Information Security Management Handbook

Sixth Edition

Edited by

Richard O'Hanley · James S. Tiller

Volume 7

CRC Press
Taylor & Francis Group
AN AUERBACH BOOK

Information Security Management Handbook

Sixth Edition

Volume 7

OTHER INFORMATION SECURITY BOOKS FROM AUERBACH

Asset Protection through Security Awareness
Tyler Justin Speed
ISBN 978-1-4398-0982-2

Automatic Defense Against Zero-day Polymorphic Worms in Communication Networks
Mohssen Mohammed and Al-Sakib Khan Pathan
ISBN 978-1-4665-5727-7

The Complete Book of Data Anonymization: From Planning to Implementation
Balaji Raghunathan
ISBN 978-1-4398-7730-2

The Complete Guide to Physical Security
Paul R. Baker and Daniel J. Benny
ISBN 978-1-4200-9963-8

Conflict and Cooperation in Cyberspace: The Challenge to National Security
Panayotis A. Yannakogeorgos and Adam B. Lowther (Editors)
ISBN 978-1-4665-9201-8

Cybersecurity: Public Sector Threats and Responses
Kim J. Andreasson
ISBN 978-1-4398-4663-6

The Definitive Guide to Complying with the HIPAA/HITECH Privacy and Security Rules
John J. Trinckes, Jr.
ISBN 978-1-4665-0767-8

Digital Forensics Explained
Greg Gogolin
ISBN 978-1-4398-7495-0

Digital Forensics for Handheld Devices
Eamon P. Doherty
ISBN 978-1-4398-9877-2

Effective Surveillance for Homeland Security: Balancing Technology and Social Issues
Francesco Flammini, Roberto Setola, and Giorgio Franceschetti (Editors)
ISBN 978-1-4398-8324-2

Electronically Stored Information: The Complete Guide to Management, Understanding, Acquisition, Storage, Search, and Retrieval
David R. Matthews
ISBN 978-1-4398-7726-5

Enterprise Architecture and Information Assurance: Developing a Secure Foundation
James A. Scholz
ISBN 978-1-4398-4159-4

Guide to the De-Identification of Personal Health Information
Khaled El Emam
ISBN 978-1-4665-7906-4

Information Security Governance Simplified: From the Boardroom to the Keyboard
Todd Fitzgerald
ISBN 978-1-4398-1163-4

Information Security Policy Development for Compliance: ISO/IEC 27001, NIST SP 800-53, HIPAA Standard, PCI DSS V2.0, and AUP V5.0
Barry L. Williams
ISBN 978-1-4665-8058-9

Information Technology Control and Audit, Fourth Edition
Sandra Senft, Frederick Gallegos, and Aleksandra Davis
ISBN 978-1-4398-9320-3

Iris Biometric Model for Secured Network Access
Franjieh El Khoury
ISBN 978-1-4665-0213-0

Managing the Insider Threat: No Dark Corners
Nick Catrantzos
ISBN 978-1-4398-7292-5

Network Attacks and Defenses: A Hands-on Approach
Zouheir Trabelsi, Kadhim Hayawi, Arwa Al Braiki, and Sujith Samuel Mathew
ISBN 978-1-4665-1794-3

Noiseless Steganography: The Key to Covert Communications
Abdelrahman Desoky
ISBN 978-1-4398-4621-6

PRAGMATIC Security Metrics: Applying Metametrics to Information Security
W. Krag Brotby and Gary Hinson
ISBN 978-1-4398-8152-1

Securing Cloud and Mobility: A Practitioner's Guide
Ian Lim, E. Coleen Coolidge, and Paul Hourani
ISBN 978-1-4398-5055-8

Security and Privacy in Smart Grids
Yang Xiao (Editor)
ISBN 978-1-4398-7783-8

Security for Wireless Sensor Networks using Identity-Based Cryptography
Harsh Kupwade Patil and Stephen A. Szygenda
ISBN 978-1-4398-6901-7

The 7 Qualities of Highly Secure Software
Mano Paul
ISBN 978-1-4398-1446-8

AUERBACH PUBLICATIONS
www.auerbach-publications.com • To Order Call: 1-800-272-7737 • E-mail: orders@crcpress.com

Information Security Management Handbook

Sixth Edition

Volume 7

Edited by

Richard O'Hanley · James S. Tiller

CRC Press
Taylor & Francis Group
Boca Raton London New York

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

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

Version Date: 20130723

International Standard Book Number-13: 978-1-4665-6752-8 (eBook - PDF)

Contents

DOMAIN 7: OPERATIONS SECURITY

Concepts

DOMAIN 9: LEGAL, REGULATIONS, COMPLIANCE, AND INVESTIGATIONS

Information Law

Investigations

Major Categories of Computer Crime

Compliance

Introduction

This is the first annual edition of the *Information Security Management Handbook* since 1994 without the guidance and the insight of Hal Tipton. Hal passed away in March 2012. He will be missed by a lot of people for a lot of reasons.

It seems that every year is an interesting one for information security, and 2012 was no different. It is interesting, too, how perceptive Kaspersky Labs, for example, was with its forecast. It also foreshadows the end of online trust and privacy. If you cannot trust digital certificates, what is left to trust?

Kaspersky Cyberthreat Forecasts

2012	*2013*
Cyber weapons	Government surveillance
Mass targeted attacks	Continued targeted attacks
Mobile threats	Mac OS X malware and mobile malware
Attacks on online banking	Cloud attacks
PPI attacks	PPI threats
Hacktivism	More hacktivism
	Problems with trust and digital authorities
	Ramsomware and extortion malware
	Espionage and other government cyberattacks

Cyberwarfare has jumped to the front pages of every newspaper, both print and virtual. Stuxnet spawned Flame, Duqu, and Gauss. While we were all focused on attacks and espionage by China, France, and Israel, Iran mounted a DDoS (Distributed Denial of Service) attack against US banks in retaliation for sanctions that appear to be working. At the same time, Iran's central bank was attacked. Added to the online attacks is the growing threat of supply chain security, and products shipped with back doors or embedded systems that let them phone home. Witness the difficulty Chinese telecom equipment suppliers like Huawei are having with gaining toeholds in the United States by purchasing the US suppliers.

While Russians and Eastern Europeans are not singled out for cyberwarfare, crime syndicates based there continue to threaten commerce and privacy.

Theft of passwords from LinkedIn and Dropbox, and what seems like daily reports of attacks on or by Facebook show the lure of social media to hackers, and the dangers to the rest of us. And while Facebook and others do not install rootkits like Sony did, its data collection efforts, combined with the apparent insecurity of the site emphasizes the growing dangers of Big Data and the Cloud.

We saw a huge increase in hacktivism as Anonymous and LulzSec launched various attacks on both government and private sites around the world.

It was only a matter of time until Mac OS X became a profitable target. Once critical mass was reached, hackers could not resist investing the time to own it.

As with Mac OS X, mobile devices are becoming even more alluring targets. We have seen the same types of attacks and malware used against PCs adapted to mobile, plus new threats like SMS (short message service) spoofing. Not surprisingly, Android, Google's open platform, has suffered the most. Plus, the growing number of apps for all platforms introduces a level of threat that is hard to estimate, but definitely growing.

M2M and the Internet of Things are creating more opportunities for hackers. From NFC (near-field communication) payments to utility sensors sending unencrypted data, this is a potentially lucrative area for fraud and identity theft. Sensor networks are now in the DIY (do-it-yourself) arena, which creates yet a new class of threats.

BYOD (Bring Your Own Device), IT consumerization, whatever you call it, is making life so much more fun for black hats. It has given new meaning to "insider threats." With portable digital devices being introduced into the enterprise, both with and without permission, we are seeing a manifold increase in threats. Clearly, policies alone are not sufficient to deal with this, and it is unclear how draconian management wants to be with forcing compliance. The products exist, but does the will to use them?

Looking at 2013, the promise of more surveillance, both from governments and online data collectors, means less privacy, even for the most careful users. Short of totally disconnecting from the grid, if such a thing is possible now, it is apparent we do not and would not have privacy.

This edition of the *Information Security Management Handbook* addresses many of these trends and threats, plus new areas such as security SDLC (software development life cycle), as well as forensics, cloud security, and security management. Chris Hare takes an in-depth look at hacktivism, identifying the motivations and the players, and providing advice on how to protect against it. Becky Herold analyzes the security and privacy challenges of social media. Sandy Bacik looks at the security implication of BYOD, and the challenges of managing user expectations. The Smart Grid offers its own security and privacy challenges as Terry Komperda explains. Noureddine Boudriga explains attacks in mobile environments.

There is new guidance on PCI and HIPAA/HITECH compliance. In addition to forensics and e-discovery, a chapter looks at cell phone protocols and operating systems from the perspective of a forensic investigator.

I have heard it said, "You can't fix stupid." So many of these attacks are successful because of clueless or irresponsible users. In what I hope is not a vain effort, Ken Shaurette and Tom Schleppenbach look at human firewall testing, social engineering, and security awareness. We also look at security and resilience in the software development life cycle, managing the security testing process, and SOA (service-oriented architecture) security.

Here is a shout out to my friend Jim Tiller, head of Security Consulting, Americas for HP Enterprise Security Services, for his help in preparing this edition. Jim's done a lot for the Handbook over the years, and I am hoping he will continue.

All-in-all, this is a good volume of the *Information Security Management Handbook*. We are working on the next edition now. If you would like to contribute, please contact me at 917-351-7146 or rich.ohanley@taylorandfrancis.com.

Richard O'Hanley

Contributors

Sandy Bacik
Lord Corporation
Cary, North Carolina

Adrian Ball
TurningPoint Global Solutions
Rockville, Maryland

Noureddine Boudriga
Réseau National Universitaire
Tunis, Tunisia

Jeff Crume
IBM
Research Triangle Park, North Carolina

Eamon P. Doherty
Fairleigh Dickinson University
Teaneck, New Jersey

Todd Fitzgerald
ManpowerGroup
Milwaukee, Wisconsin

Tim Godlove
Department of Veterans Affairs
Washington, DC

Greg Gogolin
Ferris State University
Grand Rapids, Michigan

Chris Hare
Verizon
Dallas, Texas

Rebecca Herold
Rebecca Herold & Associates, LLC
Des Moines, Iowa

Debra S. Herrmann
Jacobs Engineering
Washington, DC

David G. Hill
Mesabi Group LLC
Westwood, Massachusetts

Seth J. Kinnett
Chicago, Illinois

Terry Komperda
Illinois Institute of Technology
Chicago, Illinois

Anthony Meholic
The Bancorp Bank
Wilmington, Delaware

Mark S. Merkow
PayPal
San Jose, California

Robert K. Pittman, Jr.
County of Los Angeles
Los Angeles, California

Lakshmikanth Raghavan
PayPal
San Jose, California

Balaji Raghunathan
Infosys Limited
Bangalore, India

Ashutosh Saxena
Infosys Limited
Hyderabad, India

Tom Schleppenbach
Inacom Information Systems, Inc.
Madison, Wisconsin

Greg Schulz
StorageIO
Stillwater, Minnesota

Ken M. Shaurette
FIPCO
Madison, Wisconsin

Tyler Justin Speed
Electronics International
Eugene, Oregon

Jeff Stapleton
Bank of America
Dallas, Texas

John J. Trinckes, Jr.
PathForwardIT
Cincinnati, Ohio

Ravi Sankar Veerubhotla
Infosys Limited
Hyderabad, India

Prashanth Venkatesh
Infosys Limited
Bangalore, India

Walter B. Williams
Lattice Engines
Boston, Massachusetts

Beau Woods
Stratigos Security
Atlanta, Georgia

DOMAIN 2

TELECOMMUNICATIONS AND NETWORK SECURITY

Communications and Network Security

Chapter 1

Securing the Grid

Terry Komperda

Contents

Introduction

Before we can dive into how utility networks will evolve and how those future networks will be exposed to issues that will affect their security and continued functioning, we need to look at some history on the current networks, why they need to change after functioning properly for so long, and the benefits to be realized related to their technological advancement.

The Power (Electrical) Grid

The power grids of the twentieth century were designed to be a one-way broadcast of power from a few central generators to a large number of electrical users. At the time of the design, the main goal was to keep the lights on without any regard for energy efficiency, environmental considerations, or consumer choices. Typically, it has been a geographically organized number of integrated utilities with control based on a fixed hierarchical infrastructure. The following section is a simple illustration of a power grid that shows the main functions that a power grid performs (Figure 1.1).

Core Functions of a Power Grid

The following factors are the core functions of a power grid:

a. *Power generation*—Power is generated at a power station and can emanate from coal and nuclear plants, dams, windmills, and so on.

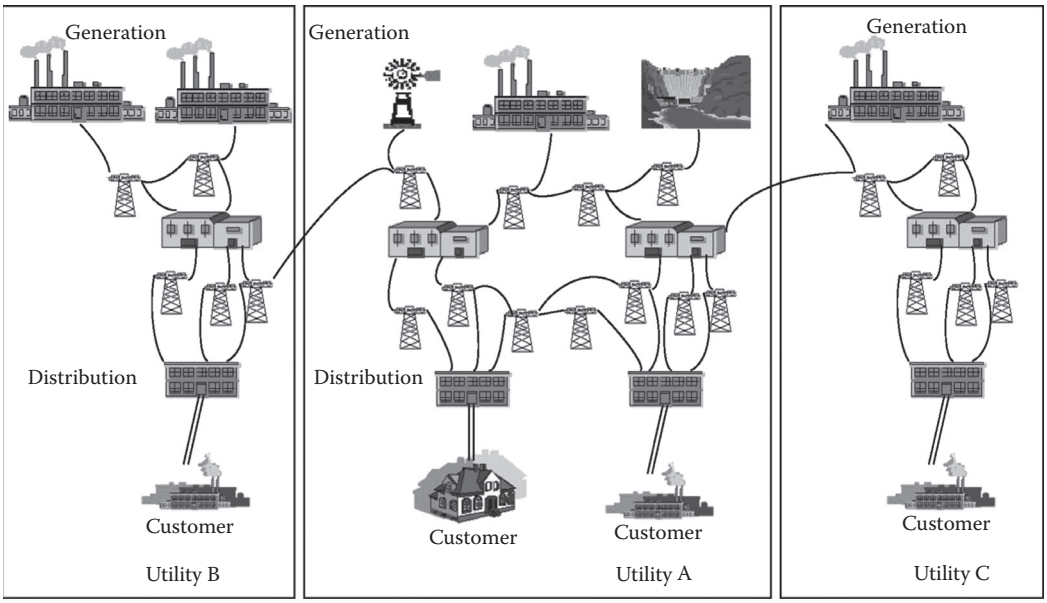


Figure 1.1 Power grid. (From Bakken, D. et al. 2003. Grid stat. Washington State University, School of Electrical Engineering and Computer Science, November 2003. Slide #4.)

b. *Transmission*—Electricity is transferred from power stations to power distribution systems at a substation. The substation is a point of monitoring and control in the grid, and high-voltage electricity is handled here.
c. *Distribution*—Medium-voltage electricity resides here, and this is where power is delivered to the end customers.

Historically, larger power companies have been granted a monopoly status and typically control all the three functions for a geographic area.

Power Grid Components

These are the major components in a power grid:

a. *Generator*—Its major function is to generate power.
b. *Substation*—This is a point of control and monitoring in the grid, and can service many generators, boost voltage, and serve as a distribution point to the customers.
c. *Control area*—This is a set or a group of substations in a geographic area covering a county to a few states. A control area performs all the three core functions and corresponds to one or a few utility companies.
d. *Grid*—A set of control areas that are synchronously controlled.

Power Distribution Topologies

Power is typically distributed in one of the three ways in a power grid:

a. *Radial grid topology*—Electricity is distributed from a substation in a pattern resembling a tree with branches and leaves. The branches and leaves receive power from a single source.
b. *Mesh grid topology*—Power is provided from other sources (other branches and leaves), and this allows a mesh grid to be more reliable than a radial grid.
c. *Looped topology*—This is a combination of mesh and radial topologies and is used primarily in Europe. This topology resists disruption in the grid, no matter where the problem occurs.

Communication Networks, Control, and Communications Protocol in the Grid

a. *Communication networks*—Frame relay, asynchronous transfer mode (ATM), public switched telephone network (PSTN), and the Internet are all used for communications in the current grid.
b. *Control*—Supervisory control and data acquisition (SCADA) is a serial system implementation used to remotely control and monitor the transmission and the distribution of power in electrical grids.
c. *Communications protocol*—The most popular utility automation protocol used in North America is distributed network protocol (DNP). It is applied through distribution and transmission networks and provides connections from master stations to substations, between devices in substations, and out-to-pole top devices.

Problems in Current Power Grids

Although current grids have worked well for many years and have had upgrades such as automatic meter reading (AMR) to remotely read meters, the communications network is hardwired, dedicated, and slow, and this has led to networks that are dangerously antiquated. In fact, the dog food industry spends more on research and development than the electrical sector does, and aging technologies have led to more blackouts, vulnerabilities, and colossal inefficiencies (Kingsbury, 2010). Additionally, the following factors are some of the other issues that are presenting themselves with:

- Distributed control systems (DCS) and SCADA that are now connected to the Internet, and when they were originally designed, the controls were not designed with public access in mind. These systems typically lack rudimentary security, and technical information and security flaws for penetrating into these systems are widely discussed in public forums and are therefore well known to attackers. In fact, many years ago, a 12-year-old broke into the computer system that runs Arizona's Roosevelt Dam. He had full control of the SCADA system that controlled the dam's floodgates (Bakken, 2003).
- The Energy Department came up with multiple scenarios for attacking the grid through SCADA systems and all of them worked.
- Continued automation is being added to substations to reduce human errors and mistakes, but the computer-controlled systems and software increase the potential for security vulnerabilities.

A recent special case confirms that current power grids are not immune to attacks and vulnerabilities.

Stuxnet

Stuxnet is the first known malware attack to target power plants. It is a worm that was introduced via a universal serial bus (USB) device in an Iranian nuclear plant. It infected a SCADA system that was considered as buffered from the attack as most of these systems are not connected to the Internet in nuclear power plants. It installs a rootkit on the control system and injects a malicious code into programmable logic controllers, reprograms them, and hides the changes. It was digitally signed with two stolen authentication certificates from two certificate authorities and this helped it to remain undetected for quite some time. Once inside, it uses default passwords (Siemens, the manufacturer of most SCADA systems, recommends against changing default passwords) to command the software and exploits four different Windows zero-day vulnerabilities to infect all sorts of computers. Siemens reported that it has discovered an additional 14 clients (power plants) that have been infected, a number of which are in Germany (Evron, 2010). This attack pretty much quietened those that argued for maintaining the current grid with proprietary protocols and systems because it was thought that the current systems were more secure (especially if they were not widely connected to the Internet).

The Case for a Smart Grid

The following factors are some of the reasons that it makes sense to evolve current electrical grids into smart grids:

- The rates can be variable based on true usage. The consumers would only pay for the power used and if more power is used (especially during peak periods), utilities could charge a premium. The customers will have to change their behavior, but they will be rewarded for saving energy.
- The consumers who have a power surplus (that they would not be using) could push the power back into the grid and sell it back to the utilities for other customers who could use it.
- The grid system could be more stable by automatically avoiding or mitigating power outages and power-quality issues and by repairing itself (self-healing) during a service disruption. This will lead to fewer brownouts and blackouts.
- Waste reduction: Cutting tiny inefficiencies can have dramatic effects on the overall grid and can lead to maintaining affordability for all.
- The grid will accommodate both renewable and traditional energy resources leading to better power quality and improved reliability.
- The grid will be able to account for new, larger potential loads in the network such as that from increasingly popular electrical vehicles.
- The reductions in the carbon footprint will help with the overall energy conservation and will promote better environmental responsibility.

The Smart Grid

Implementing a smart grid transforms the power grid from a one-way, closed, proprietary system to a modern, two-way, standards-based, intelligent system that allows operators to monitor and interact with numerous components in real time. It allows operators to detect issues and manage grid operations for faster problem resolutions and lower operating costs. A smart grid will not replace the legacy systems, but will have to incorporate them and evolve to a smarter grid over many years (and at a significant cost). Internet Protocol (IP)-based systems will tie SCADA and DCS into the evolving grid for efficient management and communication across the main stations and remote locations. IP-based systems will pose a security challenge (just like they do when deployed elsewhere) but trying to secure legacy devices (that used isolation as a security technique) will make the security job even more challenging.

Smart Grid Technologies, Systems, and Components

The following factors are some of the technologies, systems, and components that will be used in smart grids:

1. *Integrated communications*—Today, a good amount of data are still collected via the modem instead of direct communication. The implementation of direct communications can improve substation automation, distribution automation, demand response, and SCADA management. This will allow for real-time control as well as information and data exchange for optimizing system reliability, utilization of assets, and security.
2. *Improved interfaces and decision support*—The collection of extremely complex data will become difficult for humans to comprehend in a timely manner. The human machine interface (HMI) must simplify the data to enable operators and managers to make decisions quickly.

3. *Distributed grid management (DGM)*—This aims at maximizing the performance of feeders, transformers, and other components of network-distribution systems and integrates with transmission systems and customer operations. The benefits derived are better reliability, reductions in peak loads, and improvements in the capability to manage distributed renewable energy sources.
4. *Wide area situational awareness (WASA)*—This involves monitoring and display of power-system components and performance across interconnections and over large geographic areas. The goal is to optimize management and the performance of network components so that issues and disruptions can be anticipated, prevented, or responded before they occur.
5. *Sensing and measurement technologies*—These technologies evaluate congestion and grid stability as well as monitoring network and customer side equipment in terms of health and power consumption. The following factors are some of the components used:
   a. *Smart meters*—These are used to monitor usage statistics and report them to utility companies, businesses, consumers, and third-party service providers. They replace the old analog meters and record real-time usage. They can also show how much power is being used at different times of the day along with the related power costs. Two-way communication on these meters also allows for power-outage notification as well as remotely disabling the service (if necessary).
   b. *Advanced metering infrastructure (AMI)*—This remotely measures, collects, and analyzes usage statistics from smart meters. AMI is similar to advanced meter reading (AMR) but is an upgrade (two-way vs. one-way meter reading).
   c. *Phasor management units (PMU's)*—These are high-speed sensors distributed throughout the network to monitor power quality and respond automatically to power issues.
   d. *Wide area measurement system (WAMS)*—This is a network of PMU's that provides real-time monitoring on a regional and national basis.
   e. *Advanced components*—These include excess electricity storage, fault tolerance, smart devices, and diagnostic equipment. Smart (intelligent) devices are useful for providing consumption feedback to customers in the home.
6. *Home area network/business area network (HAN/BAN)*—These networks look to address demand/response and consumer energy efficiency. These networks include mechanisms and incentives for businesses, utilities, industrial customers, and residential consumers to cut energy use during peak demand or when power reliability is questionable.

Now, we know something about power grids, the evolving grid, and future smart grids, and we can look at vulnerabilities, threats, and attacks on these utility grids.

Grid Vulnerabilities

The following factors are some of the noteworthy security vulnerabilities:

- Many current security vulnerabilities are basic and include a failure to install security patches and poor password management. The fixes in these areas are inexpensive.
- Unsecure software-coding practices used in control networks and excessive allowance of portal access into networks are some of the prevailing security gaps. Poor code quality leads to bugs and vulnerabilities that can make the grid fragile and unstable as well as vulnerable to attacks.
- Ineffective passwords and lack of proper encryption for communications and databases are the common problems as well.

- Smart grid components and technologies such as smart meters and AMI/AMR networks use wireless Wi-Fi and/or BlueTooth technologies to transport the usage data from consumers back to utility companies. There are pervasive security issues with Wi-Fi networks, and many organizations have banned their use or have implemented policies that have restricted their access to corporate networks. Bluetooth is an unsecure technology, and there are known scanning tools that allow for Bluetooth device discovery, operating mode, and strength of the device. Bluetooth 3.0 uses Wi-Fi radios, and Wi-Fi can be susceptible to wireless packet sniffers.
- The utilities that do not use wireless networks can still be vulnerable if employee laptops, handhelds, and smart devices are used. A tool called Karmetasploit can turn a wireless laptop into an access point that can associate wireless clients with it. Once associated, that client can be taken to a malicious service.

Threats in the Grid

a. *Hacking*—Hackers may want to get into systems for an intellectual challenge or out of curiosity. Their actions could have negative impacts on consumers and utility companies.
b. *Theft*—The consumers can monitor their electricity usage but if that information ends up in the wrong hands, the usage can point to patterns in the home during certain times of the day. Determining when a homeowner is out of the house can lead to burglaries.
c. *Extortion*—The grid can be exploited for money and power. Extortive malware can be used to hold a system or data hostage, to extort a ransom from an owner/user. A specific service can lock a user out of the system or can prevent access to critical data (or a combination of these). Consumer access to power can be prevented and a monetary demand could be used by an attacker to restore power.
d. *Power disruption due to vengeance and vindictiveness*—A remote disconnect feature can be used by a problem neighbor, or the neighbor can also perform a physical attack on a smart meter on the side of the home.
e. *Terrorism*—This could affect a large number of people and could cause massive attention for a cause. It can occur by both digital means and physically by bombings.
f. *Warfare*—The attacks can be used during war time by an enemy to cripple a country's infrastructure.
g. *Poor patch management*—Patches would not always install correctly, and these may be found during an audit, security assessment, or by a hacker. If these occur, a customer may receive billing errors or electricity can be shut off.
h. *Intentional threats*—The angry employees could attack the consumers.
i. *Activists*—They can use the grid as an additional avenue to attack certain manufacturers (e.g., fur manufacturers).

Threats by Confidentiality, Integrity, and Availability

Flick and Morehouse (2011) discuss threats related to confidentiality, integrity, and availability.

Confidentiality—This involves protecting the information from unauthorized disclosure. In the grid, this has the greatest effect on consumers. Utility companies store names, addresses, social security numbers, and usage data. The hackers can compromise the

database through a structured query language (SQL) injection on a website used by consumers to manage their accounts, monitor usage, and make payments. The hackers could obtain credit card numbers or bank accounts from customers who use online or automatic bill payment.

Integrity—This focuses on protecting the information from unauthorized modification. If the information is modified, it has the greatest effect on utility companies in terms of fraud or service theft. Once it is determined how to hack smart meters, the information can be placed on the Internet, and customers as well as hackers can defraud the utility companies by stealing services (underreporting to lower bills) or by fooling the utility company into thinking that they are selling more electricity (over reporting to get more credits) back into the grid.

Availability—This is attained when the service is protected from unauthorized interruption. This impacts the service provider as well as customers. The threats can be from script kiddies or people the victim knows and the threat could affect power to the home. A smart meter can be attacked through the wireless configuration on the router allowing access to the wireless network to change the default password and to shutdown power. When the victim goes back in with her password to reenable power, the password is no longer known.

Privacy Threats

1. *Identity theft*—Grid identities (IDs) and other information can be placed on the Internet and can be sold. The thieves can use personally identifiable information (PII) obtained to impersonate customers for fraudulent utility use, and this could affect customer credit reports.
2. *Personal surveillance*—Sensitive personal behavioral patterns can be revealed to expose customer schedules or personal details about their lives (interactions with others, medical issues, etc.). It can be determined whether a person lives alone, whether they leave the house vacant all day, whether they are senior citizens, have small children at home, and so on.
3. *Energy use surveillance*—This focuses on meter data used to show the specific appliance used in the home. It can report the number of gadgets in the home, whether there is an alarm system (and how often it is turned on), and so on.
4. *Physical dangers*—Real-time data can be used to cause harm. Domestic violence offenders/stalkers/abusers can use the information to relocate the former victims who have an urgent and continued need for privacy.
5. *Misusing data*—Utilities could misuse the data by providing them to third-party marketing firms or a previous homeowner's smart meter may have the data that were not wiped clean when they moved, and now, there are data known about the previous users.

Potential Attacks on the Grid

Attacking Consumers

1. *Attacking smart meters*—Smart meters are a basic, low-cost technology that hackers can purchase to take apart and learn about the communications network. The customers have physical and perhaps logical access to them as well. Additionally, these could be some common attacks (Flick and Morehouse, 2011):
   - Since smart meters are accessible through wireless networks or HANs, a tool such as network mapper (NMAP) (network scanner) can be used for transmission control protocol

(TCP) pings for port scans to identify active hosts with the common services that are running.
   - NMAP can be used for probing on TCP or User Datagram Protocol (UDP) ports to determine a response on an associated port with a service running that may have a weakness that allows access.
   - If a smart meter contains a web component that allows the users to view/change the usage information, vulnerability identification and verification can be performed against the web application.
   - There can be an attempt to identify the valid credentials for a service or a web application by using a dictionary attack or through brute force.
   - Wireshark can be used to determine the typical traffic patterns, and NMAP can be used to increase the traffic and overload the meter or infrastructure to achieve a denial of service.
   - The code can be used to create a buffer overflow, or an SQL injection can be used to access a command shell through an input validation weakness in a web application.
   - Kismet (wireless sniffing tool) can be used to compromise the confidentiality, integrity, and availability (CIA) of the data going across the network. If Bluetooth is used, Kismet just added a Bluetooth component to its wireless sniffing tool. IF radio-frequency identification (RFID) is used, there are plenty of RFID sniffing kits that can be easily created.
   - A physical attack of the smart meter on the side of the building can occur. If a camera is monitoring it, the attacker can mask her identity.
2. *Attacking smart devices*—Smart devices can be purchased off of the net or can be picked up at the local retail store and studied. These devices can send the data over the wireless network or readings over power line communications (PLC) to a gateway that runs a web server. The web server can interface with something such as Google Power Meter to monitor and present the usage information.
   - Wireshark can be used to get device IP addresses and NMAP can be used to try to get responses. NMAP can also be used to identify any services that are running on TCP or UDP ports.
   - The weaknesses in web applications can be identified, and web browsers can be used with the device uniform resource locator (URL) to guess the password. Wireshark can be used to look for unencrypted passwords or encrypted passwords that are weak or contain flaws.
   - Vulnerability scanners can be used to detect whether the devices are susceptible to the denial of service attacks. A well-constructed disk-operating system (DOS) attack can cause a blackout.

Attacking Utility Companies

Companies will most likely be attacked through the use of multiple attack vectors (a combination of application, network, social engineering attacks, etc.) (Flick and Morehouse, 2011):

1. *Network attacks*—Attacks can come from systems, networks, and subnetworks. The following factors can be used to determine network-security access:
   a. *Reconnaissance*—Passive testing can be used to determine utility company infrastructure. The perimeter of the IP address ranges assigned to the utility company could be

determined. IP addresses/host names can be determined for web servers, Domain Name System (DNS) servers, e-mail servers, routers, gateways, and virtual private network (VPN) concentrators.

b. *Discovery*—Ping sweeps and port scans can be used to identify rogue systems (perhaps test systems placed in the network and never removed). Network routes can be determined, and it could be determined whether networks are properly segmented with strict access control lists (ACLs). Smart devices on the customer side could be examined to determine whether they can connect directly to the generation domain of the utility company. If they do connect all the way in, that could be the path of least resistance for a hacker. Business partner networks could be examined to determine whether they have unrestricted access into the utility networks. If they do, a hacker's path of least resistance into the utility network may be through the partner network.

c. *Vulnerability identification*—Vulnerability scanners can be used to identify services running on open ports and then interrogating the service. The scanners can inspect a service banner that can tell the name and version of the service. On the basis of version information, the scanner will determine what security patch-based vulnerabilities a service may be susceptible to. Exploiting the vulnerabilities can cause DOS, disclosure of pertinent information, or allow for remote code execution.

2. *System attacks*—If firewalls and IPS/intrusion detection system (IDS) are in place, it will make it harder for attackers to remotely attack the company; so, hackers may resort to e-mail attachments and website attacks. SCADA devices support web servers, telnet, and file transfer protocol (FTP), so that attackers may try to get in through these areas. Legacy systems will have support issues as they age and security patches may no longer be created for new vulnerabilities that are discovered. Since dial-up modems are still used (typically as the backup for outages), remote access to critical systems may be fairly easy especially if default passwords are never changed.
3. *Application attacks*—Web servers, web applications, and services will be used and may provide an avenue for the attack. Web applications can allow for a single point of failure. Web services secure login that do not accept the hypertext transfer protocol text (HTTPS) protocol (this exists today) that is vulnerable. The injection of the malicious code (scripts, commands, or queries) against web applications can allow for administrator level access to allow an attacker to perform any type of query against any table in the database.
4. *Wireless attacks*—Discovery, device profiling, and exploits can be used on wireless networks (RF, Wi-Fi, BlueTooth, and cellular), wireless clients, and access points to identify unauthorized (rogue) clients, networks, and so on. The attacks in the network can cause DOS, revealing critical information, and bypassing perimeter controls to gain access to internal networks.
5. *Social engineering attacks*—The focus here should be on security awareness and training of company employees. The attackers can try to fool employees into revealing user names and passwords or can provide the attacker with additional access by
   - Impersonating an employee and calling the information technology (IT) help desk to change the password.
   - Impersonating a vendor to obtain proprietary information or for the purpose of sabotaging the equipment.
   - Dropping USB sticks with the malicious code in strategic locations to provide a way into the network.
   - Sending phishing e-mails to solicit confidential information.

6. *Physical attacks*—This could involve gaining access to buildings without a badge, copying or taking pictures of information that is out in the open, eavesdropping, stealing unlocked mobile devices, or using an available/unlocked personal computer (pc).

The attackers do not stick to only one type of attack, and they will use a combination of the above factors to attack the utility companies.

Federal Efforts to Protect the Grid in North America

Now, the vulnerabilities, threats, and potential attacks have been identified, and we can look at the efforts to protect the grid in North America. The following factors are the Federal efforts in place to achieve the proper protections:

1. *Federal agencies*—The following agencies are concerned with protecting the grid:
   a. *Department of Homeland Security (DHS)*—The DHS is tasked with creating a national infrastructure protection plan (NIPP) for the critical infrastructure and key resources. The plan must ensure that the infrastructure and resources are safe, secure, and resilient, through prevention, neutralization, and mitigation of deliberate efforts by terrorists to exploit or destroy the grid. The objective of the effort is to strengthen national preparedness, provide timely response to attacks, and allow for rapid recovery in the event of attack, natural disasters, or emergencies.
   b. *Department of Energy (DOE)*—This organization was established by the Federal Government in the late 1970s to organize fragmented regulatory processes and to create a national energy plan. The DOE is also responsible for monitoring and reporting on the security of smart grids. The objectives of the reporting are to determine how to make the grid less vulnerable to disruption, how to restore the integrity of the grid after disruption, how nationwide emergency communications can be facilitated after a local, regional, or nationwide emergency takes place, and what grid risks must be taken into account and how the risks should be mitigated. The DOE also established the Federal Energy Regulatory Commission (FERC) as an independent regulatory body within the DOE.
   c. *FERC*—FERC regulates the interstate transmission of natural gas, oil, and electricity, and has the authority to mandate reliability standards and impose penalties for noncompliance. FERC also issues orders for emergency measures to protect the reliability of the bulk power system and critical infrastructure when the President identifies a security threat to the grid.
2. *Federal legislation*—The following factors are some of the federal legislations designed to protect the grid:
   a. *H.R. Bill 5026*—This is the Grid Reliability and Infrastructure Defense Act or Grid Act. This establishes federal authority to address emergencies (with or without notice) if the President identifies an imminent threat to the bulk power system. It also establishes measures to protect the grid against the key vulnerabilities so that we are prepared if an emergency occurs and gives FERC the authority to protect portions of the grid that serve the top 100 facilities against a cyber threat or electromagnetic weapon attacks. Additionally, it allows FERC to bypass the standards setting process and issue orders to utilities to address security vulnerabilities not addressed by the standards. Finally, it

allows for requiring entities that own or operate large transformers to ensure adequate availability of replacements to promptly restore the operation, should a transformer be destroyed or becomes inoperable.

b. *H.R. Bill 2195*—This bill directs the Secretary of Homeland Security working with other security and intelligence agencies to conduct research and determine if networks critical to the operation of critical electricity infrastructure have been compromised. It also amends the Federal Power Act to direct the Secretary to make ongoing assessments to provide periodic reports on cyber vulnerabilities or threats to critical infrastructure, including AMI, and looks to enhance domestic preparedness in case of a cyber attack.

