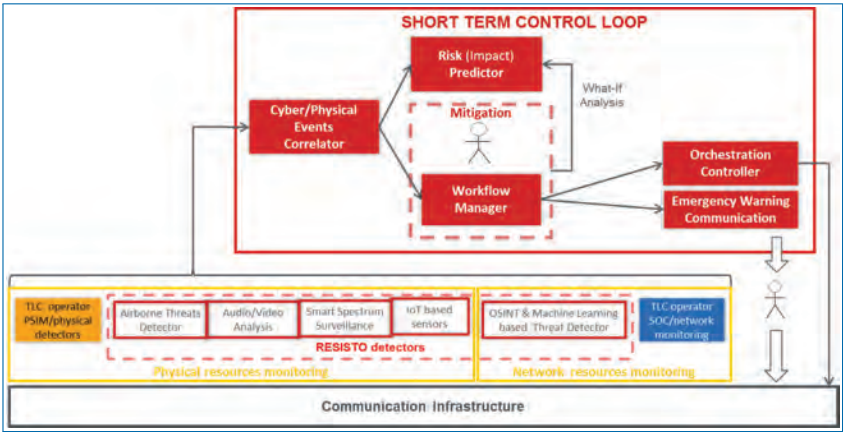

SHORT TERM CONTROL LOOP
Risk (Impact) Predictor
What-If Analysis
Cyber/Physical Events Correlator
Mitigation
Workflow Manager
Orchestration Controller
Emergency Warning Communication
TLC operator PSIM/physical detectors
Airborne Threats Detector
Audio/Video Analysis
Smart Spectrum Surveillance
IoT based sensors
RESISTO detectors
Physical resources monitoring
OSINT & Machine Learning based Threat Detector
TLC operator SOC/network monitoring
Network resources monitoring
Communication Infrastructure

Figure 17.3. Short-term Control Loop functional flow.

infrastructure data in order to collect and/or detect anomalies and provide early warnings on security attacks or events adversely impacting security;

- Evaluates the performance degradation causes related to detected anomalies and security attacks on the Communication CI and interlinked CIs, if known, based on the cascading effect;
- Supports decision-making providing a qualitative and quantitative What-If analysis tool in order to evaluate the best mitigation strategy;
- Drives response and recovery by means of action workflows (composed of directives to intervention teams, physical protection devices activation) and, mainly, of orchestrated Communication Network reconfiguration and protection function activation.

The STCL functional control flow is reported in Figure 17.3.
Input data to the STCL can be grouped into the following categories:

1. Physical events related to attacks (e.g., intrusions, damage) or to potentially dangerous events (e.g., unauthorized UAV flights);
2. Cyberattacks;
3. Communication infrastructure physical layer/HW monitoring data (e.g., power and energy consumption and HW faults);
4. Communication network QoS monitoring data (e.g., offered traffic, throughput, latencies, error statistics, …).

The sources of such data and information could be:

- Legacy Physical Security Information Management (PSIM) systems or other physical attack detectors made available by the telecommunication operator,

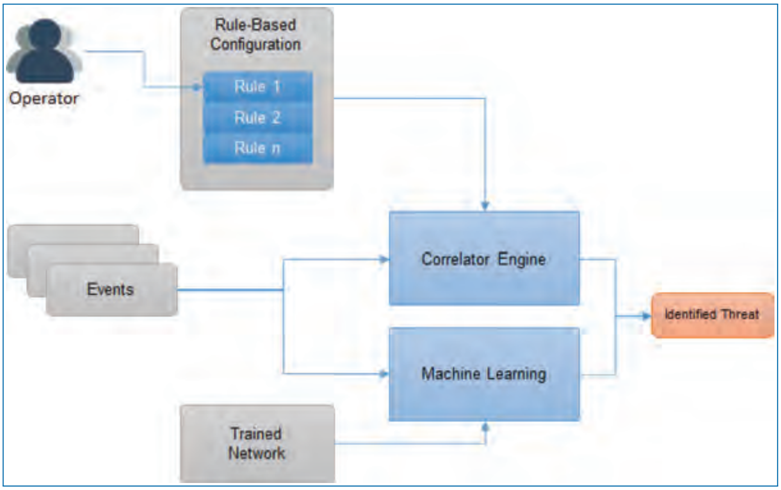


Figure 17.4. Correlator architecture.

- Legacy Security Operating Centers (SOCs) or other cyberattack detectors made available by the telecommunication operator,
- RESISTO additional physical/cyber threat detectors [e.g., airborne threats detection systems, smart spectrum surveillance, OSINT (Open-Source Intelligence)-based] described in the following.

From a functional point of view, input data are collected by the **Cyber/Physical Events Correlator**. The Events Correlator not only propagates, as alarms, externally detected and collected attack/anomaly events, but it also generates alarms on its own from apparently harmless events and monitoring data. This latter action is performed by using several event correlation techniques, such as logical, causal, and temporal correlation based on event timing.

The Cyber/Physical Events Correlator is composed by the following main components, as depicted in Figure 17.4:

- *Correlator Engine*: component correlating data source events and identifying potential threats based on a list of rules set by skilled operators;
- *Machine Learning* (ML)-based module: a component based on the application of ML algorithms for the identification of standard/anomalous behavioral models for the traffic originating from network data sources.

The Correlator Engine is mainly based on Apache Storm and Esper technologies. Apache Storm is a free open source software for distributed computing of real-time processes. Esper is a Complex Event Processing (CEP) component able to perform Event Stream Processing. This feature allows real-time or quasi-real-time detection of those events that match the stored rules. In RESISTO, rules can be updated in real time without the need of a Correlator Engine restart.

The Esper engine operates in a different manner compared to a database management system. Instead of storing data and performing queries on the stored data, it allows applications to store their own queries and directly launch them on the data. The processing mode is continuous and a reply is in real-time whenever the conditions contained within the query are met.

Esper provides two principal methods for processing events:

1. Event pattern,
2. Event stream query.

The first method is based on a language allowing specification of expression-based patterns for event matching. It analyses event sequences or a combination of event sequences based on timing factors. On the contrary, the second method offers the possibility to define queries allowing filtering, aggregation, and correlation (through join operators) as well as to analyze event streams. These queries follow the EPL (Event Processing Language) syntax. EPL is a declarative language implementing and extending the SQL-standard allowing rich expressions over events and time.

The Machine Learning (ML)-based module allows the detection of anomalous traffic situations compared with the daily recorded ordinary traffic intensity. Control flow historical data retrieved from past records are used as first baseline for training the learning procedure. Historical data repository can be increased continuously in order to tune the machine learning-based detector with respect to evolving data traffic curves.

More in details, the engine exploits a profile-based anomaly detection approach. This technique exploits the history of the normal network behavior, thus creating a normal network profile. Following this principle, "anomaly" is defined as a network behavior that is significantly different from the modeled one. One of the main advantages of profile-based approaches is that they do not require a model for the anomalous behaviors, thus allowing the detection of new and unforeseen anomalies. Moreover, the approach aims at designing an anomaly detection method which takes as input only control flow quantitative indicators such as the number of packets and bits. Let us note that this restriction on the kind of attributes exploited by the anomaly detection method is needed to fulfill the privacy preserving requirements.

Anomalies detected by the Events Correlator trigger the **Risk (Impact) Predictor**. The Risk Predictor evaluates and highlights the impacts of the potential exploit detected by the Correlator on the communication infrastructure and, mainly, on the services provided by the infrastructure. The Risk Predictor Engine is based on CISIApro 2.0 and acts at run-time on a CI model built according to different offline interlacing points of view:

- Under a reductionist perspective, each infrastructure is decomposed into a network of interconnected physical elementary entities and their behavior depends on the (mutual or not) interactions with the other reductionist elements;
- Applying a holistic approach, each infrastructure is modeled as a (logical) reality with its own identity, functional properties, and recognizable boundaries. It interacts with other similar entities according to reduced identifiable set of relationships. With such a perspective, it is easy to identify the roles that each infrastructure plays in a specific context;
- From a Service point of view, a Service Entity represents a logical element, conceptual or real, that provides an aggregate resource such as a QoS (Quality of Service) level.

Moreover, the Risk Predictor supports the decision-making process allowing a "What-If analysis" and thus simulating the application of countermeasures and reconfiguration and their impact on system resilience.

In parallel with the Risk (Impact) Predictor, the Correlator also triggers the **Workflow Manager** software engine in charge to guide the operator during the reaction and recovery phases. On the basis of the alarm type, the most appropriate workflow is selected and executed. A workflow is a conditional sequence of steps. Each step can specify a procedural action such as:

- Alert a security or technical team with an emergency message sent through the EWCF,
- Drive one or more physical actuators (e.g., lock physical gate),
- Carry out a complex O&M action on the Communication Network (e.g., activate a Virtual Network Security Function, isolate a faulty or attacked component, reconfigure a part of the network, disable a 5G slice, etc.).

The Workflow Manager inside RESISTO platform is an extremely effective tool for managing critical infrastructure security. It is based on a Business Process Model (BPM) engine for the configuration and execution of automatic or semi-automatic processes, consisting of sequences of actions and reactions, which can be triggered by a defined event. Given a certain alarm/event, it allows selecting and executing the most appropriate workflow, i.e., a conditional sequence of tasks.

The workflow execution is carried out via Activiti, an open-source workflow engine written in Java that can execute business processes described in standard

BPMN (Business Process Model and Notation) 2.0. The workflow is represented by an xml file, which is managed by the Activiti engine through its deployment in a dedicated database.

Complex actions on the Communication Infrastructure are performed by the **Orchestration Controller**. The Orchestration Controller is built around the concept of Software Defined Security (SDS) taking advantage of the Network Function Virtualization (NFV) and Software Defined Networking (SDN) paradigms of the underlying communication network. The Orchestrator Controller implements complex security functions and services composing less complex/primitive security mechanisms/functions acting on physical resources (i.e., network physical equipment) as well as on Virtual Network Functions in a NFV/5G perspective. The Orchestration Controller operates on a communication infrastructure already controlled by a telecommunication operator, so it works on top of a simple SDN Controller, on the northbound side, or on top of a more complex network Operational Support System (OSS).

The **Emergency Warning Communication** (EWC) function is activated when it is needed to send instant messages, targeted alerts, and operating instructions to specific categories of users that are present in a certain area where events like natural disasters, physical, or cyberattacks are occurring. In particular, rescue teams called to execute actions on the infrastructure can leverage on the received information. The EWC module includes a server application and either an Android one. The server exposes an interface towards the other modules of the RESISTO framework needing to communicate information concerning a physical-cyberattack to the intervention team that operates where the telecom infrastructure is located. The rescue team will leverage on the application information, both textual and visual. In particular, the position of points of interest or of the other team members is collected and visualized. The app is available on Android devices, including smartphones.

The EWCF service is implemented in a microservice architecture using Docker containers. The service has its own persistent storage and database where information regarding teams, users, and events are registered. The server is connected with an Android app installed on the user terminals of the team members. The same app can be connected to an IoT platform and will relay sensor values to the platform. In particular, the GPS position of the terminal will be collected in the IoT platform and in this way will be shared among the team members. A messaging platform is used instead to connect the terminals among themselves and with the main service that in turn receives specific messages from the Workflow Manager.

Nevertheless, the architecture is modular and can be adapted to use other external platforms.

As already explained before, **RESISTO** platform also include some state-of-the-art physical/cyber threat **detectors**.

Considering the emerging use of unmanned devices, UAVs (Unmanned Aerial Vehicles) or drones are nowadays more and more regarded as potential, human-driven, physical threats. Within RESISTO, mixed techniques involving low-cost radars combined with acoustic sensors are implemented to detect small airborne objects and moving targets. The specific **airborne threat detection system** consists of a set of tools designed and developed to detect the presence of small UAVs that may constitute airborne threats and provide alarm signals. The system can be also deployed in small unprotected areas, such as antenna telecom parks, providing additional situational awareness and perimeter defense against low-flying aircrafts.

The radar sensor currently adopted in RESISTO is a Doppler radar able to detect and track fast moving, small targets even in harsh conditions (dusk, rain or snow) at a range of several kilometers.

Detection based on the acoustic signal emitted by the UAV is based on a low cost, low power array of high sensitivity dynamic microphones. Acoustic sensors have many advantages that include non-line-of-sight, omni-directionality, passiveness, low-cost, and low-power, and play a potential key role in situational awareness. Moreover, while the equivalent radar cross section of UAVs can be rather small, due to both their small size and the electromagnetic properties of their constituent materials, their acoustic signature is directly related to the acoustic wave originating from both the engine and the propeller rotation. The acoustic microphone arrays are used as a second sensor modality to detect broadband acoustic emissions from approaching targets. In particular, rotary wing UAVs can be detected by exploiting the tonal components of the spectrum of the incoming acoustic wave related to the propellers' rotations (in the 20 Hz–2 kHz range).

Target's detection and angle of arrival estimation adopt advanced signal processing and machine learning techniques making use of radar and acoustic signal features in both the time-domain and the frequency-domain. The above sensors and tools can be either used separately or in combination, through a multiplexing console and a computer, physically connected to the sensors, that performs signal and data processing.

Video and Audio sensors are widely used in surveillance operations and protection of critical infrastructures. Intelligence algorithms are applied in audio and video streams for the real-time detection of events for the early identification of illicit activity. Pattern recognition and machine learning techniques are used to extract acoustic events (i.e., gunshot, screaming, glass breaking) or to classify persons, vehicles, and other objects that are moving within the surveilled area. Both the **audio and video analytics** modules form an intelligence **surveillance system** where the security operator is notified with an alert about the suspicious activity accompanied with important information such as location (source) and type of the event, detected objects, etc. This intelligent process reduces the effort of the operator by monitoring in a 24/7 base a huge number of sensors.

The Intelligent Audio Analytics Component (AAC) allows the detection of abnormal behavior regardless of the field of view, while also allowing the triggering of the system with the occurrence of predefined keywords. The solution implements well-established methods from the fields of audio coding, machine learning, and speech recognition and allows efficient operation on low-cost power-limited devices (or embedded systems) for the detection of screaming, glass breaking, and gunshots within the environment. The Video Analytics Component (VAC) provides the necessary functionalities for visual surveillance analytics, aiming to identify and provide methods that abstract the information of interest contained in video surveillance streams.

The design of Intelligence Surveillance System including both the AAC and VAC is based on two different processing levels, one with reduced processing capability (infrastructure based on embedded system, i.e., Raspberry PI 3) and one with enhanced processing capability (i.e. GPU-enabled computers or servers). Based on the application, audio or video, several components will be deployed on each level of processing.

Additionally, within RESISTO project, audio and video analytics modules are enhanced with some other components in order to support the smooth operation, logging, and correlation of the events.

Smart Spectrum Surveillance is a set of tools, like RADIOFILTER and RAN (Radio Access Network) MONITOR, being developed in the context of RESISTO project, which makes use of IoT Radio Frequency (RF) sensors for the detection of physical events/threats in telecom critical infrastructures.

RADIOFILTER is a stand-alone system used for the detection of non-authorized Access Points (AP), Bluetooth, and WiFi devices as well as connections in manned facilities as well as intrusion detection in unmanned facilities.

RANMONITOR is a stand-alone system used for the detection of IMSI-catchers/rogue base stations, misconfigured small cells, and interferences to the cellular network (intentional and non-intentional).

IoT Sensor networks may gather sensitive data or be used by a malicious adversary to conduct attacks. Therefore, security is a key concern for such networks, and for that reason, particular attention is paid to secure the sensors themselves. In RESISTO, the solution proposed for securing sensors is based on the premise of having a secure boot and up-to-date software in the hardware platform of each used sensor through secure periodic firmware updates. In order to secure an IoT Sensor Network, the sensors periodically poll a Firmware Update Server to query if there are secure firmware updates available. In case there are, after mutual authentication (involving an Auxiliary Authentication Server), a digitally signed firmware image is downloaded from the Firmware Update Server and its integrity and authenticity verified. To strengthen integrity verification, blockchain technology is applied. If integrity and authenticity tests are passed, the sensor can install the new firmware version. The sensor-server connection is further secured through the use of two-factor authentication.

The use of Open Source INTelligence (OSINT) techniques can help better understand potential threats surrounding a telecom organization or a specific sector, by crawling and learning from publicly available sources. The main features of the crawler are:

- Identification of Common Vulnerabilities and Exposures (CVE) that could be found on devices that can be exposed on the operator network;
- Detection of potential misconfigurations and known vulnerabilities.

The crawler stores the information concerning the type of device, the software running on it, the potential misconfigurations/vulnerabilities based on the knowledge of the operator network.

Within RESISTO, several threat intelligence sources and OSINT platforms are considered and crawled:

- The Computer Incident Response Center Luxembourg (CIRCL), a Malware Information Sharing Platform framework-based OSINT that collects threat intelligence events;
- The Instrument de Veille sur les Réseaux Extérieurs (IVRE), also known as Dynamic Recon of UNKnown networks (DRUNK), open-source framework for network recon;
- Other sources of data about vulnerabilities that can be found on Twitter;
- A machine learning platform to process the events collected by the crawlers.

17.5 Long- and Short-term Control Loops Interaction

Long- and Short-term Control Loops interact with each other by means of Resilience Indicators (RIs). The RIs have been selected in order to describe the main features of the typical resilience curve (Figure 17.5) describing the evolution of a system function performance (or provided service) vs. Time when facing a specific event type.

The selected RIs are:

1. RI_1: maximum function performance loss expressed in percentage;
2. RI_2: elapsed time between the event occurring and the recovery action beginning;

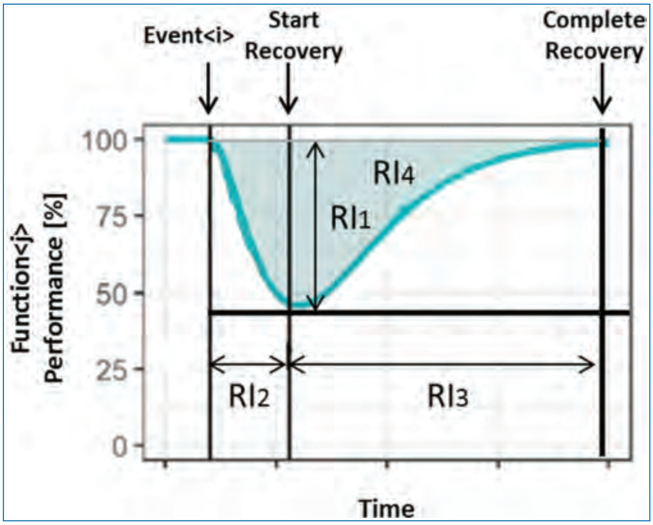


Figure 17.5. Resilience curve and Resilience Indicators.

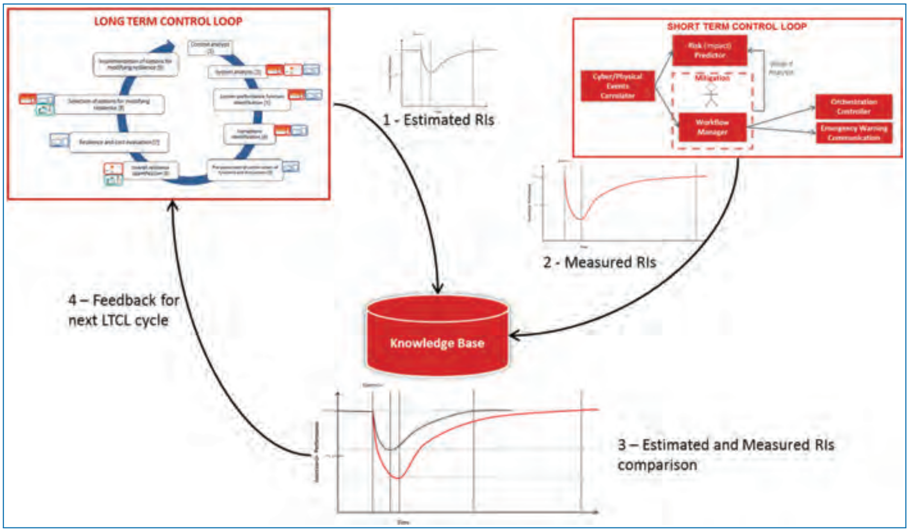


Figure 17.6. Long- and Short-term Control Loops interaction.

3. RI3: elapsed time between the recovery action beginning and the complete performance recovery;
4. RI4: total performance loss from event to complete recovery (colored area in Figure 17.5).

The interaction between Long- and Short-term Control Loops can be explained in 4 steps as explained in Figure 17.6.

Step 1: RIs estimation

During the last LTCL steps, the process:

- Characterizes (quantifies) CI «as is» resilience (step 6);
- Identifies the most critical couples [function; (threatening) event] showing RIs not in line with required Service Level Agreement (step 7);
- Selects interventions on CI in order to improve resilience for most critical couples (function; event) estimating RIs in the new «to be» configuration (step 8);
- Implements interventions (step 9).

So, at the end of each LTCL cycle, Estimated RIs are stored in a Knowledge Base (KB).

Step 2: RIs measurement

STCL operators, facing Event<i> type, measure actual RIs and store them in the KB.

Step 3: Estimated vs. Measured RIs comparison

Estimated and Measured RIs are compared to verify if the expected resilience is actually in place (no significant deviations) or not.

Step 4: Estimated vs. Measured RIs comparison

Detected significant deviations provide feedback for a next LTCL cycle to improve Critical Infrastructure resilience or estimation methods if needed.

17.6 Conclusions

RESISTO proposes a complete and integrated framework to cover offline Identification and Prevention activities as well as Detection, Response, and Recovery on-line activities. RESISTO promotes a unified approach to face physical, cyber, as well as combined physical/cyber threats to Communication CIs in order to provide a complete situation awareness and impacts evaluation allowing resources optimization and improving recovery actions efficiency.

RESISTO encompasses security analysis in a wider Risk and also Resilience analysis and management integrating both physical and cyber aspects.

RESISTO approach is scalable, developed in the context of Communications but easily applicable to different kinds of CIs.

The proposed framework is modular and based on very versatile technologies so easily adaptable to face the continuous evolution of physical and cyber threats and continuously improve the CI resilience.

RESISTO also includes a wide set of physical and cyber threatening events detectors based on state-of-the-art technologies (Machine Learning, blockchain, etc.); they could be employed in different contexts as stand-alone components as well as in integrated configurations.

Acknowledgments

The current work has in parts been supported by the EU projects RESISTO (Grant No. 786409) on cyber-physical security of telecommunication critical infrastructure.

References

L. Carlson, B. Haffenden, G. Bassett, W. Buehring, M. Collins, S. Folga, F. Petit, J. Phillips, D. Verner and R. Whitfield, 2012, "Resilience: Theory and Application," Argonne National Lab (ANL), Argonne, IL (United States).

W. Chang et al. (2019): Shiny: web application framework for R. Version 1.4.0. Available online at https://cran.r-project.org/web/packages/shiny/index.html.

ETSI OSM Northbound API—https://www.etsi.org/deliver/etsi_gs/NFV-SOL/001_099/005/02.05.01_60/gs_nfv-sol005v020501p.pdf.

C. Foglietta, C. Palazzo, R. Santini, S. Panzieri, "Assessing Cyber Risk Using the CISIApro Simulator," in Critical Infrastructure Protection IX : 9th IFIP 11.10 International Conference, ICCIP 2015, Arlington, VA, USA, March 16–18, 2015 (pp. 315–331).

Fehling-Kaschek, Mirjam; Faist, Katja; Miller, Natalie; Finger, Jörg; Häring, Ivo; Carli, Marco et al. (2019): A systematic tabular approach for risk and resilience assessment and Improvement in the telecommunication industry. In Michael Beer, Enrico Zio (Eds.): Proceedings of the 29th European Safety and Reliability Conference (ESREL 2019). ESREL. Hannover, Germany, 22–26 September 2019. European Safety and Reliability Association (ESRA). Singapore: Research Publishing Services, pp. 1312–1319. Available online at https://esrel2019.org/files/proceedings.zip.

Häring, Ivo (2015): Risk Analysis and Management: Engineering Resilience. 1st ed. 2015. s.l.: Springer-Verlag. Available online at http://ebooks.ciando.com/book/index.cfm/bok_id/2008091.

Häring, Ivo; Ebenhöch, Stefan; Stolz, Alexander (2016a): Quantifying Resilience for Resilience Engineering of Socio Technical Systems. In European Journal for Security Research 1 (1), pp. 21–58. DOI: 10.1007/s41125-015-0001-x.

Häring, Ivo; Gelhausen, Patrick (2018): Technical safety and reliability methods for resilience engineering: Taylor and Francis Group, pp. 1253–1260. Available online at https://doi.org/10.1201/9781351174664, https://www.taylorfrancis.com/books/e/9781351174664, checked on 10/17/2019.

ISO 31000, 2018-02: Risk management – Guidelines. Available online at https://www.iso.org/standard/65694.html.

ISO 31010, 2019-06: Risk management – Risk assessment techniques. Available online at https://www.iso.org/standard/72140.html.

Nadjaran Toosi, R. Mahmud, Qinghua Chi, R. Buyya, "Management and Orchestration of Network Slices in 5G," Fog, Edge and Clouds.

OMG (Object Management Group), Business Process Model and Notation (BPMN) Version 2.0.2, https://www.omg.org/spec/BPMN/2.0.2/PDF.

RESISTO (2020): Resilience enhancement and risk control platform for communication infrastructure operators. EC Grant agreement ID: 786409. Available online at https://cordis.europa.eu/project/id/786409.

USA NIST (National Institute of Standards and Technology), Framework for Improving Critical Infrastructure Cybersecurity, Version 1.1, April 16, 2018.

DOI: 10.1561/9781680836875.ch18

Chapter 18

Manage Security on 5G Communication Networks: The Software Defined Security Paradigm

By Luca Baldini, Marco Carli, Giuseppe Celozzi, Federico Colangelo, Alessandro Neri, Cosimo Zotti and Federica Battisti

This chapter is devoted to the description of the interaction between the new communication system (the 5G framework) and the emerging security paradigm, known as Software-defined Security. It can be considered as a new security model to be applied for the management of communication networks, in which security aspects are implemented, controlled, and managed at software level. The main objective is providing a clear-cut, understandable, and upgradable security model, in which novel algorithms and solutions can be integrated and optimized.

18.1 Introduction

The next generation of mobile communications, also known as 5G, is nowadays at the developmental phase and is expected to be on the market in the upcoming years. It is expected to be faster, more flexible, and secure than the existing technologies, allowing a full use of mobile network for interconnecting machines, objects, and devices. From a general point of view, the quality of a communication-based service nowadays is strongly connected (limited) to the available resources of the used telecommunication infrastructure. With 5G, the quality of the service will drive the amount of resources devoted to that service [1]. Those resources, thanks to several architectural novel concepts (e.g., network slicing), could be given by different Telecommunication (TLC) providers. The new architecture will allow up to 10 Gbps of data rate, latency of 1 ms, and up to 100× in the number of connected devices per unit area compared with the 4G LTE. However, the potentiality of 5G will be fully exploited only if security will be achieved.

On one hand, the security solutions adopted in the previous generations of wireless communications will be adopted and improved. On the other hand, new challenges will arise, and they must be addressed with 5G-specific security mechanisms. In this chapter, the novel building blocks of 5G are revised and the security challenges described. In particular, the concepts of Software-defined Network (SDN) and Network Function Virtualization (NFV) are introduced, since they are enabling technologies for 5G and thus critical to understanding the security landscape, SDN, NFV, and 5G.

18.1.1 Sotfware-Defined Network

SDN is a network paradigm aiming to allow agility and flexibility to Communication Networks (CN). The goal of SDN is to improve network control by enabling enterprises and service providers to quickly adapt to novel requirements modification: this is obtained by decoupling network control and forwarding functions, that is, by enabling programmable network control and the underlying infrastructure to be abstracted from applications and Network Services (NS).

The features of SDN are based on the following principles [4]:

1. Separation of control and data planes:

The removal of the control plane from network devices and its implementation in an external SDN controller significantly reduces the complexity of network devices, making them simpler and cheaper than devices used in actual CN whose distributed control plane functionalities require prone-to-errors long software implementation and are defined by many RFCs.

2. Logically centralized control:

Control decisions are made on an up-to-date global view of the network state rather than distributed in isolated behavior at each network hop. With SDN, the control plane acts as a single, logically centralized, network operating system in terms of both scheduling and resolving resource conflicts, as well as abstracting away low-level device details (e.g., electrical vs. optical transmission). The SDN Controller summarizes the network state for applications and translates application requirements to low-level rules. The controller may not be physically placed in one element. For coping with performance, scalability, and/or reliability issues, the logically centralized SDN Controller can be distributed so that several physical controller instances cooperate for controlling the network state and supplying applications.

3. Programmability of NS:

With programmable control, a client may exchange information with an SDN controller, either by discovery or negotiation prior to the establishment of a service, or during the lifetime of a service. The network is programmable through software applications running on top of the Controller [communicating via Northbound Application Programming Interfaces (API)], which in turn interacts with the underlying data plane devices. The SDN controller, having the complete knowledge of the network topology, may control a wide range of network devices within its administrative domain. By providing APIs, SDN enables the deployment of novel networking applications, e.g., traffic engineering. As is well known, CN devices are proprietary and closed, making it hard or impossible to develop innovative network applications.

4. Open interface:

The concept of open interface relies on the separation between functions and interfaces requiring the interfaces to be public and open to community definition. One value of SDN lies in the expectation that the Control to Data-Plane Interface (Southbound Interface) is implemented in an open, vendor-neutral, and interoperable way. In the absence of a standard open interface, one of the main SDN advantages—the interchangeability of network devices and control planes—would be taken away. In an SDN environment, the SDN controller plays a pivotal role by managing data flows into the switches/routers "below" (via southbound APIs) based on the policies and rules defined by the applications and business logic "above" (via northbound APIs). The SDN Controller translates instructions or requirements from the SDN Application layer into commands to be delivered to networking components. Conversely, the SDN controller extracts information about the network from devices and shares back an abstract view of the network to the SDN Applications, also including information about events and statistics.

18.1.2 NFV

A peculiarity of 5G infrastructure is the adoption of a virtualized network infrastructure, that is, the use of software for implementing network functions usually provided by dedicated hardware devices or components. The softwarization of these functions allows their deployment in cloud architectures and their relocation to the network edge. In this way, it is possible to cope with the severe low latency requirements required by critical 5G applications.

NFV started in 2012 as an industry initiative to virtualize network functions, thus allowing to add, move, or change network functions at the server level in a simplified provisioning process. Among others, ETSI has been a strong support and driver of the NFV standardization.

- In more details, the ETSI NFV MANO (Management & Orchestration) model [13], represented in Figure 18.1, defines as main functional blocks: NFV Orchestrator (NFVO) On-boarding of new NS, Virtual Network Function Forwarding Graph (VNF-FG), and Virtual Network Function (VNF) Packages.
- NS life-cycle management (including instantiation, scale-out/in, performance measurements, event correlation, and termination).
- Global resource management, validation, and authorization of NFVI resource requests.
- Policy management for NS instances.

VNF Manager (VNFM):

- Life-cycle management of VNF instances.
- Overall coordination and adaptation role for configuration and event reporting between NFVI and the Element Management System (EMS).

Virtualized Infrastructure Manager (VIM):

- Controlling and managing the NFVI computing, storage, and network resources, within one operator infrastructure sub-domain.
- Collection and forwarding of performance measurements and events.

