Second Edition

Handbook of SCADA/ Control Systems Security

Edited by

Robert Radvanovsky
Jacob Brodsky

CRC Press
Taylor & Francis Group

Second Edition

Handbook of SCADA/ Control Systems Security

Second Edition

Handbook of SCADA/ Control Systems Security

Edited by

Robert Radvanovsky
Jacob Brodsky

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20160323

International Standard Book Number-13: 978-1-4987-1708-3 (eBook - PDF)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

In Memoriam

Dr. Wayne Boone

and

Dr. Harold Brodsky

Dedication

This book is dedicated to our families and friends who have been supportive in the development of this book. Their patience and understanding of our efforts are an author's best backing. Our thanks go to them during this time period …

By failing to prepare, you are preparing to fail.

Benjamin Franklin

Know the enemy and know yourself; in a hundred battles you will never be in peril. When you are ignorant of the enemy, but know yourself, your chances of winning or losing are equal. If ignorant both of your enemy and yourself, you are certain in every battle to be in peril.

Sun Tzu
The Art of War

Contents

Section V: Conclusion

Foreword

Klaatu barada nikto

Increasingly, the services we rely on in our daily life, such as water treatment, electricity generation and transmission, health care, transportation, and financial transactions, depend on an underlying information technology and communications infrastructure. Cyberthreats put the availability and security of these services at risk.

Something wicked this way …

The world faces a combination of known and unknown system vulnerabilities, a strong and rapidly expanding adversarial capability, and a lack of comprehensive threat and vulnerability awareness. Within this dynamic environment, both governments and private-sector companies are confronted with threats that are more targeted, more sophisticated, and more serious.

Sensitive information, including classified government data and proprietary data from private companies, is routinely stolen. This undermines our confidence in information systems security and the ability to protect our privacy. As bad as the loss of this intellectual capital is, we increasingly face even greater threats that could significantly compromise the accessibility and reliability of our critical infrastructure.

Malicious actors in cyberspace, including nation-states, terrorist networks, and organized criminal groups, are capable of targeting elements of the U.S. critical infrastructure to disrupt or destroy systems on which we depend. Stated motives include intelligence collection; theft of intellectual property, personal identity, or financial data; disruption of commercial activities; and cyberterrorism. Criminal elements continue to show increasing levels of sophistication in their technical and targeting capabilities and have shown a willingness to sell these capabilities on the underground market. In addition, terrorist groups and their sympathizers have expressed interest in using cyberspace to target and harm the United States and its citizens. Although terrorist groups and their sympathizers may lack their own purpose, tools and techniques are readily available for purchase through black markets. This generates a very real threat to the stability and resilience of our critical control systems.

Malicious cyberactivity can instantaneously result in virtual or physical consequences that threaten national and economic security, critical infrastructure, and public health and welfare. Similarly, stealthy intruders have laid a hidden foundation for future exploitation or attack, which they can then execute at their leisure—and at a time of great advantage to their cause. Securing cyberspace requires a layered security approach across the public and private sectors. The current reliance on perimeter defense as a single solution provides a false sense of security. Similar to the Maginot line, this approach is predicated on

predictable actions on the part of our adversaries. Once the attacker figures how to drive to Belgium and the Ardennes, it is too late for the system. The landscape requires a fresh approach to defense in depth along with an active defense posture and capability.

Darmok, and jalad … at tanagra

By investing in both public- and private-sector ventures, the government and industry can establish centers that serve as "always-on facilities" for cyberincident response and management. This enables the centers to provide "actionable intelligence" for asset owners, operators, and government agencies.

President Obama's *Cyberspace Policy Review* called for "a comprehensive framework to facilitate coordinated responses by government, the private sector, and allies to a significant cyber incident." With the federal government and private industry working together to develop joint incident response capabilities, these goals may be achieved. The approach requires vigilance and a voluntary public/private partnership in order to build the capability and relationships necessary to combat the growing cyberthreat.

In addition to identifying threats and vulnerabilities, specific work must be conducted by asset owners and operators with the assistance of the vendor community to develop mitigation plans to enhance security. This includes the need to evaluate the interdependencies across critical infrastructure sectors. For example, the electric, nuclear, water, transportation, and communications sectors support functions across all levels of government and the private sector. Government bodies and organizations do not inherently produce these services and must rely on private-sector organizations, just as other businesses and private citizens do. Therefore, an event impacting control systems has potential implications at all levels and could also have cascading effects on our ability to conduct commerce or generate life-giving services.

Assessing risk and effectively securing industrial control systems are vital to maintaining our nation's strategic interests, public safety, and economic well-being. A successful cyberattack on a control system could result in physical damage, loss of life, and cascading effects that could disrupt services for a prolonged period of time. We all must recognize that the protection and security of control systems are essential to the nation's overarching security and economy. A real-world threat has already emerged that significantly changed the landscape of targeted cyberattacks on industrial control systems. Malicious code, dubbed Stuxnet, was detected in July 2010. Analysis concluded that this highly complex code was the first of its kind, written to specifically target mission-critical control systems running a specific combination of software and hardware. The analysis quickly uncovered that sophisticated malware of this type has the ability to gain access to secure systems, steal detailed proprietary information, conduct reconnaissance, and manipulate the systems that operate mission-critical processes within the nation's infrastructure. In other words, this code can automatically enter a system, steal the formula for the product being manufactured, alter the ingredients being mixed in the product, and indicate to the operator and the operator's defenses that everything is functioning normally. Looking ahead, there is a deep concern that attackers could use the information about the code to develop variants targeted at broader installations of programmable equipment in control systems.

Lacking a silver bullet

Overcoming new cybersecurity challenges requires a coordinated and focused approach to better secure the nation's information and communications infrastructures. No single

government agency has sole responsibility for securing cyberspace, and the success of our cybersecurity mission relies on effective communication and critical partnerships. Private industry owns and operates the vast majority of the nation's critical infrastructure and cybernetworks; therefore, the private sector plays an important role in cybersecurity.

Set within an environment characterized by a dangerous combination of known and unknown vulnerabilities, strong and rapidly expanding adversary capabilities, and a lack of comprehensive threat and vulnerability awareness, the cybersecurity mission is truly a national one requiring broad collaboration. Cybersecurity is critical to ensure that the government, businesses, and the public can continue to use the information technology and communications infrastructure on which they depend. We must continue to engage and collaborate in order to provide analysis, vulnerability, and mitigation assistance across the broad spectrum of industrial control systems. We must work closely with the international community in order to mitigate the risk on a global scale.

Seán McGurk
Global Manager, Intel Security

Synopses of chapters

This book is divided into five sections. The first four each consist of several chapters that represent groupings of topics, which emphasize those topics comprising functions within and throughout ICS environments; the fifth consists of conclusions.

These topics are categorically subdivided into unique and prioritized levels, beginning with Section I and its subsequent chapters, building up to Section II, and so on. Each subsequent section emphasizes a different meaning that is being conveyed such that it can be structured and remembered in an easy, cognitive fashion. A listing of each section and its corresponding chapters (with a brief summary of its description and function) is provided below.

Section I: Social implications and impacts

Chapter 1: Introduction

This chapter provides the basis for the entire book and describes some of the historical backgrounds of industrial control systems (ICS) and why it is important to critical infrastructures worldwide. There are some terms and definitions covering a brief synopsis of the intent of this book and what is to be expected from professionals who are emerging within the ICS security community.

Chapter 2: Sociological and cultural aspects

This chapter is more theoretical than most in that it identifies both background and emerging trends in the direction of the ICS security community. Some of the issues that continue to plague the ICS security community are the differences between the engineering and IT communities and the lack of proper coordination and communication between the two groups. This chapter reflects this current trend, along with other factors involving the paradigm shift from engineering to IT within the ICS security community.

Chapter 3: Threat vectors

This chapter outlines threat factors, both internal and external, to a given automated operation. Some of the factors include identifying motivational aspects and why an adversary would attempt to disrupt and perhaps even destroy a given automated operation.

Chapter 4: Risk management

This chapter applies both common and not-so-common risk methodologies and principles that can be applied to safeguard and secure an automated operation. The aim of this chapter is to provide a fundamental understanding of what risk is within the plant and how disruption can potentially cause near or completely catastrophic events to occur.

Chapter 5: International implications of securing our SCADA/control systems environments

This chapter provides an international perspective and implies that cybersecurity is non-border-specific; that is, the author of this chapter attempts to provide a representative picture of how events and incidents are related to one another for all critical infrastructures—worldwide.

Chapter 6: Aurora generator test

This chapter outlines the concepts surrounding the implications in terms of the types of physical damage and consequences that could result from a potential cyberattack. Additionally, this chapter provides a fundamental understanding of any engineering risks associated with the actual test and how it may be tied to cybersecurity.

Section II: Governance and management

Chapter 7: Disaster recovery and business continuity of SCADA

This chapter discusses methods for restoring and mitigating issues involving a *cyberincident*. Essentially, this chapter answers "what if" questions by providing a roadmap to the management of recovering automated operations to the state before the cyberincident occurs. The other half provides the "how" questions, discussing what would keep the automated operations going.

Chapter 8: Incident response and SCADA

This chapter outlines what steps should be performed as a result of a cyberincident; how management within the organization is informed; if regulated, how communications should be made to the regulating organization; and so on.

Chapter 9: Forensics management

This chapter identifies methods of determination of the events leading to a cyberincident; this includes best practices that should be applicable within any given automated operation and how this can assist the asset owner in deterministic analysis.

Chapter 10: Governance and compliance

This chapter outlines the importance and reasoning behind implementing a governance or compliance program and how it impacts SCADA and control systems environments. More critical infrastructure organizations are having regulatory requirements or guidelines

imposed on them that limit or dictate the course of operation. This chapter will outline the challenges and issues (and perhaps solutions) encountered within those operation environments.

Chapter 11: Project management for SCADA systems

This chapter identifies and focuses on SCADA and control systems implementations and the challenges often associated with them. Unlike traditional projects, SCADA and control systems' projects are uniquely different, requiring more precision and cultural understanding of the expectations of a project manager.

Section III: Architecture and modeling

Chapter 12: Communications and engineering systems

This chapter outlines the necessity for good communications within and throughout the control systems environments, while at the same time outlining fundamental engineering concepts and reasons for those environments, as well as general impacts and interactions with business and IT systems' environments.

Chapter 13: Metrics framework for a SCADA system

This chapter provides a strategic "roadmap" for the development of a secured SCADA/control systems environment and what it entails.

Chapter 14: Network topology and implementation

This chapter provides some generic, non-industry-specific examples of how an ICS network is defined and configured. Examples are not specific to any hardware manufacturer and represent general rather than specific functions that encompass an ICS network. The chapter also provides more specific functionalities involved within an ICS network, identifies key component systems that are required to secure an ICS network, and discusses why these systems are important.

Chapter 15: Active defense in industrial control system networks

This chapter defines the concept of an active role: taking the defenders' greatest strength—their personnel—and empowering them to break down barriers of communication and technology to identify, respond to, and learn from potential adversaries. This provides a strategic approach to security.

Chapter 16: Open-source intelligence (OSINT)

This chapter broaches the topic of intelligence gathered not from closed or private sources, which can cost significant amount of time and effort, but through publicly available sources. These sources provide rapid performance and vulnerability assessments of potential attackers, giving a critical edge to both private- and public-sector current and future operations.

Section IV: Commissioning and operations

Chapter 17: Obsolescence and procurement of industrial control systems

This chapter identifies current issues with ICS environments and some of the issues that arise when ICS equipment is not sufficiently maintained and kept up to date.

Chapter 18: Patching and change management

This chapter follows the obsolescence chapter and discusses why it is important to patch ICS equipment. Many of the issues that most public utilities are currently facing today involve either obsolescence issues or, more specifically, the lack of patching of key and critical systems to plant operations. Recent malware outbreaks, such as what occurred with Stuxnet, have caused many ICS security professionals to reevaluate patching methodologies within their plant operations.

Chapter 19: Physical security management

Just because ICS equipment is located within a plant or secured facility, it does not mean that there are no insider threats. This chapter provides an insight into the physical localities of ICS equipment and discusses physical security as an integral part of the holistic management of a plant.

Chapter 20: Tabletop/red–blue exercises

This chapter discusses one of the aspects of how to conduct training exercises for SCADA/control systems and provides as close to "real-life" scenarios as possible. For a tabletop exercise, the chapter outlines what is involved and how and what to set up and configure for this type of exercise. For the red–blue exercise, it describes a current program offered through the U.S. Department of Homeland Security to owner/operators of SCADA/control systems by giving students a simulated example through the disruption of real systems without any consequence for or impact on real critical infrastructures.

Chapter 21: Integrity monitoring

This chapter outlines the data that are relied upon for accurate processing and also discusses how objectives such as access rights, the integrity of operations, and data and reporting must be both valid and consistent.

Chapter 22: Data management and records retention

This chapter outlines some of the emerging issues with "data overload," especially the logging requirements that are emerging for many cybersecurity regulations and compliance guidelines today. The issue is what data are important to retain and why organizations need to retain that data.

Section V: Conclusion

Chapter 23: The future of SCADA and control systems security

This chapter provides a "future thought" in terms of one or two possible directions that ICS security can go. The authors and editors identify 5- and 10-year directions and what might be different in the future.

Appendix I: Listing of online resources SCADA/control systems

Appendix I provides a comprehensive listing of known online resources specific to SCADA and control systems security, along with a brief summary of each of their functions and purposes.

Appendix II: Terms and definitions

Appendix II provides terms and definitions used by SCADA and control systems professionals within and throughout this community.

Acknowledgments

Some materials used in this book were taken from several very reliable and useful sources. Any information that may appear to be repetitive in its content from those sources was taken to provide a more introspective perception of what defines SCADA security.

The editors wish to thank the following organizations and individuals for their contributions:

United States Department of Homeland Security's National Cyber Security Division's Control Systems Security Program
United States Department of Homeland Security's Industrial Control Systems Computer Emergency Response Team (ICS-CERT)
United States Department of Homeland Security's Federal Emergency Management Agency (FEMA)
Idaho National Engineering and Environmental Laboratory (INEEL)
Sandia National Laboratories (SNL)
Pacific Northwest National Laboratory (PNNL)
United States Department of Energy's Office of Energy Assurance
United States Department of Energy's National SCADA Test Bed (NSTB)
Government of Canada, Public Safety Canada
Government of Lithuania, Ministry of National Defense
National Institute of Standards and Technology (NIST)
Seán McGurk, Intel Security

Editors

Robert Radvanovsky is an active professional in the United States with knowledge in security, risk management, business continuity, disaster recovery planning, and remediation. He obtained his master's degree in computer science from DePaul University in Chicago, and he has significantly contributed toward establishing several certification programs, specifically on the topics of critical infrastructure protection and critical infrastructure assurance.

Robert has special interest and knowledge in matters of critical infrastructure and has published a number of articles and white papers regarding this topic. Although he has been significantly involved in establishing security training and awareness programs through his company, Infracritical, his extracurricular activities include working for several professional accreditation and educational institutions on the topics of homeland security, critical infrastructure protection and assurance, and cybersecurity. He is the owner of, and one of the lead moderators to, the SCADASEC mailing list for supervisory control and data acquisition (SCADA) and control systems security discussion fora, while working as an active participant with the U.S. Department of Homeland Security Transportation Security Administration's Transportation Systems Sector Cyber Working Group as well as the U.S. Department of Homeland Security Control Systems Security Program's Industrial Control Systems' Joint Working Group. Both of these working groups are part of President Obama's Cyber Security Initiative.

Robert's first book, *Critical Infrastructure: Homeland Security and Emergency Preparedness* (released May 2006), is a reference work dealing with emergency management and preparedness, and it defines (in greater detail) what critical infrastructure protection is. His second book, *Transportation Systems Security* (released May 2008), was designed to educate mid-level management (or higher) about aspects of holistic security analysis and management of the transportation sector. His third book, *Critical Infrastructure: Homeland Security and Emergency Preparedness, Second Edition* (released December 2009), coauthored with Allan McDougall, further evolves and incorporates critical infrastructure assurance as part of the critical infrastructure protection model. His fourth book, *Critical Infrastructure: Homeland Security and Emergency Preparedness, Third Edition* (released April 2013), further evolves and incorporates newer aspects of the critical infrastructure protection model. His fifth book project involved coediting and cowriting a book on SCADA security with Jacob Brodsky, titled *Handbook for SCADA and Control Systems Security, First Edition* (released February 2013), and works cooperatively in maintaining and promoting the SCADA and Control Systems Security (SCADASEC) mailing list.

Jacob Brodsky has been interested in computers and telecommunications since childhood. First licensed in 1975, he still maintains his amateur radio license, call sign AB3A. In 1986, he began his career at the Washington Suburban Sanitary Commission (WSSC) as

an instrumentation and telecommunications technician while attending evening classes at the Johns Hopkins University Whiting School of Engineering. He received a bachelor's degree in electrical engineering in 1991. Due to the economy at the time, he chose to stay at WSSC and has not regretted that decision one bit.

Jake has worked on every aspect of SCADA and control systems for WSSC, from the assembly language firmware of the remote terminal unit to the communications protocols and the telecommunications networks, including frequency-division multiplexing analog and digital microwave radios, the data networks, systems programming, protocol drivers, human–machine interface design, and programmable logic controller programming. In 1994 and 1995, Jake participated under a special temporary permit from the Federal Communications Commission to use spread spectrum on the air as an amateur radio licensee. As a result, he is also very much aware of the practical limitations behind the designs of spread-spectrum radio systems.

In 2007, Jake became a voting member of the distributed network protocol (DNP3) Technical Committee, and in 2012 he was elected chairman of the DNP user group. Jake has contributed to the National Institute of Standards and Technology SP 800-82 effort and to the ISA-99 effort. He is also a cofounder and moderator of the SCADASEC e-mail list. Jake is a registered professional engineer of control systems in the state of Maryland, and he has coauthored chapters on control systems for several texts, including *The Instrument Engineers Handbook Volume 3* (CRC Press, August 2011) and *Corporate Hacking and Technology-Driven Crime* (IGI Global, August 2010). His most recent writing effort was coediting and cowriting an edited book on SCADA security with Robert Radvanovsky entitled *Handbook for SCADA and Control Systems Security, First Edition* (released February 2013).

Contributors

This book was written with the community in mind; it brings about a sense of ownership, pride, and responsibility in our actions, thoughts, and movement. The contributors who are listed provided time and effort that they felt was relevant to this book, providing insight and expertise knowledge in areas of engineering, information technology, security, risk management, and more. The editors of this book would like to express their gratitude and to thank each and every contributor for their contribution toward this (and perhaps future) endeavors. Contributors' names are listed alphabetically.

Wayne Boone, CD, PhD, CISSP, CPP, CBCP, CISM, PCIP
Assistant Professor of International Affairs, Deputy Director, Canadian Centre of Intelligence and Security Studies (CCISS)

Dr. Wayne Boone was the coordinator and principal instructor of the infrastructure protection and international security (IPIS) program at Carleton University in Ottawa, Ontario, Canada. He had over 33 years of asset protection and security (AP&S) experience in the areas of force protection, critical infrastructure protection, security risk management, physical security, operations security, and information system/SCADA security, first as an officer in the Canadian Forces Security and Military Police (SAMP) branch, then as a consultant with Precision Security Consulting, and finally as an academic, instructor, and technical adviser/leader in AP&S projects through his role as a driving force in Carleton University's masters of infrastructure protection and international security program. Wayne researched at the leading edges of thinking for AP&S governance and oversight within the public and private sectors. He was active in the conceptualization and development of internationally recognized certification programs in AP&S.

Vytautas Butrimas
Cybersecurity and IT Department Adviser
Ministry of National Defence, Republic of Lithuania
National Communications Regulatory Authority Council Member

Dr. Vytautas Butrimas has been working in information technology and security policy for over 26 years, starting from his work as a government computer specialist to his present role as vice-minister at the Ministry of Communications and Informatics, Republic of Lithuania, responsible for information society development. In 1998, Vytautas moved on to the Ministry of National Defense (MoND) as policy and planning director where he participated in NATO membership preparations, managing a task force preparing Lithuania's first national military defense strategy. From 2001 to 2011, Vytautas served as deputy director of the Communications and Information Systems Service (CISS) under the MoND, where he

led two task forces preparing the first MoND Cybersecurity Strategy and Implementation Plan, and he contributed to the creation of the MoND CERT. Serving from 2011 to 2014 as chief MoND adviser for cybersecurity, he was also a member of the working group that drafted the recently approved new law on cybersecurity. Vytautas has participated in NATO and national cybersecurity and crisis management exercises, which included critical infrastructure threat scenarios. In 2007 (and again in 2012), the president of the Republic of Lithuania appointed Vytautas to the National Communications Regulatory Authority Council (RRT-Council). Vytautas has contributed to various international reports and has written several articles on cybersecurity and defense policy issues. He currently serves in the MoND Cybersecurity and Information Technology Department as senior adviser.

Jim Butterworth, CFE, GCIA, GSNA, GREM, EnCE
Chief Security Officer
HBGary

Jim Butterworth joined Soliton Systems as chief technology officer. He leads Soliton's product development in the area of cybersecurity, incident response, malware analysis, and insider threats. In addition, he provides global executive guidance in all matters pertaining to cybersecurity for Soliton clients. He is a member of the Sacramento Chapter of the Association of Certified Fraud Examiners.

Robert M. Lee
Founder and CEO, Dragos Security

Robert M. Lee is a SANS Institute Certified Instructor, the course author of *ICS515—Active Defense and Incident Response*, and the coauthor of *FOR578—Cyber Threat Intelligence*. He is also the CEO of Dragos Security, a nonresident national cybersecurity fellow at the New America think tank, a PhD candidate at Kings College London, the author of the book *SCADA and Me*, and the writer for the weekly web comic *Little Bobby*. Robert gained his start in cybersecurity in the U.S. Intelligence Community as an Air Force cyberwarfare cyberspace operations officer, where he stood up and led a first-of-its-kind intrusion analysis mission focusing on the identification of national-level adversaries breaking into critical infrastructure sites.

Allan McDougall, BA, BMASc, PCIP, CMAS, CISSP, CPP
Director, Evolutionary Security Management

Allan McDougall is a 20-year veteran security practitioner within the public and private sectors. Following his service with Canada's combat engineers, he has held senior technical advisory positions within the Federal Public Service in the security community, including the Department of Fisheries and Oceans, Canadian Coast Guard, Transport Canada, and Canada Border Services Agency. He has established himself as one of the leading contributors to transportation system security theory and has coauthored several works (including with Robert Radvanovsky, *Critical Infrastructure: Homeland Security and Emergency Preparedness* and *Transportation Systems Security*), has published several white papers on topics such as the dissolution and fragmentation of transportation networks, and has spoken at a number of universities on the protection of supply chains and related asset protection and security topics. He has served as the chair, Supply Chain and Transportation Security Council with ASIS International and was a founding member and later president

of the International Association of Maritime Security Practitioners. He is currently active in a number of industry and cyberrelated working groups.

Seán McGurk, B.ET, B.TE
Global Manager, Intel Security
Senior Vice President, Centre for Strategic Cyberspace and Security Science

Seán Paul McGurk is the global manager of critical infrastructure protection at Intel Security. He also serves as the senior vice president of the National Critical Infrastructure CSCSS/Centre for Strategic Cyberspace and Security Science, an independent research organization. McGurk holds undergraduate degrees in electronic technology and technical education. He is a member of the Information Systems Security Association (ISSA) and the Institute of Electrical and Electronics Engineers (IEEE). He has received numerous awards, including the 2011 Federal 100 Award and the 2010 and 2009 SANS SCADA Leadership Awards.

Bernie Pella, GIAC, GSLC
Principal Cyber Security Consultant, Schneider Electric, Global Cyber Security Services

Bernie Pella has more than 30 years' experience in the area of nuclear and process controls. He is currently a cybersecurity consultant for the Invensys Critical Infrastructure and Security Practice. Bernie has experience in implementing the process controls and engineering automation cybersecurity program at the Savannah River Site, a Department of Energy–owned nuclear facility in Aiken, South Carolina. He spent 19 years at the Savannah River Site in various engineering positions that included 10 years as a shift technical engineer, process controls engineer, plant engineer, and cybersecurity engineer. Bernie also obtained commercial nuclear and building automation experience after leaving the U.S. Navy in 1986. Bernie spent 8 years in U.S. Navy submarine nuclear power operations, assigned to the *USS Scamp*-(SSN-588) during a major overhaul, and he was on the commissioning crew of *USS Buffalo*-(SSN-715). Bernie is a member of the industrial control system Joint Working Group and has presented many different industrial control system topics at conferences over the last several years. Bernie used an extended study degree program and is a 2008 graduate of Excelsior College with a bachelor of science in technology.

Lt. Colonel (USAR retired) Darrell G. Vydra, B.Eng, MBA, CISSP, ISSMP, PMP
Founder and Principal, Vydra Consulting

Lieutenant Colonel (retired) Darrell G. Vydra graduated from the United States Military Academy in 1981 with a BS in General Engineering, was commissioned in the United States Army, and served on active duty for 11.5 years before he entered the United States Army Reserves in 1992. He served in a number of capacities and global locations during his 30-year career (platoon leader and battery commander in Germany, battalion commander at Fort Sheridan, IL, foreign military sales liaison officer to the Kingdom of Saudi of Arabia, deputy director of displaced persons and refugees in Bosnia and Herzegovina, chief of information operations in Afghanistan, and director of strategic marketing and professional development at CENTCOM at MacDill Air Force Base, FL). He mastered both project management and physical security practices and expertise as a military officer. He began his civilian career as a defense contractor, moved into private-sector jobs, and

quickly moved into project management, sales/marketing, and engineering positions while earning an MBA from Florida Institute of Technology in Acquisition, Procurement and Life-Cycle Management and an MS from DePaul University in Telecommunications Systems Management. He moved into information technology (IT) security operations in the early 2000s and earned his CISSP and ISSMP from ISC2 and his later earned his PMP from PMI. He now works as a consultant for Energy and Utilities, namely project managing governance, risk management and compliance (GRC) projects.

Joseph Weiss, PE, CISM, CRISC, ISA Fellow, IEEE Senior Member
Applied Control Solutions, LLC

Joseph Weiss is an industry expert on the electronic security of control systems, with more than 40 years' experience in the energy industry. He serves as a member of numerous international organizations related to control system security and has published over 80 papers on instrumentation, controls, and diagnostics, including the book *Protecting Industrial Control Systems from Electronic Threats*. He is an ISA Fellow; managing director of ISA Fossil Plant Standards, ISA Nuclear Plant Standards, ISA Industrial Automation and Control System Security (ISA99); a Ponemon Institute Fellow; and an IEEE Senior Member. He has two patents on instrumentation and control systems, is a registered professional engineer in the State of California and a certified information security manager (CISM) and is certified in risk and information systems control (CRISC).

Jeff Woodruff, CD, CAS
Departmental Security Officer, Canadian Radio-Television and Telecommunications Commission

Jeff Woodruff is a 25-year veteran of the police, security, and emergency management fields and currently holds the position of departmental security officer for one of Canada's federal government entities. In this capacity, he manages the security program, safety program, business continuity program, and emergency management program. As a former military policeman in the security branch of the Canadian Armed Forces, he served two tours in Canada's Special Operations Command, providing close security support for a counterterrorism unit. His work in physical security and operational risk management has been widely recognized within the public service community, particularly among security practitioners and professionals.

Craig Wright, GSE CISSP, CISA, CISM, CCE, GCFA, GLEG, GREM, GSPA
Vice President, Centre for Strategic Cyberspace and Security Science

Dr. Craig Wright is a lecturer and researcher at Charles Sturt University and vice president of the National Critical Infrastructure CSCSS/Centre for Strategic Cyberspace and Security Science with a focus on collaborating government bodies in securing cybersystems. With over 20 years of IT-related experience, he is a sought-after author and public speaker both locally and internationally, training Australian government and corporate departments in SCADA security, cybersecurity, and cyberdefense, while also presenting his latest research findings at academic conferences. Dr. Wright holds the industry certifications GSE CISSP, CISA, CISM, CCE, GCFA, GLEG, GREM, and GSPA, and he is working on his second PhD on the quantification of information systems risk.

Steven Young, MBA, IEM, CHS-V, IAHSS
Security Strategist and Principal Consultant

Steven has 20 years of diverse experience in the financial, medical, pharmaceutical, manufacturing, and government industry sectors (law enforcement and defense). He is an expert in leading security audits, performing resilience architecture assessments, developing disaster recovery plans, investigating security incidents, and assessing business risks/threats. He has a BA from Loyola University and an MBA from the University of Notre Dame. He is published in several industrial and law enforcement trade journals. He has served as an information security consultant for the U.S. Navy, the U.S. Coast Guard, and the U.S. Federal Reserve. He is also a licensed investigator, specializing in data breaches in two states with reciprocity in several others.

Editors' notes

This publication offers an aid to maintaining professional competence, with the understanding that neither the editors, chapter authors, or publisher are rendering any legal, financial, or other professional advice.

Due to the rapidly changing nature of the industrial control systems (ICS) security community, the information contained within this publication may become outdated, and therefore the reader should consider researching alternative or other professional or more current sources of authoritative information. A significant portion of this publication was based on research conducted from several government resources, publications, and Internet-accessible websites, some of which may no longer be publicly available or may have been restricted due to laws enacted by that country's federal or national government.

The views and positions taken in this book represent the considered judgment of the editors and chapter authors. They acknowledge, with gratitude, any inputs provided and resources offered that contributed to this book. Moreover, for those who have contributed to the book's strengths and its characteristics, we would like to say "thank you" for your contributions and efforts. For any inconsistencies that have been found, we alone share and accept the responsibility for them and will gladly make corrections as needed.

One additional note concerns the evolutionary process that we are witnessing within this community. The evolvement concerns itself with the transition from a traditional perspective—that ICSs are "islands"—to the current moment, in which those very systems are now interconnected, either privately or via open communications mediums (such as the Internet); additionally, ICSs are being treated less as an engineered automation plant asset, and more as an information technology (IT) asset, and thus we are seeing the initial witnessed efforts of a paradigm shift from engineering to IT. Part of the reason for this paradigm shift is the lack of qualified process control engineers who are technically competent in ICS design and implementation; the other part is that the term "security" has a different meaning and context within the engineering community compared to the IT community, causing continued cultural differences between them.

As there have been very few publications dedicated to this community, efforts involving establishing best practice methods, metrics, and standards continue to evolve; thus, this book represents a work in progress. Although we realize that there may be some areas that are lacking or are weak in their dissertation, please understand that we are striving for as complete a book as possible. For example, there are currently no generally accepted performance-based auditing criteria. Therefore, we have eschewed the auditing chapter as we feel that merely confirming the purchase of equipment and training of personnel does not constitute a valid security audit. For this reason, auditing has not been included in this publication.

section one

Social implications and impacts

chapter one

Introduction

Jacob Brodsky and Robert Radvanovsky

Contents

Critical infrastructure consists of both physical and cyberbased systems (along with their assets) that are essential to an economic state such that the disruption or destruction of their operations would have a debilitating impact on the security, public health, and safety of that economy. This transpires worldwide. These systems (and their assets) provide essential, yet vital, products and services to our economies, which include products such as food and critical manufactured products, or services such as our electricity, water, and wastewater treatment facilities, chemical and oil production facilities, and transportation modes. All these are essential to the operations of economies and their governments. Threats in recent years have underscored the need to protect many of our infrastructures. If vulnerabilities in these infrastructures are exploited, our critical infrastructures could be disrupted, disabled, possibly causing loss of life, physical damage, and economic losses (U.S. General Accounting Office 2007). A majority of the infrastructures worldwide are owned and operated privately by corporations.

What are "control systems," and why are they important?

Generally speaking, most control systems are computer based. They are used by many infrastructures and industries to monitor and control sensitive processes and physical functions. Typically, control systems collect sensor measurements and operational data

from the field, process and display this information, and relay control commands to local or remote equipment. In the electric power industry, they can manage and control the transmission and delivery of electric power; for example, by opening and closing circuit breakers and setting thresholds for preventive shutdowns. Using integrated control systems, the oil and gas industry can control the refining operations on a plant site as well as remotely monitor the pressure and flow of gas pipelines and control the flow and pathways of gas transmission. With water utilities, control systems can remotely monitor well levels; control the wells' pumps; monitor water flows, tank levels, or water pressure in storage tanks; monitor water quality characteristics such as pH, turbidity, and chlorine residual; and control the addition of chemicals. Control system functions vary from simple to complex; they may be used to simply monitor processes that are running; for example, from environmental conditions within a small office building (the simplest form of site monitoring) to managing most (or, in most cases, all) activities for a municipal water system or even a nuclear power plant. Within certain industries such as chemical and power generation, safety systems are typically implemented to mitigate a disastrous event if control and other systems fail.

Control systems were not always computer based. In fact, there are still many pneumatic control systems. Some are analog systems, based on operational amplifier circuits. Some are mechanical feedback systems and others are hydraulic; for example, the set point for many pressure-reducing valves is made by setting the position of a hydraulic pilot valve configuration.

In addition to guarding against both physical attack and system failure, organizations may establish backup control centers that include uninterruptible power supplies and backup generators (Library of Congress 2004).

Types of control systems

There are two primary types of control systems:

1. Distributed control systems (DCSs) are typically used within a single process or generating plant, or used over a smaller geographic area or even a single-site location.
2. Supervisory control and data acquisition (SCADA) systems are typically used for larger-scale environments that may be geographically dispersed in an enterprise-wide distribution operation.

A utility company may use a DCS to generate power and may use a SCADA system to distribute it (Library of Congress 2004).

Control loops in a SCADA system tend to be open, whereas control loops in a DCS tend to be closed. The SCADA system communications infrastructure tends to be slower and less reliable, and so the remote terminal unit (RTU) in a SCADA system has local control schemes to handle that eventuality. In a DCS, networks tend to be highly reliable, high-bandwidth campus local area networks (LANs). The remote sites in a DCS can afford to send more data and centralize the processing of that data (Radvanovsky and McDougall 2009).