Standards Bodies and Standards for Protecting the Grid

The following factors are a couple of the important standard bodies for developing standards to help protect the grid in North America:

1. *National Institute of Standards and Technology (NIST)*—This is a federal technology agency that works with the industry to develop and apply technology and standards. NIST is responsible for coordinating the development of a smart grid interoperability framework and a plan to create standards to address the remaining gaps and the integration of new technologies, testing, and certification to ensure smart grid equipment and systems that conform to security and interoperability standards. It further engages utilities, vendors, consumers, and so on to achieve standards on architecture, priorities for interoperability and cyber security standards, and any plans to meet the remaining standard needs. NIST Publication SP800-82 is one of the important publication as it addresses how to secure industrial control systems that include SCADA, DCS and PLCs, and identification of the common threats and vulnerabilities of these systems as well as any countermeasures to mitigate the risks associated with the threats and vulnerabilities.
2. *North American Reliability Corporation (NERC)*—FERC established this organization to create reliability standards and enforce them through severe financial penalties for noncompliance. NERC developed Critical Infrastructure Protection Standards (CIP) that are mandatory for utility companies to comply with. The standards cover two security categories (electronic security and physical/personnel security) and are as follows*:

   CIP-002—This covers the critical cyber asset identification and documentation concerned with assets that enable the reliable operation of the grid. The assets must be identified through a risk-based assessment approach and documentation of the assessment is necessary along with the methodologies used, evaluation criteria, and procedures/processes. Risk-based assessments must be conducted annually.

   CIP-003—This states that minimum security management controls must be implemented to protect critical cyber assets. The core requirement here is the cyber security policy, and this policy must be available to all personnel responsible for, or having access to, critical cyber assets. The exceptions to the security policy are also covered here, and the exceptions must be documented and authorized by a senior manager.

* Reliability standards for the bulk electric systems of North America. North American Electric Reliability Corporation. Critical Information Protection Standards.

CIP-004—This is concerned with the personnel and training. A security awareness program must be developed and documented for those with access to critical cyber assets and quarterly updates must occur. A cyber security training program is also required for those accessing critical cyber assets. The training must address policies, procedures, and access control related to the assets and must be performed at least annually. Also required is a documented personnel risk assessment program that verifies social security numbers and looks at a 7-year criminal background check that is reassessed every 7 years or for a cause. Finally, a list of personnel authorized to access the critical assets must be maintained and reviewed quarterly or must be updated within 7 days due to any change in personnel access rights. Physical access must be revoked within 24 h for any terminated employee.

CIP-005—This covers protecting all access points on the electric security perimeter that houses the critical cyber assets. The perimeter and associated access points must be identified and documented, and technical procedures must follow a default deny policy (only ports and services required for operating and monitoring of assets within the perimeter are enabled). Access to the perimeter must be monitored and logged and a vulnerability assessment of the perimeter must be performed annually. The results of the vulnerability assessment must be documented and an action plan to remediate the identified vulnerabilities must be established.

CIP-006—This is concerned with developing and implementing a physical security program to protect critical cyber assets. The controls to manage access to the physical security perimeter on a 24/7 basis must be implemented. In this standard are specifics related to card keys, special locks, security personnel, and authentication devices. Physical access must be monitored and logged by human observation, alarm systems, video systems, or manual logging. Physical access logs must be kept for 90 days and testing of physical security must take place every 3 years. The records of the testing must be kept.

CIP-007—Processes, methods, and procedures for securing critical cyber assets residing within the perimeter must be defined, and testing procedures must be in place to protect against adverse effects from significant changes in cyber assets. The significant changes are patches, service packs, updates, and upgrades, and procedures must be in place to test the security of the changes. A patch management program must be defined and documented, and controls must be in place for antivirus software and antimalware tools. The use of shared accounts must be identified, and these accounts must be secured in case of personnel changes. Data-storage media must be destroyed or erased prior to disposal and records of disposed or redeployed cyber assets must be maintained.

CIP-008—This looks at incident reporting and response planning. Cyber security incidents related to critical cyber assets must be identified, classified, responded, and reported. The incident response plan must include classification procedures to determine which cyber incidents are reportable, actions required to respond to the incidents, a process for escalating and reporting incidents, a process to ensure plans that are updated within 90 days of any changes, a process to ensure plans that are reviewed and exercised at least annually, and all documentation of incidents must be made available for at least 3 calendar years.

CIP-009—This looks at disaster recovery. The recovery plans for critical cyber assets must be established and must ensure that the plans follow established business continuity, disaster-recovery techniques, and practices. The plans must include definitions for roles and responsibilities associated with recovery, and recovery plans must be tested annually.

> Backup storage, restoration processes, and procedures must be documented and backup media deemed essential for recovery must be tested annually.

Although these standards are a good practice and are necessary, many legacy systems and devices are not able to meet these requirements. NERC is formalizing the procedures to allow entities to submit technical feasibility exceptions (TFEs) for cyber assets that cannot comply with the CIP standards. Also, NERC has determined that today, many utilities are underreporting critical assets to avoid compliance requirements.

Security for the Grid

There are a number of areas that must be looked at when considering security for the grid and the goal of cyber security is on prevention.

General Security Practices

- Bottom-up risk assessment should focus on authentication and authorization as it relates to substations, intelligent electronic devices (IEDs), key management for smart meters, and intrusion detection for power equipment. Top-down risk assessment should look at logical interfaces for priority areas such as electricity transportation and storage, WASA, demand response, AMI, and DGM architecture. The identified risks can be avoided, transferred, mitigated, or accepted. The risks should first be avoided or mitigated, and avoidance involves implementing information security controls.
- The grid will be composed of multiple networks that belong to different organizations and individuals and there must be a strict set of interoperability standards to facilitate communications between the networks. Firewalls should be placed between networks, and trust relationships and segmentation must be established between components and networks so that a threat or vulnerability in one network does not have a chance of spreading across the entire grid.
- Intentional attacks on consumers by utility company employees could be minimized by implementing least privilege (access is both required and authorized) and separation of duties (only the level of access required to perform a specific job). An employee who is responsible for billing, for example, should also not be able to shut off power to the home.
- The IT principles and security objective priorities of confidentiality, integrity, and availability are important, but they have different priorities in power networks as availability of the network is most important followed by integrity and confidentiality. Mesh architectures capable of self-organizing and healing can be used to ensure integrity and availability under adverse conditions. Although confidentiality is least important today, it is becoming increasingly important as more data will be collected on consumers and privacy of customer information will be stressed. The effective privacy practices will be covered later.
- Well known, open-security standards used elsewhere should be leveraged. The security mechanisms operating at multiple layers of the protocol stack should be employed to deliver a layered defense (defense in depth), and the security framework should be flexible, adaptable, and expandable to address an evolving threat landscape.
- Securing the grid requires cooperation between utility companies, the federal government, suppliers, consumers, NERC, the DHS, and the DOE to name a few.

- Data security should focus on critical asset identification and documentation, data classification, encryption of data at rest and in transit, database access monitoring, alerting and reporting, and change control and configuration management. Data security programs should be all about policy, process, and people.
- Business continuity management needs to be stressed so that incidents that threaten the continuity of operations are able to be addressed. The critical infrastructure needs to be identified (along with the associated risks) and security controls need to be implemented to minimize impacts from disasters, security breaches, and DOS. Without business continuity management, it will be difficult to survive a coordinated and targeted cyber attack.

Technical Security Practices

The following factors are the technical security practices that can be implemented to protect the grid*:

1. *Threat modeling*—The developers of software or any other solution should identify the potential attack vectors on their deployable solution. They need to develop abuse cases and ponder how their software can be used for a malicious intent. Once the potential attacks are identified, controls can be implemented for mitigation and attacks can be looked at in terms of confidentiality, integrity, and availability.
2. *Segmentation*—The goal here is to minimize the impact of a successful attack through the use of stateful firewalls, for example, not through ACLs on switches or routers. Smart meter traffic, for instance, should be contained to one geographic location so that if there is an issue, it does not spread throughout various networks.
3. *Default deny firewall rules*—These rules should apply to all inbound and outbound connections, and outbound connections should only occur from systems that have direct access to the Internet. All the other systems should be forced through a proxy for protections such as content filtering and detection of malware.
4. *Code and command signing*—Crypto hashes to validate software authorization and code integrity need to be implemented. Otherwise, attackers can run the arbitrary code on smart meters and can issue commands on the infrastructure that will be trusted. Most smart meters deployed today do not use authentication before running updates or disconnecting the service from a customer.
5. *Honeypots*—These appear to be the production systems that can be used to identify and contain attackers. They can provide alerts when successfully attacked or compromised. This can be used to understand threats and types of probes by attackers.
6. *Encryption*—This needs to be used in the transport layers where customer data are transported and also needs to be used in databases and any removable media. For example, advanced encryption standard (AES) can be used in databases and removable media and laptops should also use encryption in case they are stolen and end up in the wrong hands.
7. *Vulnerability management*—The purpose here is to determine where configuration policies, procedures, and processes are effective and where they are not. Vulnerability scanning can be used to identify weaknesses and provides information to be used for dealing with threats. Knowing where the vulnerabilities exist will allow the utility to use a risk-based approach to deal with and manage the vulnerabilities.

* *Securing the Smart Grid*—Tony Flick, Justin Morehouse. Syngress/Elsevier, Inc., 2011, pp. 153–159.

8. *Penetration testing*—This is used to validate the risks associated with the identified vulnerabilities and should be reviewed quarterly.
9. *Source code review*—This is the review of the software source code for vulnerabilities before the software is released. This is to be done on all software developed internally or by vendors. This is used during the software development phase (using the SDLC) to fix the vulnerable code before the software goes into production.
10. *Configuration hardening*—This is the hardening of the system before it goes into production. A hardened system image should be used to build the system as opposed to trying to harden an image supplied by a vendor. Penetration testing and vulnerability scanning then needs to be performed regularly on the hardened system images.
11. *Strong authentication*—To authorize access to a resource, two of the three authentication categories need to be used:
    - Something you know (such as a password)
    - Something you possess (such as a smart card)
    - Something you are (such as a fingerprint)

    This prevents unauthorized access when one of the two authentication categories are compromised.
12. *Logging and monitoring*—This allows for information necessary to identify attacks as well as being able to reproduce an event in case of an incident. For example, an unsuccessful log in attempts on a website can illustrate trying to break into a customer account. Logging and monitoring should be used in application, operating system (OS), and network levels and should be used on intrusion detection and prevention systems.

Privacy Practices

The following factors are some of the things that can be done to ensure customer privacy:

- Limit the collection of personal customer data. It should be collected by lawful means and with the consent of the customer.
- The data collected should be for a specific purpose and should be complete, accurate, and updated. The data collected for one purpose should not be used for another purpose.
- Those that handle the data need to comply with privacy guidelines.
- Privacy impact assessments (PIAs) need to be conducted on use of personal information or new forms of PII. PIAs should be conducted annually and a copy of the results should be provided to the State's Public Utility Commission Office for review.
- A clearly specified notice should be sent to the consumers before collection, use, retention and sharing of usage data, and personal information. This should also be employed when new data are to be collected for a specific purpose, as well as if the collected data are to be used for new purposes.
- The data should only be divulged to parties authorized to receive them.
- The customers should be allowed to see their personal data and should be able to request correction of any inaccuracies.
- The data should only be linked with a location or a customer's account when used for billing, certain operation needs, or for restoring the service.
- The data need to be protected from loss, theft, unauthorized access, disclosure, use, copying, or modification.
- Privacy policy statements should be made available to the consumers.

Conclusion

The customer, utility company, and environmental benefits of evolving to a smart grid further outweigh the security challenges that these networks present. The requirements and adherence for compliance do not always equate to security, and 100% secure networks are not an achievable goal. But, following sound security practices and managing the risk to an acceptable level are prudent and the industry's goal is not on compliance but on preserving national security given the vulnerabilities, threats, and potential attacks that will become apparent.

References

Bakken, D., Hauser, C., Bose, A., Gjermundrød, H., Dionysiou, I., Johnson, R., Jiang, P., Sheshadri, S., and Swenson, K. 2003. Grid stat. Washington State University, School of Electrical Engineering and Computer Science, November 2003.

Evron, G. 2010. Stuxnet: An amateur's weapon. *Dark Reading.com*, October 15.

Flick, T., Morehouse, J. 2011. *Securing the Smart Grid*. Syngress/Elsevier, Inc.

Kingsbury, A. 2010. A "smart" electrical grid could secure the energy supply. *U.S. News and World Report*, April 7.

Further Reading

Abel, A. 2004. Government activities to protect the electric grid. *Congressional Research Service Report for Congress*, October 20.

Bain, B. 2010. House moves to protect grid from cyber threats. *Federal Computer Week*, June 10.

Bindra, A. 2010. Securing smart grid from cyber attacks. Smart-grid.tmcnet.com, August 4.

Bockman, A. 2010. An irksome tale: The battle to secure the smart grid. *Huffington Post*, 1–2, September 19.

Chari, N. 2010. Securing the smart grid. Tropos Networks, September 26.

Cisco Systems. 2009. Securing the smart grid. White Paper.

Coleman, K. 2010. Protecting the smart grid from cyber attack. Defensetech.org, June 17

Condon, S. 2009. Senators aim to protect electric grid from hackers. *Cnet News*, 1–2, April 30.

Coney, L. 2010. Privacy perspective on protecting the grid and consumer data topic: Smart grid cyber security and privacy. *Smart Grid Policy Summit*, 1–17, April 8.

Dark Reading. 2010. Landis+Gyr and Safenet team for smart grid security. *Dark Reading.com*, October 20.

Echols, M. and Sorebo, G. 2010. Protecting your smart grid. *Transmission and Distribution World*, 1–2, July 1.

Edison Electric Institute. 2010. *EEI Principles for Cyber Security and Critical Infrastructure Protection*. September 9. www.eei.org.

Evron, G. 2010. Stuxnet: An amateur's weapon. *Dark Reading.com*, October 15.

Federal Energy Regulatory Commission. 2010. Smart grid standards adoption: Staff update and recommendation. Item No. A-3, July 15.

Fehrenbacher, K. 2009. Securing the smart power grid from hackers. *Bloomberg Businessweek*, 1, March 23.

Gunther, E. 2008. DNP secure authentication—Essential to smart grid progress. *Smartgridnews.com*, November 18.

Higgins, K.J. 2010. Stuxnet heralds new generation of targeted attacks. *Dark Reading.com*, September 23.

International EMP Council. H.R. 2195: Critical electric infrastructure protection. www.empcouncil.org.

IO Active. 2009. IO Active verifies critical flaws in next generation energy infrastructure. Press release, March 23. www.ioactive.com.

Luallen, M. 2009. Securing a smarter grid: Risk management in power utility networks. A SANS White Paper, 1–17, October.

Mantooth, H.A. 2010. How can we protect the smart grid? *Connected Planet*, 1, May 4.

Mark, R. 2008. Electrical grid exposed to cyber-threats. *E-week*, 1–4, September 12.

Mills, E. 2010. Securing the smart grid. *Cnet News*, 1–5, April 9.
National Institute of Standards and Technology. 2001. Security requirements for cryptographic modules. FIPS Pub 140-2, May 25.
National Institute of Standards and Technology. 2010. Smart grid cyber security strategy and requirements. Draft NISTIR 7628. Smart Grid Interoperability Panel, Cyber Security Working Group. February.
North American Electric Reliability Corporation. n.d. Reliability standards for the bulk electric systems of North America. Critical Information Protection Standards.
Oracle, 2010. Protecting the electric grid in a dangerous world. Oracle White Paper, 1–17, April.
Phys.org. 2010. Dartmouth researchers help secure the power grid. January 26. www.physorg.com.
Science Daily, 2006. Securing America's power grid. *Science Daily*, June 26.
Swanson, S.A. 2010. Securing the smart grid. *Scientific American*, 1–2, May 13.
TD World, 2010. DOE announces latest efforts to address cybersecurity. *Transmission and Distribution World*, 1–2, September 24.
U.S. Federal Energy Regulatory Commission. 2009. Smart Grid Policy—128 FERC 61,060. 18 CFR, Chapter 1, July 16.
U.S. House of Representatives. 2009. H.R. 2195, 111th Congress, 1st session, April 20.
U.S. House of Representatives. 2010. Grid Reliability and Infrastructure Defense Act—H.R. 5026. 111th Congress, 2nd Session, April 14, 2010; *Congressional Record*, 1–12, June 9.
Wikipedia. Smart grid. http://en.wikipedia.org/wiki/smart_grid.
Wikipedia. Stuxnet. http://en.wikipedia.org/wiki/stuxnet.
Wikipedia. Unified smart grid. http://en.wikipedia.org/wiki/unified_smart_grid.

Network Attacks and Countermeasures

Chapter 2

Attacks in Mobile Environments*

Noureddine Boudriga

Contents

The threats to mobility space are increasing significantly and having far greater impacts on users and organizations alike. In many ways, we face problems in mobility space security similar to what was experienced at the initial stages of the Internet. There are vast and evolving levels of communication options, growth in the number and diversity of applications, and rapidly evolving platforms where security vulnerabilities explode disproportionately year after year.

One of the reliable methods for securing systems is understanding the threats, vulnerabilities, and the systems that you are seeking to secure. Knowing more about the types of attacks and threats and employing better tactics will help in troubleshooting the security challenges with mobility.

Basic Attacks

Basic attacks can be classified into four major classes, namely illicit use, wireless spoofing, man-in-the-middle attacks, and denial of service attacks. A description of the features of the basic attacks is given as follows.

Class of Illicit Use Attacks

Illicit use is a passive attack that does not cause damage to the physical network. It involves an attacker that is close to an access point (AP) (or base station [BS]) and obtains information extracted from the traffic the attacker has exposed. Illicit use includes the following attacks:

- *Wireless network sniffing:* When wireless packets traverse the air, attackers equipped with appropriate devices and software can capture them. The sniffing attack methods include the following:
  - *Passive scanning:* This attack aims at listening to each wireless communication channel and copying the traffic flowing through it, for future analysis. It can be done without sending information and can use some tools such as the radio frequency (RF) monitors, which allow copying frames on a channel.
  - *Identity detection:* This attack consists of retrieving the identity of important entities occurring in a wireless network (such as the identity of the AP, in a wireless LAN [WLAN]) by scanning specific frames such as the frames of the following types: beacon, probe requests, probe responses, association requests, and reassociation requests.
  - *MAC address collection:* To construct spoofed frames, the attacker has to use legitimate MAC addresses. These addresses can be utilized for accessing active AP by filtering out the frames with nonregistered MAC addresses.
- *Probing and network discovery:* This attack aims at identifying various wireless targets. It uses two forms of probing: active and passive. Active probing involves the attacker actively sending probe requests with no identification using the Service Set Identifier (SSID) configured to solicit a probe response with SSID information (and other information) from any active AP. When an attacker uses passive probing, he listens on all channels for all wireless packets.
- *Inspection:* The attacker can inspect network information using tools such as Kismet and Airodump. He could identify MAC addresses, IP address ranges, and gateways.

Wireless Spoofing

The spoofing intent is to modify identification parameters in data packets for different purposes. Typical spoofing attacks include the following:

- *MAC address spoofing:* MAC spoofing aims at changing the attacker's MAC address to a legitimate MAC address. This attack is easy to launch because some client-side software allows the user to manipulate his MAC addresses.
- *IP spoofing:* IP spoofing attempts to change the source or destination IP addresses by talking directly with the network device.
- *Frame spoofing:* The attacker injects frames with spoofed content. When the network lacks authentication, spoofed frames cannot be detected.

Man-in-the-Middle Attacks

This attack attempts to insert the attacker in the middle (MITM attack) of a communication for purposes of intercepting a client's data and modifying them before discarding them or sending them out to the real destination. To perform this attack, two steps have to be accomplished. First, the legitimate AP serving the client must be manipulated to create a "difficult to connect" scenario. Second, the attacker must set up an alternate rogue AP with the same credentials as the original for purposes of allowing the client to connect to it. Two main forms of the MITM exist: the eavesdropping and manipulation MITM attacks. Eavesdropping can be done by receiving radio waves on the wireless network, which may require sensitive antenna. Manipulation requires not only having the ability to receive the victim's data but then be able to retransmit the data after changing it.

Denial of Service Attacks

Denial of service (DoS) attacks aim at denying or degrading the quality of a legitimate user's access to a service or network resource. It also can bring down the server offering such services itself. DoS attacks can be classified into two categories:

1. *The disabling services attacks:* A DoS attacker makes use of implementation weaknesses to disable service provision. Weaknesses that are used with these attacks include buffer overflow.
2. *Resource undermining:* Undermining can be achieved by causing expensive computations, storage of state information, resource reservations, or high traffic load.

The techniques used in DoS attacks can be applied to protocol-processing functions at different layers of the communication architecture. DoS attacks can threaten the services offered to mobile users (e.g., servers offering specific information, or servers of specific companies) and the communication infrastructure itself. Especially, specific access resources such as bandwidth can represent a serious problem (since it most likely will remain a scarce resource in access networks). DoS attacks can target different network layers as explained in the following:

- *At the application layer:* DoS occurs when a large amount of legitimate requests are sent. It aims to prevent other users from accessing the service by forcing the server to respond to a large number of request transactions.
- *At the transport layer:* DoS is performed when many connection requests are sent. It targets the operating system of the victim's computer. The typical attack in this case is a SYN flooding.
- *At the network layer:* If the network allows associating clients, an attacker can flood the network with traffic to deny access to other devices. Typically, this attack is performed by allowing one among the following three tasks:
  - The malicious node participates in a route but simply drops several data packets. This causes the deterioration of the connection.
  - The malicious node transmits falsified route updates or replays false updates. These might cause route failures, thereby deteriorating performance.
  - The malicious node reduces the time-to-live field in the IP header so that packets never reach destinations since they are dropped by other nodes before destination.

- *At the data link layer:* DoS targeting the link layer can be performed as follows:
  - Since we assume that there is a single channel that is reused, keeping the channel busy in the node leads to a DoS attack at that node.
  - By inducing a particular node to continually relay spurious data so that the battery life of that node may be drained. An end-to-end authentication may prevent these attacks from being launched.
- *At the physical layer:* This kind of DoS can be executed by emitting a very strong RF interference on the operating channel. This will cause interference to all wireless networks that are operating at or near that channel.

Distributed DoS Attacks in Mobile Communications

To make DoS threats worse, attackers have developed effective tools to coordinate distributed denial of service (DDoS) attacks that can be launched and coordinated from a large number of sites, systems, and devices. A DDoS attack is distinguished from a common DoS attack by its ability to launch its actions in a distributed manner over the wireless communicating system and to aggregate these forces to create dangerous traffic. According to different reports including the annual CSI computer crime and security report, the DDoS attacks have induced large financial costs to companies in recent years. In addition, they cause damage to consumer confidence in e-commerce of impacted organizations.

There are various types of DDoS attacks. They all share the same typical structure that is depicted in Figure 2.1. The attacker, in a DDoS, first gains control of several master computers connected to the wireless network by hacking into them, for example. Then the master computers gain control of more computers (zombies) by different means. Finally, a message is sent by the attacker to synchronize all zombies to send the required traffic to the victim.

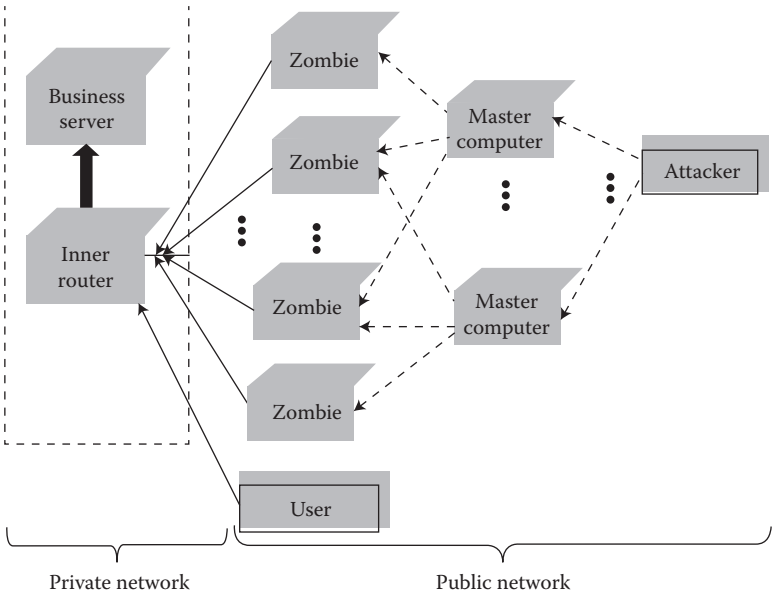


Figure 2.1 Typical DDoS structure.

In Figure 2.1, we first describe two examples of mobile systems that are targeted by DDoS attacks. Then we present some of the countermeasures that should be provided to protect against DDoS.

Targeted Environments

Two wireless communication systems are of interest to DDoS attackers, the wireless extended Internet-based networks (WEIN), where wireless technology is used only for the last mile, and the ad hoc networks (AHN), which represent, in the opinion of a large number of experts, the best architectures against DDoS attacks, since they have no central nodes and may implement severe admission policies making it very difficult for malicious users to enter into the communication infrastructure.

An example of WEIN is a network that is able to connect mobile devices to fixed networks via RF channels using the traditional Client/Server architecture and the existing transport layer protocols; for example, TCP. All the DDoS attacks achievable in the wired Internet are still feasible in the WEIN.

DDoS targeting WEIN and mobile ad hoc networks include, but are not limited to, the following attacks:

- *Attacking the wireless Internet content servers:* Since mobile devices have little computation and communication capabilities, a DDoS attack, even launched by a small number of powerful fixed terminals, can effortlessly disable a large range of mobile devices. Wireless Internet content servers, such as the WAP server, the wireless game servers, and the mail server, are often optimized for small throughput and timely response. They are particularly vulnerable to DDoS attacks compared with traditional wired servers. Furthermore, new forms of DDoS attacks may emerge taking advantage of the attractive features presented by the WEIN and ad hoc networks.
- *DDoS attacks on radio spectrum:* Often, the limited availability of radio spectrum is the bottleneck in a mobile network. Even if license-free RF bands are used and pico-cell-based (or reduced area) technologies are employed to expand transmission rates, the radio spectrum is still a scarce resource as the number of users and the demand for bandwidth is increasing tremendously. A DDoS attack can deliberately coordinate mobile devices to send out synchronized traffic to easily consume all spectrum resources or (at least) significantly reduce the capacity of any communication channel offered by the networks.
- *Attacks aiming at avoiding tracing back DDoS:* Some of the WEINs, such as the mobile IP protocol-based networks, present weaknesses that a DDoS attacker can use to launch attacks. For example, the Mobile IP protocol requires two IP addresses: the home address and the care-of address. The home address is permanently assigned to a mobile device, while the care-of address is temporarily assigned by the visiting foreign network. This allows a mobile device to send IP packets using its fixed home address, even when it is roaming, while applying the Non-Disclosure Method (NDM), which gives mobile users control over the revelation of their location information. Consequently, victim sites will find it hard to trace sources of DDoS attacks.
- *DDoS attacking devices using aggregated traffic:* Although the bandwidths used for the transmission in WEIN and ad hoc networks are much lower than those in wired networks, potential DDoS attacks are feasible mainly because of the fact that a large set of mobile devices can be involved. In particular, any wireless data packet traffic is a potential path for DDoS attacks.

Defending against DDoS Attacks

In the event of a typical DDoS attack, the victim alone cannot effectively defend itself. Cooperation among all involved parties is necessary. Typical methods to protect against DDoS attacks focus on effective coordinated technological solutions. There are three major types of coordinated technological solutions: (1) improving the security of all relevant devices; (2) enhancing the user-level traffic control; and (3) coordinating filters and tracing back methods.

Improving the Security of the Relevant Devices

Before initiating an effective DDoS attack, the attacker needs to involve enough zombie devices to secure the ability to generate sufficient traffic. An ineffective and direct countermeasure is to secure all devices to make it difficult for the attacker to install and take control of a large number of zombies. An alternative and effective solution would be to selectively secure those devices that have high traffic throughput, such as routers in the WEINs or the clusterhead nodes in the ad hoc networks.

Mobile User-Level Traffic Control

The traffic control, at the user-level, can be achieved by a set of traffic control rules. For example, the mobile user can set up a rule that fixes a daily traffic limit that is high enough not to disturb the normal activity of the user, while the unusually large traffic is stopped and may trigger an alarm (to the user or to a network administrator) for a subsequent diagnosis. The traffic control rules can also describe the data to be dropped or delayed if the network is experiencing congestion. An alternative solution can use a timestamp model to control traffic even when user devices are hacked.

This technique, however, experiences some drawbacks, including the fact that user-level traffic control rules for a specific network device need to be protected more securely than the network device itself. Edge routers in the WEINs are the perfect hosts for coordinating user-level traffic control rules. On the other hand, the designation of a host for traffic control rule coordination is more complicated in a mobile AHN, since no node (including the clusterheads) is more likely to be in a central position than another to host the rules.

Coordinated Filters and Tracing Back

Wireless Internet service providers in the WEINs can try to overwhelm the DDoS attacks by identifying the attacking traffics and stopping them using coordinated filters, whose aim is to stop the traffic as early as possible, along the attacking paths, to prevent the damage from aggregated traffic. For a mobile AHN, the filtering is not directly applicable due to the symmetric structure of the AHN. However, a dynamic voting mechanism may play an essential role to select those in charge of performing this function.

Cost-effectiveness arises as a crucial issue in the defense against DDoS attacks, because it may require the update of the current network infrastructure. Several advanced network management technologies have been proposed to address the traffic control problem. The use of these technologies will significantly reduce the costs and risks in designing future WEINs. In particular, policy-based networking (PBN) represents a promising technology for implementing the usage-based fees practices to deal with DDoS attacks. PBN provides rules that describe actions to take when specific conditions occur. These rules are able to control critical network resources such as bandwidth,

quality of service (QoS), security, and Web access across heterogeneous networks. Thus, it allows congestion to be under the control of a globally coordinated structure.

In the PBN typical scheme, two components can be distinguished for the traffic control: the policy enforcement point (PEP) and the policy decision point (PDP). While the Wireless Location Register/Authentication Center is a PDP with additional functionality such as accounting and policy information storage, the PEP accepts or denies requests appropriately, at the network border points. PDPs and PEPs can exchange policy information through secure and reliable channels to achieve efficiently their roles.

Mobile Malware

Basics on Malware

Malware (or malicious software) can be any malicious, unauthorized, or unexpected program (or code) that aims at realizing unauthorized actions on a computer, network components, or a mobile device. Some examples of the actions that a malware can perform include spying on wireless traffic, recording private communications, stealing and distributing private and confidential information, disabling systems, and erasing files. Malware can be divided into eight different categories:

1. *Worms:* A worm is a program that makes copies of itself (by various means including copying itself using email or another transport mechanism). A worm may damage or compromise the security of the visited (or infected) computer by executing special actions.
2. *Zombies:* A zombie is a program that secretly takes over another Internet-attached computer and then uses that computer to launch attacks that are difficult to trace to the zombie's creator. Zombies can be used to launch denial of service attacks, typically against targeted Web sites. The zombies can be installed on hundreds of computers belonging to unsuspecting third parties.

   They are then used synchronously to overloading the victim target by launching an overwhelming onslaught of Internet traffic.
3. *Viruses:* A virus is a sequence of code that is inserted into another executable code, so that when the regular program is run, the viral code is also executed. The viral code causes a copy of itself to be inserted in one or more than one program. Viruses are not distinct programs; they cannot run on their own and need to have some host program, of which they are a part, executed to activate them.
4. *Trojan:* A Trojan is a malware that performs unauthorized, often malicious, actions. The main difference between a Trojan and a virus is the inability to replicate itself. Like a virus, a Trojan can cause damage or an unexpected system behavior, and can compromise the security of the visited systems; but, unlike viruses, it does not replicate. A Trojan looks like any normal program, but it has some hidden malicious code within it.

   Often, a Trojan is composed of two parts, a client part and a server part. When a victim executes a Trojan server on his machine, the attacker then uses the client part of that Trojan to connect to the server and start using it based on TCP or UDP, for example. When a Trojan server runs on a victim's computer, it (often) tries to hide somewhere on the computer; it then starts listening for incoming connections from the attacker on one or more ports; then attempts to modify the system registry or use some other auto-starting method.

Most Trojans use an auto-starting method that allows them to restart and grant an attacker access to the infected machine.

5. *Logic Bombs:* A logic bomb is a programming code inserted secretly or intentionally. The bomb is designed to execute (or explode) under special circumstances, such as the amount of time elapsed since an event has occurred. It is in effect a delayed-action computer virus or Trojan. A logic bomb may be designed to display a fake message, delete data, corrupt data, or have other undesirable effects, when executed.
6. *Back Doors:* A back door is a secret entry point into a program that allows someone who is aware of the back door to gain access without going through the usual security access procedures.
7. *Phishing Scam (PS):* A PS is a fraudulent Web page, an email, or a text message that attracts the unsuspecting users to reveal sensitive information such as passwords, financial details, or other private data.
8. *Spyware:* Spyware is software that reveals private information about the mobile user or its computer system to eavesdroppers.

The first example of a mobile malware for mobile cellular phones was built in June 2004, for the Symbian operating system. The antivirus companies now have thousands of Trojans and worms for mobile phones in their antivirus databases, and new malicious programs have become a constant stream.

Let us also notice that today's mobile malwares are very similar to computer malwares in terms of the techniques they can use. However, while it took computer viruses over two decades to progress, the mobile viruses can cover the same ground in a few years. No doubt, mobile malware is the most quickly evolving type of malicious code. A short list of the actions that a mobile virus can do includes, but is not limited to, the following actions:

- Block memory cards
- Combat antivirus programs
- Infect personal files
- Modify icons and system applications
- Install "false" or nonoperational fonts, applications, and malicious programs
- Steal data and send messages to other users

A study by McAfee during late 2012 demonstrated a substantial increase over past years and reports on the threats facing mobility devices and solutions. As in the past, the Android OS is the most popular target for attackers and malware. The mix of threats and attacks included SMS—sending malware, mobile botnets, spyware, and destructive Trojans.

According to the report, McAfee's database of mobile malware increased from 1900 unique signatures in 2011 to more than 13,500 in 2012. Conversely, a report from Juniper during the same timeframe reported as many as 28,472 different forms of mobility malware. Much of the implications for Android platforms deal with drive-by downloads, botnet clients controlled by Twitter-based command and control systems (CC), and recent scenarios that infect SD cards on phones damaging or destroying data, such as photos, documents, and other stored media.

According to the Juniper mobility threat report of 2012, their researchers highlighted three key areas:

- There is more mobile malware than ever before, particularly on the Android platform. The combination of Android's dominant market share and the limits of control over the apps

being developed and appearing in stores provided ample means and incentive for malware developers to focus on the platform.