The ETSI MANO model provides the framework for provisioning VNFs and managing the NFV infrastructure. It also supports components within NFV infrastructure communicate with existing OSS/BSS systems. In fact, MANO systems oversee managing virtualized infrastructure, such as cloud systems, communication

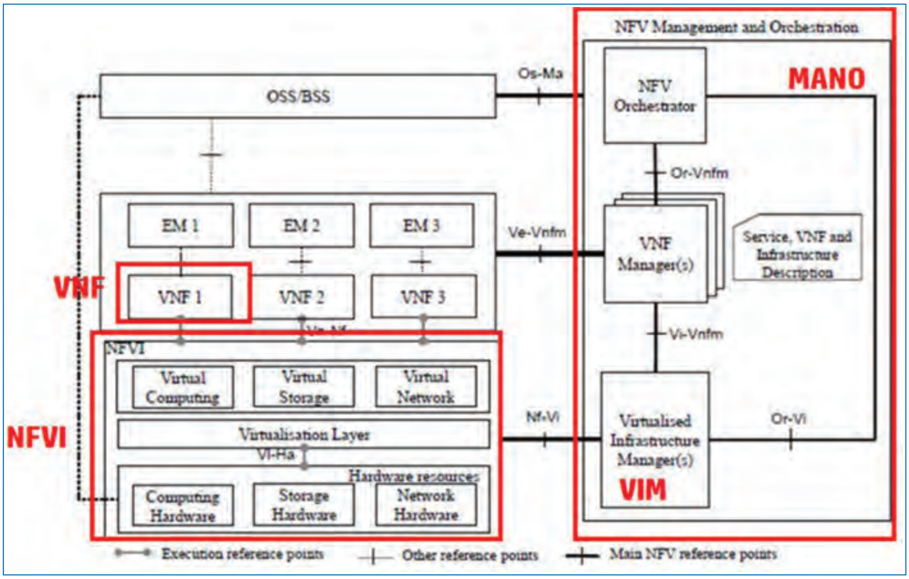


Figure 18.1. ETSI NFV MANO architecture.

and network infrastructure, VNF (implemented as virtual machine or container images), and the full life cycle of all these components.

The following list describes the interfaces that a VIM uses it to allocate, manage, and control the NFVI resources:

- Or-VNFM defines the communication procedures between NFVO and VNFM (i.e., VNF instantiation or other VNF life-cycle-related information flow).
- Or-Vi is a reference point used for exchanges between NFV Orchestrator and VIM and supports the interfaces Software Image Management and Virtualized Resources Information Management.
- Vi-VNFM defines the communication procedures between VIM and VNFM, such as resource update request for VM running a VNF.
- Vn-Nf is the only reference point not having a management functional block as boundary. This reference point is meant to communicate performance and portability needs of the VNF to the infrastructure block.

As can be noticed, even if the NFV framework identifies reference points between the NFVO and the VIM and between the VNFM and VIM, the ETSI model does not provide API specifications for the corresponding interfaces. The assumption is that the overwhelming majority of VIM implementation is based on OpenStack, and thus, the APIs exposed by a VIM are those specified by this open source community [8].

Use of RESTful APIs for management and orchestration is specified by ETSI, except for the VIM northbound interfaces where use of OpenStack APIs is assumed, as noted before. ETSI APIs specifications are available both as Text and Tables and in OpenAPI format.

18.2 Security of 5G

18.2.1 Core 5G Security Topics

The fact that 5G is designed to be a platform for a wide range of new user groups and applications does not automatically mean that it is necessary (or even desirable) for the 5G network to carry all security responsibility and related costs. On the other hand, 5G networks clearly can provide some highly valuable security services. Besides the isolation/slicing itself, many other examples of network-enabled security as a service will be attractive to multiple user groups, including network-enforced security policies, authentication, key management, and data security services [9].

Technological developments in the configuration and deployment of 5G use cases to be deployed in commercial networks are needed from an economic, competitive, and performance perspective in order to realizing critical machine-type communication or applications which belong to latency-sensitive autonomous systems 5G use cases. Which means that it shall become the reliable and trusted innovation platform for businesses and organizations to build and deliver new added-value services, but not limited to, in fact the aim is to be an enabler for digitizing and modernizing critical national infrastructures such as energy, transport, etc. In order to achieve this greater availability and improved assurances of secure communication, services must be achieved.

Protection of the privacy of 5G users based on that data sent over the system is always confidentiality and integrity protected. This complex task is achieved by combining and coordinating security controls across different domains in telecommunication networks, including radio access (e.g., radio unit, baseband units, antennas), transport networks (e.g., optical equipment, Ethernet bridges, IP/MPLS routers, SDN controller), packet core (e.g., MME, S-GW, PGW, HSS), network support services (e.g., DNS, DHCP), cloud infrastructure, and various management systems (e.g., network management, customer experience management, security management). Telecommunication networks consist of four main logical network parts: radio access network, core network, transport network, and interconnect network.

The Radio Access Network (RAN) is an instance of access network and a major part of modern telecommunications. There are many types of access networks, such as the 3GPP access networks: GSM/GPRS, UMTS, EUTRAN, NG RAN (5G), satellite, and non-3GPP access networks: WiFi or fixed (wired) access network.

The core network can provide several services to subscribers that are connected via the access network into the core, such as telephone calls and data connections. The transport network keeps the access network connected with the core, and the base stations within the radio access network connected with each other. The interconnect network connects different core networks with each other. Telecommunication networks transfer voice and data across the globe with high quality and consistency [10].

Each network part can be subdivided further into network planes, each carrying different class of traffic: signaling traffic, user payload traffic, and management traffic. The signaling plane transports messages that are used to control user sessions. For instance, the contents of a call or web page is referred to as user plane or user payload. The management plane provides functions for monitoring, troubleshooting, configuration, and optimization of networks.

5G core resilience concepts are, for example, related to network slicing that can isolate groups of network functions from other functions, in fact an organization with high security constraints can use a dedicated complete mobile network. Another example can be to differentiate and isolate low-priority IoT devices coming in high numbers to avoid interference.

SBA is another architectural concept that enhances resilience by means of software and cloud-based technologies that improve on the more static and node-centric designs of mobile networks. Thanks to this new paradigm functions can easily be scaled depending on traffic load and can be independently replaced, restarted, or isolated when failing or under attack.

Transport networks provide high-speed low-latency connectivity services between all 5G network functions. Consequently, the availability of transport networks is directly related to the availability of the 5G system and the services it provides. To ensure availability of transport services during node failure, cable or fiber breaks, or overload events, transport networks can employ various technical solutions as well as considerations during network design, including geo-redundant paths, link redundancy solutions, path redundancy mechanisms, high-availability configuration, traffic segmentation mechanisms (e.g., VLAN and MPLS), DDoS detection and mitigation, IPsec- or MACsec-based tunnels.

5G system provides secure communication for devices and for its own infrastructure including links such as front haul between distributed and central units of base stations, backhaul between access and core network, and network domain links between core network nodes. The security design principles of 4G system has been evolved to better meet the needs of new use cases. In particular, the new SBA for core network communication takes threats from the interconnect network into account from the start.

The 5G system includes eavesdropping and modification attacks protection. Signaling traffic is encrypted and integrity protected. User plane traffic is encrypted and can be integrity protected.

18.2.2 5G Security Standardization

5G RAN will have a significant impact on security, such as SDN, NFV, and edge computing as telecommunication networks are evolving towards virtualization, IoT, and Industry 4.0. The 5G 3GPP standard is agnostic, in that it is flexible enough to allow for different types of physical and virtual overlap between the radio access network (RAN) and core network, for example, from a remote device to the Core network. The separation of functions between RAN and core raises questions about competitiveness and performance.

3GPP does not typically standardize application services (such as Internet applications) since they are out of scope of 3GPP's connectivity focus. There are, however, a few exceptions: telecommunication networks have traditionally provided the possibility for two devices to connect to each other with the support of the network (e.g., to set up voice calls). In 4G networks, voice calls are set up using voice over LTE (VoLTE) service on top of the connectivity service. VoLTE uses the IP Multimedia Subsystem (IMS) also standardized in 3GPP; similar voice service is also planned for 5G. Furthermore, 3GPP standardizes the security to support these services.

3GPP standards also cover some aspects of machine type communications and IoT. Here, the focus is to provide the devices with connectivity. Consequently, the 3GPP standards cover efficient means to provide these devices with an IP point of presence. Any security issues related to the actual application is considered out of scope and needs to be taken care of over the top. For example, 3GPP's 5G system can provide a temperature controller in a refrigerated goods wagon of a train with IP connectivity, but seen from the general 5G view, the authentication of the management traffic to the controller must be addressed over the top, since the IP address may be accessible via the Internet, so anyone could send messages to the controller.

The most important security services provided by 3GPP standard are aiming to safeguard the connectivity for users, and the service availability and charging by the operator of the network.

3GPP does not standardize how 5G system functions are implemented but defines security assurance specifications to secure interoperability between the functions required to provide network connectivity. The choice to use single physical servers (physically isolated and separated) or implemented as virtual machines (VMs) in a cloud or virtualized environment (shared hardware) is up to implementation and operator deployment choices (economics). Virtualization and cloud deployments are only lightly touched in the 3GPP specifications although their security is crucial for 5G services. Virtualization and cloud though are handled in ETSI ISG NFV (European Telecommunications Standards Institute, Industry Specification Group, Network Functions Virtualization) and ONAP (Open Network Automation Platform). Furthermore, several details are not taken into consideration and standardized, but are left for implementations and deployments. The same is true for digitized society and industrial IoT aspects that are not related to the radio access connectivity and are considered out of scope for 3GPP.

3GPP standards don't strictly specify how RAN and Core functions should be separately deployed which means that within a single physical network, different configurations for different 5G use cases are possible and that all logical networks are running over one physical network can have different configurations. For functions implemented in a traditional non-virtualized fashion, 3GPP, in cooperation with GSMA, develops security assurance specifications, which sets requirements for some implementation aspects [11].

3GPP's security standardization group (SA WG3) has completed the first version of 5G security standard in March 2018 (3GPP TS 33.501). Ericsson has been one of the main drivers and contributors to the 3GPP (SA WG3). 5G security study (3GPP TR 33.899) which ran from June 2016 to August 2017 is a highly relevant reference.

Assume that in some use cases, vehicle/road safety would be dependent on 5G network security. What does this imply? Today, safety-related car systems need to follow very comprehensive standards, such as ISO 26262. This is a 10-part standard, where, for example, part six covers safety related to software. Similarly, the healthcare sector is governed by standards such as ISO 27799 and, in the US, the HIPAA (Health Insurance Portability and Accountability Act). For smart grids, demonstrated compliance with standards from the IEEE (Institute of Electrical and Electronics Engineers), the IEC, and the NIST (National Institute of Standards and Technology) may apply. If 5G security becomes a critical link in the control loop of all of these applications, would it imply that 5G networks need to be certified against (parts of) all these standards?

3GPP's 5G system standards security mechanisms are based on former 4G security mechanisms, but nevertheless include enhancements regarding encryption, authentication, and user privacy. 3GPP security mechanisms provide reliable links for non-malicious bad radio conditions. DDoS and radio jamming are not part of the standard, and protection is left for implementation and deployment, example of solution to those issues are re-route traffic to other base stations in case of jammed based station or selective dropping/throttling in case of DDoS. This means that 5G standards will only be part of a much bigger picture.

The problem of securing of mobile networks has been studied thoroughly through the years. Thus, 5G will inherit some of the work that has been done in securing the first four generations of mobile networks [2]. There are, however, novel, 5G-specific challenges that arise due to some of the key technologies and services that will be included in the 5G network.

A threat in the 5G network vision is tied to IoT devices and their impact on DDoS attacks. The volume of DDoS attacks has already been dramatically increasing in recent years, with attacks peaking at 70 GB/s. This trend has been ascribed to the rising popularity of IoT devices. Such devices often have poor security, due to either constraints of the platform, performance limits or simply a lack of focus on security. This scenario enables the creation of large botnets, which in turn enable massive DDoS attacks. Massive Machine Type Communication is one of the key use cases of 5G, and its implementation will foreseeably cause a large growth in the number of IoT devices connected to the network.

SDN and NFV also play a fundamental role in the threat landscape since they are both central part of the 5G architecture and not yet fully mature.

Concerning SDN, the softwarization of the network architecture can be considered a double-edged sword: if, on the one hand, many architectural problems that plagued network security (e.g., security network function placement) can now be addressed effectively, on the other hand, the whole network stack is now exposed to software attacks. Specifically, vulnerabilities in the software stack of the data and control plane can be exploited by malicious users with detrimental effects on the whole network. This is especially true if the attacker manages to compromise an SDN controller, as this effectively grants full control over the relative network segment.

Perhaps even more impactful from the 5G point of view are security issues of NFV. More specifically, virtualization of different services is enabled by the soft isolation (i.e., software isolation) provided by hypervisors. Vulnerabilities that enable evasion (i.e., breaking soft isolation) are thus critical and relevant for 5G, given its strong reliance on virtualized network services. Even when best practices are followed, soft isolation can be inherently insecure, as recent discoveries in the security of CPUs, such as the Specter and Meltdown vulnerabilities, have demonstrated. These vulnerabilities have meaningful impact on 5G security and especially on Network Slicing. Network slicing is seen as a key functionality of 5G networks, yet its implementation is still not completely defined. It is, however, possible to foresee that slice creation algorithms will be a potential threat surface. More specifically, a key aspect of slicing security will be avoiding the inclusion of untrusted network devices into slices with critical security requirements. Since virtual isolation cannot be considered secure at the time being, slicing algorithms will need to be able to provide physical isolation, at least for the most critical network traffic.

However, despite the novel attack surface introduced by SDN and NFV, these technologies introduce notable opportunities for security. Centralized network control and virtualized security functions can eliminate some of the biggest flaws in legacy network security functions. Examples of applications that benefit greatly of these technologies are anomaly detectors that can leverage a reliable and complete analysis of network data and DDoS mitigation system that can be deployed adaptively based on the attack patterns. However, while this scenario appears appealing, it should be noticed that these capabilities have no software implementation to date. Effort is needed in order to leverage these opportunities. More specifically, how these features are implemented is a critical point. New vulnerabilities could be introduced, or performances could be degraded, if the development and validation processes are not properly handled. Currently, multiple proposals of security mechanisms are available in the literature. While these contributions are helpful to understand and compare the performances of various algorithms for given problems (e.g., DDoS mitigation), a coherent framework in which these solutions are integrated is lacking. On the other hand, a common framework is needed in order to proceed to the deployment of these technologies into the real world. Without a proper framework, modern technologies could incur in the same pitfall of legacy networks: multiple standards that hinder the core advantages of these technologies, such as upgradability, scalability, etc. Conversely, a clearly defined, common framework allows to integrate and upgrade new security algorithms as they are developed, enabling better security of the overall system.

18.3 Software-Defined Security

Software-Defined Security (SDS) is a security framework that embraces the principle on which SDN, NFV, and 5G are based. SDS is based on a clear model of the security workflow, providing a framework to integrate existing security solutions as well as integrate novel algorithms to address future security challenges.

SDS is inspired to the approach of SDN, as the name suggests: one of the key advantages of SDN is the decoupling of the network functional planes. This separation, together with the softwarization of the control plane, enables the application of the classical software development workflow, where a solution can be developed once and then reused or updated, to networking. SDS shares a similar vision to develop a clear security workflow, composed of well-defined and separated logical functions, where security functions can be integrated and orchestrated. SDS divides the security workflow in three logical blocks: risk assessment, mitigation, and orchestration. The key idea is to divide the security workflow into three logical steps: first, assess the current situation, evaluating the need for intervention based on the risk associated to a configuration. Second, evaluate possible strategies for reaction and rank them considering the impact on the system's Key Performance Indicators (KPI). The countermeasure is then selected, either by an operator or by an automated system. Third, apply the countermeasure, translating a high-level action into a low-level set of commands that are issued to the single apparatuses.

More specifically, three logical entities can be distinguished:

- **Risk assessment** has the purpose of monitoring and evaluating the current status of a system, leveraging a variety of data from sensors as well as processed output. More specifically, risk is evaluated in terms of impact on the infrastructure KPI. This includes also the effects of cyberattacks. For example, DDoS attacks can be understood in terms of its effect on the Availability KPI. Anomaly detection algorithms can be included in the risk assessment as well. In this way, risk assessment provides system-wide situational awareness for complex scenario. Operators can then evaluate the impact of a certain event (e.g., anomaly, fault of a device, etc.) considering also the potential cascading effects and the correlation between various other events. This is particularly useful in case of a multiple-stage attack. Risk assessment also provides an evaluation of the impact of a modification of the configuration and thus supports the mitigation module in evaluating the available actions.
- **Mitigation** has the purpose of determining actions to mitigate risk and impact on the infrastructure's KPI. The idea is that an action that mitigates the effect of an attack often has a detrimental effect on performances. For example, during a jamming attack, the radio system could be made to use a more resilient modulation scheme (e.g., a QAM modulation with smaller constellation), improving robustness to noise but deteriorating the throughput. Countermeasures are selected leveraging the Risk assessment module. Specifically, multi-objective optimization is performed to balance between impact on KPIs and benefits of an action, submitting the risk associated with the post-action configuration to the risk assessment framework. A list of actions (and the correspondent gain in risk and impact on performances) is then presented to the operator which elects which action is most fitting for the current situation. Alternatively, meta-heuristics can be specified to fully automate the process.
- **Countermeasure orchestration** has the purpose of actuating the actions selected by the mitigation module, abstracting the high-level action or policy from the low-level commands that are needed to actuate the action. The main

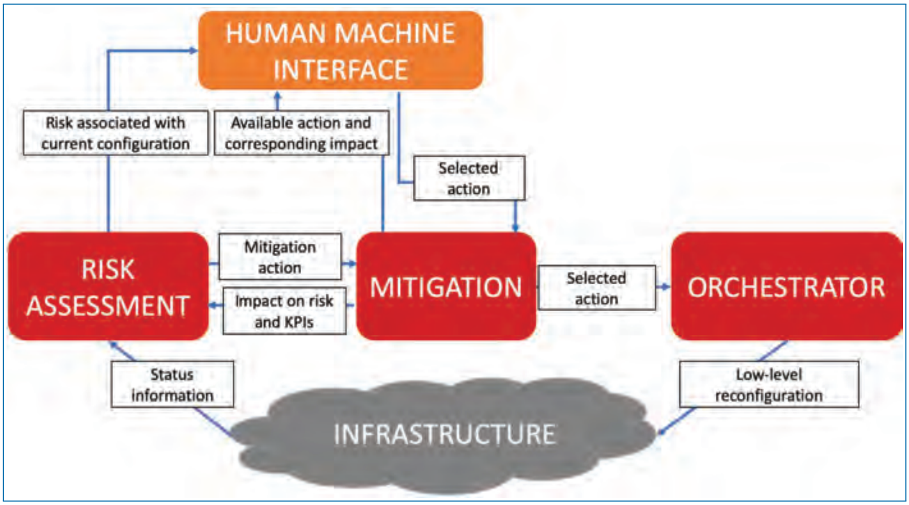


Figure 18.2. SDS conceptual architecture.

idea is to enable abstraction in the actuation of mitigation actions, removing the associated complexities. Furthermore, the orchestrator provides a way to optimize the application of a countermeasure. As an example, re-routing a flow (a potential consequence of a mitigation action) in a network topology can be achieved in many ways, depending on the complexity of the topology. In this case, multi-objective optimization can be applied to find the optimal route for every flow, depending on the associated service.

The SDS conceptual architecture is shown in Figure 18.2. The Human Machine Interface component can be considered optional, as the system can operate in a completely autonomous fashion by leveraging meta-heuristics.

Key benefits of this approach are:

- **Scalability:** Scalability is inherited from the underlying networking technology. Basically, scaling security function and services (e.g., deploy more instances of anomaly detector to address a traffic burst) can be done seamlessly in a virtualized environment.
- **Upgradability:** Security is a fast-paced field. Novel vulnerabilities are discovered continuously, and defense must be updated and upgraded accordingly to respond to threats. Anti-virus software offers an example of this problem since its effectiveness rely on steady database update. While the scenario was different for legacy networking, given the constraint of being SDS based on a clearly defined, modular, architecture, it is easy to replace/upgrade any of the logical functions. This is due to the fact that SDS entities are built with

modularity in mind, i.e., as a way to integrate different functionalities in a coherent manner, with clear endpoint.

- **Understandability:** Making security understandable is crucial. Misconfigurations and various errors are a leading cause of security breaches. As an example, in 2018 approximately 35% of the Healthcare sector breaches were caused by misconfigurations [36]. This type of incident can be mitigated by building workflows and systems that are easy to understand and act on, so operators can have better awareness of the ongoing security status of the infrastructure. Crucially, the SDS framework does not only detect threats but allow operators to properly understand actions in terms of their most important consequences.
- **Virtualization and abstraction of the security functions:** Being able to virtualize and abstract further reduces the threats coming from misconfigurations. Basically, security actions are expressed as a set of high-level actions (e.g., separation policies for network traffic or blacklisting). The SDS framework takes care of translating an abstract action into a set of low-level configurations to be applied to the nodes.

  It should be noticed that it is possible to apply SDS in a variety of contexts (e.g., smart grids, physical infrastructures) to develop a unified model. As a matter of fact, the various security domains (i.e., physical, network, system…) cannot be treated as independent, as they can interact. SDS is designed to support a holistic approach to security, focusing on virtualization of resources.

One of the most important features of SDS is its ability to quickly integrate innovations in algorithms and models to improve the security of the system. For example, a better DDoS mitigation algorithm will translate directly into better security and resiliency. The same goes for risk models and multi-objective optimization algorithms.

18.3.1 Case Study: Secure Network Slicing

SDS can be deployed in a variety of scenarios. In the context of 5G, SDS components can be adapted to provide a secure slice creation algorithm by leveraging the multi-objective optimization infrastructure. Defining a network slice implies translating an abstract specification [i.e., source, destination, Service Level Agreement (SLA), Quality of Service (QoS), etc.] into a slice template, i.e., a low-level structure that can be understood from a 5G orchestrator such as the OSM MANO. As previously discussed, this category of actions pertains to the orchestrator.

The definition of a slice can be understood in terms of a path between two nodes, with a specified SLA that depends on the type of traffic and a set of network functions that should be deployed strategically along the path. Additional requirements that must be considered include:

- **Energy efficiency:** Network nodes may need a considerable amount of electrical energy to function. Optimizing the overall number of switched-on nodes in a topology may have a considerable impact on power consumption and thus represents an important indicator to optimize.
- **Workload of link and nodes:** For each node and link, the maximum workload that can be handled is defined. However, equipment reliably operates only up to a percentage of the maximum workload capacity. From the reliability point of view, redundancy mechanisms (e.g., switching on a secondary node) should be used when the actual load reaches the limit. This represents a conflicting objective with the energy efficiency one.
- **Isolation constraint:** From a security point of view, critical services (e.g., mission critical communications during an emergency) should not share the physical infrastructure with untrusted services. Different services have specific security requirements that can require various degrees of isolation. This can impact energy efficiency as well as performances for involved, non-critical flows.

Clearly, these requirements cannot be optimized jointly. Specifically, the problem can be framed as a constrained multi-optimization problem and solved by any fitting algorithm. An example could be leveraging the Non-dominated Sorting Genetic Algorithm II (NGSA-II). It is necessary to specify what requirements should be coded into loss functions for the algorithm to optimize. Requirements can then be specified in terms of cost functions and constraints. The algorithms will in turn produce possible paths that can be ranked according to meta-heuristics or operator contribution.

Acknowledgments

This work is supported by the H2020 RESISTO project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 786409.

References

[1] Afif Osseiran *et al.*, Scenarios for 5G mobile and wireless communications: The vision of the METIS project, in Communications Magazine, vol. 52, no. 5, IEEE, 2014, DOI: 10.1109/MCOM.2014.6815890, ISSN 1790-0832 (WC · ACNP).

[2] 5G Italy consortium, 5G Italy white e-book: from research to market, https://www.5gitaly.eu/2018/white-ebook/, 2020

[4] Open networking foundation, SDN architecture Issue 1.1 2016 – https://www.opennetworking.org/wp-content/uploads/2014/10/TR-521_SDN_Architecture_issue_1.1.pdf

[8] B. Chatras, "On the Standardization of NFV Management and Orchestration APIs," in IEEE Communications Standards Magazine, vol. 2, no. 4, pp. 66–71, December 2018.

[9] https://www.ericsson.com/en/reports-and-papers/white-papers/5g-security-scenarios-and-solutions

[10] https://www.ericsson.com/en/security/a-guide-to-5g-network-security

[11] Norrman, K., Teppo, P., Mononen, K. and Nilsson, M. (2014). Setting the standard: methodology counters security threats. [online] Ericsson Review. Available at: https://www.ericsson.com/assets/local/publications/ericsson-technology-review/docs/2014/er-security-assurance-3gpp.pdf

[13] ETSI Open source management and orchestration – https://osm.etsi.org/

Part V

Sector Agnostic Issues in Critical Infrastructures Protection

DOI: 10.1561/9781680836875.ch19

Chapter 19

Detection of Innovative Low-rate Denial of Service Attacks Against Critical Infrastructures

By Enrico Cambiaso, Ivan Vaccari and Maurizio Aiello

In the cyber security field, Denial of Service (DoS) attacks are executed to exhaust the resources of the victim, by compromising availability of the targeted system, thus affecting reliability for its intended users. Last-generation threats make use of low bandwidth to target network services. In virtue of this, they are more difficult to counter, and efficient detection is still an open issue in the research field. This chapter focuses on the analysis of the functioning of low-rate DoS attacks targeting network services, with emphasis on web protocols.

19.1 Introduction

In computing, a denial of service (DoS) attack is an attempt to saturate the resources of a system, making it unavailable for legitimate users. There are many possible motivations which can lead the attacker to launch a DoS attack, hence the action of one or more people to interrupt or suspend the services offered by a generic system. The range of potential victims of a DoS attack includes instead each service connected to a computer network. In practice, the most attacked services are high-profile web services, such as banks, governments, critical infrastructures, or widely used services.

The origins of DoS attacks are strictly related to the Internet architecture. In particular, some design decisions made decades ago played a role in the game. Those decisions were driven by providing functionality rather than security, defining an architecture capable to offer a fast and cheap communication mechanism, moving packets from a source to a destination. Additional functionalities such as packet loss, reorder, or corruption management have been delegated to higher level transport protocols, deployed both by the sender and the receiver (*end-to-end principle*).

When one party of the presented model is malicious, it would act to create a damage on the other party. In this case, the adopted protocols should no longer be considered reliable. In October 1986, the Internet experienced a series of congestion collapses (Nagle, 1984), caused by the design and deployment of several TCP congestion control protocols (Floyd, 2000), unable to properly allocate resources in case of aggressive traffic flows.

In this chapter, we describe first generation of denial of service attacks in Section 19.2. Such threats act at the third or fourth layer of the ISO/OSI model, focusing in particular to flooding DoS threats. Hence, we introduce Slow DoS Attacks in Section 19.3, by describing its functioning and a categorization of known attacks. Section 19.4 focuses instead on the metrics used to characterize network traffic for slow DoS detection purposes, while Section 19.5 reports the intrusion detection framework able to detect malicious slow DoS activities. Finally, Section 21.6 concludes the chapter.

19.2 Old Style DoS Attacks

Concerning denial of service threats, a common pattern of attack consists in saturating the target host with malicious requests, blocking or slowing down the server's responses directed to legitimate users. This attack can lead to a server overload, causing a DoS.

Considering old-style DoS threats, as shown in Figure 19.1, there are essentially two types of DoS attacks: (i) *vulnerability-/exploit-based attacks* can exhaust the resources of a server by exploiting a vulnerability in the software; instead, (ii) *flooding-based attacks* simply send to the victim a high volume of data, which cannot be handled by the target.

Exploit-based attacks (also known as vulnerability, software, or semantic attacks) may involve a few well-crafted packets in the attack. As they exploit a specific

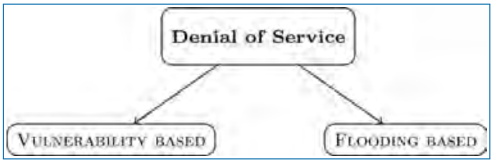


Figure 19.1. A first denial of service attacks classification.

vulnerability of the software, they are quite simple to mitigate by patching the vulnerability on the victim host or by identifying the packets related to the attack and handling them separately, or simply dropping them. In general, signature-based approaches are therefore adopted in this case to counter such threats.

Instead, flooding-based attacks overwhelm the victim's resources by sending a large amount of data. If DoS attacks are launched from a single machine, it is possible to protect from such attacks by equipping the victim with abundant resources. In this way, the attacker would need a better equipped machine in order to successfully perform the attack. In case of a distributed attack, instead, the attacker adopts a set of machines/bots whose resources overwhelm in total the resources of the server.

We will now describe in detail exploit-based and flooding-based attacks.

Exploit-based attacks

As mentioned above, exploit-based DoS attacks send specially crafted packets to the victim system, with the aim of targeting a specific software/daemon, exploiting its vulnerabilities (Tao *et al.*, 2009). These vulnerabilities/bugs may affect elements such as the operating system, the listening service, or a specific application. The success of the exploitation will typically lead to crash the targeted system.

Exploit-based attacks can potentially disable the victim machine with an extremely limited amount of network packets. Even a single packet may lead a DoS on the targeted victim (Hussain *et al.*, 2003). A well-known example is represented by *ping of death (PoD)*, an attack causing a crash on the operating system by sending a single large ICMP echo packet. Similarly, the *land attack* sends a single TCP SYN packet containing the victim's IP address as both source and destination address, hence leading to an endless loop in the protocol stack.

An efficient way to protect a system from such threats consists in frequent software patching and updating. These threats can indeed only be prevented by properly applying software updates. Although these menaces are important in the network security field, they are beyond the scope of our work. Indeed, exploit-based attacks require the attacker to address the malicious activity to the specific targeted machine, while our aim has been focused instead on analyzing "wider" attacks, able to indistinctly target different services. In particular, since vulnerability attacks have been widely considered in literature, it is stated that software updates (on the host, but also on the network, if needed) can mitigate them.

Flooding-based attacks

As mentioned above, flooding-based attacks overwhelm the resources of the victim by sending a large amount of data. Because of this requirement, an attack can be perpetrated only if the attacking host is equipped with high amount of resources (network bandwidth, especially). It is therefore clear that, being servers typically associated to high amount of resources, it is unlikely that a single node leads a DoS on a corporate server. In virtue of this, single executions of these attacks are nowadays rarely found, in favor of last-generation threats, characterized by limited resources requirements. Nevertheless, flooding-based threats are executed nowadays by organized entities, often controlling and coordinating many attacking nodes.

19.3 Slow DoS Attacks

While most level-4 denial of service attacks need a high quantity of data, this is not a characteristic of level-7 attacks, which require a low bandwidth rate. The term *Slow DoS Attack* (SDA in the following) indicates an attack whose ambition is denying a network service through a low quantity of data. We focused on such threats due to their novelty and their ability to behave similarly to a legitimate situation. Because of these reasons, such attacks are therefore often undetected. Indeed, although many research works focus on detecting and mitigating them, efficient protection systems in this context are nowadays still missing (Aiello *et al.*, 2014). In addition, the low-bandwidth requirements guided us through the exploitation of such attacks on devices equipped with limited resources.