Components of a control system

A control system typically consists of a master control system or central supervisory control and monitoring station, consisting of one or more human–machine interfaces (HMI)

in which an operator may view displayed information about the remote sites and issue commands directly to the system. Typically, this is a device or station that is located at a site in which application servers and production control workstations are used to configure and troubleshoot other control system components. The central supervisory control and monitoring station is generally connected to local controller stations through a hardwired network, or to remote controller stations through a communications network that may be communicated through the Internet, a public-switched telephone network (PSTN), or a cable or wireless (such as radio, microwave, or wireless) network (Radvanovsky and McDougall 2009).

Each controller station has an RTU, a programmable logic controller (PLC), a DCS controller, and/or other controllers that communicate with the supervisory control and monitoring station. The controller stations include sensors and control equipment that connect directly with the working components of the infrastructure (e.g., pipelines, water towers, and power lines). Sensors take readings from infrastructure equipment, such as water or pressure levels and electrical voltage, sending messages to the controller. The controller may be programmed to determine a course of action, sending a message to the control equipment instructing it what to do (e.g., to turn off a valve or dispense a chemical). If the controller is not programmed to determine a course of action, the controller communicates with the supervisory control and monitoring station before sending a command back to the control equipment. The control system may also be programmed to issue alarms back to the control operator when certain conditions are detected. Handheld devices such as personal digital assistants (PDAs) may be used to locally monitor controller stations. Controller station technologies are becoming more intelligent and automated and can communicate with the supervisory central monitoring and control station less frequently, requiring less human intervention. Historically, security concerns about control stations have been less frequent, requiring less human intervention (Radvanovsky and McDougall 2009).

Vulnerability concerns about control systems

Security concerns about control systems were primarily historically related to protection against physical attacks or the misuse of refining and processing sites or distribution and holding facilities. However, in more recent years, there has been a growing recognition that control systems are now vulnerable to cyberattacks from numerous sources, including hostile governments, terrorist groups, disgruntled employees, and other malicious intruders (Radvanovsky and McDougall 2009). Without going into too much of a dissertation about recent malware outbreaks, such as Stuxnet and Duqu, the malware Stuxnet* alone has been one of the most heavily researched, discussed, and hypothesized of any known control systems malware to date.

Several factors have contributed to the escalation of risk of these control systems, which include the following concerns:

- The adoption of standardized technologies with known vulnerabilities
- The connectivity of many control systems via, through, within, or exposed to unsecured networks, networked portals, or mechanisms connected to unsecured networks (which includes the Internet)

* Stuxnet was considered a "worm," which is a self-replicating virus.

- Implementation constraints of existing security technologies and practices within the existing control systems infrastructure (and its architectures)
- The connectivity of insecure remote devices in their connections to control systems
- The widespread availability of technical information about control systems, most notably via publicly available or shared networked resources such as the Internet

Adoption of standardized technologies with known vulnerabilities

Historically, proprietary hardware, software, and network protocols made it rather difficult to understand how control systems operated, as information was not commonly or publicly known, was considered proprietary (in nature), and was therefore not susceptible to hacker attacks. Today, however, to reduce costs and improve performance, organizations have begun transitioning from proprietary systems to less expensive, standardized technologies that use and operate under platforms that run operating systems such as Microsoft Windows, UNIX, and LINUX systems, along with the common networking protocols used by the Internet. These widely used standardized technologies have commonly known vulnerabilities such that more sophisticated and effective exploitation tools are widely available and relatively easy to use. As a consequence, both the number of people with the knowledge to wage attacks and the number of systems subject to attack have increased (Radvanovsky and McDougall 2009).

Connectivity of control systems to unsecured networks

Corporate enterprises often integrate their control systems within their enterprise networks. This increased connectivity has significant advantages, including providing decision makers with access to real-time information, allowing site engineers and production control managers to monitor and control the process flow and its control of the entire system from within different points of the enterprise network. Enterprise networks are often connected to networks of strategic partners as well as to the Internet. Control systems are increasingly using wide area networks and the Internet to transmit data to their remote or local stations and individual devices. This convergence of control networks with public and enterprise networks potentially exposes the control systems to additional security vulnerabilities. Unless appropriate security controls are deployed within and throughout the enterprise and control system network, breaches in enterprise security may affect operations (Radvanovsky and McDougall 2009).

Implementation constraints of security technologies of control systems

Existing security technologies, as well as strong user authentication and patch management practices, are typically not implemented in the operation of control systems; additionally, most control systems are typically not designed with security in mind and usually have limited processing capabilities to accommodate or handle security measures or countermeasures (Radvanovsky and McDougall 2009).

Existing security technologies such as authorization, authentication, encryption, intrusion detection, and filtering of network traffic and communications require significantly increased bandwidth, processing power, and memory—much more than control

system components typically have or are capable of sustaining. The entire concept behind control systems was integrated systems technologies, which were small, compact, and relatively easy to use and configure. Because controller stations are generally designed to perform specific tasks, they use low-cost, resource-constrained microprocessors. In fact, some devices within the electrical industry still use the Intel 8088 processor, which was introduced in 1978. Consequently, it is difficult to install existing security technologies without seriously degrading the performance of the control systems (or causing disruptions of entire control systems networks), thus requiring the need for a complete overhaul of the entire control system infrastructure and its environment (Radvanovsky and McDougall 2009).

Furthermore, complex password-controlling mechanisms may not always be used to prevent unauthorized access to control systems, partly because this could hinder a rapid response to safety procedures during an emergency or could affect the performance of the overall environment. As a result, according to experts, weak passwords that are easy to guess, are shared, and are infrequently changed are reportedly common in control systems, including the use of default passwords or even no password at all (Radvanovsky and McDougall 2009).

Current control systems are based on standard operating systems as they are typically customized to support control system applications. Consequently, vendor-provided software patches are generally either incompatible or cannot be implemented without compromising service by shutting down "always-on" systems or affecting interdependent operations (Radvanovsky and McDougall 2009).

Insecure connectivity to control systems

Potential vulnerabilities in control systems are exacerbated by insecure connections, either within the corporate enterprise network or external to the enterprise or controlling station. Organizations often leave access links (such as dial-up modems to equipment and control information) open for remote diagnostics, maintenance, and examination of system status. Such links may not be protected with any authentication or encryption (or if any exist, are considered rather weak as the individuals who configured the control systems environments wanted something easy to remember, since oftentimes they had to maintain and manage hundreds of similar devices throughout a given area of region). This increases the risk that an attempted external penetration could use these insecure connections to break into remotely controlled systems. Some control systems use wireless communications systems, which are especially vulnerable to attack, or leased lines that pass through commercial telecommunications facilities; in either situation, the method of communication performs no security methodologies whatsoever and, if there are any security measures implemented, they are capable of being easily compromised. Without encryption to protect data as it flows through these insecure connections or authentication mechanisms to limit access, there is limited protection for the integrity of the information being transmitted, and the process may be subjected to interception, monitoring of data from interception, and (eventually) penetration (Radvanovsky and McDougall 2009).

Publicly available information about control systems

Public information about critical infrastructures and their control systems is available through widely available networks such as the Internet. The risks associated with the availability of critical infrastructure information poses a serious threat to those infrastructures

being served. This has been repeatedly demonstrated by graduate students from several academic institutions over the past several years, whose dissertations reported either partial or complete relevant and sensitive information about specifically targeted infrastructures; this information, if utilized, could provide threat vector methods of attack, allowing subversive communications into and throughout these infrastructures and their control systems' networks. A prime example of publicly available information is with regard to the electric power industry, in which open sources of information such as product data, educational materials, and maps (even though outdated) are still available, showing line locations and interconnections that are currently being used; additional information includes filings of the Federal Energy Regulatory Commission, industrial publications on various subject matters pertaining to the electric power industry, and other materials—all of which are publicly available via the Internet (Radvanovsky and McDougall 2009).

Recently, other more invasive methods of determination through commercial services that probe for specific Internet functions (such as web services) somehow found either partially protected, if not completely open, control systems directly connected to the Internet (ICS-CERT 2011a).

The use of readily available and generally free search tools significantly reduces time and resources required to identify Internet-facing control systems. In turn, adversaries can utilize these tools to easily identify exposed control systems, posing an increased risk of attack. Conversely, owners and operators can also use these same tools to audit their assets for unsecured Internet-facing devices (ICS-CERT 2011a).

Internet-facing control systems have been identified in several critical infrastructure sectors. The systems vary in their deployment footprints, ranging from stand-alone workstation applications to larger DCS configurations. In most circumstances, these control systems were designed to allow remote access for system monitoring and management. All too often, remote access has been configured with direct Internet access (with no firewall) or utilizing either default or weak user names and passwords. These default and common account credentials are often readily available in public space documentation (in some cases, even on the control systems' manufacturers' websites).

Control systems are vulnerable to attack

Entities or individuals with intent to disrupt service may use one or more of the following threat vector methods, which may be successful in their attack(s) of control systems (U.S. General Accounting Office 2004):

- Disrupting the operations of control systems by delaying or blocking the flow of information through the networks supporting the control systems, thereby denying availability of the networks to control systems' operators and production control managers.
- Attempting to or succeeding in making unauthorized changes to programmed instructions within PLC, RTU, or DCS controllers, change alarm thresholds, or issue unauthorized commands to control station equipment, which could potentially result in damage to equipment (if tolerances have been exceeded), premature shutdown of processes (shutting down transmission lines or causing cascading termination of service to the electrical grid), or disabling control station equipment.
- Sending falsified information to control system operators, either to disguise unauthorized changes or to initiate inappropriate actions to be taken by systems operators—that is, falsified information is sent or displayed back to system operators who

may think that an alarmed condition has been triggered, resulting in system operators acting on this falsified information, thus potentially causing the actual event.
- Modifying or altering control system software or firmware such that the net effect produces unpredictable results (such as introducing a computer "time bomb" to go off at midnight every night, thus partially shutting down some of the control systems, causing a temporary brownout condition; a "time bomb" is a forcibly introduced piece of computer logic or source code that causes certain courses of action to be taken when either an event or triggered state has been activated).
- Interfering with the operation and processing of safety systems (e.g., tampering with or denial of service of control systems that regulate processing control rods within a nuclear power generation facility).
- Many remote locations containing control systems (as part of an enterprise DCS environment) are often unstaffed and may not be physically monitored through surveillance; the risk of threat remains and may be higher if the remote facility is physically penetrated at its perimeter and intrusion attempts are then made to the control systems' networks from within.
- Many control systems are vulnerable to attacks of varying degrees; these attack attempts range from telephone line sweeps (aka wardialing), to wireless network sniffing (wardriving), to physical network port scanning and physical monitoring and intrusion.

Consequences of compromised control systems

Some consequences resulting from control system compromises are as follows:

- Although computer network security is undeniably important, unlike enterprise network security, a compromised control system can have significant impacts within real-world life. These impacts can have far-reaching consequences not previously thought of, or in areas that could affect other industrial sectors (and their infrastructures).
- Enterprise network security breaches can have financial consequences: customer privacy becomes compromised; computer systems need to be rebuilt, and so on.
- A breach of security of a control system can have a cascade effect on other systems, either directly or indirectly connected to those control systems that have been compromised; however, not only can property be destroyed, but people can be hurt or, even worse, be killed (St. Sauver 2004).

False reports of vulnerabilities involving control systems

Not all situations are actual security incidents; in some rare cases, certain circumstances can be expounded negatively almost as bad as the threats themselves, making for a "false-positive" scenario in which there never was a given cyberincident, but is exacerbated due to press coverage and incorrect (or untimely) information gathered. For example, on November 10, 2011, the Illinois Statewide Terrorism & Intelligence Center (STIC) issued a daily intelligence notes report entitled "Public Water District Cyber Intrusion." As widely reported in the press, the report detailed initial findings of anomalous behavior in a SCADA system at a central Illinois public water district, and alleged a malicious cyberintrusion from an IP address located in Russia that caused the SCADA system to power itself on and off, resulting in a water pump burn out. ICS-CERT was made aware of the report on November 16, 2011, and immediately reached out to the STIC to gather additional information, in which ICS-CERT was provided with a log file; however, initial

analysis could not validate any evidence to support the assertion that a cyberintrusion had occurred (ICS-CERT 2011b).

ICS-CERT reached out to the affected entity, Curran-Gardner Public Water District, to gather detailed information, offering support and analytics to uncover what caused the pump to fail.* After detailed analysis of all available data, ICS-CERT, along with the FBI, found no evidence of a cyberintrusion into the SCADA system of the Curran-Gardner Public Water District in Springfield, Illinois. At the request of the utility and in coordination with the FBI, ICS-CERT deployed a flyaway team to the facility to interview personnel, perform physical inspections, and collect logs and artifacts for analysis (ICS-CERT 2011b).

There was no evidence to support claims made within the initial Illinois STIC report—which was based on raw, unconfirmed data and subsequently leaked to the media—that any credentials were stolen or that the vendor was involved in any malicious activity that led to a pump failure at the water plant. News of a potential cyberattack reached the media almost immediately and spread quickly worldwide. At the end of their analysis, both the Department of Homeland Security (DHS) and the FBI concluded that there was no malicious or unauthorized traffic from Russia, or that any foreign entities, as previously reported, had infiltrated the water utility. Analysis of what caused the pump failure has yet to be disclosed publicly (ICS-CERT 2011b).

The net result demonstrated several days of unnecessary time and resources expended in support and analysis by several organizations, in which many felt that the central Illinois water utility was penetrated, and, along with some conspiracy theorists, further complicated the situation by making false accusations that the entire scenario was a government "cover-up"—when, in fact, no threat, no intrusion whatsoever had existed.

Control systems community challenges

One of the more interesting challenges is how to address security-related issues within the SCADA/control systems community, and the sectors it supports, as SCADA/control systems enterprises do not operate in a context similar to that of their traditional IT counterparts. It is probable that one of the more significant aspects to control systems is the scope in which they dictate how issues are to be addressed (Radvanovsky and McDougall 2009).

Many technologies within the IT realm, such as SQL database transaction speeds, have traditionally been viewed by SCADA/control systems engineers as having inadequate speed for control system data storage purposes. Although the technology has made this operation outmoded (Moore's law), most opinions are difficult to shake, and thus many process control engineers continue to have difficulties accepting IT solutions within their environments. Based on some of the challenges mentioned in this paragraph, the problem is not so much a matter of data management as it is about trends and statistical analysis.

One of the larger problems is that forensics and evidentiary discovery practices are often associated with security management practices. Within control systems, these priorities are a little bit different from normalized systems, which are (usually) listed in the following order:

1. Safety
2. Availability
3. Security

* According to the ICS-CERT report, at no time were there any impacts to customers served by the water district due to the pump failure. Refer to ICS-CERT (2011b), p. xxii for the detailed report.

Note where "security" is listed: last. The reason for this is that IT-based architectures may be completely inverted from the priorities listed earlier, and thus there appears to be a conflict between what/how SCADA/control systems operate and (more importantly) how the corporation's enterprise defines its priorities. Several industries are currently attempting to either reach a compromise or figure out how both environments—IT and control systems communities—can work together. Observationally, in some industries, such as nuclear power generation, these environments may never coexist together—ever (Radvanovsky and McDougall 2009).

Some of the larger issues associated with control systems involve legacy architectures no longer supported, utilize equipment that cannot be taken off-line immediately or easily, and pose serious operational and financial risks to the companies using them. Unless these systems are interconnected with newer systems or are upgraded, there is no easy method of determining a plausible cause for any given event or incident. Outside of what may be found at the company's control center, there is little forensic data to be found, as control center computers do not lend themselves to traditional forensics analysis unless taken off-line or removed off-site. Given the nature of most control systems, if it is an ongoing operational need, it may be very difficult to remove the servers in question for an extended analysis.

Where does control systems security fit?

Of the more interesting discussions over the years, one of the more intriguing is where SCADA/control systems security fits into the overall picture. Some would like to think that SCADA/control systems security should be isolated and set apart from traditional IT-related security environments, whereas others feel that it should be combined. One perspective suggested an alternative: combining a set of interlocking circles, whereby the significant security practices, with SCADA/control systems security being the smallest and having an interconnecting function between the other two security practices, are *dead center* between significant IT and control systems practices. Although the exact number is not known, SCADA/control systems security practitioners have the smallest number of experts (even though this area is growing and evolving). To understand the scale of the number of IT security practitioners versus SCADA/control systems security practitioners, see Figure 1.1.

Future of control systems

As for where things are going, control systems will have to be segmented and configured so that high-risk sections of the control system will have to be carefully protected. These include several threats. First, ensuring that logging takes place in more than one part of a control system. When the gates of a dam are opened, there should be not only a digital signature of the operator who initiates the command at the master station from which it was sent, but also the signature of the operator at the RTU where the command was executed (Radvanovsky and McDougall 2009).

Protocols such as IEC-60870 and Distributed Network Protocol 3 (DNP3) have recently added secure authentication features to make this possible. The new specification can be found in IEC-62351.

The future holds much promise with protocols such as IEC-61850. However, it is an extremely complex undertaking that mixes many features into one layer. The maintenance management system is a nice feature with which to integrate the control systems' data, but

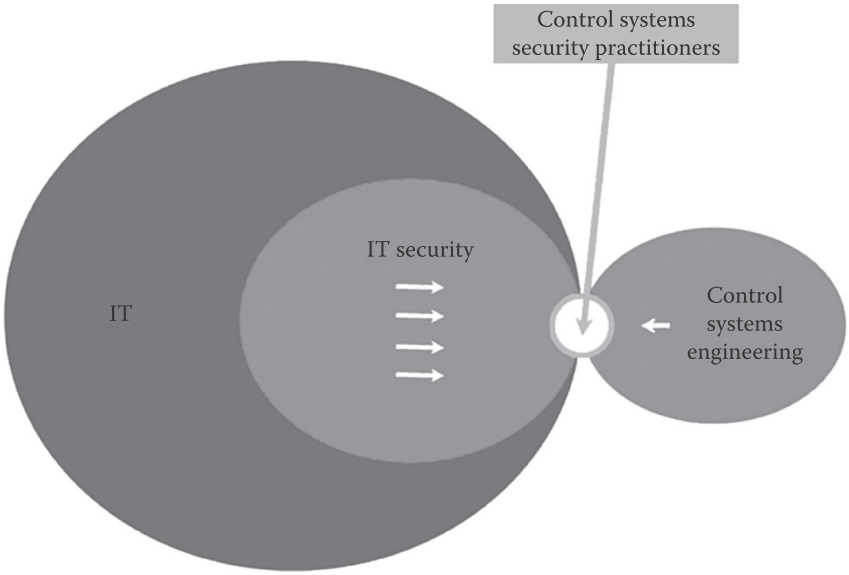

Figure 1.1 Comparative graphical representation of estimated total number of control systems security practitioners against other security practitioners. (Courtesy of Applied Control Systems.)

it may not be the best thing to place on the control systems' communications infrastructure. One of these operational elements is tactically significant and the other is strategically significant (Radvanovsky and McDougall 2009).

We may want to consider ways of segmenting and separating traffic for security reasons. This could entail reexamining the lower layers of the communications infrastructure.

SCADA/control systems' infrastructure needs to use a variety of ways to connect to remote stations. The goal is to avoid having common carrier problems disable a control system that it might depend on. Multiheaded RTU devices may be the future of many control systems.

Note the convergence of DCS and SCADA/control systems technologies. The SCADA/control systems concept originally grew from dealing with the constraints of high latency, low reliability, and expensive bandwidth. DCS concepts originally grew from the need to network everything to one central computer where everything could be processed all at once. DCSs are also getting smarter about how they distribute the functional pieces, and SCADA/control systems are handling closed loops more often as the communications infrastructure gets faster and more reliable (Radvanovsky and McDougall 2009).

This book provides a culmination of differing perspectives, ideals, thoughts, and attitudes toward securing SCADA and control systems environments. The thought is to provide a community-based effort toward establishing a strategy that can be established and utilized throughout the SCADA and control systems community. Although many of the chapters are all widely known and established within the IT, network, and security communities, to combine all three ideologies into one great big effort is a daunting task, and one in which we hope to achieve through community involvement through this book. Thus, this book is a living, breathing work in progress due to the quickly changing landscape of the SCADA and control systems security community.

References

Industrial Control Systems Computer Emergency Response Team (ICS-CERT). 2011a. ICS-ALERT-11-343-01—Control system Internet accessibility. http://www.us-cert.gov/control_systems/pdf/ICS-ALERT-11-343-01.pdf.

Industrial Control Systems Computer Emergency Response Team (ICS-CERT). 2011b. ICSB-11-327-01—Illinois water pump failure report. http://www.us-cert.gov/-control_systems/pdf/ICSB-11-327-01.pdf; http://news.infracritical.com/pipermail/scadasec/2011-November/thread.html.

Library of Congress. 2004. CRS Report for Congress. Critical infrastructure: control systems and the terrorist threat, CRS-RL31534. http://www.fas.org/irp/crs/RL31534.pdf.

Robert Radvanovsky and Allan McDougall. 2009. *Critical Infrastructure: Homeland Security and Emergency Preparedness*. 2nd edn. Boca Raton, FL: CRC Press/Taylor & Francis Group.

Joe St. Sauver. 2004. *NLANR/Internet2 Joint Techs Meeting: SCADA Security*. Columbus, OH: University of Oregon.

U.S. General Accounting Office. 2004. Critical infrastructure protection: challenges and efforts to secure control systems. GAO-04-354. Washington, DC. http://www.gao.gov/new.items/d04354.pdf.

U.S. General Accounting Office. 2007. Critical infrastructure protection: multiple efforts to secure control systems are under way, but challenges remain. GAO-08-119T. http://www.gao.gov/new.items/d08119t.pdf.

chapter two

Sociological and cultural aspects

Jacob Brodsky and Robert Radvanovsky

Contents

This chapter describes the current social aspects to implementing an industrial control system security program. Industrial control systems security is still in its infancy and, as such, there is resistance from many avenues. This chapter outlines the social hurdles, which the various groups are, and what concerns and motivates them.

It may be trite and pedantic to say this, but security begins and ends with people. This fact cannot be emphasized enough when dealing with industrial control system security. In the midst of all this high-tech gadgetry, too many act as if one could instill security with technology alone.

Although technical methods are the means to improving security, they ultimately require people to understand and use them. One can purchase many security technologies for a control system; but, unless the people who operate, maintain, and manage these systems know what to do with them, the return on the investment will be poor.

Security expenditures are not easy to justify. Responsibility for "security," specifically "cybersecurity," is not a very well-understood concept. By comparison, look at how safety works: Even if one were not responsible for a car accident, those who fail to put on a seatbelt are generally regarded as being partly responsible for the outcome. This sort of shared responsibility concept has only just begun to dawn on those who design and operate the security aspects of an industrial control system. Many operators still know little to nothing about how the control system gets data to them. They have no idea of what to do if the integrity is compromised. Many engineers still design systems without any of these features because "the customer didn't ask for it." Finally, many IT staff treat these control systems as if they were just another office application, where the computational service is the work product itself, instead of being a small part of the production effort.

Without a mandate to secure control systems, it is difficult to sell "security" to a company or a utility. The return on the investment is difficult to document. Some view it as an insurance policy; however, the data for this sort of approach is so thin that the risks and rewards are difficult to document. There are few laws mandating the accountability and

reporting capabilities of a (potentially) compromised control system. Without prescriptive standards for recording near-miss metrics, and the resulting paucity of data in common form, few have any idea where to start, what to measure, or how to adjust to various situations.

Even if there is some sort of mandate for security, it is usually defined in terms of compliance instead of a performance approach. Without ubiquitous and standardized metrics, a performance-based approach is considered by many to be insufficiently developed to be regarded as usable. This leads to a "do it because we said so" compliance approach. Unfortunately, the compliance approach is usually an investment without people or training to back it up. Those who use this approach are probably expecting that practitioners will notice some metrics along the way and somehow start building a better performance-based approach. Owning all the tools does not make one a tradesman. Likewise, mere compliance alone will not make anyone more secure.

Like the issue of safety, security is easier to bootstrap in place if it is not sold as such. It can be an employee accountability system, self-integrity monitoring, improved diagnostics, or improved longevity (through better patch management), among many other things. An artful leader will carefully craft these features into a cohesive series of investments that coincidentally improves security.

Suppose that (somehow) these initial objections were overcome, and that an effort was underway to improve security. The logical thing would be to bring the IT security and engineering groups together to build something more secure. However, both professions bring biases to the table that makes working together very contentious. Furthermore, from the operational perspective, there may be significant ignorance of the issue, as it may not have been part of the assumptions behind the design or the operations of the plant. Operations staff need to be taught what to do with these security features and how to react to alarms that these new features will raise.

The fundamental change from older, hardwired automation designs to the newer, more highly networked systems is actually quite subtle. In the past, people had to stand in front of the equipment to operate it. There was very little remote operation capability, and where it did exist, it used an inherently trusted medium: the local telephone systems of the 1970s and 1980s. Engineers and operations staff assumed that those who could access the controls were either standing in front of machinery or were standing in a limited number of places where others could see and monitor their behavior.

Some thought was given to random, nonmalicious ignorance and mistakes; but beyond that, few considered the possibility of active malice on a plant. Malicious acts would tend to hurt the person who committed them, in addition to fellow employees and the public at large. It was presumed that everyone would have a sense of self-preservation.

Gradually, computer automation became more commonplace. Staffing levels were reduced. Operational processes were made more streamlined in an effort to save or conserve money. Eventually, as networking improved, the trend toward reducing staff became even more popular, until eventually one began to read articles about how an operator or engineer, running human–machine interface (HMI)* software from his laptop, was able to save the day for a plant many hours away. Few ever considered that the very features that

* HMI is software that displays information to an operator or user about the current state of an automated process, accepting and implementing any operator control instructions. Typically, information is shown using a graphical representation format (graphical user interface or GUI). HMIs are often considered part of a supervisory control and data acquisition (SCADA) system.

made this sort of rescue possible could also (potentially) provide a venue for sabotage for the plant from halfway around the world.

Engineering perspectives and their reactions

The first reaction from engineers when discussing an industrial security threat is incredulity. Why would anyone do that? They are used to the presumption that people might act in an ignorant manner, but not an actively malicious one. The idea that someone would want to destroy infrastructure seems foreign to those who have only concerned themselves with operating and upgrading that infrastructure over most of their careers.

A response to such concerns would be to discuss the possibility that someone from another social class/country/tribe/religion/etcetera might see an opportunity to hurt the economy of those considered enemies. Or, more likely, it could be a disgruntled contractor or employee who felt that he got a raw deal. The attack vector could be the very thing they used to make remote access possible. It could be a wireless link. It could be a logic bomb. It could be a modem left behind during the construction and testing phase. Unless the whole plant was built from the ground up just a few years ago, the chances are that there are lots of poorly documented "features" that could be exploited by someone with inside knowledge.

The goal is to get engineers to realize that any opportunity to control infrastructure from somewhere else or some earlier development work is a possible source of attack. People with malicious intent against infrastructure do exist. It may be necessary to rub some noses in this ugly reality. Despite the lack of any requirement to make reports of such incidents, there is already ample public evidence that such malicious behavior does occur.

The second reaction from engineers is pretty straightforward: It was not in the design criteria, so why bring this up now? The system does what it was designed to do.

The engineers have a point in this regard. Once upon a time, when these systems were designed, they were not presumed to be attached to any other networks. There was a certain trust because the extent of the network itself was presumed to have been limited. Unfortunately, others probably followed after the original design was completed and "made a tweak" that enabled remote access of some sort.

Again, it is useful to point out that fundamental assumptions behind the design criteria have changed. The systems were never designed for anything other than physical security. Furthermore, while it is not exactly effective for one to "bolt on security after the fact," we cannot ethically leave things as they are.

From a technical perspective, the network capacity and processor speed were selected without security overhead. Introducing that extra overhead may be possible, but full review and testing is needed. The IT security people should not secure the systems without the assistance of the engineering staff. This will become a significant discussion point later, when assigning scope and performance levels.

This is also an issue with how the design took place. Engineers, especially consulting engineers, typically work in a project delivery mode. The project is designed, there are reviews, the plans are bid for, construction takes place, and then everything is tested to ensure it does what it was designed to do. At that point, everyone washes their hands of the whole thing, turns it over to the operations staff, and then goes on to something else. The system is then expected to remain virtually untouched until the whole thing is depreciated enough to warrant upgrades. And then, the cycle continues all over again.

However, security is a continuous, ongoing concern. Project-oriented engineers may get flustered and bothered by this approach because it is not a performance metric for

them. There has to be a retainer fee or a company account to which to charge the time they are going to have to spend to keep up with this stuff. Managers need to have this sort of contractual detail addressed before this objection comes up.

One way to deal with this problem, instead of contracting a firm to do this, is to hire control engineers and make them responsible for maintaining the infrastructure in conjunction with IT security. Note that this team of engineers and IT security could work under any of three major divisions: operations, engineering, or IT. It should be up to senior management to assess who has the staffing and budget to absorb these people and manage them in an appropriate manner.

Note for those who may be making this decision: Much has been written about this field for the chief information officer or chief security officer (CIO/CSO) executives. Sadly, too much of this advice has been conceived as if this was nothing but a gussied-up office system by those who have hardly even set foot on a working plant floor. The result is that many CIOs and CSOs carry some grave misconceptions over what a control system is or what it does. Do not automatically assume that a CIO or a CSO is appropriate for this task. Given this problem, another tactic is to simply acknowledge that this is an amalgamation of these three fields and to make the control systems security group independent of everyone else.

The third reaction identifies that the effort is an open-ended endeavor. Where do we stop? How do we set goals? The answer is that we as a society do not stop, but that we aim for the easy stuff first, and steadily improve from there. This is going to be a continuous process. We need to set priorities to handle the current system and figure out better designs for future systems. This may require depreciating existing assets faster than expected, and establishing different criteria for depreciation.

Managers should take note of this, and be ready to task technical staff with identifying those assets and accounting for the changes as early as possible. It is also worth noting that such security awareness is actually systems integrity monitoring and that, as such, it may have a great deal of utility for improving overall availability.

Note to those with high expectations: We must all learn to crawl before we walk. It is almost never prudent to impose full military-grade security on an existing control system overnight, no matter what fears the IT security people may have. It is dangerous, because there can be some side effects that may get in the way of critical or safety processes. Managers will encounter resistance if they push too hard. Following the inevitable accident, there will probably be testimony from license or certificate holders that these methods were not properly vetted before deployment.

To avoid this situation, ask, but do not push for better security. If there are significant objections or resistance from the people who hold licenses and certificates, particularly when the processes involve safety systems, take the time to discuss goals, methods, and timelines. These are the judgment calls we pay managers to make. It is imperative that all risks are laid on the table and discussed openly and honestly among all involved, and that the decision reasoning and outcomes are carefully documented for future reference.

The fourth reaction may be stated thus: "Well, if the Internet or remote access is bad, we'll stay away from it. Let's isolate, and all will be well." The problem with this attitude is that it will not stop malware on a flash drive or a contractor's laptop. It will not stop software logic bombs from those holding the control system hostage. More has to be done than simply isolating the networks. In any case, reporting requirements—although most are pretty minimal—are growing all the time. Engineers need to find ways to maintain some control even during periods of degraded security. This may include degraded performance strategies that do not rely on interconnections with other systems.

The fifth reaction may be stated as: "Where are the standards?" This is a good question, except that the standards are still very much a work in progress. We are going to have to forge ahead and help write better standards based on field experience. Right now, that field experience is mostly unreported or even hushed up. Many standards are underdeveloped because there is little experience to use to develop a sense of what good practice is.

It is difficult to gather field data on security systems, because there are sound reasons for not discussing incidents and accidents caused by this sort of thing. Until some sort of indemnity and limited liability is offered in return for making such reports, there is every reason to be concerned about potential lawsuits. There is a strong need for an anonymous reporting system so that everyone can learn from each other's mistakes. Defining and gathering this data is going to be one of the first tasks of the three-sided team of engineers, IT security, and operations.

Information technology perspectives and their reactions

On the other side, we have the offensive from an IT security researcher. Researchers often lack a familiarity with what they are attacking. Nevertheless, they are very good at it. Before getting started, IT security must be told, with strict authority, that the operators are ultimately responsible for everything that is officially in production. No potentially disruptive tests should be done without operations staff being aware of what is going on. There may be instances where life and limb are at stake. This is not just another office application. The product is real, and a backup cannot restore defective product.

The first reaction is: "You are relying on obscurity to protect this? There is no security through obscurity." This is true, mostly in very public arenas such as the Internet. However, in practice, there are thousands of points of data, with little understanding of the process at hand, and the automation systems that will protect key elements of the process. Real destruction (something that goes significantly beyond the nuisance level) will require subtlety. To get there, one will need specific knowledge of exquisite detail that very few besides another engineer would know. Turning things on and off rapidly may make a significant mess and trigger some downtime, but it usually does not cause a process to collapse catastrophically.

Security theory assumes information transfers without any sort of friction. That is not exactly true. While data can move that fast, the context and education to use that information do not convey so easily. The reality is that while obscurity is not security, it does represent a significant obstacle that may tip priorities from one aspect to another.

Thus, although an exposed HMI interface having an obscure backdoor password is a bad thing, a dial-up modem with access to a MODBUS interface with a remote terminal unit (RTU) may not be the worst thing in the world. The latter requires some understanding of what is present at the site to cause a problem. The former is much easier to abuse, because it includes metadata about what the site controls.

The second reaction is "What do you mean, I can't run a port scanner at full speed? An attacker would do that. This is really fragile stuff!" The answer is yes, this is all quite true, but there are some implicit assumptions here that they have not encountered before. This is where the concept of a real-time system and a near-real-time system needs to be explained.

Engineers know (or have some idea of) estimates of how much traffic should be on an industrial network. Process controllers are designed to go into a fault mode if they cannot see their remote input/output (I/O) within a very short period of time measured in tens of milliseconds. In an office, such delays might mean that a web page would take an extra

few seconds to paint. Life goes on. For industrial controllers, however, this is cause for a fault condition. This is a design feature, not a failure.

The plant floor has advantages that offices do not have: First, it is possible to baseline the appropriate traffic levels and set alarms to show if there is too little or too much traffic to some surprisingly narrow margins. Second, the processes can be coordinated so that they do something sensible when too much traffic is encountered. This will require working in coordination with the engineers. When new systems are built, they will always be vulnerable to a denial of service attack, but with judicious network design and careful limits of scope, this should be an unlikely occurrence. Some designs have already planned for this problem because the engineers may know that network traffic capacity is tight.