- Mobile malware has gotten a great deal more sophisticated and smarter. Attackers continue to hone their craft by discovering new ways to exploit vulnerabilities and human behavior for profit across multiple mobile platforms and devices.
- There is a low barrier to entry for attackers seeking to impact the mobility space. Applications have become the primary method for hackers—and application stores are fast becoming the prime delivery mechanism for infected applications. As mobile users increasingly download apps, there has been a boom in the number of developers and as a result a vast increase in the number of attackers leveraging the mobile app ecosystem.

In addition to the expansion and increasing sophistication of mobile threats, the problem does not appear to be getting much better according to a Gartner report in 2012 that highlights smartphone purchases have exceeded those of PCs. By the end of 2012, there were 650 million smartphones in use and more people will use smartphones and smart mobile devices to access the Internet than PCs.

Examples of Mobile Malware

Nowadays, mobile phones are equipped with well-performing operating systems (OS) such as the Android, Symbian, Microsoft Windows Phone, BlackBerry, and Apple IOS. These OS present interesting features such as built-in cameras, high-resolution color screens, wireless data access, MP3 players, email services, and a wide range of apps that can be used for virtually anything. They are also equipped with Bluetooth or other wireless technologies, making them directly accessible from computers and other similarly enabled devices. It is expected that mobile malware will continue to represent a significantly expanding threat as device power, complexity, and integration continue.

DOMAIN

3

INFORMATION SECURITY AND RISK MANAGEMENT

Security Management Concepts and Principles

Chapter 3

Security in the Cloud

Sandy Bacik

Contents

Looking outside the organization to gain increased competitiveness, efficiency, flexibility, and, potentially, lower cost is not new—it is simply outsourcing. So why is there so much hype about cloud computing? Well, taking virtual machines containing critical, sensitive, and proprietary information and applications off premise to public and shared environments creates new security challenges. We have relied on network perimeter defense as the method to protect information assets stored in a data center. This off premise virtual machines may also revoke compliance and breach security policies. Information technology (IT) is recognizing that competitive advantages, expanded capacity, failover flexibility, and, possibly, cost savings just cannot be passed up and are looking to cloud computing services asking

1. Will my company still have the same security policy and standards control over my applications, information assets, and services?
2. Will my company still be compliant and can I prove it to my auditors?
3. Can I prove to my company and my customers that we are still secure and meeting our service-level agreements?

All these questions and more need to be answered when thinking about, when selecting, and when using a cloud service provider. The securing of organizational information assets may now become a challenge for information, where accepting a risk may be the only possibility.

In the past, a "data center" evoked images of massive server farms and other devices locked behind closed doors, where electricity and cooling are as important as network security to maintain the confidentiality, integrity, and availability of information assets. With a data center, perimeter controls are the most common approach for data center security. This typically included firewalls, demilitarized zones (DMZ), network segmentation, intrusion detection/prevention systems (IDS/IPS), standards, and network monitoring tools. IT moved to a virtual environment with the data

center to allow device expansion to get more computing power from the underutilized capacity of physical servers. This enabled the traditional data center footprint to shrink to enable cost savings and move to a "greener" environment with server consolidation. Now, IT and the business are extending virtual machines to public clouds causing the enterprise network perimeter to evaporate and impact on the securing of informational assets. With the inability of physical segregation and hardware-based security to deal with attacks and unauthorized access between virtual machines, the need for security controls moves to focus on the server and information assets directly, rather than the perimeter.

On the surface, the security requirements for cloud computing services appear to be the same as a traditional data center—apply strong network perimeter security. However, there is no longer a formalized perimeter to protect. Some of the key items that need to be considered with cloud computing services are as follows:

1. *Administrative access to servers, applications, and information assets.* With a traditional data center, an administrator could always walk to the console or know the network path to administer a device. In the cloud environment, administrative access is now conducted across the Internet or by a third party, increasing exposure and risk. Restricting that administrative access and monitoring this access becomes more important.
2. *Virtual machines and the virtual sprawl.* We know that virtual machines are dynamic and can quickly be reverted to previous instances, paused, and restarted relatively easily. They can be cloned quickly and moved between physical servers. This dynamic sprawl might make it difficult to maintain consistent security and configurations. Vulnerabilities and misconfigurations may propagate and go undetected. It is also difficult to maintain auditability of virtual machines at any given time. In a cloud environment, we need to be able to prove the security state of a system, regardless of the location or proximity to other, potentially insecure, virtual machines.
3. *Vulnerability exploits and virtual-to-virtual attacks.* With a virtual environment of the cloud computing environment, all servers use the same operating system. The ability for an attacker or malware to remotely exploit these virtual systems can spread rapidly and increase the risk to information assets. A large virtual environment can increase an attack surface, particularly when different cloud customers on the same virtual device have different security requirements to secure their virtual environment, thus potentially opening a back door to your systems from your virtual neighbor.
4. *Dormant virtual machine security.* Unlike a physical machine, when a virtual machine is offline, any application accessing the virtual machine storage will continue to work and may be susceptible to malware. And a dormant virtual machine does not have the ability to run security scans.
5. *Patch management.* The cloud computing service nature may create issues for patch management. Once subscribed to a cloud computing service, that patch management is not in the hands of the cloud provider, but is in the hands of the cloud subscriber. What happens if a cloud tenant in the same virtual environment does not have patch management?
6. *Policy and compliance.* Where is the virtual machine located in the cloud? The organization still needs to prove compliance for things like Payment Card Industry (PCI), Health Insurance Portability and Accountability Act (HIPAA), Statement on Auditing Standards (SAS) 70, Statement on Standards for Attestation Engagements 16 (SSAE16) Service Organization Control (SOC) 1 or SOC2, and International Standards Organization (ISO) 27000, regardless of where the virtual machines reside.

7. *Perimeter protection and zoning.* The enterprise perimeter vanishes when computing power goes into the cloud. The traditional firewall and network segmentation cannot reach the cloud computing servers, or policies are no longer controlled by the tenant, and are now in the hands of the cloud provider. To establish zones of trust, the virtual machines need to be self-defending, moving the enterprise perimeter to the individual virtual machines.
8. *Rogue corporate resources.* If IT and information security is not involved with the business in selecting applications and services, it might be possible for a business unit to sign an agreement with a cloud provider and bypass IT and information security altogether. It means that organizational information assets are outside the enterprise perimeter and information security does not know when or what might have happened to them.

Things like a firewall, IDS/IPS, integrity monitoring, log monitoring, secure deployments, and auditing become more important for securing the organizational information assets.

There are different types of cloud computing service providers. Whether private or public, cloud computing networks have the following three core components:

1. *Infrastructure as a service (IaaS).* In the traditional business environment, a user's day-to-day computing resources are held on one server at one or more controlled locations. The infrastructure is fixed. With cloud computing, the infrastructure is provided to the organization in an "on-demand manner," hence the term "infrastructure as a service" (IaaS). The "as a service," or utility, element is driven by the ability to monitor resource utilization and bill the customer based on units, whether processor cycles, megabytes of bandwidth or throughput, storage read/writes, or number of virtual machines consumed. This option may be good for small to medium business without a formal IT or information security department.
2. *Platform as a service (PaaS).* This service builds on IaaS with an additional layer of capability that allows organizations to develop, build, and deploy their own applications to support their own specific business needs. This option may be good for small to medium business without a formal IT or information security department.
3. *Software as a service (SaaS).* This service allows organizations to use a fully managed application, such as ERP, CRM, file sharing, collaboration, and e-mail/calendar, over a public or private network, without owning the software or systems required to run it. The software remains the property of the service provider and the user pays for access, either by annual subscription or on a pay-per-usage basis. Much application vendor support is moving to the SaaS model.

For organizations selecting one or a combination of cloud computing services, the organization at a strategic level needs to think about and consider the following before selecting a cloud computing service provider:

1. Does the service you are considering meet the business availability need? What information can the provider give about historical and recent service availability? What investment has the provider made in resilience and high availability?
2. What service-level agreements does the provider offer? What compensation is given if the service is lost or there is a breach? Remember, this will be a service credit and will not cover consequential loss.

3. Does the service need to comply with any regulatory requirements? Where will your data reside, and if that will be outside of your operational markets, state, province, territory, or country, is that acceptable?
4. Does the service meet and exceed the requirements of the organizational IT/data security policies? Or does it fall short? It is important to understand whether the service is offered within a private or public cloud. A private cloud has inherent security advantages because the data are stored within a provider's closed environment, such as a secure data center.
5. Where is the data actually stored and who has access to the data? What happens to the data when production tasks are completed? How is it archived for regulatory requirements? How can archives be accessed? How is the data finally destroyed?
6. How viable is the service provider? It is important to select a provider with sufficient resources to provide high levels of availability, resiliency, and security that businesses require. Is cloud computing part of the provider's core business, or a new venture that could fail if it does not attract and retain sufficient customers? The cloud needs multiple, highly resilient data centers with very strong network links between them.

See Appendix A for more detailed questions to help assess the risk of selecting a cloud computing service and provider. Ensuring information security remains proactive in reviewing the responses from the cloud service provider who can assist the organization in limiting the information asset risk when using the cloud computing services.

After the cloud computing service and the provider agree upon an acceptable level of risk to the organization, there are certain things that should be considered when developing a contract for the cloud computing service with the cloud computing provider. The following are good items to ensure in the cloud computing service contract: (*Note*: Please ensure that Sourcing and Corporate Counsel are involved with the wording of these items in the organization.)

1. Prohibit the cloud service provider the right to make use of company-owned content not documented in the contract.
2. Security and privacy shall require the cloud service provider to
   a. Keep company information secure and private from other tenants
   b. Disclose current and future updated security, privacy, regulatory, and recovery policies, standards, and procedures to company
   c. Ensure secure communications for company information and service exchanges
   d. Notify the company through a formalized change control, patch management, and configuration management process of any changes potentially affecting the company service and information
   e. Provide security services such as IDS/IPS, integrity monitoring, log inspection, malware protection, antivirus protection for company information and service
   f. Provide monitoring and management of all activity, such as uploading and downloading of files, setting access permissions for individual users and entire departments, password protection and time-based restrictions for shared files, generation and export of files and user activity, and file access for internal and external users
   g. Provide separation of duty assignment for the creation, maintenance, use, and disposal of company data
   h. Provide company IT with the standard device, application, and data repository configurations for the company service

   i. Provide company with incident response services and information should an issue or incident occur with the company service or information
   j. Provide the company IT with notices of issues and incidents affecting or potentially affecting the company service and information
3. *Data location:* The terms of use shall commit the service provider to disclose the physical location and placement of company services and information, especially if regulations require the information to reside in a specific country, state, province, or territory.
4. Require the service provider to regularly back up and be able to recover company information.
5. Require the service provider to fully disclose certification of any content, including distributed or backup copies that the company has intentionally deleted from their use of the service.
6. Require the service provider to disclose any issues or incidents that may impact or may have impacted company data.
7. Require the service provider to perform an SAS 70 Type II and Statement on Standards for Attestation Engagements 16 (SSAE16) Service Organization Controls Report (SOC) 1 and SOC 2 Type I and Type II Reports on an annual basis and disclose the report, findings, and remediation plans to the company.
8. Require the service provider to a "right to audit" clause by the company of the service provider's internal controls on company data with no more than 30-day notices. The company shall work with the service provider to remediate any issues discovered.
9. Require the service provider to a "right to test" clause by the company of the company service and data within the care of the service provider without providing notice to the service provider. The company shall work with the service provider to remediate any issues discovered.
10. Require the service provider to employ data formats that make it readily possible for company to remove our data exports or copies of content, from the service and use it in other places or with other applications.
11. Require the service provider to provide assurances that applicable international, federal, state, local, tribal, and territorial laws and regulations for data production, separation of duties, and data classification are implemented and maintained.

Many cloud service providers will not commit to everything listed above, so what items are a requirement to management and information security to ensure the protection of the organizational information assets? Continue to negotiate to an acceptable level of risk to organizational information assets.

Essentially, all the items in this chapter are to ensure that information security is engaged with the business and IT when selecting a cloud service provider. The following are a summary of the steps for information security to ensure there is limited risk to organizational assets:

1. Start security planning early. Make sure the organization has policies, standards, and procedures for researching, selecting, and procuring cloud computing services.
2. Research potential vulnerabilities. Make sure the checklist for researching and selecting a cloud computing service and provider contains identifying the potential vulnerabilities with the service and provider so the organization can accept or mitigate the risk to organizational assets.
3. Make sure the policies, standards, and procedures for cloud computing services cover
   a. Information in motion, in process, and at rest
   b. Platform-specific security

Appendix A: Cloud Computing Service Provider Risk Analysis Questionnaire

1. Overview
   a. Describe the cloud computing business services you will be providing.
   b. Please provide a detailed description of how information will flow between the cloud computing services and company (include any proposed network, system, and process flow diagrams).
   c. Will company data be stored or accessed by any offshore facility? If yes, please describe.
   d. Briefly describe the security components of your cloud offering, including end-user benefits.
   e. What is your solution architecture, and how is security integrated into your cloud offering?
   f. How does your security offering help in either establishing or enforcing trust in the cloud?
   g. What unique and differentiated capabilities do you offer that help protect data and infrastructure in the cloud?
   h. How does your solution work with other providers' solutions to help build a chain of trust from the application user's interfaces to the underlying hardware?
   i. How does your security offering simplify auditing and regulatory compliance?
   j. How does your solution take the anxiety out of moving to the cloud?
   k. Why should I select your solution over the others?
   l. Do you have a method for demonstrating ROI for your cloud offering?
   m. Are there security concerns that your solution does not address that you think the industry still needs to solve?
   n. I am just beginning to investigate cloud security. What advice can you give me, and what steps should I take to make sure I am covering all my bases?
   o. What tools do you offer to establish, maintain, and protect identity in the cloud?
   p. What services do you have for federating identity between clouds (public and private)?
2. Security Policies
   a. Has management published a complete set of policies and procedures that support the information integrity objectives of the organization?
   b. Is security awareness training mandatory, and does the training include the concepts of privacy, confidentiality, integrity and availability of data, corporate security policies, information protection standards, and privacy awareness?
   c. Does your company have a formal incident response and reporting procedure in place, and is it tested regularly?
   d. Are due diligence security reviews done to ensure that systems and sensitive information are protected while under the control of other organizations (third parties)?
   e. Are external due diligence security reviews performed on at least an annual basis to ensure that systems and sensitive information are protected while under the control of other organizations (third parties), such as an SAS 70 Type II and SSAE16 SOC 1 and SOC 2 Type I and Type II Reports?
3. Organizational Structure
   a. Does your organization have a formally designated security department/program with specific missions, tasks, and functions?
   b. What is the security group's reporting relationship with senior management?

4. Personnel Security
   a. Are (criminal, prior employment, identity) checks performed on all employees that have access to sensitive information?
   b. Are there established procedures to rescind personnel access in a timely manner when access is no longer required?
   c. Is access to company information restricted and given to only those whose job functions require it? State which types of individuals have access to company data.
   d. How many individuals have access to company data?
5. Environmental Security
   a. What type of access devices are used to control the entrance to the facility? Is the device tied back to an individual?
   b. Is there 24 h onsite security or monitoring where your data center is located?
   c. Is all off-hour access logged, reviewed, and maintained?
   d. Are locking mechanisms employed on doors and windows that control access to the physical business perimeter?
   e. Do employees, contractors, vendors, and visitors carry the appropriate level of identification?
   f. Are alerts generated when physical security has been breached?
   g. Is the management of your information processing facilities outsourced to a third party?
   h. How is physical access to data centers and server rooms restricted?
   i. Are all personnel who service and maintain computer systems and support equipment properly cleared and monitored by trained, qualified escorts?
   j. Do you use cameras within the building and in the data center to detect unauthorized access?
   k. Does your data center have protections in place that minimize environmental issues such as temperature, fire, smoke, water, dust, electrical supply interference, and electromagnetic radiation?
   l. Are communication equipment located in a secure, clean, dust-free, and adequately cooled location?
   m. Are there smoke detectors under the raised floors?
   n. Is there a water detection system under the raised floor?
   o. Is there a chemical fire suppression system (Halon 1301, FM 200) installed in the data center?
   p. Is the building protected (100%) by an automatic sprinkler system?
   q. Is there a fire alarm system (with both audible and visual signals) in this facility?
6. Operations Management
   a. Are changes and new releases always tested using procedures approved by IT and internal audit functions before being placed into operational service?
   b. Do you have an incident management policy?
   c. Are incident logs maintained and reviewed?
   d. Is all off-the-shelf and downloaded software checked for viruses and determined safe for use before installation on stand-alone computers or workstations connected to the network?
   e. Are all systems that contain company information backed up on a regular basis?
   f. Are tapes tested to make sure they are complete and can be used to restore data? How often is this done?

g. Are backups of company data stored at an alternate secure site? Where are the backup tapes/media stored?
h. Who has access to offsite backup tapes?

7. Network Management
   a. Have you implemented a network firewall that separates the internal network from the DMZ?
   b. Does your firewall architecture possess the following?
      i. Stateful inspection of all incoming and outgoing packets
      ii. Mechanisms to hide internal IP addresses from the Internet (IP readdressing and IP masquerading)
      iii. Logging enabled and reviewed on a timely basis including but not limited to stateful inspection logs
      iv. ICMP disabled and blocked
   c. A rule set based on the following:
      i. Antispoofing filters are enabled (RFC-1918)
      ii. Permit rules (allow HTTP/HTTPS) to specified web servers only
      iii. Management permit rules allow SNMP traps to the management of server only
      iv. Noise drops (discard OSPF/HSRP)
      v. Deny all and alert
      vi. Deny and log
      vii. Blockage of loose source and strict source routing
   d. DNS replies to port 53 (UDP/TCP) are limited to authorized internal/external trusted sources, otherwise they are denied
   e. Egress filtering rule is enabled
   f. Stealth rule is enabled
      i. Is the management of your firewall services outsourced to a third party?
      ii. Do you perform network penetration tests on a regular basis? Define your testing intervals.
      iii. Are VLANs or DMZ architecture being used to segregate company information from other areas of the organization?
      iv. Is wireless technology used to allow users access to the internal network? If you answer yes, please define what controls are in place to prevent "snooping" of the network?

8. Information Handling
   a. Does your company have an asset classification policy?
   b. Has a clean desk policy been established at the organization?
   c. Do you have procedures for the handling and storage of information based on its classification?
   d. Is there a formal, written procedure for the accountability, control, and release of computer, network, and media equipment from the facility?
   e. Is sensitive information transmitted electronically over public communications media appropriately encrypted and properly authenticated by the recipient? Please list the encryption tools used.
   f. Do you have procedures for sanitizing electronic media for reuse?
   g. How are confidential hard-copy and soft-copy media disposed?
   h. In response to telephone inquiries, do employees furnish nonconfidential information freely but restrict confidential information to authorized inquiries only?

9. Access Control
    a. Do you have authentication mechanisms such as LDAP, two factor, or certificate-based in place to control access to network and application resources?
    b. Does each individual have a unique and identifiable login ID to access application and network resources?
    c. Do the IDs have the following characteristics?
        i. Locked after a given number of unsuccessful tries and can only be unlocked by an authorized administrator
        ii. Disabled after a period of inactivity
        iii. Not displayed on terminals or monitors
    d. Do you have the following password controls in place:
        i. Defined length (please describe)
        ii. Not displayed on screens or in reports
        iii. History is maintained
        iv. Composition rules are in effect to enforce strong passwords
    e. Please describe in detail how your organization handles authorizing emergency administrator access (i.e., sealed in envelope process, electronic process to change designees, etc.).
    f. Are there any generic or shared user IDs that are configured on a system that contains company information?
    g. Are system administrator commands traceable back to an individual user ID?
    h. Have all vendor-supplied default passwords been changed?
    i. Does the system protect passwords and security tables by encryption or some other secure means?
10. Is remote access to devices achieved via a secure mechanism such as SSH protocol (assigned to a port other than 22) or VPN?
11. Are all workstations protected by password-protected screen savers that automatically activate after a period of inactivity?
12. Do all servers and workstations utilize an antivirus solution? How often are the signatures updated?
13. Compliance
    a. What security certifications has your company earned (i.e., ISO27000, 177799, 7799)?
    b. Is your company in compliance with pertinent global information protection legislation (i.e., Safe Harbor, GLB, SOX, HIPAA, etc.)?
    c. Is a formal investigation conducted in accordance with appropriate regulations and/or policy whenever a compromise or suspected compromise is the result of a security incident?
    d. Do you have controls in place that monitor activity on all configurable systems and devices that store or process company information?
    e. Do you perform self-based security assessments? Please define frequency and type of assessments.
    f. Are external due diligence security or independent third-party reviews performed on at least an annual basis to ensure that systems and sensitive information are protected while under the control of other organizations (third parties), such as an SAS 70 Type II and SSAE16 SOC 1 and SOC 2 Type I and Type II Reports? Please define
        i. Basis and type of audit
        ii. Type of report generated
        iii. Remediation of findings

14. Business Continuity and Disaster Recovery Planning—Administration
    a. Is Business Continuity Planning (BCP) conducted on an enterprise-wide basis?
    b. Does the company have a documented BCP that is site specific to the company? If no, is the company willing to prepare a company-specific plan?
    c. Are BCP responsibilities assigned to the appropriate levels of management, specific personnel, and teams?
    d. Are there procedures in place to ensure that all plans are periodically updated to reflect and respond to changes within the company? Does the plan require that employees read and acknowledge the plan?
    e. Are third parties utilized to support the company and, specifically, the company in the performance of disaster recovery?
    f. Have the plan and test results been subjected to an independent audit (SAS 70, ISO/IEC 17799:2000)? Add detail on who audited, when the audit occurred, and what the results of the audit were.
15. Business Continuity and Disaster Recovery Planning—Facilities
    a. Does the company subscribe to multiple, diverse carriers for local and long-distance service providers?
    b. Is the data center safely located within the building (e.g., data center has no exterior walls or is not located near a dock area)?
    c. Are there multiple and physically separate connections from the public power grid substations?
    d. Are there high target buildings, structures, or tenants at this location?
    e. Does the physical cable (copper or fiber) that is provided by the telecom companies enter the building through multiple locations on the building perimeter and terminate into separate distribution rooms?
16. Business Continuity and Disaster Recovery Planning—Components
    a. Does the plan address specific expected recovery time objectives (RTOs) for company services? If no, are they amenable to customizing the plan for company?
17. Business Continuity and Disaster Recovery Planning—Alternate Sites
    a. Does the primary data center have an alternate site for data center recovery?
    b. Is the alternate site internal or external (e.g., IBM, SunGard)?
    c. Are the alternate sites "hot" or "cold" sites? For any cold sites, include projections on how fast the company can resume full operations following an emergency.
    d. When was the last fail over test? (Include (1) scope and objectives of the test and (2) the results and whether the gaps were corrected.)
    e. Are BCP and recovery procedure manuals maintained at the primary, alternate, and off-site storage locations? How are they stored (e.g., electronically, paper)?
    f. Does the plan include primary and alternate workgroup recovery procedures/sites for business processes/services?
18. Business Continuity and Disaster Recovery Planning—Testing
    a. Are annual Operational Exercise/Tests or Conference room simulations done?
    b. When was the last business/disaster recovery test performed? (Include (1) scope and objectives of the test and (2) the results and whether the gaps were corrected.)

Chapter 4

Getting the Best Out of Information Security Projects

Todd Fitzgerald

Contents

"Are we having fun yet?"

Bill Griffith 1944–
(Zippy, the Pinhead comic strip, 1979)

If the average person had to answer the question, "what do information security professionals do?," the response is likely to be along the lines of "they set up accounts and passwords," "issue access badges," "install antivirus stuff on my computer," "find the bad guys," or simply, "something with computers and tell me what I can't get to."

There is a commonality among these responses—they are all tactical activities and very operational in nature. In other words, security professionals exist to the general population to respond to requests for access and help when they cannot get to something. Those of us that have worked in the profession to protect our organization's assets know that there is much more that must be done in the context of an information security program. Just by reviewing the 11 major areas of ISO/IEC 27002 (ISO, 2005), it is clear that an effective security management system should address security policy, an information security organization, asset management, human resources security, physical and environmental security, communications and operations management, access control information security acquisition development and maintenance, information security incident management, business continuity management, and compliance processes. Beyond these areas, ISO/IEC 27002 consists of 133 controls which must be considered to develop the appropriate security management system and supporting architecture. Now this sounds much more extensive than the establishing of an account, provisioning access to a particular resource or resetting a password!

So why is it important to make the distinction between a security "operation" and the requirements for building an effective information security program? The reason is that we cannot build an effective security program with the same methods or mentality that is used to respond to a ticket to establish a user account. These requests follow a predetermined operational process such as user submits a ticket, information security reviews the ticket to ensure the appropriate approvals are included, information security sets up an account with appropriate access according to the application and user profile, the user is notified of the access, and the ticket is closed. This scenario is played out many times per day, for many users, and can be measured as to the efficiency and effectiveness of the process. Building an information security program, on the other hand, requires that each initiative is managed using techniques to ensure that the effort is kept on budget, on schedule, and is of high quality (delivers the expected results to the business). These techniques are well-developed and known as the field of "project management."

Effective project management can increase the likelihood that the security initiative that is being implemented will deliver the results that are expected. If these projects were managed according to operational constructs, such as that noted in the prior example, the project would be managed through performing the tasks in a disorganized manner, whereby the tasks performed today would not necessarily be the right tasks, in the right order, or by the right individuals. Imagine building a house and on the first day the carpenter, cement pourer, electrician, and bricklayer all showed up, but what was really needed was the surveyor before any of the other individuals were needed! Security projects also need the discipline that the field of project management can deliver.

Project management has been around much longer than the Information Security field, and includes the construction of such projects as the Giza Pyramid, the Parthenon, Colosseum, the Taj Mahal, and other great buildings which required that they be delivered in a systematic way (Lessons From History, 2012). Modern project management emerged in the 1950s after WW2, as there was a need to organize large quantities of resources and personnel to reach critical goals in short period. Several projects such as the Golden Gate Bridge and the building of the Hoover Dam were created to address the large unemployment situation and required project management to complete efforts on this large scale. Of more recent times, as many as 1000 construction workers worked every day for several years to build the Milwaukee Brewer baseball stadium with a 92-ton

covered retractable roof. Imagine 1000 people working on an effort without well thought-out plans of what the deliverable would look like.

No Need to Reinvent the Wheel

Fortunately, the project management discipline has evolved to where there are practices which have been tested over time, and if followed correctly, should result in more satisfying projects that meet the right expectations. The security profession itself essentially exists to limit the risk of disclosure, destruction, or loss of information assets that are entrusted to the organization. Projects also carry risks which can be mitigated, and just as security professionals limited the risk through the implementation of policies, processes, and supporting technologies, project risk can be mitigated through the use of established project management practices. This is analogous to the security professional keeping the backup tape on top of the server being protected or taping the PIN number to the back of a security token. There are established methods which should be leveraged.

One of the most notable organizations that has promoted a standardized project management discipline is the Project Management Institute (PMI). PMI has published the Project Management Body of Knowledge (PMBOK®) which represents "good practice," meaning that the knowledge described are applicable to most projects most of the time, and there is widespread consensus about their value and usefulness (PMI, 2012). PMI has established the gold standard Project Management Professional (PMP) certification to recognize individuals that have a working knowledge of the PMBOK (PMI, 2008) and have experience in leading projects. Any information security professional that is leading projects to implement new security deliverables would be well advised to seek out this certification or ensure that large projects are leveraging the resources of the Project Management Organization (PMO) to complete the projects.

The subsequent sections cannot do adequate justice to the detailed PMBOK, which is almost 500 pages of very rich information. Even then, each of the subject areas within the PMBOK can be elaborated further in books in their own right. However, key considerations as applied to managing an information security project are highlighted here. Every project will be different and have different challenges, but what project managers bring to the table is a consistency in approach and a method to uncover the appropriate issues at the right time, before they become issues that are too large for the project to overcome to make its deliverable date. For example, if individuals assume that the order was placed for an Internet connection 2 weeks prior to the launch of a web application, yet the individual assumed that someone else was placing the order, it may be too late to make the deliverable date if there was a 30-day lead time required. In this situation, a diligent project manager would be aware of the tasks required and who was responsible for implementing at the correct time and followed up to ensure that the tasks defined were completed. Project management does not leave activity completion to chance and operates through a series of tools, software, soft skills, templates, and methods to manage the project to completion.

What Is a Project?

PMI defines a project as "a temporary endeavor to create a unique product, service, or result" (PMI, 2008, p. 5). Being that this is temporary in nature, this infers that there is a discrete beginning and a discrete end, an end that comes once the project's objectives are met. This means that objectives need to be defined, otherwise, how will it be known when the end is reached? As the old

saying goes, "If you don't know where you are going, all roads will get you there." In the information technology field, it seems common to see a project morph into ongoing maintenance, which really should be considered as future enhancements or ongoing support. Projects on the other hand are typically driven by a date set by the business to launch a new product or service, or to reduce costs of operating a particular process. Security projects produce a unique policy, process, or implementation of a technology to permit enhanced protection of the information assets.

Project Planning

The first step in the development of a good security project is to develop a project charter. The charter documents the business needs of the organization and why the security project will meet those needs. The charter also provides a vehicle by which to authorize the project and provide visibility to the work that the information security department is doing on behalf of the business. For example, if the security department wants to install appliances to monitor the vulnerabilities that exist on the servers and workstations, the project charter would define what the business reasons for the project are. These reasons should not be defined in terms of technical reasons, but rather on what is the benefit to the business. The reasons could be saving the company time and money that would be incurred by individually scanning using other tools, reducing the number of vulnerabilities in the environment, addressing zero-day threats more efficiently and effectively, identifying rogue software, and so on. The best reasons would be those that are nonsecurity department or nontechnical explanations which support business objectives. This is sometimes hard to do, but the more that the security project is tied to a key business objective, the more likely that funding will be allocated and supported until project completion. For example, the protection of the servers that host the organization's key 24×7 application from being attacked by a known exploit, thus increasing the availability of the application would be a stronger business case than the security department wanting to have visibility into the vulnerabilities.

Develop Scope Statement

Once the protect charter has been developed from reviewing existing contracts, vendor Statement of Works (SOWs), and existing processes, a scope statement is defined, which reflects the deliverable requirements, project boundaries, and generally how the scope will be managed throughout the project. Scope Creep, or the adding of functionality without regard to the impact on resources and costs of the project, can easily set in if the scope is not defined. This does not mean that the project cannot be flexible to maintain changes to the project (through predefined change management), as resources and costs can be adjusted throughout the project with proper approval. When projects are extended beyond their timeframe, resources that were needed for other initiatives are now suddenly unavailable and the impact of the unavailability of these resources needs to be consciously reacted to. In a perfect world, all security projects would start and end on time with exactly the budget decided upon. The world is not perfect, and Mother Nature constantly changes our plans with disasters and weather changes—same is true with projects. However, we do not wake up each morning without an idea of what we are going to do for the day—we have a scope statement, even if it is not written down—we go to work, will work 8–10 hours, make some work decisions, complete some projects, start other projects, drive home, make dinner, play with the kids, go to their activities, watch the news, and call it a night. In other words, we approach our day with some objective and various steps (tasks) by which we will meet the objective. This may

change depending upon the day and the objective for the day, but we typically start the day with an objective.

The security project scope statement scope definition subsequently becomes the basis for future project decisions. As items are introduced that are "out of scope," these can be added. If time is of essence in a project, such as a new mobile application depends upon having the security coding practices for mobile applications defined before the product can complete development, then it becomes more important to hold the line on permitting new functionality to be added to the security project. The new functionality could delay the effort beyond the time desired to launch the product. Security professionals also tend to be on the conservative side of risk, and are more likely to want to extend the project to get close to 100% security in data at rest, in transit, and ensure it is vulnerability-proof. While this approach greatly enhances the security posture, it could also be damaging to the security organization's reputation and contribute to being viewed as slowing down the process. One way to mitigate this perception is define project scopes that are truly manageable within a reasonable timeframe, and work to build the best security possible within those constraints. Additional projects or enhancements can be added and scheduled once the completion of the original project scope is achieved.

Develop Work Breakdown Structures

Work Breakdown Structures (WBS) break down the major security project deliverables into more manageable components. The average human brain can only keep track and manage seven things at one time before becoming overloaded. Concepts are much easier to understand when the lower level details are summarized into higher level categories. Take seven steps to be a good car salesman, (1) meet and greet, (2) qualification, (3) trade-in appraisal, (4) go get "Managers Offer," (5) test drive, (6) close, and (7) stay in contact. Even if we are not car salesmen, the steps can be understood at this level. However, there are many detailed techniques and tasks, from writing up offers, taking the customer for a test drive while waiting for the manager's offer (even though they know right away what they will offer), to tricks of downgrading the trade-in and building up the new car in the customer's eyes. Categorizing the security tasks, also known as "chunking" into higher levels, can help create the right focus. For example, developing a set of ISO27000 policies may involve areas such as reviewing existing policies, examining laws and regulations, researching current industry trends, examining internal infrastructure, developing policies, and rolling out and training the users on the new policies. Each of these areas would have multiple activities or tasks associated with them.

Define Activities

Activities are defined by subdividing the work pages in the WBS (i.e., project deliverables) into smaller activities through decomposition. Once a project has been created, these lists of activities can then be used as templates for future projects. For example, the implementation of cloud-based Software as a Service (SAAS) with the requisite security considerations can be used as a template, or model containing the appropriate activities for a future, similar project. Over time, a library of security project templates can speed up the project and help realize the benefits of having a consistent security project process. This also reduces the project risk that the project undertakes. Activities may be defined through the use of Rolling Wave Planning, which attributes more detail to the near-term tasks, and leave the other, future tasks at a higher level. As the project moves toward completion, the activities are elaborated in more detail. Sometimes, this is necessary as

there are many unknowns at the beginning of a project, as well as new discoveries along the way that may not have been anticipated at the beginning of the effort.

Sequence Activities

Once the activities are known, then the activities need to be sequenced with the proper precedence relationships. Some items can be done concurrently, such as the configuration of a firewall while the incident processes are being developed, however, some items have a sequence that must be followed in order. If performing a disaster recovery test, the operation system would need to be installed before the databases, middleware, and applications, and after the virtualized servers were made available. Project management software packages typically use the Precedence Diagramming Method (PDM) of developing the activity dependencies and use combinations of four different types of dependencies:

- Finish-to-Start—Before the successor activity can start, the first (predecessor) activity must finish.
- Finish-to-Finish—Before the successor activity can finish, the processor activity must finish.
- Start-to-Start—The initiation of the successor activity depends upon the initiation of the predecessor activity.
- Start-to-Finish—The completion of the successor activity depends upon the initiation of the predecessor activity.

The resulting diagram represents the relationships between the different activities. Lead and lag times are applied, whereby when an activity has lead time, there is the ability to accelerate the project completion. For example, an organization writing an information security policy document could have the social media policy written with a lead time of 20 days, as reflected by a Finish–Start dependency with a 20-day lead time. If this was finished prior, this could accelerate the start of the entire policy. A lag reflects a delay in the subsequent successor activity. A lag of 5 days would indicate that the successor cannot start until 5 days after the predecessor has completed. This could represent the implementation of a processor burn-in period to ensure that the processor was operational and stable before deploying applications on the equipment.