In this section of the document, we describe in detail the SDA term meaning, hence describing available threats and their characteristics, providing an appropriate taxonomy of such kind of attacks.

19.3.1 The Slow DoS Attacks Term

The term *Slow DoS Attack* refers to a DoS attack which makes use of low bandwidth rate to accomplish its purpose. An SDA often acts at the application layer of the network protocol stack. The characteristics of this layer are easier to exploit in order to successfully attack a victim even by sending few bytes of malicious requests.

Note that the term *slow* does not necessarily imply that SDAs “send data slowly”: even if in some cases this can be true, the term *slow* historically comes from a famous attack in this context, known as *Slowloris*, which has been particularly important during the Iranian protests of 2009 (Giralte *et al.*, 2013).

In literature, similar definitions have been given to *low-rate DoS* (*LDoS* or *LRDoS*) (Macia-Fernandez *et al.*, 2008, Guirguis *et al.*, 2006, Liu *et al.*, 2012, Tang, 2012), *application layer DoS* (Yu *et al.*, 2007, Xie and Yu, 2009) or (in case that either HTTP or XML protocols are exploited) *HTTP/XML DoS* attacks (Chonka *et al.*, 2011).

Relatively to the purpose of an SDA, this is often to cause unavailability of a victim host by seizing all the connections available at the application level of the victim. Under the SDA condition, all service queues are busy and any new incoming request from legitimate users is discarded, therefore causing a DoS. The attacker attempts to force the server to process only his own requests by filling the service queues only with attack message requests.

Assume that somehow the attacker manages to make the server reach its saturation state; in this situation, whenever a position is freed in any of the service queues, the attacker attempts to take it again, before any other (legitimate) user can. Thus, the aim of the attacker is to achieve the maximum number of positions in the service queue, at a high-rate, thus maximizing the probability of seizing all the positions as quickly as possible. However, it is desirable to use low-rate traffic instead, for two main reasons: because this approach allows the attack to be carried out with much fewer resources, and also because it can more easily bypass protection mechanisms that rely on the statistical detection of high-rate traffic (Siris and Papagalou, 2006, Huang and Pullen, 2001, Gil and Poletto, 2001).

An SDA may also exhibit an ON-OFF nature, which comprises a succession of consecutive periods composed of an interval of inactivity (called off-time), followed by an interval of activity (called on-time). This characteristic may play an important role for the seriousness of the threat, since the attack traffic assumes a form which is statistically similar to the one of legitimate traffic; moreover, servers implicitly support users with slow or intermittent connection bandwidth. For instance, let's think to a user searching a specific term on Google: in this case, an active query would be accomplished (ON status), thus interrupting communication activities for some seconds/minutes for consulting obtained results (OFF state), hence making an additional query, if needed (ON state, again).

Once the SDA has seized all available positions in the service queues, the attacker slows down the connections from or to the victim by exploiting the characteristics of either a specific protocol (i.e., HTTP, FTP, DNS, etc.) or the application software (i.e., PHP, SOAP, etc.). The connections are thus kept active as long as possible, by sending minimum amounts of data per time unit.

19.3.2 Slow DoS Attacks Categorization

We will now report a categorization for Slow DoS Attacks (Cambiaso *et al.*, 2013). According to (Howard and Longstaff, 1998), "a good classification has to be considered as a common language for the research in a particular field," while formally a taxonomy can be defined as "the study of the general principles of scientific classification" (Igure and Williams, 2008). At the beginning of a scientific study relative to a new field, a good taxonomy is considered as an "important and necessary prerequisite for systematic study" (Howard and Longstaff, 1998; Lindqvist and Jonsson, 1997), Indeed, a mere collection of objects is not useful for an exhaustive study, unless the objects are properly classified (Landwehr *et al.*, 1994, Lindqvist and Jonsson, 1997).

As described above, an SDA uses low bandwidth to reach its objectives. To do that, making the server unavailable for legitimate requests, it should exhaust some kind of resources. In general, the access to a server can be denied in two different ways: either exhausting its internal main resources (CPU, RAM, kernel buffer space, network capacity of its interface card, available sockets, available process IDs, etc.) or exhausting resources that are functional to it (bandwidth of the network where the server is attached or external devices necessary to the server like switches, routers, gateways, power supply, etc.) In general, these are the targets of the old style attacks like high bandwidth floods (i.e. ping floods). Old style SYN flood threat (Safa *et al.*, 2008) affects the kernel buffer space: the attack aims to fill all the available connections (for each connection a proper structure is created in memory), whose maximum number is pre-configured. Looking more deeply on the slow DoS field, the object of the taxonomy, we have analyzed existing as well as potential attacks.

A first categorization has been focused on distinguish between *practical* and *meta* attacks (Cambiaso *et al.*, 2012): while practical attacks are concrete, implemented and often released as a software, meta attacks are defined at a higher level, providing the guidelines to perpetrate a generic threat. In this case, a concrete implementation is missing, since it depends on the targeted server. During our study, we have analyzed both the threats, providing and implementing innovative menaces belonging to both the categories.

Concerning a different categorization we have provided (Cambiaso *et al.*, 2013), an SDA which affects CPU or memory of the victim aims to force it to do expensive operations. We have called such attacks *Delayed Responses DoS* attacks. On the network side, high bandwidth attacks are not SDAs by definition, so they will not be considered; thus the only options are (i) to directly attack the network infrastructure (*Network Oriented DoS*) or (ii) to seize all the available connections of the victim, at the application layer if this is easier than at the transport layer. The only way to seize all the connections is to occupy them, keeping them busy (avoiding connection close that could free-up resources for legitimate client). Once a connection is open, it is kept busy until it is used; when no more traffic flows in the channel, it is automatically closed by timeout. The trick is to try and keep the connection open for a long period of time by sending the least amount of bytes per time unit. For this purpose, the attacker could exploit three different entities: (i) the malicious client, (ii) the victim's application timeout, or (iii) the server under attack itself.

(i) *The client*: even if it may seem extravagant, the attacker could deliberately close an established conversation in order to plan an immediately subsequent reoccupation of the same channel. We introduced the *Resource Occupation Planning DoS* category to include such attacks.

(ii) *The timeout*: the attacker could postpone indefinitely the expiration of the timeout used by the daemon application running on the victim host, using various techniques (analyzed hereinafter). Such threats are commonly categorized as *Timeout Exploiting DoS*.

(iii) *The server*: in some cases, the attacker could trick the server making it believe that the client resources are different from the reality, in order to accomplish a DoS. Such category is identified as the *Resource Distortion DoS* attacks.

Notice that the categories described above are *not* mutually exclusive. Hence the categories can also be mixed, and a single attack may belong to more than one category. Also, categories may be related among themselves: for instance, Delayed Responsive threats are also included in the Timeout Exploiting DoS category.

19.4 Detection of Low-rate DoS Attacks

In the last century, communication has evolved and Internet became the most relevant communication medium. As today Internet-connected computer systems play a vital role in modern society, they are often subject to intrusions and attacks. Therefore, Internet has to be kept a safe place, providing an appropriate security layer to its users. Intrusion detection techniques are executed to identify malicious activities targeting a specific network or host.

Intrusion Detection Systems (IDS) can be categorized into *anomaly detection* and *misuse detection* (Cambiaso *et al.*, 2016), while anomaly detection systems, such as IDES, flag as anomalous each activity that significantly deviates from normal usage profile, misuse detection systems, such as IDIOT (Kumar and Spafford, 1995) or STAT (Ilgun *et al.*, 1995), profile well-known menaces extrapolating attack signatures characterizing an intrusion.

Currently, building an effective IDS is no easy task. An anomaly-based approach may use intuition and experience to identify statistical measures (Lunt, 1993), while a misuse approach first analyzes and categorizes attacks and vulnerabilities, thus defining specific rules and patterns to identify a running threat. Once the signature of a particular menace is obtained, a potential execution of the same attack could successfully be detected. Nevertheless, since such signature-based approach cannot detect novel attacks, it should not be considered a complete solution.

Although there are many Slow DoS Attacks, the possibility to correlate some specific performance patterns to a given attack or category would be useful in detection systems (e.g. IDS) for recognizing such menaces. Since intrusive behavior often shows anomalies from legitimate behavior, we exploit anomalies to detect possible intrusions in an information system. Indeed, in this paper we adopt a Statistical Based Intrusion Detection approach (SBID in the following), which belongs to the anomaly detection category (Ye *et al.*, 2001). This approach works by characterizing "legitimate activities" based on the analysis of some chosen parameters and then by investigating unknown traffic. In particular, if the observed traffic falls outside the scope of the legitimate one, it would be flagged as anomalous (Farshchi, 2003). A SBID approach provides the ability to detect both known and novel malicious activities. Through a SBID approach, a statistical characterization of some legitimate activities product (i.e., network traffic, system calls, email sent, socket connections, etc.) is created. An unknown activity product is then compared with the legitimate one, in order to ascertain its eventual abnormality.

These approaches adopt "parametric" or "non-parametric" techniques: in the first case, the distribution is assumed as known (for instance, we could assume a Gaussian distribution of the traffic), while in the latter case there is no knowledge assumption of the underlying distribution (Chandola *et al.*, 2007). Since SBID systems report all anomalies as malicious activities, such approaches may signal false alarms in case the anomaly is caused by a legitimate behavioral irregularity. Because of this, SBID systems are often used in conjunction with pattern recognition techniques, in order to serve a better working Intrusion Detection System (Ye *et al.*, 2001).

Analyzing available detection and mitigation methodologies for network attacks, a protection system may involve hardware or software components. While applying an hardware protection [such as a load balancer, a network proxy, or an hardware accelerator, able for instance to protect from specific Slow DoS Attacks (Aiello *et al.*, 2014)] could successfully mitigate some categories of attacks, it can be considered a workaround rather than a good solution. Indeed, since such hardware appliances have not been designed for this purpose, their resources would be allocated for this additional activity. Moreover, such approaches usually do not provide a built-in functionality aimed at detecting a working attack on the network.

For instance, if we consider the HTTP protocol, which is particularly exploited by network attacks such as SDAs, current software protections systems are organized as software modules. Although there are several different modules available, their functioning is based on two basic principles: (i) limit the maximum number of simultaneous connections coming from a particular client and (ii) apply specific server-side timeouts (Aiello *et al.*, 2014). Even if such solutions represent some sort of mitigation techniques, in (Aiello *et al.*, 2014) we demonstrate the inefficacy of the available modules in protecting from particular attacks such as, e.g., distributed attacks.

If we analyze the design process for building an IDS, following core activities are usually involved (Cambiaso *et al.*, 2016):

1. First of all, a behavior is observed, usually on the server host or on the server's network path. Based on this observation, a network traffic *representation* phase is accomplished to select specific parameters to extrapolate, in order to identify potential network anomalies.
2. Subsequently, an *analysis* algorithm is applied to the extrapolated data, by using approaches and metrics deriving from different research fields such as machine learning, neural network, statistics, game theory, etc…
3. Finally, a *characterization* phase is accomplished, by defining a proper threshold distinguishing a legitimate situation from an anomalous one. Also in this case, different research areas may be involved.

After characterization is properly defined, the intrusion detection system is able to detect anomalies on the network. In this chapter, we focus on the definition of a representation scheme able to identify Slow DoS Attacks, based on a set of characteristic attack metrics derived by the functioning of the attacks described above. Hence, we detail the intrusion detection framework by reporting information on how to retrieve such metrics, also considering a reference protocol like HTTP.

It is worthy mention that our subdivision is driven by a "component-based" approach (Jifeng, 2005), in which (software) components are reused to build and to maintain a system. In our context, referring to the activities executed by an IDS mentioned above, we believe that an intrusion detection system may be built as a set of components communicating and integrating among themselves. To this end, we describe an attack representation model/framework for researchers and designers of intrusion detection systems. Although we focus on Slow DoS Attacks, the framework and detection algorithms can be applied to efficiently protect from other categories of threats.

19.4.1 Characteristic Attack Metrics

As mentioned above, when building an intrusion detection system, the first activity which is accomplished is relative to the observation of the phenomenon, with the aim of *represent* a possible anomaly. This representation is accomplished by accurately selecting parameters which have to be extracted from captured data (i.e., a PCAP packet capture file). These parameters have to be potentially able to represent an anomaly, distinguishing it from a legitimate/accepted behavior.

A first approach is based on selecting parameters characterizing Slow DoS Attacks exploiting application timeouts with extreme efficiency. In virtue of this, we describe an intrusion detection framework for application layer threats identification. Although particularly suitable to the slow DoS field, we have analyzed that retrieving these parameters is an expensive task (since they work at the application layer, thus requiring connections reconstruction), hence making real-time detection difficult in practice. Therefore, we have focused on reducing retrieval costs, by working at lower layers of the ISO/OSI model.

Δ-parameters

After the connection between client and server has been established, the client sends a request to the server. The request is interpreted by the server in order to generate a response to send back to the client. After the first request-response exchange, two possible events could characterize the connection: (i) the connection is closed or (ii) the connection is kept alive (*persistent connection* Fielding *et al.*), in order to reduce the connection overhead for any additional request-response between the same client/server pair.

Since each connection potentially is persistent, a first extrapolated parameter is the N_{req} parameter, reporting for each connection the number of requests included in each connection stream.

We define as *connection slot* the portion of a connection which refers to the time passing between the start of a request and the end of the relative response on the same stream. According to Figure 19.2, let us define $t_{start\_connection}$ the connection start identified by the 3-way-handshake completion, t_{start_req} the starting time of a request, t_{end_req} the ending time of a request, t_{start_resp} the starting time of a response, and t_{end_resp} the ending time of a response. From these values we can extrapolate data relatively to the time passed before sending the first request (Δ_{start}), the duration of a request (Δ_{req}), the duration of a response (Δ_{resp}), and the time passed between the end of a request and the start of the relative response (Δ_{delay}). While the Δ_{start} parameter is associated with a single connection, other parameters are related to each connection slot, which also includes the time passed between the end of the response and the start of the next request of the same stream (Δ_{next}). In addition, since connections slots are part of a connection, each connection is uniquely identified by the C_{id} value (in our case, a sequential integer value), and each connection slot is associated with the C_{id} connection it belongs, plus the slot index S_i on the connection.

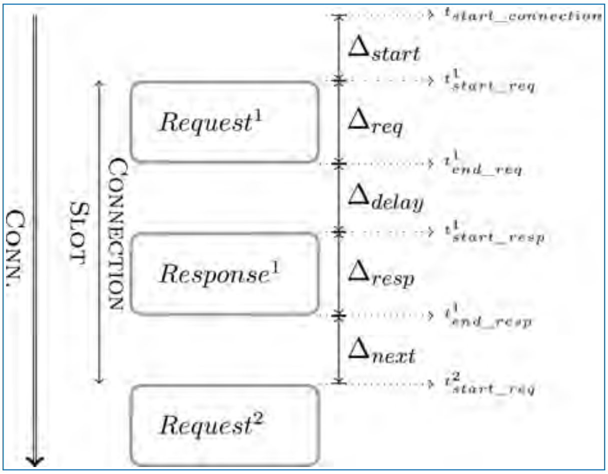


Figure 19.2. TCP connection stream for a request-response protocol-based connection.

Fixing a connection stream, let us define $i = S_i$ the connection slot index on the stream. According to Figure 19.2 (where $i = 1$), which depicts a scheme of the parameters, we define:

$$\Delta_{start} = t^1_{start\_req} - t_{start\_connection} \tag{19.1}$$

$$\Delta_{req} = t^i_{end\_req} - t^i_{start\_req} \tag{19.2}$$

$$\Delta_{delay} = t^i_{start\_resp} - t^i_{end\_req} \tag{19.3}$$

$$\Delta_{resp} = t^i_{end\_resp} - t^i_{start\_resp} \tag{19.4}$$

$$\Delta_{next} = t^{i+1}_{start\_req} - t^i_{end\_resp} \tag{19.5}$$

By choosing these parameters, we are able to extrapolate behavioral features for application layer attacks. For instance, it is known that Slow DoS Attacks like Slowloris (Giralte *et al.*, 2013) split HTTP requests by sending request packets delayed during the time, thus being typically characterized by high Δ_{req} values. Instead, the Apache Range Headers (Cambiaso *et al.*, 2013) attack makes the Δ_{delay} parameter assume high values, since requests sent to the server need particularly intensive calculations to produce an appropriate response. Similarly, Slow Read attack (Shekyan, 2012) simulates a tiny reception buffer to slow down the responses of the server, thus being characterized by high Δ_{resp} and p_{resp} values. Relatively to the Δ_{next} parameter, it is instead exploited by the Slow Next attack (Cambiaso *et al.*, 2015).

19.4.2 Assumptions

In order to properly define a detection model, we have to define the model behavior at limit cases. In particular, we make the following assumptions:

- A connection which does not start with a request is ignored until a request is found on the same connection; this may happen when traffic capture operation begins after a full request has been sent and the relative response is captured;
- Due to the nature of elements like network, communication medium, or response production times, traffic measurements always provide:

$$\Delta_{delay} > 0 \qquad \Delta_{next} \neq 0 \tag{19.6}$$

19.4.3 Messages Overlapping on the Same Connection Stream

If we focus on the Δ_{next} parameter for a single TCP connection stream, in case $\Delta_{next} < 0$ a connections slots overlapping occurs. Moreover, in this case connection persistence is adopted and connections may include more than a single request. For example, in Figure 19.3, the request next to the current one on the same connection stream is (even partially) received before the full response to the current request is sent.

Although in Figure 19.3 no overlappings are shown between $Request^{i+1}$ and $Response^{i}$, since $t^{i+1}_{end_{req}} < t^{i}_{start_{resp}}$, overlappings may occur. In particular, relatively to a single TCP connection, an overlapping occurs when Equation 19.7 is satisfied.

$$\begin{aligned}&\exists j \in \mathbb{N}^{+} \exists k \in \{j-1, j+1\} | (t^{k}_{start_{req}} < t^{j}_{start_{resp}} < t^{k}_{endreq})\\ &\vee (t^{k}_{start_{req}} < t^{j}_{end_{resp}} < t^{k}_{endreq})\end{aligned} \tag{19.7}$$

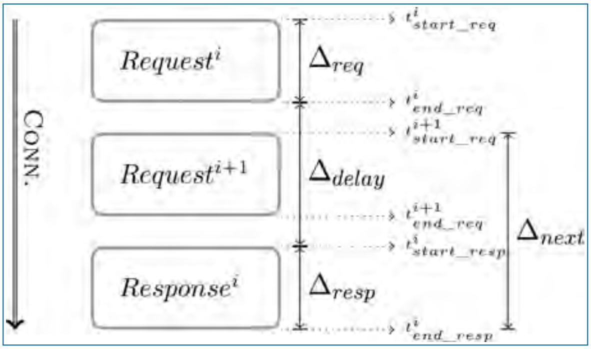


Figure 19.3. TCP connection stream in case of $\Delta_{next} < 0$.

For instance, this kind of overlapping may occur in the HTTP protocol: considering a single TCP connection stream, a client requests to the server a particular resource such as a web page. After the page content is received and parsed by the client, a set of resources bounded to the same hosting server (i.e. pictures, scripts, or stylesheets) may be found in the page. In order to correctly show the web page such additional resources have to be obtained, hence additional requests have to be sent to the server. In case a persistent connection (HTTP 1.1) is used, a series of subsequent requests is usually sent to the server through the already established TCP connection.[1] In this case, the requests sent from the client and next to the first one may be overlapped to the receiving of the responses to the previous requests.

Our feature selection is based on the fact that each connection is composed by a sequence of requests and responses. This fact is not always true; as described above, in case of messages overlapping at the application layer some connections may adopt a full-duplex communication thus resulting in a simultaneous communication between client and server. The model has therefore to adapt itself to correctly identify the start/end of a request/response. For instance, in case of an HTTP 1.1 connection, the client may sequentially send two requests/questions to the server. In this case, the response/answer to the first request would be overlapped with the second request sending, on the same TCP connection stream.

In this context, although a first implementation may identify changes in the packets direction for a common connection stream, such solution may generate inaccurate data. Indeed, although in this case packet inspection at the application level is not needed, if message overlapping occurs, improper data would be generated, due to the (possible) frequent change of direction relative to two different and overlapping messages on the same channel. Therefore, in order to carefully extrapolate connection slots data, an external protocol-dependent module may be needed.

19.4.4 Additional Parameters

Relatively to a single connection slot, described in detail in the previous section, following parameters could also be extrapolated for detection purposes:

- s_{req} to identify the request size, in bytes
- p_{req} to identify the amount of TCP packets that compose a request
- s_{resp} to identify the response size, in bytes
- p_{resp} to identify the amount of TCP packets that compose a response.

[1] Actually, some browsers use multiple connections in order to speed up displaying.

In particular, let us note that p_{req} and p_{resp} values depend on the data-link layer protocol adopted.

From a combination between these parameters and the Δ-parameters, we are able to retrieve composed parameters. For instance, the amount of bytes per second sent during a request/response may be defined as reported in Equation 19.8:

$$r_{s_{req}} = \frac{s_{req}}{\Delta_{req}} \qquad r_{s_{resp}} = \frac{s_{resp}}{\Delta_{resp}} \tag{19.8}$$

Similarly, the ratios representing the amount of packets per second sent during a request/response are reported in Equation 19.9:

$$r_{p_{req}} = \frac{p_{req}}{\Delta_{req}} \qquad r_{p_{resp}} = \frac{p_{resp}}{\Delta_{resp}} \tag{19.9}$$

It is also possible to obtain the average size per packet, as reported in Equation 19.10:

$$\mu_{s_{req}} = \frac{s_{req}}{p_{req}} \qquad \mu_{s_{resp}} = \frac{s_{resp}}{p_{resp}} \tag{19.10}$$

19.5 Intrusion Detection Framework

The proposed innovative intrusion detection framework (Cambiaso *et al.*, 2016) is based on the parameters introduced in Sections 19.4.1 and 19.4.4, and it provides a representation framework specific for DoS attacks working at the application layer. Since we believe that a component-based implementation of an IDS is fundamental, the proposed model not only provides important parameters able to identify application DoS attacks, but it also simplifies researchers' work, allowing them to use an already defined representation system.

The model is based on the concept that each connection is composed by a sequence of requests and responses. This fact is not always true; as described above, in case of messages overlapping at the application layer some connections may adopt a full-duplex communication thus resulting in a simultaneous communication between client and server. The model has therefore to adapt itself to correctly identify the start/end of a request/response. For instance, in case of an HTTP 1.1 connection, the client may sequentially send two requests/questions to the server. In this case, the response/answer to the first request would be overlapped with the second request sending, on the same TCP connection stream.

In this context, although a first implementation may identify changes in the packets direction for a common connection stream, such solution may generate inaccurate data. Indeed, although in this case packet inspection at the application level is not needed, if message overlapping occurs improper data would be generated, due to the (possible) frequent change of direction relative to two different and overlapping messages on the same channel. Therefore, in order to carefully extrapolate connection slots data, an external protocol-dependent module may be needed. We now describe in detail a module implementation, by using as a reference protocol the HTTP protocol.

19.5.1 HTTP Model Implementation

The aim is to analyze traffic data, such as a PCAP packet capture file, representing the rough network packets on a network/host relatively to a time interval. The analyzed rough data are only relative to the packets directed to the application layer. From such packets, it is possible to rebuild and extrapolate a list of (captured) connection streams. This choice allows us to easily retrieve needed capture files by sniffing the network on the server needing protection. Nevertheless, some issues are related to such approach: due to network traffic dump limits, a request (or similarly a response) composed by a single packet leads to:

$$t_{end\_req} = t_{start\_req} \implies \Delta_{req} = 0 \tag{19.11}$$

since capture files associate a packet to a specific reception time. For instance, this fact may occur in case of a single packet including the entire request payload. Although this issue may be considered an important limitation, a more accurate retrieval is not relevant in this context, since our model focuses on attacks targeting the application layer, while in case of a single packet composing the request (or similarly for a response), we expect potentially long Δ_{req} values for attacks targeting lower layers. In addition, a more accurate data retrieval would operate on the protected server (for instance, by operating at the kernel level, by intercepting sent and received messages), thus excluding a central node analyzing an entire subnetwork.

Another important consideration is relative to packets payload. In particular, the inspection of the messages directed to the application layer is needed for protocols allowing message overlapping (see Section 19.4.3). Indeed, in this case an analysis of the payload is needed in order to identify the starting/ending times of a request/response on a mixed stream. For instance, in case of the HTTP protocol the end of requests is identified by analyzing packets payloads and looking for the `\r\n\r\n` string. Instead, the `Content-Length` value sent by the server in the response header is needed to identify the response end. Conversely, in case messages overlapping is not supported by the protocol, inspection is not needed, and it is possible to identify requests/responses times by analyzing the direction flow of the packets. Although the packet inspection requirement is a limitation of the proposed approach, gaining access to the server needing protection should not be a problem, hence messages decryption should be possible (i.e., making use of private encryption keys).

19.6 Conclusions

In this chapter, we have analyzed the cyber security topic related to last-generation threats. We focused on Slow DoS Attacks, emerging denial of service threats making use of minimum attack resources to make a network services unavailable. Such threats are considered particularly dangerous for different domains including critical infrastructures. We have described in detail how such attacks work, and we have reported a categorization of such threats, in function of the approach adopted by the malicious entity. By analyzing Slow DoS Attacks functioning, we have also identified their weaknesses: in particular, the exploitation of specific server-side timeouts may expose the attacker for detection purposes. In this context, we have defined a set of metrics able to characterize legitimate network traffic and providing the ability to detect slow DoS threats. For metrics retrieval purposes, we have also considered as a reference protocol the HTTP protocol, widely adopted in different contexts. The metrics adopted can be adopted on other protocols as well. Similarly, the application of a wide variety of intrusion detection algorithms can be adopted to identify running Slow DoS Attacks on the network.

Acknowledgments

Part of this work has been carried out in the scope of the FINSEC project (contract number 786727), which is co-funded by the European Commission in the scope of its H2020 program.

References

A. Chonka, Y. Xiang, W. L. Zhou, and A. Bonti (2011). “Cloud security defence to protect cloud computing against HTTP-DoS and XML-DoS attacks.” *Journal of Network and Computer Applications*. 34(4): 1097–1107. URL: %3CGo%20to%20ISI%3E://000291846700008.

A. Hussain, J. Heidemann, C. Papadopoulos (2003). “A framework for classifying denial of service attacks.” *Computer Communication Review*. 33(4): 99–110. URL: %3CGo%20to%20ISI%3E://000188215800010

C. E. Landwehr, A. R. Bull, J. P. McDermott, W. S. Choi (1994). "A taxonomy of computer program security flaws." *ACM Computing Surveys (CSUR)*. 26(3): 211–254.

E. Cambiaso, G. Papaleo, G. Chiola, M. Aiello (2013). "Slow DoS attacks: definition and categorisation." *International Journal of Trust Management in Computing and Communications*. 1(3): 300–319.

E. Cambiaso, G. Papaleo, G. Chiola, M. Aiello (2015). "Designing and Modeling the Slow Next DoS Attack." *International Conference on Computational Intelligence in Security for Information Systems (CISIS)*.

E. Cambiaso, G. Papaleo, G. Chiola, M. Aiello (2016). "A Network Traffic Representation Model for Detecting Application Layer Attacks." *International Journal of Computing and Digital Systems*. 5(1): 31–42. URL: http://www.uob.edu.bh/uob__files/651/v%5C%205-1/paper%5C%204.pdf.

E. Cambiaso, G. Papaleo, M. Aiello (2012). "Taxonomy of Slow DoS Attacks to Web Applications." *Recent Trends in Computer Networks and Distributed Systems Security*. 195–204.

G. Macia-Fernandez, J. E. Diaz-Verdejo, P. Garcia-Teodoro (2008). "Evaluation of a low-rate DoS attack against application servers." *Computers & Security*. 27(7–8): 335–354. URL: %3CGo%20to%20ISI%3E://000261635200010.

H. Jifeng, X. Li, Z. L. (2005). "Component-based software engineering." In: *Theoretical Aspects of Computing-ICTAC 2005*. Springer. 70–95.

H. Safa, M. Chouman, H. Artail, M. Karam (2008). "A collaborative defense mechanism against SYN flooding attacks in IP networks." *Journal of Network and Computer Applications*. 31(4): 509–534. URL: %3CGo%20to%20ISI%3E://000262946600009.

J. D. Howard, T. A. Longstaff (1998). "A common language for computer security incidents." *Sandia Report: SAND98-8667, Sandia National Laboratories*. URL: http://www.cert.org/research/taxonomy_988667.pdf.

J. Farshchi (2003). "Statistical-based intrusion detection." *Retrieved October*. 10: 2009.

J. Nagle (1984). "Congestion control in IP/TCP internetworks."

J. Yu, Z. Li, H. Chen, X. Chen (2007). "A detection and offense mechanism to defend against application layer DDoS attacks." *Networking and Services, 2007. ICNS. Third International Conference On*. 54–54.

K. Ilgun, R. A. Kemmerer, P. A. Porras (1995). "State transition analysis: A rule-based intrusion detection approach." *Software Engineering, IEEE Transactions On*. 21(3): 181–199.

L. C. Giralte, C. Conde, I. M. De Diego, E. Cabello (2013). "Detecting denial of service by modelling web-server behaviour." *Computers & Electrical Engineering*. 39(7): 2252–2262.

M. Aiello, E. Cambiaso, M. Mongelli, G. Papaleo (2014). "An on-line intrusion detection approach to identify low-rate DoS attacks." *Security Technology (ICCST), 2014 International Carnahan Conference On*: 1–6.

M. Aiello, G. Papaleo, E. Cambiaso (2014). "SlowReq: A Weapon for Cyberwarfare Operations. Characteristics, Limits, Performance, Remediations." *International Joint Conference SOCO'13-CISIS'13-ICEUTE'13* pp. 537–546.

M. Guirguis, A. Bestavros, I. Matta (2006). "On the impact of low-rate attacks." 5: 2316–2321.

N. Ye, X. Li, Q. Chen, S. M. Emran, M. Xu (2001). "Probabilistic techniques for intrusion detection based on computer audit data." *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions On*. 31(4): 266–274.

R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, T. Berners-Lee, 'RFC 2616 – Hypertext Transfer Protocol – HTTP/1.1 – Available at https://tools.ietf.org/html/rfc2616." URL: https://tools.ietf.org/html/rfc2616.

R. Tao, L. Yang, L. Peng, B. Li, A. Cemerlic (2009). "A case study: Using architectural features to improve sophisticated denial-of-service attack detections." *Computational Intelligence in Cyber Security, 2009. CICS'09. IEEE Symposium On*: 13–18.

S. Floyd (2000). "Congestion control principles."

S. Kumar, E. H. Spafford (1995). "A software architecture to support misuse intrusion detection."

S. Shekyan. 2012. "Are you ready for slow reading? – Available at https://community.qualys.com/blogs/securitylabs/2012/01/05/slow-read." URL: https://community.qualys.com/blogs/securitylabs/ 2012/01/05/slow-read.

T. Lunt (1993). "Detecting intruders in computer systems." *In Proceedings of the 1993 Conference on Auditing and Computer Technology*.