It would be prudent to review this situation with the engineering staff to find out what is already in place and to integrate some form of operator alarms to handle this class of problem. New designs should have improved fallback control schemes to handle a saturated network on a programmable logic controller/distributed control system (PLC/DCS) or a supervisory control and data acquisition (SCADA) system. IT security will need to work with the engineering team to identify the risks and to help develop strategies to deal with this problem.

It may not be practical to remove denial of service attacks against control systems, but it is possible to detect the problem and limit the damage.

The third reaction is "Centralize all security into one great big glass room/box/network switch for ease of monitoring." While it is indeed convenient to bring security together into one room, this is the sort of policy that works better in an office than on the plant floor. In an office, if the central security services are not available, nothing happens. The bureaucracy stops. This is not a good thing—there will certainly be a loss of money—but it is unlikely that someone will lose life or limb as a result.

However, if the security server denies access to a controller, if a single switch is misconfigured with everything, the process will continue to do something; perhaps that something will be very undesirable or even deadly, but it will continue with or without the control system. Inertial energy, chemical energy, thermal energy, and so forth do not magically disperse when the control system fails. The security systems need to be as resilient as the rest of the control system process. The IT security people will need to find ways to distribute security in a safe and resilient manner.

Managers need to make it abundantly clear that engineers work very hard to avoid single points of failure. After all that careful investment, there is not going to be one great big central thing that can fail at once and bring the whole operation to its knees. This is particularly true for license and key servers. The security systems will need to be distributed throughout the plant or SCADA system.

The fourth reaction is "We must push patches; there is no time to review anything." Once again, not so fast. Engineers, contractors, and senior operators tested things very carefully before turning them over to an end user; pushing a patch is indeed a very dangerous thing to do. Processes are typically broken up into parallel pieces. If possible, a patch will be deployed to a parallel segment of a process to evaluate it for stability, performance, and interoperability. If parallel segments are not available, then one of two common operations are possible: First, keep extra operators on-site to run things manually in case the update goes horribly wrong, or wait until a parallel segment is available, or until conditions are light enough that the infrastructure can afford to take a chance in case things go very badly.

Such conflagrations do not happen very often, but when they do, things can get ugly very quickly. Make sure the IT security people know that they are going to be given training

so that they can help out with this effort and lend a helping hand in case a process goes awry. Note to managers: Care and ownership of one's actions is improved a great deal when staff has to not only admit to their misdeeds but also clean them up as well. The cost of training them with all the safety and process narratives will be greatly repaid in job performance.

The issue can be summarized by saying that patches should be pulled (by an operator and possibly others), not pushed, through the automation networks. This issue will become less of a problem as the development cycle for control systems focuses toward a more continuous, less disruptive, less project-oriented management.

That said, a policy where operations and engineering do not patch at all is unacceptable. Patching will improve the performance and life cycle of all parts of the control system. Evaluation of each patch release is something that everyone should be part of.

The fifth reaction is "Use strong passwords and authenticate everything." Few will argue with the authentication aspect, but strong passwords are often forgotten under stress. Use other methods for identity validation: biometrics or card/radio frequency identification (RFID) access (something you have/something you are [made of]). Passwords, if used, must remain very simple and easy to remember under stress. This limits their utility for obvious reasons. Locking people out in high-stress situations is a recipe for disaster, and besides, it is a security risk all by itself.

The sixth reaction is "The protocol is insecure by design." You can start and stop a controller with just one packet! We have got to fix this stuff! The answer is that protocols such as MODBUS, DF1, Profinet, or Common Industrial Protocol (CIP) were never designed to be exposed to untrusted or public networks. This is where we will need the expertise of the IT security specialists to help document the network topology, and set up virtual private networks (VPNs) where there is no other way to get the data from one place to another and back.

Eventually, some day, standards committees may include authentication in these protocols, but few are there now, and it takes time to do this correctly. The author knows this firsthand, from having seen the deliberations over the years that it took to develop a secure authentication feature set for the DNP3 (IEEE-1815).

The old joke about the civil engineer and the soldier rings true here: Engineers are paid to build things; soldiers are paid to destroy them. Similarly, engineers are paid to make things work; IT security researchers are paid to break things. Teaching them to chase a single goal with the same equipment is not easy. It is imperative that everyone focuses on the goal of making the system work more reliably. The security researchers need to recognize that their part of the equation is simply part of the whole control systems endeavor: making things more durable and reliable so that the system works better under adverse conditions. Engineers need to realize that the IT security researchers are not the enemy. By focusing everyone on the ultimate goal of better resiliency and reliability, we all win.

Finally, when these two groups understand each other, they will need to promulgate some actual user interfaces that the operations people can act on.

Operations perspectives and their reactions

Operators seek consistency. Usually, they do not like changing how things are done. With change, there will be complaints.

The first reaction from operations is that they probably had some very nice remote access in the past. Why should they not have access to their plant from the World Wide Web? It will be up to IT security, engineering, and management to decide how to make this work securely. One point worth making is that even if everything works in a perfectly

secure manner (unlikely, but consider this for the sake of argument), we still do not know if the system is being accessed by the employee or perhaps a vindictive child or spouse, that the employee is not drunk or high, or that someone is not holding a family member hostage to force the issue.

When people have to be on-site to issue controls, one can use physical security to augment the other security features. Remote access defeats that layer of security. The operations staff needs to understand that something is needed to replace that implicit layer of security.

The second reaction is "What does this mean? What do we do when this stuff barks at us?" The immediate need is to explain that if you get alarm X from system Y, you call person Z and say the following things to them. This is, basically, how to call for help. However, underneath it all, this is a very important concern. The alarms and the systems designed by engineers and augmented by IT security will not be used by either of them. Real security begins on the front lines with the foot soldiers: operations. It is imperative that they understand what the new security features are, why they are needed, and what they can do for them. There is useful diagnostic and alert information embedded in those alarms that can improve recovery time from a bad situation.

Furthermore, this can be used to track when employees or contractors are jacked into the network. If the operations people were not notified, they have grounds for taking action against those who are not coordinating with them.

The third reaction is "What is this Big Brother stuff? I don't want my name on this stuff!" This comes out of an abundant distrust of the automation systems. Some of these very concerns were expressed when flight data recorders were first introduced to the airline industry.

The first issue is how the data will be used. Managers will need to be ready with policies that the operations staff will find reasonable. Nobody wants to be rated by the machines they work on. A reasonable compromise would be to use the data to improve training, for forensic purposes after an incident, and for preventing unauthorized intrusions.

An interesting side issue may arise when using biometrics such as fingerprint readers. This is where the IT security staff should explain the basics of what a hash function is, and how passwords and other access information are hashed before it is stored in the computer. This way, even if the hashed information is revealed, no one is likely to reconstruct the original fingerprint, retinal scan, or whatever token was used to access the data.

The second issue is one of job performance. It would be a mistake to think that a control system could tell you who is good at doing what. That is like having the autopilot rate the pilot. Management can use these systems to figure out who has done what, but they should not use it in any way for performance reviews. This point needs to be brought home to the operations staff.

The fourth reaction is "Why should we care how well this stuff works? If it breaks, we'll run things manually." The problem here is that, like modern airliners, the performance requirements are such that running things by hand for extended periods of time is no longer particularly safe or practical. Does anyone have an attention span good enough to keep a large furnace running properly by continually monitoring and adjusting the heat output, the air intake, and the fuel intake? We use automation because it is not financially feasible to staff places with lots of people to run things manually hour after hour, day after day.

Ultimately, as we become more reliant on the control system, we need to know how well the control system is doing its job. We need to know how healthy it is. And, if something is amiss, if a baseline of performance has changed, operators (and the IT security

and the engineering staff) need to know. In other words, we need the operators to evaluate the control system continuously.

The fifth reaction is "What do you mean, we need to keep track of the contractors?" If they're incompetent we dismiss them! This flies in the face of reality. Contractors, or even company visitors, can leave all sorts of malware or back doors behind without even realizing they have done it. The people most likely to stumble across such anomalies are the operators themselves. IT security and engineering staff need to give the operators tools to track and hold staff accountable for what is left behind because they are the ones who will need to know what happened, and who to call to fix things.

The sixth reaction is one of resigned defiance: "Do what you must, but keep it out of our way, and don't get in the way of profitability." This is the most important point of all. This is often lost on everyone but the operators; the reason control systems exist is to improve quality, capacity, reliability, and availability. Whatever it does, a security system should not get in the way of these goals.

In other words, while security is important, it is no less important than the reliable and safe production of an inexpensive product on time. The purpose of security is to ensure that this can continue. As such, one point to make is that security systems can improve awareness of what is going on with the plant and its control system.

This is a primary selling point for SCADA and control systems' security features: self-integrity monitoring. The more we know about how well the control system is working, the better our processes can be controlled, and the more reliable our operation will be.

But, beyond that, there are some common issues of how to achieve that goal.

Penetration testing

If you do not attempt to penetrate the defenses, you will simply have to take the attestations of others that it will perform adequately when the time comes. Manufacturers can claim all sorts of things, but only by actually hiring someone to penetrate a system or product can you actually know where software flaws and other issues may be a problem.

That said, many IT security people prefer to perform penetration testing against real live systems, on the theory that this is the best way to find out at full scale whether the security system performs as designed. This can work in an office, where data can be backed up or restored in a jiffy. However, in a control system, there will be real product on the floor with real consequences. The machines may really come apart from a successful attack. Nobody really wants this to happen.

Just as we take samples of concrete and test them for strength during construction, we can test the individual pieces of a control system in a lab. Not surprisingly, many larger companies have such test labs, if for no other reason than to test integration of newer products on older systems. These labs could pull spares from stock and test them with the original running firmware against various security attacks.

Penetration testing can be a frightening, eye-opening experience. The author has personally observed a test where a safety integrity level (SIL) rated controller was attacked and frozen in its current state with a primitive local area network denial (LAND)* attack. Although a private security researcher may not get much traction with an original equipment manufacturer (OEM), the customers of that OEM usually do. The alliance between

* A LAND attack is best described as a denial of service. The attack consists of a TCP/IP packet with both the source and destination addresses of an SYN packet set to the victim's address. Unless the victim's software is able to recognize this attack, it will reply to itself endlessly. It was first reported on November 20, 1997.

customer and security researcher is thin at the moment, but it has every reason to grow and prosper in much the same way that insurance companies evaluate how crashworthy a vehicle is by actually purchasing one and destroying it.

Penetration testing also depends on how well chosen the access methods are and how easily they can be cracked. In the case of a certificate authority (CA) server, it has to be properly configured with up-to-date software that cannot be easily corrupted. As long as there is a backup CA server, it should prove fruitful to attack one to see what expectations an end user can have of it.

An alternative to attacking live equipment is to try out an attack on a virtualized platform of some sort. This is a brand-new approach that has not received much attention until now, because of issues regarding time of day accuracy in the guest operating system. However, even if the original software is working on real hardware platforms, one can still test the entire system on a virtualized platform in a private LAN.

These results should be shared with care. Above all, they need to be reported to a computer emergency response team (CERT)* agency and kept confidential, not only for the duration it takes to effect a patch but also for a certain time thereafter, to give the end-user community time to patch the most critical parts of their systems.

Network mapping and scanning

In and of itself, tools such as NMAP,† used for scanning and discovering network nodes and open ports, are not bad. However, the commonplace defaults for such tools are toxic for a control system or SCADA network. It is not uncommon for older equipment to be running with 10 Mbit half-duplexed hardware, and for that equipment to seize up in the presence of more than 3 Mbps of traffic. Recall that in the earlier days of networking, it was more commonplace to use a hub instead of a switch and that, because collisions were repeated to all ports on the hub, it was expected that networks would be incapable of more traffic than 30% of 10 Mbps or 3 Mbps.

Thus, when these devices were exposed to full duplex switches that could spew a sustained 10 Mbps of traffic, the equipment would often go catatonic or worse, even overwriting parts of their flash memory. There are documented cases where a nuclear power plant (Browns Ferry Unit 3) had to SCRAM‡ the reactor because they lost control of the cooling water pumps. The problem was believed to be someone accidentally inserting the wrong cable in a switch. This caused a significant broadcast storm to be propagated toward both 10 Mbps interfaces that happened to be the motor controls for the cooling water pumps.

* CERT agencies may go by different names in different countries, but the ultimate purpose is pretty universal: They are agencies that track computer problems and assist with negotiating a well-known outcome with the manufacturer. At some point, they will publish the links to the fix. This is very helpful to those with software and firmware from many vendors who seek one source for easy resolution and tracking of outstanding problems. Typically, CERT agencies are supposed to share information with each other, although some may have an easier time dealing with their domestic software firms than others.

† The NMAP tool is a program designed to scan a series of IP addresses or port numbers to see what responds. This tool is very useful to confirm that only the appropriate services at a network address are online or that no extraneous services are enabled. It is also useful for discovering hidden or forgotten addresses on a network.

‡ The acronym SCRAM has traditionally been used to refer to an immediate, emergency shutdown of a nuclear reactor. Though it is unknown what the acronym actually means, it has been used to describe a sudden and abrupt halt or shutdown of any given critical operation, and not necessarily associative with nuclear power generation.

The astute reader may be wondering why this older equipment has not been updated yet. The problem is that it is often embedded in large, expensive, and critical pieces of equipment. One does not just replace the interface of such equipment without a significant engineering and recertification effort. The network interface may have been state of the art when it was designed. Unfortunately, such equipment is purchased and financed with the expectation that it will last for 20 years or more.

A careful scan of the network (eliminating port scans in sensitive areas) would be educational. Also note that default speeds for port scanning are set with typical office computing platforms in mind. Usually, there are software switches that can slow down the scan to something that can reasonably coexist with the rest of the control system. The IT security and engineering staff will have to establish guidelines for where, when, and how often such scans should be done.

Nevertheless, these scans are invaluable. Often, old network equipment thought to be removed is still online. Scanning will find it. Sometimes one can find network ports open to control equipment that nobody has documented. This is where it is wise to scan a few spares and then make some inquiries to the OEM.

The more manufacturers that hear this sort of thing, the less likely they will be to think that they can "hide" a back door in a product simply by not documenting a port number.

Some features include web servers that were either not turned off or were poorly documented in the first place. It is not uncommon for plants to receive entire skids of equipment containing an embedded PLC with metered pumps. The PLC's primary interface may be known, but there may be others that are not. Those interfaces can be used for attack.

Traffic monitoring

It is common practice in the office world to use smart switches that can be queried to obtain statistics on how much traffic is coming from what port and can segment traffic in two groups of virtual LANs (VLANs) so that broadcast traffic does not go everywhere. It has done wonders for office computing performance and it can do the same for a working control system. However, there are some features that should be used with care.

First, because this is a switch, not a hub, one does not hear all the traffic all the time. One only hears traffic addressed to that specific port. A broadcast or multicast packet or an address with the Internet protocol (IP) address of something on that port is the only traffic to be expected.

It is commonplace for security staff to monitor traffic from various ports and VLANs. However, one must ensure that the switch backplane speeds and port speeds are up to the task. In an office, one would not usually notice a slightly slower web browser or a slower database response caused by network congestion, but on a busy control system, it would be noticed.

Second, while intrusion detection tools for Nessus and other open-source packages are available, they still are not as familiar with commonplace industrial protocols. Furthermore, not everything runs on Ethernet media. There are still RS-485 serial networks, long-distance twinaxial networks, and many more unique interfaces, such as HART.* It is important that such networks be identified, documented, and reviewed regularly, because the intrusion detection tools are simply not available for these interfaces.

* For information on the HART protocol, see http://www.hartcomm.org.

Who are the threats?

Most security people like to discuss the infamous man-in-the-middle (MITM) attacks because they are impersonal, or an evil hacker lurking in a basement somewhere. This is an easy sell because we have all imagined sociopaths like this before. And, although they do exist, they are comparatively rare.

A variant of this popular theme are the nation-state actors. The infamous Stuxnet malware was probably developed by a nation-state with resources. The only thing worth mentioning about nation-state threats is that if the control system is too difficult to act on, there are usually other methods. Someone with a decent hunting rifle could do significant damage to a substation before anyone could respond. The old joke about running from a bear applies; you do not need to run faster than the bear, you only need to run faster than your fellow campers. Likewise, if physical security and background checks of contractors and personnel are not maintained, having super-high-security cyberassets are not going to make much difference. In other words, to defend against nation-state actors, you need all security to be up to that level, not just the "cyber" part.

This brings us to the most common and the most insidious actors: insiders. There is a saying in the business—the most dangerous people on an industrial site are usually standing right next to you every day. While we commonly invoke an "evil" third party as the rationale for installing security, the most numerous and dangerous threats are actually the employees themselves.

Imagine a contentious situation regarding a union, and negotiations are not going well. Would it be outrageous for someone to have an "accident" which would cause significant damage and financially force the issue with the company executives? How would you stop a situation like this?

Imagine a contractor who thinks he was cheated on his last job with this customer. He installs a logic bomb in the controller code he wrote. How would you stop a situation like this?

Imagine a sociopath with a need to prove himself. He sets up a dangerous situation and then shows everyone how he "saved the day"—only, it does not go so well.

The reason why employees and contractors are so dangerous is because they know the process intimately and think they can weasel their way around the process. A hacker living in his parents' basement might not know what to do with an old dial-in modem used for a MODBUS connection to a PLC in the field. But these people just might.

It is imperative that someone develop extensive code review and storage systems for the PLC equipment in every control system. It is also useful that there be more than one system available to download and upload code from a controller. The reason for this became apparent with the infamous Stuxnet malware attack. The application environment was attacked in such a way that it would silently insert extra code into a controller. Since that code was both downloaded and uploaded from the same development work stations, nobody would have a chance to notice the extra software this malware inserted. Source code control systems (SCCSs) can mitigate this problem.

Engineers, particularly those who integrate embedded devices for control systems, like to think in terms of a project-oriented approach. They tend not to think of the whole life cycle of the software. The long-term value of an SCCS for configuration data is often lost on them. The IT departments, on the other hand, tend to get very bureaucratic with the SCCS and its features, requiring extensive training and complex models to manage software versions.

Somewhere between these two extremes is a happy medium. Someone who inserts a logic bomb in an embedded device can be discovered through review of the SCCS. Patches can be reviewed very easily with the aid of an SCCS to show all of the configurations that a patch is likely to face in the field. The ultimate goal for an SCCS is to have a clear, unambiguous record of what is supposed to be in the control system embedded devices.

Summary

Control systems security is not simple, nor is it easy. This chapter represents distilled experience of having dealt with the mindsets that various professions bring to the fore. Many behaviors are defensive and bureaucratic. We cannot afford knee-jerk reactions to these perceived threats. Management planning is key to bringing these professions together in a productive manner. Those who throw people into a meeting room with no guidance have no reason to expect good outcomes any time soon.

chapter three

Threat vectors

Jim Butterworth

Contents

Cyberspace operations

Cyberspace consists of many different nodes and networks. Although not all nodes and networks are globally connected or accessible, cyberspace itself continues to become increasingly interconnected and warehoused in the cloud. Computer networks make possible geographic travel, although electronically, at the speed of light, able to circle the globe in milliseconds.

We can isolate our networks using protocols, firewalls, encryption, and physical air gaps between network segments; however, the very purpose of the network is to interconnect; to accomplish efficiency, data sharing, and collaboration. Therein lies the challenge for a mature nation as they plan for sustainability to operate among the threat actors, fight through probes, reconnaissance, and successful incursions into their computer networks, computers, and data stores.

This chapter serves as a primer for building and maintaining a robust cyber operations capability that meets the growing threat to national networks, critical infrastructure, and a nation's most precious commodity … the information necessary for e-commerce, public service, finance, and defense. There is not a single industry that is not touched by cyberspace; therefore, it is incumbent on the stewards entrusted to protect it with vigilance, speed, and decisiveness.

Scoping threat vectors

The employment of cyber capabilities serves to enable, protect, and ensure continued operations in and through cyberspace. Such operations include computer network operations and activities to operate and defend a nation's interests globally. The types of people, process, and technology employed to attain these operations change at an alarming pace, as is required to remain in cadence with the myriad of threat actors placing you directly in their crosshairs.

The traditional military industrial complex philosophy of leveling the playing field does not apply in cyberspace, where but a few talented and determined foes can penetrate and wreak havoc on a company, a critical system, an intelligence agency, or even a government itself. Recent news stories highlight the anonymity that these threat actors can use to attain their goals, making the task of defending exponentially more difficult to achieve.

Globalization of the battlefield

IPv6 was driven out of necessity as the world simply ran out of addressable space. As global presence grew and nations moved their information online, seeing the benefit of an interconnected world, Internet assigned numbers authority (IANA) was forced to look into the sunset of IPv4 and devise a means to usher in a seamless means to remain connected.

Legacy network protocols, operating systems, applications, and equipment will remain connected, which is unavoidable. These older devices are reliant upon IPv4 to communicate, and are most likely incompatible with the IPv6 standard. While IPv6 has been available for several years, it has not gained wide acceptance by the networking community. A global consortium* recently announced their goal to accelerate the deployment of IPv6 at the Internet level by having several thousand Internet Service Providers, edge device manufacturers, and application developers to make IPv6 the default protocol, instead of relying on IPv4 as the default protocol.

The primary benefit to an IPv6 standard is the increased address space. Initial reports that IPv6 would usher in tighter security controls have proven false, with many reviewers reaching the conclusion that IPv4 with IPsec configured could be just as secure as the IPsec configuration within IPv6. Additionally, IPv6 traffic could be tunneled through an IPv4 message header, further solidifying IPv4's continued reliance.

If IPv6 eventually makes its way onto the world stage as the default protocol, legacy devices and applications will require modified sockets in order to communicate. If the operating system manufacturers have publicly stopped supporting aging operating systems, who then will be tasked with modifying the underlying network layer to ensure

* "Internet Society" and their test day entitled "World IPv6 Launch," which was initiated on June 6, 2012. Refer to http://en.wikipedia.org/wiki/World_IPv6_Day_and_World_IPv6_Launch_Day and http://www.internetsociety.org/ipv6/archive-2011-world-ipv6-day.

operability with IPv6, and who will conduct the code review to ensure there are no gaping holes or potential flaws that could grant unauthorized access?

Critical infrastructure protection and threat vectors

The lion's share of legacy networks exists in the industrial control systems (ICS) industry, largely due to the continued reliability and safety of these systems. The unintended consequence lies on our inability to patch, update, or conduct a technology refresh without the cooperation of vendors, service providers, and governmental agencies to ensure adequate funding exists, regulations and standards are put in place and enforced. Of paramount importance is that any infrastructure upgrades must be designed with security intrinsically baked into the ICS of tomorrow. In the United States, the Federal Energy Regulatory Commission (FERC) and North American Electric Reliability Corporation (NERC) have recently updated their CIP guidelines. In June 2011, the National Institute of Standards and Technology released Special Publication 800-82, Guide to Industrial Control Systems (ICS) Security. This is an example of where regulations and compliance are leading the development of advanced supervisory control and data acquisition (SCADA)/ICS technologies, such as Smart Grid.

Considerations must be made to not only design secure systems (programmable logic controllers, remote telemetry units, intelligent electronic devices) but also ensure the point-to-point communications protocols between them are not left to "off the shelf" distributions of Bluetooth, 802.x, infrared, or other network layer protocols. A determined foe will exhaust every possible avenue to gain entry, looking for devices that have embedded wireless wide area network (WWAN) antennas and processors, bridging wireless protocols with an external device designed to negotiate and proxy communications between these mediums, checking online repositories of exposed devices, the list of potential access points extends far beyond what is traditionally viewed as such. With just a bit of research and creativity an attacker can, with relatively low-tech and affordable modifications, decide to survey and lie in wait for the opportune moment to seize access to a system they can use as their base of operations against you. Cyberattack is designed to be clandestine and stealthy, and rest assured that future threats will rely upon bleeding edge exploit development, requiring defensive measures on par with the "art of the possible" to an attack enabler. A shining example of this are Stuxnet and Flame, both having been in clandestine operation for years without detection. Although the underlying payloads were designed for different purposes, Stuxnet, designed to induce uncontrollable failure in nuclear enrichment centrifuges, and Flame, designed to collect intelligence that would enable future operations. Presuming both of these payloads have been in operation for several years, it should make the reader curious about what undetected payload is currently operational and what its intended purpose is.

Computer network operations

How does a nation build and retain a talented and mature cyber workforce? It is this author's opinion that successful cyber operations are 65% human skill, operating 35% advanced technology solutions. Overreliance on automated detection, executive dashboards, and solutions that are only as efficient as yesterday's threat will certainly ensure continued vulnerability to the threat of tomorrow. Terms such as "advanced persistent threat" are good for categorizing a determined foe and make for good PowerPoint slides.

It misrepresents, however, the nature of the problem. Malicious code is a vehicle used to carry out computer network operations and is always designed by a human.

Automation in information processing enables vast amounts of computer instructions to be computed, culled, analyzed, and reported. The process, however, is wholly reliant on human interaction in order to program the algorithms that the process will use. This is an important consideration in that in all computing operations it takes human ingenuity to enable it. In computer network operations, it takes human skill to attack, exploit, and defend. Human knowledge that is aligned to a specific goal in mind, whether originating from nation-state efforts, privatized cyberterrorist groups, or random hobbyists using your network as their proving grounds. The end result is the same; unwelcome access, influence, and the ability to potentially cripple operations.

Computer network operations: Defend

Defense is more than collecting and aggregating the infinite alerts and events that automated sensors generate. Proper defense is not about keeping the adversary out; rather, it is about being able to successfully sustain critical operational functions while running in a degraded status. Stoic watch floors full of monitors and dashboards, "alive" and displaying the health of a network make for fantastic visions of advanced operations yet can convey a false sense of security. Their implementation oftentimes falls short of being able to detect, dynamically adjust, and provide real-time access to the information and access necessary to fend off or fight through an ongoing attack. Look for vendors and providers that are willing to open application programming interfaces (APIs) to share information and alerts in near real time, so that your frontline defenders can close the time gap from detection to subsequent action.

Computer network operations: Exploit

The art of digital exploitation can take either passive or active forms. Human involvement in cyberspace will leave traces. Despite the growing use of applications designed to provide anonymity such as virtual private server (VPS) networks, proxy servers, and bulletproof noncompliant servers located around the globe, they introduce a diplomatic and legal challenge the likes that will not be addressed or solved any time soon. National legislation takes years to adopt, and international treaties take decades to reach, leaving the defense of cyberspace to the owners of the systems and network themselves, employing the knowledge and expertise resident within their own teams.

The exploit operations gained from exhaustive and thorough digital analysis of discovered malware, internal characteristics of code structures, behavioral analysis, and the digital footprints in the sand left on an exploited host pay tremendous dividends in getting you closer to solving the person behind the keyboard problem. Who is your attacker? What is their motive? What is their technological capability? Can you maneuver within their attack cycle to mitigate the impact and sustain operations? Is the attacker using deceptive techniques themselves, such as planting flags, to throw you off in another direction? Cyberwarfare is similar to asymmetric warfare where a force of unequal size and firepower can successfully engage in conflict with a superior force. A control system engineer's responsibility is the daily care and feeding of the process under their charge, not to conduct cyber or asymmetric warfare with an intruder. Furthermore, engaging in tactics to disrupt the adversary on anything except an owner's systems could be construed as offensive in nature and subject the defender to legal action. Asymmetric warfare calls for

an equal application of unconventional measures to equalize and tip the scales in your favor, if not tip completely knock the scale off the hinges. The defender's inability to take decisive measures gives the edge to the attacker. If analysis revealed the public location of the attacker's pass through server, it is highly probable that the server is the property of an unwitting party and any attempt to access would be unlawful. The attacker is keenly aware of the legal framework and privacy laws in the United States and routinely operates both domestic and international points of presence in order to exacerbate and cross the jurisdictions of investigating agencies.

To successfully defend, you must learn as much as possible about your primary adversary and threat against your interests. Simply penetration testing public-facing sites to find potential entry points does not yield enough information about your adversary's weaknesses. You must employ human skill and expertise; dare I say the "art of human hacking?" Behind every virus, Trojan, worm, remote access Trojan (RAT), botnet, dropper, or exploit payload is the person who built it. They are responsible, and the human psyche is far easier to exploit and manipulate than thousands of lines of evanescent code in memory, designed to operate from a segment of memory that is configured at runtime as a temporary clipboard that will never cache its contents to disk. The growing talent of open-source intelligence collection yields a tremendous amount of valuable information; however, there is no business argument that makes a person of this skillset valuable, save the information they can provide to security teams.

Computer network operations: Attack

The single-most important element of these operations is nonattribution. As outlined in the earlier section, even your "developers" may tend to reuse structures and routines in their custom efforts. All too often, these highly specialized groups of experts live within a black world, keeping their operations tightly locked down in the interest of nonattribution. To introduce a paradigm shift from this approach, imagine if skilled exploit/defense analysts were able to "have a go" at the result of a payload. This is similar to war-gaming exercises, where military forces play out their continuity of operations plans and adapt according to the environment, and unforeseen circumstances. It affords them an opportunity to hone their craft before they need to use it. Code reuse in malware is common due to its modular construction and reliance upon the x86 architecture. Application exploit development is reliant upon specific memory offsets of an application given a specific patch level. Once an application is patched, the memory allocation of the vulnerable point may change, rendering the exploit inoperative. This is not the same as the payload that is delivered and installed following a successful exploit. The exploit is designed to enable access, while the payload is designed to retain access. What the analyst would expect to see will differ depending on what class of malware it is. Getting back to the human in the loop, the malware coders are not waking up every day designing new innovative ways to exploit the x86 architecture. Once an operational payload is designed, they will continue to repurpose to functional blocks within other payloads. Collecting digital intelligence on the code assembly and structures can reveal patterns that can be used to identify and correlate other processes with these functions built in.

Digital intelligence

Digital information takes many forms, depending on their medium and placement within the OSI model. This could vary widely from standard radio frequency transmissions used

in computing like Bluetooth/Wi-Fi, it could be the cellular networks we are continuing to increase our bandwidth and hence computing mobility atop. To be proficient at analyzing the many artifacts that fit the category of digital intelligence, an analyst must be adept at Unicode, Code Pages of many languages, file compression techniques, encryption schemes, hexadecimal encoding, byte offsets, file signatures, code bit shifting, identify the list of file formats, file and byte offset math, communication and messaging encapsulation protocols, keying and encryption algorithms, and many more—the requirements are staggering.

Types of sources of digital intelligence

We deal in both static and dynamic computing environments, composed of petabytes of stored files from standard computing assets and users, all the while expecting to be able to detect and handle any alert that triggers a threshold. Different uses and gems are derived from the many differing types of data. Are you dealing with memory resident malware that is designed to never write to disk? To ensure evanescent memory code is properly preserved, the responder needs to ensure that their memory-imaging tool is able to preserve the entire memory space, including the kernel-protected area. Failure to do so will result in smear, where recompiling memory introduces ghosts where instructions pointing to specific memory locations no longer exist, rather have been allocated and are in use by another process. Once the plug is pulled, the traces of the code disappear when the 5Vdc is removed from the memory chips.

Methods and procedures

How you gain access to intelligence is as varied as the types of digital intelligence that exist and equal in scope to the medium being chosen. RF exploitation requires advanced receiver technology. To secure digital communications at all points between transmission and reception, system designers will use techniques such as spread spectrum, encryption standards that use a combination of key-based or time-based authentication, compression or obfuscation of the data stream, and even point-to-point tunnels that use a master certificate authority to remain in sync. In the case of malicious code, in an effort to thwart reverse engineering of their code, authors will use packing schemes that obfuscate the contents of their code at rest. Oftentimes, these packers use a salt or some other form of bit shifting in order to scramble the data stream. Decryption of proprietary packers and encryption algorithms requires hefty computing resources, best adopted in a parallel computing structure for expedient results. As stated earlier, if a malware specimen is going to execute its payload, it will have to unpack itself into normal programming language. This is the point where the code is at its most vulnerable. Many analysts rely upon static code review of a binary or executable exported off of a system. The most accurate and telling time to analyze, however, is on the infected machine, as the payload is already resident.

We tend to traditionally view collection of digital intelligence as a row of lab computers, connected to source and destination hard drives, imaging the cell phones, video cameras, removable drives, CD/DVDs, hard disks, etc., that are all part of an intelligence effort. This will never be replaced, and analytic process advancements are being developed and fielded by vendors to assist the investigators in ascertaining the raw intelligence in a smooth process, in a fraction of the time. Using multiprocessors and multithreading of computing resources makes this possible.

Methods and procedures: Collection

When you do undergo collection operations, ensure that your process is commensurate your end goal. Clandestine or black bag collections require far more consideration than fear of being detected by your target. Oftentimes there are electronic, physical, and human interaction aspects to these types of operations. "Smash and grabs," concealed as a traditional crime of thievery, gains you the hard evidence. Passive taps, snarfing the airwaves, there are many creative and successful methods to collect intelligence. I would submit that the easiest method is directed against the human target, which as our own analysis of internal intrusions would prove time and again. The weak link and primary target in cyberattacks continues to be the end user. This is largely a result of the success the attacks have had when the end user is targeted as the attack vector. Exploits still require that they are executed in order to run, and one very effective method to accomplish this is to deceive a human operator.

Methods and procedures: Open source

Astroturfing is a phenomenon that has grown tremendously in the past few years. With the rise of WikiLeaks and groups such as Anonymous, LulzSec, and other organized #AntiSec movements, it is more important than ever to monitor these groups and be able to identify Astroturfing when it happens. This allows your organization to get ahead of the curve, plan your message accordingly, and handle any blow back from disinformation campaigns.

Methods and procedures: Deception

Pirate Pad, TOR, VPS, Proxy, Trac phones (amateur) ham radio, and personal management all have an inherent flaw. On the Internet, as much as they would like us all to believe, there is no such thing as true anonymity. A packet is structured and delivered, a fake e-mail account used to deliver a single message, has an originating IP that was used to sign up. It is a matter of putting talented open-source analysts at work, collecting as much information as they can about your threat. You're on their watch list, why shouldn't they be on yours?

Honeypot and honeynet technologies have their place in a defense-in-depth architecture. It is far easier to catch a bee with honey than it is with vinegar. In order for them to give the appearance that there is an entire infrastructure behind them, these technologies tend to rely upon virtualization, and modern malware is designed to recognize virtualization and either self-destruct, or will have built-in routines designed, upon detection, to invoke a harmless behavioral signature that will leave the sensors to weigh it as a benign low-level threat.