Determine Resourcing

Now that the activities have been defined and sequenced, the activities must be resourced with the appropriate headcount and skills needed to complete the task. Depending upon the skill level of the security analyst, Information Technology individual, business partner, and so on available for the project, this could impact the time required for the task. Many times security consultants are used to supplement security projects, not only to increase the amount of available staff, but also to shorten the deliverable time of the project due to their security expertise. While the rates of the consultants will most likely be higher than the rates of in-house employees, use of external resources is advisable when there are expertise gaps or new technology is being introduced. Vendor (product)-specific resources tend to command an even higher rate than specialized security consulting or the Big Four, however, these resources are usually only engaged for minimal periods, particularly during security product installations. Due to the expense of these resources, it is advisable to ensure that the project plan has accurately identified the correct activities and their timing (sequencing), so that the resources are engaged for limited times. For example, implementing a

Security Information Event Management (SIEM) tool may require 5 days of upfront installation and tuning expertise to implement basic functionality of the product, with future budget allowances for tuning efforts at that juncture of the project or once the product is being supported by the organization and the implementation project has been completed.

Estimate Duration

The duration of the activities can now be estimated, using such techniques as Analogous Estimating (using a prior, similar activity of similar scope to estimate the project), Expert Judgment, often guided by historical information, Parametric estimating (using a metric to estimate activity duration, such as the number of firewall rules to be analyzed × time per rule), and 3-point estimates (most likely, pessimistic, and optimistic) which take an average of the three different estimates. Building reserve time into the estimates is known as Reserve Analysis, whereby a buffer may be needed. This is particularly used when embarking on an unknown technology, or adding a new environment where the actual numbers of file permissions that must be reviewed to bring the new environment into compliance are unknown.

Determine Project Schedule

Schedules are then developed and by using the early start and finish dates and the late start and finish dates, the flexibility of the schedule is calculated. If there is a positive amount of time between the early and late dates, then there is free/slack time referred to as the "float." These tasks have more time to slip and still keep the project on schedule. The "Critical Path" is that part of the precedence diagram which contains zero or negative float, and as such, will impact the project schedule if these days slip. Several techniques to minimize the schedule time can be tried, such as crashing the schedule, whereby different tradeoffs of cost and schedule are attempted to shorten the project schedule (possibly using more resources at an increased cost) while keeping the same scope. The schedule may also be Fast Tracked by performing several activities in parallel. For example, this may involve ordering the hardware security tokens and issuing them to the end users for use at future date while setting up the database information for launch. Resource leveling is a technique to keep the resources at a constant rate during the project, which may extend the project time, but reflect the reality of the people necessary for the implementation. A new tape encryption policy, process, and technology implementation may require the infrastructure team at the same time as a major application is being deployed. Rather than have this team allocated 80 h a week for 3 weeks, the resource leveling may reduce the workload to 20 h a week for 12 weeks to permit the other, greater priority implementation.

Cost estimate accuracy increases throughout the project. This is because as the project continues, more of the unknowns and their impacts are answered. The greater the ambiguity, typically at the beginning of the effort, the greater the margin for error. Projects in the initiation stage are usually estimated by a rough order of magnitude of −50% to +100%, and later estimates in the project are narrowed to the −10% to +15% range (PMI, 2008). Cost and schedule variances need to be managed throughout the project lifecycle. Cost variance is the earned value (EV), or the amount budgeted for the scheduled activity completed less the actual cost (AC). In other words, has the project returned the expected (budgeted) value for the money spent? Schedule variance communicates the difference between what has been earned thus far versus what has been planned. Indexes can also be used to show a quick indicator as to how far off the cost estimates are versus actual expenditures.

Develop RACI Chart

Human resources need to be planned to ensure that the right individuals are onboard and it is clear who is accountable and responsible. Charts such as the RACI chart, named as the chart indicates the activity and who is Responsible®, Accountable (A), should be Consulted (C), and who needs to be Informed (I). Typically, only one person should be held accountable for a particular activity, whereas there may be multiple individuals responsible.

Project Execution

Project execution involves assembling the appropriate project team to accomplish the tasks and coordinating those resources to complete the activities identified in the planning stage. While the project execution stage typically has the longer duration, the activities to properly plan the project should not be short changed, as failure to properly plan usually leads to surprises, delays, and implementation of a lower quality. Project management performed by a project manager is a skill that could be honorably viewed as a profession similar to that of a policeman directing traffic at a busy intersection during a stoplight outage. Cars are approaching from many directions and need the guidance of the traffic cop to determine when to go and when to stop. If the traffic cop lets traffic from one direction never have the opportunity to proceed, the drivers will get very frustrated while they wait. If the policeman lets all the cars, or activities, advance at the same time, there are bound to be constraints and collisions in the intersection, which also slows down progress. The policeman also may be receiving information from other policemen at different checkpoints to understand how those cars or activities are progressing so that he/she can adjust his own flow through the intersection. In other words, the project manager, like the traffic cop/policeman, must make many decision based upon what is occurring in the environment. This is important, because project management is much more than creating a project plan, Gantt chart, deliverables, or project milestones and then forgetting them. A project manager must be thoroughly engaged throughout the project to ensure that tasks are being completed and the plan is appropriately modified.

The project manager may be someone within the Information Security department or part of the PMO as previously mentioned. What is important is not where the project manager resides, but rather that the person running the project is taking on the role and discipline of the project manager. Are the activities well defined? Have the activity sequences (precedence diagrams) been created with realistic time frames and are they agreed to? Does the project have the right skill sets available at the right time? Note that none of these questions are technical security questions, but rather are project management questions to keep the project on track.

Where the Rubber Meets the Road

A successful project will spend a good deal of time in upfront planning, and with the proper resources, the ability to execute will increase. Gartner classified product vendors into four quadrants, ranging from a niche player to one that has the vision and leadership with the "ability to execute." Our projects should be viewed the same way—maybe we have the appropriate vision, but not the right resources to execute. Or, maybe we have the right technical resources, but our vision of the project and project management practices are not up to snuff. Project execution is where we bring the planning together and create a new service, increase the protection of the information

assets, or implement a new process to satisfy the objective stated in the original project charter. Projects fail more often not because of the technology product selected, but rather from inadequate project management or not being aligned with the business objectives.

Project Monitoring and Controlling

Project monitoring occurs thought the project and is not a phase in the project. Project changes need to be controlled and authorized to ensure that the project timeline and functionality are not adversely impacted. Corrective actions can be taken only if issues are raised and escalated in a timely manner. The RACI chart identified earlier can be referenced to determine where the issues need to be raised and who has the decision-making authority. Monitoring needs to ensure that the scope, schedule, costs, quality, risk, and human resources are appropriately monitored and adjustments are made where necessary to enhance the likelihood of project success.

Regular Updates

Status meetings should be held at a minimum on a weekly basis. Sometimes very technically oriented resources may regard these as a waste of time, or "another meeting to attend," whereas the regularly scheduled meetings have a way of making everyone more accountable. The status meetings bring visibility to the issue areas and those activities that are in trouble that need additional focus. No one likes to show up for a status meeting unprepared, so these meetings have a catalyst effect of keeping the project moving in the right direction. These meetings are not the place to perform a deep dive into a particular technical security discussion; however, they should be used to identify additional meetings that need to be scheduled. The status meetings also permit the collaboration necessary between departments, which may/may not occur naturally with the formal meetings to serve as the impetus.

Matrices containing responsibilities, dates, and tasks should be color-coded with red (behind schedule/in danger of meeting schedule), yellow (potentially in danger), green (on schedule) to provide a quick read as to the status. These status need to be honest assessments, as it does the project no good to highlight an activity as green for 10 weeks, and then in the last week turn it to red. Rarely are events not known earlier in the effort and this type of scenario typically indicates that the activity was not being managed or actively worked on until just prior to the deliverable date.

Project quality is monitored through the use of quality tools such as cause and effect diagrams, control charts, flowcharting, histograms, Pareto charts, run charts, scatter diagrams, statistical sampling, and defect repair reviews. Quality audits of the project review the scope change requests, corrective actions, actions to mitigate project risk, and so on.

Communications

Project communications take on many forms, as there are many stakeholders. There are the executive sponsors, which need a high-level understanding of whether or not the project will meet the agreed upon timeframes. Issues of cost overruns may surface and need approval for additional funding to proceed. The parties from the other departments, such as the business units being supported, network, infrastructure, database administration, systems development, physical security, facilities, and so forth depending upon the project, need to be informed of the project status. There may also be a core team that needs to have information on a more frequent basis. If this is a large

project impacting many other users, a weekly communication via email may be necessary. Finally, depending upon the complexity and scope of the project, a steering committee may be necessary for the rollout.

The old adage that "you can't ever over communicate" applies here. Communication must be two way, and during these meetings, it is beneficial to ask the difficult questions, "Does anyone have any concerns with this effort?" Each organization can identify those that are very vocal in their positions, which tend to be the more extraverted individuals. Key information that could impact the project may be known by those that are more reserved, or introverted, and need to be drawn out during the meetings. Therefore, it is more productive to "draw out" those responses from those individuals.

Project Closing

Once the new security controls are implemented, we are done! Not so fast. While it may appear desirable to move the resources to the next big security initiative, care should be taken to properly close the project. Were all the proper signoffs obtained? What were the final costs? How did the actual schedule/budget correspond to what was planned? What items were surprises during the project? What were the lessons learned that we should not repeat on the next project? Were other enhancements identified during this project? There are many questions which should be answered and appropriately documented.

To really leverage the work from the project on the next project, the project team should have maintained a project repository that can be leveraged for subsequent projects. Many times a new template may be developed, such as one for recording minutes, or capturing the technical environment guests/hosts/servers for a cloud computing endeavor. These templates can be standardized and made available to future projects. Failing to do so would be a missed opportunity to increase the effectiveness.

As busy as teams are, key resources will be pulled into other initiatives. Recognizing these individuals is a step that is very important, not just to thank them for the work done on this project, but they may also be supportive of the next security initiative if they feel that they had some ownership in the deliverable. Resources that work on security projects extend well beyond the security team, so they need to have a sense of "what is in it for me" to be re-engaged for the next effort. They need to see some value for themselves and their department in working in the effort to become fully committed in the next project. This recognition will be increasingly more important, as Generation Y becomes a larger part of the workforce, where project experiences tend to be more important than long-term job security. The security teams also need the involvement of this generally technically savvy demographic.

A project will always have events that are unexpected during the project—the scope may increase because of a new time-to-market marketing initiative that needs a hardened server implemented in less time than originally planned, a key resource suddenly leaves the company, or someone forgot to order the leased line which has a 90-day turnaround and it is needed in 30 days. Each of these items can increase the project risk, risk that the project will not be completed on time, within budget, or with the quality expected. These risks must be managed and tracked, so that accountability is maintained and new alternatives and corresponding dates are managed. This is analogous to tracking the security vulnerabilities across the environment and ensuring that the appropriate departments are reducing the risk by implementing mitigating controls in a timely manner.

Final Thoughts

The information security field has evolved greatly in the last 10–15 years, and then again, one could argue that it is still at the same place it was. As laws and regulations have increased with the introduction of the Health Insurance Portability and Accountability Act (HIPAA) Final Security Rule (2003), the Federal Information Security Management Act (FISMA, 2002), Payment Card Industry Data Standard, Gramm–Leach–Bliley Act (GLBA), Health Information Technology for Economic and Clinical Health Act (2009), the National Institute of Standards and Technology (NIST) publications, and others, there has been an increased focused on security controls. There has also been an increased focus on the soft skills of the information security officer to build effective teams, influence decision making, market their services, build relationship, and so forth. What has not received as much focus has been the focus on effective project management for information security projects. While this is emerging, as indicated adding Program Management as the 18th family of the NIST 800-53 control series (NIST, 2009), there needs to be a greater recognition that just as security operations control activities need resources capable of interpreting the security events that are coming across the wire, security projects need to have allocated project managers to achieve effective results. This may be a project manager from the PMO, or taken on as a role by a security professional. In either case, it is important that the individual has the appropriate training and experience needed, as defined by the criticality of the project. With these skills in place, the information security team can achieve the projects to carry out the vision and strategy. Otherwise, the vision will remain just that, a vision.

References

ISO/IEC 17799:2005 *Information Technology Security Techniques—Code of Practice for Information Security Management*. International Organization for Standardization (ISO), http://www.iso.org/iso/en/prods-services/popstds/informationsecurity.html

Lessons from History. 2012. *The History of Project Management*. http://lessons-from-history.com/history-project-management-page

National Institute of Standards and Technology (NIST). August 2009. *Special Publication 800-53 Rev3: Recommended Security Controls for Federal Information Systems and Organizations*. http://csrc.nist.gov/publications/nistpubs/800-53-Rev3/sp800-53-rev3-final_updated-errata_05-01-2010.pdf

PMI. *A Guide to the Project Management Body of Knowledge (PMBOK Guide)*—Fourth Edition. Project Management Institute, Newtown Square, PA, 2008.

PMI. Project Management Institute, 2012. http://pmi.org

Chapter 5

Mobility and Its Impact on Enterprise Security

Prashanth Venkatesh and Balaji Raghunathan

Contents

As part of its report titled "User Survey Analysis: Impact of Mobile Devices on Network and Data Center Infrastructure," Gartner, who surveyed respondents from enterprises with 500 or more employees and an in-house data center in the United States, the United Kingdom, Germany,

Australia, Brazil, Russia, India, China, and Japan, in October–November 2011, found that 90% of these enterprises have deployed mobile devices, with smartphones being most widely deployed and 86% of these enterprises planned to deploy media tablets this year.[1]

"Consumerization of IT," which is all about how consumer technology, which includes mobile devices such as phones and tablets as well as PCs, is rapidly proliferating into the enterprise has changed the traditional IT environment in a big way.[2]

This chapter describes the challenges brought into the realm of enterprise security by the rapid adoption of mobile technologies and how leading enterprises are addressing these challenges.

Drivers for Adoption of Mobile Technologies in the Enterprise

A decade ago, mobility for an employee meant having a company laptop and having access to the required data that would enable them to work offline. This evolved into employees getting access to corporate network through virtual private network (VPN) or remote access tools. Today, the world has moved on from PCs and laptops to mobiles, tablets, and PDAs.

Mobile devices play a key role in rapid and ubiquitous access to data within the enterprise. This can be leveraged by the enterprise to enhance customer experience, and improve partner and employee productivity and engagement.

The rapid adoption of mobile devices such as smartphones and tablets has also meant a paradigm shift in the way the employees, partners, and customers of the enterprise expect to be engaged.

To enhance customer experience and retain customer loyalty, enterprises develop mobile applications for their customers and allow them to download them on to their device. These mobile applications may provide access to data that lie within the corporate network and the applications may also allow the data to be stored on their personal devices.

Partners also expect the enterprise to allow access to data from their mobile devices.

Given that mobile devices enable rapid access to corporate data, employees too expect "productivity-enhancement" applications and tools to be made available through mobile devices. Enterprises are continuously seeking newer ways to enable their employees to access information ubiquitously.

The various business benefits that a mobile work force can bring in to the organization have resulted in the enterprises equipping their employees with mobile devices (or allowing these employees to get in their own mobile devices) and making corporate applications and data accessible from these devices.

Enterprise Mobility Ecosystem

Figure 5.1 provides an overview of how a mobile device can access enterprise data and the ecosystem around this.

Thus, the basic mobility ecosystem involves

- Mobile applications that provide access to enterprise data
- Devices and operating systems (OS) to consume the enterprise data
- Protocols to share this data with other devices
- Telco carrier (data) network to transmit this data
- Enterprise applications/services and data stores

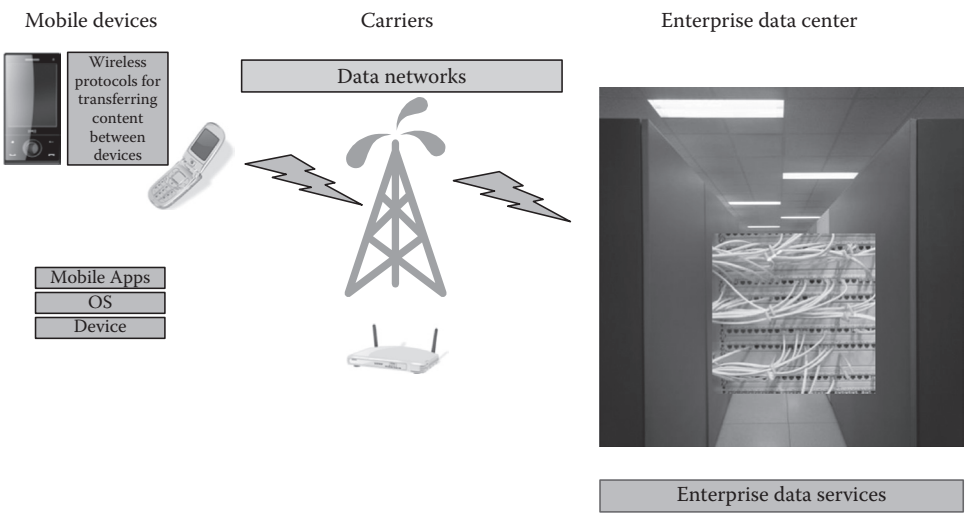


Figure 5.1 Ecosystem for a mobile device to access enterprise data.

In addition to these, appstores (application stores), from where mobile applications are typically downloaded, can also be considered as a component of the mobile ecosystem.

The integration of this ecosystem into the enterprise IT infrastructure makes it imperative for the enterprise IT to procure additional systems, and have additional policies, processes, and tools to manage, control, and secure the enterprise assets and communication points.

Key Challenges in Managing, Controlling, and Securing Access to Enterprise Data from Mobile Devices

Device and Technology Diversity and Heterogeneity

Mobile technology is still evolving at a rapid pace. The ecosystem is filled with heterogeneous possibilities and diversity and a wide range of options for the end consumer. Mobile devices can be a tablet or a smartphone or low-end cell phone. The handset can be from different vendors (Apple, RIM, HTC, Samsung, Nokia, etc.) and each vendor can have multiple models. The OS on these devices can vary (IOS, Android, Windows Phone, etc.). Devices also have a range of protocols to choose from, for any data sharing or file sharing (Wi-Fi, Bluetooth, etc.) and a host of data network types (GPRS, Edge, 3G, 4G) they can connect to. The applications can be either native or hybrid or can be accessed by a browser. The data for these applications can be from any enterprise data store (SAP, Siebel, RDBMS, etc.).

Table 5.1 captures the heterogeneity and diversity of the ecosystem. This heterogeneity results in the following challenges:

- How should different devices, platforms, and OS be supported by the enterprise?
- How should communication between the device and enterprise network be secured?
- How should diverse devices be managed and controlled?
- Can one tool support all these devices, platforms, and OS, or should we go in for multiple tools?

Table 5.1 Mobile Device and Ecosystem Heterogeneity

Ecosystem Component	*Options*
Device	Smartphones, tablets
	Handset choices: iPhone versus Android phones (HTC/Samsung/Nexus) versus Blackberry versus Windows Phone
Device operating system	iOS versus Android versus Blackberry versus Windows Phone
Device protocol	Wi-Fi, Bluetooth
Data network	GPRS, EDGE, 3G, 4G
Mobile application	Native apps versus mobile web versus hybrid apps
Enterprise application/data store	ERP, CRM, portal, database

- How should security patches and policies be applied on the devices?
- How should applications be delivered?
- How should applications be managed?
- How should application development and testing for applications on multiple platforms be supported by the enterprise?
- What is the minimum device configuration that should be supported by the enterprise?
- What is the minimum OS version that must be supported by the enterprise?
- What is the minimum permission that should be granted to applications?

BYOD (Bring Your Own Device)

Initially, enterprises used to provide their own handsets to employees, which were hardened as per company policies. Most companies allowed their employees to check e-mails or store contacts, but did not allow employees to download rich applications. With the advent of powerful devices with rich features, more and more employees prefer to use their own devices for work purpose.

More enterprises permit their employees to bring their own devices mostly as an employee engagement initiative. On the one side, this initiative helps organizations to save capital cost (on procuring devices for their employees), while on the other side, it helps employees to use the device that they are comfortable with.

However, BYOD brings with it a host of security, privacy, and legal concerns.

- What is the level of control that enterprise IT can exert over a personal device?
- What is the level to which the personal device must be managed?
- How should theft of devices be handled?
- Should applications be allowed to store enterprise data on personal devices?
- Which are the handsets (makes, models) and OS versions that should be permitted?
- How many devices can the user connect to the enterprise network?
- How should applications be delivered on to devices?
- Can the user be restricted from accessing specific social networking and other sites?

- What level of support should be provided to personal devices?
- What is the level of restriction that can be placed on the device without annoying the user?
- Given that native apps are much more difficult to be controlled by IT as compared to web apps, can the user be restricted from accessing native apps?

Traditionally, the enterprise IT infrastructure team is used to manage all the OS in an enterprise. With the advent of BYOD, heterogeneous systems are introduced into the network.

Different mobile OS support different ways to manage device and application security. For example, if we need to install an application on Android OS, we would have to give either all the permission to the list or cancel install. This is not the case with Apple iOS-based devices. We can choose not to give permission to a specific service and still install the application.

Also, the levels of security might vary if the same device is jail broken and allows installation of applications from unrecognized application sources.

Additional Security Vulnerabilities to Be Handled

The introduction of mobile access to enterprise data brings into its fold a list of additional vulnerabilities and threats to the enterprise. Figure 5.2 provides a list of additional vulnerabilities and threats the enterprise IT must be prepared to handle.

OS vulnerabilities: Many OS vulnerabilities have led to the compromise of the device. For example, one of the vulnerabilities in Safari web browser allowed access to phone features such as address book. Some of the major vulnerabilities found in iOS are CVE-2012-0674, CVE-2011-3442, and so on.

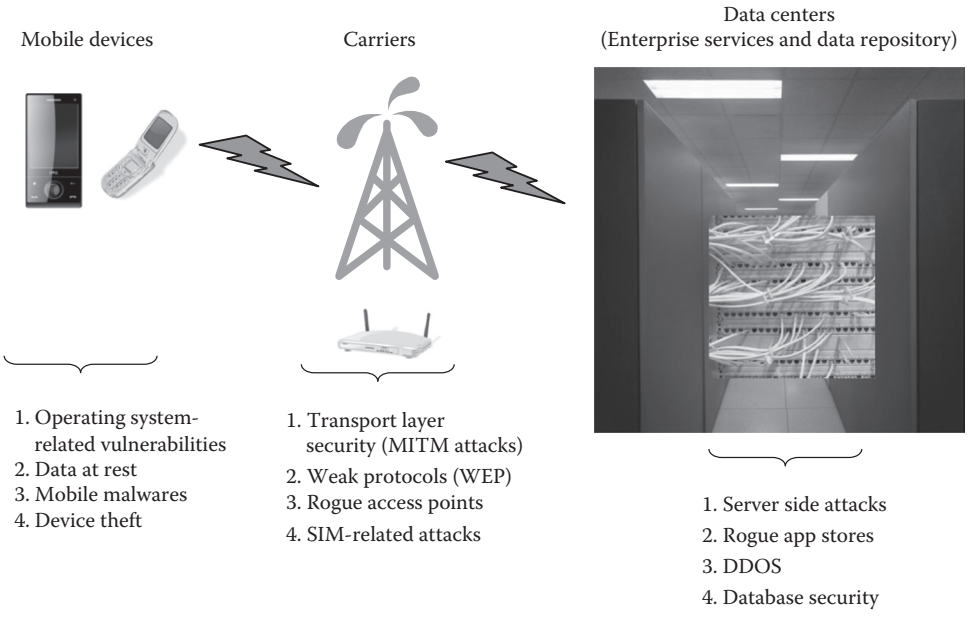


Figure 5.2 Device-level vulnerabilities.

Android OS has been a target for malware writers and hackers for some time now and many enterprises still do not prefer Android for enterprise usage due to reported major vulnerabilities. There are specific websites (http://www.cvedetails.com) that highlight some of the major vulnerabilities found in Android OS. Most vulnerabilities found in Android revolves around access bypass and privilege escalation. Some of the common malwares found on Android are Zitmo Android Edition mobile version of banking malware (Zeus).[3]

Data at rest: Applications might store user data such as passwords to applications such as Facebook locally on mobile phones. In case of a device theft or unauthorized access, it might be possible that someone can steal the data. Data at rest vulnerability has to be addressed by the application.

Mobile malwares: Mobile device users are increasingly vulnerable to malware attacks. Changing the device settings, tracking users based on the location of the devices, collecting sensitive data from the device, spam messages, and so on are the malwares which are increasingly threatening unsuspecting users.[4]

Device theft: Device theft remains a biggest worry for any organization supporting mobility. Mobile device management (MDM) solutions provide support for remote disk wipe for lost devices. There are applications that send an SMS to a registered number whenever a SIM card in the phone is changed automatically, and thus help in tracking the lost devices.

Carrier-Level Vulnerabilities

Transport layer vulnerabilities (MITM attacks): Man In The Middle (MITM) attacks are those that are targeted toward the network layer. MITM attacks allow an attacker to sniff the network traffic and gain access to the sensitive data. There are various ways to perform MITM attacks but the attack can be avoided by using SSL or HTTPS at the transport layer. Strong algorithms with good key size can prevent an attacker from decrypting the data.

Weak protocols (WEP): WEP is considered to be insecure, and encrypted data transmitted using these protocols can be easily sniffed and decrypted. WEP is considered to be weak as it uses RC4 cipher with a key size of 40 bits and a 24-bit initialization vector (IV) repeated across the message stream.

Rogue access points: Usually at public places such as hotels and airports, attackers can set up rogue access points to provide free Internet and lure innocent users to get connected to websites. A variety of tools can be used to act as a proxy and capture Internet traffic that can contain user credentials to websites. For example, a tool such as SSL strip can be used to capture credentials from sites using https protocol.

SIM-related attacks (SIM cloning): SIM cloning is a process of creating a copy of the original SIM. There are various tools and software available on the Internet that help in SIM card cloning. There are many methods through which SIM card details can be stolen. Essentially, the solution to prevent SIM card cloning has to be implemented at the service provider's end so that innocent customers are not charged for usage by cloned SIM cards.

Vulnerabilities at the Enterprise (Server-Side Vulnerabilities)

Server-side attacks: Enterprises are vulnerable to server-side attacks such as SQL injection and buffer overflow even when the access devices are mobile devices. Enterprises have to secure these servers in the same way as it is done when accessed by web browsers. Firewalls and IDS (intrusion detection systems) still need to be used as protection mechanisms to ensure that only the necessary services are exposed to the world and the ones exposed are monitored.

Rogue appstores: Appstores are the preferred mechanism to allow users to download mobile apps. Care needs to be taken while installing apps from unofficial appstores that can turn out to be rogue appstores. Mobile malwares are largely spread using rogue appstores, which combine authentic application with viruses. A user might install it thinking it to be a genuine software, but behind the scenes, it can monitor SMS or steal personal information.

Different OS allow different levels of access to third-party appstores. Apple iOS devices can only install applications from iTunes appstore. One needs to jailbreak it to install third-party apps. Android allows installation of third-party apps.

DDOS: Cyber criminals own a vast network of botnets and use them for performing distributed denial-of-service (DDOS) attacks against websites. DDOS is a sophisticated method of generating malicious traffic and makes servers unavailable to authentic users. It is a difficult attack to prevent as the source of attack might come from different destinations. Recent attacks by hacktivists have concentrated on using DDOS to bring down servers.

Database security: Appstores that store customer data should take extra precautions and ensure that customer data are protected. PCI compliance provides a comprehensive list of controls that should be implemented to ensure that credit card and personal identification information (PII) is protected for the data at rest, transit, and display. In addition to PCI, there are other legislations such as HIPAA, PIPEDA, European Data Protection Directive, and so on, which mandate adequate protection to PII and PHI (Protected Health Information).

Tools Leveraged by IT Departments in Leading Enterprises for Addressing Mobile Technology Challenges

Table 5.2 provides a mapping of the approach or tools enterprises used to address key challenges and vulnerabilities.

MEAP

Mobile Enterprise Application Platform (MEAP)[5] is a platform that provides enterprises the capability to mobile-enable their business processes securely. MEAP essentially provides enterprises the capability to

Table 5.2 Approach Used by Enterprise to Address Mobile Technology Challenges

Challenge	*Approach Increasingly Being Adopted by Leading Enterprises*
Device and technology diversity and heterogeneity	MEAP
BYOD	MEAP + MDM + enterprise mobile appstores
Device-level vulnerabilities	Remote patching, OS upgrades
Carrier-level vulnerabilities	Transport layers security and data encryption
Server-side vulnerabilities	Firewalls, IDS, application firewalls

- Deploy mobile applications across heterogeneous devices and OS
- Manage the handset and OS heterogeneity of mobile devices
- Synchronize data between smartphones and enterprise servers
- Increase mobile developer productivity by providing a rapid application development toolkit (mobile application development platform) that supports drag-and-drop controls for application development
- Deploy policy-based configurations on mobile devices

In addition, a MEAP provides a host of important management and security features such as backup and restoration of critical data, distribution of software, and sending out automatic updates to applications and antivirus software to mobile device with minimal impact to the user. It can optimize distribution over low-bandwidth connections by helping compress applications. It also supports security features such as Power-On Password and password lockout and can also lock down Bluetooth ports wherever necessary. It also helps in automatically configuring device settings.

MDM

The need to support BYOD has resulted in organizations accelerating their deployment of MDM solutions in their enterprise. Some of the important features that MDM supports include

1. Remote administration
   a. Over-the-air distribution of necessary applications and start/stop required services
   b. Firmware upgrades
   c. Remote device tracking, wipe in case of device theft or loss
   d. Remote password resets
2. Data security
   a. Security at rest: Ensure that the data stored in a mobile (file system and database) are encrypted. Only authorized person should be able to access the data.
3. Third-party app installation
   a. Ensure that only permitted apps from recognized appstores are allowed.
   b. Access control on apps to ensure that only appropriate services are allowed to be accessed.

Enterprise Appstores

To ensure that enterprise users are able to search and download authorized enterprise mobile applications on their devices, more and more enterprises enable the distribution of applications only through enterprise appstores rather than public appstores.

Given the rapidly evolving nature of the enterprise mobility space, the borders between MEAP, MDM, and Appstores are getting redrawn.

The development components in MEAP are being increasingly referred to as MADP (Mobile Application Development Platform), while new tools for Mobile Application Management and Mobile Content Management are being adopted by the enterprise. The device, application, and content management tools are collectively being grouped as Enterprise Mobility Management tools.

Best Practices

Tackling Heterogeneity

Mobility space is evolving, and it is always better for enterprises to design their applications and tools for these evolutionary challenges. One of the key evolutionary challenges is device and ecosystem heterogeneity. Mobile apps must be designed for heterogeneity. Similarly, enterprise tools need to support heterogeneity. MEAP and MDM can address today's challenges, but enterprise IT must be prepared to invest on procuring additional tools in the future to adequately address evolutionary challenges.

BYOD Precautions

Adequate tools to enable remote device wipeout in case of device theft are mandatory before allowing BYOD. When implementing BYOD, enterprise IT must ensure that the enterprise IT policy does not restrict the employee's right to use his or her personal device for their legitimate personal needs. BYOD does not mean absence of providing technical support for devices. Enterprise IT must be equipped to provide reasonable level of technical support to personal devices of employees. Enterprise IT must set up a mobility center of excellence to constantly monitor the mobile technology evolution, evaluate new devices, OS, and management tools, monitor loopholes and risks in mobile technologies, and advise on the permissible list of devices, minimum device configurations, and OSs, which can be used by the employee to access enterprise data and frame BYOD guidelines and policies.

Precautions to Address Additional Vulnerabilities

As a good practice, employees must not be allowed to connect their smartphones to access points that use WEP protocol for data transmission. While allowing an employee-owned device to be connected to the enterprise, it should be ensured that certificate-based authentication is used with WPA2 enterprise protocol.

Any device that is to be connected to corporate network should be registered with the enterprise IT. Certificates should be pushed to the device at the time of registration and subsequently used for authentication.

Employees must be allowed to download enterprise applications only through the enterprise's own appstores and not through public appstores.

Enterprises should restrict jail-broken devices as it increases the risk of rouge app getting installed on the phone

Mobile applications must address data at rest vulnerability.

Conclusion

The need to enhance employee productivity and enhance customer experience as well as the need to provide access to enterprise data to employees, partners, and customers have brought mobile devices into the enterprise IT. The introduction of mobile technologies into the enterprise comes with additional challenges, vulnerabilities, and pain points for enterprise security. The

enterprise IT needs to be equipped with a mix of tools, procedures, processes, and best practices to address the challenges and vulnerabilities. Taking a strategic view as opposed to a piece-meal view is needed to address the challenges of device and technology heterogeneity, BYOD, and other vulnerabilities.

References

1. http://www.gartner.com/it/page.jsp?id=2048617.
2. http://blogs.msdn.com/b/b8/archive/2012/04/19/managing-quot-byo-quot-pcs-in-the-enterprise-including-woa.aspx.
3. http://www.cvedetails.com/product/19997/Google-Android.html?vendor_id=1224).
4. http://articles.economictimes.indiatimes.com/2012-06-17/news/32281927_1_mobile-malware-mobile-devices-android-platform.
5. http://www.sdn.sap.com/irj/sdn/go/portal/prtroot/docs/library/uuid/80701ab1-7153-2e10-6db9-d57b8d0b4b7d.

Chapter 6

An Introduction to Digital Rights Management

Ashutosh Saxena and Ravi Sankar Veerubhotla

Contents

Introduction

Digital rights management (DRM) is a collection of access control and encryption technologies used by publishers, copyright holders, and hardware manufacturers to limit and regulate the usage of digital content. With rapid information sharing and unauthorized distribution of high-value content, digital piracy has grown multifold. This piracy, which was initially confined to the entertainment industry involving movies or music, is now spreading to the e-books and software market as well. With the increase in electronic reading devices, smartphones, and tablets, the problem of piracy is only getting worse.

In 2011, Frontier Economics estimated that the U.S. Internet users annually consume between $7 and $20 billion worth of digitally pirated recorded music. The same report estimated that the total value of counterfeit and pirated products impacting G20 economies in 2015 would be in the range of $1220–$1770 billion, which was $455–$650 billion in 2008. According to the ninth annual Global Software Piracy Study (2012) by Business Software Alliance (BSA), which is a software industry lobbyist group dedicated to combating digital piracy, 57% of global PC users acquire pirated software, up from 42% in 2011. The BSA also pointed out that the cost of unchecked digital piracy in the software industry was $63.4 billion in 2011 whereas software worth of $59 billion was illegally downloaded in 2010.

The rise in these figures, year on year, is alarming and poses a tough challenge to all the concerned industries that are losing revenues. Hence, it is necessary for legal and IT security professionals to be better equipped for combating piracy.

To counter the piracy menace, DRM solutions are widely practiced, which range from software approaches to hardware designs. Early DRM solutions were proven not-so-strong, but today, with technological advancement, the situation is much better and favorable to the publisher. Today, DRM solutions use strong encryption techniques combined with watermarking or fingerprinting to track the usage. Moreover, users can consume the protected content either in the online or in the offline mode. Apart from choosing the proper DRM solution, implementing the solution correctly is also equally important. Cost, performance, usability, and scalability are few of the additional parameters to be considered for selecting a right DRM solution, apart from security aspects.

Digital Rights Management

Background

In the predigital era, the people's ability to use and alter the content was limited. But in this networked digital era, it is possible to do just about anything to the digital content instantly and with minimal cost. Today's digital contents are in various forms such as documents, e-books, audio, video, games, and software binaries. In general, any business that needs to control access to its content or intellectual property documents is a potential user of DRM. Thus, there is a need for a technology that enables the secure creation, distribution, and management of digital content.

DRM places a digital lock on the digital asset to regulate the usage, thereby protecting it from being misused. DRM systems can be used to specify user rights such as read, play, edit, copy, and print. These rights are enforced during the consumption of the content by a trusted client. DRM has been an active area of research for decades. Previously, the main intent of DRM was to prevent the user from making illegal copies of the proprietary content and limit its distribution to only those who pay. Hence, early DRM solutions focused only on different encryption techniques to solve the issue of unauthorized copying. This was the case with the first-generation DRM solutions

where it was comparatively easier to bypass the DRM restrictions. Today's second-generation DRM solutions are far better as they are capable of protection, identification, monitoring, and tracking for digital assets.