T. M. Gil, M. Poletto (2001). "MULTOPS: A data-structure for bandwidth attack detection." *Usenix Association Proceedings of the 10th Usenix Security Symposium*. 23–34. URL: %3CGo%20to%20ISI%3E://000174316300003.

U. Lindqvist, E. Jonsson (1997). "How to systematically classify computer security intrusions." 154–163.

V. A. Siris, F. Papagalou (2006). "Application of anomaly detection algorithms for detecting SYN flooding attacks." *Computer Communications*. 29(9): 1433–1442. URL: %3CGo%20to%20ISI%3E://000237994600019.

V. Chandola, A. Banerjee, V. Kumar (2007). "Anomaly detection: A survey." *ACM Computing Surveys (CSUR)*. 41(3): 15.

V. Igure, R. Williams (2008). "Taxonomies of attacks and vulnerabilities in computer systems." *Communications Surveys & Tutorials, IEEE*. 10(1): 6–19.

X. Liu, G. Cheng, Q. Li, M. Zhang (2012). "A comparative study on flood DoS and low-rate DoS attacks." *The Journal of China Universities of Posts and Telecommunications*. 19: 116–121.

Y. Huang, J. M. Pullen (2001). "Countering denial-of-service attacks using congestion triggered packet sampling and filtering." *Tenth International Conference on Computer Communications and Networks, Proceedings*: 490–494. URL: %3CGo%20to%20ISI%3E://000172717000070.

Y. Tang (2012). "Countermeasures on Application Level Low-Rate Denial-of-Service Attack." *Information and Communications Security*: 70–80.

Y. Xie, S. Z. Yu (2009). "Monitoring the application-layer DDoS attacks for popular websites." *Networking, IEEE/ACM Transactions on*. 17(1): 15–25.

DOI: 10.1561/9781680836875.ch20

Chapter 20

Resilience Analysis and Quantification for Critical Infrastructures

By Natalie Miller, Mirjam Fehling-Kaschek, Gael Haab, Andrea Roland, Katja Faist, Alexander Stolz and Ivo Häring

The resilience analysis performed in RESISTO follows an enhanced risk and resilience management process based on the ISO-31000 standard [1]. The main inputs needed for the resilience quantification are gathered at separate steps of the management process: a precise understanding of the system context and the system itself including all subsystems and components and their interconnections; a collection of all relevant system functions to quantify the loss of performance due to a disruptive event; a comprehensive list of potential threats and hazards including information about their effect on the system; a list of critical combinations of system performance functions and threats taking account of relevant resilience dimensions [2] not explicitly covered by the risk and resilience management process; a risk and resilience quantification of combinations found to be critical taking into account risk and resilience evaluation criteria; and finally a list of potential counter actions and mitigation strategies for the listed threats.

The resilience quantification is performed via a network simulation based on all the collected information with the network simulation tool CaESAR [3], which was developed to simulate cascading effects in interconnected critical infrastructures. This allows to identify weak points of the system in terms of a critical resilience behavior for all combinations of performance functions and threats. In addition, the resilience improvement by different mitigation options can be tested via the network simulation.

20.1 Introduction

Telecommunication networks are becoming increasingly complex as connections with other critical infrastructures increase and the technology becomes more advanced. Additionally, these networks need to be able to handle certain disruptive events like terrorist attacks or natural disasters such as hurricanes or flooding. It is important that even as these networks grow and change, often with large legacy subsystems remaining operational, any risks are mitigated and the networks have a large resilience against the disruptive events.

Resilience does not yet have a universal definition; however, many of the proposed definitions are similar. For this chapter, resilience is defined as a cyclical process (resilience cycle or timeline, similar to the catastrophe management cycle) with five main phases: prepare, prevent, protect, respond, and recover [4]. Taking appropriate actions in each of these phases for overall risk control creates a more resilient system or network. Within this context, classical risk management focuses on the phases before and including the event, whereas resilience defined in a narrow sense cares about actions during and post event. Resilience as understood in the present chapter covers all phases before, during, and post disruptive events.

The aim of this chapter is to introduce a risk and resilience management process and discuss how it is currently being implemented in the RESISTO project, a project with the aim of improving the resilience of telecommunication networks against a variety of threats. This is done by using a resilience management process and a network simulator software, namely CaESAR.

This chapter is structured as follows: Section 20.2 introduces the risk and resilience management process used in RESISTO, including a general overview of the process as well as a more detailed look at how the process is being applied in RESISTO. Section 20.3 goes into more detail about how certain steps are completed in RESISTO, focusing on the tabular method used to obtain information on system components, system performance functions, threats and mitigation measures. Section 20.4 discusses how resilience is quantified in the project, by introducing the CaESAR software and the resilience indicators. Section 20.5 concludes the chapter with a summary of RESISTO's risk and resilience management process, an outlook of the next steps for the project and the reach this project and the methods discussed have on risk and resilience management and quantification.

20.2 Risk and Resilience Management

20.2.1 Methods

Risk management is the process of handling different risks that may be experienced by different organizations. There exist various types of risk management standards and frameworks, which have been developed and applied by organizations throughout the years. The standards propose how to manage risk and explain the managing processes regarding the risk. Similarities can be drawn between the risk management processes in the most used standards. Some examples of standards are ISO 31000 [5], ISO 31010 [6], COSO ERM, and CoCo [7].

In this chapter, the risk management process that will be discussed is the process defined by ISO-31000. ISO-31000 is a general risk management approach that is applicable to any sectors or industries [5]. The standard discusses risk management principles, a framework as well as the process of risk management.

Resilience management differs slightly from risk management; however, many of the steps remain similar. Resilience management could be said to be an extension of risk management, taking the analysis further to obtain a better understanding of the system, the threats or disruptions that may occur, and modification measures.

An increasing awareness toward resilience management has occurred in the last years, also in the context of ISO 31000 [8]. Clear guidelines on the resilience management processes are needed in order to implement resilience in a correct way. Efforts have been made to construct such guidelines that are applicable to different types of systems in various fields. This has resulted in a wide range of concepts, principles, and approaches regarding resilience management [9]. However, a generic approach is still lacking [1], even if many conceptual approaches have been formulated, see, e.g., [10] and [11] for the context of the present chapter.

20.2.2 Risk Management

The ISO-31000 standard creates an iterative process for managing risks (see Figure 20.1) that includes determining the context and scope, completing a risk assessment, and then treating the risk [12]. It has been used in many civil security application domains including urban security [13] and major public events such as football games [14].

Within the risk assessment, three steps are defined: risk identification, risk analysis, and risk evaluation. These in total five steps are complimented by other activities such as communication and consultation, monitoring and reviewing, and recording and reporting. It has been shown that the overall process can be modeled using the semiformal Systems Modeling Language (SysML) [15].

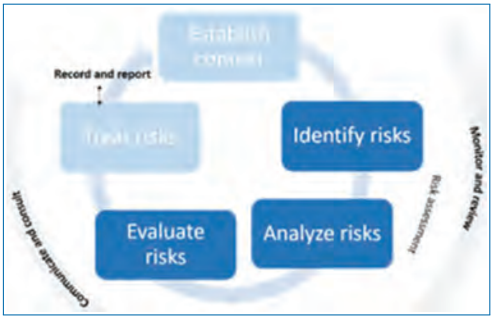


Figure 20.1. The risk management process as defined by ISO31000 (adapted from [5]).

The first step of risk management is determining the context, scope, objectives, and criteria. This step creates the environment the rest of the process is completed in. The scope and criteria are defined by the organization and are required for the management process as they define which and how many risks will be considered in the process in a given context and how they are evaluated in terms of overall risk acceptance, respectively [5].

Once the first step is completed, the risk assessment begins. This assessment consists of three parts: risk identification, risk analysis, and risk evaluation. In the risk identification step, risks are recorded. Risks that are relevant for the process are the ones that will hinder the objective defined, i.e., they are risks on objectives. The last step of the risk assessment is the evaluation step. In this step, it is determined if the risks need any action to be taken to reduce them. Decisions that can be made in this step include taking no action, determining treatment options, analyzing the risk further, using the present controls, or changing the objectives of the process [5].

The final step is the risk treatment. If risks are deemed too high to be acceptable, mitigation measures are determined. This step includes determining the mitigation measures but also implementing and testing to see how well they perform. Some examples of treatment options defined by the standard include reducing the likelihood or consequences, removing the source of the risk, or dividing the risk by utilizing contracts or insurance [5].

Tools and activities that aid the five-step process include communication and consultation. These tools are important as they can help stakeholders understand the process and results [5]. They are completed throughout the entire process. Communication focuses on understanding risk, while consultation is focused on decision-making and the information that decision-making requires. Other tools include monitoring and reviewing, and recording and reporting. Monitoring and reviewing ensures the assessment is completed to the highest quality. Recording and reporting helps to communicate outcomes that were determined or found through

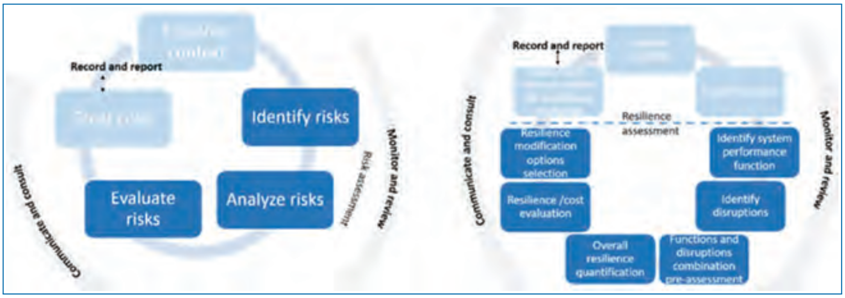


Figure 20.2. The two management processes differ in number of steps but the general idea remains the same. Resilience as used here covers classical risk control to better prepare for, avoid and protect from damage events as well as resilience to better respond, recover, learn, and adapt, since it considers all resilience cycle phases as well as (technical) resilience capabilities. Adapted from [16].

this process. Actions for risk management and decision-making information can also be reported and recorded [5].

20.2.3 Resilience Management

The resilience management process, initially defined by [1], extends the ISO 31000 standard to be applicable to resilience. Originally created from the ISO 31000 version of 2009, it still holds valid for the updated 2018 ISO 31000 version. The resilience management process is a nine-step process, almost double the number of steps when compared to the risk management process. A side-by-side comparison of the two processes can be seen in Figure 20.2 (adapted from [16]).

In the resilience management process, the context and system analysis are separate, unlike in the risk management. This is done to make sure a full understanding of the system, and its subsystems or connections outside the system are identified [1]. Once the first two steps are completed, the resilience assessment begins. Like the risk assessment, it is split into three parts: identification, analysis, and evaluation. However, within the three parts, there are further steps.

In the identification step, the system performance functions and the disruptions are determined. Mentioned in [1], when the performance functions are determined, qualitative and quantitative descriptions should be included. An inventory should be made to include all the functions found. Performance functions that are found can be either dynamic or static, allowing for a deeper understanding of the system. An inventory of disruptions should also be created.

For each disruption or threat that is determined, for each system performance function, the effects should be determined for all resilience cycle phases, i.e., before, during, and after events. This means that in particular the losses should be known, which layers will be affected, and what uncertainty they have.

To this end, within the resilience analysis steps, the combinations of functions and disruptions are pre-assessed and the overall resilience is quantified. Within this pre-assessment, a matrix form may be utilized with the system functions as one dimension and the disruptions, or threats, as the other. This matrix is explained in more detail in Section 20.3.5. This can then be utilized to determine the critical combinations. The overall resilience quantification of these critical combinations is completed in the next step. The resilience quantification takes things deeper, investigating disruption combinations, cascading effects and the overall system [1].

After the resilience is quantified, the resilience evaluation can be completed. This consists of a resilience/cost evaluation and selection of modification options for improving the resilience. The resilience evaluation has specific steps to follow including a comparison of resilience performance values to historic values, investigating the performance loss, and determining if the new resilience level is acceptable or not [1]. The procedure of selecting improvement options is also defined in [1] and includes creating an inventory of options, and completing the resilience management process again to determine the benefits. This will lead to the selection of final measures that can then be implemented while considering potential secondary (unintended) effects of selected risk control and resilience measures.

Once the resilience assessment is completed, the final step of implementing the modification options is completed. This step would focus on the development of the improvement methods and how to implement, operate, and maintain them with the system.

The resilience management process has the same tools and activities that the risk management process does: communication and consultation, monitoring and reviewing, and recording and reporting, within an overall iterative as well as incremental process.

20.2.4 Risk and Resilience Management in RESISTO

RESISTO's [17] long-term control loop completes this risk and resilience management process. Figure 20.3 (adapted from [18]) below shows the resilience management process and how RESISTO completes each step. Further, much less detailed application cases of the resilience analysis and management process for urban transport, coupled critical infrastructure of a region and a national electricity grid can be found in [1], as well as for an indoor localization system in [19]. The process has also been proposed to be applied to local electricity distribution grids after the last transformer [20].

The selection of candidate tabular methods for improving resilience has been described in [21] by considering several resilience dimensions, including the five

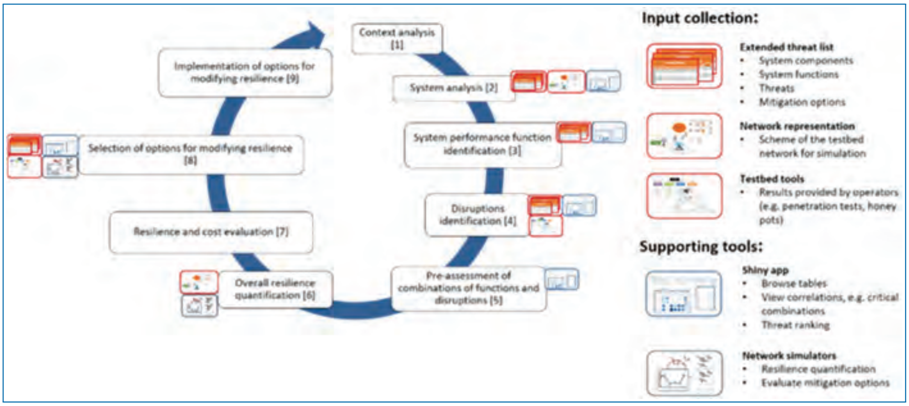

Context analysis [1]
System analysis [2]
System performance function identification [3]
Disruptions identification [4]
Pre-assessment of combinations of functions and disruptions [5]
Overall resilience quantification [6]
Resilience and cost evaluation [7]
Selection of options for modifying resilience [8]
Implementation of options for modifying resilience [9]
Input collection:
Extended threat list
• System components
• System functions
• Threats
• Mitigation options
Network representation
• Scheme of the testbed network for simulation
Testbed tools
• Results provided by operators (e.g. penetration tests, honey pots)
Supporting tools:
Shiny app
• Browse tables
• View correlations, e.g. critical combinations
• Threat ranking
Network simulators
• Resilience quantification
• Evaluate mitigation options

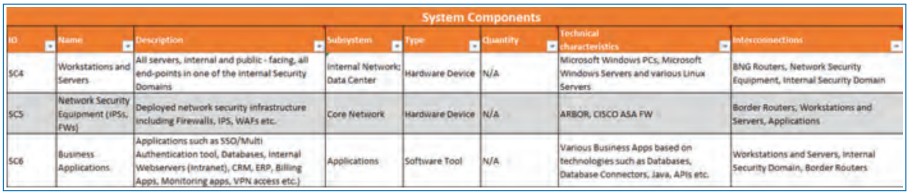
Figure 20.3. The RESISTO resilience management approach (adapted from [18]). This graphic includes the different collection methods and supporting tools necessary to complete certain steps of the process.

step risk management process, however, not its extension to system performance function-based resilience assessment.

Steps 2 through 4, and step 8, all use a tabular approach (in Figure 20.3—the extended threat list) to obtain the necessary information from end users. Steps 2 through 5 also use the Shiny app that helps with visualization of the tables and connections. Additionally, step 2 uses the testbeds and network representations to complete the system analysis. Step 4 additionally uses the testbed tools such as penetration tests and honeypots to identify disruptions. To complete step 6, the resilience quantification, the network is represented with testbeds, and simulations are completed to determine the resilience. The resilience modification options, step 8, utilize all of the steps previously mentioned including the tabular approach, shiny app, testbed tools and network simulators. More details on each of these steps and how RESISTO completes them can be found in the following sections.

20.3 System and Threat List

In this section, a tabular approach used to collect, assess, and interpret the information obtained from the resilience management process is introduced (see also [16]). Four tables were created including system components, system functions, threats and improvement measures. These tables correspond with a specific step in the resilience management process. After the tables are completed with all the necessary information, a Flexible Resilience Analysis Template (FRAT) analysis is completed utilizing the Shiny package written in the R programming language [22].



System Components

ID	Name	Description	Subsystem	Type	Quantity	Technical characteristics	Interconnections
SC4	Workstations and Servers	All servers, internal and public - facing, all end-points in one of the internal Security Domains	Internal Network; Data Center	Hardware Device	N/A	Microsoft Windows PCs, Microsoft Windows Servers and various Linux Servers	BNG Routers, Network Security Equipment, Internal Security Domain
SC5	Network Security Equipment (IPSs, FWs)	Deployed network security infrastructure including Firewalls, IPS, WAFs etc.	Core Network	Hardware Device	N/A	ARBOR, CISCO ASA FW	Border Routers, Workstations and Servers, Applications
SC6	Business Applications	Applications such as SSO/Multi Authentication tool, Databases, Internal Webservers (Intranet), CRM, ERP, Billing Apps, Monitoring apps, VPN access etc.)	Applications	Software Tool	N/A	Various Business Apps based on technologies such as Databases, Database Connectors, Java, APIs etc.	Workstations and Servers, Internal Security Domain, Border Routers

Figure 20.4. Screenshot of the system components table that was filled in by end users.

20.3.1 System Components

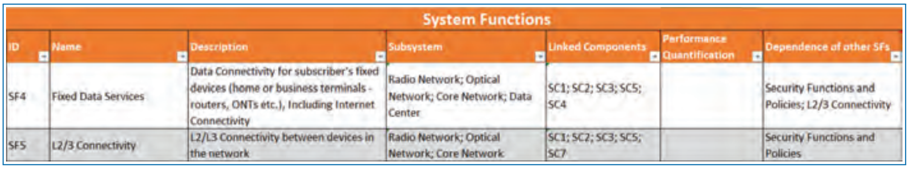
Information on system components (SCs) is obtained from process step 2: system analysis. In this step, end users report the SCs in a table format (see Figure 20.4). Each component reported includes information such as the ID, name, description, subsystem, type, quantity, technical characteristics, and interconnections.

Many of the columns allow for any text to be written inside, such as technical characteristics. This column includes any information that is necessary to investigate how disruptions affect the component and how the component functions. For example, this can include the hardware brand or series, how much energy it consumes or data it needs.

A few columns have drop down options for selection instead of free text. For example, the subsystem column has seven different options to choose from. The options include five different networks: radio, optical, satellite, core or internal network, or other subsystems such as the data center, or applications. The type column has five options to choose from: built structure, mechanical, software tool, interconnection, or hardware device. This is done to simplify the columns and allow these columns to be better organized.

From this list of components, a model of a telecommunication network is created. This is necessary for the rest of the risk and resilience management process. The model needs to be as realistic as possible, including as many components as possible, allowing for a full picture of the network to be created. However, the increased number of components also increases the complexity. In RESISTO, it needed to be decided to which level the analysis will include. To help with this decision, a column for identifying which subsystems each component belongs to was created. This table is also an important input for starting to model a quantifiable model for simulation as described in Section 20.4.

The final list of SCs for RESISTO covers the entire range of types and subsystems. The types and descriptions of the components were filled out by end users in much detail; however, other columns were missing information. This includes the quantity of the components. Obtaining a complete set of information regarding the system analysis is important for the entire resilience management process.



System Functions

ID	Name	Description	Subsystem	Linked Components	Performance Quantification	Dependence of other SFs
SF4	Fixed Data Services	Data Connectivity for subscriber's fixed devices (home or business terminals - routers, ONTs etc.), Including Internet Connectivity	Radio Network; Optical Network; Core Network; Data Center	SC1; SC2; SC3; SC5; SC4		Security Functions and Policies; L2/3 Connectivity
SF5	L2/3 Connectivity	L2/L3 Connectivity between devices in the network	Radio Network; Optical Network; Core Network	SC1; SC2; SC3; SC5; SC7		Security Functions and Policies

Figure 20.5. Screenshot of the system functions table filled with end-user input.

20.3.2 System Functions

Step three of the resilience management process is identifying system performance functions. Similar to the system components table, this was also completed with end-user input. The system functions (SFs) table consists of eight columns, to get a general idea of the different SFs. The information reported can be seen in Figure 20.5 and includes ID, name, description, subsystem, linked components, performance quantification, dependence of other SFs and comments.

Most of the SFs columns are the same as the columns for the SCs. The subsystem drop-down menu is the same, including all seven options. The linked components column allows for the identification of SCs that are needed for the function to be at full performance. This column is a drop-down menu with all of the system components previously defined. The performance quantification column helps define performance rates. The last new column is the dependence to other SFs which is also a drop-down menu that allows for the selection of other SFs.

The quantification of resilience is possible once the system performance functions are identified. For each SF defined, the performance has a specific and unique quantification. An example of a system performance function is the L2/3 connectivity, see the last line in Figure 20.5. As stated in the screenshot, this connection connects different devices together. This describes the congestion of the network. If packets become stuck in the nodes and cannot reach their destination because the node or arc capacity is exceeded, the system is said to be congested. This can occur, of course, also in a fully connected network.

The quantity L2 connectivity is defined with Equation (20.1) according to [23]. In this equation, $B*$ is the largest number of paths going through a node in (betweenness centrality) the network and S is the steps the packet takes before going missing. The last variable, μ, comes from how the capacity of the nodes is defined. The exponentially distributed time required for one packet to be fully served is the capacity. This distribution has a mean of $1/\mu$. The result of this equation (ρ_c) is the mean value of packets created at each node before congestion occurs.

$$\rho_c = \frac{\mu(S-1)}{B*} \tag{20.1}$$

Another example of a connectivity is the L1 connectivity. The L1 connectivity provides L1 radio and fiber optic (FO) links between equipment. It highlights whether the system is connected or not.

If everything is connected, the components can communicate with each other. The connectivity can be quantified using:

$$0 \leq connectivity = \frac{size\ of\ the\ giant\ component}{size\ of\ the\ network} \leq 1 \tag{20.2}$$

For RESISTO, the size of the network and giant component is defined by the number of nodes each has. The giant component is the largest connected cluster of nodes within the network.

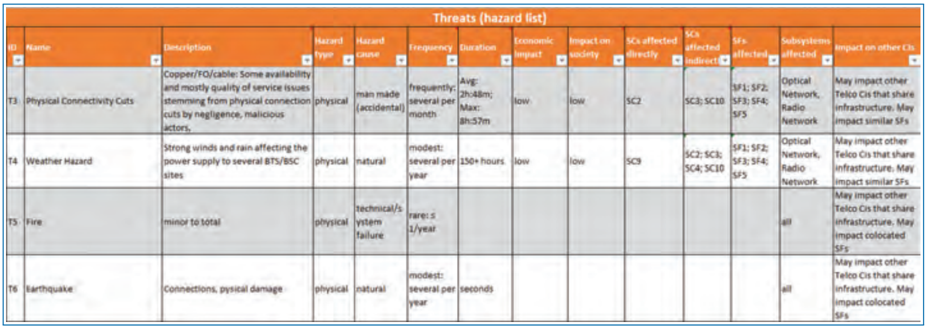
Common SFs for networks that are delivering a service are related to how well that service is delivered. For telecommunication networks, this can be mobile data services, fixed data services, or voice services. In RESISTO, the voice service is used as a SF. To calculate the performance of the voice service, the L1 connectivity was used. Nodes inside the largest connected component are investigated to determine if they are required for voice service. The number of nodes that are required that are in the largest connected component is divided by the total voice service nodes. Besides this, a SF can simply be the percentage of working components.

To quantify this performance, the percentage of undamaged components needed for the service can be used. Other quantifications can be used including standards or Quality of Service (QoS) as defined by the RESISTO partners or literature. For the voice service, the QoS is defined as having less than 150 ms for one way delay, less than 4 s for request/response delay, having a loss of less than 3 percent and jitter less than 50 ms [24]. Each service can have a QoS, which would need to be integrated into the model.

20.3.3 Threats

Step four of the resilience management process identifies different disruptions. End users report threats that make up the third table, see Figure 20.6 for an example. The information reported with the threat is more extensive than the other tables, including the hazard type, cause, frequency, duration, economic impact, societal impact, SCs that are affected either directly or indirectly, SFs that are affected, affected subsystems, and impacts on other CIs.

Drop-down menus in the threats table include the hazard type, hazard cause, frequency, economic impact, and the SCs, SFs, and subsystems expected to be affected. The hazard type can be defined as physical, cyber, or cyber-physical (including



Threats (hazard list)

ID	Name	Description	Hazard type	Hazard cause	Frequency	Duration	Economic impact	Impact on society	SCs affected directly	SCs affected indirect	SFs affected	Subsystems affected	Impact on other CIs
T3	Physical Connectivity Cuts	Copper/FO/cable: Some availability and mostly quality of service issues stemming from physical connection cuts by negligence, malicious actors.	physical	man made (accidental)	frequently: several per month	Avg: 2h:48m; Max: 8h:57m	low	low	SC2	SC3; SC10	SF1; SF2; SF3; SF4; SF5	Optical Network, Radio Network	May impact other Telco Cis that share infrastructure. May impact similar SFs
T4	Weather Hazard	Strong winds and rain affecting the power supply to several BTS/BSC sites	physical	natural	modest: several per year	150+ hours	low	low	SC9	SC2; SC3; SC4; SC10	SF1; SF2; SF3; SF4; SF5	Optical Network, Radio Network	May impact other Telco Cis that share infrastructure. May impact similar SFs
T5	Fire	minor to total	physical	technical/system failure	rare: ≤ 1/year							all	May impact other Telco Cis that share infrastructure. May impact colocated SFs
T6	Earthquake	Connections, pysical damage	physical	natural	modest: several per year	seconds						all	May impact other Telco Cis that share infrastructure. May impact colocated SFs

Figure 20.6. Screenshot of the threats collected by end-user input.

both physical-cyber and cyber-physical) due to the focus and aim of the RESISTO project.

Hazard causes can be defined as man-made (accidental), man-made (attack), technical/system failure or natural. The frequency can be defined as Equation (20.3).

$$1/year \leq frequency \leq 10/week \tag{20.3}$$

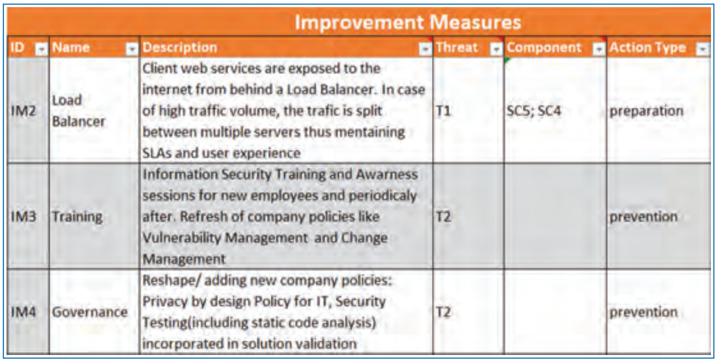
The economic impact has four different options: high, medium, low, or none. The SCs and SFs columns allow users to select options from the previously filled in tables. The subsystems remain the same as the previous tables.

Methods to determine threats include preliminary hazard analysis (PHA), historic event analysis (e.g., using databases of historic terroristic events with a high resolution of targets and tactics, see, e.g., [25]), expert opinion gather [26], or root cause analysis (fault tree analysis).

There are different categories for the threats. Some threats that are known on the level of (partial) performance function loss for telecommunication networks include loss or disruption of services, quality degradation, and data loss or leakage. This includes link problems as well as service problems. Threats that are specific to the service include problems with access, authentication or authorization.

The advantage of considering such types of threats is that they are to a large degree independent from the physical or cyber root cause. This allows to take also account of hitherto unknown disruptions, at least with respect to post-event resilience improvement and to some degree, i.e., to counter unexampled or black swan events, see also the discussion in [2].

Other threats that were collected from user input for RESISTO include Distributed Denial of Service (DDoS) attack, data exfiltration, physical connectivity cuts, weather hazards, fires, earthquakes, and power shortages.



Improvement Measures

ID	Name	Description	Threat	Component	Action Type
IM2	Load Balancer	Client web services are exposed to the internet from behind a Load Balancer. In case of high traffic volume, the trafic is split between multiple servers thus mentaining SLAs and user experience	T1	SC5; SC4	preparation
IM3	Training	Information Security Training and Awarness sessions for new employees and periodicaly after. Refresh of company policies like Vulnerability Management and Change Management	T2		prevention
IM4	Governance	Reshape/ adding new company policies: Privacy by design Policy for IT, Security Testing(including static code analysis) incorporated in solution validation	T2		prevention

Figure 20.7. Screenshot of the improvement measures table that was filled in by the end users.

20.3.4 Mitigation Measures

The last table is one for mitigation options and includes columns such as ID, name, description, subsystem, component, action type, and comments (see Figure 20.7). The action type clarifies the purpose of mitigation measure and categorizes it into one of seven different phases (preparation, detection, prevention, protection, stabilization, recovery and improvement). For each mitigation measure, they are assigned to a specific threat and associated with different components.

Examples of mitigation measures include anti-DDoS appliances or load balancers. Determining the effect of the mitigation measures on the resilience can vary in difficulty. The more technical measures, such as adding batteries to system components, can be easier to quantify than an improvement measure of increasing staff training.

20.3.5 FRAT Analysis

Once all four tables have been filled out by end users, the FRAT analysis can begin. As mentioned, the analysis was completed using R and the shiny package. Using the shiny package means that a web application can be implemented for the analysis.

The FRAT analysis includes three main parts: connections, threat ranking, and correlation. The web application allows for easier visuals of the user inputs of the SCs, SFs, and threats. For the connections in particular, by clicking on one SF, all of the SCs that are connected to it, and the threats that it is vulnerable too, are highlighted. This can be seen in Figure 20.8.

The threats can also be ranked as part of the FRAT analysis. The threats are ranked based on Equation 20.4.

$$Score = FQ \cdot (EI + SI) \qquad (20.4)$$

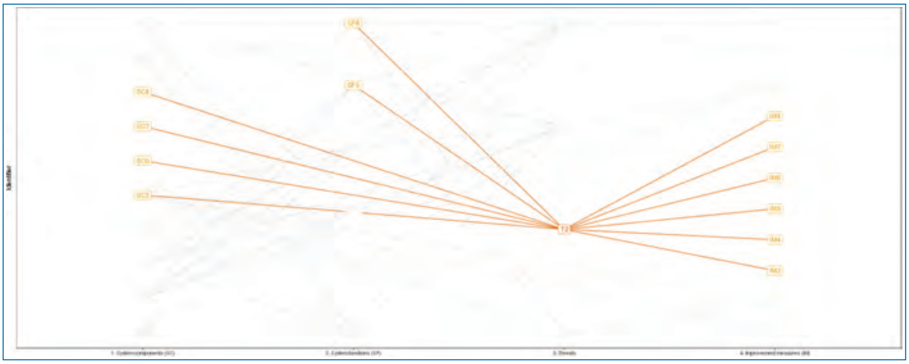


Figure 20.8. The Flexible Resilience Analysis Template (FRAT) Shiny app will display the connections between the system components, system functions, and threats. In this example, the T2 threat (data exfiltration) is highlighted. It has connections to different system components such as workstations and servers, and the business applications. It also connects to two system functions: L3 connectivity and security functions and policies. Improvement measures related to this threat include training, governance, and alerts.