Computer incident response teams

Intrusions, sabotage, data theft, information exposure, and code manipulation will continue to occur in cyberspace. The geographically separated, yet electronically connected, world of cyberspace makes responding to these incidents, a sometimes-difficult task to achieve. Speed, mobility, and global omnipresence on our own heterogeneous networks require that we establish and maintain an infinite digital reach into our assets.

Field operations

There are times when response teams must deploy on-site, due to either an air-gapped network or as protective measures, such as creating isolated virtual local area networks (VLANs), are put in place to ensure the safety and operability of the rest of the infrastructure. Data on a network, unless specifically logged, do not remain for after action analysis. Data in memory are most certainly volatile, and as time passes the likelihood and possibility of operating system overwrites, fragmentation, or other computing actions introduces risk into the preservation process.

Development of flyaway kits, rapid response teams, forward operating or staging locations of equipment, or placing into the network/system administrators hands the tools, capability, and knowledge to preserve information rapidly, prior to taking protective and defensive measures. This statement presumes the incident will not cause further harm to personnel, endanger lives, or amount to a mission kill if you have to temporarily isolate or take down a system.

Understanding that TCP/IP is a connection-oriented protocol, once a computer network connection is terminated, or isolated from communicating, the connection will be torn down as a part of the protocol. This means that a response team may lose the ability to collect the volatile information on process connections, who and what is connecting inbound/outbound, and other information relating to an ongoing attack. In a control system environment, where there are as many measurement and test mnemonics as there are true control signals, the loss of any signal may cause a sensor placed as an interlock to invoke a safety circuit that prevents overload. Interfering with status signals can be as effective as interfering with the actual control signals themselves.

There has been decades of exposure within the IT industry to computer forensics and the necessity to preserve data using industry-accepted methods. Preservation is very critical for field operations, as it will take time for rapid response teams to deploy and arrive on-site.

Remote operations

Technology has also advanced to the point where it is completely feasible to conduct an entire investigation remotely. Software exists today that allows forensically sterile reach into your end points; to preserve and analyze data far faster than a response team can physically deploy on-site.

There is also a benefit to having these sensors and capabilities pre-deployed, in that your ability to seize on a critical alert, event, or other anomalous behavior can immediately be re-acted upon, thereby lowering your overall risk. Our assets will always be vulnerable. Determining the patch status of the operating systems across your enterprise is a necessary process in determining your vulnerability to the threats that are known today. It is called a zero day for a reason, and some of the nastiest exploits are yet to be discovered and are currently installed on many networks, around the globe, without regard for any specific industry.

Support to response teams

Incident response teams will need back-end support, either through passing back malware to specialized labs and expertise to conduct reverse engineering on a piece of suspiciously behaving process or driver, or providing remote access to the repository of evidence being

collected so that remote examiners or analysts can begin to operate in parallel, using distributed processing technology to cull through and extract the necessary information to respond to the threat.

Support efforts can best be thought and planned for as master-, journeyman-, and apprentice-level skillsets. Some of the more advanced cyber-elite skills require a few master-level experts. Incident response requires a journeyman who has a breadth and depth of knowledge of computer network topology, ports, and protocols, and a varied exposure to operating systems from a forensic perspective. Finally, apprentice-level skills could be considered as imaging teams, evidence custodians, incident yeoman, and analysts using automated processes and procedures to extract actionable intelligence and data from evidence repositories.

Malware and emerging threat actors

Recent highly publicized events have run the gamut from highly developed and sophisticated attacks to exploitation of embarrassingly basic lack of patching to attain breach success. Attack vectors range from application exploits, the tried and still true structured query language (SQL) injection, introducing logic flaws during code execution, bypassing internal authorization mechanisms, escalating privileges, or exploiting the end user to allow the attack to begin from within the house instead of going through the front door.

Malware: Delivery

All too often an incident responder will uncover during an investigation a rogue file or e-mail attachment. This is typically something a very adept journeyman can identify and recognize as a threat. However, what is oftentimes the case is that they have stumbled upon the delivery mechanism, or "the dropper," which is designed as a single-use bullet to make an outbound connection to a transient location somewhere on the Internet, controlled by your attacker. Upon successful exploit, the victim system/user's computer will make an outbound connection, shimmed either via DynDNS, DNS2TCP, or straight out SSL or HTTP, to download the actual payload necessary for the attacker to begin their operations.

Dropper analysis will usually yield where, by IP or URL, the payload was retrieved from, but a swift adversary will have ensured their own anonymity and survivability by using an unwitting public-facing exploited server as a temporary base of operations. They have thousands of exploited computers at the ready, enabling them to quickly shift the landscape and render your investigative efforts dead in the water. Once the delivery of the payload is successful, they will oftentimes discontinue the use of that exploited server, hedging their bets that your investigative team will be unable to gain access to it in order to conduct analysis. Both law enforcement and legal involvement take time, and it gives the attacker ample opportunity to change their modus operandi, erase their tracks, and carry on with the next phase … launching the payload and establishing a foothold in your network.

Delivery can be accomplished by a variety of means, many of which rely upon the deception of a human in the loop. USB drives dropped in a parking lot, or handed out at conferences; crafty e-mail attachments, social engineering a user in their private life on Facebook, LinkedIn, or some other social networking (SN) medium, with the expressed purpose of figuring out the means which will yield the highest likelihood of success. Unfortunately, it is my opinion that the user represents the greatest threat to our ability to

intercept and stop delivery. The user vulnerability reaches further than a lack of education. Although we can desire so, they are not expected to be the front line of defense against an attack. User education will stop some attacks, and when it does, the attacker will up the ante and begin to target our public-facing application and back-office developers, researching and singling them out as humans, knowing they contain the information required to do great harm.

Malware: Payload

The payload is the "sauce" that makes persistent access possible. They are usually stealthy in nature, deceptively designed to conceal their true purpose, hence making identification and eradication very difficult. Understand that the professional attacker is not going to rely upon the standard, already been analyzed and signatures written for, methods of retaining control over your machine(s). They adapt their tools and methods with target specificity in mind.

They can employ packing techniques, bit shifting of data at rest, obfuscation on the wire, hiding in plain sight, and a myriad of other deceptive and oftentimes troublesome tactics for our investigative teams. The "Holy Grail," however, is memory visibility and analysis in real time. Code must execute in memory, leaving the code itself exposed for our own analytic capabilities.

Memory detection and analysis is the digital battlefield of today and tomorrow. For a malicious piece of code to work, it must be running and to do so will occupy memory space. To occupy memory space is to interact with the host operating system kernel to achieve the desired outcome. There are only so many commands, structures, calls, routines, etc., that an operating system uses, and unless the malicious code has the ability to dynamically change the underlying kernel upon reboot, and there are instances of rootkits out there that can and have accomplished this in the real world, the fact remains that the malicious process itself is exposed when it is running in memory. It is also important to understand one last point with regard to malicious payload.

It can be designed as a single-use payload, designed to detect the presence of certain conditions and therefore launch; it can be designed to sleep and awaken at certain cycles; it can be designed to accept normal DNS query/response traffic to reconfigure itself. A payload can be a logic bomb, or a RAT designed to provide continued stealthy access into your network. Determining payload purpose is a master-level skill, and there are very few individuals that can accomplish this in support of a real-time investigation.

Malware: Command and control

Presuming the payload is designed for continued RAT access, the attacker must then establish a means to command and control the payload, all the while remaining undetected and nonattributable. Most control mechanisms of payload are noninteractive, meaning a command will be either sent or retrieved by the payload, and reconfigured on the fly to execute the revised operational request. The essence of command and control (C&C) is low and slow. One would tend to think that it would be beneficial for an attacker to configure their payload to operate during "non-peak" hours, to avoid detection. Yet what better way to conceal a single connectionless user datagram protocol (UDP) packet than to determine peak traffic times on your perimeter and configure the payload to sneak out a single, well-crafted domain name system (DNS) query? The attacker just needs to issue

single commands to the payload, which is automated to perform internal reconnaissance, collection of data, further penetration, privilege escalation, exploit du jour.

Threat trends

While it is commonly known that many nations have either expressed interest in or have already developed advanced cyber operations capabilities, the threat landscape is by no means limited to the adversarial nation-state attack. In many regards, a more serious threat is the rise of the #AntiSec movement, as their intention is public disclosure and media exposure. Astroturfing is not likely to subside any time soon, and it is a more likely scenario that due to the lack of law enforcement action, or legal implications to the perpetrators of recent highly publicized attacks, this underground movement will be viewed by many individual or splinter groups as an unregulated frontier to carry out their motives. As it stands today, they are largely correct in their assumptions that the international diplomatic community lacks the integrated and collaborative efforts to remove their cloak of anonymity and render swift justice in an unregulated and widely interpreted swath of "privacy rights," erring on the side of preserving an individual's right to privacy with regard to their activities on the Internet.

In the absence of global leadership and cooperation in this domain, an organization is essentially left to defend itself and take the necessary action to protect their assets. Participation in the Internet is voluntary and connecting a computer online, storing your data in the cloud, or otherwise taking advantage of the interconnected world we now live in is an essential way of life today, and it is prudent to remain vigilant and responsible for what an organization chooses to place or expose online.

chapter four

Risk management

Wayne Boone (revised by Allan McDougall)

Contents

What is risk?

Risk is an inherent part of business, and even life in general. It is something that we live with on a daily basis. That is because there will always be some form of obstacle or impediment that stands between us and achieving our objective(s). It can be as simple as traffic standing in the way of our crossing the street or as complex as working through interconnected regulatory requirements to succeed in international business. We also tend to respond to risk on a daily basis. We use resources such as people, time, consumables (gasoline, paper, food, water, electricity, etc.), buildings, equipment, information (including information systems), and processes or procedures to overcome obstacles (threats and vulnerabilities, as we will discuss) and reduce the potential for failure. We may make decisions to have other persons handle tasks to which they appear more suited. We may use different tools or better quality materials in our production. We may insist on more reliable information from those providing assessments. All this is to say is that we tend to take steps to avoid risk, to reduce its impact on our lives or to reduce its probability of

occurring. Since we cannot anticipate all impediments and make preparations to overcome them, there will always be some uncertainty that we will succeed. According to Cardenas et al. (2009, p. 1434), "obtaining perfect security is impossible." One might even argue, given the economic lessons over the last 10 years, that attempting to achieve perfect security can be disabling for a nation and its economy. For the purposes of this chapter, that uncertainty can be considered to be risk, and dealing with residual risk is risk management. "Protecting SCADA systems is a tricky task" (Gold 2008, p. 39) and requires as close to "100% proof against both modern and old security threats" (p. 40). Considering the environment in which supervisory control and data acquisition (SCADA) systems typically operate, mission success defined as service delivery according to mandates, regulations, policy, and, perhaps most importantly, user expectations would indicate that what is being "done" has a relatively high value and, therefore, there may be more uncertainties that could potentially impede success. The business of identifying these uncertainties or risks that can impact commodities or services supported by SCADA systems is an ongoing task. Personnel take steps to see that such risks are identified, analyzed, assessed, and treated in some manner to reduce them to a level that is acceptable to those senior management individuals accountable for service delivery. This cyclical process can be considered to be risk management. How an operation approaches the issue of risk management can be the determining factor between significant success and catastrophic failure. The challenge is that "risk" and "management" are both terms that are terribly overused in a number of contexts. This chapter will address the concept of "risk management" from an asset protection and security (AP&S)* perspective.

Objective of this chapter

The objective of this chapter is to explain the AP&S risk management process at a conceptual level as applied to SCADA systems and their supporting environments. The individual elements of risk management will be covered, including mission analysis (what business you are trying to do), scope (how much you are trying to do and in which environment), asset valuation (what useful or needed things that you will use to do something and what deliverables or results you are trying to produce or achieve), threat assessment (what or who are the "bad guys" who want to prevent you from doing what you want to do), vulnerability assessment (what are the "holes" or weaknesses in your assets that could let the bad guys in), risk analysis (how bad is it in general if the bad guys exploit the holes), and, finally, risk assessment (how bad is it *to us*) as it applies to risk management. Ultimately, the extent of risk management that is conducted is an expression of management's decision of how it wishes to address or treat identified risks. This chapter stops short of the development of specific security safeguards, controls, and countermeasures (which can be considered synonymous). As a caveat, this chapter is not meant to be a primer on AP&S risk management. There are several excellent books and articles

* AP&S is an inclusive term that has been coined in critical infrastructure protection (CIP) literature and is equally applicable in information system and corporate security environments. This term acknowledges that protection of assets is often inadequate, since this concept does not include assurance, continuity, and resilience in many people's lexicon. Also, security as a term often connotes the traditional security guard in a physical environment, another limiting concept. AP&S refers to all measures taken through the risk management life cycle, including mission analysis, asset valuation, threat assessment, vulnerability assessment, risk assessment, and, thereafter, safeguarding implementation to protect against, mitigate the effect of, deter, absorb, isolate, respond to, recover from, and restore all services and capabilities after an attack or major interruption to operations.

that focus on risk assessment for the practitioner, and the harmonized threat and risk assessment (HTRA) produced by the Royal Canadian Mounted Police (RCMP) and the Communications Security Establishment Canada (CSEC) provides tactical guidance for those who are required to conduct threat risk assessments (TRAs). While practitioners will enjoy reading this chapter as a refresher of basic principles, both they and line managers will benefit more from the conceptual treatment of this topic, along with some lateral thinking and application of the principles. In this manner, it is intended that practitioners will hone their analytical skills, and managers will better understand the significant level of effort and resources that go into establishing and maintaining an effective risk management program. The overall expectation is that they will collaborate in their mutual interest to protect valued assets supporting mission success.

AP&S risk in theory

Risk itself can be challenging to define, since perspective factors highly into how it is approached. In the financial community, for example, addressing risk can lead to both positive outcomes (profit) and negative outcomes (losses). In the AP&S context, the concept of risk generally refers to negative or undesirable outcomes, which must be addressed to ensure mission success. Generally, AP&S risk can be described in terms of the exposure of an organization to losses that result from a threat agent exploiting a vulnerability to cause injury to some asset. This is often expressed by the following expression:

$$R = f(M, AV, T, V)$$

where risk (R) is a function (f) of mission importance (M), asset values (AV), threats (T) in terms of their capability, opportunity, and intent (COI) (will be explained), and vulnerabilities (V). While not strictly mathematically sound, if the mission is more critical operationally, the threats are more dedicated, and/or the vulnerabilities (gaps or holes) are greater, then the risk is greater. Conversely, if the mission, commodity, or service provided is less important to clients, if the assets used are not valuable in terms of their according to availability, integrity, or confidentiality (AIC),*—in that order of importance according to Cardenas et al. (2009)—if no threat is inclined or able to attack, or if there are no gaps in the assets' protective posture, then, arguably, there is no risk. Any increase to one of these factors (without corresponding decreases in other categories) leads to an increase in risk and must be addressed.

AP&S risk management is not an exact science; rather, it is considered more of an art, because it is ultimately a qualitative process. Even supporting quantitative approaches (such as *annualized loss expectancy*) are based on a range of assumptions and subjective decisions rendered by people with varying amounts of AP&S training, education, experience, and critical logic. For example, it may be challenging to determine the full hard (financial) and softer (maintenance, performance, opportunity, etc.) costs associated with a valve that actuates as part of a pipeline. Does it include the replacement, installed, or

* In traditional AP&S parlance, confidentiality or protection from unauthorized disclosure of sensitive information or other assets is paramount, followed by integrity and availability. However, when discussing SCADA systems and national critical infrastructures (NCIs), availability is considered the most important security function, followed by integrity (protection from unauthorized modification), and then confidentiality; while it is important to protect the privacy of individuals and the sensitivity of information such as intellectual property operating data, and so on, this is less important than having services accessible on demand and of an assured quality.

initial price of the valve, or the prices associated with a component part, or its calibration, or its removal of service? What are the costs associated with not doing something else when working toward getting a valve up and running, which could include requirements analysis, approval, choice of product, procurement, shipping, and arranging installers who may have to learn about the product, with supervision of installation, quality assurance, and testing? Notwithstanding this complexity of determining hard and softer costs, a valve is relatively simple. Now, consider the value of a key operating official or the chief executive officer (CEO) of the company. That individual's value could be based on their salary dollars, the cost of hiring a new person, lost opportunity costs associated with going in a certain corporate strategic direction, or in the value accrued by the CEO's support for the AP&S risk management program (which would include the provision of capable staff and other resources). These examples indicate the overall qualitative nature of AP&S risk management, supported by some supporting quantitative risk assessments. Typically, discussions and decisions become more quantitative and fiscally oriented as one ascends the "corporate ladder" (what is the bottom line?) as busy executives discuss relative numbers. Unfortunately, when expressing AP&S risk, the best that can be presented is a relative assessment, such as that provided by a Likert scale of, for example, negligible, very low, low, moderate, high, very high.* In all cases, assessment criteria and assumptions for each scale must be very clearly defined and communicated to those who conduct the assessment and to those who receive the reports if the risk management advice is to be successfully communicated.

There is a tendency today, in the era of fiscal restraint, to have to show some measurable empirical value. One must be cautious with this approach, as there are a range of risks that are dealt with on a preventive basis that cannot be easily defined in this manner. Consider that many laws hold executives accountable with respect to whether or not their organization has taken all reasonable steps to prevent harm. What is the dollar value of taking steps to meet this accountability? That may be calculable. What is the dollar value of the possibility that senior executives may face incarceration for failing to maintain their duty of care? That may be more difficult to calculate—the only certainty is that the senior executives will certainly want a voice in that matter.

Generally, risks are defined in terms of the *likelihood* of a threat exploiting a vulnerability to impact negatively on the AIC of assets supporting business activities, production, or service delivery, and the resultant *impact* to the organization.† Lowrance (1976) uses the terms *probability* and *severity* in defining risk, and Cardenas et al. (2009) uses the terms *likelihood* and *consequences*, but these may all be considered synonymous with likelihood and impact. It is at this point that confusion may emerge with respect to the concept of risk. When considering likelihood, one is dealing with probability. Probability can be described in terms of the number of times a specific outcome or condition occurs given a total number of events. For example, flipping a two-sided coin leads to a probability of 50% as long as all the flips are random. Typically, deliberate attacks and accidents affecting entities supported by SCADA systems are not random, in that conditions must be in place for the attack or incident to occur; nor are natural events such hurricanes or floods completely unpredictable. Therefore, AP&S risk management is based on an accurate assessment of probabilities of negative events occurring, and taking appropriate mitigative action.

* A tip for providing more precise risk assessments is to use an even-numbered scale (typically four or six). This addresses the tendency to take the "safer" middle value instead of conducting more in-depth information gathering and analysis.

† As found in the *Protection of Assets Manual* Section 1.3.0 (ASIS International, n.d.).

Appropriate, in this case, refers to those measures that mitigate risk to a level acceptable to senior management and within the confines of what is considered to be legally acceptable.

A consideration here is that probability tends to be analyzed, assessed, and communicated in terms of simple individual risk events, without considering the effect of interrelated or aggregated outcomes on (potentially) complex systems. Consider weather events, and the concept of the 100 year storm. In many cases, people may look at the name and think that the storm need be considered in terms of a frequency of once per 100 years. There may be a tendency to discount this threat event thereafter, based only on history. However, with climate change, some areas have suffered a number of these 100 year storms over the past decade and there are new parameters defining what the 100 year storm may look like. This indicates that historical frequencies of threat events require continual reassessment for applicability in a certain industry, geographical location, or operating environment. From updated risk assessments may arise the requirement for changes to safeguards to ensure the AIC of valued assets supporting mission success.

The second consideration is how to describe the impact to the organization. This will be described further in the section "Scope of risk management," but it is important to understand that impact can be influenced by the perspective, location, and mission of those impacted. If you are driving a car that becomes involved in an accident, your impacts may be described in terms of health (you and those in the vehicle with you) and in terms of the costs associated with property damage. To the driver behind you who is caught in the traffic disruption, the impact may be more aptly measured in terms of the delays suffered waiting for the accident to be cleared and potentially lost earnings (such as could result from missing a meeting or a deadline to provide a service or product). Since time is an asset, it is being consumed without apparent return on investment.

Some aspects of impacts can be quantified; others can be assessed only qualitatively, and some others may be assessed as a hybrid of the two. Quantifiable impacts typically are more clearly measurable and demonstrable—as long as they can be assessed against an agreed scale or set of specifications according to a standard, "a set of useful metrics" (Zhu and Sastry 2010, p. 4)—if you will. Quantitative impacts utilize a specific number of units within that scale (e.g., dollars, number of products produced, or amount of service provided), which can be compared and, given the same conditions of a risk event, can be repeated. Other impacts are less quantifiable. Consider the loss of an employee in an accident. How does one measure the impact of such an event when the value of the asset is so difficult to quantify? It is certainly different if you are the parent or spouse of that individual, as opposed to a disinterested researcher or loss-prevention specialist analyzing the victim as part of a statistical group. How is the value and impact affected if the individual had a significant amount of corporate or technical memory that had not been written down? Outcomes of civil actions fall into qualitative impacts because of subjectivity and perspective applied to a factual event. Probability, in this case, is a result of precedent, common law, or a standardized means of calculating injury, which provides some degree of predictability.

What is certain in AP&S risk management is that risks are ultimately qualitative and must be acknowledged as such by both AP&S analysts and senior management. Many definitions, therefore, are not necessarily the most easily utilized. One of the clearest and most operationalized definitions within critical infrastructure protection (CIP) can be found in the Masters of Infrastructure Protection and International Security (MIPIS) program at Carleton University's AP&S risk management course—that the risk to an organization can be described in terms of a factor associated with "a threat agent exploiting a vulnerability to cause damage to an asset supporting a mission, resulting in some form of loss of AIC,

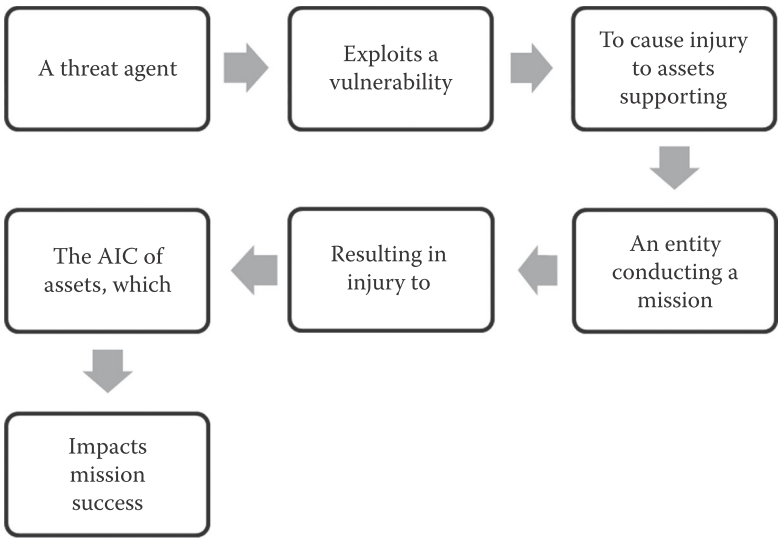

Figure 4.1 Description of AP&S risk broken down.

in turn resulting in operational impact to the mission." This structure of risk assessment fits into the concept of risk management well, in that it identifies and examines the major elements that lead to the losses to an organization. This is shown graphically in Figure 4.1. Note that each step can be isolated for analysis. More information on this will be presented later.

What does mission success mean?

Before one can answer this question, it is important to understand fully what is meant by mission. The mission of an organization is often simple to understand at the highest level; it may even be expressed as a motto on a poster or coffee cup, but such typically flowery and fluffy language may not define adequately what product, commodity, or service is provided, how much of it, how important it is to the community, region, or nation, and how reliably it is to be provided. To properly analyze the mission and draw salient conclusions for effective risk management, it must first be understood to the requisite level of detail. This is a matter of returning to first principles, and can be answered by two simple questions. The first is "why are we here?" and addresses the strategic level. The answer to the first question may be to provide a service (if part of a federal department) or it may be to generate wealth for the business owners in the production of commodities or products (if a privately owned enterprise). The motivation can be both monetary and more altruistic or patriotic, especially when considering those national critical infrastructures (NCIs)* supported by SCADA systems, for which the meeting of national objectives on security, sovereignty, economic prosperity, or health and safety may be their mandate. A follow-on question in this case may be "what do we do to help?"

The second question involves "How do we do that?" and exists at the operational level. The answer to this question describes the key activities or business lines of the enterprise. For a manufacturer, it may be to "deliver high-quality product X capable of meeting or exceeding the requirements of Specifications A–E for a specified period of time at a

* NCIs are those goods and services that have a very high AIC requirement based on their contribution to national objectives.

reasonable cost on client demand." From this mission statement, the various supporting, complementary, and interrelated activities within the organization can be identified, further decomposed, and analyzed at the tactical level. It is at this level that AP&S-relevant observations can be made and risk management-relevant conclusions be drawn.

The mission statement may be derived from the requirements of a parent organization, and may or may not be customized or interpreted for a subsidiary or regional facility. In those cases, the parent organization's mission statement is reviewed and the specific supporting business lines (operational) or functions (tactical) performed by the subsidiary organization are linked directly to the higher (strategic) mission statement. A critical path for expressing delivery mandates is thus formed.

Mission analysis

Once the mission statement has been captured and isolated, mission analysis can be undertaken. This is necessary to identify the indicators of mission success. Once again, this is a matter of asking simple questions and working toward detailed answers. Information to answer these questions is typically gleaned from reviewing business and AP&S documents, interviews, and site visits (observation). From the strategic mission statement, key business lines will emerge, such as those subordinate organizations in our example that prepare to build the product, fabricate the product, ensure the quality of the product, market the product, deliver the product, and support both employees and corporation. Each of these business lines should have its own mission statement or summary of key business functions, ideally linking functionally and understandably to the higher-level mission statement. By identifying each of the qualifying elements that are used to define a successful outcome, analysis will begin to lead to some AP&S-relevant findings that will contribute to risk assessment and overall risk management. An effective method is to ask the question, "So what?" from an asset valuation, threat, or vulnerability perspective. Since the overall objective of risk management is to apply an appropriate level of protection to assets in support of mission success, a lot of the answers to "So what?" will indicate that the AIC of an asset needs protecting. In our example, the organization must deliver a "high-quality product" (refining the goal toward something more achievable) that "meets or exceeds the requirements of Specifications A–E," specifications being precise, measurable, and consistent with both functional and quality criteria. From the statement, it can also be shown that the product must be deliverable on demand (transport the product) and must be produced for a reasonable cost (considering the costs to train, equip, supervise, and compensate employees within the business lines and to purchase all raw materials and consumables). Some examples of emergent considerations for AIC for each business line are broken down as follows:

- Prepare to build the product—so what? Need a
  - Trusted supply chain
  - Quality raw materials
  - Trusted staff to process invoices
  - Secure site to store materials
  - High-quality equipment, consumables, and processes (e.g., billing, receiving, etc.)
- Build the product—so what? Need a/an
  - Secure and safe facility
  - Trusted staff to build the product
  - Trusted, repeatable processes

    - Effective supervision (by people) and monitoring (by information technology (IT) and SCADA systems) of all activities
- Ensure the quality of the product—so what? Need a
    - Trusted staff performing as trained and reporting deviations
    - Input materials that meet requirements right up to the time of use
    - Secure sight with an eye toward preventing unauthorized activity or unwanted (such as counterfeit) materials
    - Trusted and routine reporting lines
    - Trusted policies and procedures that permit interruption of operations for quality reasons
- Market the product—so what? Need a
    - Current assessment of business intelligence
    - Protected customer database
    - Trusted vendors
- Deliver the product—so what? Need a
    - Trusted and protected supply chain
    - Trusted transportation personnel and vendors
- Support both employees and corporation—so what? Need a/an
    - Set of processes for fair treatment
    - Honest and fair recruitment processes
    - Credible and sustained awareness, training, and development
    - Efficient and accurate remuneration processes
    - Trusted processes for advancement based on merit
    - Protected and safe working environment

These decomposed subsets are business processes that require assets whose AIC must be assured through a risk management program. This analysis will provide the framework for further risk-related analysis and assessment. Also, by taking this approach, the tasks (tactical) and objectives (operational) that need to be met to achieve the ultimate goals expressed in the mission statement (strategic) can be isolated and analyzed. From the statement given earlier, the measurable criteria are defined in "Specifications A–E." The criteria that are used to measure whether or not the objectives are being met could then be defined in several ways, for example:

1. Must meet functionality and quality requirements.
2. Must do so in a way such that the client is not left waiting.
3. Must take into account elements such as cost. In this manner, we can validate the strategic role of the business, as expressed in its supporting business lines and functions.

Ethical or moral considerations

Some persons confuse "why" an enterprise exists by attempting to overlay moral, ethical, or altruistic dimensions (social responsibility) onto government or private industry enterprises, typically in favor of a personal or group agenda. While this is appropriate to an extent, it can be taken too far. The first clear goal of a private industry business is to generate wealth for its stakeholders. This is a key difference between the private sector and the public sector. In the private sector, the focus is on wealth, whereas, in the public sector, the focus is (hopefully) on delivering a quality service function to improve the lives of the population. In both cases, it should be clear that the first goal is to be able to achieve

the mission (and thereby generate wealth and provide needed services) as effectively and efficiently* as possible, regardless of personal preferences and beliefs.

There is an important risk management nexus to the ethical or moral dimension of an enterprise. In AP&S doctrine, all advice provided is considered to be apolitical and "politically incorrect." All recommendations for, and application of, approved safeguards must be apolitical in that they must map only to meeting the residual risk levels approved by senior management and are consistent with industry best practices, training, and education. According to Chittester and Haimes (2004, pp. 4, 5), "the level of acceptable risk depends on the critical nature of the system's mission and the perspectives of the individuals or groups using the information." They are politically incorrect in that they are statements of the supportable facts and do not get looked at through the lens of ensuring appropriate representation of demographic groups, and so on. It should be clear to the reader that this does not translate into "being abrasive," but only clinical in application.

In this manner, AP&S risk managers may find themselves in a temporary dilemma between, on the one hand, limitations on safeguard implementation that are imposed by senior management (after all, all protective safeguards have an inconvenience or hard cost associated with them) and, on the other hand, their best assessment of the most appropriate safeguards to be implemented to meet the residual risk targets approved by senior management. Fortunately, this is easily resolved. The primary role of the AP&S practitioner is as an adviser to senior management on residual risk. If the adviser communicates successfully to senior management the residual risk and any concerns after approved safeguards are implemented, even if that residual risk is higher than that which the AP&S practitioner considers prudent based on training, education, experience, and industry best practices, then the practitioner's job is done. It should be clear that there is a legal threshold here—if the AP&S practitioner notes that there is a clear violation of law or something being done that jeopardizes the life safety of the population, he or she may well be compelled to act, even without management support. This is a difficult call to make, and usually only made once within an organization, but it should be clear that the AP&S practitioner cannot simply hide behind management accountability when there is a clear and verifiable risk in this context.

Once the practitioner has expressed those concerns and senior management has acknowledged the advice provided (and thereby accepted the residual risk in question), the dilemma is resolved. Assumption of AP&S risk is a management function, not a technical one; the practitioner simply works within the residual risk targets set by senior management and implements the approved safeguards. An ethical consideration emerges only if the protective posture becomes too ineffective for the AP&S practitioner to tolerate, after which there is no choice but to vote with one's feet and seek alternate employment.

AP&S risk management in support of business and social responsibility

It is important to remember that all enterprises, public or private, manage risks every day. There are many types of risks, including financial, cultural, legal, business, partner,

* If one differentiates effective (doing the right thing) from efficient (doing things right), then it may be argued that private industry attempts to maximize efficiency (reduce overheads, maximize and exploit capabilities of staff, operate as a meritocracy) in its goal toward effectiveness (mission success being fiduciary). Government, on the other hand, focuses on effectiveness in reflecting Canadian values over pure operational efficiencies. Merit may take second place to hiring for gender equality, ethnic diversity, bilingualism, and so on.

operational, sales, and reputational, to name a few. Haimes and Chittester (2005, p. 1) note that "Prudent management of any business, whether in government or the private sector, calls for making cost-effective decisions regarding the investment of resources. Investing in the assessment and management of risk associated with cyber attacks, and thus, with information assurance, is no exception." AP&S risks to the AIC of valued assets contributing to mission success are just others to be managed within the overall process of enterprise risk management (ERM), which is a senior management function. All risk management programs exist only to support business lines, which, in turn, exist only in support of mission success, however defined in the enterprise's mission statement.

The alignment of business activities with societal norms (including ethical, altruistic, and moral) occurs on at least three levels. The first of these is the *legal* or *regulatory* level. While the business seeks to generate wealth, the government (representing and protecting the people) sets in place certain constraints and restraints* that limit how the business can achieve that goal. These are generally defined in terms of *criminal* acts between the individual and the state when the business does not act honestly. The second layer can be described as *civil* constraints and restraints—generally defined in terms of *negligence* and *tort* between individuals. In these cases, the company's failure to take all reasonable steps to prevent harm to another can lead to costs associated with *civil liability*. A third element involving social and cultural norms is a matter of projecting and protecting a *positive brand*. This brand is important if an enterprise wishes to be perceived as a positive and contributing member (or at least not as a destructive member) of the community, the region, and possibly the nation. Compliance and conformity with these and other societal norms such as environmental consciousness, charity, and community support refine what are considered to be acceptable boundaries for corporate activities, meeting objectives, and achieving goals.

A paradigm case of business and social accountability rests with those NCIs assuring national security, sovereignty, economic prosperity, and the health and safety of citizens. Overwhelmingly privately owned, these NCIs comprise those physical or logical networks that, if destroyed or disrupted, would cause serious injury to those assets supporting the NCIs' missions and also to those national objectives that have been deemed to be essential to our way of life. This includes transportation, energy, water, manufacturing, government, IT, and telecommunications; essentially, all services, goods, and commodities that are provided in the quantity, time, and quality that is consistent with the populace's expectations.