Types of Rights

Rights are creations of law. Property is a bundle of rights, protected and guaranteed by a government. Examples include real property, personal property, intellectual property, and so on. Intellectual property (IP) is a general term for intangible property that is an outcome of an intellectual endeavor or the creation of the mind. Intellectual property right (IPR) is the legal recognition of the ownership of IP. In general, the following forms of IP are recognized: copyrights, patents, registered design, trademarks, know-how, and confidential information.

Copyright allows the creators of a *work* to control the use of their material, such as making copies, distribution, or its use in public domain. However, copyright cannot protect ideas and other forms of IP. A patent issued for an invention permits the inventor the right to stop others from making, using, or selling; offer to sale; and import the invention without the permission of the inventor. When a patent is granted, the invention becomes the property of the inventor, which, like any other form of property or business asset, can be acquired or licensed.

In the case of digital content, it is very easy to replicate and distribute it. Thus, it is important to identify the rights, which are applicable to digital content for its legitimate usage and distribution. Essentially, DRM is the management of these digital rights. Such rights consist of permissions on how the content can be used and constraints such as duration for which the content can be accessed. Digital rights can be classified into two main categories—static rights that do not change with time and dynamic rights that may be altered by the application of content usage policies.

DRM Principles

DRM principles ensure that the desired goals for content protection are attained. DRM controls the access to sensitive content by including information about the user rights for the content in the form of a license. Initially, the digital content will be encrypted by the publisher, with a random secret key to prevent the unauthorized copying and misuse of the content. In some scenarios, such as multimedia encryption, selective encryption techniques are employed. These techniques partially encrypt the content, by choosing important portions of the content or by making a random choice. The encrypted content is distributed to users for consumption along with a user license. The publisher, who owns the content, grants the user rights applicable for it.

DRM systems comprise of many different subsystems that handle content distribution, licensing, rights handling, and deterring mechanisms. In some instances, the DRM server plays a dual role as a repository for content and a license server. In this scenario, the server holds the content and decryption keys, and is responsible for rights management and license distribution to authorized consumers on behalf of the publisher.

Protocols and Industry Standards

Various organizations, for example, ContentGuard, Open Mobile Alliance (OMA), W3C, and Open Rights Group, are working toward establishing DRM standards. The OMA, which is a consortium of wireless, IT industry, and mobile manufacturers, released the OMA DRM. The aim of the OMA DRM is to facilitate a controlled consumption of digital content by allowing content providers to

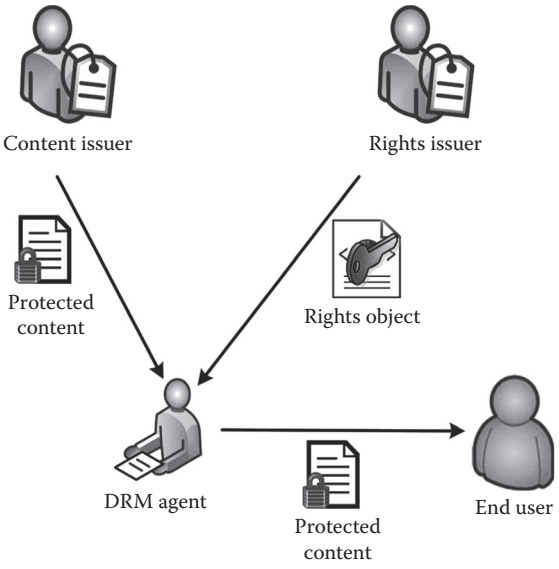


Figure 6.1 Functional architecture of OMA DRM.

express usage rights by granting permissions on the digital content, which in turn defines how the content can be consumed. Figure 6.1 shows the functional architecture of the OMA DRM.

OMA DRM 1.0 supports forward-lock, combined delivery, and separate delivery.

- *Forward-lock*—This prevents the content from leaving the current device by blocking forward (to another user) option.
- *Combined delivery*—This mode packages the digital content and rights together for delivery.
- *Separate delivery*—This mode provides the content and user license as two separate files. In this case, by changing the license file, new rights can be acquired.

OMA DRM 2.0 extends its predecessor and controls the content and rights separately. A few of its characteristics are as follows:

- Content is encrypted using a symmetric key either in
  - DRM content format (DCF) for discrete media
  - Packetized DRM content format (PDCF) for streaming media
- DRM licenses are handled as rights objects (RO) and acquired through rights object acquisition protocol (ROAP). RO is an XML document created to specify permissions or constraints associated with the content.
- Public key cryptography-based techniques are used to authenticate devices and bind RO to devices.

OMA DRM 2.1 has several additional features on top of OMA DRM v2.0, which includes

- *Metering*, mainly intended for information gathering on how the content is used, thereby enabling the rights issuers to collect royalty based on the actual usage of the content.
- *Content differentiation*, describing a mechanism to control the content consumption. For example, this mechanism can prevent the music track to be used as a ringtone.

- Support for user *editable metadata*, besides content issuer-defined metadata.
- *RO upload functionality* that enables users to upload rights from their old device to a rights issuer, which can then be downloaded to their new device.

The OMA DRM defines a general framework for downloading rights to devices, but the OMA secure removable media (SRM) standard allows users to move and consume rights on a different device. Rights expression language (REL) is used to express digital rights and achieve interoperability by DRM vendors. Major RELs include ODRL specification from OMA and XrML specification from ContentGuard.

DRM Practices

The DRM solutions that are being practiced today range from the software approaches to the hardware designs. Each approach has its own benefits and limitations. Techniques such as licensing, watermarking, and fingerprinting are few, generally used on the software front, whereas in hardware, security is relied upon the external hardware objects such as dongles and SIM cards.

Software

To improve the DRM protection, initially, the content can be watermarked, before encrypting and distributing it. Content protection and deterring mechanisms that monitor and track content include watermarking and fingerprinting. Deterring measures do not aim at preventing copyright violation but make copyright violations detectable, verifiable, and thus prosecutable.

Watermarking techniques hide a message or copyright information in the content. In the event an unauthorized copy is traced, the content owner can recover the watermark and use it as an evidence to sue the culprits. The main requirements of watermarking techniques are robustness, imperceptibility, and security. Robustness is the ability of a watermark to survive intentional and inadvertent distortion. The watermarking process must not affect the fidelity of the content and for this reason the embedded watermark must be imperceptible. Watermarking must be secure to prevent unauthorized detection, embedding, or removal. There are several watermarking techniques such as least significant bit (LSB) insertion, and discrete cosine transform (DCT)- and discrete wavelet transform (DWT)-based methods.

The LSB insertion technique is the simplest method for embedding watermark. In color images, each pixel has three components, namely, red, green, and blue. Assuming 3 bytes are allocated for a pixel, each of these colors has 1 byte, or 8 bits. In the LSB technique, watermarking information is embedded into red, green, or blue bytes by storing 1 bit of information in each least significant bit. So, for each pixel, we can hide 3 bits of watermarking information, in the LSBs.

Watermarking in the frequency domain involves the modification of the image (or media) in the transform domain. In the DCT method, the image is first transformed into the frequency domain by the use of DCT. Subsequently, the DCT coefficient values are modified by adding watermark information. The inverse transform of the marked coefficients forms the watermarked image. The DCT allows an image to be broken up into different frequency bands, making it much easier to embed watermark information into the middle frequency bands of an image to withstand compression and noise attacks.

The DWT-based methods are similar to other transform domain methods such as DCT embedding, but can model human visual system (HVS) more accurately. In the DWT-based

methods, initially, the image is decomposed into higher- and lower-resolution bands by choosing a particular wavelet. The watermarks are then embedded into high-resolution bands where the HVS is less sensitive.

Fingerprinting techniques are another form of watermarking techniques. Fingerprinting makes the digital copy unique by embedding a unique identification or serial number into it. Fingerprinting has an ability to track back the culprits who circumvent the copy protection by analyzing the pirate copy under circulation. Watermarking deals with embedding identification information in a cover, such as a video or audio signal robustly, whereas fingerprinting mainly concentrates on creating a unique identification number, which, when embedded in a digital copy, can track the culprits who circulate unauthorized copies. Some of the fingerprinting methods do not alter the digital copy; instead, they use the natural properties unique to the content. Software licensing solutions also use similar techniques to protect software.

Hardware

There are many kinds of DRM solutions available in the market. In a device-based DRM, the rights management comes from the mandatory usage of customized players and unique global device identifiers, for example, the international mobile equipment identity (IMEI) number. However, this kind of DRM scheme is constrained by its inflexibility, especially in the mobile environment. Alternate approaches use field programmable gate array (FPGA) and application-specific integrated circuit (ASIC)-based security platforms or smart card-based design where a chip is the core component of the scheme. This chip carries all the intelligence to identify, encrypt/decrypt, and store the required data. But the complexity and cost associated with this kind of DRM schemes are high.

Working of DRM Systems

DRM Architecture

Because of the rapid increase in counterfeit and illegal distribution of digital content, there is a considerable need for DRM solutions in the market. The major responsibilities of a DRM system include secure delivery of content, prevention of unauthorized usage, enforcement of user rights, and monitoring of the content. In a typical DRM system, each customer obtains the encrypted content from a distribution center on their network or over the Internet. To decrypt and access the content, a license, containing the user rights and the access key, needs to be obtained from a license server. The license server authenticates the customer based on their credentials and returns a license file. The customer's workstation must have a DRM client trusted by the DRM system to render the content after enforcing the rights as per the license. The DRM systems support online and offline models for the consumption of the content, so that appropriate content delivery and licensing methods can be used.

a. *Online model.* In this model, the DRM client has to connect to the DRM server to consume the content. Once connected, the end user authenticates to the server and gets a license through a secure session. The license that includes the secret key for content decryption along with the information on rights is transferred to the client during content consumption. The DRM client decrypts the content in memory and enforces the rights specified in

the license. By changing the rights information on the server, the publisher can grant new rights or extend existing privileges to the user. This online model provides the greatest flexibility when assigning rights to any combination of users and controls the usage of content such as view, copy, edit, and forward.

b. *Offline model.* The offline DRM model is applicable to the scenarios where the DRM client has limited connectivity to the DRM server. In this case, the client can choose to work offline after the initial connectivity to the DRM server. The DRM server establishes the trust among the client's workstation, the end user, and the DRM client using a reliable mechanism such as digital certificates. The end user obtains a DRM license and stores it locally, which will be used every time the content is consumed.

DRM Components

DRM systems incorporate many different integrated mechanisms and functionalities such as content protection, distribution, licensing, payment systems, access control, rights handling, and deterring mechanisms.

The DRM solutions come in various flavors but most of them have the combination of stages defined as follows:

- *Packaging.* Normally, encryption methods are used to protect the digital content before user-specific rights are granted. In some cases, watermarking techniques are also used in packaging the content.
- *Content distribution*. The DRM-protected files are delivered to the customers through the Internet, e-mails, or through a physical medium such as CD/DVD. Using superdistribution, the encrypted content can be shared with anybody without any restrictions. However, as licenses are customized and cannot be transferred, each customer has to acquire a new license for using the content.
- *License service.* Specialized servers are placed to authenticate legitimate users through an Internet connection to allow or deny access to the DRM-protected content using various authentication mechanisms. DRM licenses that generally contain cryptographic keys are used to manage protected content. Legitimate users use these keys to unlock their files. The choice of how to provide these cryptographic keys to the users will vary with the DRM scheme. As such it is not advisable to give these keys in plain to the end user, since it may lead to the circumvention of the DRM protection.
- *License acquisition.* Users generally need to pay a fee to the publisher through a payment gateway or prove the legitimacy by an authentication process before acquiring a license from the license server.
- *Communication protocols.* Various components in the DRM system need to communicate with each other using a set of predetermined commands or protocols to make the solution work. User authentication, rights acquisition, and rights revocation are examples of such activities.
- *Rights enforcement.* User rights need to be enforced at their end by means of a trusted application.
- *Tracking.* In some scenarios, DRM systems may need to monitor the use of content. This is mainly carried out by fingerprinting methods. Broadcast encryption systems that allow targeting of an encrypted message to a privileged group of receivers can also track the contributors of a pirate decoder.

- *DRM client.* This client, also referred to as a DRM agent, is a trusted application that resides in an end user's workstation. It acts as an interface between the end user and the corresponding DRM server(s), handling user authentication, license acquisition, content rendering, rights enforcement, and tracking. This client can be a dedicated application, a plug-in to existing applications such as Adobe PDF reader or a customized browser.

Limitations of DRM

Given the variety of devices (with different form factors) used to consume digital contents and the lack of a common ground, the interoperability of DRM systems is a major problem today. A variety of organizations are working on DRM standards; however, there is a lack of a common standard. DRM vendors either follow different DRM standards and rights expression languages or create one of their own. This makes the interoperability of DRM systems difficult. The content protected on one platform may not be used on any other platform. Similarly, users cannot choose their favorite media player for rendering the DRM-protected content due to interoperability issues.

Another major challenge is related to the deployment of DRM technology for the end users. End-user environments are heterogeneous and mainly untrusted. DRM clients generally enforce the rights by means of customized content rendering software and using cryptographic keys. End users will not be interested to install customized software at their end without any incentives or perks. The protection of cryptographic keys and licenses is very difficult in an untrusted environment. The circumvention of the DRM protection is possible by ripping the audio, converting digital content to analog forms, capturing the decryption streams, or using print screen options for readable content.

As a genuine concern, archival of the DRM content will be difficult due to the fact that the technology may become obsolete after a couple of years or a particular DRM vendor may go out of business. However, this problem can be partly addressed by using proper key escrow mechanisms and establishing decommissioning procedures for DRM.

Implementation of DRM Systems

The implementation of a DRM system involves an initial risk assessment for identifying possible threats to content, development of use cases, defining requirements, RFP creation, and vendor evaluation to finally create a DRM system.

Identification of Scope

In this phase, the stakeholders' requirements are to be identified and documented. This may incorporate different perspectives of publishers and end users. The publisher is mainly concerned about content protection and its secure distribution whereas the end users may wish to consume the content on a wide variety of platforms with ease. It is important to identify the classification, format, volume, usage, and life cycle of the content to understand how the content is to be protected and which digital rights are to be managed. It is also essential to consider and understand the business model for content dissemination, which may include paid downloads, subscriptions, rentals, pay per view, try-before-you-buy, and so on. The DRM solution shall support the required business models and address the technical challenges to sustain it. Thus, the

activities for the identification of scope mainly involve distribution of questionnaire, conducting interviews, data-gathering techniques (see Annexure), and an initial risk assessment to identify the threats for the content.

Analysis of Requirements

Publishers may wish to protect a variety of content formats. For example, in the case of e-learning, the website might host audio/video clips apart from html and PDF content. If a download option for the content is provided to the end users, the DRM protection must be persistent on the user's environment. Even though the end user wishes to use the content on multiple devices, including mobile gadgets, the publisher may wish to limit the access to desktop or a set of devices. A music company may wish to distribute music albums and other creations of its artists over the Internet. However, they may be reluctant to distribute specialized software for content protection such as trusted DRM clients along with it. In the case of conflicting requirements from stakeholders, a trade-off needs to be achieved for a successful outcome.

Implementation Choices

There are many DRM solutions available in the market for off-the-shelf use. These solutions include Microsoft Active Directory Rights Management Services (ADRMS) for Microsoft Office documents, Microsoft PlayReady content access technology for music, video, ringtones, images, or games, ADOBE Content server for PDF content, LockLizard's Lizard protector, and IBM's WebGuard for web content protection. Many other DRM solutions and alternatives can be found on the Internet. Content publishers or organizations willing to implement DRM solution have to make an appropriate choice between buy and build, once the scope and the use cases are identified. Buying a DRM solution from the market is advisable when an existing DRM solution meets their requirements and is cost effective and reliable. Otherwise, a custom DRM solution needs to be built to suit their needs. If the organization understands the DRM technology and has the resources, they may consider building the solution on their own. However, one has to be careful with the developmental costs and rework. An alternate approach would be to customize an existing solution to suit the needs of the organization and integrate it.

Evaluation Framework

This section presents an evaluation framework for DRM solution as an illustration, which is based on major DRM requirements such as flexibility, efficiency, interoperability, and security. However, based on business needs and operating models, this framework may be amended or customized to add or remove new components.

a. Flexibility requirements
   Content format. Digital content is available in a wide variety of formats. It should be possible to protect these content formats with the DRM solution.
   Assignment of rights. DRM should be able to enforce rights at the granular level, for a selected user, on a selected content.
   Consumption of content. Consumers shall be able to consume the content with ease. Device restriction for consumption may be needed on the business need.
   Platform support. The DRM solution shall be portable to a variety of platforms.

b. Efficiency requirements
   Robustness. The DRM architecture shall be robust to support load balancing and clustering in the case of a large user base and high-volume transactions.
   Complexity. The DRM solution should be fairly simple and be easily implemented. It should also support time-tested and common business models for content distribution. Complex systems tend to fail easily.
   Network traffic. The DRM solution should not congest the network traffic. In the case of huge files, it is desirable to protect them as they are, rather than uploading them to the DRM server. License acquisition and distribution also accounts for network traffic.
c. Security requirements
   Availability. The protected content shall always be available to all the legitimate users. If a user license is lost, a backup copy should be provided to the legitimate users.
   User rights management. Rights management should be granular to the required extent. It is desirable to have a support for revocation of rights.
   Multiuser environment. Users need to authenticate to the DRM server for fetching licenses. In the case of multiuser environments, only authorized users should be able to access the licenses stored locally.
   Communications security. All the communications from DRM client to server shall happen over a secure channel such as SSL. Every time the content needs to be consumed, user authentication to server over https is needed.
   Tamper-proof license. In the event of a tampered license, the DRM client shall reject the license file.
   Protection of content. Content shall be protected with reliable encryption schemes. The cryptographic keys shall be adequately protected using access controls or using key wrapping.
d. Interoperability
   Rights expression. The DRM system may use popular rights expression languages such as ODRL or XrML for interoperability.
   Third-party integration. The solution shall support third-party DRM client integration using an API.

Match the Business Needs

At the end, it is important to match the business needs with the capabilities of DRM solutions. If the DRM solution is not implemented properly, it will not cater its intended purpose; instead, it may become a roadblock for future activities. Prior to DRM implementation, a checklist of requirements is to be prepared and each solution may be carefully evaluated and weighted. Additional care may be taken to get support from vendors for doing pilot or prototype implementations and determine the right solution.

Conclusion

DRM aims to protect, distribute, manage, and enforce user rights associated with the use of digital content. Content delivery methods ensure that the content is properly distributed to legitimate users and prevent unauthorized copying. Users need to acquire consumption rights from the publisher by paying a royalty or fee. During consumption, the end users are authenticated

to the DRM server by a trusted client, which is also responsible for enforcing user rights. DRM is an evolving and promising technology but has not yet reached its perfection. There are merits as well as shortcomings associated with it. DRM allows new content to be made available in a safe and trusted environment. It enables industry and content owners not to encode their works in proprietary formats. For a foolproof content management system, additional access controls and data loss prevention (DLP) techniques can be used in conjunction with DRM systems.

Annexure: Sample Data Gathering Templates

Content Type Requirements—This checklist helps to identify if the solution supports all the content types.

Requirement	*Vendor's Compliance*
Single media—Text, audio, still images (PDF/DOC/TXT/JPEG/MP3/e-book, etc.)	★★★★☆
Multimedia—Animation, video, games	★★★☆☆
Executable code—Dynamic link libraries, Java class files, executable file	★★★☆☆
Stream data—Video/audio for broadcasting	★★★☆☆
Designs and drawings—CAD files, proprietary formats	★★★☆☆

Note: Low -★☆☆☆☆; High -★★★★★.

Business Model Requirements—This checklist helps to identify whether the solution supports the popular business model(s).

Requirement	*Vendor's Compliance*
Try-before-you-buy—Enables evaluating prior to purchasing	★★★☆☆
Pay per use model—Facilitates the customer with more licenses than they had purchased. It can support Time based Volume based	★★★★☆
Subscription/rental—Facilities the customer to subscribe to a content for a specific duration	★★☆☆☆
Lending—Facilitates the customer to lend or borrow content	★★☆☆☆

Note: Low -★☆☆☆☆; High -★★★★★.

Content Protection and Security Requirements—This checklist helps in gathering the content protection and security requirements.

Requirement	*Vendor's Compliance*
Technology used—Has the current technology been used in the solution?	★★★☆☆
Encryption—What kind of encryption techniques are used?	★★★☆☆
Watermarking—What kind of watermarking techniques are used?	☆☆☆☆☆
Fingerprinting—Is it possible to uniquely identify each digital copy?	☆☆☆☆☆
Authentication mechanisms—What kind of authentication mechanisms are used?	★★★☆☆
Secured communication—Is data secured during transmission?	★★★☆☆
Granularity of rights—To what level can rights and permissions be assigned?	★★★★☆
User restrictions—To what level can device, IP, or geography restrictions be enforced?	★★★☆☆

Note: Low -★☆☆☆☆; High -★★★★★.

Business Enabling Requirements—This checklist helps in identifying additional requirements and prerequisites for going live.

Requirement	*Vendor's Compliance*
Time to market—How much time will it take to launch the solution in the market or go live?	★★★☆☆
Distribution—How easily can one distribute the content?	★★★☆☆
Payment gateway—Is there a secure payment gateway for monetary transactions?	★★☆☆☆
Royalty—What is the possibility that publishers do not lose the royalty for the content used?	★★★☆☆
Maintenance—How easy is it to maintain the solution? Is it cost effective?	★☆☆☆☆
User acceptance—What would be the user acceptance level for the solution?	★★★☆☆

Note: Low -★☆☆☆☆; High -★★★★★.

Further Reading

Adobe. ACS, Adobe Content Server. http://www.adobe.com/products/

Business Software Alliance. 2011. *2010 Eighth Annual BSA Global Software Piracy Study*. http://portal.bsa.org/globalpiracy2010/

Business Software Alliance. 2012. *2011 BSA Global Software Piracy Study*, Ninth Edition. http://portal.bsa.org/globalpiracy2011/

Cox, I.J., Miller, M.L., and Bloom, J.A. 2002. *Digital Watermarking*. Morgan Kaufmann Publishers, San Francisco, CA.

International Chamber of Commerce. 2011. Estimating the global economic and social impacts of counterfeiting and piracy. http://www.iccwbo.org/Advocacy-Codes-and-Rules/BASCAP/BASCAP-Research/Economic-impact/Global-Impacts-Study/

Koved, L. System and method for supporting digital rights management in an enhanced javaTM2 runtime environment. U.S. Patent 20020161996, filed Feb 23, 2001, and issued Oct 31, 2002.

LockLizard. Lizard Protector web security. http://www.locklizard.com/html_security_features.htm

Microsoft. 2012. ADRMS, Active Directory Digital Rights Management. http://msdn.microsoft.com/en-us/library/cc530389(v=vs.85).aspx

Microsoft. 2012. PlayReady. http://www.microsoft.com/PlayReady/Default.mspx

Mourad, M., Munson, J., Nadeem, T., et al. 2000. WebGuard, A system for web content protection. IBM white paper. http://domino.watson.ibm.com/library/cyberdig.nsf/papers/D2CC8887A94BCFF585256A01006F8727/$File/rc21944.pdf

ODRL, Open Digital Rights Language. http://odrl.net

OMA-DRM, Open Mobile Alliance Digital Rights Management. http://www.openmobilealliance.org/technical/release_program/drm_v2_0.aspx

Open Rights Group. http://www.openrightsgroup.org/

Rosenblatt, W., Trippe, W., and Mooney, S. 2001. *Digital Rights Management: Business and Technology.* M&T Books, New York, NY.

Veerubhotla, R.S. and Saxena, A. 2011. A DRM framework towards preventing digital piracy. In *Proceedings of IEEE, 7th International Conference on Information Assurance and Security*, Malaysia, pp. 1–6.

W3C. http://www.w3.org/

XrML. http://www.contentguard.com/

Zeng, W., Yu, H., and Lin, C.-Y. 2006. *Multimedia Security Technologies for Digital Rights Management.* Academia Press, Burlington, MA.

Chapter 7

Information Security on the Cheap

Beau Woods

Contents

An informal survey of the stories on information security budgeting suggests that 40–60% of organizations are cutting or holding their budget the same over the last year. And of those that are increasing their budgets, most increases are small. This is at a time when the actual threats are on the rise, security has risen in importance to the organization, vendors are daily coming out with solutions to problems we did not know we had a year ago, and hiring and keeping good people is becoming increasingly difficult and expensive.

Thankfully, effective information security does not have to be overly expensive. Although many of us feel we do not have the budget we need to do our jobs, it is alright. Many of the things

we need to do would not cost a lot. That is not what we are hearing in the media and from vendors, but it is true nonetheless.

Effectiveness is almost never achieved by doing a single thing and this is true for information security effectiveness as well. Instead, it helps to use many different approaches and tactics, which together can make your program effective. And it also helps to shape or form a philosophy that will guide your decisions and your activities.

I have tried to encapsulate the philosophy behind the specific technical guidance into a few guiding principles. These principles are meant to act in concert with each other to build and enhance each other. I feel that the philosophy here is the one that will let you get the most out of your limited resources*. Some of the best-secured organizations I have seen have used these methods to be very effective and efficient. In fact, some have found it hard to spend their budget on more than one occasion!† And one found that they did not have enough work for everyone; so, they had to move him elsewhere. Imagine that!

Here are the guiding principles that I feel will allow you to be most effective and efficient with your resources:

1. Plans should precede actions.
2. The basics are highly effective.
3. Minimize the diminishing returns.
4. Look at the most efficient things first.
5. Use the power of iteration.
6. Demonstrate your success.

"Plan" Is Not a Four-Letter Word

Planning is about laying out your priorities so that everyone (including you) knows where you are going. A friend of mine was once asked in a job interview what she would do first if she did not have time to prioritize her actions. She responded, "that's exactly the time when you need to prioritize most!" She was right, she got the job, and is excelling. If you do not impose some kind of prioritization, you can only succeed by sheer luck. Documenting these priorities and decisions allows yourself and others to have a clear vision of how things should look.

> I have always found that plans are useless, but planning is indispensable.
>
> **—Dwight D. Eisenhower**

But planning is hard and takes time. It is much easier and more fun to just start working. And that is the problem. Planning requires a block of uninterrupted time, it takes freedom

* But with that said, this is your security program. I would not have to live with the consequences and I would not be by your side to help you out. So, it is important to make the philosophy work for you. In the end, results are your measuring stick, rather than following a particular dictum.

† Oh do not worry, they ended up spending it all. They refresh their MacBook Pros every year and always have enough to fund pet projects for research to stay ahead of the curve.

from distractions, and you need space and clarity. Probably, the biggest hurdle is just getting started on what can seem like a monumental task. But planning does not have to be a huge undertaking.

Plans should be short and straightforward. These are documents to help you and your organization know what is foundational and what is most important. Depending on the size of your organization, the documents may be as short as a page or as long as three to four pages. This also helps you to get started, knowing you do not have to climb a mountain.

Now that you know what you are working toward, you still need a way to get started and to keep up with things. You need to find a time and a place to work. And it needs to be free of distractions for a single block of time. I recommend blocking out time at the beginning and the end of the week—the first session, so that you are up to a speed on what matters and what you are working on, the second session to recap, measure your performance, and lay the foundation for the next week.

> Things which matter most must never be at the mercy of things which matter least.
>
> **—Goethe**

But however you do it, remember that this planning is likely the most important thing you will do all week! So, do not skip it and say you will come back to it later. Do not let others steal it or cut into it for less important things. Do not get distracted in the middle and do something else. Keep this time firmly, as it will be the thing that gives you the best chance at success.*

Focus on Fundamentals

When I was playing basketball as a kid, I used to try out lots of flashy moves. And playing pickup games around the neighborhood, it worked great—I had a great reputation as a shot blocker and could sometimes dunk the ball. But I was not very good at actually playing competitively and I did not make the cut for the high school team. That is because I did not practice things such as dribbling, shooting, strategy, and the basic fundamentals.

Often times, the most expensive products and projects are those that tackle the latest threats, tactics, and techniques. Although they are only used as a fraction of the time to compromise organizations, they grab headlines and take up a greater part of your mindshare. Consequently, these things are seen as a much bigger problem than the basic issues.† And because of that, too many organizations end up skipping the fundamentals to chase the latest threats. If you do that in basketball, like I did, your security program would not make the cut.

---

* For more on helping to carve out time and space to work, as well as on the power of planning, one of the most popular and best books on the topic is *Getting Things Done*, by David Allen.

† For more information on this, see two well-known effects in cognitive psychology: the recency effect and the availability heuristic.

Now, I am not going to rehash all the security fundamentals here. You have seen them all before dozens of times. If you want to know them, read up on certifications that have been around for a long time, such as CISA, CISM, CISSP, and others. These are broad surveys of information security that take into account both academic and practical perspectives. Instead, I will try giving you some practical tips on where you can get the most bang for just a little bit of expense. The courseware mentioned above does not take into account that prioritization.

> You've got to get the fundamentals down because otherwise the fancy stuff isn't going to work.
>
> **—Randy Pausch**

And do not forget other fundamentals, as well. Things such as good project management can make a huge difference in the effectiveness of whatever you do. Regular, routine maintenance is the same. These things never grab headlines unless they are for failures. The fundamentals such as these are often overlooked; yet, they can mean the difference between being effective and wasteful.

Minimize Diminishing Returns

At some point in your security spending, you will come to an inflection. That is, at some point, the next dollar you spend on some initiative will make you less secure than the dollar spent elsewhere. But it is hard to know where that point is. And it is even harder, after you realize you have reached it, to pull back. Especially if the spending has already been approved. Nevertheless, it is important to think about where that point is.

> Roughly 80% of the effects come from 20% of the causes.
>
> **—Pareto Principle**

As a thought exercise, think about the implications if 80% of your security improvements come from 20% of your spending. How do you identify the 20% and make sure that it is fully realized? How can you reduce the 80% of less productive spending and use it elsewhere? What things can you do to increase the effectiveness or decrease your costs? Should you spend $250,000 to do those antivirus upgrades or should you instead spend $50,000 to harden your endpoints? Is the additional $80,000 web application firewall going to be as effective as the $15,000 code audit?

> Example: A client came to me asking if I could evaluate a planned security initiative. His management was asking him to spend $1M procuring and implementing whole-disk encryption on his mainframes, based on their legal

council's advice. In going through this exercise, I developed several alternatives that were clearly as good and much cheaper, as well as some that were as effective and expensive, but clearly absurd. One I remember fondly was hiring guards to stand by the mainframe to prevent against physical theft—the main risk countered by fully encrypting the drive. Another was simply buying an additional insurance as well as isolating the system in the network. In the end, a simple and cost-effective solution won out, saving lots of money without substantially increasing risk versus the original solution.

This kind of thinking is more than just an exercise. You can also use this approach to create a "dream budget" that is much larger than your actual or projected one, fill it with everything you would like to have, and then slice it down to the 20% that is going to be the most effective. You may also look at the ways you are spending your time and simply drop or delegate anything that is not effective. Look at where your team is working and cut that back similarly. And this may mean you stop funding some projects that are consuming more money and resources than they are returning in security improvements.

Pick the Low-Hanging Fruit

If you are picking a few fruits from a tree, the lowest ones are the ones you will go for. If the ones within easy reach are not ripe or are damaged, you will look for fruits higher and higher until you find the lowest fruits that meet your criteria. This is a common-sense approach of eating that people and animals have used for hundreds of generations. It is easy to generalize this concept to achieving goals.

A simple life is not seeing how little we can get by with—that's poverty—but how efficiently we can put first things first.

—Victoria Moran

Given a particular goal, some ways of achieving it will be easier than others. If the goal is large enough, there will be many different things that contribute to achieving it. Some will be easier or more difficult; some will be effective whereas others will not. One of the keys for being effective and efficient is continually tackling the easiest tasks that get us the closest or the fastest. That does not mean avoiding the hard things; it just means first do the thing that is easiest among the equally effective alternatives.

It is easy to see how this analogy fits the information security management. But it is hard to keep the principle in mind as we are planning projects and going about our day-to-day activities. And sometimes, it is hard to see which fruit hangs the lowest. Too often, we try to solve the biggest problems or implement the most wide-scale solutions. But that is not always required, especially when we can make our low-hanging fruit tastier over time than the fruit way at the top of the tree—and we can do that more quickly, easily, and with lower risk of failure.

Iterate to Dominate

The process of successive improvements over time is extremely powerful. This is the same basic approach that underlies evolution, Agile software development, and the business strategy of Kaizen. The basic structure of these processes is to take an initial state and improve it in successive generations based on feedback, prioritizing improvements where they are most needed. The same concept can be applied to an information security program.

The goal here is to have good security across the board. Do not put in Fort Knox security in one area and none in another. Instead, get them all up to a minimum baseline, then improve gradually, and over time, you will have a very effective program. Essentially, you are putting the money where it is needed the most and you are minimizing the risk of missing something. All the while, building a great program from the inside out represents a defense in the strategy of depth. Rome was not built in a day. The greatest creations in the universe evolved over billions of years and trillions of iterations. You may not have that much time, but you can certainly use the power of iteration to build something great.

One key point in the power of iteration is to first do no harm; that is, when you are going through your first iteration, err on the side of caution. Do not do things that may impact business productivity, uptime, or other areas that would garner unwanted attention. Affecting the business is the surest way to doom even the best intentions.

Visibility for the Win

When was the last time the chief executive officer (CEO) or the board extolled the virtues of your security program? If you are like most organizations, it is not often. But it should be! You are probably doing a lot of great things. But it is usually a matter of how many people know what those things are and what they mean for the organization and for them. Making yourself visible helps in a number of ways, from building support to buying forgiveness to improving your actual and perceived effectiveness.

> If a tree falls in the woods but nobody hears it, did it make a sound?
>
> **—Philosophical Exercise**

You should be transparent and communicative outside your direct chain of command. That means talking with people often and in different ways, not just up and down the organization, but also sideways. You should spend some time speaking with the business units to understand them, get their input, and to talk about new ideas. This helps you understand where they are coming from, where they are going, what most concerns them, and ways to help. And it lets you vet ideas and approaches so that you have a better chance of success when you propose them formally. It also shows others that security is approachable and understandable, and that you and your group are responsive for their needs. All this builds political capital.

It is also important to talk with people in a way they can clearly understand you. That is part of what being an effective communicator means. Again, this is not just vertical in the organizational chart, but is also horizontal. A big part of this will come from understanding other

business units, their concerns, and their priorities. Even though most security managers tend to prefer to speak or think in either Greek or business, it is very beneficial to have a good command of both languages or to be bilingual. This is the best way to make sure you are understood throughout the company.

We cannot always win, and our wins cannot always be monumental. But it is important to keep the communication lines up to stay visible within the organization. And it also helps you shape and mold your ideas to be more successful. And in doing so, you will learn more about the business units you support, as well as they will learn more about what you are doing to help them. At the same time, learning how to speak to them in their language and hopefully transferring some of your passion for information security to them.

Putting It into Practice

Now that we have covered some general suggestions, how do you put those things into practice? Here are a few projects or programs you can implement that will allow you to provide more security with less money. If you are already doing one or more of these things, that is great! You may safely skip that section, although hopefully, if you read it, you will find it valuable as well.