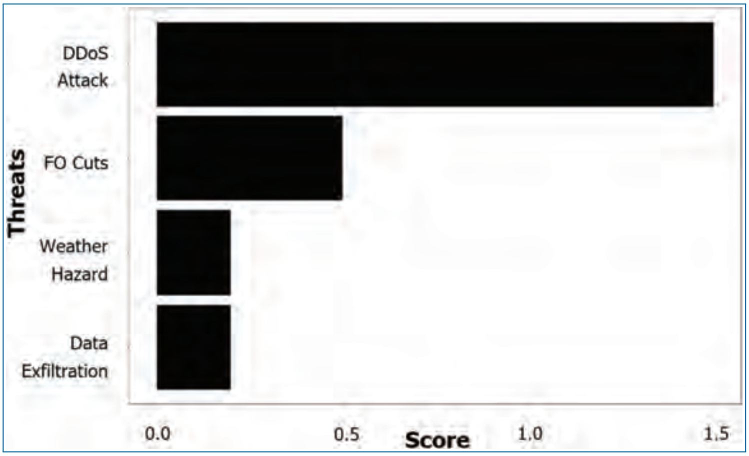


Figure 20.9. The threat ranking of the different threats as seen in the Shiny web application. This is the score of the different threats, calculated with Equation (20.4). The threat with the largest ranking is a DDoS Attack, and the lowest threat is the data exfiltration. The values on the x-axis are the scores each threat has.

This equation combines the social impact (SI) and the economic impact (EI) and multiplies it with the frequency of occurrence (FQ). It should be noted that this equation can be defined by the user, depending on where the focus is on the analysis. For each of these factors, the end users selected a value from a scale provided (i.e., for frequency values of never, rare, modest, frequent, or very frequent could be selected). These values need to be converted to numerical values. This is done by the user. Figure 20.9 shows the final score and thus the threat ranking in the Shiny app.

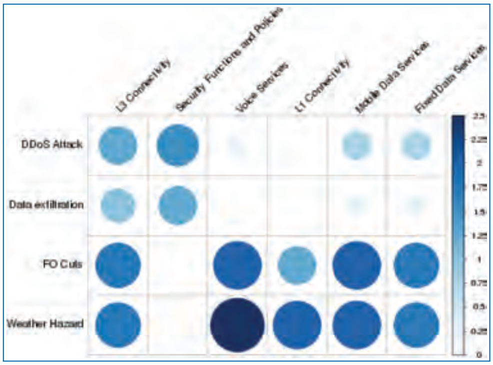


Figure 20.10. The correlation matrix from the Shiny app. This matrix has the threat on one axis and the system function on the other. The larger, darker circles indicate a more critical correlation.

Lastly, the FRAT analysis is used to complete step five of the resilience management process in RESISTO. This step completes a pre-assessment of combinations of functions and disruptions. This is done by creating and analyzing the correlation matrices of the system functions and threats. The matrix can show how threats may affect specific functions more than others, see Figure 20.10, where the larger, darker circles indicate a larger affect. The combinations that result in a large effect will then be bookmarked for further use in the resilience management cycle, where the resilience quantification will be completed.

20.4 Resilience Quantification

From the FRAT analysis and Shiny app, steps two through five of the resilience management process are completed. The next step is therefore the overall resilience quantification, step six. To complete the resilience quantification, a network scheme and network simulation is needed, since semi-quantitative tabular assessments starting out, e.g., from Figures 20.9 and 20.10 or similar are not considered as sufficient due to the highlight non-linear and coupled nature of the CI telecommunication under consideration.

The network schemes are provided through testbeds from the end users. It was determined that CaESAR is the simulation tool that fits best with the objectives of RESISTO, in terms of resilience assessment and quantification. The resilience is then specifically quantified by using resilience indicators.

20.4.1 CaESAR Simulation Tool

CaESAR (**Ca**scading **E**ffects **S**imulation in urban **A**reas to assess and increase **R**esilience) was originally created as a tool to determine the resilience of critical

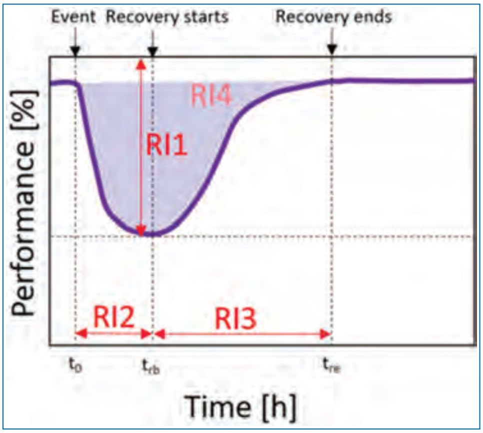


Figure 20.11. An example performance time curve similar to ones that are an output from CaESAR. In the figure are also the resilience indicators (see the next section for more information). The performance function, and the threat, can be chosen during the simulation.

infrastructures against cascading effects and other threats within the EU project SnowBall [27]. CaESAR works by creating a network model and incorporating threats and their consequences, the resilience calculations and any improvement methods [3].

Once the networks have been implemented in the tool, damages can be inserted and cascading effects can be computed. The recovery of the network can begin once the damage effects are finished, and no more cascades are occurring, if assuming that the response and recovery starts after the cascade are completed. In the real world, there may be a buffer time between a disruptive event's impact and cascading effects and the recovery. An example of a reason for this buffer time can be apprehension about the repair crews' safety and sending the crews into the environment too early. Resilience can then be quantified for the entire process.

This approach allows for mitigation measures to be evaluated in terms of the improvements (or lack thereof) of resilience. This is a cyclical process where after each iteration, critical components are identified and mitigation strategies selected, and then implemented into the networks.

As mentioned, the resilience can be quantified considering the entire process until the end of the recovery phase. The final result of this is a resilience performance time curve (Figure 20.11). This curve then gives the resilience indicators, explained in the following section, that are used to validate the simulation results.

The performance measure can be any of the mentioned system functions, such as L1 connectivity, the percentage of working components, or a function related to any of the services provided to consumers. The curve covers the entire resilience process, from before the event occurs to the response during the event and the recovery after the disruptive event ends.

The curves are also specific to a threat. At the moment, threats that are modeled in CaESAR include a failure of 20 percent of the nodes, failures of specific nodes, or assigning a probability of failure for nodes at different times to see a time-based attack.

20.4.2 Resilience Indicators

Resilience indicators (RIs) are used to link the two control loops, one offline and one online, in RESISTO. These indicators can be defined at different points throughout the process, occurring at different resilience phases. Four main resilience indicators can be defined as seen in Figure 20.11, see, e.g., the figures and terminology provided for performance and non-performance system functions in [2].

The point where the performance loss is the largest can be defined as RI1. The time it takes for recovery to start after the event has occurred is defined as RI2. RI3 is the recovery time and RI4 is the performance loss spanning the time since the event began to the time recovery was completed, assuming that full recovery has been reached (this can be calculated using integration of the curve). Many such similar quantities are accessible, see, e.g., [28].

These resilience indicators are specific to the event that was modeled and the SF. Once the RIs are known, a matrix structure is used to store them. This matrix of values can then be used to compare to the real-time RIs and validate the model.

20.5 Conclusion

The chapter gave a summary of a well-defined stepwise approach to analyze and manage jointly risk and resilience of critical infrastructure systems resorting to system performance functions and by considering all resilience cycle phases before and during disruption.

For the approach, a minimum set of tables and matrices to be used for successful implementation was discussed. The focus was on an overall tabular and matrix approach that allows to determine in an efficient way which parts of the resilience quantification needs more careful consideration, modeling, and quantitative simulation. It was shown how to realize such an approach within a web-based application software.

The FRAT approach allowed in particular to identify critical combinations of single threats and performance functions, for which sample quantitative simulations were provided. The performance function quantification can be analyzed with a set of appropriate resilience indicators, including, e.g., time till maximum loss after start of disruption, time to recovery, maximum decay of performance, overall loss or improvement of performance post recovery and learning. Such resilience indicator quantities can be used single and in combination for quantitate assessments by considering the system without risk control and resilience improvements and with selected improvement measures to select the best options for implementation.

The approach was applied to telecommunication infrastructures, but is suited more generally for socio-technical systems, including smaller technical systems. Within the presented approach, a semi-quantitative overall risk control and resilience quantity was presented taking account of superposition of risks.

In future work, besides using the network computation as input for the overall system risk sum, also further network effects could be simulated going beyond assessing the effect of single events till full recovery. An example includes to compute a stationary balance between threats and ongoing counter and improvements measures to simulate the ongoing real-time efforts that make networks operational already today. Further potential improvements of future work include the consideration of dependencies on other networks already within simulations.

Acknowledgments

This work is supported by the H2020 RESISTO project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 786409.

References

[1] I. Häring *et al.*, "Towards a Generic Resilience Management, Quantification and Development Process: General Definitions, Requirements, Methods, Techniques and Measures, and Case Studies," in *NATO Science for Peace and Security, Series C, Environmental Security, 1874–6519, Resilience and Risk: Methods and Application in Environment, Cyber and Social Domains/edited by Igor Linkov and José Manuel Palma-Oliveira*, I. Linkov and J. Palma-Oliveira, Eds., Dordrecht, Netherlands: Springer, 2017, pp. 21–80.

[2] I. Häring, S. Ebenhöch, and A. Stolz, "Quantifying Resilience for Resilience Engineering of Socio Technical Systems," *European Journal for Security Research*, vol. 1, no. 1, pp. 21–58, 2016, doi: 10.1007/s41125-015-0001-x.

[3] S. Hiermaier, S. Hasenstein, and K. Faist, "Resilience engineering – How to handle the unexpected," in *Poised to Adapt: Enacting Resilience Potential Through Design, Governance and Organization*, Liège, Belgium, 2017, pp. 92–97. Accessed: Jan. 27 2020. [Online]. Available: https://www.resilience-engineering-association.org/wp-content/uploads/2018/06/REA-Proceedings-Final-Version.pdf

[4] K. Thoma, B. Scharte, D. Hiller, and T. Leismann, "Resilience Engineering as Part of Security Research: Definitions, Concepts and Science Approaches," *Eur J Secur Res*, vol. 1, no. 1, pp. 3–19, 2016, doi: 10.1007/s41125-016-0002-4.

[5] *Risk Management – Guidelines*, ISO 31000:2018, 2018.

[6] *Risk Management – Risk Assessment Techniques*, ISO – IEC 31010:2019, 2019.

[7] L. Belascu and A. Horobet, "The Standardization of Risk Management Practices at the International Level," *Ovidius University Annals, Economic Sciences Series*, vol. 0, no. 1, pp. 6–11, 2015. [Online]. Available: https://ideas.repec.org/a/ovi/oviste/vxvy2015i1p6-11.html

[8] C. Lalonde and O. Boiral, "Managing risks through ISO 31000: A critical analysis," *Risk Manag*, vol. 14, no. 4, pp. 272–300, 2012, doi: 10.1057/rm.2012.9.

[9] B. Adini, O. Cohen, A. W. Eide, S. Nilsson, L. Aharonson-Daniel, and I. A. Herrera, "Striving to be resilient: What concepts, approaches and practices should be incorporated in resilience management guidelines?," *Technological Forecasting and Social Change*, vol. 121, pp. 39–49, 2017, doi: 10.1016/j.techfore.2017.01.020.

[10] I. Häring, B. Scharte, A. Stolz, T. Leismann, and S. Hiermaier, "Resilience Engineering and Quantification for Sustainable Systems Development and Assessment: Socio-technical Systems and Critical Infrastructure," in *IRGC Resource guide on Resilience*, Lausanne: Lausanne, EPFL International Risk Governance Center (IRGC), 2017, pp. 81–89. Accessed: Jan. 27 2020. [Online]. Available: https://beta.irgc.org/wp-content/uploads/2018/09/Haering-et-al.-Resilience-Engineering-and-Quantification.pdf

[11] I. Häring, B. Scharte, and S. Hiermaier, "Towards a novel and applicable approach for Resilience Engineering," in *Proceedings of the 6th International Disaster and Risk Conference*, Davos, Switzerland, 2016, pp. 272–276. Accessed: Jan. 27 2020. [Online]. Available: https://beta.irgc.org/wp-content/uploads/2018/09/Haering-et-al.-Resilience-Engineering-and-Quantification.pdf

[12] M. Leitch, “ISO 31000:2009 – The new international standard on risk management,” *Risk Analysis: An Official Publication of the Society for Risk Analysis*, vol. 30, no. 6, pp. 887–892, 2010, doi: 10.1111/j.1539-6924.2010.01397.x.

[13] D. Baumann, I. Häring, U. Siebold, and J. Finger, “A web application for urban security enhancement,” in *9th Future Security, Security Research Conference: Berlin, September 16–18, 2014 Proceedings*, Berlin, Germany, 2014, pp. 17–25. Accessed: Jan. 27 2020. [Online]. Available: http://publica.fraunhofer.de/eprints/urn_nbn_de_0011-h-513078.pdf

[14] U. Siebold, S. Hasenstein, J. Finger, and I. Häring, “Table-top urban risk and resilience management for football events,” in *Safety and Reliability of Complex Engineered Systems: Proceedings of the 25th European Safety and Reliability Conference, ESREL 2015, Zürich, Switzerland, 7–10 September 2015/editors, Luca Podofillini, Bruno Sudret and Božidar Stojadinović, Enrico Zio, Wolfgang Kröger*, ETH Zurich, Switzerland, 2015, pp. 3375–3382.

[15] C. Schoppe, I. Häring, and U. Siebold, “Semi-formal modeling of risk management process and application to chance management and monitoring,” in *Safety, Reliability and Risk Analysis: Beyond the Horizon/editors, R.D.J.M. Steenbergen, TNO, Delft, the Netherlands [and three others]*, Amsterdam, Netherlands, 2014, pp. 1411–1418.

[16] M. Fehling-Kaschek *et al.*, “A Systematic Tabular Approach for Risk and Resilience Assessment and Improvement in the Telecommunication Industry,” in *Proceedings of the 29th European Safety and Reliability Conference*, Hannover, Germany, 2019, pp. 1312–1319. Accessed: Feb. 14 2020. [Online]. Available: http://publica.fraunhofer.de/documents/N-572367.html.

[17] RESISTO, *Resilience Enhancement and Risk Control Platform for Communication Infrastructure Operators: EC Grant Agreement ID*: 786409. [Online]. Available: https://cordis.europa.eu/project/id/786409# (accessed: Jan. 27 2020).

[18] M. Fehling-Kaschek and K. Faist, “Risk and resilience management process for cyber-physical threats of telecom CI,” RESISTO, Aug. 2019. Accessed: Feb. 14 2020. [Online]. Available: http://www.resistoproject.eu/resources/

[19] I. Häring *et al.*, “Analytical engineering process to identify, assess and improve technical resilience capabilities,” in *Safety and Reliability. Theory and Applications: ESREL 2017 (Portoroz, Slovenia, 18–22 June, 2017)*, Portorož, Slovenia, 2017, pp. 1069–1079.

[20] S. Tomforde, P. Gelhausen, C. Gruhl, I. Häring, and B. Sick, "Explicit Consideration of Resilience in Organic Computing Design Processes," in *ARCS Workshop 2019; 32nd International Conference on Architecture of Computing Systems*, Technical University of Denmark, Copenhagen, 2019. Accessed: Jan. 27 2020. [Online]. Available: https://www.vde-verlag.de/proceedings-en/564957007.html

[21] I. Häring and P. Gelhausen, "Technical safety and reliability methods for resilience engineering," in *Safe Societies in a Changing World: Proceedings of ESREL 2018*, Trondheim, Norway, 2018, pp. 1253–1260. Accessed: Jan. 27 2020. [Online]. Available: https://www.taylorfrancis.com/books/e/9781351174664/chapters/10.1201/9781351174664-158

[22] W. Chang, J. Cheng, J. J. Allaire, Y. Xie, and J. McPherson, *shiny: Web Application Framework for R*. [Online]. Available: https://cran.r-project.org/package=shiny# (accessed: Jan. 10 2020).

[23] J. Duch i Gavaldà, "Structure and Traffic on Complex Networks," Dissertation, Universitat de Barcelona, Barcelona, 2008. Accessed: Jan. 13 2020. [Online]. Available: https://www.tdx.cat/handle/10803/21775?locale-attribute=en

[24] J. Mirkovic *et al.*, "Measuring Impact of DoS Attacks," in *Proceedings of the DETER Community Workshop on Cyber Security Experimentation*, Arlington, Virginia, 2006. Accessed: Jan. 13 2020. [Online]. Available: Proceedings of the DETER Community Workshop on Cyber Security Experimentation.

[25] U. Siebold and I. Häring, "Terror event database and analysis software," in *Fraunhofer Symposium Future Security: 4th Security Research Conference: September 29th–October 1st 2009, Karlsruhe, Germany*, Karlsruhe, Germany, 2009, pp. 85–93.

[26] A. Mosleh, "Ask the expert: Plenary talk," Jun. 19 2018.

[27] SnowBall, *Lower the Impact of Aggravating Factors in Crisis Situations Thanks to Adaptative Foresight and Decision-support Tools: EC Grant Agreement ID: 606742*. [Online]. Available: https://cordis.europa.eu/project/id/606742# (accessed: Jan. 27 2020).

[28] S. Hosseini, K. Barker, and J. E. Ramirez-Marquez, "A review of definitions and measures of system resilience," *Reliability Engineering & System Safety*, vol. 145, pp. 47–61, 2016, doi: 10.1016/j.ress.2015.08.006.

DOI: 10.1561/9781680836875.ch21

Chapter 21

CISIApro Critical Infrastructures Modeling Technique for an Effective Decision-Making Support

By Chiara Foglietta and Stefano Panzieri

Modeling critical infrastructure interdependencies is mandatory to assess the consequences of adverse events such as natural disasters, failures, and also cyberattacks. However, interdependencies can be exploited during the recovery phase for increasing the effect of the countermeasures. In this chapter, we present CISIApro 2.0, an agent-based simulator that assesses the consequences of negative events on interconnected infrastructures, describes as devices and services. The output of CISIApro 2.0 is the set of possible devices and services which are affected by an adverse event. The simulator has been tested using a telecommunication network.

21.1 Introduction

Critical infrastructure concept (or essential services, following the European law) has changed during the last 40 years: critical infrastructures are large and geographically extended systems that are a fundamental part of our lives. The actual definition of critical infrastructures is defined in terms of national security, leading to 17 sectors, including also agriculture and food systems, national monuments and icons, commercial and government facilities. We prefer the concept of "lifeline systems," (O'Rourke, 2007), defined for evaluating the performance of large, geographically distributed networks during the crisis caused by adverse events, such as natural events or cyberattacks. Lifelines are grouped into six principal systems: electric power, gas and liquid fuels, telecommunications, transportation, waste disposal, and water supply. Those systems are tightly linked with the economic well-being, security, and safety of our lives.

Lifeline systems all influence each other. Lifeline systems are dependent one to the others, primarily by physical proximity and operational interaction. However, the pervasive introduction of telecommunications leads to the introduction of the Industrial Internet of Things (IIoT) and 5G also in critical infrastructures. Telecommunications are mandatory for remotely telecontrolling sites and devices through SCADA (Supervisory Control And Data Acquisition) networks. SCADA networks are now connected to Ethernet-based networks, thanks to open-source protocols.

Assessing risk in critical infrastructure is a well-known problem. The Department of Homeland Security (DHS) defines risk as "the potential for an unwanted outcome resulting from an incident, event, or occurrence, as determined by its likelihood and the associated consequences" (Committee *et al.*, 2013). Risk is thus traditionally defined as a function of three elements: the threats to which an asset is susceptible, the vulnerabilities of the asset to the threat, and the consequences potentially generated by the degradation of the asset. Risk management involves knowing the threats and hazards that could affect an asset, assessing the vulnerabilities of the asset and then evaluating the impacts on the asset. Based on these characteristics, it is possible to develop specific indicators and metrics to assess the risk to critical infrastructure.

Also, the concept of resilience (Bruneau and Reinhorn, 2007), like the concept of critical infrastructure, is evolving: the resilience of an organization reflects the degree of preparedness and the ability to respond to and recover from a disaster or, in general, a negative event. Because lifeline systems are intimately linked to the economic well-being, security, and social fabric of a community, the initial strength and rapid recovery of lifelines are closely related to community resilience (Bruneau *et al.*, 2003).

The concepts of risk and resilience are similar and they are tight connected: improving the resilience of the system means decreasing risk. Risk is usually organized in terms of preparedness, mitigation measures, response capabilities, and recovery mechanisms; the traditional components of resilience are anticipation, absorption, adaptation, and recovery. Risk is usually related to a possible metric for understanding the consequences of adverse events; resilience is the ability to decreasing the effects of adverse events. In this chapter, the two concepts are exploited for understanding the consequences of adverse events (such as natural disasters, cyber-attacks or faults) and the consequences of restoration or mitigation actions.

21.1.1 Contributions

The modeling approach exploited in this chapter is based on the Mixed Holistic Reductionist (MHR) approach, where each infrastructure is divided into components (reductionist layer), services (service layer) and holistic nodes (holistic layer). This approach is then applied using an agent-based simulator, called CISIApro 2.0. This simulator can represent the consequences of adverse and positive events in an interdependent scenario. This simulator runs in real time connected to a SCADA control center to acquire updated information on faults and connected to an Intrusion Detection System (IDS) to acquire actual threats and on-going cyber-attacks. CISIApro 2.0 can integrate heterogeneous data to improve the situational awareness of operators and their decision-making process.

21.1.2 Organization

The paper is organized as follows. Section 21.2 reviews the literature on modeling activities for critical infrastructure interdependencies; in Section 21.3, the Mixed Holistic-Reductionist (MHR) approach is presented, with Section 21.4 presenting the agent-based simulator called CISIApro 2.0; the case study is detailed in Section 21.5, in terms of components and results; and, finally, conclusions and future works are discussed in Section 21.6.

21.2 Literature Review

In literature, there are three main methodologies for the modeling approaches of critical infrastructures: agent-based simulation, input-output analysis, and network modeling. In literature, it is also possible to find heterogeneous and/or unclassified approaches (Gopalakrishnan and Peeta, 2010).

The agent-based simulations consider each infrastructure as complex adaptive systems, composed of agents representing single aspects in the infrastructure itself. Different agents can be modeled at different degrees of abstraction based on the proposed level of resolution modeling. The main advantage of agent-based simulation is the ability to arise synergistic behaviors when agents are starting to interact together (Rinaldi *et al.*, 2001).

The second approach is based on the Input-Output economic analysis introduced by Leontief in the early 1930s but then adapted to modeling infrastructures. Haimes and Jiang developed the linear Input-Output Inoperability Model (IIM) to study the effect of interdependencies on the inoperability of interconnected networked systems (Haimes and Jiang, 2001). For example, we consider a two-system model. When a failure of subsystem 1 leads to subsystem 2 to be 80% inoperable, and a failure of subsystem 2 makes subsystem 1 to be 20% inoperable, the effect of functionality loss due to an external perturbation can be calculated by solving the Leontief equations. The main advantage of the IIM and its improvements is related to the simplicity and flexibility of the proposed approach. Usually, IIM is limited to the economic costs of interdependencies.

In the last years, researchers explored new approaches for modeling infrastructure interdependencies. The most promising approach is based on graph and network theory. In this approach, infrastructures are represented using abstract graphs made of nodes and arcs, standing for links between components in the infrastructures. The main advantage is to exploit closed-form expressions and numerical simulations to characterize their topology, performance, and uncertainty.

Several works reviewed the proposed approaches for modeling interdependencies among critical infrastructures; the reader can refer to Eusgeld *et al.* (2008), Satumtira and Dueñas-Osorio (2010), and Ouyang (2014) for more details on this problem.

21.3 Modeling Interdependencies with MHR Approach

In this document, we propose an approach for helping during the modeling phase. The Mixed Holistic Reductionist (MHR) (Digioia *et al.*, 2012) approach was created to exploit the advantages of both methods: holistic and reductionist. The main aim of the MHR approach is to give a possible guideline to properly model critical infrastructures and their interdependencies.

In holistic modeling, infrastructures are seen as singular entities with defined boundaries and functional properties, generating a global and overall analysis. Seeing an infrastructure as a single element aims at identifying and characterizing the different infrastructures and their geographical level. At this level, the amount of data needed for modeling activities is very low and can be found in public databases.

On the other hand, the reductionist model emphasizes the need to fully understand the roles and the behavior of individual components to truly understand the overall infrastructure. The reductionist approach drills down to each component in terms of inputs and outputs. At this level of abstraction is easy to find dependencies between the equipment and single components.

Different levels of analysis are required in modeled systems and their boundaries are lost in the event of complex case studies. With the MHR model, relationships between infrastructures could be seen at different levels through either a top-down or bottom-up approach. The other main advantage is to model infrastructures at a different level of abstraction considering the amount of available data.

The connection point between the two levels of abstraction, i.e., holistic and reductionist approaches, is the quality of service (in the following, abbreviated as "service") evaluation which is a key element for operators. This layer describes functional relationships between components and infrastructure at different levels of granularity. In MHR, services to customers and to other interconnected infrastructures are explicitly considered as a middle layer between holistic and reductionist agents.

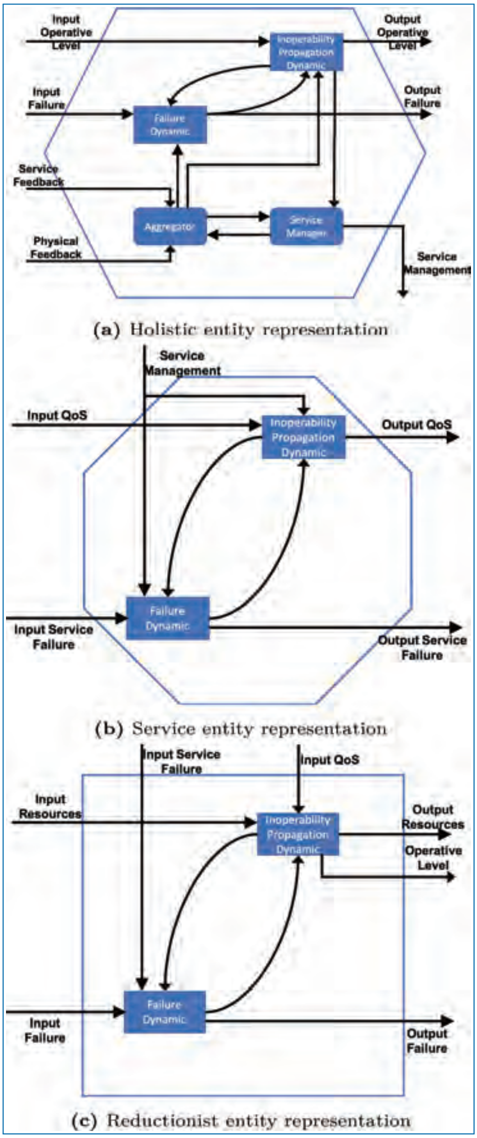
The MHR allows us to reach the right level of detail with minimal data and collected information. Some important considerations can be summarized in the following:

- Each infrastructure is modeled starting from the identification of components and their interactions.
- Each layer is defined with an appropriate level of abstraction based on information coming from end-users, stakeholders and open documents.
- Each component (we called it entity or agent) must be described in a way to decouple it from other components: the behavior of the component must depend on the valued explicitly exchanged with the other components.
- The simulator, that implements the MHR approach, must be able to represent any type of agentâĹ™s behavior for adapting to the specific reference scenario.

The MHR approach allows defining three different typologies of entities: holistic entities, service entities, and reductionist entities.

A Holistic Entity (Figure 21.1(a)) represents the infrastructure as a whole (or its general organizational divisions) in order to have a model that can consider the global dynamics between infrastructure possibly one might think of representing behaviors related to policies, strategies, etc.

A Service Entity represents a logical or organizational element, that provides an aggregate resource as the remote control: the remote control generally refers to a solution that provides supervision, by means of software and data collection. Data can be collected through telecommunication network or field equipment in case of a geographically distributed infrastructure. In Figure 21.1(b), a service component



(a) Holistic entity representation

(b) Service entity representation

(c) Reductionist entity representation

Figure 21.1. MHR representation of different entities at different abstraction layers.

is depicted considering the classical model of an agent in CISIApro 2.0. Some examples of service are:

- The ability to supply customers
- The ability to produce resources
- The ability to change the topology
- The status of some specific and important components.

Finally, with a Reductionist Entity, we can represent, with the right degree of abstraction, all physical or aggregated entities of the overall system. In Figure 21.1(c), the representation of a reductionist component is depicted. The picture does not explicitly consider a cyber threat: this malicious event can be represented in the same way as an input failure with a suitable "cyber dynamic."

The MHR approach allows the developer to represent a complex scenario into components that have different functionalities. The layers allow to model a complex scenario, made of several interconnected infrastructures, with different abstraction levels: an infrastructure can be modeled in all its features (reductionist, service, and holistic layers), another can be modeled using only the holistic layer, without any kind of problem apart from the granularity and the precision of the results.

21.4 Dynamic Risk Propagation using CISIApro 2.0

CISIApro 2.0 (Critical Infrastructure Simulation by Interdependent Agents) (Foglietta *et al.*, 2015; Masucci *et al.*, 2016) is a software engine able to calculate complex cascading effects, considering (inter)dependencies and faults propagation among the involved complex systems. CISIApro 2.0 can also consider mitigation and restoration actions to assess their positive consequences. CISIApro has been designed from scratch in 2011 within the H2020 ATENA project (Adamsky *et al.*, 2018) for improving the modeling process of the interdependencies among infrastructures. During the H2020 RESISTO project (H2020 RESISTO Project, 2018), CISIApro has been updated to version 2.0 adding some important functionalities related to the modeling of telecommunication infrastructures.

CISIApro 2.0 is an agent-based simulator, where each agent has the same structure. In particular, each agent receives resources and failures from the upstream agents and spreads it to the downstream ones, as depicted in Figure 21.2. The layers

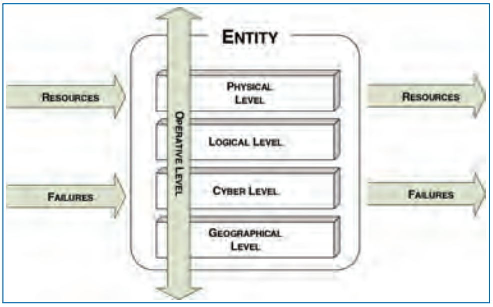


Figure 21.2. CISIApro 2.0 entity representation.

are obtained from the propagation of a resource or a fault. A resource is a service or a data produced and/or consumed by the agent, represented in CISIApro 2.0 as an entity. The entity produces or receives also failures (in general, malfunctions) representing a physical failure or a possible cyberattack. The malfunctions are spread among the agents following different propagation models that take into account the class of the interdependencies (i.e., layers) and the reliability of the information. The considered layers are physical, logical, geographical, and cyber.