While the private sector owns and operates a significant portion of the critical infrastructures of the nation and is responsible for the provision of these essential goods and services contributing to national objectives, it does not follow that these enterprises have become accountable directly to the populace for the provision of uninterrupted, high-quality goods and services. As noted earlier, the primary role of private industry is to generate wealth for its stakeholders. The concept of making a reasonable return on investment while working in service to the nation is not inconsistent or in conflict. The burden of compliance for a private enterprise is simply to operate within the various legal, civil, and social constraints and restraints and to produce the goods and services in a quantity, quality, and timeliness outlined in contracts with the government. The government retains all accountability to its citizenry for meeting national objectives. Communicating to the

* A constraint is considered something that must be done; for example, all products must be sold by year end. A restraint is something that may not be done; for example, there must be no casualties or injuries during construction of a new production line.

NCIs the expected levels of performance, including standards of protection of the AIC of supporting assets, is a government responsibility and one to which the AP&S practitioner contributes significantly within the NCIs' risk management programs. While responsibility for the provision of a capability can be delegated, accountability for results cannot. This is especially true in the cases of government oversight of its NCIs. Supervision of performance, periodic monitoring and auditing, setting training standards, timely communication of threats, and information on vulnerabilities or changes to mandatory requirements are all essential elements of accountability.

In summary, following industry best practices for AP&S provides a secure and safe operating environment for the enterprise, and also contributes to legal compliance, protection from civil law suits, and a positive brand. In this manner, the AP&S risk management program definitely contributes to ERM and mission success, however defined.

Scope of risk management

As discussed earlier, when considering the basic elements of risk, the perspective and expectations of the individual or organization affected by the risks is important to understand. Consider the issue of critical infrastructure and who is responsible and accountable, both for individual service provision and in aggregate. In comparison, if one asks a citizen who requires a specific good, commodity, or service who is responsible for ensuring that the service is available and of expected quality and quantity, the reply will likely be "the company, of course"—the result of the service agreement between the individual and the company.

Regarding the provision of critical infrastructure services, the private company may fully understand and appreciate the expectations or service-level agreements with government if they are stated explicitly (which, in many cases, they are not, due to a lack of governmental oversight mechanisms). Companies, ever mindful of the financial bottom line, may prioritize how those services are to be achieved and to what extent they are achieved—particularly in the case of widely distributed services. Finally, as noted earlier, the government may require that the company providing critical infrastructure services comply with legislation and regulations to ensure that the service is available to some quantifiable extent (typically a percentage of "uptime" and "quality of service") and hopefully take steps to ensure that those criteria are met. In each of these cases, the concept of scope factors significantly. Clear delineation of roles and responsibilities, agreed to by all stakeholders, is essential to agreement on the scope of services provided, to provision of service, and to reducing any gaps in the protective posture of the NCI providing those services. The AP&S risk management program contributes to ensuring the provision of services and, ultimately, the mission success of the NCI. Risks within the NCI and among NCIs (since they are interdependent in many cases) may be influenced significantly by the actual ability to meet enough of the mandated or expected (by government) demand for critical services for the organization to remain viable, if not profitable. Finally, from the government perspective, a risk necessarily has a much larger scope, perhaps regional or national, in which case it may focus on and manage the ability of many companies to maintain an appropriate level of a critical service within a community—requiring the elimination of any one company as a single point of failure (SPOF) in the provision of an essential service to an individual, a community, a region, or a nation.

Thus, it can be seen how the extent to which scope can define how risk will be assessed and managed; scope becomes a limiting factor. From the corporate perspective, it may be communicated that the risk is being assessed in relation to the *ability of the corporation to*

remain viable, if not profitable, in meeting its service delivery mandates from government. From the government perspective, the risk may be assessed twofold: first, in relation to the *trust of the community that a certain service will be available* on demand and to an appropriate quantity and quality to meet collective needs, and second, in relation to the *ability of the government to ensure, through service-level agreements (SLAs) and oversight, to continuity of service* in the expected quantity, time, and quality, to all citizens requiring it. From the individual's perspective, the risk may well be defined in relation to his or her *trust in the delivery and quality of that service at the home.* Each of these statements implies a reassessment of, and perhaps changes in, the company's objectives to be met and the goals to be achieved.

The reason that scope and perspective have been emphasized to such an extent in any chapter on risk management is that inadequate consideration of these two elements by risk analysts, senior management, and other stakeholders has led to misunderstanding of risk management recommendations and subsequent decisions that did not protect adequately the assets supporting the provision of critical goods and services. In short, clearly understanding how perspective and scope shape the focus of any risk assessment will be a very positive and significant step toward being able both to present and to argue a case for a protection posture—be it at more senior management tables, with peers, other NCIs, government oversight bodies, or the public being served. To assist in communicating or transmitting the existence of risks in the control system domain, four basic steps are offered:

1. Express the risk at the equipment level, describing the impacts in terms of the losses of its immediate functions. This level is perhaps best understood by the operators and engineers, both of whom must "buy in" to the risk assessment to convince line managers/supervisors and senior management.
2. Extrapolate the assessed impacts associated with a specific loss of function in terms of how they would affect the local system. This will get the attention of line managers and regional managers, who are responsible to headquarters or the main office for meeting AIC requirements.
3. Communicate how the local or individual system's loss would translate to the larger system of systems at a corporate level. This moves the risk into the strategic level and, by definition, becomes a senior management concern from a purely business perspective.
4. Finally, identify any potential outside issues associated with impacts at the community, regional, or national levels. This will concern senior management from an ethical, moral, or societal perspective, which is also their responsibility as a good corporate citizen.

This layered, bottom-up approach to scoping and expressing risks to mission success capitalizes on many strengths, including the analytical skill of the AP&S practitioner based on his or her training, education, and experience coupled with a growing collection of like-minded stakeholders through the tactical (operator), operational (line or regional manager), and strategic (senior decision-maker) levels of activity. An example of this approach, when considering the valve that helps mix a certain chemical into paint to help it bond more effectively onto metal, follows:

- Based on the assessment by capable engineering and design staff, there is a significant risk that this valve would not function as intended (integrity risk) and would likely not mix the needed chemical into the paint (availability risk). The engineer or operator would likely be the first to notice this.

- This loss of service would result in paint that would appear to be bonded appropriately to the metal during a quality assurance check but would become less bonded when exposed to water, thereby causing the paint to chip prematurely (integrity risk). This would not come to light until noticed after time by the consumer.
- The premature chipping of the paint would become a quality of vehicle issue in the eyes of the consumer, devaluing the company's product in terms of being competitive against similar makes and models (a business risk). Social media and word of mouth would communicate this risk to the community, to the region, and perhaps to the nation.
- As a result of this, one could reasonably expect a drop in sales (perhaps evolving into a business survival risk). However, it would not likely impact the safety systems on the vehicle and, therefore, would not likely gain the attention of the government regulator from a vehicle safety perspective. Nonetheless, senior management quickly becomes implicated if a bottom-up approach is adopted to scope and communicate risk.

This approach is effective, applicable in any system, is repeatable, and gets a clear, validated message to senior management regarding key risks. It presents a clear and logical link that allows the individual conducting the risk assessment to identify *what was assessed* and how findings relate to the *local*, *system*, *corporate*, and *outside* objectives and goals.

Asset value

As noted, assets of several types are necessary to achieve mission success, whether in service delivery or the production, processing, movement, or storage of commodities or products. These assets have value in terms of AIC, which means that they must be accessible on demand in sufficient quality and quantity and that they must be protected from unauthorized modification and unauthorized disclosure. They also have monetary value in that they must be purchased, installed, maintained, operated, updated, and finally disposed of. This monetary value is of interest to us, and also to a threat agent who would steal the asset, render it unusable to us, or corrupt its utility so that it is untrusted thereafter. Perhaps the most valued assets when considering SCADA systems are pieces of information; therefore, "data collection, control, communication, and management, which are essential for the effective operation of large-scale infrastructures, are being performed by SCADA systems. These work remotely to improve the efficiency and effectiveness of the control, operations, and management of critical physical infrastructures" (Chittester and Haimes 2004, p. 2).

Asset valuation

Asset valuation is, simply, the process of determining how important (qualitatively and quantitatively) an asset is to mission success in terms of AIC and, also, how important the asset is to a potential adversary. This will indicate how likely it is that an adversary will attack an asset, which is a key step in threat assessment, discussed later in the chapter. Quantitative asset valuation focuses on the total cost of ownership of an asset throughout its life cycle. Qualitative asset valuation focuses on what exactly the asset does in the various processes leading to mission success, and how critical the asset is to completing a process. Several examples are cited in the following.

It is important to keep the issue of perspective and scope in mind during the asset valuation process. The reason for this is simple. Consider the panel through which electricity enters a home. To an individual, it may be a critical part of the home's infrastructure, in that if it fails or catches fire, this results in a catastrophic situation—an absence of power, which, depending on the time of year, can be deadly or extremely costly. From a community or regional perspective, a similar type of panel can be more valuable if it is contributing to the recovery of electrical services after a blackout as part of the community that sells electricity back to the grid through alternate means (such as solar). This panel could also be more valuable to keep up and running and in good operating condition, since a failure could cause a fire resulting in damage to an infrastructure on which many households depend, or injury to several workers due to higher voltages involved and the technical complexity of the system. At the level of the federal government, a fire in an individual home may be significant if it reveals a design flaw in the panel that could affect a larger part of the population, all of whom trust the government to oversee the implementation of standards to ensure that vendors provide products that work correctly and meet the expectations of citizens. Government oversight action could include triggering a recall of the equipment or direction to the company to conduct emergency repairs. Thus, it is indicated that it is important to keep in mind the consideration of perspectives and scope in asset valuation.

Asset valuation in support of mission success

The achievement of goals and objectives is the result of work completed and the resultant provision of services or the production of goods and commodities. This is usually the product of processes that are brought together in systems. These processes can be defined in terms of the following:

- The creation, transmission, processing, and protection of information to make informed decisions, whether it is to open a valve or to open a regional office.
- The efforts of personnel to analyze information from all sources and make informed decisions to take some kind of action, such as overriding the automated opening of a valve, responding to an anomaly, or hiring new staff.
- The equipment and supplies that are consumed in the process, such as petroleum oils and lubricants (POL), stationery, toner cartridges, shop supplies, or light-emitting diode (LED) light bulbs.
- The physical equipment that provides the service, builds the product, and actuates or measures an action. It also includes the occupation and use of building spaces appropriate to the work being conducted. Examples include the switches in a rail yard, navigation systems for ships, satellite communications among road carriers, specialized diagnostic equipment, and the environmentally controlled buildings and offices in which this equipment is found, such as hospitals, power stations, emergency operations centers, and IT server rooms.
- The implementation of formal (hopefully written and understood) supporting activities, including policies, standards and procedures, training programs, and oversight mechanisms, all of which are intended to assure consistent, timely, high-quality services, commodities, and products.

All of the foregoing are assets, which are shown nested in the following in relation to the processes that they support (Figure 4.2).

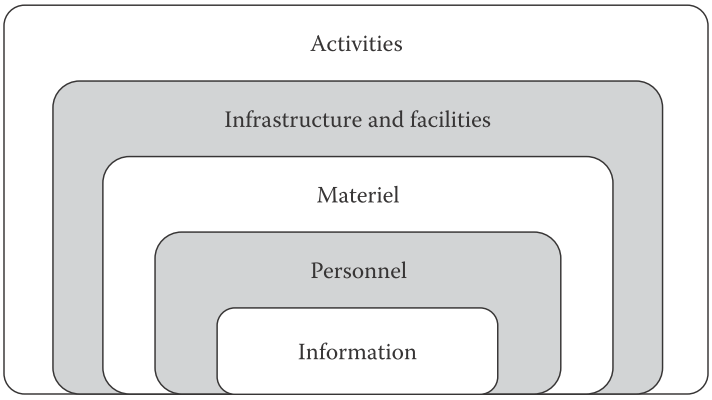

Figure 4.2 A taxonomy of asset usage.

Within the CIP doctrine, these asset groups can be organized according to the mantra of *personnel*, *materiel* (objects and consumables), *infrastructure and facilities*, and *information and activities*. For the sake of brevity, this will be referred to as the "unique level" in that it deals with a singularity—one person, one asset, one building, one piece of information, or one supporting activity. This is essential for effective risk assessment and management.

Many of these will also be the product of work or will require services that support them. This is the case with various forms of control systems. Again, the business of business is to generate wealth, not to operate a control system. The purpose of the control system is to help the company generate that wealth effectively, efficiently, and safely. So, when we are discussing the security around control systems, we are looking at an infrastructure that most likely supports an organization's critical path (but may not, depending on what business line it supports), but is, itself, often interpreted as being *critical infrastructure* because of the impacts associated with public safety (Figure 4.3).

The first layer identifies a general business line; for example, production operations (the assembly line). There are a series of discrete business functions comprising that business line; for example, each of the stations that prepare (paint, fold, drill, etc.) components to be assembled further down the line. Several automated systems (infrastructure and activities) contribute to the production process by performing a specific task or process. Each of the systems and processes, as one descends in the diagram, is an asset, and supporting the processes are additional assets as shown. Personnel oversee processes and intervene as necessary. Information is passed, analyzed by systems, and overseen by people. All processes take place in facilities and hopefully follow written procedures to produce, activate, actuate, move, or provide something (activities). Material is consumed, IT and telecom networks support communications and information exchange. Individual components (infrastructure) consume materiel, send information, are managed, changed, or maintained by people, reside in facilities, and perform a function that is essential to the provision of a mandated good or service.

Considerations for asset valuation

The valuation parameters of these assets can be refined in a number of ways. Remaining true to the business model, the values of the assets must be linked directly to the business processes and service delivery/production mandates that they support. Again, scope and perspective must be considered in asset valuation, since a misstep can lead to significant errors in the subsequent assessment or management of risk; some assets may turn out to

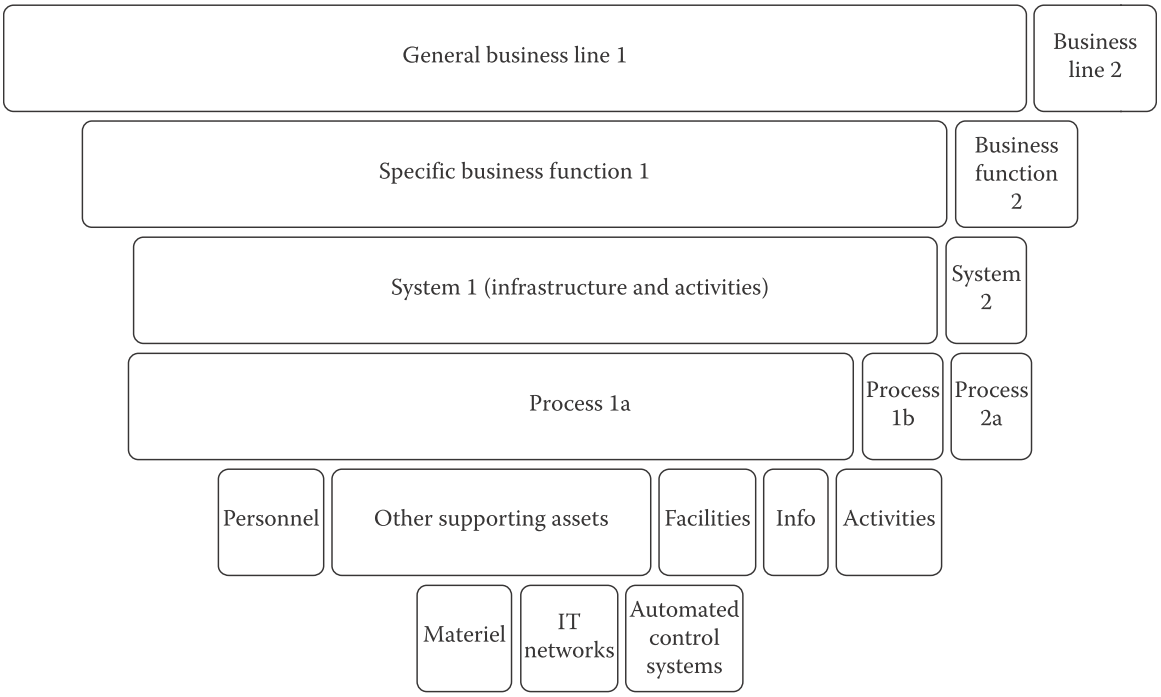

Figure 4.3 How assets support business functions.

be overprotected, which is inefficient, while others may be underprotected, which is ineffective. One approach involves identifying assets according to the following:

- At the unique or individual asset level, how does the loss of the asset affect the availability of the service (in terms of drops in production, etc.) or the integrity of the service (in terms of quality)?
- At the unique level, what are the confidentiality concerns associated with the unauthorized disclosure or loss of control over information that is directly related to the asset?
- How would these losses at the unique asset level affect the larger system, community, regional capability, or corporate entity (SLAs, legal or regulatory contracts, reputation, etc.)?

For example, in further consideration of the valve mixing a chemical into the paint for a piece of metal, one might argue that the loss of the valve entirely could lead to a shutdown of the painting line for a period of 5 h while it was replaced. The cost of this disruption would be, approximately, the cost of replacing the part, any installation/testing/calibration costs, and the lost production time while employees stood idle and no processing is being conducted (in the absence of redundant systems). Some of these costs may be recovered from returning the part for refurbishment or repairing in-house (reducing the costs associated with having to purchase a new part). The loss of the line, however, means that certain items may not be delivered on time, which is a cascading effect of the risk. Again, scope factors significantly here—the focus starts tactically or locally, but quickly rolls up to the level of the company. In this case, one might consider any penalties for late shipment, the potential losses associated with customer cancellation, or the loss of credibility or reputation in terms of the ability to deliver a product. Finally, downstream costs

may involve having to repair vehicles that are found to have unacceptable paint jobs, the cost of protecting the brand, and the potential losses of brand value.

It is important to appreciate the nexus between the disruption and the value of the asset. It is not linear. When one considers how that component affects the system, including how its loss affects the process both upstream (toward the start of the process) and downstream (toward its final outcome), one may observe a *cascading* impact, because it acts like a house of cards—remove one card and the overall structure (system) begins to topple. The value of the asset, once compromised, must also be understood in terms of the overall impact at the unique asset, process, system, corporate, and societal levels. As with our chemical valve in the painting process, the monetary cost at a unique level may be rather insignificant (a couple of dollars), but it may be much more significant at a corporate level (many individual sales lost, representing thousands in lost profits, damage to reputation, etc.).

This becomes even more profound when dealing with safety systems. Consider the various measurement tools that activate safety systems in the nuclear industry. If those fail (*en masse*—this is very conceptual), then the unique cost may only be a few hundred dollars. If the item fails and, as a result, the safety system fails to prevent a significant radiation leak, then the impact could be measured in the millions of dollars in terms of liability to the company and much more in terms of the loss of territory and citizens within the affected area.* These can be referred to as *escalating* impacts, in that they operate differently at unique, process, system, corporate, and societal levels.

In summary, the proper valuation of assets, considering their importance in terms of AIC to the enterprise as well as the adversary or threat, is an essential component to be considered in the risk management process. Assets have value only to the extent that they support the operations of the enterprise. Once this has been determined, the AP&S risk analyst can compare these findings with those of the mission analysis and begin to formulate ideas regarding the extent of existing risk and to visualize appropriate safeguards to mitigate those risks to a level acceptable to senior management. The next step, the assessment of threats, will further paint the risk picture.

Threats: Introduction and categorization

The concept of threats is reasonably straightforward; it is their assessment and treatment that become complex and, possibly, complicated. A threat can be defined generally as any condition or action, typically negative, which can cause injury to the AIC of an asset by exploiting some vulnerability. For example, a thief may take advantage of a weak lock to steal items or a heavy snowfall may cause damage to the structure if there is a weakened roof. The challenge is often that individuals and organizations alike often fail to take the time to actually (1) identify potential threats in sufficient detail, (2) analyze how those threats tend to operate in terms of their COI to act, or (3) assess the threats relatively qualitatively, having limited understanding of the full impact or effects of a threat event. Chittester and Haimes (2004, p. 2) describe threat as "the intent and capability to adversely affect (cause harm or damage to) the system by adversely changing its states."

Threats within the AP&S domain are often grouped into three broad categories—the deliberate, accidental, and natural. Within the CIP specialty of AP&S, a fourth threat type

* This is why safety systems often rely on layers of protection in terms of redundancy—to prevent a single asset from failing and allowing for a catastrophic impact. Within the nuclear industry, there are multiple layers of controls that are overlapped and layered to ensure that these kinds of events are extraordinarily rare.

is emerging in the literature, that of deterioration. This phenomenon is interesting, because it can be considered either a risk (a result of a threat exploiting a vulnerability) or a threat (which can exploit a vulnerability to cause a risk). As a risk, deterioration can be considered the result of a threat exploiting vulnerabilities; for example, in the case of bridges the threat could be natural (exposure to the elements), man-made (salting roads), or accidental (construction staff cutting corners, incorrect maintenance), and major vulnerabilities could be inadequate inspections or a lack of spending on preventive maintenance. The result, that is, the risk, is then the deterioration. Since all AP&S risks are expressed in terms of their effects on the AIC of assets, deterioration can be considered both an integrity and an availability risk. However, deterioration can also be considered the first link in a chain of cascading risks; for example, in the case of a deteriorated bridge, when it could cause an accident if it fails, and thereafter cause a disruption in transportation, supply chains, and manufacturing (and possibly IT/telecom if conduits are routed across the same bridge).

As a threat, deterioration (or more specifically, a deteriorated infrastructure) can exploit the same vulnerabilities to cause the same cascading risks noted earlier. For the purposes of this chapter and follow-on study, deterioration will be considered a threat.

Deterioration (or alteration) in the *Dictionary of Civil Engineering* (Kurtz 2004) refers to defects or (negative) changes in the texture of a work resulting from mechanical, physical, chemical, or atmospheric causes (threats). The *McGraw-Hill Dictionary of Engineering* (2003) definition is, perhaps, more precise, referring to a decline in the quality of a structure over a period of time due to chemical or physical action of the environment. From the *ASTM Dictionary of Engineering Science and Technology*, 10th edition (2005), deterioration results in a need for repair due to physical or mechanical breakdown, and is a permanent impairment of the physical properties. The constant in all these is that the infrastructure no longer maintains the same robustness and resilience that it was intended to maintain, meaning that, as the demands placed on it approach its overall capacity, the likelihood of failure increases at what can be described as an increasingly unpredictable rate. Given the current state of infrastructure, the understanding of deteriorating structures is an increasingly important area of study to the AP&S practitioner.

Threats can also be described as failure scenarios when applied to SCADA systems. According to Bobbio et al. (2010, p. 1346), "A failure scenario consists in the identification of the sequence of adverse events that have produced an anomalous and undesirable behavior …, the identification of services that have been impaired (in terms of continuity, readiness, performance, response time) during the sequence of adverse events and the set of interconnected networks that … have contributed to their degradation."

There are several characteristics that distinguish threats in general and apply to these four threat types, including COI. Again, while not mathematically sound, it can be argued that if one or more of these are missing, then the attack or the threat event will not likely be successful.

Capability refers to the extent to which the threat agent possesses the knowledge, skills, abilities, and resources to launch an attack, including "ability and capacity to attack a target and cause adverse effects" (Chittester and Haimes 2004, p. 2). Opportunity refers to how possible it is to get close enough to the target to launch an attack. This includes the receipt of information regarding vulnerabilities of the target's assets; routing information of targeted IT systems for cyberattacks; transportation, infiltration, and exfiltration (if required) routes for physical attacks, and so on; essentially, anything that can get the threat agent into the proximity of the valued assets to be attacked. It may also be referred to in terms of the attacker having the time and space to commit the attack without fear of response or disruption. Intent is, perhaps, the most difficult to gauge, and refers to the level

of commitment of the adversary to actually launch an attack, including "the desire or motivation of an adversary to attack a target and cause adverse effects" (Chittester and Haimes 2004, p. 2). Intent can result from cultural, ethnic, criminal, or religious indoctrination, the influence of a charismatic leader or family member (as in the Khadr case), or peer pressure.

Another challenge with intent involves the conditions that operate at the fringe of rationality, such as we find with those with significant mental challenges or that have been radicalized. In these cases, the ability to protect the infrastructure using deterrence and similar factors is often offset by the attacker's willingness to trade everything for success. This reveals another vulnerability or gap in our own defensive posture—the ability to assess the potential of violence that may or may not be present in a certain environment. This is the subject of ongoing efforts, and has resulted in a number of tools (such as the WAVR-21 assessment and similar structures). The challenge is that such tools are still at the point where those using them must possess significant education if there is to be an assurance of effectiveness.

In addition to categorizing threats by type and by characteristics, AP&S analysts also group them as being either internal or external (Cardenas et al. [2009] refers to them as outsider and insider attacks). An internal threat, such as an employee, contractor, or authorized visitor, has some or great knowledge of the organization, including its operational processes and its security posture. An internal threat has been granted access privileges to physical and electronic assets, and therefore possesses both capability and opportunity to launch an insider attack. According to Gold (2008, p. 40), "70% of attacks tend to be internal to the organization concerned. This is especially true with SCADA-based systems."

From a protection perspective against internal deliberate threats, corporate efforts typically revolve around ensuring the loyalty and reliability of the insider through background checks, appeals to patriotism or to "the team," or routine supervision and fair compensation to minimize any intent to launch an attack. The latter two, however, are areas of constant pressure—particularly as supervisors' workloads (including administrative tasks) increase and economic pressures continue to cause organizations to look for opportunities to adjust their balance sheets. An external threat has no legitimate access to assets, and must therefore build the capability, opportunity, and intent (COI) before attacking. In the case of deliberate external threats, all are developed with the assistance of intelligence which is typically gathered through reconnaissance of the target facility and information gathering from insiders and other knowledgeable people. This can occur accidentally through social engineering or deliberately through bribery, extortion, blackmail, subversion, or threats. In the current environment of standardization, there is a growing vulnerability that an attacker can identify a less protected area and, based on the need for compliance with a standard only, gather useful information for an attack against a more sensitive location.

The deliberate attack involves a willful intent to cause direct harm against assets to impact the AIC of an enterprise. The accidental attack does not involve intent, but rather negligence, inattention, distraction, fatigue, or overwork. In the case of the latter, there could be an intent by senior management or line managers to overtask or overwork their employees, thereby introducing the conditions for an internal or external accidental threat to occur and cause harm directly; that is, a hazard. This can lead to additional issues, such as legal liability, particularly where the demands placed on the organization move further and further away from the expected maintenance and operations of the equipment and processes.

A natural threat causes harm without intent by its nature and often affects the environment in which the entity operates, particularly within the realm of control systems. It may

Threat types	External	Internal
Natural	Earthquake, tornado, flood, tsunami, tropical storm, hurricane, thunderstorm, blizzard/snow/ice storm, hail, volcano eruption, landslide, erosion, wildfire, high wind, extreme temperature, disease, drought, animal attacks, meteorite, asteroid	
Deliberate	Terrorism, crime, sabotage, subversion, hostile military action, insurrection, state- or corporate-sponsored espionage (personal or electronic), cyberattacks, political activism, hoaxes, poisoning	Employee sabotage, theft, strike, work action (work-to-rule, slowdowns, stoppages, delay of access)
Accidental	Cut cable or water pipe (backhoe threat), wildfire, spill of dangerous material, poisoning	Error, loss or improper use of equipment, improper maintenance, slips and falls, spills, flooding, fire, poisoning
Deterioration	Erosion, rust/corrosion, weather fatigue	Wear, neglect, stress/structural fatigue, aging equipment or material

Figure 4.4 Threat categories.

also affect the area surrounding the infrastructure, meaning that the ability to respond to the event can be deteriorated significantly. Consider a serious storm—individuals needed to respond to an event may not be able to reach the facility. This is also a concern for business continuity planners who, from time to time, need to explain that plans may need to remain at the employee's home where they can be accessed if the facility cannot be.

Deterioration, as a threat, can be deliberate (e.g., willful decision not to maintain an infrastructure) or accidental (e.g., inadequate or nonroutine inspection or maintenance). The former case is a particular vulnerability where budget cycles and politics are linked—the cost of the maintenance of the infrastructure may lead to deficits, which, in fiscally restrained periods, are not politically acceptable. In the latter case, there will typically have been a change in some aspects of the infrastructure; for example, in the case of a bridge, it could be increased traffic, use of a new type of ice melter, different paving techniques or materials, a different paint type, and so on. Figure 4.4 summarizes the threat types and offers additional examples.

Analysis of threats

As noted earlier, analysis answers the question, "How bad is it?" Regardless of the threat under analysis, one must consider the likelihood of a threat agent exploiting a vulnerability to cause injury to an asset (risk), and the general impact of a successful attack. Threat assessment takes it one step further, and answers the question, "How bad is it to us?"; that is, the results of applying threat analysis to the assets, processes, systems, and enterprises under risk assessment. One method to conduct further threat analysis is described in the following.

Understanding that the threat is the act or condition that provides the vector or path for injury to be caused to an asset, it is now useful to consider further the nature of the threat agent. He or she can be described in terms of what they actually *do* to cause the injury to the asset—such as a burglar committing a theft or an IT cracker breaching the firewall of a corporate enterprise system. From the commission of the act, which has a certain likelihood based on the COI discussed earlier, three important elements for threat analysis emerge:

1. The threat itself in terms of the nature of the injury involved and resultant impacts (such as theft leading to unauthorized disclosure or loss of assets)
2. The threat agent performing the actions that lead to the threat manifesting itself (such as the burglar committing the act of theft)

3. The threat vector that describes the physical or logical path that is taken by the threat agent to successfully launch an attack (which will be discussed more in the section on vulnerability)

Challenges to threat assessment

In applying these three elements to the realm of control systems, one needs to be cognizant of the various different kinds of threats at the unique asset, system, and corporate layers. It is not sufficient simply to be cognizant of one form (say physical or technical) and ignore the others; this could lead to an incomplete assessment and introduce gaps (vulnerabilities) into the protective posture due to incomplete risk assessment. This is particularly true when dealing with high-availability systems in organizations that may be involved in operations with a significantly potential insider threat; for example, that of an employee or another given full and unmonitored access privileges to controlled areas and sensitive assets. These kinds of insider threats may become particularly grave because, as mentioned, they will typically have advanced or extensive knowledge of operations (and the controls that protect them), access to sensitive, high-value or other significant resources (such as keys or token to gain access, money and negotiables, and control consoles), and abilities to launch an attack and cause an impact (having often been trained specifically on the system, understanding the extent of monitoring and auditing of security-related events that take place, and provided with lists of what not to do). They may also act on behalf of an outside individual with ulterior motives, such as through the introduction of a USB device in return for money, where the attack is intended to cause other forms of harm.

To counter this, it is often proposed that the various members of the operations and AP&S (e.g., the corporate, IT, and continuity staffs) communities maintain routine liaison to share threat information regularly and as events occur, so as to generate a clear picture of likely threats to organizations that are similar in location, lines of business, size, sensitivity, value of assets, and so on. This information sharing is a necessary element of threat analysis, but is often defeated due to stovepipes within organizations or convoluted reporting chains. The premise is that all threat information is simply data, and the more the better, whether it is received from open (public, nonsensitive) or closed (private or government, sensitive) sources. At the highest sensitivity levels of information regarding a specific threat in terms of its COI, it is often the source of the information that leads to the closed and sensitive classification of the information, and not the content. Some information from open sources can be factually the same as from closed sources; it is the confirmation from trusted sources that verify the accuracy of the information, which better contributes to risk assessment and choice of safeguards under risk management. Typical closed sources include confidential informants, interception of signals such as telephone conversations, imagery from satellites, collated reports featuring analysis and assessment of COI that are prepared by the military and lead security departments, and so on.

A typical weakness (vulnerability) in the threat assessment process is the reluctance of some government agencies, private enterprises, and individuals to share information, regardless of the operational requirement to do so bidirectionally with public and private industry, especially in the case of NCIs that are working in the national interest. As discussed, some information is highly sensitive based on the source, even though the content is much less sensitive, or even unclassified. In other cases, the reliance on open sources without checking to determine whether or not the information is reliable

and credible swings the vulnerability pendulum to the other extreme—where too much information (some of it just noise) clouds the situation. Private industry requires only an assurance from the government of the veracity and accuracy of the information, not the source. Information can also be "anonymized," that is, stripped of specific names and locations while retaining the essence of the threat details, likelihood assessment, and impact assessment. This, however, can pose challenges, as organizations move toward an increasing integration of geographic information systems to plot events in attempts to detect patterns or areas of concentration. Periodic operational security awareness sessions and reminders will go a long way to ensure that even the redacted or stripped threat information is protected from those without formal access approval, requisite security clearance, and the need to know. While the greatest fear of government agencies may be unauthorized disclosure by private industry, there is a reciprocal fear. Private industry, in many cases, is afraid of at least two things: first, that the government will fail to protect adequately their intellectual property and trade secrets from competitors; and second, that the government, learning more about the workings of an individual enterprise (including NCIs, interestingly enough), may impose additional regulations, policies, or taxes that could impede the freedom of the enterprise to operate. A subset of this is that private-sector enterprises, often already regulated, are rather reluctant to share vulnerability information with their regulators—particularly where the regulator entrenches its position with an "enforce everything" rule. Without the mutual confidence to share and protect each other's information, the threat assessment process remains incomplete.

A key concept relating to the sharing of both threat and vulnerability information is that of trust. As alluded to earlier, trust is essential to information sharing, comprehensive threat analysis and assessment, accurate risk assessment, and the appropriate, cost-effective implementation of safeguards. It is interesting to consider that all trust is personal; individuals will not typically share information unless there is mutual, personal confidence that the recipient actually needs the information, that sharing contributes to the common good (an integrated protection posture within and among enterprises, especially NCIs), and that the information will be protected adequately. That is why relationship building is so important among threat analysts; it is more likely to guarantee a continual flow of threat information. How is trust earned? The author suggests that, from an AP&S perspective, first and foremost, be good at your job. This requires training, education, and experience in your AP&S specialty. With demonstrated competence comes confidence from your peers. Also, you will be more able to communicate your information requirements to your peers, as well as to your and their senior management, making reasoned arguments based on a full understanding of protection requirements at the strategic, operational, and tactical levels. If the respective senior managers open the conduits, it remains only for the line managers, intelligence staffs, and AP&S analysts to begin sharing information of mutual interest, knowing that it is valued and both the source and information will be protected. In this manner, threat assessments will have more quality, which will contribute to the quality of the subsequent risk assessment.