The goal of this part of the chapter is to pass on pragmatic advice for securing your company without spending a fortune. The practical tips below do not necessarily lend themselves to a direct mapping against the six principles above, but you should definitely keep them in mind when you go to put these things in place in your organization. To summarize the first part of the chapter in a more practical and concise way:

1. Planning—Thinking about what will and what will not work, anticipating objections, and walking through processes to spot issues. Remember the maxim of "measure twice, cut once" saves time and money.
2. Basics—Do not underestimate the power of the basic steps to provide big benefits at a low cost.
3. 80/20 Rule—Look for places where you can get the biggest bang for your buck and focus on these first.
4. Low-hanging fruit—Sometimes, knocking off 10 easy things is more effective and less costly than taking on one large project.
5. Iteration—The most powerful processes in the universe leverage iteration to be both very effective and very efficient. You should do the same.
6. Communicate—Talk to and about your colleagues and peers across the company.

System Hardening

System hardening is the practice of configuring systems so that they run with the least functionality required to perform their business function. Simply put, that means stripping systems down to bare essentials without breaking anything. Most often, this is done prior to deploying the system, but you can also retroactively harden systems already deployed. Hardening trades slightly increased the up-front effort for much lower maintenance effort and cost at the same security level.

Hardening includes

- User account restrictions on workstations—By using limited user-type accounts, rather than the administrator type, it is almost impossible for malware to infect a system, increasing security and decreasing support costs. Support costs are further cut by preventing employees from accidentally changing important settings and having to call the help desk.
- Disabling unnecessary system services—Most services that are running on workstations and servers are critical to serving their business function. However, many are not and expose security vulnerabilities. One organization studied lost thousands of hours of productivity because a network worm attacked one of these services.*
- Enabling host-based network security—Modern operating systems and most endpoint protection suites have a host-based firewall built in with central management capabilities. This is a great tool for blocking network attacks, or if nothing else, then for detecting them. And you can use free tools such as Splunk to monitor logs and alerts—more about this below.
- Changing default insecure settings—Many systems are deployed with default passwords and other insecure settings. On a sufficiently large network, it is almost guaranteed that you have some default passwords or other insecure settings. A lot of the time, appliances and network devices ship with insecure default settings and passwords (though these are changing); so, really pay attention to these. And if you have not audited your gear for these low-hanging fruits, now would be a good time.

System hardening is probably the most effective tool you have for protecting your information technology (IT) systems. By hardening systems, you protect their critical components from accidental or intentional changes. To stay under the radar, persist on a system, and carry out its primary tasks, malicious software requires access to specific system resources. The same access would be required for an employee to accidentally or intentionally change the system in ways you would not want. In many environments, this prevents most of the desktop problems that the IT team sees. If given the choice between hardening and antivirus on a system, many IT security professionals would prefer hardening both in terms of effectiveness and cost savings. But doing both is preferable to either for many reasons, such as compliance and defense in depth.

Getting started on system hardening sounds intimidating, but it does not have to be. That is one reason why most large companies have implemented the key components of system hardening, but not so much within smaller organizations. But sometimes, even larger companies have not gotten started yet. If you are in either of these categories, here are some tips:

- At some point, there will be a need for someone to perform an administrative task on workstations. Many organizations have begun delegating these privileges to a limited number of individuals within a business unit, using a different account than their normal account. This takes some of the workload off of IT staff, empowers businesses to be reactive, and shifts the ownership for decisions that increases buy-in for support and security decisions.
- Iteration and testing will be key to preventing things from breaking. Start with the simplest systems, learn lessons there, and move on to trickier systems later. Also, start with less controversial or resource-intensive changes first, then go back, and tighten further as you can do more testing.

* The service was UPnP and the network worm was Zotob, in 2005. This same service was also used by the Conficker malware in 2009.

- The simplest systems to harden will be the systems exclusively using standard software, such as Microsoft Office, Adobe Reader, some light web browsing, and so on. Most of these software packages have been thoroughly tested in a hardened environment.
- If there are certain areas that are more standardized than others, you can usually harden these first to get the biggest bang for the buck. This also allows you to visibly demonstrate a successful rollout.

Pro Tip: It is effective to come up with some very basic hardening guidelines for your systems to allow you to guide the internal staff and vendors on what is expected of them. Keeping the guidelines flexible keeps the security from getting in the way of business. And these standards are something that can be easily validated automatically through vulnerability scanning. And having these documented allows you to go back behind vendors to make sure they have not just charged you a boatload of money to get you hacked. The key items to include in the guidelines for implementation and updating are

- Disable default accounts and change the default passwords.
- Disable the components of an application or a system unless they are required for proper functionality.
- Ensure that all components (including the underlying operating system) are using the latest versions with any security updates applied when implemented. For operating system patches, notifications should be sent for any specific security patches or updates that could affect the normal operation of the system.
- Ensure that the configurations are in line with the published hardening guides.
- Where deviations from these standards exist, they must be documented as well as potential risks, alternate controls in place, and residual risks. The business unit will make a determination whether to accept the risks or to look for alternate solutions.

There are dozens of system-hardening guides you can use, available for servers and workstations. The three that are most likely useful to you are available from Microsoft,* the National Institute of Standards and Technologies (NIST), and the Center for Internet Security (CIS).† Of the three, the Microsoft and CIS versions tend to be the most updated and are the easiest to read. Some excellent tools are available to help with this, as well, but are beyond the scope of this

* The same concept also applies on Linux, OS X, Unix, and other operating systems, but most companies run Windows desktops; so, that will be the focus here.

† Although they are a bit hard to find on the site without bumping into membership and payment applications, the CIS "Benchmarks" as they are called are free and as of this writing are accessible directly at http://benchmarks.cisecurity.org.

chapter.* It is important to remember that these are guides, though, and you should adapt them to work best in your environment.

Patch Management

Keeping your organization's software updated is a difficult task. Organizations tend to fall into one of two camps—either they cover nearly everything, but spend piles of money; or they patch nearly nothing, but do it for free. But there is a middle ground. It is possible to find a good balance of security and cost.

> **Clarification:** Patch management specifically refers to software updates provided by third-party vendors. Many people conflate patches with the software vulnerabilities they fix. But it is important to distinguish between patches, software vulnerabilities, and other types of vulnerabilities such as weak passwords. There are many ways to reduce the risks from software vulnerabilities; patches are only one, though usually the best option.

You can patch less often than a lot of people think. Although 0-days get a lot of press, they are rarely used in actual attacks. Many organizations have seen this and have applied a quarterly patch cycle. For patches that need to be applied sooner, they have emergency procedures that allow for quick decision making and action. This combination cuts out most of the effort and cost, is easier to manage, and can often give you the same level of security.

> **Info Bit:** 0-Day vulnerabilities have become difficult and expensive to find in most software. Because of that, attackers do not tend to use them, except for very high-value targets. And even then as a means of last resort, after other methods have failed. Instead, they will try guessing passwords, exploiting older vulnerabilities, spear-phishing attacks, and other means.

Some patches, however, will need to be applied more quickly than that. In some cases, attackers are thought or proven to be exploiting vulnerabilities associated with newly released patches. Or a working exploit code may be available to them. Organizations should have plans for emergency patch cycles, just for these types of events, which allow them to quickly respond and patch or apply workarounds. This is where a lot of people drop the ball—they do not plan for emergencies and so, they get caught flat-footed. That is the kind of thing that drives companies to the dichotomy of either overspending or doing nothing.

* CIS also provides tools to automatically apply their guidance to systems and to test your system against their benchmarks, although these require a membership. Microsoft provides similar tools for free, such as the Security Configuration Manager (SCM), although these may require more resources to implement and administer than the guides. Nessus, as discussed later, can also be used to help with the hardening process, as well as monitor systems for deviations from your standard.

Similarly, some systems will routinely need to be patched sooner. The systems exposed to the Internet have the greatest likelihood of being attacked. Therefore, they should be handled differently. Some organizations decide to patch all these systems immediately, with minimal testing. Whereas some organizations simply lower the threshold for what is needed to trigger an emergency patch cycle.

In managing a patch cycle, then, it is important to decide which patches should be applied to which systems and when. Which patches should be emergency fixes and which patches can wait for the patch cycle? In general, the patches that need to be applied immediately are those that fix an issue that is actively exploited or where the working proof of concept (POC) code exists. In most cases, the systems exposed to the Internet should be patched first to reduce the likelihood of becoming compromised. And it is important to make sure you consider the alternate methods of reducing risks (such as disabling certain vulnerable services temporarily), in case patching is deemed riskier than not patching. These alternate methods can be a great way to buy time needed to do proper testing or in case a patch is not yet available.

Pro Tip: Several organizations have begun using a few guidelines to easily manage patches across several thousand workstation environments. These work well especially when the deployments largely have the same used cases and software. This can mean, for example, a patch cycle of <1 month, almost no time testing but with the agility to apply emergency patches almost immediately. This approach might not work for everyone, but it will work for many organizations.

- Watch blogs, forums, mailing lists, and vendor notifications for discussions about conflicts with updates for about a week after patches are released. You probably would not see any issues and those you do see will be quickly resolved by the crowd-sourced expertise and by the experience of thousands of IT professionals.
- Start updating systems or groups of systems where any conflicts would cause a minor impact. Update about half of the systems in a physical or a logical area on the first pass so that the impact is not too severe if a conflict or an issue is found.
- If no issues arise, deploy the updates to more systems. Iterate this way until all the systems are updated. Smaller sample sizes with faster iterations are usually preferable to larger deployments at the same time.

This approach reduces your work effort as well as buys credibility when there are few to no issues. So, when an emergency patch comes up, the management will trust that you know what you are doing, which buys you some leeway if any issues come up with these.

To make the patching process as responsive as possible, you should have a preapproved method to make decisions and to act quickly. In many organizations, this consists of having a group (such as a CERT—or others) empowered by the management or the board to handle this entire process. In others, particularly smaller organizations, the key decision maker will be briefed and informed of

their role ahead of time. When an emergency patch situation comes up, then, that person can quickly decide after being briefed of the particular situation. Then the appropriate action can be taken.

> **Pro Tip:** The idea is that right now, you want to get a program up and running effectively and quickly. You will likely know of the top of your head, the relative criticality; so, for now, there is no need to set up a framework for determining criticality, but that will come later. There are plenty of models you can use; my advice is to try some of them and see which model works best for you, then adapt them over time to exactly meet your needs.

Another area where organizations either overspend or think they need to is tools to deploy patches. Many of these tools are costly to obtain, but also cost a lot to maintain and to use properly. Instead, for most vendors, patching can be automated and simplified with free tools. Microsoft, for example, has a tool called Windows Server Update Services (WSUS), which allows organizations to patch Microsoft and some third-party software. Adobe* and Java,† two of the most frequently patched software packages, have guides for IT administrators on rolling out updates across the organization in a way that is easy and unobtrusive.

Vulnerability Awareness

The goal of vulnerability awareness is to make more informed risk decisions and to prioritize these better. The biggest waste of money in a security organization often comes as a direct result of poor vulnerability awareness. The two things that kill most organizations' budgets are false positives and not understanding the context of the vulnerabilities they do have. These lead organizations to overspend and spend in the wrong areas. The critical components for keeping costs down then will be knowing your vulnerabilities, as well as knowing how vulnerabilities, threats, and patches will affect you. That information will then feed into your decision-making process, allowing you to make the best use of your resources.

Tracking all this can be time consuming and intimidating, especially for someone without a strong technical background. And it can be very expensive, costing hundreds of thousands of dollars to go with the top-of-the-line tools, and at least that much in resources to manage the process. Fortunately, a little bit of knowledge and preparation can cut the time and money to a fraction.

Vulnerability Scanning

Vulnerability scanning is the process of performing active reconnaissance on your network to discover security issues. This is usually done either by probing systems over the network for known vulnerabilities or by logging in and analyzing the configuration. The types of vulnerabilities that can be identified using these scanners are large and include missing patches, weak passwords, devices that have not been hardened, default configurations, and more. You can even use them to measure each system against a baseline configuration to determine policy compliance.

* http://www.adobe.com/devnet/flashplayer/articles/flash_player_admin_guide.html
† http://java.com/en/download/help/silent_install.xml

Pro Tip: Avoid the biggest problems by running scans in authenticated mode or in credentialed mode. This reduces false positives, meaning you are not chasing vulnerabilities that are not real. This reduces false positives on Windows machines nearly completely. And it also reduces the risk of having the systems you are scanning that become unstable or sluggish. I recommend setting up a specific domain account for scanning, using a strong password, only allowing it to login from the scanning system and disabling it when not in use.

Most large organizations will already have vulnerability scanning programs. If you do not, most of the tools have a free trial period for testing. When using these tools, the safest, fastest, and most accurate data will be gained by configuring the scanner to authenticate to each system, download configuration and other information, and report on what it finds. Microsoft's Baseline System Analyzer is excellent for Windows systems. And Tenable Nessus is a great low-cost all-purpose scanner.

Cautionary Tales: First and foremost before you start performing these types of network scans make sure you have any permission you need. If not it could get you in trouble or even fired. And before you start scanning, make sure you coordinate with the appropriate groups. Some stories from my own experience:

- Consider support resource availability. When one legacy device became sluggish during an overnight scan the support call had to go all the way to the Network and Server Architect to fix. He was understandably grumpy on Monday morning.
- Consider worst case scenarios. In performing a scan on a machine that was 15 years out of date, the system crashed hard and took 2 days to repair. Fortunately the precautions put in place isolated the system and its backup continued to function normally.
- Consider bottlenecks. A vulnerability scan on one system generated so many log events that it saturated a slow network connection. This bottleneck was found as an unintentional side-effect of system testing.

Also keep in mind that vulnerability scanners don't catch everything. In several assessments, I've penetrated networks by looking up default passwords for VPNs. These were companies who regularly performed internal scans and had third-party penetration testing. Scanners don't know everything, and one thing they do poorly is guess passwords. Similarly, things like layer 2 networking flaws and passwords stored on FTP servers are not usually detected. It's important to understand those gaps and compensate by manually building password lists, inspecting FTP sites or shared folders for sensitive information and looking for layer 2 issues, for example.

Threat and Vulnerability Intelligence

New vulnerabilities and threats emerge daily. Keeping on top of all the information is difficult, but can be done inexpensively. Vulnerability announcements keep you up to date on what systems within your environment may have security issues you have not previously identified. The information about the latest threats allows you to put your vulnerabilities into context and prioritize.

The targeted subscriptions are often very expensive, but there are alternatives. Many companies offer commercial vulnerability feeds, patch release information, and threat intelligence updates. These are typically customized specifically for your systems and your thresholds, but can run into hundreds of thousands of dollars.

Free alternatives exist, but require some up-front setup on your part. Organizations such as the U.S. CERT* and the Internet Storm Center (ISC)* offer RSS feeds and mailing lists. And most vendors themselves have alerting mechanisms for when new threats, vulnerabilities, and patches are announced. You can use your e-mail program or other software to track the up-to-the-minute alerts. And you can build filters, you can review only the information that applies to you, as well as route alerts to team members who oversee certain technologies or platforms. These feeds are not nearly as in-depth as the commercial threat intelligence feeds, but they save quite a lot of money.

Analysis

Now, it is time to analyze and prioritize issues based on our information. Analyze vulnerabilities in the context of the threats. If this is your first time performing this type of work, the biggest risks will likely jump off the page at you. Even if not, looking for the lowest hanging fruit will give you the best use of your security resources.

- *Is there any way in from the outside?* Systems that are accessible from the Internet represent a high risk, since attackers can use these to gain a foothold in your network. Also, consider systems that are accessible by business associates or partners in this category.
- *What are the most severe vulnerabilities?* Look for findings that have "remote code execution"† possibilities, a high common vulnerability scoring system (CVSS) score,‡ and easily acquired exploits.§
- *What systems would have the greatest impact?* Look at findings from your most critical business systems, those which have the most sensitive data and those which could give attackers access to the aforementioned systems. Infrastructure systems often contain the "keys to the kingdom" for attackers.

You will also want to look at broader issues and patterns than just the tactical, technical findings. The first priority will be the really critical findings from your initial analysis, but beyond that, you want to make sure that these things do not keep coming up. Look for patterns that

---

* But with that said, this is your security program. I would not have to live with the consequences and I would not be by your side to help you out. So, it is important to make the philosophy work for you. In the end, results are your measuring stick, rather than following a particular dictum.

† This simply means that an attacker can gain access to the system over the network.

‡ CVSS scores range from 0 to 10 and confer severity and exploitability of the vulnerability. These are listed on most vulnerability findings. http://www.first.org/cvss.

§ To find these, you can search for the vulnerability name and "exploit" on Google. Or use the Metasploit lookup tool http://www.metasploit.com/modules/.

expose the root causes and you can likely find simple solutions to them. Here are some of the patterns most organizations I have assessed tend to fall into:

- Only a few outliers—mostly servers. You are doing a good job overall. The usual line is that these systems are going away soon, cannot be patched for one reason or another, and so on. It is best to take some other indirect measure, such as turning the server off when not in use, segregating it from the rest of the network, or protecting it with a firewall.
- Most servers have issues, workstations have few. You are doing a good job on the workstations and you likely have standardized hardened builds and regular patch cycles. You may be able to shift some time from workstation maintenance to cleaning up the servers.
- Workstations have issues, mainly with third-party software. You are doing a good job on patching and standardized hardened builds with the workstations. But third-party software is still a problem. Put some more resources here to update the most widely distributed, most critical issues.
- Many unknown and nonstandard systems show up. These are probably personal devices, vendor-managed devices, and others that you may not know were there. This can be a big problem and you should see what your policies say with regard to vendors, consultants, and contractors securing their systems, as well as employees connecting their laptops or mobile devices.

Security Awareness

An effective security awareness program is a force multiplier. Educating your employees on security means empowering them to make decisions as well as giving the tools to make the right decisions. This makes your job a lot easier because you do not have to think of everything. And you are also extending your eyes and ears out into the organization, to have others identify security problems for you. Not only that, but you can build security champions within the organizations who can be the local gurus for each group to take some of the load off you. So, shifting some resources here from other areas can significantly cut your costs, giving you a much more effective security program overall.

No wait! Do not skip over this section just yet. I know you have heard everyone say that this is a key area that needs work in most organizations, but I am not going to give you the same old advice you have heard before. I am as tired as you are of hearing about this problem, but I am not hearing any good advice. But the truth is pound for pound increasing awareness is the cheapest and most effective way to improve security for every organization I have worked with!

Virtually every security professional says that security education is the biggest area for improvement in most organizations. The usual feedback is that the company already does training or spends a lot on computer-based training systems. They cannot tell if it is effective, but suspect it is not and feel like they are already doing the best they can think of. There is a difference between training and education in this context—one should be an action, the other should be a result.

To say it in a different way, your goal should be education and awareness, not training—an effective result, rather than just more action. To achieve that goal, you have to get your employees to internalize the information, meaning they understand the concepts well enough to act appropriately in new situations. You cannot accomplish this goal by simply doing training, dropping notes in the newsletter, and putting up posters. You have to adopt multiple learning and teaching modalities.

Learning and teaching styles vary among people; each person has their own strength. The key to better security awareness is to reach each person in the way that works best for them. But it also means using the ways you are most effective and developing the areas where you are not. And you can help your team develop as well. If you are not a great public speaker, for example, performing training may not be the best way for you to get your point across. But if someone else in your group enjoys it, they may jump at the opportunity to do it, leaving you free to use your most effective style.

Here are some suggestions on the different ways to engage and hopefully educate:

- *Group lunches:* Everybody likes lunch! Informal "lunch and learn" or "brown bag" sessions are a great way to get people together to discuss security. If you can do these once or twice a week for different groups, you will start identifying regulars who may be great champions within the groups. If you can provide lunch or desserts, that will raise your attendance.
- *Bring in outside pros:* The local security people are always looking for continuing professional education credits (known as CPEs), which can be presenting or talking with a group about security. Get in touch with the local groups[*] and offer these, plus a free lunch or a similar goodwill gesture and you are likely to get several people taking you up on the offer. Especially in Q4 when they are due for the year.
- *Information on staying safe at home:* Giving employees information—whitepapers, links, training, and so on—on how to stay safe online when they are at home is always well attended. Especially if you talk about keeping kids safe. And the information is much more likely to be absorbed this way because it is relevant. There are some great resources available on the Internet about this.[†]
- *Meet with business unit leaders to talk about security topics in their area:* This is easy to do over lunch or dinner, not just in the office. Showing the business leaders you are thinking about their problems and already looking at how to solve them can make them more likely to see and understand the problems you are talking about.
- *Posters and newsletter updates:* These are fairly basic but still effective ways of getting the word out. Posters should be eye catching and should be changed regularly. Having a regular column in a newsletter is great, but make sure it is interesting. I would avoid tips because they seem self-serving, unless they are in a piece about staying safe at home that make them relevant.
- *Use anecdotes and metaphors:* Even if the story is atypical or the metaphor is inexact, these things actually get the point across. One I like to use is the story of Ali Baba, who overhears the secret phrase to the cave of wonders. It is something we all learn as a kid and teaches us about password security.
- *Security mentoring or tutoring:* Make the offer, at least, even if no one takes you up on it. If someone is interested enough about security to seek out a one-on-one session, find a way to give it to them. This can help you identify and develop the talent or champions within an organization.

---

[*] You can simply search the Internet for groups such as local Defcon chapters, ISACA, SANS Mentoring, ISC2, Infraguard, CitySec, and others in your area. There is usually a forum or mailing list you can subscribe and post to.

[†] A few are: Stop. Think. Connect. http://www.stopthinkconnect.org/National Cyber Security Alliance. http://www.staysafeonline.info/in-the-classroomFederal Trade Commission. http://www.ftc.gov/bcp/edu/microsites/idtheft U.S.-CERT. http://www.us-cert.gov/home-and-business/

- *Provide a learning environment:* For the IT staff as well as for others, allow them to set up a learning environment with the old equipment. There are videos online teaching the basic IT and security skills. This can be useful to teach employees about computers as well as security. And it can be an outlet for people who are naturally interested and may try learning on their own. Instead, you are building a new talent internally to your organization. This is how I got started in security; so, it is a topic near to my heart.
- *Ride-alongs:* Giving employees the opportunity to see what your IT or security program looks like from the inside is a great way to foster cooperation and understanding. And by the same token, allowing IT and security staff to see a day in the life of the employees they support can be very valuable. You will likely find that your solutions are better in line with business and are more palatable to them after these have been going on for a while.

Security awareness is also an area where it is easy to demonstrate your effectiveness! It is pretty easy to build metrics around security awareness through regular testing. And it is also easy to get some positive anecdotes where people exceeded expectations! That means you can track your improvement and get visible wins within the organization.

Testing security awareness program effectiveness can take a couple of formats. The first and the most often done is written testing—usually an online multiple-choice test after a computer-based learning session. The results are usually used for compliance, but you can use them to show the effectiveness of a revamped program. The other way to test is to do live-fire security awareness testing through social engineering.

Pro Tip: Social engineering is often referred to as testing the human firewall. That is, simulating an unauthorized person trying to get access or information that they should not have. This is done through phishing e-mails, pretext phone calls, or simply by walking in the front door. It is a method commonly used by real attackers because they are so often successful. The most effective way for an attacker to hack into your organization is to get an employee to do it for them! Even the strongest technical controls are no match for a clever employee.

Even easy-to-spot scams can lead employees to giving out too much information. When performing these tests, a simple e-mail asking people to log into a foreign system with their credentials had about a 30% success rate. Promising an iPhone to do so raises that rate to nearly 100%! In my experience, the most effective way of preventing these attacks and alerting the security team quickly is by regularly testing employees. On most failed attempts to a social engineer, an employee they tell me that they do not want to fall for a simulation—not that they think it is a security risk!

You can perform social engineering yourself or you can have someone for external performance testing. The sites such as phishme.com allow you to simulate phishing attacks against your organization. The internal employees can also perform some of these tests by sending e-mails from outside or making phone calls. And you can have the external security professionals come in and test to raise the bar. It is so cheap, easy, and effective that I would recommend testing and reporting monthly.

Pro Tip: Social engineering testing is about improving your program. Use the feedback to first set a baseline, then improve both the program and internal operational processes over time. You can also identify highly vulnerable areas or methods of attack so that you can implement more targeted security controls, saving money and effort.

Social engineering is not about identifying people who break the policy. When reporting results, it is usually best to share the results anonymously or statistically. If, over several months, trends start to emerge within certain groups or with individuals, call those to the attention of the right people and work to solve the issue and not place the blame. Otherwise, the testing is likely to be seen as the problem, not the weakness.

Review and Strengthen Password Security

A large percentage of compromises happen because of bad password security.* And that is really too bad because there are some cheap and easy things you can do to fix that. In most organizations, some very simple steps can make a huge difference.

One of the biggest issues is default accounts and passwords, particularly on the perimeter network. In my time doing penetration testing, I compromised several large organizations this way. On large enterprise systems such as Oracle and SAP, default credentials are very common, even if you have had professional consultants to implement them. Build your hardening procedures to take into account default credentials, hold the internal staff and contractors to these, and go behind and check that they have been done.

Pro Tip: Automated scanning tools are great, but have some limitations when looking at default passwords. But there are ways around each of these limitations.

- The built-in lists have only about a dozen default accounts. That will cover about half of the systems you will end up testing. To cover the other half, you can download lists from the Internet. Or, you can manually search for default passwords for your systems and use these.
- Automated scanners are not able to test all systems this way. So some—especially web applications. These are also often missed by independent vulnerability assessments or penetration tests, but not by most hackers. Test these manually.
- Some systems will automatically lock out your scanner when running these types of tests. Instead of scanning, look up the default credentials and test these manually.

* According to the Verizon Data Breach Report, 2012, nearly all breaches involve some form of password guessing or using stolen passwords.

Most organizations by now have implemented password strength requirements on most systems. These should match up with the defined policy and procedures. But make sure you have no outliers—there are usually a few systems that do not enforce strong passwords or do not match with the standard procedures. It will help with your security as well as auditability to maintain a list of policy/procedures and enforced requirements for each system.

Sometimes, there are also accounts that have weak or easily guessable passwords. These are usually passwords that adhere to the requirements only barely. For example, "Password1" or "!passworD" or similar. Sometimes, the accounts are IT administrative, system, or vendor accounts.

Fortunately, tools and techniques exist to check the password strength.* Password-cracking tools, such as Cain & Abel, John the Ripper, and Ophcrack, allow you to easily break weak passwords. These tools are free and are often used by attackers. Another technique is to track password hashes across accounts and over time. If you find identical hashes, you know the password is the same. That will let you check for weak or reused passwords but without having to know what the actual password is.

Password reuse is the practice of using the same password on multiple systems or quickly changing back to an old password after a forced change. It is very common and very risky. In a recent meeting with some executives, one of my colleagues discussed a high-profile breach of a social-networking service. The colleague mentioned that lots of people were now having to change their bank passwords, since they were the same. The meeting was paused while several of the executives went and did just that. To solve this issue in your organization, you should require frequent password rotation and disallow using the previous passwords.

Pro Tip: Password vaults are software packages that allow for the safe storage of passwords on a computer. I strongly recommend that organizations must consider them. That is because your employees, IT staff, and vendors can use these tools to use strong, unique passwords without having to remember them all or write them down. This can also reduce support costs from locked-out accounts and resetting passwords.

One free tool is called KeePass. While it is not as robust as some of the other commercial tools, it will allow you to do things such as set the password strength, tie the database to Windows accounts, create strong passwords, and other features to help with security, usability, and management.

Final Thoughts and Conclusion

Hopefully, you have obtained good information from this chapter. Maybe from a new approach to try, a new tool you did not know about, and a technique that will work well in your organization or whatever. My goal here has been to give you a way to think differently about using your resources, not just to give you specific ideas.

So, I will give you one last piece of advice that is nonintuitive: You can use your budget constraints as a tool to your advantage.

---

* As stated above, always obtain permission to use any hacking tools before you download them. Simply possessing some of these tools can get you fired from most companies, no matter what the reason.

- *Force upgrades:* Work with business units that want to update their systems and help them to build a case that it is cheaper to upgrade than to not. Older systems need more security and compliance workarounds and often, these exceed the cost of the upgrade or impact the business to the point where lost revenue or the margin would be too high. Both you and the business unit win.
- *Allow more tightening:* You can make the argument that because you do not have money for tools or people, you need to tighten things down more. Many of the tools in the market compensate for security deficiencies in areas such as firewalls, desktops, servers, and web applications. If the choices are poor security, more spending, or tightening down controls, you may actually get political support for more restrictions as long as they do not impact business. At the very least, you raise the profile of the issue and get others involved in the risk decisions.
- *Invest in savings:* If you can show that investing in security gives savings, you may be able to get the support for nontraditional ideas. At one employer, we were paying for a company to transport and dispose off old workstations and servers. We proposed building an internal program to do this that would reduce the risk of data loss and would be less expensive. We ended up saving even more money than we thought, as we were able to internally repurpose several pieces of hardware for testing laboratories and nonessential areas!

Reduced budgets do not have to mean reduced security; it just means you have to change the way you think about doing things. Several top companies have started during downturns, and many companies manage to improve in lean years. By building sound principles and practices during these times, you can make the recovery even better than it would have been otherwise.

Chapter 8

Organizational Behavior (Including Institutions) Can Cultivate Your Information Security Program

Robert K. Pittman, Jr.

Contents

Introduction

Throughout life, many of us have commuted to and from our place of work, traveled during vacation to other states and perhaps other countries, spent quality time with our family, as well as found time for entertainment at a sporting event, play, beach, or concert. Many of these trips and activities place you in an environment where people exist. People are unavoidable regardless of where your commute or to which place the travels take you. In an assembly of people, there exists a culture that requires an in-depth knowledge and insight to ascertain how you approach those individuals. To identify that, an appropriate approach is warranted, but it is challenging when

establishing or cultivating an information security program, regardless of what sector (e.g., public, private, and nonprofit) you are employed into.

Many well-known theorists like Douglas McGregor (created the Theory X and Y model), Edgar Schein (created the Organization Culture model), as well as psychologist Abraham Maslow (created the Hierarchy of Needs five-level model), and numerous others in the field of organizational behavior and culture research have brought this topic to the forefront because of their immense research and the value it brings to organizations worldwide.

The worldwide sprawl of organizations use government businesses essentially supported by the public. The public comprises its citizens, its constituents, and its businesses such as nonprofit organizations and corporations, including government agencies at all levels. A relationship among everyone involves the citizens with regard to the government and associated organizations interrelating their programs and services on behalf of the public. At least, this is one of the goals of government, since it provides the services that a corporate business would not even consider. The plethora of services being provided to the public are social services, general governance, health care, and public safety.

Some of the countless government social service programs include the addressing and supporting of low-income families, foster care, emancipated youths, and general relief payments for food and housing for the disadvantaged. Other services consist of property value assessment, property tax payment, requests for a birth certificate, marriage license, or a death certificate, as well as simply registering to vote, which, in part, constitutes general government services. Throughout their lives, citizens will require health-care services. Medical and mental health care, including public health issues, will always be of the highest concern to all levels of government for all ages. It may seem obvious that public safety services are at the top of the list with health care as well. The security of our homeland, borders and ports protection, law enforcement, and protecting our loved ones are the areas where the government visibly plays a significant role.

All these aforementioned government services are provisioned externally to the citizens. The perspective on services provided internally would be contrary to corporate services, in terms of the existence and loyalty of a significant amount of an employee's labor unions (i.e., Civil Service Rules), attractive sustained retirement packages, consistent health and dental benefits, career and job advancements within the same government level where opportunities exist at different departments, branches, and agencies, and are knowingly supporting a cause or the application of Greek philosophy using Aristotle's definition of greater good.

Regardless of what lens we use to view the government and the corporate, obvious differences do exist. Much of the government services provided to the public are unique as compared to the corporate-provided services. These unique government services (e.g., social work, probation officer, and librarian) are predominately performed by government employees who are usually passionate about their work, including a strong will to deliver a good service. However, later in this chapter, you will become acquainted with the lack of similarities from the perspective of institutions.

By viewing the government at the 80,000 foot level and viewing through the looking glass, differences exist from an employee and an organization perspective. The differences exist between the local government (i.e., county and city), organizations, and corporations (e.g., corporate stock shares that are well compensated). Obviously, corporate stock shares and job security are some of the differences. Therefore, the establishment of an information security program differs significantly between the local governments and corporations.

Establishing an information security program in the local government involves an array of focal points that must be addressed during the initial 6–18 months by the chief information security officer (CISO), chief security officer (CSO), or information security manager (ISM).

In some recent information security forums and industry writings, the term chief risk officer (CRO) may have a significant role as well. It is imperative that these focal points are addressed in terms of having them established and adopted by the organization:

- Enterprise information security policies
- Information security steering committee
- Enterprise information security program
- Enterprise information security strategy
- Organization health posture based on an information security risk assessment
- Enterprise and departmental (or agencies) computer emergency response teams (CERTs)
- Enterprise security engineering teams (SET)

Each of the above focal points can be categorized as your "Lucky 7." Throughout this chapter, these points will be referred as Lucky 7. The security professional who addresses these points will be "lucky," and the others will not be as "lucky" in terms of continued employment with that particular organization, since the primary responsibility exists in the information security unit. This may sound harsh. However, at the end of the day, the organization's business and services being provided to the citizens and constituents have the expectation that their confidential, sensitive, and personally identifiable information (PII) are secured and protected. It is the job of the information security professional to accept the challenge and responsibility to insure that the organization stays away from any press or media release announcing a data breach, or perhaps, a breach of trust. As information security practitioners are aware, there has been a plethora of announcements in the press and media on organizations (public and private sectors) that experienced computer security breaches. These are in corporate America, colleges and universities, health-care organizations, as well as the 26 million veterans' records with PII that was the responsibility of the federal government Veteran's Administration (public sector) and T.J. Maxx's 45.7 million credit and debit card owners (private sector) that occurred in 2005. However, the all-time record breach occurred 4 years later during 2009 with Heartland Payment Systems that now leads all the hacks which hit or affect the financial services industry (private sector) with 130 million credit and debit card account numbers.

Organizational Governance

It seems more apparent that the public sector leverages a security-related event to promote an information security program or, at the minimum, obtain a funding source to support a project or an initiative. Despite the consequences of failure or compromise, security governance is still a muddle. It is poorly understood and ill defined and, therefore, means different things to different people. Essentially, security governance is a subset of enterprise or corporate governance. Moreover, one could identify governance as security responsibilities and practices, strategies and objectives for security, risk assessment and management, resource management for security, and compliance with legislation, regulations, security policies, and rules.

Information security governance is "the establishment and maintenance of the control environment to manage the risks relating to the confidentiality, integrity, and availability of information and its supporting processes and systems."

From a local government perspective, in terms of county government that is governed by a five-member board of supervisors and the chief executive officer (CEO), the CISO, departmental

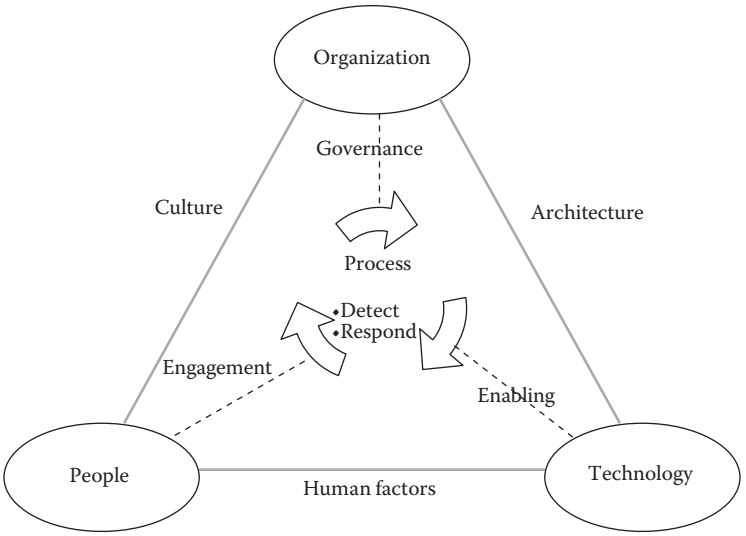


Figure 8.1 Information security strategic framework.

information security officers (DISOs), and the Information Security Steering Committee (ISSC) or a security council comprise the information security governance.