The ability to produce resources is summarized by the concept of operational level, depending on the availability of received resources, on the propagation of faults, and on the functionality of the entity itself.

The operational level of each agent can be considered as a risk metric. Usually, the risk is a numeric value, from the impact severity, the likelihood of occurrence or threat, and the vulnerability analysis. In CISIApro 2.0 applications, the likelihood of occurrence is usually considered more connected to the concept of the trustworthiness of the information. For each entity, the user can add also a vulnerability variable, but in the following case study, we suppose that the vulnerability depends only on the distance from the source and on the persistence of the attack itself. The operational level of each agent is associated with a risk level: the risk is the amount of harm due to specific events, such as a cyberattack, and can be evaluated as

$$\mathit{Risk} = 1 - \mathit{Operational\ Level} \tag{21.1}$$

where 1 is the maximum value of the operational level. A higher value of the operational level means a lower risk. Therefore, the operational level represents a dynamic risk assessment considering the cascading effects of adverse events, i.e., natural disasters, failures or cyber attacks. This value is normalized for each infrastructure considering the quality of service towards customers and other infrastructures.

CISIApro 2.0 is mainly composed of two modules, as depicted in Figure 21.3. The first one is the off-line tool known as “CISIApro 2.0 Design” that allows the design and implementation of complex and highly interdependent scenarios as represented in Figure 21.4. While the second one is the on-line tool called CISIAmat (or “CISIApro 2.0 Run”) which exploits Simulink Mathworks for the real-time engine actually connected to near real-time data sources.

CISIApro 2.0 is a software platform based on a database-centric architecture in which the database plays a crucial role, called in Figure 21.3 “CISIApro 2.0 DB”. This means a centralized asynchronous design that allows horizontal scalability where each element of the risk propagation architecture, independently, interfaces with the centralized database to acquire the last data from the field and the output of CISIApro 2.0.

“CISIApro 2.0 Run” engine provides an impact evaluation of detected anomalies. In order to mitigate the effects, the decision-maker, also supported by a workflow manager, can choose among different sequences of possible reaction

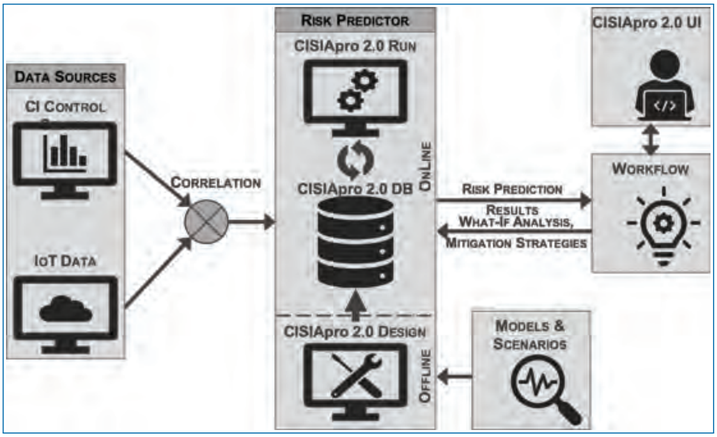


Figure 21.3. CISIApro architecture.

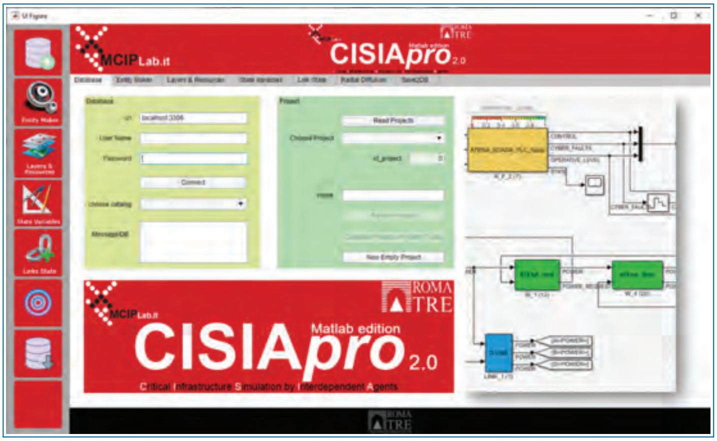
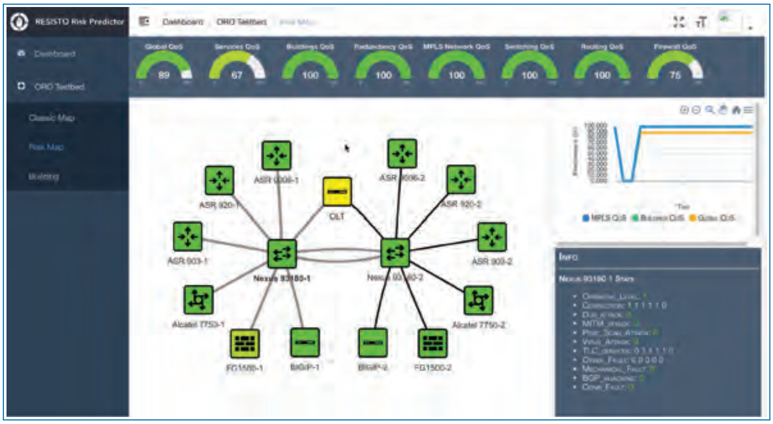


Figure 21.4. Graphical user interface of CISIApro 2.0 design.

strategies, taking into account also the output of CISIApro 2.0. CISIApro 2.0, starting from the actual scenario and QoS (Quality of Service) levels of involved devices, simulates What-If scenarios to provide useful information for the decision-maker with respect to forthcoming critical situations.

21.5 A Telecommunication Scenario

The proposed scenario is built in a redundant telecommunication architecture: most resources, both physical equipment and virtual machines are indeed doubled. This scenario simulates a production network and includes most of the monitoring capability and capacity of the production networks. We consider two critical services in the use case: voice and data communications over 4G and fixed networks.



Info

Nexus 93180 1 State

- Operating_Level: 1
- Connection: 1 1 1 1 1 0
- DoS_attack: 0
- MITM_attack: 0
- Port_Scan_Attack: 0
- Virus_Attack: 0
- TLC_service: 0 1 1 1 1 0
- Cyber_Fault: 0 0 0 0 0
- Mechanical_Fault: 0
- BGP_hijacking: 0
- Cond_Fault: 0

Figure 21.5. CISIApro 2.0 User Interface.

The scenario is equipped with various network elements ranging from high-speed backbone routers to mobile and B2B (Business to Business) access routers. The devices create a Multi-Protocol Label Switching (MPLS) network and are aggregated using high-speed links on the pair of Nexus switches, see Figure 21.5. The MPLS layer contains routers from a wide range that are used to deliver backbone functionality router (such as CISCO ASR 9006), mobile Router (i.e., CISCO ASR 903) and B2B Routers (i.e., CISCO ASR 920). MPLS Network will also integrate an OLT (Optical Line Termination) to simulate an attack/outage on the part of the network.

The security fabric and data-center layer are achieved using a few next-generation security devices and application controllers as:

- Fortinet FortiGate (URL Filtering, Centralized Antivirus, Intrusion Detection and Protection System, E-mail filtering, Layer 4 Firewall)
- F5 BIGIP (Web Application Firewall).

The testbed also includes various servers that run VMWare and OpenStack hypervisors for virtualized solutions and data-center services emulation. In Figure 21.5, the User Interface of CISIApro where information is displayed: the network topology with colors from red (complete unavailability of the device) to green (device properly working), the values of considered services placed in the upper side of the picture, and the complete set of state variable for each entity (in the right bottom side of the picture).

The scenario considers two possible events:

1. An unintentional fiber cut severs the connections between the two core routers in the scenario. The fiber cut is followed by a large-scale DDoS (Distributed Denial of Service) attack on one of the border routers.

2. A human actor enters one of the core network buildings and attempts to connect to a border router, access its administrative console and maliciously change a route to one of the servers hosting a critical part of the core network.

The consequences in both cases are related to the risk of connection loss and service delivery failure. Other consequences are related to the infrastructures that are connected to the telecommunication to properly deliver commands to the field devices.

21.6 Conclusions and Future Works

This chapter describes CISIApro 2.0 simulator: CISIApro 2.0 is an agent-based simulator aiming at assessing the consequences of adverse events in an interdependent scenario. CISIApro 2.0 has two distinct phases: the first is the modeling activities and the second is the real-time simulator that evaluates the consequences of the adverse events connected to heterogeneous data sources. The output of CISIApro 2.0 is exploited in the decision-making process, to improve the operator situation awareness and to make better decisions knowing which are the consequences of actual events.

The presented scenario is a small part of a larger case study in which CISIApro 2.0 in under validation with the help of stakeholders and end-users. Actual developments are related to the modeling activities for 5G networks in terms of network function virtualization and software-defined networks.

Acknowledgments

This chapter is partially supported by the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 786409 (RESISTO – RESIlience enhancement and risk control platform for communication infraSTructure Operators).

References

Adamsky, F., M. Aubigny, F. Battisti, M. Carli, F. Cimorelli, T. Cruz, A. Di Giorgio, C. Foglietta, A. Galli, A. Giuseppi, *et al.*. 2018. "Integrated protection of industrial control systems from cyber-attacks: the ATENA approach." *International Journal of Critical Infrastructure Protection*. 21: 72–82.

Bruneau, M., S. E. Chang, R. T. Eguchi, G. C. Lee, T. D. O'Rourke, A. M. Reinhorn, M. Shinozuka, K. Tierney, W. A. Wallace, and D. Von Winterfeldt. 2003. "A framework to quantitatively assess and enhance the seismic resilience of communities." *Earthquake Spectra*. 19(4): 733–752.

Bruneau, M. and A. Reinhorn. 2007. "Exploring the concept of seismic resilience for acute care facilities." *Earthquake Spectra*. 23(1): 41–62.

Committee, R. S. *et al.* 2013. "DHS Risk Lexicon: 2010 Edition (Washington, DC: Department of Homeland Security, September 2010), 26." *As of December* 29.

Digioia, G., C. Foglietta, S. Panzieri, and A. Falleni (2012. "Mixed holistic reductionistic approach for impact assessment of cyber attacks'. In: *2012 European Intelligence and Security Informatics Conference*. IEEE. 123–130.

Eusgeld, I., D. Henzi, and W. Kröger (2008. "Comparative evaluation of modeling and simulation techniques for interdependent critical infrastructures." *Scientific Report, Laboratory for Safety Analysis, ETH Zurich*: 6–8.

Foglietta, C., C. Palazzo, R. Santini, and S. Panzieri (2015. "Assessing cyber risk using the CISIApro simulator'. In: *International Conference on Critical Infrastructure Protection*. Springer. 315–331.

Gopalakrishnan, K. and S. Peeta (2010), *Sustainable and Resilient Critical Infrastructure Systems: Simulation, Modeling, and Intelligent Engineering*. Springer.

H2020 RESISTO Project. 2018. "Resilience enhancement and risk control platform for communication infrastructure operators." URL: http://www.resistoproject.eu/.

Haimes, Y. Y. and P. Jiang. 2001. "Leontief-based model of risk in complex interconnected infrastructures." *Journal of Infrastructure Systems*. 7(1): 1–12.

Masucci, D., C. Palazzo, C. Foglietta, and S. Panzieri. 2016. "Enhancing decision support with interdependency modeling'. In: *International Conference on Critical Infrastructure Protection*. Springer. 169–183.

O'Rourke, T. D. 2007. "Critical infrastructure, interdependencies, and resilience." *BRIDGE-Washington-National Academy of Engineering*. 37(1): 22.

Ouyang, M. 2014. "Review on modeling and simulation of interdependent critical infrastructure systems." *Reliability engineering & System Safety*. 121: 43–60.

Rinaldi, S. M., J. P. Peerenboom, and T. K. Kelly (2001. "Identifying, understanding, and analyzing critical infrastructure interdependencies." *IEEE Control Systems Magazine*. 21(6): 11–25.

Satumtira, G. and L. Dueñas-Osorio. 2010. "Synthesis of modeling and simulation methods on critical infrastructure interdependencies research." In: *Sustainable and Resilient Critical Infrastructure Systems*. Springer. 1–51.

DOI: 10.1561/9781680836875.ch22

Chapter 22

Modern Innovative Detectors of Physical Threats for Critical Infrastructures

*By Rodoula Makri, Panos Karaivazoglou, Alexandros Kyritsis,
Michael Skitsas, Nikolaos Koutras, Javier Valera,
and Jose Manuel Sanchez*

Nowadays, the types of threats against Critical Infrastructures are becoming more sophisticated imposing the use of equally modern detection measures. The involved aspects are too important when considering both direct physical threats and physical threats that enable malicious impact to the cyber domain as well. The Chapter begins with an overview of the current situation in Critical Infrastructures in terms of detecting physical threats, attacks, or hazards and continues by introducing modern detecting techniques covering a wider range of threats. These vary from systems with sensors for airborne threats along with audio and visual analytics up to using the wireless networks themselves as sensing systems by exploiting their networking features.

22.1 Introduction

The most common impression when discussing in general terms about physical security in critical infrastructures (CI) is that of dealing mainly with the protection of building sites and internal equipment from theft, vandalism, natural disasters (i.e., floods, earthquakes), manmade catastrophes, and accidental damage (e.g., electrical surges, heavy rains, and lightning) or unintentionally destructive acts. In this context, physical security requires solid building construction, suitable emergency preparedness and procedures, reliable power supplies, adequate climate control, and appropriate protection from intruders. In order to accomplish building sites be safeguarded in a way that minimizes the risk of resource theft and destruction, decision-makers must be also concerned about regulations governing equipment placement and use, product handling, and relationships with outside contractors and agencies [1].

However, physical security is often a second thought when it comes to information security. Since physical security has technical and administrative elements, it is often overlooked because most organizations focus on "technology-oriented security countermeasures" to prevent hacking attacks [2, 3]. Hacking into network systems is not the only way that sensitive information can be stolen or used against. Physical security must be implemented correctly to prevent attackers from gaining physical access and cause physical damages and consequently endanger information systems. In this context, cyber threats, apart from relative direct actions, can be also seen as a result of physical security breaches ("cyber-physical threats"). To this end, the physical element of security is often disregarded; hardware damages or vandalism could occur while working with administrative and technical controls as well, while organizations often focus on them, and as a result, security breaches may not be discovered right away [4].

Physical security is often thought as only controlling personnel access to facilities; however, its relation to achieving data center availability goals is more than crucial. Physical security professionals are more concerned about the physical entrance of a building environment and what damages a potential intruder may cause. As new technologies such as biometric verification and remote management of security data become more widely available, traditional card-and-guard security is being supplanted by security systems that can provide identification and tracking of human activity in and around the data center. The challenges of implementing physical security are much more important now than in previous decades; laptops, tablets, and smartphones all have the ability to store sensitive data that can be lost or stolen. CI organizations are obliged to safeguard personnel, information, equipment, IT infrastructure, data, facilities, systems, and company assets (and all information and software contained therein). Thus, before investing in equipment, they must carefully evaluate their specific security needs and determine the most appropriate and cost-effective security measures for their facility [5].

22.2 Current Mechanisms and Sensors for Physical Security

A brief overview of the current physical security solutions, mechanisms, and sensors for protecting CI assets is given in this Section. Usually, the most sophisticated physical security mechanisms are implemented for the protection of central buildings and headquarters. In most of the cases, integrated unified security systems are implemented addressing mainly visual inspection and access control, following a predefined plan.

The first step in a security plan is to identify the areas, rooms, and entry points that need different rules of access. Thus, different levels of security are employed depending on potentially stringent access methods to achieve added protection. By this way, an inner area is protected both by its own access methods and by those of the areas that enclose it. In addition, any breach of an outer area can be met with another access challenge at a perimeter further in. This can be employed in areas that might have concentric boundaries (i.e., site or building perimeter, computer area, and computer rooms along with equipment racks) or with side-by-side boundaries (i.e., visitor areas, offices, utility rooms) [5].

Thus, the strategies used to protect the organization's assets need to have a layered approach. It is harder for the attackers to reach their objective when multiple layers must be bypassed to access a resource. Various access levels and fragmentation/separation of the building in zones are then employed, depending on the foreseen physical threats and the type of area to be protected so that to prevent unauthorized people to enter or access the site and use equipment. The main physical security mechanisms along with related sensors or components currently used per case are described in the following:

22.2.1 Access Control and Personal Identification

Access control takes place at main entrances (input/output) especially for personnel and vehicles and in internal or other critical areas (i.e., data center). For the organization's staff and visitors, various methods may be of use involving personal identities, entrance cards, and, in most sophisticated cases, biometrics and related measures. For critical or high security areas, simultaneous monitoring by visual surveillance and closed circuits (CCTV) usually takes place along with checks of incoming/outgoing people, recording of activities, tamper alarms, etc. Usually, guests must be accompanied while entrance may be allowed only to predefined program-based areas. Almost in all cases, CCTV and Access Control systems are interconnected. Depending on the desired security level, special measures include detection of explosives, chemicals, or even of weapons and metal objects. Intercommunication of the entrance checkpoints with the control room usually takes place, while silent panic buttons in all positions are also available.

Concerning entrance cards, a large variety exists on the market including personal cards with data and photo, limited validity cards for visitors/suppliers, smart cards with onboard processor, magnetic stripe cards (with a simple magnetic strip of identifying data) or magnetic spot cards (barium ferrite cards), bar-code cards, and infrared shadow ones. The types of interaction with the relevant card readers may be of swipe, insert, flat contact, and even with no contact as in proximity cards or proximity tokens. Apart from their ability to be reprogrammed, the above types of cards present resistance to counterfeiting and ability to allow access only in permitted areas (floors) per employer. The programming process may be separate for visitors and vendors, while an escort mode may be also used in certain cases in conjunction with anti-passback mode.

Keypads and coded locks are also in wide use. They are reliable and very user-friendly, but their security is limited by the sharable and guessable nature of passwords as personal access codes (PAC) or personal identification numbers (PIN) are implied.

The most sophisticated tools are the biometrics sensors, which are used mostly for verification of the identity rather identification of a person. Biometric scanning techniques and relevant sensors have been developed for several human features: fingerprint scanners, iris pattern, face biometrics models, retina pattern, handwriting, and voice recognition systems. Biometric devices are generally very reliable. The main sources of unreliability for biometrics is the possibility that a legitimate user may fail to be recognized ("false rejection") and the erroneous recognition, either by confusing one user with another or by accepting an imposter as a legitimate user ("false acceptance"). Nowadays, a wide range of biometric devices exist on the market, used either independently as stand-alone or in combination to other measures like smart cards, which is highly likely to become mainstream. The main considerations when choosing a biometric technique are the equipment cost, the failure rates, and the user acceptance, especially in cases where specific conditions should be regarded (i.e., distance from the sensors, light, etc.).

In parking areas and at all entrances, gardens, or facilities that vehicles may access, the checking is being held by reading of the vehicles' plates with high definition and surveillance cameras (Vehicles' plate recognition system). Recording of the surrounding area and parking management systems can also be applied, using license plates databases of employees and access control procedures for transporter/carriers.

Explosives Detectors can also be used in parallel to access control in high security areas or at the entrances. Different detection systems (i.e., gates) for explosives and hidden objects (i.e., weapons) are applied with units capable of detecting any such material or metal within hand-luggage's or carried by visitors, highlighting also the material with color. Advanced detectors can be employed guaranteeing minimum detection time while interconnections with the operator's security system provide remote control and silent alarm capabilities.

22.2.2 Perimeter Defense Mechanisms

In demanding situations, sophisticated methods such as perimeter defense are also applied. These include both internal (headquarters or main buildings) and external cases (outer area perimeter) incorporating a variety of sensors and combined detection/protection systems.

Building perimeter: This refers mostly to major buildings of the infrastructure and headquarters. The building perimeter usually includes a combination of the following measures: access control on inputs/ outputs of the building and of sensitive critical areas, surveillance and tracing of interior violations, motion detectors and magnetic contacts, webcams around the building, tamper alarm even glass (window) breaking detection.

Outer perimeter: On the other hand, the outer perimeter defense may incorporate the whole infrastructure area by employing full monitoring and tracking of the point of violation, cameras and sirens alert, visual surveillance of the perimeter, enhanced lighting in dark areas and hidden (i.e., infra-red) illumination along with electronic fencing systems.

The sensing systems that are usually employed in these detection and protection mechanisms against physical threats include the following:

Access control systems: With biometrics or card readers as already described earlier.

Optical (Camera-based) surveillance systems: These include the following sensors:

- *High-definition IP cameras for outdoor use*: capable for day/night operation, wide zones (i.e., 10 m) coverage, motion detection, enabling parallel monitoring (in control rooms or checkpoints) with automatic recordings and alarm;
- *Optical surveillance with high-analysis cameras*: Megapixel cameras are deployed around the building, supervising all surroundings. They offer the ability to analyze individual images from a single camera, resulting up to

10 times (and more) larger coverage than a simple camera, with multiple digital extended zoom;

- *Panoramic surveillance of the environment*: Pan, tilt, and zoom (PTZ) cameras are installed on the roof featuring high resolution of 520 TVLines, high sensitivity less than 1 lux, and powerful zoom (×35 or more). They enable overriding their control from the control room and interface with the perimeter tampering system.

Apart from the high-resolution cameras and access control system interconnections, the perimeter lighting can be also enabled through infrared (IR)/invisible lighting elements. The camera-based surveillance systems can also accommodate vehicle license plate recognition, are often combined with silent panic buttons or with electronic fencing system, and assist to the intercommunication of the entrance checkpoints with the control room.

Intrusion detection systems: Intrusion detection sensors provide an all-around glazing at the ground and lowest floors. The functionalities include the surveillance of openings (doors, windows) and sliding doors, while the relevant sensors include magnetic contacts, motion detectors, and crystal breaking detectors among others.

Electronic fencing systems: these are more sophisticated and more expensive solutions which incorporate invisible underground (buried) sensing cables. The fencing systems create an invisible detection field, not affected by vegetation or the natural environment. The detection range varies with usual widths of around 3 m and with accuracy of less than 1 m. Graphic on-line representation can be visible to the central control, while there is the possibility of partial or total activation of the fence or interfacing with CCTV.

Associated operation procedures: The above sensing and detection systems are usually combined with associated operation procedures of the CI's security staff. Security Guards patrols take place through specific route and time planning based on patrol scenarios, crosschecked with CCTV and intrusion detection systems. Evacuation plans are also being practiced involving specific measures, i.e., automated output counting units and similar.

22.2.3 Discussion on Current Physical Security Mechanisms — Identification of Gaps

The above detection and sensing systems are administrated as standalones or within unified security management systems (SMS), either outsourced to specialized companies or managed inherently by each operator or both. The SMS often involve central control rooms and are designed to be compatible with International Security

Standards and the organizations' procedures, providing functionalities such as integration with IT applications and TCP/IP LAN networks, client–server remote monitoring, encryption of network data and data integrity, double routes or dual flows, checks with internal and external blacklists and provision of direct, through complex reports. However, it should be noted that usually the SMS, although conducting the orchestration of the sensors and surveillance mechanisms, only indicate the faults providing alarms, while the faults/alarms' handling and tackling is being managed by the operator instead.

It is clear that the abovementioned physical security aspects affect all types of CIs which in turn also make use of telecom infrastructures. Therefore, large interdependences with the telecom CIs are shown for delivering data or notifications to the SMS which might not be noticed if the telecom infrastructure is breached. Security equipment deployed on site can have different levels of integration with the telecommunication system; many of them rely on the possibility to use the site LAN and local networks via ethernet or via WiFi. This, however, makes them open to any attack to the telecom infrastructure and availability.

For the Telecom CIs, the issues of both physical and cyber threats are of major importance since they greatly affect one another. Nevertheless, although cyber threats are given the major attention, which is reasonable since data security is a primary factor and are usually handled by the telecom provider's Network Operation Centre (NOC), the physical security threats in telecom CIs are not regarded evenly. Usually, the most sophisticated physical security mechanisms, as described earlier, are implemented for the security of the central telecommunication buildings and headquarters where large number of the telecom organization's personnel is employed and is present on day-to-day basis, along with the core of the telecom assets (i.e., NOCs and main backhaul fiber optics terminals).

Although the above hierarchy is considered reasonable due to the critical units involved, it should be noted, however, that this is rather not the case for all the telecom assets that a telecom provider possesses. Telecom pillars, antenna parks, or even fiber optics terminals in remote and rural areas are not given equal security treatment as in large central telecom buildings, since it is more expensive. Sophisticated security methods are rarely observed in decentralized structures; the situation is even more critical for assets such as antenna parks or facilities at remote islands or mountainous regions. There, security guards may be employed, while ordinary wire fencing is the main protection measure.

On the other hand, cost is one of the most important parameters, since the investment in advanced security mechanisms for the many remote assets could turn to be high enough, limiting the cost-effectiveness, not considering the needed resources in time and effort for the relevant implementations. Thus, it is most common for major telecom organizations to invest in disaster or redundancy centers, back-up infrastructure and networks, along with failover techniques for the main services, facilitating management from a central station to relevant checkpoints all over the country. In general, redundancy networks, disaster centers, or failover capabilities are common, and almost obligatory, for all types of CIs, tackling not only issues associated with physical intrusion but also with potential losses due to overloads or natural disasters such as floods, lighting, and earthquakes.

It should be highlighted, though, that despite the offered control operations, the main functionality pursued by current security systems, mechanisms, and sensors apart detection is deterrence. Thus, the principles governing the basic procedures of the usual physical security plans are detection, deterrence, response, and recovery/re-evaluation. As mentioned earlier, in current CI's security systems, detection is seen more as a physical detection of an intruder, attempting to illegally access premises and facilities. To this end, the overall security measures and sensing systems employed mainly focus to address the personnel identification and access control and thus are mainly meant for deterrence purposes. The deterrence principle is conducted through the organization's policy, operating procedures, and control, and similarly, the response principle is mostly addressed by the guard staff through the same procedures. Recovery and re-evaluation mainly affect the operational actions of the security staff.

However, the procedures used mainly address a rather limited implementation than a holistic tackling of the principles addressed. Additionally, the use of all the above security mechanisms depends on their relevant cost which is being regarded as an important factor. For example, limitations and weaknesses can be noticed in most of the current cases: i.e., reluctance in employing the newest commercial models, possibly due to cost reasons, may result in inefficient access management or poor integration of ambient information. Furthermore, in case of inherently managed security systems, best practices are difficult to be followed, resulting in limited protection of entry/exit points or even limited scalability maintaining only a minimum coverage of the surroundings (i.e., only the main building).

Moreover, newest trends in physical security such as pattern recognition and machine learning techniques are rarely employed to classify persons, vehicles, and other objects moving within the controlled area and to extract profiles and relevant semantic information for event processing and further analysis with correlation platforms. Current processes often require large effort of the operator and the security personnel when monitoring a huge number of sensors on a day-to-day basis.

Another important issue that needs to be taken into account is that, due to the increase in malicious actions nowadays and the large amount of services provided, the threats in modern CIs involve quite more complex aspects than the ordinary physical security systems can handle. Furthermore, all kinds of issues endangering

CI facilities, including airborne threats, are on the table for that matter and need to be confronted. As malicious acts and terrorism threats turn to be more advanced and sophisticated nowadays, it seems that current approaches should be enhanced with more flexible solutions that could integrate more advanced mechanisms along with an increased degree of assessing resilience. Additionally, the impact that physical threats may have on cyber aspects (cyber-physical threat) seems often to be disregarded or dealt separately, and this is too important especially for the telecom CIs. Presently, physical intrusion is not only meant for theft or to cause physical damages but rather to enable hazards and damages at the cyber domain of a CI, i.e., to install malicious dormant software or enable hacking actions. To this respect, modern and novel techniques are needed to address cyber-physical threats along with holistic solutions to provide suitable correlation of physical and cyber events.

22.3 Modern Detectors of Physical Threats in CIs

In the previous Sections, an overview of the current security and protection sensors and mechanisms against physical threats in CIs was described, while the gaps in the implementation were identified. As discussed, the increased security requirements nowadays imply the employment of new approaches and solutions. In this context, modern innovative sensing mechanisms are presented for the detection of physical and cyber-physical threats. The presentation will focus on the functionalities of various platforms of active and passive sensors for direct detection (i.e., physical intrusion, airborne threats, etc.) along with wireless networks acting themselves as sensing units. All these techniques are implemented within the framework of the EU H2020 RESISTO project [6].

22.3.1 Audio and Video Analytics Sensor Platforms and Monitoring Tools

Video and Audio sensors are widely used in surveillance operations and protection of critical infrastructures. However, the emerging sophisticated types of threats impose the demand for adding intelligence, resulting in integrated audio and video analytics tools. In this context, intelligence algorithms are applied in audio and video streams for the real-time detection of events and for the early identification of illicit activity. Pattern recognition and machine learning techniques are used to extract acoustic events (i.e., gunshot, screaming, glass breaking) or to classify persons, vehicles, and other objects that are moving within the controlled area of the infrastructure.

A smart surveillance system that constitutes from video sensors (cameras), embedded on other computational units and video analytics algorithms, consists the Video Analytics Component (VAC). Video segments of interest are generated (upon the detection of alert) and stored for further use and notification of security personnel. Various types of CCTV cameras (i.e., fixed Outdoor IP-based CCTV camera and a PTZ camera) can be used to provide seamless image of the area of interest, supporting Pan, Tilt, and Zoom operations so that the camera is moved to a desired location.

Highly sensitive microphones, such omnidirectional and array ones (microphones operating in tandem), can be used as well for the Audio Analytics Component (AAC). The omni type is meant for acoustic event detections, while the array is meant to locate the source of the sound event. Both types can be attached to an embedded PC (e.g., Raspberry 3), where the first level of detection algorithms is executed. Beyond the acoustic event detection, audio analytics are enhanced with intelligent algorithms capable for localization of the source of the detected event (estimating the position of the acoustic source). This feature is used as an input for the steering of the video sensors (i.e., PTZ video sensor) in order to adjust their position to the source of the acoustic event. The whole tool is developed by Aditess Ltd, Cyprus within the framework of the RESISTO project.

By this way, the added intelligence transforms the sensors to an integrated audio/video-based surveillance system; each component can be also used as standalone or be integrated as an overall sub-system to a holistic integrated platform. The added value of this integration, between audio and visual sensors, lies in the ability to provide to the security operator a real-time picture of the field where the event occurred, through a generated video clip and an alert notification. Furthermore, cross-correlations of audio and video analysis results are developed in order to provide more accurate and precise alerts. This intelligent process reduces the effort of the operator by monitoring in a 24/7 base a huge number of sensors. Additionally, the early detection of events and the ability to extract semantic information (i.e., type and location of event, illegal access in restricted area) can provide useful data to event processing and correlation platforms for further analysis.

22.3.2 UAV Platform-based Sensors and Mini-UAV Systems

Drones and mini Unmanned Aerial Vehicles (mini-UAV systems) are an emerging new concept worldwide which, with a relatively low cost, are used in a variety of applications, commercial to surveillance ones. Apart from existing commercial platforms, customized platforms can be designed, such as of multirotor, helicopter, and fixed-wind types, as those developed by Aditess Ltd, Cyprus, within the framework of the RESISTO project.