The threat analysis effort focuses on one very basic question—"What or who is attempting to injure (deliberately) or is responsible for the injury of persons, materiel, facilities, infrastructures, information, and activities?" The focus of this question is always on operations and determining what injurious influences may occur (proactive), have been detected (alarms and indications), have occurred (reactive), may have shown indicators, or may be emerging within the physical and logical realms of operations. This approach

has two benefits if supported by effective information sharing. First, it keeps the various groups aware of what kinds of threats are present in the environment so that they can take a more holistic approach to prevention, preparation, mitigation of vulnerabilities, and preparations for response to a threat event. Second, it increases the number of "eyes and ears" that can give the overall organization the ability to detect the approach or presence of a threat. This is called situational awareness in AP&S doctrine and is based on the following principles:

- All stakeholders understand and comply with baseline security safeguards and additional safeguards implemented as a result of a TRA. This means that they understand the residual risks to operations, and work within those boundaries. It also means that they understand what constitutes "normal" behavior in the operating environment—"business as usual," if you will—especially with respect to physical and logical access to valued assets.
- Knowing what constitutes business as usual, all are able to identify anomalies in operations, which are "not business as usual," and understand that it is their responsibility to challenge unknown persons conducting reconnaissance, attempting unauthorized access, and isolate or cease all unknown processes (within their levels of expertise and pursuant to policy and by following formal procedures).
- Since all anomalies to operations are likely to have an AIC nexus, reporting all such unusual incidents to line managers and to departmental or company security officer staff.

Through establishing technical and professional competence in AP&S, especially in threat assessment, as well as developing situational awareness and instilling mutual trust within an enterprise, among like enterprises, and also among collaborating enterprises (such as NCIs), more threat data will be made available to all, more comprehensive collation and analysis will be conducted by individual groups of threat specialists, more accurate and useful results (assessments) will be produced, and more threat products (threat assessments, intelligence summaries, etc.) will be shared among operational stakeholders. This will permit more accurate risk assessments to be conducted of individual facilities, infrastructures, and enterprises, which, in turn, will result in more informed decision making regarding the implementation of safeguards. The overall result will be a more appropriate cost-effective protective posture, and one which will lend itself to integration of safeguards within and among facilities and infrastructures, and among enterprises (government and private industry). Continued trust and the trusted sharing of useful products will be considered a success, and in business, as in threat assessment, success breeds success. More and better products will be shared by more and better threat analysts.

The terms of reference, charter, or "marching orders" for such a group of AP&S threat analysts would be straightforward to establish (assuming that all practitioners understand their roles, as discussed earlier). One key requirement (after trust) is courage on the part of both practitioners and senior management to open up their fingers and give up their tenuous hold on sensitive information in the outdated and mistaken impression that "knowledge is power" in AP&S, especially in threat assessment. While this concept may still be valid in politics, the author opines that it has no place in risk management, especially with respect to NCIs. Given the consequences of a breach or a successful attack on national objectives, in most cases the restrictive and exclusive "need to know" principle must be replaced with the more inclusive (within the threat assessment cohort) "need to

share" principle, subject to the caveats and anonymization techniques discussed earlier.* Once the technical competence of the potential recipients of threat information has been established, and once trust is instilled, it remains only for the managers to park their egos and start the bidirectional information flow in strict conformance with the details of the information-sharing agreements among the group.

The goal is to achieve a broad representation of the AP&S and operational communities that can be influenced by threats. The ideal is to have each of the major organizations represented at the group by staff who are cognizant of the information-sharing requirements, authorized to speak about sensitive matters regarding the organization, and, most importantly, authorized to share threat information with all members of the group. As an example of the potential dynamics of such a professional body of threat analysts, individual representatives of the group could provide a routine and periodic overview (in real time) of what their organization has been contributing to operations and the challenges that they have faced. This would indicate the requirement to meet regularly to exchange ideas and information. In defining, describing, and analyzing those challenges, the speaker would use the framework of deliberate, accidental, natural, or deterioration threat types, taking into account both logical and physical domains. For example, the human resources organization may report that the online application system used to provide the initial screening of applicants (a personnel security measure) has shown signs of becoming unstable periodically (which might result in a false positive in showing a person to be trustworthy when he is not). The engineers responsible for the control system may indicate that they have been experiencing a much higher rate of replacement activities due to damaged equipment in a certain area, and the two seemingly disparate items may very well be collated and analyzed to determine that a deliberate threat event has occurred. It is important, in these meetings, that the information presented is accurate and critical (i.e., based on observation and analysis), nonaccusatory (this is not about performance reviews), as comprehensive as possible, and, perhaps most importantly, useful to others.

Part of the outcome of such meetings is a more defined and explained threat in terms of knowledge, skills, abilities, adaptability, resources, intent, commitment, and proximity. What is being established is a standardized, deterministic, and consistent approach to describing, collating, and analyzing threats to promote clearer understanding for subsequent assessment. With a clear understanding of threats, the analyst can then compare them to vulnerabilities to determine the further likelihood of a threat event taking place as it exploits those vulnerabilities.

In summary, threats are the most uncertain element in the risk equation, since, unlike the mission, assets, and vulnerabilities, the organization does not "own" the threats. Further, there is no apparent limit to the intent of a threat actor to launch an attack. Therefore, it is essential that the fullest picture as possible be amassed by threat analysts. It is clear that they cannot do this in isolation; they must collaborate and share threat information, unencumbered by outdated concepts of security clearances and other impediments to bidirectional information flows. Threat data can be sanitized through various methods, after which it will require courage on the part of senior management to release it. All recipients must be trusted to use the threat information responsibly, to share it with trusted colleagues, and to protect it appropriately throughout its life cycle. In this manner, the most accurate and current threat picture

* It should be clear that the need to share is based on operational requirements for an organization to be given access to information; it is not the simple act of making all information available to everyone.

will be possible, which will, in turn, improve the quality and utility of the subsequent risk assessment.

Vulnerabilities

A vulnerability, as put forward in the MIPIS program and other credible institutions that have a strong risk management approach, is described as a gap, weakness, or "lack" of something in an asset. These gaps are inherent in states of the asset (Chittester and Haimes 2004, p. 11) and in many cases of SCADA systems are the result of not seeing "security as a major integral part of the system" (Patel and Sanyal 2008, p. 401). These weaknesses can be exploited by a threat to cause a loss to the AIC of valued assets supporting the mission. This potential for loss is a risk, the extent of which must be assessed and safeguards applied to mitigate it. Since security and protection can never be absolute, and since not all risks can be mitigated completely (due, in great measure, to the uncertainty in assessing threats), there will always be some risk remaining. This is residual risk, which is assumed by senior management to be part of the cost of doing business. So, vulnerabilities are a key component of the risk equation, and also of risk management. Fortunately, vulnerabilities are perhaps the easiest to mitigate.

The primary reason that vulnerabilities can be mitigated is that they are "owned" by the enterprise. This is a key element, as asset values are likely relatively stable and threats are often outside of management control. All vulnerabilities are inherent, or else emerge, typically as an act of omission, not commission. All vulnerabilities exist or reside in assets, which are owned or controlled by the enterprise; specifically, senior management. Therefore, senior management has full control and discretion over addressing vulnerabilities in their enterprise. Since, by definition, vulnerabilities are a weakness, inadequacy, or lack of something that presents a "hole" to be exploited by a threat, they must be expressed in negative terms. The treatment of vulnerabilities has often proven difficult, however, because they are not approached clinically, dispassionately, and critically, but often in terms of a more accusatory approach that tends to devolve into unproductive, or even defensive, entrenchment of organizations. Figure 4.5 demonstrates a possible hierarchical structure around vulnerabilities.

The subjects of fragility and deterioration, however, are beginning to challenge this approach. In these cases, there can be vulnerabilities that emerge as the result of direct actions. For example, an increasing loading is a direct act that puts additional strain on an item and brings it closer to failure. These conditions, however, are often the result of systematic or management decisions, not just individual acts.

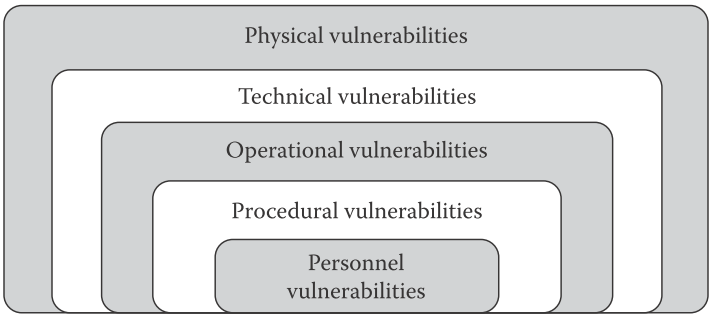

Figure 4.5 Taxonomy of nested vulnerabilities.

A fundamental vulnerability in any organization concerns the personnel (the inner layer of the taxonomy), and this may be the reason for organizations "circling the wagons" against the vulnerability analyst when he or she starts discussing weaknesses of individuals. While the intent is not personal, many people find it difficult to hear that they are not yet capable, even though it is true. Starting at the bottom of Figure 4.4, typical personnel vulnerabilities include the following:

- Lack of proper security clearance prior to being granted access to sensitive information. This results in a security breach in all cases.
- Lack of or inadequate technical training prior to assuming duties. This results in a capability gap while the individual learns "on the job," making errors and possibly causing accidents along the way.
- Ego and inability to acknowledge that one is not yet capable. This vulnerability can lead to anger, resentment toward the AP&S staff, and the hiding of other vulnerabilities. Without the maturity and courage to disclose fully the extent of additional training, education, and experience required, personnel will not be able to improve their operational capability.
- Inadequate supervision. Some senior managers in organizations think (erroneously) that "a manager can manage anything" and put untrained, uneducated, and inexperienced personnel in charge of competent practitioners. These managers simply do not have the capability to manage, guide, and correct technically competent staff, especially in AP&S. Another instance of inadequate supervision occurs when managers simply do not follow up on the activities of their subordinates and do not know what or how much work is being done; quality assurance often does not even make the cut as a business function.
- Lack of security awareness program. While senior management is ultimately accountable for protecting the assets supporting mission success, all personnel are responsible for protecting the assets entrusted to them as part of displaying due care. If they do not know what is expected of them to protect sensitive information, high-value equipment, the secrecy of how they operate, or to physically protect themselves, then there will be insufficient assurance of the AIC of assets, which could impact operations.

It is important that personnel vulnerabilities be addressed first, since many of the other vulnerabilities could cascade and be exacerbated due to weaknesses at the level of the individual. It must be stressed that these are not typically *personal* weaknesses, or individual flaws, but *personnel* weaknesses, which are institutional. There is no intent in vulnerability analysis to impugn any individual, but only to identify gaps that could be exploited by a threat. Vulnerability analysts are, after all, corporate resources whose primary role is to support operations.

If personnel vulnerabilities remain, there will be some uncertainty as to whether effective policy, standards, and procedures will be formally captured, or whether they will remain in the "corporate memory" or in "Sam's head." If no one but Sam understands how to operate or maintain a control system, for example, and Sam gets hit by a bus, this represents a SPOF, which, in the author's opinion, is the most serious type of vulnerability when discussing SCADA systems. Procedural vulnerabilities include the following:

- Lack of or outdated distributed security policies, standards, and directives. Policies should be approved by senior management as an expression of the importance of protecting valued assets that support operations. It is preferable that all key security

policies such as corporate (physical, personnel, operational), information system, emergency management, and continuity of operations security policies be contained in one document. This assists in addressing any vulnerabilities associated with conflicting or incomplete direction.
- Inconsistent or conflicting procedures. At the process level, it is critical to ensure consistent, repeatable performance by all operators; otherwise, an apparently minor lack of attention to an anomaly could escalate very quickly to affect the whole process.

If the correct performance of individuals cannot be assured in light of inadequate training and procedures, then there could be significant operational impact. Operational vulnerabilities include the following:

- Lack of alignment of individual operational processes. This could result in one process working against another, thereby introducing more operational vulnerabilities.
- Lack of training in hazard and accident prevention.
- Inadequate personal protective protection equipment. This is either a personnel or an operational vulnerability and could lead to injuries which could render key personnel unavailable to do their jobs.
- Lack of cross-training of personnel. This could lead to SPOFs if key personnel with unique knowledge or skills are unavailable for work.
- Lack of communication among and within business lines. The classic "silos" impede information flow, understanding, and overall operational effectiveness, and could introduce "holes" in the overall corporate posture that could be exploited by an internal or external threat.
- Lack of operational security, which typically means maintaining the confidentiality of the workings of the organization, from strategic direction, to operational-level business lines, to tactical operation of equipment. It also refers to maintaining an operational focus to work activity and ensuring that no actions are taken which could affect the efficiency, reputation, or credibility of the organization.

Vulnerabilities in the first three types could start to have compounding effects on operational effectiveness; when technology is added to the mix, it can become even more serious. Technical vulnerabilities include the following:

- Lack of hardening of IT systems supporting operations. Hardening includes antimalware, intrusion detection or protection systems, disabling all unnecessary ports and accesses to the system, timely and complete patch management, encrypting open communications where warranted, and continuous monitoring of activity to identify anomalous actions.
- Lack of physical separation of IT systems and lack of integrated management. According to Haimes and Chittester (2005, pp. 3–4), "The need to store business information has added a new function to SCADA: the management information system (MIS). MIS enables managers and customers in remote locations to monitor overall operations and to receive data that facilitates the making and review of high-level business decisions. The … SCADA—the engineering process control subsystem and the MIS—could be in conflict at times … the PCS has dominance … integrating security into the SCADA system more difficult. The situation is further complicated by company hierarchy; … the MIS is under the control of the chief information office, while the PCS is controlled by engineering." "This integration of SCADA networks

with other networks has made SCADA vulnerable to various cyber threats" (Zhu and Sastry 2010, p. 2).
- Inadequate configuration management. Doctrinally, all changes to an approved system have security implications; accordingly, if all changes do not go through a formal assessment process for operational and security concerns, then new vulnerabilities or instabilities in the network or control system could be introduced.
- Inappropriate clipping levels. These settings, to determine when an anomaly should set off an alarm, could lead to more vulnerabilities, and possibly an attack, if they are set too openly.
- Infrequent maintenance. Not checking and maintaining equipment regularly could lead to failures, which may affect operational schedules.

Finally, if vulnerabilities exist in overall operations, the attitude of line personnel and management could be transmitted to the physical posture of the organization. Physical vulnerabilities could include the following:

- Inadequate physical access control. This could include leaving doors and windows insecure (including propping doors open for smoke breaks), not challenging unknown individuals, and so on.
- Lack of defense in depth. This could include not having perimeter fencing, signage, or reception areas.
- Not physically locking and controlling valued assets, such as IT systems, negotiables, IT server rooms, control rooms, consumables such as fuel, high-value equipment and spare parts, and so on.

Thus, it is seen that vulnerabilities do not exist individually or in a vacuum; rather, they can spread and either introduce new ones or exacerbate the magnitude of existing vulnerabilities. The greater the number, type, and extent of the vulnerabilities, the greater potential exists for threats to launch a successful attack, resulting in risks to the AIC of valued assets, with consequent operational impact. As with threats and asset valuation, vulnerability treatment is another instance where practitioners and professionals must consider the needs of operations first.

It is important for the vulnerability analyst to understand the concept of a temporal vulnerability, one that changes over time, such as the fragility of infrastructure in different seasons or the ability of an individual to withstand fatigue when working long hours. Most temporal vulnerabilities are a result of deterioration, whether accidental or deliberate, of a capability, as indicated in the aforementioned examples. When paired with deterioration as a threat, the risk is potentially compounded.

Understanding how these vulnerabilities emerge is critical to understanding risk. Consider a physical example of a building completely surrounded by a deep ditch over which persons take a footpath. If the threat is a vehicle-borne improvised explosive device (VBIED) that cannot get close to the facility because of the ditch, what changes in the vulnerability to this kind of attack can be discerned? There are questions to be answered here; for example, can the truck use the footpath or use bridging materials that may be readily available that can be used by the truck to cross the gap? At the same time, perhaps the driver of the truck is aware of the physical obstacle from previous reconnaissance, and will also bring materials that can be used to breach the obstacle. To counter the potential for a threat to exploit a vulnerability, the individual must understand the potential threat event and the extent to which conditions that are observed reduce the means, opportunity,

or motive of the threat agent to launch an attack. This can be triaged by using a hasty method of linking the capabilities, opportunities, and intent associated with the threat to the means, opportunity, and intent facilitated by the environment (i.e., vulnerability).

While this approach is applicable directly to physical networks, it is also applicable to logical networks. IT equipment may be susceptible to threats exploiting vulnerabilities and causing risks that involve destruction, disruption, or corruption of equipment. At the logical level, it may include opportunities for malicious or otherwise disruptive information to cause havoc with the system, through exploiting such vulnerabilities as a lack of separation (from other networks, from other sensitivities of information, or other operating environments), inadequate hardening controls (such as firewalls or intrusion detection systems), or even inadequate training of personnel (which could cause accidents).

The description and representation of a vulnerability, therefore, must map directly to the threat (which can exploit it to cause a risk) and to an asset (which both houses the vulnerability and is impacted by the risk should a threat successfully exploit a vulnerability). This link can be analyzed in terms of the following:

- *The capabilities gap*: Describing how the vulnerability facilitates access by the threat to the asset to gain some capability desired by the threat agent (such as hijacking an IT transaction or service).
- *The opportunity gap*: Describing how the time and space available to the threat agent to exploit a vulnerability has been changed so that the attack has a greater probability of success.
- *The intent gap*: Describing how conditions found would reasonably lead an attacker (based on past tactics, motivation, and similar factors) to conclude that the rewards associated with successfully exploiting a vulnerability outweigh the risks of failure, of being identified as the attacker, or of being apprehended.

This description would also benefit from an understanding of the organizational breadth and depth associated with any vulnerability. Although all vulnerabilities are "owned" by the enterprise, since they map directly to assets used to achieve objectives, there are differing parameters that describe the mitigative effect that the organization can exert on the vulnerability to address it. These parameters can be described in descending order of effect as follows:

- *Span of control*: Exists when AP&S analysts in the organization have full, direct contact with the asset, have full authority from senior management (typically in policy), and have the technical capability to change that asset's structure, location, magnitude, or environment to reduce the exploitability of the vulnerability. This is the most effective situation in terms of being able to respond to the detection of a vulnerability because all decisions are reached internally at the lowest operational level and are most likely to be in line with the requirements, objectives, and goals of the organization.
- *Span of influence*: Exists when there is less direct control by specialist AP&S staffs, when decisions must be coordinated among various business line owners within an organization, or when vulnerability mitigation decisions must be coordinated with one or more other organizations. This situation seeks to acquire the range of action as per the span of control parameter, but must also ensure that the concerns of the other organizations are addressed. The AP&S analyst must influence the other organizations' operations and AP&S staff that vulnerability mitigation actions are

in the best interests of all. *Memoranda of understanding* or *agreement* are often used to establish the acceptable ranges of action in a specific case of vulnerability mitigation, taking into account all operational, financial, and cultural impacts of any measures taken.

- *Span of awareness*: Exists when processes are in place to identify and analyze vulnerabilities, as well as take preparatory steps toward mitigation, such as communicating their existence and assessment of magnitude to all stakeholders or hiring technically capable consultants. In this parameter, the organization cannot yet influence the environment or vulnerability, but has detected it to the point where it can begin to respond. The use of *bulletins*, *technical advisories*, and other communiqués issued by the intelligence section within the organization's security group could fall within this parameter.
- *No influence*: Exists where the organization relies on assets owned by another organization, or is not authorized or not technically able to access the assets to identify, analyze, or take mitigative action against vulnerabilities. No formal or informal relationship exists between the organizations and there is no trust established between them. Uncovering potential vulnerabilities is typically the result of an investigation of operational or performance impacts that are not otherwise explainable. Many organizations operate with areas in which they have no influence or awareness, especially in distributed operations having little direction from the center. This includes distributed and decentralized IT infrastructures. In all instances of this parameter, there is an absence of formal policy, hierarchy, or architecture; also typically missing is a cadre of trained operations or AP&S staff. This situation is best described as chaotic, nondeterministic, and inefficient. Staffs are not aware of the mission of the enterprise, nor of its main objectives, and are incapable of taking action on behalf of the mission in the absence of information or authority. In this parameter, it is the role of vulnerability analysts, supported by their AP&S managers, to identify the presence of vulnerabilities and commence building the relationships, understanding, and trust with the various business line owners and senior management to establish spans of awareness, influence, and, ultimately, control.

It is important to remember that these parameters must all "roll up" to the highest and most effective span of control parameter before trusted change can be effected; specifically, the taking of mitigative action to minimize the magnitude of the vulnerability.

Once the relevance of the vulnerability to the organization is established with respect to mission threat and asset, the vulnerability analysis (how big the gap is) has evolved into a vulnerability assessment (how significant the gap is to my operation). The focus of the vulnerability assessment is on taking the technical and operational details of the vulnerability (in terms of how it functions) and determining their relevance to the assets involved and the threats identified. It is at this point that we can begin to see the formation of the overall risk picture. The second part of the vulnerability assessment involves identifying the relevant level of control that the organization can bring to bear on the vulnerability.

In summary, vulnerabilities are weaknesses, gaps, or "lack of" something in an asset that could be exploited by a threat agent to cause a risk to the AIC of that asset, and thereby have a negative impact on mission success. Vulnerabilities are perhaps the best element of the risk equation on which to focus protection efforts, since they are typically within the physical, logical, and operational control of the enterprise.

Risk assessment and management

Once risk has been analyzed ("How bad is it?") and assessed ("How bad is it to us?"), something has to be done about it. The application of safeguards by security professionals, and the assumption of residual risk by senior management, is what risk management is all about. The management processes of "defining security roles of personnel, establishing rigorous management processes, … implementing security policy [at the] technical, operational, quality, and system [levels]" (Patel and Sanyal 2008, p. 401) all contribute to risk management. To be most effective, risk management must be proactive (Schneier 2003), as it deters, prevents, protects against, and mitigates adverse events before they occur. According to Patel et al. (2008, p. 483), "Risk assessment is … usually the most difficult and error prone step in the risk management process." That is why it is essential that risk analysts be trained, educated, and experienced to achieve usable results.

Risk management applied

As described in the introduction to this chapter, risk is a function of mission, asset values, threats, and vulnerabilities. Having objectives to achieve (mission), there will be some deliberate, accidental, natural, or deterioration elements (threats) that can exploit weaknesses or gaps (vulnerabilities) in an asset to cause an unwanted impact or uncertainty of a negative result that can affect the AIC of an organization's assets, thereby affecting mission success. Risks, once identified, analyzed, and assessed, must be treated; specific safeguards will be discussed in the next chapter. Applying risk management is simply putting into place the programs that can implement safeguards and treating with the residual risk, since "there is no such a thing as perfect security or prevention product … [which would be] extremely expensive both in economic and operational sense but also technically and socially infeasible. The arm-race between protections and attacks is a continuous up-hill battle" (Zhu and Sastry 2010, p. 2). The remainder of this chapter will cover those programmatic elements which serve to apply risk management to an enterprise.

One key step, often overlooked, is identifying the actual owner of the risk. Only this individual has the ability to make decisions on what courses of action are to be taken and where the triggers and thresholds for further action are going to be set. Too often, one looks at the risk management decisions to see that detached committees, working groups, or even individuals have essentially usurped the risk owner's role, diminishing his or her ability to maintain their accountability. There is a significant need to ensure that those making recommendations understand who owns the risks and collaborates with those risk owners to understand the basis of previous decisions.

Once risks have been assessed, they must be treated in a programmatic manner. Chittester and Haimes (2004, p. 10) suggest that three questions can assist in decision making:

1. What can be done and what options are available?
2. What are the associated trade-offs in terms of all costs, benefits, and risks?
3. What are the impacts of current management decisions on future options?

The answers to these questions will drive the programs for risk management, of which there may be many. Each contributes to mitigating (or reducing) and thereafter managing (maintaining) risk at a level acceptable to senior management. These components are introduced in the chapter offering a deeper treatment of safeguards and countermeasures.

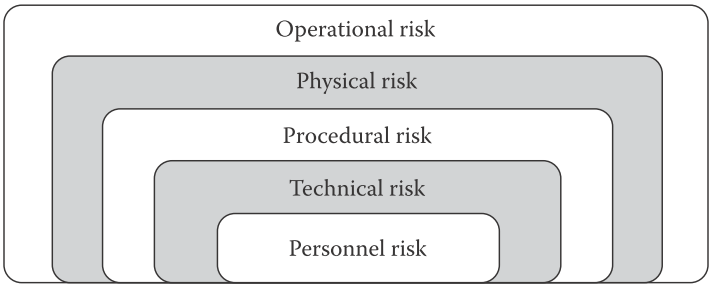

Figure 4.6 Nested risk taxonomy.

Effective risk management is indicated by the presence of processes and capabilities in the organization's AP&S program that will continually address the categories of risk (Figure 4.6).

These risks are nested in a suggested order of priority. As noted earlier, all risks map to some loss of the AIC of valued assets. Since employees and staff are arguably the most critical asset to meeting mission objectives, risks to them are considered to be the most significant. Trusted and capable personnel can mitigate all other risks; conversely, untrusted or incapable staff can exacerbate all other risks, thereby having the most serious impact on mission success. Risks to personnel most frequently result in absenteeism due to injury through accident or workplace violence, or reduction of productivity due to errors, inadequate motivation, training, or supervision. Processes and capabilities within the AP&S program that would be appropriate to manage these risks include the following:

- An AP&S policy suite (policy, directives, standards, procedures, guidelines)
- An AP&S awareness program, including rewards for compliance and sanctions for noncompliance
- Periodic spot checks by AP&S staff (also an operational process)
- An occupational safety and health program
- An emergency response program

Having addressed personnel risks programmatically, arguably the next most important risks for the organization to manage are technical risks, since technology (IT, telecom, SCADA, etc.) permeates virtually all organizations, although with the advent of voice over IP (VOIP), the line between IT and telecom is becoming blurred. Technical risks typically result in unauthorized disclosure or modification of sensitive information, denial of IT service, equipment malfunctions, incorrect processing sequences on the production line, and so on. Processes and capabilities within the AP&S program that would be appropriate to manage these risks include the following:

- An information system security program that features a policy suite; monitoring (real or near real time) and auditing (periodic snapshot) of security-related system activity; hardening; and certification and accreditation of all IT and telecom systems

Once a trusted cadre of staff is established and trusted systems are implanted, the next set of risks to be addressed programmatically is procedural. Risks could result in errors affecting operations, or in not taking correct and corrective action on the processing line, with resultant work stoppages. Processes and capabilities within the AP&S program that would be appropriate to manage these risks include the following:

- A process mapping program that formally records all business processes, interdependencies, and steps to operate
- Formal written procedures that can be used to teach and evaluate the performance of AP&S practitioners

The next set of risks concern the physical environment or "protective shell" of any operation. Risks could result in unauthorized access to the facility and subsequent risks to availability as a result of theft of assets, sabotage of equipment, injury to staff, and so on. Risks from damaged equipment, especially IT and telecom, could accrue from unreliable heating, ventilation, air-conditioning, or refrigeration systems. Processes and capabilities to address these risks could include the following:

- Formal access control programs that feature electronic access control systems, wearing of badges, or challenging of all unknown persons or those without badges
- Regular maintenance programs for heating, ventilation, air-conditioning and refrigeration (HVACR) systems

Finally, operational risks affect the overall ability of the organization to meet its service delivery or production mandates. These are perhaps the most significant risks, and also the "umbrella" risks under which all the previous risks contribute. Operational risks could arise from the unauthorized disclosure of intellectual property or trade secrets, from production impacts in not getting services or products to the customer on time, and so on. Reputational, financial, and branding risks could also be included within operational risks. Processes and capabilities to address these risks could include the following:

- Routine reporting programs to senior staff for both operational and security-related incidents, followed by programs of formal, collaborative analysis of incidents
- Employee indoctrination and awareness programs to inculcate all with a sense of operational focus

Superimposed on all of these risk treatment programs are security intelligence and incident investigations programs. The former serve to provide current threat information as part of the risk management process, while the latter serve to validate all components of the overall risk management program. Both will contribute to determining the most appropriate safeguards to implement, as will be discussed in the next chapter.

Risks, by their nature, are imprecise, potential, and unverifiable until they are realized. Thereafter, they can be analyzed and adjustments made to the security posture. Part of the challenge in corporate-level risk management is that both senior management and line employees seek refinement and detail in the guidance and advice that they are given—but do not understand that this refinement and detail does not necessarily produce an exact value of return on investment. Senior managers want a quantitative expression of security return on investment, but this is not a linear relationship of X dollars providing Y protection from risk. As noted earlier, risk management is an art and not a science; the majority of threats contributing to risks are nontechnical, so it is not possible to apply quantitative, technical solutions to address all risks. This reality is quite unsatisfying to busy senior managers who are most comfortable in comparing values in spreadsheets. In some cases, this is why security risk management gets short shrift in ERM; it is less predictable, therefore easier to disregard in the short term. If not considered, however, security risks will very likely be realized in some form, and will have a significant effect on

operations. Line employees, likewise, often demand clear proof and justification for implementing safeguards, which in all cases pose some inconvenience. They often cite a lack of historical precedent; so, if it has not (yet) happened here, why worry? Unfortunately, this is one of the fundamental challenges to an AP&S practitioner, that of "selling" the product of security in the absence of a direct impact nexus. Successful advisors are able to take security incidents that have befallen other organizations and extrapolate or apply them to the reluctant organization. But, it is acknowledged that precision in the likelihood or impact of the future risk events is not possible.

It may also be that the senior management team lacks the necessary mindset and openness to listen actively to reports on current security risks, which typically fall outside of routine risk management ranges and thresholds—itself a significant corporate vulnerability. The fundamental point to understand with risk is that it must be an honest and, as far as possible, accurate reflection of the conditions as they are found or expected. This requires trained, educated, experienced, and convincing AP&S specialists to meet those criteria, and also "a common language for risk management that may be used for describing risks" (Stoneburner 2006, p. 485).

The goal, therefore, should be to remain true to scientific principles where such principles can be applied (typically to the technical threats, vulnerabilities, and risks), but understand that there will be several risk types where scientific principles either do not apply or cannot provide the necessarily level of refinement. Once that point has been reached, then the practitioner must be able to put forward a reasonable, defensible, and confidently logical argument as to why a certain selection or decision is put forward for consideration. Reasoned arguments emerge as a result of considering risk from both historical data and also from making reasonable forecasts or predictions based on a strong situational awareness and currency with threat and intelligence information in the industry. Too often, a program manager or other administrator will argue that there is no threat (and therefore no risk) because there are no statistics or reports associated with the risk. Sophists tend to use this argument because it fits their own agendas—usually associated with making the case that nothing needs to be implemented (thereby reducing inconvenience) and no additional funding needs to be expended. A lack of historical data does not mean that the organization is not at risk. It can mean, simply, that no attack has taken place *yet*; or it can mean that no monitoring or auditing processes are in place to capture the information necessary to identify risks. It can also mean that the risk is defined differently or categorized differently within an operational system, perhaps under performance or quality of service parameters. It could also be a case of lack of communication among the various risk analysts in an organization; when risks are considered independently or in isolation among the various business lines and systems in an enterprise, the risk is often only partly identified within the organization, not fully understood in terms of the various impacts among business lines, and, therefore, not addressed with an integrated, strategic, business perspective. Finally, it can also mean that the risk under consideration is the result of something very infrequent (with, therefore, a lack of records) or something very new (such as emergent technology). In an effective risk management program, the practitioners conduct "worst-case" analysis (low-likelihood/high-impact events) and remain current on the technology, including threats and vulnerabilities.

Effective risk management means being able to synthesize all of the work mentioned earlier and accomplish four things. These are the following:

1. Ensuring that the relationships between mission, asset, threat, and vulnerability are mapped appropriately to the operations and requirements of the organization. This means being able to link those relationships among all business lines within an

enterprise to the requirements of parent organizations and other subsidiaries, and to all upstream and downstream stakeholders, especially customers and clients.
2. Ensuring that this approach is used consistently and appropriately for all forms of risk—documenting challenges in arriving at conclusions where they arise. Integrating risk management among all of these entities requires a deterministic, formal approach. This will provide a common picture from which to operate securely.
3. Ensuring that management has agreed to scales that can be used to communicate the outcomes of the risk assessment process in a meaningful and actionable way. Haimes and Chittester (2005, p. 1) remind us that "business and government still insist, and justifiably so, on the need for a way to evaluate, with some metrics, the efficacy of risk assessment and management associated with cyber attacks on telecommunications and supervisory control and data acquisition (SCADA) systems." Determining risk is but an intermediate step in risk management, and has value only to the extent that it will result in mitigative measures, which will be discussed in the next chapter. Again, consistency of terminology, of degree or significance of threats, vulnerabilities, or risks, is key to mutual understanding and integrated, cost-effective program implementation.
4. Ensuring that management communicates target residual risk, or risk appetite, early in the risk management process. By imposing any conditions that would result in senior management's nearly automatic conclusion that a level of risk is too high to accept, AP&S analysts will be able to efficiently determine appropriate safeguards and not waste time on risk management strategies when the appetite for risk is low. One method of assisting senior management in determining their risk tolerance is to provide the results of the vulnerability assessment, so that management understands how much influence it has on reducing the risk, since it "owns" the vulnerabilities more than the other elements of the risk management equation.

This last factor is linked directly to how management will choose to treat the risks that it faces. Options will be influenced by a number of factors. The first may be the level and nature of risk and how it translates into losses (in terms of AIC) to the organization. The second major factor will be the span of control that the organization can exert over the assets, threats, and vulnerabilities involved. This will guide the specific risk treatment actions that are taken by the company's senior management. These can be described in terms of the following:

- Directly *mitigating* the risk in terms of reducing any one of the values associated with asset value, threat, or vulnerability through various steps, including:
- Reducing the individual asset value by eliminating single points of failure (hot spares, inventory) or increasing the resiliency of infrastructure (redundancy), thereby reducing potential losses.
- Taking steps to reduce the threat in an area by engaging specially state-approved bodies that can engage in law enforcement or similar activities, and by sharing threat information among stakeholders and neighbors. This may result in an overall improved protective posture that will reduce the intent for a threat to act in a specific area.
- Addressing vulnerabilities by reducing the means, opportunity, motive, or perceived benefit to the attacker.
- *Sharing* the risk among organizations through the formation of communities that, through their collective efforts, have a greater impact than if they acted independently for the same level of effort. Councils, industry associations, and working

groups may contribute to understanding in this respect. Thereafter, through formal contractual agreements, individual senior managers can accept shared risk, especially in operating integrated systems, programs, and services.