A federated organizational structure is the norm for the majority of the local government organizations. If the discussion is about the county or city government, numerous business units or departments serve unique and differing business purposes. Because of these unique business units, comprehensible governance is vital to the success of any information security program. This governance involves a strategic organizational framework (Figure 8.1) that provides a clear illustration of the involved players. The "Security Triangle" or Figure 8.2 is a framework that is doable for the CISO organization regardless of whether their information technology (IT) is decentralized, centralized, or a managed-security service. Additionally, a local government organization can be deemed as an organization with 30 or more corporations, in terms of having 30+ county departments with distinct businesses as they serve their respective constituents.

The organization's senior management must support the Security Triangle; however, articulation of this support should be achieved by the development of board-adopted policies. These policies are similar to the corporate world where the board of directors and CEO can adopt policies. However, an information security council or steering committee can approve information standards and procedures. The use of an advisory council or a committee provides a weaker connotation that is not supportive of establishing or sustaining an information security program.

Organizational Culture and Behavior

Bruce Schneier is an internationally renowned security technologist and author, as well as the go-to security expert for business leaders and policy makers. Currently, he is the chief security technology officer for BT-Managed Security Solutions. In his book *Beyond Fear, Thinking Sensibly about Security in an Uncertain World*, he explains that security is all about people: not only the people who attack systems but also the people who defend these systems. If we are to have any

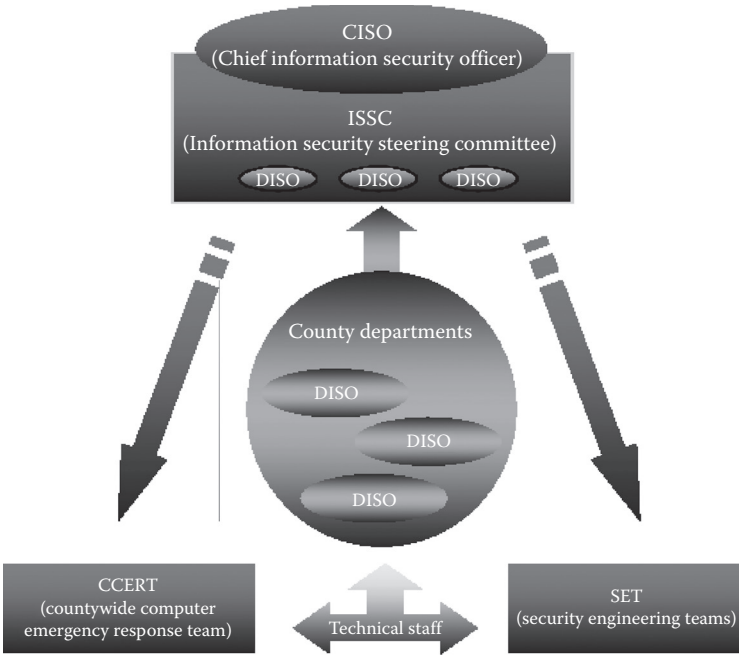


Figure 8.2 Information security strategic organization, the "Security Triangle."

hope of making security work, we need to understand these people and their motivations. We have already discussed attackers; now, we have to discuss defenders.

Schneier also states that good security has people in charge. People are resilient. People can improvise. People can be creative. People can develop on-the-spot solutions. People are the strongest point in a security process. When a security system succeeds in the face of a new, coordinated, or devastating attack, it is usually due to the efforts of people. (See the section: "Organizations Are Institutions" for more detail.)

If it was not obvious prior to reading this chapter, it should be obvious now that people play a significant and critical role as part of any information security program. Moreover, the same people at times can bring about challenges as well. However, the culmination of people defines organizational behavior and its culture. Organizational culture is the culture that exists in an organization, something akin to a societal culture. It is composed of many intangible phenomena, such as values, beliefs, assumptions, perceptions, behavioral norms, artifacts, and patterns of behavior. The unseen and unobserved force is always behind the organizational activities that can be seen and observed. Organizational culture is a social energy that moves people to act. "Culture is to the organization what personality is to the individual—a hidden, yet unifying, theme that provides meaning, direction, and mobilization."

Organizations are assumed to be rational–utilitarian institutions whose primary purpose is the accomplishment of established goals (i.e., information security strategy and initiatives). People in positions of formal authority set goals. The personal preferences of organization employees are restrained by systems of formal rules (e.g., policies, standards, and procedures), authority, and norms of rational behavior.

These patterns of assumptions continue to exist and influence behaviors in an organization because they repeatedly have led people to make decisions that "worked in the past." With

Table 8.1 Public Sector versus Private Sector and Corporate Organizations

Public Sector	*Private Sector*	*Public Corporation*
Director	Owner	Board of directors
Deputy director/branch manager	Vice president	Executive management
Division chief	Manager	Middle management
Section manager	Manager	Supervisory management
Associate	Employees	Employees

repeated use, the assumptions slowly drop out of people's consciousness, but continue to influence organizational decisions and behaviors even when the environment changes and different decisions are needed. They become the underlying, unquestioned, but largely forgotten reasons for "the way we do things here"—even when the ways may no longer be appropriate. They are so basic, ingrained, and completely accepted that no one thinks about or remembers them. (See the section: "Organizations Are Institutions" for more detail.)

In the public sector (Table 8.1), it seems that almost every employee has worked at least 20 years or more. In retrospect, they may have only worked many years less. Reality sits in when many employees consistently echo the aforementioned phase, "this is the way we do things here." As the CISO, attempting to implement one of your many information security initiatives, this echo seems to sound loudly increasing exponentially with the employees that will be affected by implementing a change to their environment. This type of behavior illustrates the presence of a strong organizational culture.

A strong organizational culture can control organizational behavior. For example, an organizational culture can block an organization from making changes that are needed to adapt to new information technologies. From the organizational culture perspective, systems of formal rules, authority, and norms of rational behavior do not restrain the personal references of organization employees. Instead, they are controlled by cultural norms, values, beliefs, and assumptions. To understand or predict how an organization will behave under varying circumstances, one must know and understand the organization's patterns of basic assumptions—its organizational culture.

Organizational cultures differ for several reasons. Some organizational cultures are more distinctive than the others. Some organizations have strong, unified, and pervasive cultures, whereas others have weaker or less-pervasive; some cultures are quite pervasive, whereas others may have many subcultures existing in different functional or geographical areas.

In contrast, there are some "prescriptive aphorisms" or "specific considerations in changing organizational cultures," when this occurs, your information security program (i.e., Lucky 7) along with its processes and practices will flourish with positive outcomes:

- Capitalize on propitious moments
- Combine caution with optimism
- Understand resistance to culture change
- Change many elements, but maintain some continuity and synergy
- Recognize the importance of a planned implementation
- Select, modify, and create appropriate cultural forms

Table 8.2 Stakeholder-Desired Behavior

Organizational Target Group	*Desired Behavior*
Board of supervisors/city council	Endorsement
Executive management	Priority
Middle management	Resources
Supervisory management	Support
Employees	Diligence
Constituents/consumers	Trust
Security program	Execution

- Modify socialization tactics
- Locate and cultivate innovative leadership

Altered organizational culture is merely the first—but essential—step in reshaping organizations to become more flexible, responsive, and customer driven. Changing an organizational culture is not a task to be undertaken lightly, but can be achieved over time.

Organizational cultures are just one of the major tenets that constrain the establishment and building of an information security program. As a CISO or an information security practitioner within your organization, you should have at least some interaction with the various target groups (i.e., stakeholders) to grasp an awareness of their behavior. Table 8.2 illustrates the expected behavior of the target groups and the desired behavior to assist in driving your program. This chart will provide benefits when you assess each stakeholder from your perspective.

Organizational Culture and Behavior: Millennials Generation

This is the twenty-first century where organizational culture and behavior must not be discussed without the inclusion of the Millennials generation. A generation is a group of people who share common experiences and a common collective memory based on key events that occurred during their lifetime. The collective memories of a generation lead to a set of common beliefs, values, and expectations that are unique to that generation. A generation does not grow up to become like their parents, but instead continues through their lifetime with a separate and distinct set of beliefs and expectations formed by the shared experiences.[1]

The Millennials is the generation born during 1981 and after is also known as Generation Y, Nintendo Generation, Generation 2001, Net(X)t, Nexters, or Net Generation.[1] Therefore, if you are 31 years of age and younger (age in the year 2012), you are part of the Millennials generation.

Today's workforce is commonly thought of consisting four (4) distinct generations, as illustrated in Table 8.3. The end dates of each generation vary slightly among different researchers. However, these generations comprise our challenges in sustaining and cultivating an information security program based on these generational differences since the workforce is a critical element in mitigating the risk. For example, your information security program spent enormous amount of

Table 8.3 Generational Descriptions

Generation	*Birth Years*[a]	*Ages in 2012*	*Shared Life Experiences*
Traditionalists	1922–1945	67–90	Great Depression, New Deal, World War II, and the GI Bill.
Baby Boomers	1946–1964	48–66	Great Society, general economic prosperity, expansion of suburbia, VietNam, President Nixon, sex, drugs, and rock 'n roll.
Generation X	1965–1980	32–47	Divorce, lack of faith in institutions, AIDS, Sesame Street, MTV, crack cocaine, Game Boy, and the PC.
Millennials	1981 and after	31 years and younger	Internet, always connected, September 11th, and the War on Terror.

Source: Adapted from Ron Zemke, Claire Raines, and Bob Filipczak, 2000. *Generations at Work: Managing the Clash of Veterans, Boomers, Xers, and Nexters in Your Workplace*. AMACOM Books: New York.

[a] Range of birth years may vary slightly depending on the information source and demographic model.

time developing technical standards specific to Microsoft Windows operating systems and applications, perhaps, using the Security Triangle (see Figure 8.2) and the information security strategic framework (see Figure 8.1)—where little or no time was spent on security requirements for Mac-operating systems.

It is my belief, that the Millennials seduced organizations and institutions into the Bring Your Own Device (BYOD) and social media movement that has gained significant and strong attention by all organizations' senior management that specifically includes chief information officers (CIOs), CISOs, and CSOs.

Like the majority of human behavior models, the generational model above provides a way of thinking about people, but does not imply the mutually exclusive categories. Not all people behave in the way that their common generational experiences would lead us to expect. For example, not all Millennials speak in technical jargon and not all Traditionalists are clear communicators.[1]

Some of the exceptions to the model are explained by the crossover effect and the end dates of the generation. The crossover effect occurs when some event is so important that it affects more than one generation, so multiple generations share that common experience. For example, the tragedy of 9/11 impacted all workplace generations and is part of the shared memory. Another source of discrepancy is the difficulty of setting the exact dates for the ending of one generation and the beginning of a new generation, as shown by the different dates used by various researchers when describing the four distinct generations.[1]

These shared experiences and collective memories of each generation lead them to have a common set of core values, beliefs, and behaviors, which in turn affect their expectations in the workplace. These beliefs and values stay with them for life, offering challenges to manag four distinct groups within the workforce. Some of the generational differences in values are shown in Table 8.4.

Table 8.4 Generational Core Values and Expectations/Beliefs

Generation	*Core Values*	*Expectations/Beliefs*
Traditionalists	Dedication/sacrifice Hard work Conformity Law and order Respect for authority Delayed reward Duty before pleasure Adherence to rules	Security from the institution Promotions based on longevity Loyalty to the organization Wait to be told what to do Respect based on position/title
Baby Boomers	Optimism Team oriented Personal gratification Health and wellness Personal growth Youth Work Involvement	Live to work (70–80 h work week) Optimistic and driven Relentless pursuit of goals Change is painful but inevitable Rule the workforce, created the culture Prefer a casual, team-oriented environment
Generation X	Diversity Thinking globally Balance Techno literacy Fun Informality Self-reliance Pragmatism	Work to live (work/life balance required) Security from options versus commitment Promotions based on ability versus seniority Mobility versus stability Computers are part of everyday routine Attitude/fun important in the work place Expect a say and want to be heard
Millennials	Optimism Civic duty Confidence Achievement Social ability Morality Street smarts Diversity	Team oriented (socialize/work in groups) Hard working Multi-taskers Prefer structured environments Acknowledge and respect position/title Want a relationship with their boss Relate best with the Traditionalists

Source: Adapted from Ron Zemke, Claire Raines, and Bob Filipczak, 2000. *Generations at Work: Managing the Clash of Veterans, Boomers, Xers, and Nexters in Your Workplace*. AMACOM Books: New York.

Given all these differences in the generations, the key question is how to manage your information security program where governance has been established, as previously discussed in this chapter.

Somewhat unrelated, but related, some companies have set up reverse mentoring programs. For example, Deloitte and Touche L.L.P. matched the executive Baby Boomers with younger consultants. The executives have learned technical skills (e.g., Internet and email), but even more importantly have "learned valuable insights into how both younger staffers and the marketplace perceived the firm."[2] Jack Welch, the former General Electric Co. Chairman and CEO, told his senior managers to actively look for relationships with younger employees to improve company performance.[2] These types of programs or techniques have shown the value of treating generational differences as a competitive advantage to enhance creativity rather than as a source of conflict and misunderstanding.[1]

For your information security program, CISOs, CSOs, and ISMs must pay particular attention to the generational individuals comingling within your teams and groups. A tremendous return on investment is in terms of the interaction among the various individual differences as related to their specific generation. These differences become sources of strengths rather than sources of conflict.[1]

An awareness of generational differences can help CISOs, CSOs, and ISMs encourage more creative and productive interactions among employees. This generational awareness can also help employees to more effectively work with the generational environment.[1] We must be able to dissect these differences via the awareness tenets:

- Identifies competencies necessary to be successful in each task/job/project
- Recognizes what makes each generation "tick" in the workplace
- Blends the competencies and the generational qualities in a way that inspires, motivates, and leads employees to achieve the universal organization or institutional goals (strategic imperatives)
- Provides rewards to employees who are consistent with each generations' motivations, expectations, and values
- Recognizes generational differences
- Respects differences in generational outlooks

The above tenets play a huge role in sustaining and cultivating your information security program where a greater focus on Millennials is prudent and warranted. In Table 8.4, one of their core values is social ability as well as an expectation oriented toward a team approach that incorporates a form of socialization (e.g., work and focus groups).

Cultural values and norms play a substantial role in the development of an individual's personality and behaviors—that is, socialization. Social factors reflect such things as family life, how their downtime is spent, and the many types of formal and informal groups in which people participate throughout their lives—friendship groups, athletic groups, and formal workgroups. The demands of differing situational factors emphasize or constrain different aspects of an individual's personality.[3] Finally, in today's work environment, the expectations to have access to social media websites (e.g., FaceBook and Twitter) while at their workplace support their belief system or ecosystem; however, it clashes with their organization's policies, norms, behavior, and culture.

Organizations Are Institutions

When discussing organizational culture, it is moot if we are referencing the government or the corporation. To extend the conversation on culture, a CISO must have a clear understanding of

institutions. The term "institution" designates organizations of every kind. To avoid confusion, it is useful to distinguish between institutions and organizations. Institutions are the rules of the game; organizations are corporate actors, that is, groups of individuals bound by some rules designed to achieve a common objective.[4] There can be political organizations such as political parties, educational organizations such as universities, social service organizations such as a government department, or economic organizations such as firms. Therefore, organizations, when interacting with other organizations or individuals, submit to those general social rules called institutions, that is, they are equally constrained by the general rules of the game.

Essentially, institutions are simply normative social rules or rules of the game in a society, enforced through the coercive power of the state or other enforcement agencies that shape human interaction.[5] This interaction is vital to establishing and sustaining your information security program. Institutions exist based on an individualistic approach. The first class of reasons refers to the motivational possibilities of *Homo sapiens* and the second class refers to the cognitive ones. Motivation main assumption is that every individual strives to increase his utility or, in other words, that every individual strives to better his condition by all means available to him. It becomes obvious that the conflicts between individuals are bound to arise. From the perspective of the observer, however, such social problems are clearly identifiable and their basic characteristic is that the utility obtained by some kind of individual behavior depends in one way or another on the behavior of other individuals.

Some stylized social problems have been identified in game theory or the game of trust. When asking the question on why not solve social problems ad hoc because, in a way, every problem situation—and thus every social problem is unique, the response is associated with the cognitive structure of the human mind. The human mind is far from being a perfect tool, able to perform all the difficult computations needed for solving problems that arise from the interaction with other minds. Because of a restricted cognitive capacity, every individual mobilizes his energies only when a "new problem" arises and follows routines when the individual classifies the problem situation as a familiar one. This distinction is rooted in the limited computational capacity of human beings and is a means to free up an individual's mind from unnecessary operations so that the individual can deal more adequately with the problem situations arising in the individual's environment.[5]

The individual's environment could be deemed as complex. This precisely indicates that the individual's limited cognitive capacity makes their environment appear rather complicated for the individual and in need of simplification to be mastered. This refers to both the natural and the social environment of the individual. Because of the perceived complexity of the social environment, people consciously or unconsciously adopt rules as solutions to the social problems rather than deciding each time anew how to act and react to the settings where coordination with other individuals is needed. Rules, in general, "are a device for coping with our constitutional ignorance"; they are the "device we have learned to use because our reason is insufficient to master the full detail of complex reality."[6]

A very productive and very widely used distinction among the types of institutions is based on the criterion of the enforcement agency of institutions. Institutions are commonly classified according to this criterion as shown in Table 8.5.

The most important feature of conventions is their self-policing character. After they have emerged, nobody has an incentive to change rules that everybody else sticks to. In game theory, conventions are usually analyzed with the help of what are known as "coordination games." Examples of such rules are traffic rules (e.g., traffic speed signs are regulatory), industrial standards, forms of economic contracts, language, and so on. The moral rules are largely culture

Table 8.5 Informal and Formal Institutions

	Mechanism	*Character*
Informal Institutions	Conventions	Self-policing
	Moral rules	First party
	Social norms	Third party: Social forces (i.e., individuals of the group)
Formal Institutions	Law	Third party: State

Source: Adapted from Mantzavinos, C. 2001. *Individuals, Institutions, and Markets*. Cambridge University Press, Chapter 6.

independent because they provide solutions to problems that are prevalent in every society, as Lawrence Kohlberg has shown in his famous empirical research.[7] These rules (e.g., policy and standards) are critical to establishing and sustaining your information security program, as well as assist in culture change. The mechanisms for the enforcement of moral rules are entirely internal to the individual, and therefore no external enforcement agency for rule compliance is required. The typical examples of moral rules are "keep promises," "respect other people's property," "tell the truth," and so on. These have a universal character. However, their existence does not necessarily mean that they are also followed, and in fact, many individuals break them. (Thus, the empirical phenomenon to be explained is the existence of moral rules in a society, which are followed by part of the population). On the contrary, social norms are not of a universal character and they are enforced by an enforcement agency external to the agent, usually the other group members. The mechanism of enforcement refers to the approval or disapproval of specific kinds of behavior. Social norms provide solutions to problems of less importance than moral rules and regulate settings appearing mainly at specific times and places.

The enforcement agency (i.e., conventions, moral rules, and social norms) of each different category of the informal institution (Table 8.5), as well as the specific enforcement mechanism (i.e., approval or disapproval of specific kinds of behavior), is different. The common element to each type of informal institution is critical when all emerge as the unintended (i.e., not planned) outcome of human action.[5] It may be obvious that this outcome may be favorable or unfavorable depending on the situation as it relates to the previous described equation.

Their mechanism of emergence is thus an evolutionary process of the invisible hand type. This process starts as an individual perceives his situation as constituting a new problem because the environment has changed, and then in an act of creative choice, the individual tries a new solution to this problem. Both the problem and its solution are of a strictly personal nature, and the solution is attempted because the individual expects it to increase their utility.

Whereas informal institutions emerge from the unintended results of human action in a process that no individual mind can consciously control, law or the sum of the social rules that are called formal institutions are products of collective decisions. The state as an organism* creates law, by constructing the conscious decision of its organ's new legal rules or by providing a means of suitable adaptation—the existing informal rules with sanctions.[8]

* See the famous definition of the modern state of Max Weber, 1919 and 1994, p. 36.

Informal institutions are generated through an invisible-hand process in a way endogenously from within the society. The formal institutions are the outcome of the political process, which is imposed exogenously onto the society from the collective decisions of individuals who profit from resources: political, economic, and ideological. On the other hand, this is reflective of an institution adopting policies and standards to address a baseline for their information security practices for compliance by its employees. There is no necessity that informal and formal institutions complement each other in such a way that a workable social order is produced or even more for the economic development of a society to take place.

The Information Security Executive in the Organization

The late UCLA legendary men's head basketball coach John Wooden, wrote "there is a choice you have to make in everything you do, so keep in mind that in the end, the choice you make makes you." Nowhere is this more evident than the relationships that are established throughout your organization, as well as external to the organization. Surround yourself with people who add value to you and encourage you. At the minimum, having photographs or prints hanging from your office walls of individuals who have achieved greatness, regardless of what the industry, will provide an added psychological benefit when tough decisions must be made. If the opportunity presents itself, where you are able to visit my business office or home office, this psychological benefit is apparent. In plain sight, you will see the memories and fondness of greatness from the first African American athlete to enter the major league baseball—Jackie Robinson, Muhammad Ali (former name Classis Clay)—who changed the culture of boxing to a style, business, and character (e.g., charisma) that was not seen previously, and Louis Armstrong (nicknamed Satchmo)—once proclaimed as the greatest musician to ever live was the first jazz musician to appear on the cover of *Time Magazine* on February 21, 1949. These individuals without any doubt changed their respective institution's culture overnight. The establishment of strong relationships is an excellent indicator of a strong CISO; however, staying visible in the organization is equally important and that will provide the path to have a positive information security culture throughout your organization (e.g., department, agency, and division).

People can trace the successes and failures in their lives to their most significant relationships. Establishing relationships are part of our livelihood in terms of family, personally, professionally, and business. Moreover, as the CISO, when establishing an information security program and chairing an ISSC meeting with your security peers or colleagues in your organization, those relationships are imperative to your success. The effective CISOs have learned how to gain the trust and confidence of the executive team. The CISO must remember that security is easier to sell if the focus is on the benefits of the company. Sometimes, while selling security, analogs associated with personal and home practices will provide clarity and additional reinforcement.

The CISO is the information security executive (i.e., senior management), regardless, if we are referencing public, private (i.e., corporate American), or nonprofit sector organizations. Regardless of what sector, an organization's CISO must address the big picture and must rely on timely and actionable risk information that enhances their ability to make decisions that will drive the local government efficiencies and operational effectiveness.

In the local government, many CISOs are using a matrix reporting structure and either report to the CIO or the CEO, and ultimately to the city manager, board of supervisors (Board), or City Council (Council). Actually, this matrix model can only function in this manner as long as no

operation responsibilities are incorporated. In other words, the daily operational activities and tasks may collide, at the minimum, with the strategic and tactical mindset of the information security practitioner. This model has brought this author numerous successful implementations of information security projects and initiatives, including program sustainability.

However, there are many other ways to organize the security function and they all have advantages and disadvantages. Strong CISOs understand that the organization of the security or the successes is not important. The key to success is the support structure the CISO is able to build among the executive team. However, the manner in which the security is organized will change the methods and processes a CISO will use to be successful. The effective CISOs will adapt their approach to the most advantageous organizational structure. The two most common primary organizational structures are (1) matrix structure, in which the CISO is an enterprise-level (or corporate-level for the private sector) organization and the security staff report in the business lines, and (2) the CISO has direct or indirect (e.g., dotted-line organizational structure model) responsibility for the implementation and operations of security.

Smart CISOs understand that they do not need to have all the security staff in their direct reporting line. Be ready for decentralization. Being a strong CISO is not about how many staff you manage; it is about how many staff you can influence. Drive the difference of security any way you can—through direct staff, matrix staff, and supporting staff—to reach the security program goals and initiatives. Large organizations have already implemented a matrix organization or are seriously reviewing how to manage the business lines more effectively. Be prepared to manage in a matrix organization.

Regardless of the reporting structure, decisions are made to eliminate press clippings in tomorrow's local newspaper or, perhaps, the national news. The CISO cannot be risk adverse. All information security practitioners should think quantitatively. This does not necessarily mean doing calculations. Rather, it means thinking about things in terms of the balance of arguments, the force of each of which depends on some magnitude.

Some local government organizations are forward-thinking companies that have recognized that business and IT executives (e.g., CIO, CISO, or Chief Technology Officer) need to establish standardized, repeatable ways to identify, prioritize, measure, and reduce business and technology risks, both collaboratively and effectively. Moreover, security executives who were accustomed to working in their own silo must now consider all business-related risk areas to align initiatives (e.g., business and applications, system migration projects, and customer-based applications to enhance e-Government/e-Services) properly with exposures.

Collaboration and communication is sunshine on its brightest day. Team relationships and/or team meetings are training gold nuggets. If the opportunity exists, inviting individuals to attend selected meetings within your security program can go a long way to helping them understand the scope and breadth of security. Make them an honorary member of the team. This has been done on several occasions to break through the myopic barrier. In addition, if other groups will let you attend a team meeting or two, go for it. This seems very simple, and is, but can be unbelievably powerful.

It is very true that there is success in numbers from an empirical perspective, where teams can drive your information security program. Two types of teams should be implemented to support an information security program: proactive and reactive.

We call the proactive measure teams the SET. All these teams develop and review policies, standards, procedures, and guidelines. These teams are usually experienced and knowledgeable in terms of the technical, cultural, and organizational perspectives. These teams address host strengthening and isolation, policy and operating procedures, malware defense, and application

security, to name a few. However, there will be opportunities where a proactive team will be formed to address a point-in-time project. For example, our implementation of an Internet content filter was a win-win because of the formulation of an SET from the development of the technical specifications to enterprise deployment. Once deployed throughout the organization, the team was no longer required.

A reactive team addresses an enterprise-wide CERT. This team reacts to situations that potentially affect or have affected the enterprise network, servers, applications, workstations, and so on. This is reactive in nature. However, the use of a structured methodology while responding, resolving, and reporting the incident is vital. The use of a well-maintained and clearly written documentation (e.g., narratives, matrixes, and diagrams) for responding to incidents and using a standardized incident-reporting form is crucial. It may be obvious that by defining the various types of information security, incidents to report will provide one of the numerous performance metrics that can be established to measure a portion of the operational aspects of your program (Table 8.6).

Table 8.6 Stages of Group/Team Evolution

Stage	*Dominant Assumption*	*Socioemotional Focus*
Group/team formation	*Dependence:* "The leader knows what we should do."	*Self-Orientation:* Emotional focus on issues of a. Inclusion b. Power and influence c. Acceptance and intimacy d. Identity and role
Group/team building	*Fusion:* "We are a great group/team; we all like each other."	*Group/Team as Idealized Object:* Emotional focus on harmony, conformity, and search for intimacy Member differences are not valued
Group/team work	*Work:* "We can perform effectively because we know and accept each other."	*Group/Team Mission and Tasks:* The emotional focus is primarily on accomplishment, teamwork, and maintaining the group in good working order Member differences are valued
Group/team maturity	*Maturity:* "We know who we are, what we want, and how to get it. We have been successful, so we must be right."	*Group/Team Survival and Comfort:* The emotional focus is preserving the group/team and its culture Creativity and member differences are seen as threats

Source: Adapted from Schein, E.H. 2004. *Organizational Culture and Leadership*, 3rd Edition.

Information Security Policies, Standards, Procedures, and Guidelines

One of the major critical components of an information security program is the formulation, collaboration, and adoption of information security policies. These written policies cannot survive without the associated supporting standards, procedures (some private sector organizations use standard operating procedures or SOP), and guidelines. Personally, having clear, distinct, and physically separated policies, standards, and procedures would provide benefits to your overall information security program.

Charles Cresson Wood, well known in the information security industry as a leader for information security policy development, has emphasized segregating information that has different purposes. Specifically, one should formulate different documents for policy, standards, procedures, and guidelines. This structure provides numerous benefits to the responsible owner of these documents in terms of the ease of modification to maintain currency and relevance, reviews and approvals are more efficient, and requests for any single type of document can be distributed for a need to know the basis that protects the security and privacy of the written information, where applicable.

Policy is defined as the rules and regulations set by the organization (including addressing institutional issues). Policies are laid down by the management in compliance with the applicable law, industry regulations, and the decisions of enterprise leaders and stakeholders. Policies, standards, and procedures are mandatory; guidelines are optional. However, policies can be used to clearly define roles and responsibilities of the information security program, including the CISOs, steering committee, and so on. Moreover, policies are written in definite, clear, and concise language that requires compliance. The failure to conform to the policy can result in disciplinary action, termination of employment, and even legal action.

Information security policy governs how an organization's information is to be protected against breaches of security. The familiar examples of policy include requirements for establishing an information security program, ensuring that all laptops are deployed with automatic hard-disk encryption software, employees' Internet usage, security awareness and training, malware (e.g., antivirus, antispam, and antispyware) defense, and computer incident reporting for employees, to name a few.

Information security standards can be an accepted specification for software, hardware, or human actions. These standards can be de facto, as well, when they are so widely used that new applications routinely respect their conventions. However, the written format is preferred and recommended from the perspective of an information security professional and IT professionals including auditors.

A software standard can address a specific vendor's solution for antivirus software protection. In fact, from a defense-in-depth perspective, an organization may be standardized on two vendor's solutions. If a particular organization has implemented all Cisco systems, incorporated (Cisco) network devices, they could conclude that their hardware standard for the network infrastructure is Cisco. There are many standards to address human actions or even their behavior. For example, to address a potential computer security breach, a standard will address actions to be performed by specific employees' roles, responsibilities, and timelines for an appropriate response.

Procedures prescribe how people are to behave in implementing policies. For example, a policy might stipulate that all confidential and private data-network communications from employees working or traveling and desire to connect externally to the enterprise network must be encrypted. This would constitute the previously identified software and hardware (perhaps an adopted standard for communicating externally) required to be implemented based on the policy. The

corresponding procedure (the "how-to") would explain in detail each step required to initiate a secure connection using a particular virtual private network (VPN) or some other technology.

Policies, standards, and procedures as previously stated are mandatory. However, guidelines are not mandatory. Guidelines could be used as a documented standard or procedure where invariably, in the future, they could be transformed and adopted into a standard or a procedure. Establishing guidelines assists in identifying the usefulness and the trial of specific security controls for future adoption. For example, if an organization prefers the use of Windows Mobile operating system for all mobile devices, however, there is a small community within the organization that prefers the proprietary Blackberry device. Therefore, one may have to satisfy both the communities. A guideline would be feasible to address the appropriate security controls for the Blackberry device, where a standard would address the appropriate security controls for all Windows Mobile devices. Eventually, the Blackberry security controls guideline would be transformed into a standard after a greater acceptance within the organization was achieved. This eliminates the use of a de facto standard in this example.

All documents should use the suitable policy resources, including the aforementioned Charles Cresson Wood's Information Security Policy Made Easy, government (e.g., National Institute of Standards and Technology [NIST], National Security Agency (NSA), security guidelines, and RFC 2196), industry bodies (e.g., International Standards Organization (ISO) 17799/27002, control objectives for information and related technology [COBIT], and Committee of Sponsoring Organizations [COSO]), and commercial (e.g., Microsoft) organizations in preparing to formulate policies and standards.

The writing style should state what employees can do and what they cannot do, use of short sentences, written at a 10th-grade level similar to the model newspapers use, review and improve (i.e., sunset date), or adapt policies regularly, circulate drafts showing changes in policies to stakeholders and interested participants prior to adoption, and articulate the major changes to the senior management (e.g., Department Heads, Counsel, CIOs, and Privacy Officers) within the enterprise.

The Information Security Organization

The organizational culture and behavior—the CISO as the information security executive—and the organizational structure are the dependent variables in establishing an information security program. The framework that has been proved at numerous local governments west of the Mississippi River, regardless of the workforce size, is the "Security Triangle" (Figure 8.2). This framework has paid dividends in having clearly defined roles and responsibilities, while addressing defense and offense strategies. In other words, these strategies are the previously stated reactive and proactive teams that allow for continual collaboration with stakeholders vertically and horizontally throughout the public sector organization.

The following Information Security Strategic Organization diagram (i.e., Security Triangle) depicts an example from a local government (i.e., county government). It illustrates the CISO at the top of the organization that may report to a CIO or CEO, as previously stated. The ISSC is composed of the DISOs. This will provide a forum for all information security-related collaboration and decision making. This deliberative body will weigh the balance between the heightened security and departments performing their individual business. The ISSC responsibilities will be to

- Develop, review, and recommend information security policies
- Develop, review, and approve the best practices, standards, guidelines, and procedures

- Coordinate interdepartmental communication and collaboration
- Coordinate countywide education and awareness
- Coordinate countywide purchasing and licensing
- Adopt information security standards

The DISOs are responsible for departmental security initiatives and efforts to comply with countywide information security policies and activities. They also represent their departments on the ISSC. To perform these duties, the DISO must be established at a level that provides management visibility, management support, and objective independence. DISO responsibilities include:

- Represents their department on the ISSC
- Develops departmental information security systems
- Develops departmental information security policies, procedures, and standards
- Advises the department head on security-related issues
- Department security awareness programs
- Conducts information security and privacy self-assessments/audits

The Countywide Computer Emergency Response Team (CCERT) will respond to the information security events that affect several departments within the county with actions that must be coordinated and planned. The CCERT comprises of membership from the various departments that are often members of the departmental computer emergency response team (DCERT). The CCERT team meets biweekly to review the latest threats and vulnerabilities, and to ensure that the membership data are kept current. The CISO participates in their activities, as well as leads the response to cyber-related events. The efforts include improved notification and communication processes, and ensuring that weekend and after-hour response are viable. Additionally, training will be conducted to provide forensic capabilities to the CCERT team members, but specific to incident response in terms of maintaining the chain of custody of electronic evidence.

The information security strategic framework (Figure 8.1) developed to support a local government is designed to address the organization, people, processes, and technology, as they relate to information security. The strategy is based on the principle that security is not a one-time event, but must be a continuously improving process, an emergent process that addresses changes in business requirements, technology changes, new threats and vulnerabilities, and the need to maintain currency with regard to software release levels at all levels within the security network, server, and client arena. It is also based on the realization that perfect security is an impossible goal and that efforts to secure systems must be based on the cost of protective measures versus the risk of loss.

As the CISO or ISM, many of these protective measures are identified in an information security strategy, as a necessity. A documented strategy that is annually reviewed is imperative to ensure the currency of the projects and initiatives for that particular fiscal year. It is prudent that as the information security practitioner, you align your security projects, initiatives, and activities with the annual budget process of the organization. This will provide a means and awareness to the senior management that funding is mandatory to sustain a true information security program that will reduce the risk. This strategy must clearly articulate the mission and vision of the program. Additionally, information security program goals and objectives are articulated in the strategy, in terms of short- and near-term timelines. For example, your high-level goals can be derived from the 12 information security domains that are articulated in the ISO 27002 standard. The objectives will support the stated goal that should apply to your organization's required level

of security protections. The strategy will assist in the CISOs ability to achieve the stated goals and objectives over a defined period.