These UAV platforms can be optimized for certain surveillance and inspection needs with various payloads such as lightweight miniature cameras and electro-optic sensors (i.e., daylight, IR, thermal cameras), GPS modules, chemical sensors as well as intelligent web-based software for the coordination of UAV platforms and sensor setup in order to fulfill every security requirement. This configuration includes the selection of the appropriate platform and payload along with the communication with the ground control station (GCS), based on several criteria including the mission analysis (path, time, covered distance, range), communication (Air to Ground and Ground to Infrastructure using LTE/4G or physical network if applicable), and risk metrics among others.

The Mini-UAV systems can also enable navigation in controlled civilian airspace according to relevant aviation standards and certification requirements that will become applicable, addressing homeland security applications and flight operations. Especially for rural and remote areas, friendly drones and UAVs can enhance the video surveillance system with aerial images and be used to evaluate the emerging anti-drone technologies.

22.3.3 Active and Passive Detectors of Airborne Threats (i.e., UAVs)

The rapidly proliferating use of unmanned devices, in many aspects of commercial and everyday life, has brought about new and emerging challenges. In many cases, such as regulating air traffic and security, early detection of such objects is more than crucial. Furthermore, drones and UAVs can be maliciously used as potential types of human-driven physical threats against civilian critical infrastructures imposing the need for relevant detection measures. To this respect, especially for civilian CI cases, the focus is given specifically in detecting airborne threats as UAVs and drones, flying at rather short ranges from the targeted CI. The aim is to be able to use low-cost sensors that can provide early warnings of an airborne threat and potential intrusion, to a central security platform.

In light of the emerging use of unmanned devices, UAVs or drones are nowadays more and more considered as potential human-driven physical threats. Anomalies and airborne threats to CIs or specific telecom ones (i.e., remote antenna parks and/or telecom pillars on high rooftops) are mainly monitored through visual methods (i.e., cameras); the threat has to be in rather close proximity to be detected which leaves less time for reaction. Counter-UAV technology has already seen extensive use and a continuously growing interest. Main challenges include detection effectiveness, false negatives/positives, distinguishing legitimate and illegitimate drone use, legal framework, and lack of standards [7].

The specific airborne threat detection system, developed by the Institute of Communication and Computer Systems (ICCS), Greece, within the framework of the RESISTO project, is a set of tools designed and developed to detect the presence of UAVs as airborne threats and to provide alarm signals. The system consists of active and passive sensors, namely radar and acoustic sensors, respectively [8]. These system's components can be used either separately or in combination. The detectors are a small lightweight continuous wave Doppler radar combined with acoustic sensors, such as array of high sensitivity microphones for the reception of signals emitted by such UAV platforms. Those signals may be in the microwave/RF (e.g., remote controls) or the sound/acoustic frequency region (e.g., drones propulsion). Hence, the focus is on exploiting the combined imprint of UAVs as captured by microphone and radar systems.

A Doppler radar uses the Doppler Effect to derive velocity data about moving objects at a distance. An accurate measurement of the original transmitted frequency and the reflected return frequency (echo) determines the Doppler frequency shift, which is a direct indication of the target's speed, the radial component of a target's velocity relative to the radar. The measured speed is relative to the range of the target (radial distance) along with the target's angle of arrival. The Doppler CW radar is able to detect and track fast-moving targets providing good visibility in harsh conditions (dusk, rain, or snow). In contrast to optical or infrared systems, Doppler radars detect more accurately small-sized objects in the (optimal) 3 GHz–30 GHz frequency band with adequate sensing ranges. However, it should be noted that radar's detection capability depends on the target's Radar Cross Section (RCS: the effective aperture of the target) and the ability to distinguish between small objects and clutter that the sensor will also pick up especially for objects with low RCS as the drones are [9]. Generally, the RCS reflects the degree that the target can be detected and depends on a variety of parameters such as the relative position of the target in respect to the radar, the frequency of operation along with the relative incident and scattering angles, and, most importantly, the target characteristics, namely the aperture/profile, its material, and dimensions relatively to the wavelength. Low RCS makes the detection within cluttered environments very demanding with trade-offs between false alarm rates and detection probabilities [10]. To this respect, low RCS of airborne objects, such as UAVs and drones, is tackled with advanced signal processing and the combined use of other sensors such as the acoustic ones.

The acoustic sensors used in the present system are a set of high sensitivity dynamic microphones forming an array. Acoustic sensors have many advantages that include non-line-of-sight, omni-directionality, passiveness, low-cost, and low-power, playing a potential key role in situational awareness; since they do not depend on the target's size, but rather on its acoustic signature, i.e., the sound of the engine. Nevertheless, acoustic sensing depends on the environmental conditions and related sources of acoustic attenuation (e.g., temperature, wind speed, and direction). Acoustic microphone arrays are used as a second sensor modality to detect broadband acoustic emissions from approaching targets, either forming linear arrays or diagonal (i.e., 4 microphones arranged in a cross format), which yield to be the most optimal ones. By exploiting the target's strong emitted sound harmonics mainly in the 20 Hz–2 kHz range, moving targets can be detected and tracked regardless of their size by acoustic sensors [11].

Having captured the sound signal of the target, signal detection and processing methods can be used. For example, time-domain waveform cross correlation of the captured waveform with a previously recorded sound waveform of the target as a reference along with performing in parallel Harmonic Line Association in the frequency domain, in order the necessary results to be extracted. Advanced signal processing and machine intelligence/machine learning techniques are applied to the radar and acoustic data, both in the time-domain and the frequency-domain to achieve detection and to estimate the target's angle of arrival and range/velocity. In terms of the detection and tracking, the radar sensors detect the presence of the small unmanned aircraft by its radar signature (often employing algorithms to distinguish between other small, low-flying objects), while the acoustic ones detect drones by recognizing the unique sounds produced by their motors. The system can detect the target's movement when the drone is approaching or moves away; even when the drone is at hover mode, the derived conclusion is that the radar mainly detects the fast movement of the propellers. Neural network techniques are expected to increase the ability of distinguishing low RCS targets and to advance the overall performance.

Combinations of the two methods are beginning to emerge as integrated solutions for monitoring the airspace over critical infrastructures: aligning the results and theoretical basis with modern system implementations, progressing by this way the relevant state of the art for a low cost, low power combination. Recent technology trends show that detecting low RCS moving targets can be made feasible by implementing mixed techniques [12]; the emission of high-frequency waves (active methods) may be complemented by receiving the acoustic output (passive) of such systems to accomplish detection. These emerge as a promising solution also combined with visual methods (i.e., cameras) or electro-optical systems which detect the UAVs based on their visual signature and infrared (IR) ones based on their heat signature.

The whole system may act either as stand-alone or as plug-in module in a wider platform architecture providing alerts and potential intrusion events corresponding to the presence of potential moving airborne threats. The use of multiple detection elements is intended to overcome the inherent limitations of each technology and to increase the probability of a successful detection, given that no individual method is entirely failproof. Therefore, the employed combination of active and passive sensors may provide additionally situational awareness and perimeter CIs defense against low-flying threat aircrafts.

22.3.4 Wireless Networks as Sensing Systems Against Physical Threats

Considering the emerging 5G telecom infrastructures, Internet of Things (IoT) networks are expected to be expanded in the near future and to dominate in everyday life with the use of wireless networks and wireless sensor networks distributed in large areas and infrastructures. Thus, it is of great importance that these telecom facilities are capable of adequately tackling physical as well as cyber-physical threats. While sensing networks with advanced features are foreseen, it is reasonable that these networks obtain additional security functionalities, in order to be capable of detecting and responding to risks and physical threats, being either human driven or as consequences of natural disasters.

This notion presupposes certain added functionalities both in the hardware and software/firmware of these wireless networks as current and future parts of CIs so that to act by themselves as detecting systems. In the following, two examples of such types of applications are being given, as developed in the framework of the RESISTO project by Integrasys SA, Spain, also involving features of the Guard-Time SA KSI Blockchain, Estonia. These applications are the RADIOFILTER and RANMONITOR ones and are briefly described below:

RADIOFILTER Tool

RADIOFILTER is a tool which offers detection, location, and reporting of WLAN-based threats and attacks to critical infrastructures protected assets.

The tool is based on a network of N-distributed passive Secured Cyber Sensors, deployed at an infrastructure, which continuously scan the data-link traffic and relay this information to a central processing node (Central Node). This setup enables the system to monitor the data-link (layer 2) traffic parameters in 802.11 WLAN networks in order to detect, locate, and report 802.11 WLAN-based threat events. The events and useful related information can be visualized through an external Web User Interface for stand-alone use. The integrity of the cyber sensors' firmware is taken care of by a firmware update server connecting to an external Guardtime KSI server for generating KSI-blockchain signatures of the sensor firmware as a cryptographic timestamp. These firmware security anchors are used to check that the new firmware version has not been maliciously tampered with.

The specific RADIOFILTER application is ideal in cases that a CI operator needs to protect the inner part of a building, whereas a WLAN Network (i.e., 802.11) is installed, from confidential information stealing and service disruption risks. In such case, the operator's internal WLAN network needs to be protected from unauthorized devices installed. These intrusion devices are not the access points, devices, and client connections whitelisted by the CI operator, which have location and connectivity restriction and are allowed inside the infrastructure (centralized, ad hoc, local, and internet allowed for each WLAN client). In other words, the WLAN only accepts a specific set of client devices and any unauthorized device connection is dealt as an unauthorized intrusion.

Following a relevant survey, the RADIOFILTER radio monitoring sensors are placed within the inner part of the building. Each monitoring cyber sensor is powered by a small single-board computer (i.e., Raspberry) and is able to monitor the IEEE 802 b/g/n/ac, (2.4 GHz and 5 GHz bands) WLAN technologies. The sensor also includes a secure element used to encrypt and store critical data such as root certificates and firmware security anchors. The system collects data captured by the cyber sensors from all the access points, devices, and connections in the infrastructure; estimates a specific access point or device location using a fingerprinting method based on machine learning techniques; checks the different infrastructure whitelist databases for event detection; and generates events in case that an unauthorized device attempts to be connected.

By this way, various cases of unauthorized connections can be detected. For example, A non-whitelisted mobile phone with WLAN connectivity (Unauthorized Device Event Detection) or a non-whitelisted WLAN Access Point (Unauthorized Access Point Event Detection) may be a threat event. As soon as the mobile phone's WLAN connection or the Access Point's connection are activated, these events are detected, as the BSSID (MAC Address) of the device or Access Point are not in the relevant operator's whitelists. The same takes place also in the case that a non-whitelisted WLAN ad hoc connection between authorized clients (Unauthorized Connection Event Detection) is used. Thus, a new threat event detect message is generated and displayed on the RADIOFILTER Web User Interface as a flashing indicator upon the location's map.

RANMONITOR Tool

RANMONITOR offers detection and reporting of threats and attacks to LTE Radio Access Network (RAN). The main RAN attacks that the tool is able to monitor are the following: full-band or partial interference, protocol-aware jamming, rogue base station, IMSI-catcher (International Mobile Subscriber Identity), and poorly configured base stations.

The tool is based on one or several Radio-Cyber RAN Sensors that can be deployed anywhere, which passively scan the cellular radio spectrum as well as the downlink control channel information and relay this information to a central processing node (Central Node). This setup enables the system to monitor the cell parameters and status in LTE RAN networks in order to detect, locate, and report threat and attack events which can be visualized also through a Web User Interface for stand-alone use.

The RANMONITOR concept of operation is similar to the RADIOFLTER's one; however, the sensors are different since they are suitable for LTE RANs networks. Each monitoring radio-cyber RAN sensor powered by a standard computer module allows for appropriate signal acquisition and processing while the monitoring functions are performed by a radio cellular modem and by a Software Defined Radio (SDR) platform able to scan all surrounding channels. The RANMONITOR tool is suitable for LTE mobile network operators in order to protect the availability of the service from any degradation in a set of cells located in a specific area, in other words, to protect the network users against attacks generated from, i.e., IMSI-catcher devices.

Following a survey to provide an estimation of the proper location, the RAN sensors are suitably placed as well as the general parameters of the system are configured so that upon operation a set of threats and attack events to be detected by the tool. These threats include jammer performing a protocol-aware jamming attack, an unintentional interferer causing full-band or partial interference, as well as rogue base stations.

In the case that an unauthorized LTE cell (rogue base station) attempts to get connected and starts broadcasting its physical level information parameters, the RANMONITOR tool extracts the base station identification parameters checking with the mobile operators' network information, cell list, and cell map databases. Thus, an Unauthorized Cell Event is detected by RANMONITOR tool since this cell was not in the RAN cell list. The new event and its related information appear at the tool's events log table, and also the rogue cell is indicated at the RANMONITOR Web User Interface, including the detected cells and band visualization at the moment of the intrusion.

22.4 Conclusion

The present chapter describes the status of the sensors and security measures that are currently used in critical infrastructures while, additionally, presents novel detection mechanisms to enhance detection, protection, and security against intrusions and modern sophisticated physical threats. As it is seen, sensors and tools with various maturity level are offered in order to fill in the gaps in physical security in existing CIs and to provide advanced features in detection and protection processes addressing the modern needs in confronting risks and attacks. The concept of employing wireless networks as sensing networks by themselves, using firmware methods is also presented. The foreseen tools involve applications of emerging technologies in order to address intrusion events in light of the nowadays situations both in attacks and in technology trends.

Apart from the detection systems' description, the aim of the specific chapter, concerning security, is also to prove the concept that new types of sensing systems for the detection of modern threats can be successfully integrated to an overall holistic platform that would address the whole circle from prevention, detection, protection, response, and mitigation in critical infrastructures enhancing in the end their overall resilience.

Acknowledgments

This work is supported by the EU-H2020 RESISTO project, which has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No. 786409 (http://www.resistoproject.eu/).

References

[1] https://nces.ed.gov/pubs98/safetech/chapter5.asp IES/NCES, National Center for Education Statistics, US Department of Education.

[2] S. Harris, "Physical and Environmental Security" in CISSP Exam Guide 6th ed., pp. 427–502, USA McGraw-Hill, 2013.

[3] D. Hutter, "Physical Security and why it is important" GIAC (GSEC) Gold Certification, SANS Institute InfoSec Reading Room, Copyright SANS Institute, https://resources.infosecinstitute.com/importance-physical-security-workplace/, https://www.sans.org/reading-room/.../physical/physical-security-important-37120, 2016.

[4] S. Oriyano, "Physical Security" in Cehv8: Certified Ethical Hacker Version 8 Study Guide, pp. 393–409, Indianapolis, IN USA: Wiley, 2014.

[5] S. Niles, "Physical Security in Mission Critical Facilities," White Paper 82, rev. 2, APC White Papers, part of the Schneider Electric white paper library produced by Schneider Electric's Data Center – Science Center, https://it-resource.schneider-electric.com/.../wp-82-physical-security-in-mission-critica, 2015.

[6] http://www.resistoproject.eu/

[7] A. H. Michel, "Counter Drone systems," Center for the Study of the Drone at Bard College, Edited by Dan Gettinger, February 20, 2018, http://dronecenter.bard.edu/counter-drone-systems/.

[8] Al. Kyritsis, R. Makri, M. Gargalakos and N. Uzunoglu, "Active and Passive Methods for the Detection of Drones and Small Airborne Objects," 4th International Conference on "Operational Planning, Technological Innovations and Mathematical Applications OPTIMA 2017" pp. 190–191, Hellenic Military Academy, Athens, Greece, 25–26 May 2017.

[9] M. Ritchie, F. Fioranelli, *et al.* "Micro-Drone RCS Analysis," pp. 452–456, 2015 IEEE Radar Conference, Johannesburg, October 2015.

[10] Y. Bar-Shalom, *et al.* "Tracking in a Cluttered Environment with Probabilistic Data Association," Automatica, 11: pp. 451–460, 1975.

[11] T. Pham and L. Sim, "Acoustic detection and tracking of small, low-flying threat aircraft," 23rd Army Science Conf., Orlando, FL, 2002.

[12] W. Shi, G. Arabadjis, B. Bishop, P. Hill, R. Plasse and J. Yoder, "Detecting, Tracking, and Identifying Airborne Threats with Netted Sensor Fence," The MITRE Corporation Bedford, Massachusetts, U.S.A, Chapter in Book: "Sensor Fusion – Foundation and Applications," Dr. Ciza Thomas (Ed.), ISBN: 978-953-307-446-7, InTech, Available from: http://www.intechopen.com/books/, 2011.

DOI: 10.1561/9781680836875.ch23

Chapter 23

The Ethical Aspects of Critical Infrastructure Protection

By Marina Da Bormida

Critical infrastructures across different sectors are being strongly affected by the introduction of the IoT paradigm, CPS systems, intelligent digitally empowered devices, Big Data analytics, AI, and machine learning. Alongside an array of benefits, this transformational path also poses not only additional risks to their operation and security but also legal and ethical challenges and concerns for developers, practitioners, participants, and policy-makers, ranging from data protection and privacy preservation, to dataveillance, social cooling and dictatorship of data, to data ownership and access aspects, to safety, responsibility and liability, algorithmic bias and others.

The regulatory landscape is fragmented and runs at a much lower pace than technological development. Novel "soft law" tools, capable of giving granular and practical guidance, as well as ethics-related standardization initiatives, like the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, provide complementary rules and useful insights to traditional legal instruments to overcome or mitigate the given challenges raised by these technologies.

Other driving factors towards legal compliance and ethically-sound design, development, and operation of such developments are the Privacy and security by design and ethics & rule of law by design approaches, the regulatory sandboxes, and the cross-fertilization of law and technology, such as certain forms of automated compliance tools [6].

23.1 Introduction

Critical infrastructures that support the operation and development of our societies across several sectors, like finance, healthcare, and energy, are being strongly affected by the introduction of the IoT paradigm, CPS systems and intelligent digitally empowered devices, such as sensors, to robots, smart wearables, smartphones, and drones, as well as by other emerging ICT technologies, like Big Data analytics, AI, and machine learning. Furthermore, critical infrastructures' digitalized and interconnected operations, business processes, and decision-making imply large collection and processing of increasingly amount of field data, often exchanged between relevant stakeholders within the given value chain. The boundaries between the physical and digital worlds are vanishing, and the digital control of physical processes is a reality.

This transformational path has multiple benefits, in terms of increasing the efficiency and sustainability of current practices and better performance gains, for instance, unfolding a range of possibilities to discover, manage, orchestrate, and control physical space to realize coordinated behaviors within and across devices.

At the same time, this metamorphosis also poses additional risks to critical infrastructures' operation and security. A novel range of cybersecurity challenges sums up to the traditional physical security ones faced by critical infrastructure operators, giving rise to the emergence of integrated approaches for critical infrastructures security, simultaneously protecting cyber and physical assets.

The introduction of integrated security systems into critical infrastructures poses ethical and legal challenges for developers, practitioners, participants, and policymakers.

In conjunction with the array of expected benefits of these systems, unintended negative effects might occur and need to be avoided, or at least minimized, by thinking ahead, while at the same time ensuring that these technologies can benefit everyone, upholding legal concepts and ethical values, protecting human safety, physical integrity, dignity, intimacy, autonomy, and self-determination [1, 2].

One of the challenges related to these systems is to maximize security, and therefore the utility of the overall systems towards this direction, while protecting human rights, preserving ethical values, and respecting the regulatory framework.

The ethical dimensions of these systems need to be explored in an attempt to balance them with the protection of the critical infrastructures against physical and cyberattacks within the EU. Ethical risks, including data-related risks, have to be mitigated both at design time and run time, ensuring that architectures are safe and secure but also adhere to and promote European values (e.g., democracy, privacy safeguards, equal opportunities). Fair, trustworthy, ethical, and regulatory frameworks aimed at ensuring the compliance to the legislation enforcing these values should be perceived, rather than as restrictive, as an opportunity and competitive advantage, even more if taking place alongside technological developments.

In fact, an adequate ethical and legal framework, properly tackling with human-centered challenges and which would ensure that the solutions and services are designed and used in an ethical manner, is therefore critical to ensure trust in the security ecosystem around critical infrastructures, which, in turn, is essential to the acceptability of the technological artifacts, offering services and experimentation opportunities to the whole range of stakeholders across the critical infrastructure value chain [1].

The next paragraph will focus specifically on some key questions and ethical challenges, which can provide a basis for the development of an ethical and legal framework on systems relying on holistic approaches for critical infrastructures security against cyber and physical threats.

23.2 Legal and Ethical Challenges

As underlined in the previous paragraph, the technological changes related to the Cyber-Physical System (CPS), Artificial Intelligence (AI), Internet of Things (IoT), and integrated security systems introduction in the critical infrastructures, while carrying the potential to yield new solutions and opportunities for business, government, and societies, also generate new risks, concerns, and challenges in multiple contexts.

IoT, AI, blockchain and Distributed Ledger Technology (DLT), collaborative and intelligent devices, cyber ranges, cyber-physical systems are expected to optimize and make more secure and more efficient the critical infrastructures' processes. These processes and the operation of the system are fuelled by digital assets like digital twins, operational data, and machine learning models. They manifest both in the cyberspace and in the physical world, depending on underlying cloud-based infrastructure and other operational and information technology infrastructures, often geographically spread.

Being such digital assets and infrastructure increasingly interconnected, automated, and geographically distributed, not only the security challenges are greater but also ethical concerns and the risk of non-compliance with internationally recognized human rights, such as the right to privacy. The same apply to AI-supported technology with, for instance, facial recognition and emotion detection.

The increasing fragmentation in the legal and regulatory landscape at global, regional, and national level contributes to make the situation even more complex and to the emergence of novel accountability challenges.

Without claiming to be complete, the following list provides hints on some of the most pressing legal and ethical issues and concerns that need to be addressed, ranging from privacy and data protection rights, to liability, inequality, discrimination, algorithmic bias and non-transparency, safety, personal autonomy, and identity [3, 7, 11, 12].

Data protection and privacy

CPS extract, collect, and share vast amounts of data to operate effectively, including sensitive information, especially in the healthcare and financial sectors. This raises privacy concerns.

The areas of interest or concern and possible issues and challenges include:

- Data practices in relation to obtaining and ensuring informed consent
- Ensuring transparency of the process by which the tools collect, process, and make use of personal data, including the terms of use of algorithms
- Materialization of the concept of privacy by design and by default in IoT, CPS, and AI applications
- Concepts of sensitiveness and vulnerability, especially in case of patients and/or people under constant direct observation or surveillance
- Sharing of private individual information collected by IoT devices with other systems and preventing the potential misuse of data
- Data collection and processing during the research, development, and testing of AI-empowered tools and CPS
- Tackling inverse privacy and safeguarding personal data rights, filling the gap between the rights enacted by the GDPR (and its 28 national implementations) and the average understanding of their implications, both from citizens and businesses, as well as their operationalization in IoT and AI settings, where sticky policies, dynamic user consent, and other developments could be further explored to to develop legally compliant, smart solutions.
- The awareness of the kind of data that is being collected and processes is often scarce, and this diminishes an individual's power and freedom.
- Considering that the human-data relation is asymmetric, individuals can feel powerless in the relation to data, and there is the risk of leading to a loss of control over the access to one's own personal data, including the so-called right to be forgotten, which is considered in the EU as one of the pillars of an individual's control over their personal data.

Dataveillance, social cooling, and dictatorship of data [5]

The risk of dataveillance and intrusive big data practices, due to the availability of more and more data sources and the easier and faster data analysis to generate insights. For instance, for addressing the security challenges posed by the critical infrastructures protection, one's position can be tracked over time, through tools like the ubiquitous use of Closed-circuit Television (CCTV) circuits, coupled with Global Positioning System (GPS) positioning in mobile devices, as well as the use of credit cards and Automated Teller Machine (ATM) cards for payments and withdrawals.

People's awareness of the possibility of being watched at any moment might result, as shown by field experiments, in the so-called social cooling, which is a side effect of Big Data, and refers to the individuals' attitude to conform to the expected norm, especially considering that our society makes extensive use of scoring systems, where critical life changing opportunities are increasingly determined by such scoring systems, often obtained through opaque predictive algorithms applied to data to determine the value of an individual or social group. This is capable of limiting people's desire to take risks or exercise free speech. Over the long term, these self-censorship, risk aversion, and waiver to the exercise of free speech might "cool down" society and produce increased social rigidity and have an impact on human ability to evolve as an inclusive society, where minority views and vulnerable people are still able to flourish.

In strict correlation with dataveillance and social cooling, another ethical concern arises. Despite the undoubted advantages of digital identities, for example, in terms of possibility to access to online contents and all related services through them, the widespread use of such identities makes possible retrieving from the web publicly available information on an individual and generating insights. This might determine the dictatorship of data, with discriminating effects, based on the representation of a person as portrayed by his/her data, as opposed to the real self. In other words, individuals are treated as mere aggregates of data and are therefore no longer respected.

Data ownership and access aspects [1, 8, 10]

Data ownership, control, and access aspects need to be investigated, as regards the claimed property right on data and information, in relation to human data interaction and interconnected devices, that is the case of data retrieved by the sensors of the objects connected to the Internet of Things, with even more complexity when the information is personal or financial data. Radio-frequency identification (RFID), GPS, and Near Field Communication (NFC) technologies allow to track the geographic place where a person is and his movements from one place to another, without his knowledge.

Ubiquitous devices embedded in daily lives in a IoT landscape, primarily collect data that is about or produced by people, either explicitly produced by themselves (such as location data in case of sharing location while running through wearable accessories) or implicitly inferred by the sensing infrastructures, in cases such as monitoring critical infrastructures. Data collection and processing serves them in a broad range of purposes in everyday life in connection, for instance, with the operation of the critical infrastructures in the health, energy and financial sectors, ranging from personal healthcare to tailored smart city services for energy savings, processing data on energy footprint of an individual's home or other situational context. In relation to the unprecedented amount of data collected by these devices, the fundamental research questions are who owns this data and who might have access to it.

The data ownership claims are also related with the risk of data monopolies and with the theme of asymmetries of powers.

In fact, data ownership might be referred also to proprietary data, not only to personal data: data producers have the interest to remain in control of their data and to retain their rights as the original owners and therefore demand for the recognition of ownership claims. However, the legal framework is uncertain and fragmented, and it is difficult to apply legal categories: for instance, data is an intangible good difficult to define, and it is not clear the legal concept itself of data ownership. Many questions arise, such as if the EU's existing law provides sufficient protection for data and, if not, what more is needed; if data is capable of ownership (sui generis right or copyright law); if and which is the legal basis for claims of ownership of data. Meanwhile, there are solutions, such as those reflecting the IDSA Data Sovereignty paradigm, that provide the factual exclusivity of data through flexible and pragmatic tools, combining agile contracting with enabling technological artifacts, able to provide certainty and predictability.

Accessibility of information

In relation to accessibility of information, a cyberattack in IoT employed in critical infrastructures, which makes the system vulnerable, might have a direct influence on people's lives, and this might happen in electric heating systems, bank and insurance IT infrastructures, food distribution networks, hospitals, transport networks, and many others.

Safety, responsibility, and liability [11, 12]

One of the main concerns, especially in relation to AI and human–machine interaction, refers to safety aspects, which are especially important as the complex, intelligent, and self-learning CPS increasingly operate in close proximity to humans.

Furthermore, also finding the initial cause and the allocation of liability might prove complex. In case of malfunctioning, who can we hold accountable and responsible for failure? Which is the position of the developer or producer of the CPS?

The theme of liability, including the identification of who is responsible – and liable – for failures and insurance instruments for products/users, is a key issue for CPS systems and their integrated security solutions to reach their full potential, especially in contexts with multiple stakeholders and decisions being made by artificial intelligence.

Increase of Digital divide [1]

Another concern regards the difficulty of some individuals in understanding and accessing services delivered through the use of these new technologies, not being familiar with them.

Algorithmic bias

Another issue pertains to the risk of algorithmic bias and in general the risk of discrimination, manipulation, misuse, and technological determinism.

23.3 The Regulatory Landscape

23.3.1 The Legal Context

The legal and regulatory framework relevant in relation to the design, deployment, and operation of integrated security systems into critical infrastructures characterized by the wide use of the IoT paradigm, CPS systems, and intelligent digitally empowered devices and other emerging ICT technologies, like Big Data analytics, AI, and machine learning, is complex, fragmented, and significantly different in each of the domains concerned (financial, energy, and healthcare sectors).

The following notes are intended only to provide an overview of the main general pieces of legislation applicable across several domains and at EU level and need to be integrated with sector-specific and national-wide surveys dwelling upon the legislation underlining the security of the infrastructure employed in each sector.

The main pieces of legislation, partially overlapping among themselves, refer to Human Rights Law, Data Protection Law, Telecommunications Law, Information Technology Security Law, Law on Trust Services, Identification, Authentication, Intellectual Property Law, Critical Infrastructures Law, include:

- Regulation 2016/679/EU 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;
- Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of

privacy in the electronic communications sector (ePrivacy Directive), which is expected to be replaced by a Regulation on Privacy and Electronic Communications, whose proposal is currently following the approval process;

- Directive (EU) 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 (NIS Directive);
- Directive 96/9/EC of the European Parliament and the Council of 11 March 1996 on the legal protection of databases;
- Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union;
- Directive 2016/1148 concerning measures for a high common level of security of network and information systems across the Union
- Regulation (EU) No 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC (eIDAS Directive)
- The set of Communications on Critical Infrastructures, including Communication 786/2006 on a European Programme for Critical Infrastructures Protection, the Communication 163/2011 on Critical Information Infrastructure Protection 'Achievements and next steps: towards global cyber-security' and others;
- Directive 2008/114/EC on the identification and designation of European Critical infrastructures and the assessment of the need to improve their protection;
- Set of Communication on data economy and Artificial Intelligence, framing the issues and discussing the evidence collected through public and targeted consultations, as well as dedicated support measures, such as the Communication COM(2017)9 "Building a European data economy," the Communication COM (2018) 237 on "Artificial Intelligence for Europe," COM (2019) 168 "Building Trust in Human-Centric AI" and other;
- The Charter of Fundamental Rights of European Union.

23.3.2 The "Soft Law"

In addition to the legislation and official regulatory instruments, complementary regulatory tools should be explored, shifting from a vision of mere legal compliance towards exploiting the possible benefits of the "soft law," considering its relationship with the traditional legal instruments and its possible role in a landscape of increasingly and dynamic cross-fertilization of regulations and technology. Soft law is capable of providing important safeguards on issues like transparency and accountability, while, due to its flexibility, can be quickly adapted to the rapidly evolving technological artifacts, thereby ensuring alignment of the current legislative system, which is developing at a much slower pace.

"Soft law" instrument, in a broader sense, relevant in this context includes the Big Data Value Associations' Position Papers, such as "Towards a European Data Sharing Space. Enabling data exchange and unlocking AI potential," published on April 2019; "Data Protection in the era of Artificial Intelligence. Trends, existing solutions and recommendations for privacy-preserving technologies," published on October 2019; and others.

Likewise, the following two works elaborated by expert groups appointed by the European Commission need to be taken into account:

– The Ethics Guidelines for Trustworthy Artificial Intelligence, presented on 8 April 2019 by the High-Level Expert Group on AI;
– The Report on Liability for Artificial Intelligence and other emerging technologies, released by the European Commission's Expert Group on Liability and New Technologies—New Technologies Formation ("NTF").