- *Transferring* the risk to another entity, through either contracting out the requirement to return risk levels to acceptable levels, or having another party assume responsibility for dealing with the consequences of the event, such as an insurance company or a contracted security guard force. It should be reemphasized that this approach does not absolve senior management from accountability for decisions as to how those risks are treated. Transferring risk may still leave the organization open to a range of legal actions (in terms of failing to take all reasonable steps to prevent harm) or to a loss in terms of branding, reputation, and so on.
- *Accepting* the risk, where those accountable have made an informed decision that the level of risk to the AIC of operations does not conflict with the organization's requirements, nor does it represent potentially unacceptable losses. According to Haimes and Chittester (2005, p. 2), "The level of required information assurance, or conversely the level of acceptable risk, depends on the critical nature of the system's mission," which maps back to the section on mission analysis.
- *Avoiding* risks through changing locations of operations that place adequate time and distance between the operations of the organization and identified key threats, so as to make them less relevant.
- *Ignoring* the risk by choosing to reject the arguments offered by trained, educated, and experienced AP&S risk analysts. This is never considered to be prudent or demonstrative of due diligence, both necessary qualities of senior management. This approach could lead to legal issues such as negligence or failing to act in line with an appropriate duty of care.

The concept of the span of control also factors significantly in terms of determining how the organization wishes to respond. Where there is adequate span of control, the organization may decide to act unilaterally and inform its various stakeholders. This is efficient and, as long as the advice of trusted and capable AP&S analysts is taken, the most effective course of action. As this span of control diminishes, such as would happen where an agreement exists regarding the use of distributed and networked assets, the restrictions on unilateral freedom of action decrease.* This is where carefully defined and crafted agreements become important, as they reduce the potential for friction among interested or implicated organizations that can occur where expectations are less than clear. Where there is little more than a span of awareness, the organization may be limited to taking steps to learn more about potential risks so that cogent arguments can be made to influence, and then control, treatment of risks. In all cases, however, the degree of control that can be exerted is a factor of capacity to respond effectively to the identification of risks and implement appropriate controls.

Managing more complex risks

Part of the value in taking a formal and deterministic risk management approach lies in the ability it gives security practitioners to put forward consistent and understandable recommendations to senior decision makers regarding the management of risk, regardless

* This is perhaps most prevalent in NCIs, with multiple ownership, operational responsibilities, distances involved, and complexity of architectures.

of how complex, complicated, integrated, new, or diverse. Often, it may be a simple case of reiterating the regulatory or policy requirements to comply with relevant and appropriate best practices. This compliance, however, should not be interpreted as leading to effective or appropriate security in the wider sense, since compliance with baselines is the lowest form or protection; there will typically be peculiar threats and vulnerabilities that are not addressed adequately by general baselines. These are identified and assessed in a TRA, so additional safeguards would be based on that same TRA. This is the essence of threat-risk-based security. Baselines may provide overprotection in some cases, but in many more cases provide underprotection. It is in analyzing the delta of protection requirements and proposing risk-based safeguards that the AP&S practitioner provides the real value added to a protection posture.

Compliance with baselines as a risk management approach is safe and defensible by security managers ("I was just following policy"), but does not provide the value added, or expected, by accountable senior management. It may demonstrate "institutional" due care for assets, but in most cases not appropriate due care, given the diverse threats and vulnerabilities in many systems and enterprises. While the line manager may escape scrutiny with this argument, the senior managers will not. Although a rules-based compliance approach to AP&S addresses known and set questions and then applies predictable, sound, proven generic controls to address known and generic (if not current or emerging) threats and vulnerabilities, in many contexts this approach would itself constitute a vulnerability, because it introduces a gap in analysis. It does not allow for the identification and analysis of new missions, assets, threats, or vulnerabilities that can lead to risks. And, since compliance-based safeguards are typically open-source industry best practices, they will be well known by an adversary, who can study and analyze them to determine the best threat vectors (routes to the asset), strategies for vulnerability exploitation, and specific targets of an asset in terms of AIC; for example, destruction of a production line, denial of service attack on a SCADA system, corruption of data through masquerading, or stealing company secrets. It also leads to an attacker being able to engineer his or her way through the existing baseline safeguards—understanding that attacks need not always be technical, since social engineering may have a greater potential for attack success if baselines only are employed. Security awareness programs mandated by baselines are typically not current, not taken seriously, nor is it assured that all employees participate if a threat-risk-based approach is not implemented, because there will be little new or captivating threat or vulnerability information to pique their interest. If it is relatively certain that a company has not implemented threat-risk-based safeguards above baselines, then that company increases its susceptibility to attack, since it is seen as a weak link.

Complex risks may be described as those that feature the following:

- Emerging technology as the attack vector or as the target.
- Multiple and diverse threat sources; for example, a physical, social engineering, and concurrent cyberattack, or a distributed denial of service attack.
- Extreme motivation and disregard for collateral damage on behalf of the threat agent; for example, terrorists, criminals, the deranged, state-sponsored actors, or the excessively greedy. These risks could result in extensive property damage or contamination.
- Multiple and diverse assets targeted, perhaps concurrently.
- Multiple offices or production facilities targeted, perhaps concurrently.

Complex risks require complex analysis by well-trained and capable AP&S analysts, preferably those who have the trust and authority of their senior management to conduct

extensive, often intrusive, and normally time-consuming analysis. Complex risk analysis also typically requires extensive coordination and liaison among stakeholders at all levels; this will require authority from senior management to "sidestep" routine (and bureaucratically inefficient) chains of command or reporting relationships. Trust by senior management in the technical, operational, and corporate capability of the risk analysts is essential for complex risks to be addressed adequately. Both AP&S practitioners and line managers can collaborate and actually break the chain of events that lead up to complex risks.

Consider a basic cyberattack on a discrete (unconnected) computer network such as a traditional SCADA system. This attack may be broken down into a series of steps, much like the processes used by the organization's own operations, and may include the following mental analysis on the part of the adversary:

- I must be able to identify where the system is housed and gain some level of access to it.
- I must determine if the assets that I want or those that I want to impact are actually there, and if the attack will meet my objectives.
- I must confirm the level of protection that is afforded those assets and if that level of protection changes with time or other factors.
- I must be able to pass through the perimeter controls, typically comprising a fence and a guard post, perhaps with some closed-circuit video equipment.
- I must be able to get into the building, hopefully without alerting anyone.
- I must be able to get past the receptionist (perhaps using social engineering).
- I must be able to gain access to the restricted area in such a way that I remain undetected for 15 min, which I estimate is required to launch the attack.
- I must be able to turn on one of the workstations.
- I must be able to use my cracking tools on the workstation to escalate my privileges and gain access to the files that I want to steal or corrupt to the operating systems or applications that I want to infect or change.
- I must be able to locate the files.
- I must be able to download the files without being detected or that provides me with 10 min before a response is made so that I can escape.
- I must be able to leave the restricted area with my USB key without being detained.
- I must be able to leave the facility.
- I must be able to download the file from my own computer.
- I must be able to break through any encryption placed on it.
- I must be able to exploit this information for my own purposes.

In thinking like an adversary and decomposing an attack into individual threat vectors, the AP&S risk analyst can isolate

- The business processes that could be affected
- Intermediate or final assets targeted
- Types of complementary or contributing threats that could be brought to bear
- Different vulnerabilities that may be exploited in isolation, concurrently, or in succession to bring the attacker closer to the targeted assets

This case study is not intended to be an in-depth coverage of safeguards, but is, rather, an illustration of how risk management processes can be effective if utilized by capable practitioners in a deterministic manner. From this decomposition, there emerge several points along the threat vector where the attack can be disrupted. For example, the

attacker may have to pass through physical access control points at various stages of a layered defense that would prevent him or her from ever reaching the computer terminal. Similarly, even if the adversary makes it to the terminal, the USB ports can be disabled as part of workstation hardening to prevent the use of removable media. The terminal might involve technical controls, such as strong identification/authentication procedures that do not allow a terminal to operate unless the username and a complex, routinely changed password are entered. There may be a program of random searches of the person to prevent the unauthorized removal of media. And the list goes on. By fully understanding how the attack is likely to take place given the nature of the threat, the next step is to reduce vulnerabilities through the manipulation of means, opportunity, and motive or intent for the threat agent to act. The organization may also seek to manipulate the adversary's perception of the asset value through implementing stringent safeguards; for example, requiring highly sensitive documents to be stored on-site only on hard media, copied to prevent destruction, and stored off-site in secure locations after being strongly encrypted, requiring special software to open them. By manipulating the values of assets, threats, and vulnerabilities, risk analysts can either break the attack chain or reduce the impacts associated with an attack to acceptable levels.

This decomposition approach for complex risks also allows for a degree of efficiency to be realized. By comparing various threat models and vectors, analysts can identify overlaps that could allow the organization to apply a single safeguard that mitigates a number of different threat vectors. Some care must be taken to ensure that there is an appropriate balance of redundancy and resiliency (key elements in establishing layers of defense) in the security controls on the one hand, and efficiency and minimization of inconvenience on the other. In essence, the security practitioner must be able to work across the various communities in his or her organization to balance not only an appropriate number and type of controls, but also an appropriate level of operational impact within the organization. What is important is that doing nothing is not a preferred option when the mission is important and when valued assets are involved. Regardless of whether the threat is natural, deliberate, or accidental, action is preferred. This also applies to deterioration as a threat. Monitoring of deterioration of a facility or infrastructure and assessment of its extent drives one of three management decisions: do nothing, rehabilitate, or replace (Morcous et al., 2003). Maintaining current inventories, infrastructure condition databases, and maintenance data, along with having trained inspectors follow inspection intervals consistent with projected deterioration rates, are essential to addressing deterioration. These can all be considered programmatic activities, and are indicative of the components of an effective risk management program.

Risk management: Pulling it all together

In the management of risk, we have looked at the risk assessment and management processes in detail and then identified how those various elements interact. This interaction is important, not only in determining the nature and level of risk, but also in terms of later analyzing different attack vectors (a threat plus the route that it takes to exploit a vulnerability) that can be subjected to certain safeguards or controls so as to deter or disrupt the attack. Having identified these points, the concept of spans of control has been introduced in terms of the organization's ability to add, change, or remove factors that can impact the likelihood or gravity of a threat event. Finally, we have looked at communicating risks (including their elements) to overcome the challenges associated with analyzing threat events that cascade through systems or that escalate toward higher levels of impact. The

next step is for the practitioner and management to decide on the controls that will be considered appropriate to the identified risk, and that mitigate risk to a level acceptable to senior management in terms of operational impact and tolerable in terms of social and cultural norms. Hentea (2008, p. 4) refers to this as "the process of assigning priority to, budgeting, implementing, and maintaining appropriate risk reducing measures." In all cases, it is senior management who ultimately decide the safeguards that are implemented and who is accountable for the residual risk to operations.

References

ASIS International. (n.d.). *Protection of Assets Manual*. Available online. Alexandria, VA: ASIS International.

Bobbio, A., Bonanni, G., Ciancamerla, E., Clemente, R., Iacomini, A., Minichino, M., and Zendri, E. (2010). Unavailability of critical SCADA communication links interconnecting a power grid and a telco network. *Reliability Engineering and System Safety*, 95(12), 1345–1357.

Cardenas, A. A., Roosta, T., and Sastry, S. (2009). Rethinking security properties, threat models, and the design space in sensor networks: A case study in SCADA systems. *Ad Hoc Networks*, 7(8), 1434–1447.

Chittester, C. G. and Haimes, Y. Y. (2004). Risks of terrorism to information technology and to critical interdependent infrastructures. *Journal of Homeland Security and Emergency Management*, 1(4), 1075–1075.

Gold, S. (2008). Look after your heart. *Infosecurity*, 5(8), 38–42.

Haimes, Y. Y. and Chittester, C. G. (2005). A roadmap for quantifying the efficacy of risk management of information security and interdependent SCADA systems. *Journal of Homeland Security and Emergency Management*, 2(2), 117.

Hentea, D. M. (2008). Improving security for SCADA control systems. *Interdisciplinary Journal of Information, Knowledge, and Management*, 3(12), 4.

Kurtz, J. (2004). Chapter D–D1/D2–Dynstat apparatus. *Dictionary of Civil Engineering: English–French*. New York: Kluwer Academic/Plenum.

Lowrance, W. W. (1976). *Of Acceptable Risk: Science and the Determination of Safety*. Los Altos, CA: William Kaufmann.

Morcous, G., Lounis, Z., and Mirza, M. (2003). Identification of environmental categories for Markovian deterioration models of bridge decks. *Journal of Bridge Engineering*, 8(6), 353–361.

Patel, S. C., Graham, J. H., and Ralston, P. A. S. (2008). Quantitatively assessing the vulnerability of critical information systems: A new method for evaluating security enhancements. *International Journal of Information Management*, 28(6), 483–491.

Patel, S. C. and Sanyal, P. (2008). Securing SCADA systems. *Information Management & Computer Security*, 16(4), 398–414.

Stoneburner, G. (2006). Toward a unified security/safety model. *Computer*, 39(8), 96–97.

Zhu, B. and Sastry, S. (2010). SCADA-specific intrusion detection/prevention systems: A survey and taxonomy. In *Proceedings of the First Workshop on Secure Control Systems (SCS)*. Stockholm: Team for Research in Ubiquitous System Technology.

chapter five

International implications of securing our SCADA/control system environments

Vytautas Butrimas

Contents

Introduction

As someone* occupied with government information technology (IT) and national security policy for the past 25 years, I have worked in a changing cyber security environment that started from dealing with the first hackers invading our IT systems with viruses such as the "Michelangelo" virus of 1991 to worrying about cyber criminals, socially motivated hacktivists, and possible activities of "cyber terrorists" to state-sponsored cyber attacks, not limited just to IT systems. The appearance of Stuxnet† and the "denial of computers" attack perpetrated against energy company Saudi Aramco strongly indicated that critical infrastructures (CIs) that support national economies and the well-being of modern society were now new targets for cyber attacks. Additionally, the extensive expansion of the capabilities of modern industrial control systems (ICSs) made possible by the advances in information and communication technologies (ICT) and their application to the management of complex systems running CIs has introduced, together with increased efficiencies and cost savings, serious dependencies, and vulnerabilities. Vulnerabilities that, due to a lack of understanding of the interrelatedness of increasingly complex systems, have given rise to unintentional incidents. Vulnerabilities that, if known by the "bad guys," may be exploited to execute intentional cyber related attacks, attacks which are now possible due to the entry of IT into the formerly isolated and proprietary world of ICSs (supervisory control and data acquisition [SCADA] systems). The new threats emanating from cyber space have provided new and broad challenges that range beyond the national level to the international level. CIs today have a cross-border or international dimension. Failure at a national level can affect a connected neighboring country. While some worthy and effective efforts are being made by national governments and industry in terms of laws, regulations, and standards, they fall short in meeting the international dimension of today's cyber threats. SCADA and ICS environments can no longer be considered safe from today's dynamic threats emanating from cyber space. This chapter will address implications of any changes to cyber space environments that have taken place within the last few years that now require international responses in the form of self-restraint, acceptance of responsibility, and cooperation. Possibilities for moving forward into the future—at an international level—will also be discussed.

In the past 5 years, a new concern has developed for the cyber security of CIs belonging to the energy, transportation, water, manufacturing, and telecommunications sectors. The public appearance of Stuxnet in 2010, and its subsequent analysis, revealed it to be a nation-state manufactured cyber weapon targeting specific control systems belonging to CIs, indicating that the cyber security environment had changed in a significant way. Up until then, the protection of the confidentiality, integrity, and availability of electronic data generated, transmitted, and processed in information systems was the key focus of the work of IT security professionals. The threats emanating from cyber space used to be a collection of "the usual suspects"—cyber espionage, cyber crime, and computer hackers. Policies were developed to ensure security of the chosen objects which needed to be protected from the perceived threats emanating from cyber space.

* Evaluations and ideas presented within this chapter exclusively belong to the author and is not considered an official position of the Ministry of National Defense of the Republic of Lithuania or any institution with which the author is associated.

† Stuxnet is cyber related malware discovered sometime in June 2010, and was designed to render industrial programmable logic controllers (PLCs) inoperable; in this case, the malware specifically targeted Siemens Series 7 devices.

From 2001 to 2011, I was responsible for information and communication security (INFOSEC/COMSEC) at the Communications and Information System Service under the Ministry of National Defense in Lithuania. The years 2001–2004 were especially intense, as my service was tasked with the IT and telecommunications work to join NATO in 2004. We had to demonstrate that we had fully implemented INFOSEC/COMSEC (later to be called "information assurance" and popularly called "cyber security" today) policies before anything was allowed to connect our national IT and communications infrastructure to NATO's systems. However, no NATO security policy or anything in my experience up until then ever provided even an inkling of an idea that there were other CIs that were just as vital to national defense and our ability to perform as a member of the NATO alliance. In late 2010, I was asked to write about the state of the cyber security of our energy infrastructure. It was assumed that this would be an easy task, thinking that IT systems were the same as those used in the energy sector. Needless to say, the writing of "The cyber security dimension of critical energy infrastructure" (Butrimas and Bruzga, 2012) proved to be both a humbling and very enlightening professional experience. CI protection is not just about information security and protecting documents, but about the reliable and secure monitoring and control of the real-time processes found in the energy, transportation, utility, and manufacturing sectors vital to the economy and the well-being of society. While disruptions to information systems could lead to one form of danger, disruptions to the control systems of CIs were potentially far more serious to a nation's national security and could affect other nations as well. Failure of an electric grid, gas pipeline, or traffic control system had cross-border or international ramifications that required international cooperation. This all seemed clear enough to me after writing the white paper; however, it was soon realized that there was a problem. This realization was not shared among any of my IT and security policy colleagues in government nor with colleagues in any other governments. Cyber security was basically understood in terms of the confidentiality, integrity, and availability of electronic data found within the information systems of governments, banks, businesses, public websites, and the computers of private individuals. The vulnerability of CIs and ongoing processes found within, for example, the electric grid used to supply power for those information systems and computers to operate were not within scope of government security policy makers. In fact, to those involved in developing government cyber security programs and strategies, ICS did not even exist. In discussing what needs to be protected, the term critical *information* infrastructure was used. Nobody seemed aware of the other critical (non-information-centric) sectors that were vulnerable to intentional and unintentional cyber incidents.

The alarm which should have sounded in the international community after the first appearance of malicious state cyber activity directed at the CI of another state went unheeded. This situation, with some exaggeration, is similar to what would have happened had the world continued to concentrate on fighting organized crime while ignoring the invasion of Poland or the Japanese surprise attack on Pearl Harbor. Governments did take steps at the national level to address the newly exposed vulnerabilities of CIs to cyber incidents and attacks. The U.S. Department of Homeland Security (DHS), in addition to establishing a national computer emergency response team, US-CERT, also created a dedicated CERT for ICSs. While national-level efforts were underway, the international borderless dimension of cyber space required new efforts at the international level to deal with vulnerable cross-border interdependencies exposed by this new threat. However, the efforts at the international level on cyber security policy among states and within international organizations continued to focus on dealing primarily with cybe rcrime and the antics of socially motivated hacktivists. This resulted in a dangerous gap between

efforts to formulate national policies and efforts to formulate a comprehensive international cyber space and security policy. Recognizing and dealing with this gap has created a very broad challenge, not only at the national level for industry and government, but internationally as well. National efforts were not enough in protecting a system that was interrelated and interdependent with other systems in cyber space. Gas pipelines, power grids, and submarine communication cables today cross borders and reach across to other continents. A failure in one section of the grid can ripple and cascade across to affect other networks and systems belonging to CIs in other countries. Additionally, those same national systems are vulnerable to external attacks originating from other parts of the world. Management of these global-level complexities can only be done through international cooperation.

2010–2014: The security environment of cyber space changed for CI and ICS: Stuxnet, Saudi Aramco, Snowden, Havex-Dragonflies, and Bears Oh My

In terms of the cyber security of ICSs and the CIs they support, we live in a "post-Stuxnet world" today. One may ask what is so unique about this malware that was discovered years ago when thousands of new pieces of malware are discovered every day? Without going into technical descriptions (Langner, 2013), Stuxnet was the first publicly known (Russell, 2004) nation-state-developed malware which was specifically targeted against the control system of a critical industrial process. The malware effectively deprived operators of the "view" and "control" of centrifuges belonging to a controversial uranium enrichment facility. It achieved this by intercepting and inserting false data sent to the operators telling them that systems were functioning normally, when, actually, they were not. To put it more simply, the effect was similar to what would happen to a driver of an automobile whose mechanisms were manipulated to steer the car over a cliff. The driver feels no alarm nor reason to take action since the view of the road they see ahead is "normal." Even if they tried to take action to save themselves, they would find that they had no control of the steering wheel, brake pedal, or engine.

The appearance of Stuxnet can be said to be the equivalent of a "Hiroshima moment" for cyber security and international relations. The first known execution of a cyber attack by one nation-state against the CI of another nation proved that conflicts among states were now being executed in the cyber domain. It was recognized that this technology was now being applied to disrupt and destroy machinery and industrial processes. This operation, which was probably politically motivated (to keep Iran from making atomic weapons) also introduced a new problem of cyber weapons coming into the hands of lesser-skilled hacktivists, criminals, and even terrorist groups (Simonite, 2012). Unfortunately, the Stuxnet code made it to the Internet where it could be freely copied and analyzed. The methods could be studied and the code adapted to execute new and destructive cyber attacks. The makers of Metasploit also seem to have taken notice of Stuxnet, as new versions now have modules that apply to ICS (Selena, 2012). CIs that were, up until then, largely living in their own isolated world of closed communications networks and obscure proprietary technologies became a new area of interest for hackers. For example, SCADA/ICS began appearing at popular hacking conferences as a topic of interest. The website of Black Hat Asia 2014 featured a course on "attacking SCADA" at the top of the list. The course is intended to "provide students with the knowledge that they need to safely perform penetration testing *against live SCADA environments"* (Parker, 2014). Shodan is also being used to search

for connected ICSs. One individual has even published, in their Twitter feed, screenshots of control system workstation panels that they stumbled on. They can also be accessed on Google,* and provide inspiration for others to seek out and even try to "touch" the controls of a critical system exposed on the Internet.

Not just hackers and governments were seeking ways to exploit the newly exposed vulnerabilities and do physical harm to ICSs of national CIs. For the first time, it was plausible to think about the possibilities of true "cyber terrorism." This technology was now available to terrorists groups lacking the skills to develop their own cyber weapon of mass destruction (WMD). The apparent success of the Stuxnet operation contributed to not only a new recognition of the vulnerability of CIs, it also provided the international security policy community with a new problem: what to do about nation-states playing cyber games with each other's CIs?

The implications of this new form of malicious cyber activity should not be lost on anyone. It raises the issue of whether one can trust the safety and reliability of systems used to monitor and control critical processes that are now so vital to our economies and societies' well-being. This is far from the concerns raised by distributed denial of service (DDOS) attacks on websites executed by hacktivists, the theft of financial information by cyber criminals, or the stealing of industrial secrets by industry competitors or spies. At the same time, it must also be remembered that the impressive work that led to the development of Stuxnet was supported by espionage and intelligence assets that only a nation-state could have provided. Details of the operations of the targeted facility had to be fully understood to develop and execute the cyber attack that was Stuxnet, and to ensure that the attack would not execute anywhere other than at the facility that was targeted. As one commentator on Stuxnet has said, the intelligence was so good "they knew the shoe size of the operators working at the plant" (Langner, 2011). One asks the question: is it OK to allow this kind of malicious cyber activity to continue without some kind of international response to punish the perpetrators, or at least agree on some rules of the game? International criticism of the Stuxnet operation was muted. Perhaps some thought it served some useful purpose in keeping Iran from making an atomic weapon. What is little appreciated is that the majority of potential targets for Stuxnet-type attacks are not in just the Middle East, but in the developed countries found in Europe, North America, and parts of Asia that are developing modern CIs—potential targets that are far less protected (not located in underground facilities) and more vulnerable (more possibilities for penetration) to Stuxnet-type attacks.

Saudi Aramco

In December 2012, another nation's CIs were cyber attacked. Saudi Arabia's oil company, Saudi Aramco, experienced a targeted cyber attack on its computer systems. This cyber weapon, called Schamoon, succeeded in executing a "denial of computer" (DOC) attack, wiping clean over 30,000† computer hard drives belonging to servers and workstations. The attack appeared to have been limited to the administrative part of the company and not the CI parts involved with the production and processing of oil. Although the attack did not affect the ICSs, it did cause havoc for the management of the business of the company. Even pipeline operations are dependent on management's world of contracts and timetables. As one commentator said, the

* https://www.google.com/search?q=dan+tentler+shodan+screenshots&rls=com.microsoft:lt-LT&tbm=isch&tbo=u&source=univ&sa=X&ei=RNBZVO_vH-eM7Abr8oAQ&ved=0CEEQsAQ&biw=1323&bih=662.

† Curiously, 30,000 is the same number used in describing a similar attack that occurred in South Korea.

company must have had a difficult time without this special information when there were orders to be processed and tankers waiting in the harbor when this attack occurred (Eugene Kaspersky Press Club, 2013). For the Saudis, this cyber attack was taken as an attack that threatened not just its critical energy infrastructure but its economy (AL Arabiya News, 2012). Although there was no conclusive proof, it was suspected that another government's cyber power was responsible (Perlroth, 2012a). The lack of international response further reinforced the message that cyber attacks are an attractive and highly effective tool to inflict damage on an adversary at low cost in terms of liability, preparation, delivery, and minimal collateral damage. The problem is getting worse, as there were indications that these attacks were counterstrikes in retaliation for earlier attacks (Perlroth, 2012b).

From Snowden to Sandworm

The next key event indicating a change was taking place in the cyber space environment occurred a few months after the Saudi Aramco cyber attack. This was the revelation of government electronic spying and surveillance by former U.S. National Security Agency (NSA) employee/contractor Edward Snowden that began in May 2013. Taking aside the issues of the breaches in the privacy of persons and government leaders which were raised by Snowden's revelation, it is the intelligence gathering and surveillance capabilities possessed by governments exposed by Snowden that are worthy of comment here. The revelations indicate that government capabilities include possibilities not only for passive measures to collect intelligence information, but also for active measures once a system's software or hardware has been penetrated by one of the catalog of available tools (Applebaum et al., 2013). These capabilities go beyond just massive monitoring of worldwide telecommunications traffic, but also penetrating the hardware and software supply chain, making the offensive capabilities truly worldwide in scope. If one recalls how much intelligence was required to develop and execute Stuxnet, the capability to develop a successor is more than feasible. In fact, it may be just too tempting not to do so, especially if one is the leader of a nation whose efforts to achieve a foreign policy objective by traditional means is continually being frustrated. The possibilities of making use of these intelligence gathering and surveillance capabilities combined with the proof of concept that was Stuxnet make for a very dangerous "cyber cocktail" capability with implications for the future of the cyber security of ICSs. The "Eye of Sauron" (to paraphrase *The Lord of the Rings*) has focused its attention on ICSs. Just one example of this at the time of this writing (November 2014) comes from one of the first published analysis of "Sandworm" which indicates that "Sauron's Eye" is looking for where ICS equipment manufactured by GE and Siemens is located (Hultquist, 2014). This reflected similar activity that most likely took place during the development of Stuxnet to meet the specifications of its Siemens-based warhead and location of its intended target.

Havex/Dragonfly/Energetic Bear

The cyber event of the summer of 2014 was the Havex (aka Dragonfly, Energetic Bear) malware attack. It illustrates an unsettling trend regarding a new ICS attack vector (which, in this case, are watering-hole attacks* on vendor websites) and the sinister nature of cyber

* A "watering-hole attack" is an attack method used against a target of a specific group (organization, industry, or region). Through this method of attack, the attacker guesses or observes which websites the group most often uses and infects one or more of those websites with malware; the eventual outcome is that one or more members of the targeted group become infected with the malware.

espionage. According to reports from the DHS industrial control systems computer emergency response team (DHS ICS-CERT), (Alert, 2014) this reported malware targets the software/firmware download websites of manufacturers of industrial control systems. Compromised vendor software that customers download from these sites may allow attackers to access customer networks, including those that operate CIs. Commentators are comparing this malware to Stuxnet, as they also indicated that the sophistication demonstrated and the choice of target pointed to nation-state involvement (Perlroth, 2014). This is really bad news, especially to those in the energy industry and other sectors of CIs. One respected colleague in the ICS world commented that "this is the tip of the iceberg." The news gets worse. According to an analysis conducted by Symantec, this malware not only provided a platform for conducting cyber espionage activities, but also provided the "attackers the ability to mount sabotage operations against their victims," and if the attackers had used the sabotage capabilities available, "could have caused damage or disruption of the energy supply in the affected countries" (Symantec Security Response, 2014). This should cause many who tend to accept cyber espionage as being part of traditional spying to pause and consider its ramifications. In cyber space, the cyber spy wears two hats. To remove the spy hat and put on the saboteur black mask only requires the press of the <ENTER> key. This is not about the spying of Mata Hari; it is about the activities of the cyber space spy/saboteur equivalent of James Bond. If James Bond gets the order to kill someone, he will, and has all the resources of the state and "Q" to help him inevitably succeed in his given mission.

In terms of the cyber exercise in which I participated in 2012, one of the difficulties encountered was finding ICS specialists to deal with solving a problem presented in the scenario. Specialists from traditional national CERTs with Microsoft Windows, Cisco, and Linux certifications were available, but what was lacking were ICS specialists with engineering diplomas who were more familiar with the affected equipment.

The problem of unintentional cyber incidents in CI

It is not enough to worry about protecting critical systems from intentional cyber attacks. Many readers of this chapter perhaps are also aware that unintentional cyber incidents also take place within ICS space. One of the causes of unintentional cyber incidents in ICSs comes from the great success in terms of better management and cost savings coming from digitalization of control equipment and entry of IT into ICS environments. IT's strengths of automation and remote management have allowed for the creation of complex systems of systems, providing integrated services over a wide territory. However, together with the good side of all this, there is also a bad side in terms of new vulnerabilities and potential points of failure. The term "cyber fragility" has been used to describe this situation in much depth by Ralph Langner in his book, *Robust Control System Networks* (Langner, 2012). The IT security professionals coming to work in ICS environments are new to this environment, and do not always understand the ICSs they are hired to secure in the same way that the ICS engineers who designed and operate them do. This false sense of "I know what I am doing" can have surprising and potentially dangerous outcomes. A good example is the emergency shutdown of the reactor at the Hatch nuclear power plant in 2008, which was caused by a software update on a single computer belonging to the control system. This was a complete surprise for the administrators, who had to question whether they were adequately knowledgeable about their operating environment to do their jobs. There are other surprises to consider coming from honest software programming errors.

Dangers of programming errors affecting the heart of cyber space: Heartbleed and Shellshock

In addressing vulnerabilities arising from the complexity of modern ICSs, there is also the issue of software programming used to enable our use of cyber space. The Heartbleed bug is a programming error in a popular OpenSSL library that is used for providing cryptographic services such as Secure Sockets Layer/Transport Layer Security (SSL/TLS) used to ensure secure communications over networks (Heartbleed Bug, 2014). "By attacking a service that uses a vulnerable version of OpenSSL, a remote, unauthenticated attacker may be able to retrieve sensitive information such as secret keys. By leveraging this information, an attacker may be able to decrypt, spoof, or perform man-in-the-middle (MITM) attacks on network traffic that would otherwise be protected by OpenSSL" (Homeland Security, 2014). If one considers that this vulnerability, which was exploitable for 2 years before it was discovered, was bad enough, how about a programming error in another vital part of cyber space management that was only discovered after 20 years. This is what happened with the discovery of the Bash shell vulnerability popularly called Shellshock. This is about a vulnerability discovered in the Bash interface shell used to access the depths of operating systems. Personal computers that used Bash could be subject to attacks using this vulnerability. However, this shell program is also used in networks that monitor and control processes found in CIs (Saarinen, 2014a). Bugs and patches to these well publicized vulnerabilities resulting from programming errors of long ago soon became available after they were disclosed. The problem is that it is likely that many more such unknown errors are waiting to be discovered. The vulnerabilities yet to be discovered in the software that runs our critical systems seem to be endless. Microsoft issues vulnerability patches every month. In November 2014, it issued a record number of fixes during its "Patch Tuesday" (Saarinen, 2014b).

One of the most important things to remember in terms of unintentional incidents stemming from cyber fragilities of ICSs in the context of this article is that knowledge of these vulnerabilities can be used by the "cyber samurai" to plan and execute cyber attacks on ICSs. Attacks and incidents that perhaps occur unintentionally are difficult to investigate due to a lack of ICS forensic capabilities. As ICS industry opinion leader Joe Weiss indicated that a major cyber incident in ICSs is likely to happen; however, we will probably never know whether it was achieved with malicious intent or not (Elinor, 2010).

Something wrong at the international level (United Nations and European Union) in terms of dealing with changes to the cyber security environment

One would think that for the international community there have been enough alarms and wake-up calls for action to be generated. What has been their response to the examples of malicious cyber activities of states listed above? In September 2010, I attended the UN-mandated Internet Governance Forum (IGF) in Vilnius. In the midst of concerns to preserve privacy and open access to the Internet, there was no attention given to some of the unsettling events occurring in cyber space during the previous 5 year mandate of the IGF. Estonia pulled out its national Internet plug after it had experienced a cyber attack in 2007, and, later, cyber was attacks were used to compliment a traditional armed attack during the Russian–Georgian war of 2008. News about Stuxnet had first appeared in IT professional circles 4 months earlier. Regardless of these unsettling actions, indicating that

nations were engaging in malicious cyber activities, the IGF meeting simply concentrated on concerns of digital rights of privacy and universal access to the Internet. In the fall of 2013, the European Union held its Information and Communication Technologies conference (ICT 2013) in Vilnius. Once again, there was very little appreciation for what had happened in cyber space during the previous two summers (e.g., cyber attack on Saudi energy company Saudi Aramco in 2012 and revelations on the extent of government electronic spying and surveillance in 2013). Stuxnet, the attack on Saudi Aramco, and Mr. Snowden's revelations about the large-scale surveillance activities of governments raised serious security issues for the international community to address. However, in response, very little was being done about it in international fora and by organizations created to promote international security and peace. For some reason, perhaps thinking that some other organization will tackle the problem or just not being aware of what was happening, these fora were not considering the unsettling trends in cyber space. I suspect that what is really missing is an appreciation of the technological implications of the dynamic threats to ICSs emanating from cyber space. The UN-appointed Group of Governmental Experts on Developments in the Field of Information and Telecommunication in the Context of International Security is dominated by ambassadors and diplomats, working mostly for arms control agencies and ministries of foreign affairs. No one who can be called an ICS opinion leader who could address the implications of IT technology on modern ICSs is listed in the annex list of group members. I will return to the issue of IT bias in terms of dealing with the cyber security of ICSs later in this chapter.

A few words about Internet governance, the multistakeholder myth, and the ITU

The UN's International Telecommunications Union (ITU) organized the World Conference on International Telecommunications (WCIT) at the end of 2012 in Dubai. This was a most interesting conference, as the ITU tried to foster some updates to the way world telecommunications were to be regulated. For example, there were proposals to update the regulations to include something that was missing from the last time the regulations were approved in 1989: the Internet. While the WCIT meetings failed to reach an agreement on an updated set of telecommunications regulations to cover the Internet, it illustrated another issue: the growing divide between East and West in regard to Internet governance issues. It was evident that there was a growing concern among non-Western nations—in particular Russia and China—over the West's domination (in particular by the United States) of the way the Internet was managed. Democracies tended to support a multistakeholder approach (minimal government involvement) to Internet management, while more authoritarian governments sought more government controls over content and use. While Internet freedom advocates were joyous over the failure of the "UN to take over the Internet," (Klimburg, 2013) a dangerous split remained between East and West over the management of cyber space (Gewirtz, 2012). It also represented another failed opportunity by the international community to raise and deal with the issue of the malicious activities of states in cyber space. The West's position in favor of a decentralized "multistakeholder" approach to Internet governance sounds dishonest in the face of the malicious cyber activities of states in cyber space. Some of the very same states fighting to keep the decentralized multistakeholder model were also taking advantage of the assumed trust behind this system of governance by engaging in malicious behaviors in cyber space. This was similar to cowboys saying that they should be free to roam the prairie without the restrictions that could be enforced by sheriffs, while, at the same time, they engaged in cattle rustling. The multistakeholder model depends very much on a certain degree of trust among the

stakeholders. The model will be discredited further if stakeholders choose to take advantage of this trust by engaging in cyber misbehavior, as seen in the case of the penetration of Belgacom. This serious intrusion on a key telecommunications provider in Europe and a major manager of international submarine cables has been linked to the work of a friendly nation that promotes multistakeholderism (Koot, 2013). Those that argue that the ITU should stay out of Internet governance are sounding more like outlaws calling for fewer sheriffs. Maybe it is not such a bad idea that the ITU is trying to address a problem that is not being addressed elsewhere—the malicious activities of states in cyber space. Those that are against the ITU's efforts at governance seldom indicate which alternative organization should address this issue.

Approach of NATO

In the last week of April 2007, I attended a NATO Cyber Security Workshop co-sponsored by the U.S. Department of Defense (DoD) and Microsoft held in Redmond, Washington. It was an excellent workshop for becoming familiar with NATO's approach to cyber security and vision for the way ahead. It also featured presentations from Microsoft on the virtues of the recently released Windows Vista operating system for the military. Microsoft also announced its Government Security Cooperation Program and invited NATO member governments to join. There were interesting aspects to the announced program. It was revealed that China had just signed (Russia had previously signed) on to the program and had been given access to Microsoft operating system source code. Although it was mentioned by this author, there was no reaction to the apparent contradiction between publicized cyber incidents associated with these two countries and providing them with access to one of the most popularly used and bug-filled operating systems in the world. Later, a presenter from Estonia came up to the podium and announced that he would depart from his planned presentation because "my country [was] under cyber attack." There we were, all the top NATO cyber security practitioners sitting in one place; yet, upon hearing this announcement, we could only look at each other in amazement. No one had any idea what to do, since there were no official policies or agreements in place that could address what had just happened. Later, NATO did come up with a cyber defense concept and offered to sign memoranda of understanding (MoU) with individual member states that included the possibility of sending "rapid reaction teams" for cooperation in cyber defense, which Lithuania signed in 2010.

This meeting of private industry, government, and NATO illustrated a lack of a comprehensive and coordinated policy toward cyber security. Microsoft's providing access to its operating system source code to nations with bad reputations for abusing cyber space seemed to contradict efforts being taken to improve the security of cyber space by Microsoft and the workshop participants.

NATO and other international organizations have different understandings of what needs to be protected and from what cyber threats. Protecting communications and information systems from cyber attacks by establishing CERTs is not enough to deal with protecting what is truly critical from the threats emanating from cyber space today. A good example that illustrates this comes from a 2014 summer conference held in Vilnius, commemorating Lithuania's joining NATO in 2004. I asked Mr. Sorin Ducaru, the NATO Assistant Secretary General for Emerging Security Challenges, "Has NATO evaluated what would happen to its ability to perform its mission if the critical infrastructure that it and member states depend upon to function was degraded by a cyber incident or cyber attack?" To illustrate, I reviewed what happened to the Carmel tunnel in Israel (part of the

main highway to the seaport of Haifa) in the fall of 2013 (Hamadia, 2013). The operators were forced to close the 6 km tunnel for 2 days because a cyber attack knocked out the tunnel video camera surveillance system, fire control, and air-conditioning systems. This was not a "denial of service" but rather a "loss of view and loss of control" of critical processes required to ensure safe and efficient operation of a tunnel. I asked him to imagine the impact, in terms of a military operation, of the closing of a key transportation link for a military convoy of supplies that is forced to stop and wait for a tunnel to be declared safe? What effect would such a delay have on a nation's ability to participate in a mission and how would that effect NATO's operations? A tough question, and perhaps too tough to answer in a question and answer session after a long day. However, the Carmel tunnel cyber attack has one point that is missed by many. The "first responders" to the attack site did not come from a traditional CERT. They came from Cyber gym,* an organization that specializes in the security of ICSs. It was Cyber gym that determined what had happened and had the skills to contain and manage the incident. A CERT staffed with Windows/Linux/Intel/CISCO certificates hanging on their office walls did not have the skills to deal with an attack on an ICS in the Carmel case. This is an important point that needs to be considered, for it is a mistake to think that it is enough just to have a cyber security program with CERTs to deal with an attack on a website or malware on an information system. Sadly, this seems to be the mindset and set of assumptions behind the concept of CERTs. To deal with the full range of cyber threats to IT and ICSs, the appropriate range of skill sets is also required. Policies developed at the governmental level in focusing on the threat to IT systems are not enough to deal with the cyber threats of today. All parts of the cyber defense structure need to be accounted for. As with building a house, it needs not only a strong roof and walls, but also a good foundation.

The NATO summit in Wales conducted in September 2014 did include more attention on cyber defense. However, ambiguities continued to remain in terms of the alliance's understanding of what needs to be protected, from what cyber threats, and how to address them. In reading the published summit declaration in terms of cyber defense, NATO's chief focus is on protecting its own networks while the responsibility for protecting national systems are left to the nations themselves (Wales Summit Declaration, 2014). However, NATO does seem to recognize the possible threats from cyber attack on the CIs of its members, and will include consideration regarding an alliance response on a "case by case basis." What were not addressed were cyber attacks on the CIs of member states by other allies. Cyber attacks or intrusions performed by allies directed at the telecommunication sectors of Belgium and Germany (Müller-Maguhn, 2014; Gallagher, 2014) have been reported in the press. If it was proven that a cyber attack was successfully executed against the CIs of a fellow ally, would Article 5 of the NATO treaty be invoked?

Some hopeful signs have appeared during the time of this writing (November 2014) from NATO that it may be "getting it" in terms of cyber securing CIs. In October 2014, the newly accredited NATO Energy Security Centre of Excellence in Vilnius held the first tabletop exercise of its kind that included cyber attacks on the energy sector in one of the exercise scenarios (NATO, 2014). The participants of nine countries, including partners from Qatar and the United Arab Emirates, all concluded that the exercise was very useful. The lessons learned will be used to organize a full-scale exercise in the future. On the other hand, NATO may still be in a situation of the left hand not knowing what the

* http://cyber gym.co.il/, "Cyber Gym™ is the global leader in cyber defense solutions for critical and sensitive production, governmental, infrastructure and utility organizations including Finance, IT, TELECOM, ICS and SCADA environments."

right hand is doing. Soon after the above-mentioned NATO tabletop exercise, I attended the Innovative Energy Solutions for Military Applications (IESMA) 2014 conference sponsored by the NATO Energy Security Centre of Excellence and the government of Georgia, and supported by the NATO Science for Peace and Security Programme. It was an excellent conference on the latest and greatest innovative applications of technology applied to energy efficiency. Unfortunately, the word "security" was missing not only in the vendor's exhibit hall and product brochures, but also in the words used in the panels and discussions (IESMA, 2014). In a short intervention during the question and answer period, I tried to point out the dangers of innovation based on "insecurity by design." The excellent new products and savings of energy from technical innovation in the energy sector are made possible by advances in information and telecommunication technology. These technologies have a vulnerable side which is exploitable by malicious actors. Cyber security must be considered right from the beginning of the design phase before providing this new equipment to soldiers, sailors, and airmen going into harm's way. The high officials from NATO, including the department responsible for emerging threats, stated that technology was a separate issue from security, and that this was the reason why security was not being stressed at this conference; simply put, security was too big an issue to cover. It made no difference to this kind of thinking, even after pointing out that the energy sector has experienced multiple and serious intentional and unintentional cyber incidents. To this audience, the wake-up calls of Stuxnet, the attack on Saudi Aramco, the Idaho National Laboratories 2007 "Aurora" experiment, Black Energy, and Sandworm never existed.

This also brought home that there was a divide in the fundamental comprehension between IT and ICSs. An official NATO response to one comment indicated that one of the exhibitor's products (a deployable energy management system) sent its unencrypted data over the Ethernet, Bluetooth, wireless, and to smart phones; a NATO official said that encryption (if really needed) was "no problem." Another said that the equipment being discussed belonged to much smaller systems (found in mobile bases) and that the vulnerabilities being pointed out are not that easily exploitable. Later, during the coffee break, I approached the first official and tried to explain that encryption is not to be taken lightly when designing a control system that is to be run in real time. Encryption could cause unexpected problems.

OSCE (2015) makes an attempt at confidence-building measures for states to follow in cyber space

In May 2011, during the Lithuanian chairmanship of the Organization for Security and Co-Operation in Europe (OSCE), I was invited as "Lithuania's national cyber security expert" by the Lithuanian Ministry of Foreign Affairs to an OSCE conference on cyber security. During this conference, the OSCE decided to apply its expertise in arms control to cyber space. It subsequently created an informal working group (OSCE, 2012) to develop proposals for confidence and security building measures (CSBMs) for states to follow in cyber space. This was an exciting moment for me, as I actually participated in some of these early discussions, which took place from the summer of 2011 until the fall of 2012. While many proposals were discussed, nothing that would in any way put limits or restraints on malicious state activities in cyber space could be discussed. I know this, because I was one of those who made such a proposal (Digital Dao, 2012). Sitting in the meetings, it was noticed that while many able representatives from member nations

were in attendance, they were mostly career diplomats whose experience in working with IT and (cyber)communication issues varied greatly or was based on previous work in nuclear or conventional arms control issues. Some nations only sent their local OSCE mission representatives, who mostly sat quietly, while others sent higher-level diplomats, who rigidly maintained an approved policy position rather than engage in an open discussion of the issues involved. It became clear that there was a significant lack of general, shared knowledge about the technical aspects of cyber security and its application to a foreign policy issue. Something very important was missing in these discussions on CSBM proposals. No one wanted to mention or discuss what Stuxnet represented, nor its implications. Here was an example of one nation's malicious cyber activities being directed at the CIs of another nation. This destabilizing activity was even being practiced by some nations represented in the workgroup. In fact, raising the issue of restraint by states while eliciting some nodding of heads by some representatives immediately raised concerns for cyber superpowers, who were publicly declaring in other communications and fora that cyber space was considered an "operational domain." The hostile reaction to any discussion on restraint and transparency seemed almost childlike, as if some valued toy was going to be taken away by a parent. It represented another failed opportunity in another international forum to deal with an obvious topic that no one wanted to discuss.

The OSCE, however, did come out with a curious document, called *Guide on Non-Nuclear Critical Energy Infrastructure Protection from Terrorist Attacks Focusing on Threats Emanating from Cyberspace* (OSCE, 2013a). It was curious for two reasons. One was the distinction being made between cyber threats directed at nuclear and nonnuclear power plants. As one colleague from the ICS world remarked to me, both nuclear and nonnuclear power plants use the same control systems and are equally vulnerable to cyber incidents and attacks. It really begs the question: Shouldn't we also be very concerned about what would happen to a nuclear plant if its control systems are hit by a cyber attack? The full plant failures resulting from loss of power to run control systems leading to reactor meltdowns (such as occurred at Fukushima, Japan) can be caused by a cyber event and not just by earthquakes and tsunamis. Another curious part of the guide was the use of the cyber terrorist model. Throughout the time of the writing of this guide, and after it was published, there was little evidence of "cyber terrorism" executed by what many consider to be terrorists of the "Al-Qaeda" brand. On the contrary, cyber attacks on critical energy sectors pointed to nation-state, not terrorist, involvement. I did manage to become a member (and stay to the end) of the task force and contribute to preparing the guide, but was unsuccessful in changing the title of the guide and including the activities of states as one of the sources of cyber threats. I was successful in reducing the IT bias contained in the early drafts of the guide by successfully proposing that, in addition to ISO 27000 standards, that more relevant standards for ICSs and the energy sector be included, such as IEC 62351 and IEC 62443. Language addressing the peculiarities in cyber security practices found in both the IT and ICS realms also found a place in the text. For example, this statement about risk: "Risk needs to be understood with an appreciation for the peculiarities in security practices found in the ICT and Industrial Control System (ICS) realms" (OSCE, 2013b).

A bias toward protecting IT and information systems from cyber criminals, hacktivists, and "cyber terrorists" seemed to be the only point policy makers had in common. This bias, or lack of awareness, was contributing to the lack of ICS language in legal instruments on cyber security. This was not the first time I noticed this missing element at a meeting on cyber security.

Closer to home: Experience in dealing with cyber security questions in Lithuania

In Lithuania, analogous experiences applied in dealing with national cyber security issues. Meetings were held in our government to discuss preparation for cyber exercises. Scenarios were proposed and one of them was to include a cyber attack on Lithuania's electric grid. One representative from a participating government ministry said that such a scenario was totally impossible, as our grid was "not connected to the Internet"! To someone aware of Stuxnet, this is a disturbing statement coming from a government official participating in the development of our national cyber security policy. In another cyber exercise scenario discussion, it was proposed that a Stuxnet-type cyber attack would result in lost electric power to half of the capital city of Vilnius. The exercise leaders agreed with this, but only on the condition that the location of the Ministry of National Defense would not be in the part of Vilnius experiencing the blackout. If this happened, the communications center used for the exercise would not be able to be used in the exercise. To this, I could only reply that one can only hope that if real conflict took place, our enemies, in executing a cyber attack on our electric grid, would also be sure to leave the ministry's power supply untouched. I ran into similar differing levels of understanding when working on various national task forces dealing with cyber security issues. The lack of a general base of knowledge about cyber security made it very difficult to answer fundamental cyber security policy questions such as:

- What needs to be protected
- From what cyber threats
- How to protect them

There was also reluctance, when discussing the cyber security of CIs, to invite representatives from the electric, gas, and other utilities. One can imagine how difficult it was to develop realistic scenarios that included a cyber attack on the electric grid or pipelines without the participation of national or regional operators.

In considering the responses of governments and international organizations to the increasingly sophisticated and dynamic threats emanating from cyber space, it is difficult to understand the presence of these "blinders" when seeking to determine what needs to be protected and from what threats. In spite of growing evidence of the attacks on increasingly vulnerable CIs, the emphasis continued to be focused on protecting government and business information systems from cyber criminals, hacktivists, and cyber spies. This is quite ironic, for if you ask anyone working in government or in an international security organization about the importance of CIs, the reply would be that it is very important. One would think that the main purpose behind the work of government officials is to ensure that the people they serve are protected from harm from malicious cyber incidents and not the other way around.

Why is the cyber security of ICSs/SCADA not being included in the discussions on securing cyber space and CIs? Reading the documents produced by these organizations, their understanding of cyber security appears to be lacking. Cyber threats tend to be characterized as external, in terms of outside attacks by criminals and targeted espionage attacks by states or state-sponsored actors on IT systems. CIs are mentioned but the fundamental understanding is basically IT based. The cyber vulnerabilities and exploits in the energy sector seem to be unconsciously lumped together with vulnerabilities, exploits, and attack vectors associated with traditional IT attacks (DDOS, spear phishing, social engineering, etc.). Targeted attacks on control systems of the Stuxnet variety (not to mention

the unintentional incidents that take place) do not seem to be factored in. It is a mistake to assume that those writing the documents and making statements on cyber security and cyber defense in these organizations have ICSs in mind. Terms like SCADA, representatives of manufacturers of these systems, engineers, and the awareness of designers of SCADA's different approach to cyber security and specific cyber threats rarely appear. The assumptions do not fully apply to ICSs and, therefore, the documents dealing with the problem and the strategies to address them only cover part of the issue. This is similar to the realization I faced when writing the above-mentioned article on the cyber security dimension of critical energy infrastructure. This erroneous assumption coming from an IT bias is quite common when cyber security based on IT is so dominant.

I will use one anecdote to illustrate what I mean by IT not seeing ICSs. In Lithuania there is a very well-known painter, Aloyzas Stasiulevičius. He has had a long and successful career as a painter. His unique place in Lithuanian painting comes in part from his main theme that he uses over and over again—the city of Vilnius. He paints the same scenes in different ways and in different colors, but the theme is almost always Vilnius. The story goes that one weekend, Lithuanian painters gathered in a national park by the beautiful Lake Aisetas. Great paintings were accomplished, with depictions of lakes, forests, and wildlife scenes. When they came over to look at Stasiulevičius, who was painting beside a lake, they all remarked at his work—"Look, it is Vilnius!" IT cyber security specialists seem to be stuck with the same vision when they approach ICS cyber security. They see just the IT part and do not notice that ICSs are different. This mindset tends to dominate so much that, when policies are created for ICSs, there is so much that is missing. A good example of this is the integrated management system policy of the Slovak Republic's electricity transmission system operator Slovenskaelektrizacnaprenosovasustava a.s. (Integrated Management System Policy, 2014). Among the standards listed, only ISO 27000 is listed for information security management. There is no mention of any standards having to do with operating ICSs; for example, no mention of IEC 60870 and ISA 99/IEC 62443 (ISA99 Committee, 2015). If the writers of these documents and designers of critical systems are not aware of these relevant ICS standards, then they are just left out by default. The result is that much is missing from these documents that could be used to prevent and limit the possibility of a bad event occurring in CIs. The bottom line is that these efforts do not result in ensuring that everything that is truly critical is protected.

What is considered to be critical infrastructure seems so obvious, but …

The dependence of our economies and well-being of our societies on a safe, reliable, and increasingly hi-tech-based infrastructure consisting of energy, finance, telecommunications, transport, and other utility sectors has been recognized by governments for a long time. The availability of the services provided by these sectors, if disrupted or discontinued for longer than a few hours, would have damaging effects on the economy and society. That is why their availability at all times is considered critical. The United States, among other countries, has a good understanding of what a CI is and what it means to its national security. This goes as far back as November 9, 1965, when President Lyndon Johnson, in a letter, cited a report on the blackout of 1965 which took place on that date. He wrote in his order for the preparation of the report, "Today's failure is a dramatic reminder of the importance of the uninterrupted flow of power to the health, safety, and well-being of our citizens and the defense of our country" (Federal Power Commission, 1965). This early recognition in 1965 by the U.S. government of the importance of CIs is

further reinforced today in the existence of the DHS ICS-CERT, which is probably the one of the few nationally-backed CERTs of its kind dedicated to the cyber security of industrial control systems which form the backbone of today's CIs.

While the 1965 blackout was not the end result of a cyber incident per se, it was caused by an unintentional "programming error"* in one of the electromechanical links (a relay) belonging to the internationally operated electric grid providing electricity to the northeastern United States and parts of Canada. The important message in terms of this article is that after the 1965 blackout, the U.S. and Canadian governments, as well as industry, after careful analysis of what was wrong with the system, implemented remedies at national and industry levels to ensure that such a failure would not easily reoccur. This included creation of the Northeast Power Coordination Council,† which would later become affiliated to what is known today as the North American Electrical Reliability Corporation (NERC), and the passage of the Electric Reliability Act in 1967. This resulted in more rational management and in improved reliability in the power industry, as well as setting the stage for developing the more complex and interconnected power systems of today. It also resulted in making management systems more complex and vulnerable to new threats from cyber space.

As long as the scope of affected systems is limited to one nation or is inside an isolated system, this model, in terms of local measures taken by industry, government, and a cooperative neighbor to prevent another blackout or failure in CIs, falls short in the changed cyber security environment of today. Today's threats in cyber space have a global or borderless character. The interdependence of CIs crosses borders. A cyber attack or incident leading to a failure of CIs may have its origin in another country or as a result of a political conflict among nations.

Much has been covered so far about the response of governments and the international community to the growing cross-border cyber based threats. What has been the response of industry, especially from the various CI sectors?

Response of industry

In talking about the public–private partnership between governments and CI sectors, the lessons of Stuxnet and the Saudi Aramco incident have not been learned. My experience as a guest speaker for a conference where energy sector representatives from industry and government participated serves as one illustrative example. In May 2014, I was invited to speak at an energy sector conference in Vilnius, sponsored by the Estonian Chamber of Commerce. It was opened by the Prime Ministers of Lithuania, Latvia, and Estonia. Participating were representatives from governments and private energy companies in Lithuania, Latvia, Estonia, Sweden, Poland, and energy-related NGOs.

What surprised me most was that, until my session, there was no mention of the word "security" in any of the presentations. Terms like "critical infrastructure bottlenecks" were used to describe the lack of transmission capacity on the grids or pipelines between countries. Polish industrial boiler manufacturer Rafako (2015) gave a vendor brochure-type presentation, filled with pictures of their products, descriptions of their experience in providing turn-key systems, and of sites where their products were installed. One picture showed a "condenser," a two- or three-story high cylinder-shaped object sold to nuclear power plants. Generally, security as a problem was only mentioned in the context of "supply"; for example, in terms of what could happen if Russia stopped fulfilling its gas supply contracts. Other

* The relay was mistakenly set to trip at a much lower power level than could be safely transmitted by the capacity of the power lines.
† https://www.npcc.org/default.aspx.

presenters covered the financial aspects (market forecasts) and the online bidding and selling of energy. China was cited as an "island of economic stability" in the East.*

My presentation (allowed for just 15 min) came after lunch, when the prime ministers and some of the morning presenters had already left, and I asked the audience to recall the huge condensers that were shown earlier. I pointed out that they were controlled by things called program logic controllers (PLCs)† and belonged to complex systems of monitoring and control called SCADA. These "systems of systems" have now, for various reasons, become vulnerable to unintentional and intentional cyber incidents that have caused major damage and loss of life. Nobody said they had heard of "Aurora" and were equally clueless about both the Google incident and the Idaho National Lab experiment.

In terms of "bottlenecks" and "problems with infrastructure," I asked them to imagine that one day, while looking at their online market transaction screens, they suddenly found that the screen had "frozen" or that a network/server error message ("try again later in about an hour") appeared while they were trying to make a bid. This could happen because of an attack on a website that hosts the online transaction system (I believe servers for the Nord-Balt Energy Pool Spot Market are located in Oslo, Norway). What would they do when something went wrong with the operation of the pipeline or electric distribution system, resulting in an interruption in the supply of gas or electricity? These incidents may be caused by unintentional or intentional failures in the control systems that are used to run devices that form these CIs. I also told them that classes are now being offered at Black Hat with titles like "Hacking SCADA" (Parker, 2014).

Cyber security in the sectors belonging to a given CI seem to be understood in different ways. A very interesting study on the cyber fragilities of traffic light control systems came out, which contains a surprising finding illustrating how some manufacturers in the industry look at cyber security: "A clear example can be seen in the response of the traffic controller vendor to our vulnerability disclosure. It stated that the company *has followed the accepted industry standard and it is that standard which does not include security*" (Ghena and Beyer, 2014). Similar perception problems exist in other critical sectors. In looking at air traffic control (ATC) systems, the willingness to implement solutions for improved resiliency of ATC systems to a cyber attack hinges on cost, and in some cases, a certain state of denial. As Camilleri writes in his study of the cyber preparedness of the aviation industry, "Most are already familiar with many of the issues of unencrypted radio communications. As most aerospace and defense contractors also originally developed the same civilian equipment for military aviation systems, they are also aware of the solution to these problems—simply to encrypt all communications traffic in air between aircraft, as well as on ground. But the FAA and airline industry argue otherwise" (Camilleri, 2014).

At the NATO IESMA 2014 conference I described earlier above, I made a point of visiting the vendors' and manufacturers' (including Honeywell, BAE Systems, Bredenoord) exhibition hall and asking about the connectivity and security of their equipment. I was provided with a lot of information about the way the equipment can be accessed over Ethernet, Bluetooth, wireless, smartphones, and even remotely over the Internet. The manufacturer's representative at the booth did not include any presentation on how secure their products were from threats emanating from cyber space. A heating, ventilation, and air conditioning (HVAC)‡ vendor expressed surprise over hearing that the Target

* Which provoked a question from me: "What about Japan?"

† A digital computer used for automation of typically industrial electromechanical processes, such as control of machinery on factory assembly lines (https://en.wikipedia.org/wiki/Programmable_logic_controller).

‡ https://energy.ces.ncsu.edu/heating-ventilation-and-air-conditioning-system-hvac-defined/.

Corporation's financial system was cyber attacked through the company building's HVAC system (Target Hackers Broke in Via HVAC Company, 2014). Most vendors were very interested in learning more about what Stuxnet was. I spelled out the word to several vendor representatives who said they would investigate this and respond back to me. I suggested that the first question they asked their engineers back home was whether they had heard of Stuxnet and if they had read any of the analysis published by Ralph Langner. I told them that if their engineers answered negatively to those questions, then some attention and enlightenment would be in order for the good of their esteemed company's products.

In dealing with the malicious cyber activities of states, neither the international community nor industry appears to have a coherent understanding of the serious implications of the new threats emanating from cyber space being directed toward a CI. In international fora described above, there is either a reluctance to talk about any limitations on these activities, or they are not even recognized as an issue (some other organization's problem to deal with). The examples of the energy and airline industry used above may not be representative of the views and understanding of all the sectors of CIs as a whole, but without government leadership in fostering and in coordinating an effective international response, the threat to CIs from debilitating cyber attacks continues to exist.

Since Stuxnet (and other similarly produced malware that has followed) was the work of a nation-state directed at the CIs of another nation-state, the level of response required to address this kind of attack is beyond the local capabilities of a national government, regulatory body, or industry. Efforts to address this new form of cyber attack have to come from efforts of national structures in coordination with other members of the international community. As a former White House policy director, Jason Healy, remarked in his book, *A Fierce Domain: Conflict in Cyberspace 1986–2012*, "as Smart Grid and other technologies interlink the Internet with real infrastructure—made of concrete and steel, not silicon—the consequences of attacks will be far worse, especially from more covert nation-state conflicts … a further trend, which suggests that there will be more covert disruptive conflicts between governments, as each nation realizes its own advantages in disrupting adversaries on-line" (Healey, 2013, pp. 85–86).

Recommendations

Short/medium-term and long-term recommendations

In terms of making recommendations, a parallel two-track approach will be used: one for the immediate and short term and the other for the long term. I would like to borrow a term from the practice of medicine which has been useful in helping us to understand cyber space in the past (e.g., the use of the term "virus" to describe the actions of malicious software). My wife is a rheumatologist and she treats many patients using a two-track method by first prescribing symptom-relieving medication to reduce the immediate pain and discomfort felt by the patient. This addresses the immediate concern of the patient to feel better right away. However, the path to a cure also requires attention to the disease process, which will continue regardless of the effects of the short-term pain medication. For this reason, she also prescribes another drug (a disease-modifying antirheumatic drug [DMARD]) that treats the actual disease over a longer period of time. I will propose some solutions to the international issues of protecting SCADA systems in the same way, by proposing short-term solutions (providing immediate relief of "symptoms"), which can address some of the immediate concerns over dealing with current threats emanating from cyber space, and long-term solutions (cyber security "DMARDs") that can address

core issues to ensure the reliability and safety of these vital systems, which form the technical platforms we depend on for our modern economies and way of life.

Short/medium term

Proposals for addressing the misbehavior of states in cyber space

1. Commitment to restrain from malicious cyber activities directed against critical civilian infrastructure (financial, energy/utility, transportation, and telecommunications).
   Rationale: The desire to protect national economies and civilians from financial loss or physical harm should be common to all nations. Certain state activities in cyber space can lead to misperceptions and instability in relations among states. For example, the placement of "logic bombs" or "back doors" in a nation's CI infrastructure can be mistaken for "preparation of the battlefield" activity and could lead to rapid escalation of tensions. Cyber activities directed against the CIs of another nation-state can also have significant cross-border and even regional effects, due to the integration of financial systems, power grids, pipelines, and other modern CIs.

   Something similar has already been mentioned in other proposals made by representatives of both Eastern and Western countries. One comes from the nation closely associated with Stuxnet. Richard Clarke, former adviser on national security for several U.S. presidents, has applied his extensive experience in nuclear arms control issues to the realm of cyber space in his recent book, *Cyber War*. Read his proposal for a cyber war limitation treaty (Clarke, 2010). Language prohibiting the use of cyber weapons against CIs is also included in the Shanghai Cooperation Group proposals for an international code of conduct sent to the UN in 2011 (Ministry of Foreign Affairs of the People's Republic of China, 2011).

   Restraint is not enough of a pledge; it also requires an acceptance of responsibility to meet one's obligations, which leads to Proposal 2.
2. Commitment to national cyber space liability: States agree to accept responsibility for malicious cyber activities taking place within their cyber space jurisdictions or transiting through them.
   Rationale: Nations need to agree on minimum obligations to secure their national cyber space. Emphasis should be placed on the state's obligations to react to incidents originating from or transiting through their cyber space jurisdictions. Nations should ensure, for example, that national internet service providers (ISPs) and law enforcement agencies take appropriate steps against individuals, groups, and/or information and communication technology equipment found to be participating in a cyber attack. This also implies that nation-states agree to develop a capacity for dealing with cyber security matters. This means establishing appropriate laws and structures (national CERTs, law enforcement organizations, etc.) needed to implement the commitment.

   This is also not a new idea. Scholars in the United States have been discussing the merits of nation-states accepting responsibility for what goes on in their cyber jurisdictions. Examples of this policy thinking include Chris C. Demchak and Peter Dombrowski's paper covering cyber borders and jurisdictions. They argue that cyber space is no longer a public commons or prairie where all can roam and do as they wish. There is so much development and interest at stake for a nation's security that the establishment and control of "cyber borders" is an important

step toward ensuring protection of their CIs from cyber based threats (Demchak and Dombrowski, 2011).

Related to responsibility and liability is the problem of attribution. The level of difficulty to carry out cyber attacks and the probability of getting caught must be raised higher. The establishment and control by a nation-state of its cyber borders will make it more difficult for cyber attacks to pass unnoticed. However, the unsuccessful effort up until now of placing the blame needs to be shifted from trying to identify who is actually attacking to identifying "what nation, if any, is responsible" (Healey, 2013, p. 265). It is the nation-state that should be held responsible for ensuring the control of its cyber borders and for making sure that malicious cyber activity originating or transiting through its cyber jurisdiction is monitored and controlled. The full burden of responsibility for reacting to and investigating an attack should not be placed on the victim but on those closest to and capable to react to the incident.

3. Monitoring of implementation of agreed commitments as listed: Nation-states agree to create a coalition of willing experts and institutions to monitor and advise on violations of these two agreements.

   Rationale: Some means must be available to monitor and inform participating nation-states of malicious cyber activities taking place or transiting through their cyber jurisdictions. An institution consisting of experts that can monitor and provide objective evaluation of violations to commitments should be established. This will provide for a capability to apply pressure on nations that are slow or reluctant to act on reported malicious activity taking place in their cyber jurisdictions.

Again, this is nothing that should be new to anyone working in international relations; this is not naive idealism. In questions where the need is recognized and where it really matters, nation-states have banded together and signed international agreements and conventions. This has been especially so with prohibiting the use of weapons of mass destruction. One possible model for dealing with the production and use of cyber weapons by nation-states is the International Convention on Chemical Weapons. Still perhaps remembering their use in World War I, and in recognition of the advances in technology that could facilitate the use of chemical weapons and amplify their potential for harm, a convention entered into force in 1997. Over 190 nations have signed it, representing most of the world's population. Associated with the agreement, the Organization for the Prohibition of Chemical Weapons (OPCW) was created to monitor and follow up implementation of the convention (OPCW, 1997). The convention on chemical weapons can serve as a useful model when considering implementation of the three above-mentioned proposals.

The Asia Pacific Computer Emergency Response Team coalition (APCERT) offers an example of regional cooperation. APCERT is made up of CERTs and ISPs from Japan, China, and South Korea. APCERT treats "the Internet and its health as a connected common shared infrastructure" (Ito, 2011). The coalition has been successful at addressing cyber incidents arising from political conflicts among its members.

One example of an ad hoc, yet effective global response to a perceived common threat in cyber space, is the work of the Conficker work group in 2008–2009. Governments appear to have failed to recognize the growing danger to the Internet from the creator of the Conficker worm, and the growing number of infected computers that could be commanded into action at any time. The fight to save the Internet from this new and potentially destructive worm was taken up by a group of volunteers that included private