23.3.3 Ethics-driven Standardization Initiatives [9]

The main ethics-related standardization projects and initiatives to be considered for materializing EU human factors, ethical principles, and values prioritizing human well-being in the next generation of critical infrastructures' integrated security systems relying on CPS, IoT, and other intelligent and autonomous applications and devices (in addition to other standardization initiatives, such as for security) are, besides the Technical Reports elaborated by ISO/IEC JTC 1/SC 42, above all the Standardization projects of the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, whose mission is "to ensure every stakeholder involved in the design and development of autonomous and intelligent systems is educated, trained, and empowered to prioritize ethical considerations so that these technologies are advanced for the benefit of humanity."[1]

By following the indications of these projects, developers and operators will be guided to create and use the cutting-edge solutions in a way explicitly honoring the inalienable human rights and the beneficial values of their users, thereby maximizing the increase of human well-being as a key metric for progress and social sustainability.

[1] https://standards.ieee.org/industry-connections/ec/autonomous-systems.html

Among the most relevant, there are:

- IEEE P7000^{TM}—Model Process for Addressing Ethical Concerns During System Design
- IEEE P7001^{TM}—Transparency of Autonomous Systems
- IEEE P7002^{TM}—Data Privacy Process
- IEEE P7003^{TM}—Algorithmic Bias Considerations
- IEEE P7006^{TM}—Standard on Personal Data AI Agent
- IEEE P7007^{TM}—Ontological Standard for Ethically driven Robotics and Automation Systems
- IEEE P7008^{TM}—Standard for Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems
- IEEE P7009^{TM}—Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems
- IEEE P7010^{TM}—Wellbeing Metrics Standard for Ethical Artificial Intelligence and Autonomous Systems
- IEEE P7012^{TM}—Standard for Machine Readable Personal Privacy Terms

23.4 Ethical and Legal Framework and Safeguards

There is the need for a coherent legal and ethical frame to delineate the limits of legal and ethical compliant behaviors and to provide responses for tackling the risks of integrated security infrastructures and surrounding technological developments for the protection of critical infrastructures, rooted in common human values and multi-stakeholder involvement, and relying on inclusiveness, adaptivity, agility and fitness for purpose, and thereby functional to the achievement of the sustainable development goals, as recently emphasized by the United Nations Secretary-General's High-level Panel on Digital Cooperation.

Such a framework should take the form of a code of conduct for researchers/designers and users and should be based on the principles enshrined in the EU Charter of Fundamental Rights (such as human dignity, autonomy and human rights, non-discrimination and non-stigmatization, the integration of persons with disabilities and of elderly people and other) and on existing ethical practices and codes.

The values enshrined in the EU Charter of Fundamental Rights represent the normative framework on which a common understanding of the ethical risks associated with the operation of robots could be built. Still, judgments about the ethical soundness of robotics applications depend significantly on the specific context of application and the findings of the respective risk assessment process.

The production of realistic and workable codes of conduct for each domain provides a number of advantages, because they, as a typical "soft law" tool, are capable of offering practical guidance and tackling in meaningful, flexible, and practical ways the issues and ethical challenges of CPS, IoT, and AI breakthroughs in each of such domain. They are also aligned with the legislative support for the self-regulation and accountability instruments (e.g., General Data Protection Regulation (GDPR) Regulation (EU) 2016/679, art. 40). An example of this codes, though referring only to Data Protection, is the Cloud Security Alliance Code of Conduct for GDPR Compliance.

The production of these codes and, in general, the ethics reasoning in relation to integrated security infrastructures for the protection of critical infrastructures need to be based on a prioritization approach and resulting balancing operations, in order to let fully reaping the benefits coming from technological progress in conjunction with the safeguard of human rights and ethical values [5].

In line with the European Group on Ethics (EGE's) in Science and New Technology, aiming at identifying criteria of accountability and oversight in order to protect the ethical values and the freedom of individuals together with security, without giving up on any of the rights and interests at stake, there is the need to find a compromise and a balance between, on the one hand, the interest in strengthening the public safety and in protecting critical infrastructures and the set of related human rights, and, on the other hand, the need to safeguard other human rights, such as privacy, data protection, freedom of expression, freedom of association, freedom of movement, due process and non-discrimination [13].

This balance between opposite interests is particularly relevant in case of critical infrastructure: their protection against physical and cyberattack has a growing role also in national security issues. Attacks on one of them are able to produce huge consequences, in terms of damage economies, cause disasters and other possible serious impacts on health, safety, security, or economic well-being of citizens or even preventing the effective functioning of governments in the Member States.[2]

A rich jurisprudence and a long history of scholarship both in ethical and in legal philosophy confirm the balancing and prioritizing of rights. For instance, in relation to the fundamental right to the protection of personal data under Article 8 of the Charter of Fundamental Rights of the European Union, the Court of Justice of the European Union (CJEU) stated that "is not, however, an absolute right, but must be considered in relation to its function in society."[3]

[2] "Critical infrastructure protection in the fight against terrorism" – COM (2004) 702.

[3] Among others, CJEU, Joined cases C-92/09 and C-93/09, *Volker and Markus Schecke GbR and Hartmut Eifert v. Land Hessen*, November 2010, par. 48.

In addition to the elaboration of Code of Conducts for a more agile and adaptive governance, as well as to the prioritization approach in ethical reasoning and operations, other driving factors and countermeasures for minimizing ethical risks and ensuring legal compliance include:

- Ethical and Legal oversight, for instance through the appointment of Ethics Boards or Ethics and Safety Officers
- Accountability Mechanisms, such as due diligence and certification of ethical and legal compliance, relying also on ad hoc metrics and on Fundamental Rights Impact Assessment and mechanisms afterwards to allow for feedback on any potential infringement of such rights. This impact assessment should be additional to the Data Protection Impact Assessment (DPIA) regulated by the GDPR and should be developed according to the set of indications adopted on it by the EC.
- Privacy and Security by design and Ethics & Rule of law by design
- Adequate redress mechanisms in place in case of damages caused by products and services
- Regulatory sandboxes, where innovative services and tools can be tested and experimented in real regulatory conditions (but with possible exceptions from some rules) in a gradual and controlled way before going to the market, pursuant to a specific testing plan agreed and monitored by the competent authority
- Exploiting and further advancing the cross-fertilization of law and technology, such as in terms of solutions aiming to translate and automate legal provisions into computer language, and then allow some form of human control or intervention to slightly modify the parameters in the computer language translation of legal requirements of compliance: Privacy-enhancing Technologies (PETs), sticky policies, dynamic user consent, blockchain-enabled transactions and smart contracts move in this direction, as well as in general, certain forms of automated compliance tools.

23.5 Conclusion

Beyond the identification of the main areas of potential legal and ethical concern and the associated challenges and the respective relevant pieces of EU legislation that might need to be reviewed or considered, the analysis leads to these conclusions:

- Every attempt to conceive and tackle with the legal and ethical challenges associated with the multifaceted emerging technologies concerned needs to be set on each specific sector and, sometimes, fine-tuned on a case-by-case basis, reflecting on how much security it is reasonable to expect or claim in any given domain, and what is seen as responsible behavior.

- The ethical and legal framework would not need to take a legally binding form but, preferably, should take the form of a domain-specific code of conduct, to be prepared through an holistic, prioritization approach supported by a multidisciplinary exercises, able capture and shape a pluralist conception of law, ethics, and technology.
- The code of conducts and the other "soft law" instruments are expected to be capable of providing granular and practical guidance to all the relevant concerns previously identified, which include privacy and data protection issues, but also data ownership, certification, safety, liability, and much other.
- By reflecting on such issues and conceiving and implementing adequate safeguards and mitigating measures, such as the appointment of ethical officers, the fundamental rights impact assessment, the alignment with ethics-related standardization outcomes and EC-promoted guidelines, and, overall, by ensuring legal compliance and upholding ethical values into the new technological developments at stake, the positive benefits can be reached while mitigating and the negative side effects potentially eliminated, thereby fostering societal acceptance.
- CPS, IoT, AI, ubiquitous data streams, integrated and holistic security infrastructures are neutral even though they can give rise to a more complex world in which human beings will need to improve their ability to predict and understand the machines and their risks and effect on well-being and human rights. Regulatory sandboxes, as safe and controlled environment where innovative services and tools can be tested and experimented in real regulatory conditions before going to the market, are useful tool empowering a better understanding of legal and ethical implications of new technological developments.
- Promising avenues come from the expected advancing of the cross-fertilization of law and technology, including solutions aiming to translate and automate legal provisions into computer language: PETs, sticky policies, dynamic user consent, blockchain-enabled solutions and automated compliance services and techniques.

References

[1] High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI, 2019.

[2] BDVA – EURobotics, Joint Vision Paper for an Artificial Intelligence Public Private Partnership (AI PPP), 2019.

[3] BDVA, Data Protection in the era of Artificial Intelligence. Trends, existing solutions and recommendations for privacy-preserving technologies

[4] Data Ethics Commission of the Federal Government, Opinion of the Data Ethics Commission, 2019.

[5] European Group on Ethics in Science and New Technologies, Ethics of security and surveillance technologies. Opinion. 28, 2014.

[6] ENISA, Privacy and Data Protection by Design, 2014.

[7] European Parliamentary Research Service, Ethical Aspects of Cyber-Physical Systems, 2016.

[8] Max Planck Institute for Innovation and Competition, Data Ownership and access to data. Position statement. Research Paper No. 16-10, 2016.

[9] Institute of Electrical and Electronics Engineers (IEEE), Ethically Aligned Design, 2018.

[10] A. Mashhadi, F. Kawsar and U. G. Acer, "Human Data Interaction in IoT: The ownership aspect," 2014 IEEE World Forum on Internet of Things (WF-IoT), Seoul, 2014, pp. 159–162, doi: 10.1109/WF-IoT.2014.6803139.

[11] Expert Group on Liability and New Technologies – New Technologies Formation, Liability for Artificial Intelligence and other emerging technologies, Report, 2019.

[12] European Commission, Liability for emerging digital technologies, Accompanying the document, 2018.

[13] Council of Europe, European Commission for Democracy through Law (Venice Commission), *Opinion on Video Surveillance in Public Spaces by Public Authorities and the Protection of Human Rights*, 70th Plenary Session, 16–17 March 2007.

Epilogue

The world has recently witnessed large-scale security incidents such as the WannaCry ransomware attack, the Mirai botnet, and various notorious attacks against industrial plants. These incidents indicate that despite the ever-increasing investments in security solutions, industrial organizations and critical infrastructures remain vulnerable against adversarial attacks. Several of the proclaimed vulnerabilities of critical infrastructures stem from their complexity and their cyber-physical nature. Modern critical infrastructures comprise both cyber and physical assets and, as such, can be considered as large-scale cyber-physical systems. Hence, the conventional approach of addressing cybersecurity and physical security separately is no longer effective. On the contrary, more integrated approaches that address the security of cyber and physical assets at the same time are required. Even though the merit of such integrated approaches is acknowledged, their implementation is in its infancy.

This book has presented integrated (i.e., cyber and physical) security approaches and technologies for some of the most important infrastructures that underpin our societies. Specifically, it has presented advanced techniques for threat detection, risk assessment, and security information sharing, based on leading edge technologies like machine learning, security knowledge modeling, IoT security, and distributed ledger infrastructures. Likewise, it has introduced how established security technologies like SIEM, pen-testing, vulnerability assessment, and security data analytics can be used in the context of integrated Critical Infrastructure Protection. Moreover, certain chapters of the book have dealt with the ever important operational, business, and ethical aspects of critical infrastructure protection.

The novel methods and techniques of the book are exemplified in the scope of case studies involving pragmatic critical infrastructures in four industrial sectors, namely finance, healthcare, energy, and communications. In this way, the peculiarities of critical infrastructure protection in each one of these sectors have been adequately discussed and addressed based on sector-specific solutions. The presentation of security systems and technologies for these four different sectors provides opportunities for understanding the commonalities and the differences of security systems in the various sectors.

The advent of the fourth industrial revolution (Industry 4.0) is expected to increase the cyber-physical nature of critical infrastructures as well as their interconnection in the scope of sectorial and cross-sector value chains. Therefore, the demand for solutions that foster the interplay between cyber and physical security and enable Cyber-physical Threat Intelligence is likely to explode. In this book, we have shed light on the structure of such integrated security systems, as well as on the technologies that will underpin their operation. We hope that Security and Critical Infrastructure Protection stakeholders will find the book useful in planning their future security strategies.

Index

About the Editors

John Soldatos (http://gr.linkedin.com/in/johnsoldatos) holds a PhD in Electrical & Computer Engineering from the National Technical University of Athens (2000) and is currently Honorary Research Fellow at the University of Glasgow, UK (2014-present). He was Associate Professor and Head of the Internet of Things (IoT) Group at the Athens Information Technology (AIT), Greece (2006–2019), and Adjunct Professor at the Carnegie Mellon University, Pittsburgh, PA (2007–2010). He has significant experience in working closely with large multi-national industries (IBM Hellas, INTRACOM S.A, INTRASOFT International) as R&D consultant and delivery specialist, while being scientific advisor to high-tech startup enterprises, such as Innov-Acts Limited (Nicosia, Cyprus) and Innovation Sprint Sprl (Brussels, Belgium). Dr. Soldatos is an expert in Internet-of-Things (IoT) technologies and applications, including IoT's applications in smart cities, finance (Finance 4.0), and industry (Industry 4.0). Dr. Soldatos has played a leading role in the successful delivery of more than sixty (commercial-industrial, research, and consulting) projects, for both private & public sector organizations, including complex integrated projects. He is co-founder of the open source platform OpenIoT (https://github.com/OpenIotOrg/openiot) and of the Edge4Industry (www.edge4industry.eu) community. He has published more than 180 articles in international journals, books, and conference proceedings. He has also significant academic teaching experience, along with experience in executive education and corporate training. Dr. Soldatos is regular contributor in various international magazines and blogs, on topics related to IoT, Industry 4.0, and cybersecurity. Moreover, he has received national and international recognition through appointments in standardization working groups, expert groups, and various boards. He has co-edited and co-authored three edited volumes (books) on Internet of Things topics, including IoT for Industrial Automation, IoT Analytics, and IoT Security.

James Philpot graduated from Bath University in with a Bachelor's Degree in Politics and International Relations in 2015. During his studies, he spent a year working abroad for an NGO confederation in Brussels. After his studies, he began working for a consultancy, focusing on project management support. He was subcontracted to a 15-month cultural heritage project and provided communications support for various other projects and clients. After the conclusion of this project, he moved on to work for a cultural heritage network, helping implement a Creative Europe network support grant. He joined EOS (European Organization for Security) in the spring of 2019 and has been working on European projects in crisis management and security for health services, as well as coordinating the working groups on cybersecurity and security screening and detection technologies.

Gabriele Giunta holds a degree in Computer Science, at the Engineering Department of the University of Palermo (Italy). He has been working in the Engineering Ingegneria Informatica S.p.A. R&D Lab since 2000 in multi-sector Italian and European research projects, playing several roles, such as solution designer, system architect, technical coordinator, and project manager. Now, he is the research manager and head of the "Smart Transport and Infrastructure" Unit within the IS3 R&D Lab ("Intelligent Systems and Social Software for Security, Enterprises, Transport, Infrastructure") at ENGINEERING (http://www.eng.it) dealing with Critical Infrastructure Protection, Smart Mobility and Logistic solutions, mostly focused on the systemic view. In the past years, he was involved in several Italian Ministry and EC co-funded research projects, such as MAIS, DISCoRSO and NEXOF-RA and he coordinated Easy Rider, SECURE! and STORM. He is currently involved in DEFENDER (http://defender-project.eu) as Project Coordinator, in FASTER (https://www.faster-project.eu/) as Technical Coordinator, in InfraStress (https://www.infrastress.eu/) as WP Leader. My main areas of interest and relevant expertise include Software Engineering, Software Architecture Design, System Integration, Complex Event Processing, Data Mining, Information Fusion, Human Computer Interaction, Knowledge Modeling and Representation, Business Process Modeling and Management, Service Oriented Computing. He has been involved as co-author of scientific papers in International Conferences, Journals and Books.

Contributing Authors

Habtamu Abie
Norwegian Computing Center, P.O. Box 114 Blindern, NO-0314 Oslo, Norway
habtamu.abie@nr.no

Maurizio Aiello
Consiglio Nazionale delle Ricerche (CNR), IEIIT institute, Via De Marini, 6, 16149 Genoa Italy
maurizio.aiello@ieiit.cnr.it

Silvia Andernello
CSI Piemonte, Corso Unione Sovietica, 216, 10134 Torino, Italy
silvia.andernello@csi.it

Faten Atigui
CEDRIC Laboratory, Conservatoire National des Arts et Métiers (CNAM), Paris, France

Luca Baldini
Ericsson, via Anagnina 203, 00118 Roma, Italy
luca.baldini@ericsson.com

Federica Battisti
Department of Engineering, Universitá degli Studi Roma Tre
federica.battisti@uniroma3.it

Maria Belesioti
Fixed Network R&D Programs Section – Hellenic Telecommunications Organization S.A. (OTE)
mbelesioti@oteresearch.gr

Edward-Benedict Brodie of Brodie
THALES Research & Technology, France
edward-benedict.brodieofbrodie@thalesgroup.com

Fabrizio Bertone
LINKS Foundation, Via P.C. Boggio, 61, 10138 Torino, Italy
fabrizio.bertone@linksfoundation.com

Elisabetta Biasin
KU Leuven Centre for IT & IP Law
Sint-Michielsstraat 6 box 3443, 3000 Leuven, BELGIUM
Elisabetta.Biasin@kuleuven.be

Andrea Bisegna
Fondazione Bruno Kessler
Via Sommarive 18
38123 Trento, Italy
a.bisegna@fbk.eu

Svetlana Boudko
Norwegian Computing Center,
P.O. Box 114 Blindern, NO-0314
Oslo, Norway
Svetlana.Boudko@nr.no

Guilherme Brito
NOVA School of Science and
Technology (FCT NOVA)/
UNINOVA Centre of Technology and
Systems (CTS), Caparica, Portugal
guilherme.brito@uninova.pt

Cédric Buron
THALES Research & Technology,
France
cedric.buron@thalesgroup.com

Denis Čaleta
Institute for Corporative Security
Studies
denis.caleta@ics-institut.si

Enrico Cambiaso
Consiglio Nazionale delle Ricerche
(CNR), IEIIT institute, Via De
Marini, 6, 16149 Genoa Italy
enrico.cambiaso@ieiit.cnr.it

Serban Carata
UTI Group, Romania

University "Politehnica" Bucharest,
Romania

Roberto Carbone
Fondazione Bruno Kessler
Via Sommarive 18
38123 Trento, Italy
carbone@fbk.eu

Marco Carli
Engineering Department, Universitá
degli Studi Roma TRE, Roma, Italy
marco.carli@uniroma3.it

Mariano Ceccato
Università di Verona
Strada le Grazie 15
37134 Verona, Italy
mariano.ceccato@univr.it

Giuseppe Celozzi
Ericsson, via Anagnina 203, 00118
Roma, Italy
giuseppe.celozzi@ericsson.com

Ioannis Chochliouros
Fixed Network R&D Programs
Section – Hellenic
Telecommunications Organization
S.A. (OTE)
ichochliouros@oteresearch.gr

Federico Colangelo
Engineering Department, Università
degli Studi Roma TRE, Roma, Italy
federico.colangelo@uniroma3.it

Ioan Constantin
Orange Romania
ioan.constantin@orange.com

Antonello Corsi
Engineering Ingengeria Informatica
S.p.A.
antonello.corsi@eng.it

Marina Da Bormida
R&I Lawyer and Ethics Expert Via
Assarotti 8/3A, 16121 Genoa, Italy
m.dabormida@eurolawyer.it

Ilesh Dattani
Assentian Partners, London, U.K.
ilesh.dattani@assentian.com

Samantha Dauguet-Demailly
AIRBUS CyberSecurity, Élancourt,
France
samantha.dauguetdemailly@airbus.com

Corrado De Santis
Studio Tecnico BFP S.r.l.
casanova@studiobfp.com

Giovanni Di Orio
NOVA School of Science and Technology (FCT NOVA)/ UNINOVA Centre of Technology and Systems (CTS), Caparica, Portugal
gido@uninova.pt

Alberto Dognini
RWTH Aachen University, Institute for Automation of Complex Power Systems (ACS), Aachen, Germany
adognini@eonerc.rwth-aachen.de

George Doukas
National Technical University of Athens, 9, Iroon Polytechniou Str., Zografou Campus,15780 Greece
gdoukas@epu.ntua.gr

Dimitris Drakoulis
INNOV-ACTS LIMITED, 27, Michalakopooulou, Str, Nicosia, Cyprus

Nikos Drosos
SINGULARLOGIC, Greece
ndrosos@singularlogic.eu

Katja Faist
Fraunhofer Institute for High-Speed Dynamics, Germany

Mirjam Fehling-Kaschek
Fraunhofer Institute for High-Speed Dynamics, Safety Technology and Protective Structures, Germany
mirjam.fehling-kaschek
@emi.fraunhofer.de

Maurizio Ferraris
European Projects, GFT Italia Srl, Genoa, Italy
maurizio.ferraris@gft.com

Sabin Floares
UTI Group, Romania

Chiara Foglietta
Engineering Department, University of "Roma Tre"
chiara.foglietta@uniroma3.it

Simon Fossier
THALES Research & Technology, France
simon.fossier@thalesgroup.com

Dušan Gabrijelčič
Jožef Stefan Institute, Slovenia
dusan@e5.ijs.si

Beatriz Gallego-Nicasio
Atos Spain, Calle de Albarracín, 25, 28037 Madrid, Spain
beatriz.gallego-nicasio@atos.net

Teni Gasparini
ELES d.d.
teni.gasparini@eles.si

Marco Gavelli
LINKS Foundation, Via P.C. Boggio, 61, 10138 Torino, Italy
marco.gavelli@linksfoundation.com

Giorgia Gazzarata
DIBRIS, University of Genoa, Genoa, Italy, and CINI, Rome, Italy
giorgia.gazzarata@dibris.unige.it

Lev Greenberg
IBM Israel – Science and Technology LTD, Derekh Em Hamoshavot 94, Petah Tikva, 4970602 Israel
LEVG@il.ibm.com

Marian Ghenescu
UTI Group, Romania
Institute for Space Science, Romania

Ilias Gkotsis
Center for Security Studies (KEMEA)
P. Kanellopoulou 4, 101 77 Athens, Greece
i.gkotsis@kemea-research.gr

Bárbara Guerra
EDGENEERING Lda, R. Abranches Ferrão, 10 11 C, 1600-001 Lisbon, Portugal
barbara@edgeneering.eu

Gael Haab
Fraunhofer Institute for High-Speed Dynamics, Germany

Fayçal Hamdi
CEDRIC Laboratory, Conservatoire National des Arts et Métiers (CNAM), Paris, France

Ivo Häring
Fraunhofer Institute for High-Speed Dynamics, Germany

Hussain Ijaz
Beia Consult International
ijaz@beia.ro

Erik Kamenjašević
KU Leuven Centre for IT & IP Law
Sint-Michielsstraat 6 box 3443, 3000 Leuven, Belgium
Erik.kamenjasevic@kuleuven.be

Ioannis Karagiannis
INNOV-ACTS LIMITED, 27, Michalakopooulou, Str, Nicosia, Cyprus

Panos Karaivazoglou
Institute of Communications and Computer Systems of the National Technical University of Athens, 9, Iroon Polytechniou Str, 15773 Zografou, Athens, Greece (GR)
pkaraiv@esd.ece.ntua.gr

Nikolaos Koutras
ADITESS Advanced Integrated Solutions and Services Ltd, Byzantiou 40, 2064, Nicosia, Cyprus (CY)
management@aditess.com

Alexandros Kyritsis
Institute of Communications and Computer Systems of the National Technical University of Athens, 9, Iroon Polytechniou Str, 15773 Zografou, Athens, Greece (GR)
al.kyritsis@gmail.com

Nadira Lammari
CEDRIC Laboratory, Conservatoire National des Arts et Métiers (CNAM), Paris, France

David Lancelin
AIRBUS CyberSecurity, Elancourt, France
david.lancelin@airbus.com

Francesco Lubrano
LINKS Foundation, Via P.C. Boggio, 61, 10138 Torino, Italy
francesco.lubrano@linksfoundation.com

Eva Maia
Institute of Engineering, Polytechnic of Porto (ISEP/IPP)
GECAD–Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development
egm@isep.ipp.pt

Rodoula Makri
Institute of Communications and Computer Systems of the National Technical University of Athens, 9, Iroon Polytechniou Str, 15773 Zografou, Athens, Greece (GR)
rodia@esd.ece.ntua.gr

Pedro Maló
NOVA School of Science and Technology (FCT NOVA)/ UNINOVA, Caparica, Portugal
pmm@uninova.pt

Alessandro Mamelli
Pointnext Advisory and Professional Services, Hewlett-Packard Italiana Srl, Cernusco s/N, Milan, Italy
alessandro.mamelli@hpe.com

Marco Manso
EDGENEERING Lda, R. Abranches Ferrão, 10 11 C, 1600-001 Lisbon, Portugal
marco@edgeneering.eu

Vasiliki Mantzana
Center for Security Studies (KEMEA)
P. Kanellopoulou 4, 101 77 Athens, Greece
v.mantzana@kemea-research.gr

Konstantinos Mavrogiannis
Singularlogic S.A, 3, Achaias and Trizinias, Str, Nea Kifisia, Greece

Alessio Merlo
DIBRIS, University of Genoa, Genoa, Italy
alessio.merlo@unige.it

Salvatore Manfredi
Fondazione Bruno Kessler, Via Sommarive 18, 38123 Trento, Italy
smanfredi@fbk.eu

Roxana Mihaescu
UTI Group, Romania
University "Politehnica" Bucharest, Romania

Natalie Miller
Fraunhofer Institute for High-Speed Dynamics, Germany

Antonello Monti
RWTH Aachen University, Institute for Automation of Complex Power Systems (ACS), Aachen, Germany
amonti@eonerc.rwth-aachen.de

Vasiliki Moumtzi
VILABS ltd, 6, Vasili Vryonides str. Gala Court Chambers 3095 Cyprus
mova@vilabs.eu

Cosmin-Septimiu Nechifor
SIEMENS, Corporate Technology, Romania
septimiu.nechifor@siemens.com

Alberto Neri
Leonardo SpA

Alessandro Neri
Engineering Department, Universitá degli Studi Roma TRE, Roma, Italy
alessandro.neri@uniroma3.it

Barry Norton
Milestone Systems A/S
bno@milestone.dk

Mathias Normann
Milestone Systems A/S
mjn@milestone.dk

Federica Pascucci
Department of Engineering, Università degli Studi Roma Tre
federica.pascucci@uniroma3.it

Gabriel Petrescu
Beia Consult International
gabriel.petrescu@beia.ro

Paolo Petrucci
ASLTO5 – Azienda Sanitaria Locale TO5 – Chieri (TO) – Italy
petrucci.paolo@aslto5.piemonte.it

Stefano Panzieri
Engineering Department, University of "Roma Tre"
stefano.panzieri@uniroma3.it

Ariana Polyviou
Research and Development, INNOV-ACTS LTD, Department of Management and MIS University of Nicosia, Nicosia, Cyprus
polyviou.a@unic.ac.cy

University of Nicosia, Makedonitissis 46, Nicosia, Cyprus
apolyviou@innov-acts.com

Isabel Praça
Institute of Engineering, Polytechnic of Porto (ISEP/IPP)
GECAD–Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development
icp@isep.ipp.pt

Silvio Ranise
Fondazione Bruno Kessler
Via Sommarive 18
38123 Trento, Italy
ranise@fbk.eu

Andrea Roland
Fraunhofer Institute for High-Speed Dynamics, Germany

Abhinav Sadu
RWTH Aachen University, Institute for Automation of Complex Power Systems (ACS), Aachen, Germany
asadu@eonerc.rwth-aachen.de

Jose Manuel Sanchez
INTEGRASYS S.A. Boabdil 6, Camas, 41900, Sevilla Spain (ES)
jose.sanchez@integrasys-sa.com

Mari-Anais Sachian
Beia Consult International
anais.sachian@beia.ro

Francesca Santori
ASM Terni S.p.A
francesca.santori@asmterni.it

Giada Sciarretta
Fondazione Bruno Kessler
Via Sommarive 18
38123 Trento, Italy
giada.sciarretta@fbk.eu

Aidan Shribman
IBM Israel – Science and Technology LTD, Derekh Em Hamoshavot 94, Petah Tikva, 4970602 Israel
Aidan.Shribman@il.ibm.com

Samira Si-said Cherfi
CEDRIC Laboratory, Conservatoire National des Arts et Métiers (CNAM), Paris, France

Michael Skitsas
ADITESS Advanced Integrated Solutions and Services Ltd, Byzantiou 40, 2064, Nicosia, Cyprus (CY)
mskitsas@aditess.com

John Soldatos
University of Glasgow and
INNOV-ACTS LIMITED,
Glasgow, UK
john.soldatos@glasgow.ac.uk

INNOV-ACTS LTD, Nicosia, Cyprus
jsoldat@innov-acts.com

Omri Soceanu
IBM Israel – Science and Technology
LTD, Derekh Em Hamoshavot 94,
Petah Tikva, 4970602 Israel
Omri.Soceanu@il.ibm.com

Alexander Stolz
Fraunhofer Institute for High-Speed
Dynamics, Germany

George Suciu
Beia Consult International
george@beia.ro

Olivier Terzo
LINKS Foundation, Via P.C. Boggio,
61, 10138 Torino, Italy
olivier.terzo@linksfoundation.com

Alessandro Tomasi
Fondazione Bruno Kessler
Via Sommarive 18
38123 Trento, Italy
altomasi@fbk.eu

Francesco Tresso
CSI Piemonte Corso Unione
Sovietica, 216, 10134 Torino, Italy
francesco.tresso@csi.it

Ernesto Troiano
European Projects, GFT Italia Srl,
Genoa, Italy
ernesto.troiano@gft.com

Ivan Vaccari
Consiglio Nazionale delle Ricerche
(CNR), IEIIT institute, Via De
Marini, 6, 16149, Genoa, Italy
ivan.vaccari@ieiit.cnr.it

Javier Valera
INTEGRASYS S.A. Boabdil 6,
Camas, 41900, Sevilla
Spain (ES)
javier.valera@integrasys-sa.com

Luca Verderame
DIBRIS, University of Genoa, Genoa,
Italy
luca.verderame@unige.it

Luca Viarengo
CSI Piemonte Corso Unione
Sovietica, 216, 10134 Torino, Italy
luca.viarengo@csi.it

Emanuele Viglianisi
Runtastic GmbH Pluskaufstraße 7
4061 Pasching bei Linz, Austria
emavgl@gmail.com

Artemis Voulkidis
Power Operations Limited, United
Kingdom
artemis@power-ops.com

Xiao-Si Wang
Future Security & Cyber Defence,
Applied Research, British
Telecommunications plc
selina.wang@bt.com

Nikolaus Wirtz
RWTH Aachen University,
Institute for Automation of Complex
Power Systems (ACS),
Aachen, Germany
nwirtz@eonerc.rwth-aachen.de

Theodore Zahariadis
National and Kapoditrian University
of Athens
zahariad@uoa.gr

Cosimo Zotti
Ericsson, via Anagnina 203, 00118
Roma, Italy
cosimo.zotti@ericsson.com