Conclusion

Today's information security professional practitioner is increasingly being challenged in numerous facets that are warranted based on the numerous human behaviors (e.g., negative and cultural generations' disparities), technological threats, and vulnerabilities that exist in the world. Government organizations are among us and throughout the world that experience constant probing for various reasons, as well as for major software houses. While discussing the local government or private sector organization challenges, some specific areas are unique to the government, such as the diversity of businesses and services under a single organization (i.e., county or city government), the type of businesses that warrants differing security and privacy protections, multiple legislations and regulations that sanction departments within a local government organization, and perhaps, most of all, the organizational and institutional culture issue because of the Civil Service Rules that provide the difficulty when employee termination is considered.

The CISO responsibilities range by establishing and sustaining positive relationships with the executive management, learning about the organization's culture and behavior, including institutional culture and generation differences, constantly being visible and communicating the security message throughout the organization, having formulated clearly defined policies, standards, and procedures, and establishing a governance structure that comprises and establishes a successful information security program.

In today's global society, a career path definitely exists for information security practitioners that would ultimately lead to holding a position as a CISO or CSO. This chapter as well as other chapters in this book should provide dividends throughout your career as a practitioner. However, business acumen, IT experience, and enormous leadership and organizational skills are a few of the major tenets in striving to be an outstanding and successful CISO. On the other hand, IT training curriculum does not usually include managerial skills such as leadership, team building, collaboration, risk assessment, and communication skills including negotiations, as well as psychology and philosophy courses.

Robert K. Pittman Jr. is a public sector employee who exceeds 32 years in IT that includes over 16 years of information security experience, as well as being the CISO since 2008. He will be receiving his doctoral degree (with honors) in public policy with his field of interests in organizational behavior and culture from the University of Southern California in May 2013.

References

1. Patota, N., Schwartz, D., and Schwartz, T. 2007. Leveraging generational differences for productively gains. *Journal of American Academy of Business*, September 2007, 11(2), 2.
2. Maher, K. 2003. *Career Journal: The Jungle in Wall Street Journal (Eastern Edition).* November 11, 2003, p. B10, New York, NY. Retrieved November 1, 2012, from http://proquest.umi.com/pqdweb.
3. Schermerhorn, J.R., Jr., Hunt, J.G., Osborn, R.N., and Uhl-Bien, M. 2010. *Organizational Behavior*, 11th Edition. Hoboken, NJ: John Wiley & Sons, Inc.
4. Coleman, J. 1990. *Foundations of Social Theory*. Cambridge, MA: Harvard University Press.
5. Mantzavinos, C. 2001. *Individuals, Institutions, and Markets.* Cambridge, UK: Cambridge University Press. Chapter 6.

6. Hayek, F.A. 1960. *The Constitution of Liberty*. London: Routledge.
7. Kohlberg, L. 1984. *Essays on Moral Development*, Vol II: *The Psychology of Moral Development. The Nature and Validity of Moral Stages.* New York: Harper & Row.
8. Gemtos, P.A. 2010. *Institutions*. The Sage Handbook of the Philosophy of Social Services. Retrieved from http://www.mantzavinos.org/wp-content/uploads/2011/05/Mantzavinos-chapter-19.pdf, p. 405.
9. Zemke, R., Raines, C., and Filipczak, B. 2000. *Generations at Work: Managing the Clash of Veterans, Boomers, Xers, and Nexters in Your Workplace*. New York: AMACOM Books.
10. Schein, E.H. 2004. *Organizational Culture and Leadership*, 3rd Edition. CA: Jossey-Bass.

Chapter 9

Metrics for Monitoring*

Sandy Bacik

Contents

A security policy architecture document should not be written unless it applies to protecting an enterprise asset and unless executive management is willing to enforce it. Another thing to remember is as a security policy architecture document is written, how is it going to be monitored and enforced? Therefore, what specific items or activities can be monitored that are documented within the security policy architecture document. The details from the security policy architecture documents are what the enterprise can use to develop and document security metrics or the return on security investments (see Figure 9.1).

A metric is defined as the quantitative standard of measurement that coincides with a specific method, procedure, or analysis, or it can be a set of related measures that facilitates the quantification of security characteristics. Security metrics can be tracked for costs and benefits. The security metrics can answer questions such as

- What is expected from the technology and security teams?
- How much is my security team costing the enterprise?
- How much is being spent on asset protection relative to costs and time? (This will enable better decisions for risk and security technologies.)
- How protected are the enterprise assets? Is the management enabled to make better risk decisions on enterprise assets?
- What security metric information should be presented to assist the business?

* From Sandy Bacik, *Building an Effective Information Security Policy Architecture*, Copyright 2008 Taylor & Francis Group, LLC.

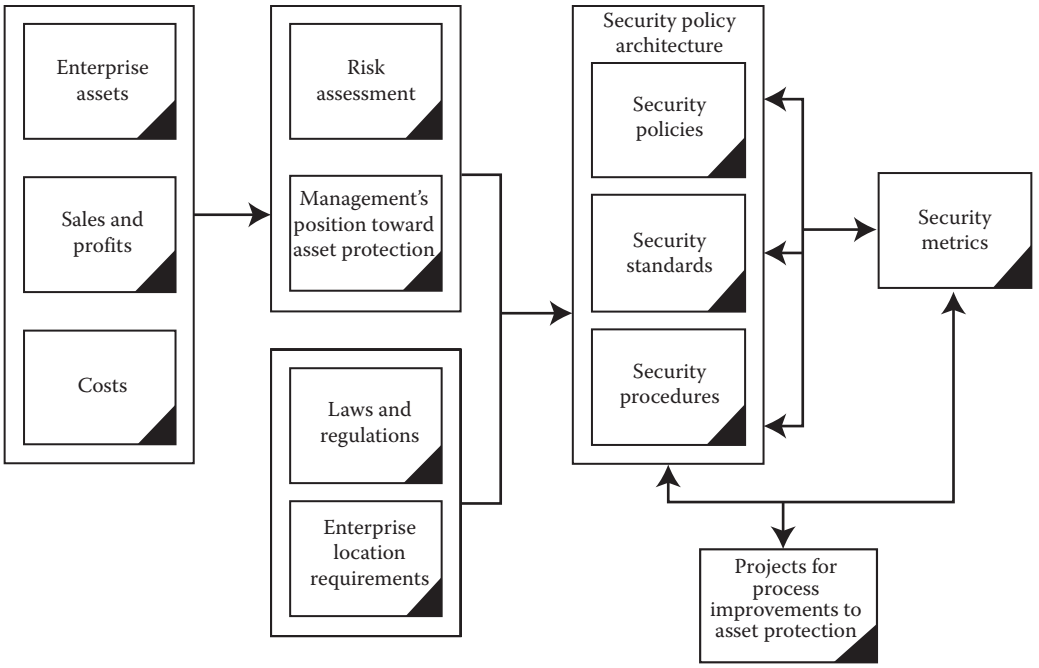


Figure 9.1 Feeding metrics.

Monitoring for Enforcement

The policy architecture enforcement can be a problem, especially if a physical security policy states that there is no piggybacking for building the entrance. Unless the enterprise has mantraps for building the entrance, how is the enterprise going to enforce that physical policy statement unless there is a security guard who watches everyone and every building entrance. Or if there is a logical policy statement that nonenterprise devices cannot connect to the enterprise network, who will enforce this?

Although scripts for every network switch and router in the enterprise can be written and implemented, is it practical for the enterprise to perform this type of monitoring? There is the option of network access control and unless the baseline traffic has been defined, there may be many false positives reported. Enforcement is sometimes the hardest part of the security policy architecture. Security policy architecture monitoring cannot be done as a project. It must be done as an ongoing business process. The enforcement needs to be measurable and repeatable.

As policy architecture documents are being written, the writer needs to consider the enforcement of the document. How is compliance going to be monitored? How is the reporting of a violation going to be performed? Who is going to be performing the enforcement? A security professional reports the facts and does not put any emotions into what is being reported or done by a user. In a rare incident, the security professional may have to report to an executive for a policy violation.

The security professional must have the supported documentation to formally report the incident. This comes through routine monitoring and metrics. For policy architecture documents that cannot be automatically monitored, such as logical security, routine walk-throughs need to be conducted and regular user awareness training needs to be conducted. The more the enterprise

builds security into the culture, the easier anomalies will stand out for irregular policy monitoring. As walk-throughs are performed for policy architecture compliance, the security professional needs to document the following items:

1. Date of the review
2. Name of the reviewer
3. Enterprise location of the review, such as which facility, department, or building area
4. Topic of the review
5. Name of the staff talked to
6. Findings and areas of the risk
7. Mapping the document topic to regulatory requirements

Many executives know that the enterprise cannot be managed if there is nothing against which to measure. Before regular monitoring takes place, baselines for the security policy architecture compliance need to be documented. In many countries, privacy laws require enterprises to disclose what is being monitored and logged with regard to employee activities. The security team needs to document business reasons for monitoring the employee activities. The basic business reasons for monitoring various business areas are as follows:

- Business transactions or to ensure that processes are being followed
- Regulatory compliance
- Enterprise systems functioning effectively and preventing unauthorized access
- Training and service standards
- Preventing criminal activity

An enterprise needs to remember that assets are being monitored and the monitoring cannot invade an employee's privacy. Enterprises need to define monitoring with specific limitations, such as

- Track data, rather than communication contents, unless filtering for the loss of intellectual property.
- Use periodic checks, rather than continuous monitoring, such as system configurations.
- Target enterprise areas that are high risk.

The enterprise needs to define monitoring. For example, monitoring can be defined as communication interceptions, validation of systems and their configurations, or the logging, recording, reviewing, and auditing of data. A simple monitoring document can comprise a few statements, such as

> There are legal, privacy, and regulatory requirements and conditions with which the enterprise must comply. There is also a business requirement for monitoring because the enterprise must protect all of its assets. The XXX department may monitor and record communications to establish the existence of facts for regulatory and legal practices, to prevent and detect crime, to detect and investigate unauthorized use and access to enterprise systems, and to secure the use of all enterprise assets. Monitoring will be conducted and documented with business and legal requirements. Note that monitoring and logging without authority may be a criminal offence, so such activities must be agreed to by the management team.

Table 9.1 Sample Baseline Metrics

Topic	*Metric*
Results	Viruses stopped
Results	Spam messages filtered
Results	System patches implemented
Results	New accounts created per operating system/application/remote access
Results	Accounts removed per operating system/application/remote access
Results	Accounts disabled per operating system/application/remote access
Results	Number of Internet/internal probes logged
Results	Number of Internet/internal port scans
Results	Summary of firewall rule usage
Results	E-mail messages processed
Results	Accounts and passwords reset per operating system/application
Results	Certifications maintained by the staff
Results	Number of unauthorized wireless access points
Results	Number of new network devices (workstations, routers, switches, and firewalls)
Results	Number of retired network devices (workstations, routers, switches, and firewalls)
Results	Number of new wireless devices (cell phones, BlackBerrys, and personal digital assistants)
Results	Number of retired wireless devices (cell phones, BlackBerrys, and personal digital assistants)
Results	Top 10 hosts for source/destination port scanning, network probing, and e-mail
Results	Number of administrative/root accesses for operating systems/applications
Costs	Estimated costs of a virus infection per hour of downtime, including recovery time
Costs	Estimated costs of a breach of customer information, including recovery time
Costs	Estimated costs of an internal network breach, including recovery time
Costs	Estimated costs of a breach of staff personal information, including recovery time
Costs	Estimated cost per person for security administration, including recovery time
Costs	Estimated cost of vulnerability exposure being compromised, including recovery time
Costs	Estimated cost per operating system and application for each patch for security vulnerabilities
Costs	Estimated costs to perform compliance and security controls (monitoring and auditing)

Table 9.1 (continued) Sample Baseline Metrics

Topic	*Metric*
Quality	Number of security incidents
Quality	Number of remediated security vulnerabilities
Quality	Number of compliance controls
Quality	Number of control deficiencies
Quality	Number of privacy incidents
Quality	Number of employees, contractors, consultants, and partners
Quality	Number of new hire/termination/transferred employees, contractors, consultants, and partners
Quality	Number of staff (employees, contractors, consultants, and partners) who have attended the security awareness training
Quality	Number of visitors and vendors on-site

After enterprise monitoring statements have been defined to assist with security policy architecture enforcement, defining baselines and metrics is the next step.

Baselines

A baseline is a standard by which things are measured or compared. Establishing a security baseline will give enterprises a benchmarking for the progress of asset protection. Before the baseline can commence, the following types of questions must be answered:

- What specific information/data will be collected and why?
- What are the business drivers/requirements for the collected information/data?
- When and how will that information/data be collected?
- Who will be responsible for the collection and disbursement of the information/data collected?
- What is the team attempting to show with the collected information/data?
- Who will receive the information/data collected and in what format will it be presented?
- Can the team ensure that the presentation is simple and clear?
- Whether a third-party application is going to gather these data or a homegrown script basic application, requirements need to be established for the gathering and processing of the baseline.

A sample listing of the metrics that can be collected can be seen in Table 9.1.

Routine Metrics

Once the baseline is established, regular metric processes need to be performed. The security team needs to establish specific procedures or work instruction on how the information and data will be

collected and where they will be distributed on what specific schedule. This will allow the management to see the standard progress in asset protection. The regular metrics will give the security the foundation for justifying additional security software and additional staff for the security program. The regular metrics also can be used when presenting user awareness training. Metrics can improve accountability through collection, analysis, and reporting of relevant performance-related data. The procedures should be a step-by-step document on how and where the scripts or applications are executed. The security metrics program can be set up on a repeatable cycle. See Table 9.2 for a sample of the security monitoring schedule.

1. Define metrics and thresholds.
2. Document the source from which the metrics are gathered and determine how accurate these data are.
3. Collect and transform these data using automated or manual tools.
4. Report on the results and ensure that the results can be duplicated with these data.
5. As the environment or situations change, review and revise what metrics are being monitored and reported.

Reporting

Metrics reporting can be classified as operational and business metrics, because metrics serve different audiences. Operational reporting can include the following:

- Number of security policy violations by risk
- Number of devices deviating from enterprise security configuration standards
- Number of abandoned accounts
- Number of different types of attacks against the enterprise from external sources
- Number of viruses, worms, and Trojans that were blocked

Business reporting includes items such as the following:

- Amount of downtime and cost caused by misconfigurations or wrongful implementations
- Dollars overspent or underspent in security projects
- List of outstanding or accepted risks
- Amount of change, better, worse, or the same, since the last reporting period

The enterprise culture will determine if the metric reporting will be done through spreadsheets, PowerPoint presentations, or executive dashboards. The key to all the presentations of data is that they are in summary form and are relevant to the objective of the presentation. Spreadsheets and PowerPoints need planning for their presentation. Many times, it is just a process of gathering the summary information from multiple points and putting them into a centralized location. In contrast, dashboards can be used for up-to-date ad hoc reviewing of metric reporting. Dashboards can be used by the management to get a quick view of the security metrics within the organization.

Most metric reporting is a combination of operational and business and is presented in various forms, depending on the audience. The security team needs to define the methods for reporting and what information will go into each type of presentation when the security metrics program is being developed.

Table 9.2 Sample Monitoring Schedule

Security Monitoring Task	*Frequency*	*Time Period*	*Duration*
Server, domain, and NT security issues	Alert/D/W/M		
Firewall and IDs log	Daily/alert		
Review network device logs for security violations by accounts	Daily/alert		
Review network device logs for security violations by superuser accounts	Daily/alert		
SQL server log	Daily/alert		
Virus activity	Daily/alert		
Attend change management meetings	Weekly	1, 2, 3, 4 W	M
Attend IT department staff meetings to monitor the critical project status reports and decisions, provide control guidance, and so on	Weekly	1, 2, 3, 4 W	M
Exchange database—authorized user review	Weekly	1, 2, 3, 4 Q	M
Exchange e-mail security issues	Weekly	1, 2, 3, 4 W	M
MS security configuration tool/SecEdit	Weekly	1, 2, 3, 4 Q	M
Oracle database	Weekly	1, 2, 3, 4 W	M
Oracle server log	Weekly	1, 2, 3, 4 W	M
SQL server database	Weekly	1, 2, 3, 4 Q	M
Monitor third-party security compliance for access to information assets	Monthly	1, 2, 3, 4 Q	M
Provide information security status reports to the management	Monthly	1, 2, 3, 4 Q	M
Review the administrative password changes for network and application accounts	Monthly	1, 2, 3, 4 Q	M
Review the inactive account activity for network and application accounts	Monthly	1, 2, 3, 4 Q	M
Internally run the external network audit/security analysis tools—NMAP, NESSUS	Quarterly	1, 2, 3, 4 Q	M
Maintain and distribute employee emergency contact cards	Quarterly	1, 2, 3, 4 Q	L
Provide the summary of security status reports about information security program for reporting to the executive team	Quarterly	1, 2, 3, 4 Q	M

continued

Table 9.2 (continued) Sample Monitoring Schedule

Security Monitoring Task	*Frequency*	*Time Period*	*Duration*
Review the physical security	Quarterly	1, 2, 3, 4 Q	L
Review to ensure user account password changes—systems and applications	Quarterly	1, 2, 3, 4 Q	M
Run DBMS security analysis tools	Quarterly	1, 2, 3, 4 Q	M
User group privileges for privileged groups and IT-owned groups—final tools to be determined	Quarterly	1, 2, 3, 4 Q	M
Validate contractor and remote user access to the network	Quarterly	1, 2, 3, 4 Q	M
Collect and review disaster recovery test records and report to the management	Semiannually	First and third Q	M
Data classification processes	Semiannually	First and third Q	M
Maintain, monitor, and test the business continuity planning	Semiannually	Second Q and fourth Q	H
Maintain, monitor, and test the disaster recovery planning	Semiannually	Second Q and fourth Q	H
Monitor IT disaster recovery testing and support IT audit's review of IT disaster recovery plans	Semiannually	First and third Q	M
Annual internal information security awareness	Annually	Third Q	L
Maintain information security templates, guidelines; for example, data/information security	Annually	Second Q	M
Maintain the corporate information security policy architecture	Annually	Second Q	M
Password strength testing	Annually	Second Q	M
Review business unit business continuity plans	Annually	Third Q	L
Review business unit disaster recovery plans	Annually	Third Q	L
Review the IT and security job functions/ descriptions and match them to the production access for the separation of duties test	Annually	End of the first Q	M
Scrub network, application, e-mail, voicemail, and wireless account information	Annually	End of the second Q	H
Support the coordination of the annual network security/risk/vulnerability review	Annually	Second–third Q	H

Table 9.2 (continued) Sample Monitoring Schedule

Security Monitoring Task	*Frequency*	*Time Period*	*Duration*
Maintain and coordinate network incident response	Ongoing	As needed	M
Monitor the implementation of system patches, configuration management	Ongoing	1, 2, 3, 4 Q	M
Monitor compliance with the enterprise information security policies and procedures among employees, contractors, alliances, and other third parties, and refer problems to the appropriate department managers or administrators	Ongoing	1, 2, 3, 4 Q	M
Audit logs (system, application, security, registry)	Troubleshoot	As needed	
Coordinate incident response investigations	Troubleshoot	As needed	M
Internet gateway	Troubleshoot	As needed	
VPN access	Troubleshoot	As needed	
Architecting network design changes	Ad hoc	As needed	M
Monitor IT measurements "audit certifications" of work completed by areas: database management, help desk, network, operations, and telecommunications	Ad hoc	As needed	M
Provide information security training to new employee classes	Ad hoc	As needed	M
Review the security features of new computing systems to ensure that they meet the security policy requirements	Ad hoc	As needed	M

Note: The blank items are not being done and are the possible gap items.
Frequency: D, daily; W, weekly; M, monthly
Time period: Q, quaterly; W, weekly
Duration: L (Low) = A few hours per occurrence (unless the frequency is high then it is considered a Medium duration)
M (Medium) = Up to 1 week total commitment
H (High) = Anything greater than 1 week

Policies, Standards, Procedures, and Guidelines

Chapter 10

Security Implications of Bring Your Own Device, IT Consumerization, and Managing User Choices

Sandy Bacik

Contents

Bring your own device (BYOD) or bring your own technology (BYOT), two of the newer acronyms within information technology (IT) has unleashed the age of IT consumerization. With the proliferation of various types of mobile devices and the consumerization of IT, corporate networks are becoming more challenging to manage and are more at-risk every day. The days of a static, one-size-fits-all policy applied to company-owned assets are over. Today, information security needs to understand what devices are trying to connect to their networks, so that they can provide secure layered access to those that are authorized. Today's mobile devices no longer include just a cell phone. Devices such as iPhones, Androids, smartphones, iPads, and tables are minicomputers; they have processing power, storage, communications, applications, and, more importantly, they are not always owned by the organization. The risk to organizational assets is now more distributed and uncontrolled. Information security needs to understand the risk and provide controls to limit this increasing risk to organizational information assets.

Let us take a step back. In the days of company-purchased and company-maintained assets, setting information security controls to limit access through authorization and authentication was a challenge, but could be maintained, because the organization knew where the assets were placed and who they were assigned to. Company-purchased devices were only for business use and some of the controls used included

1. Authorization from the management before purchase
2. Authenticated access
3. Separation of duties with role-based access
4. Monitoring of assets to be based on a known list of devices
5. Wiping information from a device when it is taken out of service

The separation of personal life and business life was relatively easy. There was one set of policies, standards, and procedures no matter which location the company placed the purchased devices. Today, many IT and information security departments are struggling to define and implement a BYOD strategy, especially when they know the devices are already accessing the corporate network and information assets.

With constrained budgets, staff reductions, and outsourcing, how IT and information security protect the organizational information assets is more complicated. And today, many IT groups underestimate by quite a bit the portion of employees using their own devices for accessing organizational information assets and organizational business. The proliferation of company-owned and personal mobile devices are not likely to abate any time soon. In today's environment, many professionals use at least two personal devices to access organizational information assets and systems. The most common scenario is a laptop or tablet and a smartphone. While the use of mobile technologies has the potential to transform the organizational business model, making it possible for an organization to be more agile in serving its internal and external customers, it can also disrupt IT and how information security protects the organizational information assets. New employees expect complete freedom to use mobile devices, whether they are company-owned or employee-owned. New employees are expecting privacy, while the information security department needs to provide governance over the organizational information assets. Information security can no longer avoid these common mistakes with the consumerization of IT:

- Mobile applications
- Leaving passwords up to the users
- Missing out on the cost–benefits or reduced total cost of ownership
- Failure to define and enforce device requirements
- No monitoring follow-up over the life of the device, malicious endpoints, and wanted/unwanted applications
- No defining of prerequisites for access
- No remediation plan for BYOD incidents

One of the first things information security needs to consider when deploying secure BYOD is a simplified workflow that can accommodate all types of devices and work with the existing infrastructure. This workflow while simple will allow organizations to start small, measure progress, and then support better additional requirements on the base platform as needed (see Figure 10.1).

This gradual workflow allows the controlled implementation of a BYOD strategy. Yes, it does need to start with prohibiting, so that foundational policies, standards, procedures, and controls can be established for a baseline. As with any project,

- piloting of devices
- limiting the access to information assets
- defining which organizational resources and information assets the BYODs can access

should be part of the deliverables and tasks.

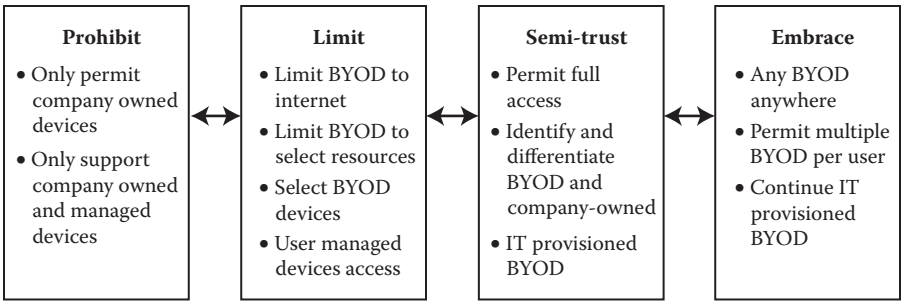


Figure 10.1 Accommodating BYOD.

On the basis of policies, standards, procedures, and controls, information security can start trusting the BYOD access to some information assets, then allowing the organization to start embracing the BYOD concept.

The benefits of such a strategy shift are both tactical and strategic. The most obvious benefit is reduced cost of purchasing company-owned devices. Information security needs to investigate the current types of company-owned mobile devices and what types of layered security controls are in place to protect information assets. Then what other devices have been seen or found in the employee community. After the majority types of mobile devices have been identified, requirements can be developed to select a mobile device management (MDM) system to centralize the administration of mobile devices. One thing to note, many organizations are a Windows-based environment and many mobile devices are not Windows-based operating systems; so, simple Windows access controls and group policies will not always work. The next few paragraphs do not endorse a specific vendor for an MDM, but help an organization to define requirements, if the organization chooses to implement an MDM. An MDM brings the following values to organizations using mobile devices (company-owned and non-company-owned):

1. Quickly configure devices
2. Centralizing security policies and enforcement rules
3. Enabling certain applications
4. Assist with issues of the troubleshooting device
5. Assist in finding a lost device
6. Can wipe the personal data when needed
7. Can wipe only corporate data
8. Can push applications and updates when needed

With an MDM, information security can categorize the devices and then devise and apply the appropriate network access policy to each category based on the risk profile. Using company-owned mobile devices as an initial step, a baseline inventory can be conducted to detail an initial inventory of what is connecting to the network, and then possibly denying access to any new device for network connections. Now, the information security group can define and apply appropriate policies, standards, and procedures whenever a device connects to the network, whether it is a company-owned or a non-company-owned device.

Moving toward an MDM solution and enforcing some of the access control on a non-company-owned device, every global employee's concern might be, can "big brother," the organization

1. Locate where I am at midnight on a Saturday night?
2. See which personal applications I have installed?
3. Wipe my personal music and pictures without asking?
4. Read my personal e-mails?
5. View my browser history?

Here are some things that information security, executive management, human resources, and corporate counsel can do to satisfy corporate governance and end-user privacy:

1. Educate employees on mobile device risks.
2. Educate employees on how to protect their mobile devices.
3. Set the ground rules and ask the employees if they will comply before permitting access.
4. Stay away from personal data.

The most successful BYOD programs

1. Keep the users informed on why it is important to have an MDM solution and what is an information asset
2. Explain what IT can and cannot do to their personal device
3. Have formal acceptable use policies
4. Educate users about the additional risks posed by mobile devices
5. Communicate business needs to employees
6. Provide security for all data—personal, company, customer, and intellectual property
7. Include mobile policies in new employee training
8. Explain that device choice is an employee benefit
9. Set the expectation of mobile device responsibility from the start

From a strategic and tactical perspective, information security needs to define the guidance on how to consider employee-owned mobile devices by asking certain key questions to assist the organization in permitting employee-owned devices to access organization resources and information assets. The Appendix contains a set of security, policy, and risk considerations. Outside the security items, from an operational perspective, the organization will need to think about reimbursement for personally owned devices and will need to document it in a policy or a standard. A stewardship agreement that many organizations use when distributing company-owned assets should have a similar companion stewardship agreement for employee-owned mobile devices connecting to the company network and information assets. When defining any type of stewardship agreement, both human resources and corporate counsel should be involved in developing the content. After many of the questions in the Appendix have been answered, the company can start its strategic BYOD direction with the basic critical policies and standards that document things similar to the below-listed requirements:

1. Be registered to and monitored by the company MDM system.
2. What is going to be monitored on a company-owned and personally owned device when it is registered and why.
3. Notify IT before a non-company-owned mobile device hardware is being exchanged or replaced to ensure that company information has been removed from the non-company-owned mobile device and unregistered from the MDM.

4. Notify IT when a non-company-owned mobile device is lost to ensure that the non-company-owned mobile device is unregistered from the MDM.
5. Tethering a non-company-owned device to a company-owned device is prohibited.
6. Any default- or vendor-supplied password on the non-company-owned mobile device that may be already set for the system needs to be changed to a nondefault value.
7. All wireless connections shall use strong encryption.
8. The non-company-owned mobile device shall have, at a minimum, the following company security controls enabled on the non-company-owned mobile device, such as
   a. Encrypting the device.
   b. A password/passcode that matches the company standard with a password/passcode rotation and uniqueness.
   c. A password/passcode that meets the company's password length.
   d. Screen locking and requiring your password/passcode to regain access if you are inactive for a specific amount of time.
   e. Non-company-owned mobile devices shall be configured to report to the MDM on a regular basis.
   f. Non-company-owned mobile device activities shall be logged and monitored by IT to ensure compliance with company policies and procedures.
   g. Non-company-owned mobile devices shall meet the company authentication requirements.
9. The non-company-owned mobile device shall have an approved operating system (OS) version installed.
10. Accounts and access privileges shall be unregistered and company information shall be removed from the non-company-owned mobile device in a timely manner when an individual no longer has a business need to access a company information asset from the non-company-owned mobile device.
11. It is strongly recommended that non-company-owned mobile device shall install
    a. A firewall where practical
    b. An antivirus or antimalware package with regular updates
12. The non-company-owned device shall be set up to receive updates automatically from vendors of your purchased software.
13. If connecting from a noncompany location, establish at least a Secure Sockets Layer (SSL) connection before accessing any company information asset or electronic resource. Wireless routers in home locations are provided with public hot spots that often have poor protection.

This not only ensures consistent security, but also saves the IT staff hours, if not days, previously spent fighting security brush fires. Rather than having to manually identify and provision each new device, the IT staff are free to concentrate on important, more strategic projects that will further the business goals. An accurate and complete profile of the devices on the network enables better integration of those silos of security. IT need not rip out and replace the equipment or buy expensive upgrades. IT staff can leverage the technologies already in place to create a more holistic security posture. Ideally, this integration enhances the effectiveness of the previous technology investments.

The explosion of mobile devices is both a boon and a bane to the security of organizational information assets. These devices, which are powerful yet inexpensive and easy to use, help employees to be more productive no matter where they are. They can increase collaboration and the flow of information, enabling an organization to react more quickly to market conditions and customer

needs. And they are becoming so ubiquitous that organizations often do not even have to buy the devices—employees are bringing their personal smartphones or tablets to work. Saying "no" to business use of personal devices for accessing organizational information assets and systems may no longer be an option and information security needs to adapt to how it protects organizational information assets. Today's growing demand for anytime, anywhere network access has expanded to include the use of personal mobile devices such as laptops, tablets, smartphones, e-readers, and more. This BYOD phenomenon is changing the way IT organizations and users address network access security. For IT organizations, BYOD means supporting a variety of devices and their operating systems, and maintaining an expected level of service. To keep the costs low, it must be easy to securely onboard new devices and quickly identify and resolve problems.

Managing User Choices

Whenever managing user choices come into play, the wild, wild West of cowboys appear and information security needs to be able to guide the wild, wild West to decisions where the risk level is acceptable to the business. Yes, many times, it seems like information security is the mediator when the user community starts asking for new features, hardware, and applications. But how can and does information security manage user choices to ensure that the risk level is acceptable to the business.

When it comes to monitoring the enterprise environment, information security may have many options on how this is done and what is used to perform the monitoring. Yet, the user community does not want "Big Brother" watching what they do. Using the leverage of management, information security can make monitoring more palatable to the user community by performing the following items:

1. State the specific business purposes for monitoring.
2. Clearly state the ownership of company computers, networks, files, and e-mail.
3. Clearly outline the forms of communication considered as illegal, prohibited, and unacceptable.
4. Clearly outline the websites considered as illegal, prohibited, and unacceptable.
5. Define the acceptable use of company networks and e-mail.
6. Set clear boundaries for the personal use of company networks.
7. Inform employees of the specific types of monitoring activities that will be used.
8. Explain how monitoring activities are advantageous to employees, clients, and the company.
9. Determine the consequences for policy violations.

As the business users need additional services, applications, or hardware to perform their daily responsibilities, information security needs to be ready to assist the business in determining the risk of the additional services, applications, or hardware to the current infrastructure. This means information security should have a list of requirements or risk concerns as it applies to each separate item. For example, referring to the chapter on cloud computing, while the information security group may not know what the business is looking for, it can guide the business to think through the risks of the requested service, application, or hardware.

Information security needs to have guidelines and standards formally documented and available to the business, when the business is ready to look for additional tools. These standards and guidelines need to be documented, publicized, and available to the business for use and

understandability. When the business does not know what the security or interoperability requirements are, the business will go out and purchase software for what they feel fits their needs, sometimes without regard to postsale and implementation support or learning curves needed by IT for the ongoing support. Information security and IT need to remain business friendly in guiding the business in selecting the service, application, or hardware that fits into the current infrastructure, while protecting the other organizational information assets. While this is not an all-inclusive list of topics, here is a list of topics that should be documented and available for the business and potential vendors to use when selecting new services, applications, and hardware:

- *Transmission Control Protocol (TCP)/Internet Protocol (IP) Standard:* This will allow a vendor and the business to determine what changes will need to be done to the infrastructure for the new service or application to be implemented and supported.
- *Third-party access standard:* This allows a vendor to know what will be required, if they need to remotely connect on-site to the new application or hardware.
- *Standard enterprise hardware and software:* This allows the business to know what is currently available and will guide them in determining what service, application, and hardware can be supported and secured within the organization.
- *Application of security requirements:* This allows the business and vendor to review the selected service or application to ensure things such as auditing, logging, monitoring, reporting, administrative access, and separation of duties can be done with the service or application.
- *Purchasing security requirements:* This allows the business to negotiate with the vendor to ensure that the organizational information assets will be protected when the service, application, and hardware are implemented.
- *Security service requirements:* While this is similar to the application of security requirements, this documents the security services required to be used when implementing a service or an application.

To ignore or to not work with the business in defining requirements and guidance in selecting a service, application, or hardware, the business choices may make it nearly impossible to support, implement, and secure the organization infrastructure and information assets. Information security needs to stay in front of the business to guide them into looking at the risk and the interoperability with the existing infrastructure.

Appendix: Questions to Assist in Determining a BYOD Strategic Direction

1. Should this strategic and tactical direction be global or regional?
   a. What are the privacy implications?
   b. What are the monitoring implications?
   c. What are the logging and auditing implications?
   d. How does each region handle mobile devices, company-owned and non-company-owned?
2. Eligibility considerations for employee-owned mobile devices
   a. Questions to be answered
      i. Are all (global) employees eligible for employee-owned mobile access to organizational resources and information assets?

   ii. Will the organization restrict access based on an employee's role, responsibilities, title, manager approval, geography, or other organizational considerations?
   iii. What is the business reason for requesting access?
   iv. Will the organization restrict access for employee-owned mobile devices to particular company resources, applications, or data? If so, which resources, applications, and data?
   v. Will the organization support any employee-owned device?
b. Policy considerations should be clearly addressed
   i. All employee-owned device eligibility requirements
   ii. Any employee-owned device support limitations
   iii. Employee or role risks and responsibilities
   iv. Any corporate resource, application, or information asset access limitations
   v. Any processes for obtaining management approval

3. Security considerations
a. Questions to be answered
   i. What is the organization's policy and process for handling a lost or stolen personal device?
   ii. What is the organization's policy and process for handling the decommissioning of a device (e.g., if a user switches to a new device, a change in the user's role/title deems them no longer eligible for access, the user leaves, is terminated by the company, etc.)?
   iii. Will the organization wipe the whole device, only corporate information assets and applications, or both? Who assumes the liability of the loss of data?
   iv. Will the organization allow the employee to initiate wipe action(s) themselves (e.g., through self-service portal)?
   v. What are the requirements to wipe a non-company-owned device?
   vi. Will the organization set and enforce the use of a whole device password? Will the password configuration match the organizational policy? Or will the organization only set and enforce the password on the personal mobile device?
   vii. Will the organization require limits on the use of cameras, browsers, Bluetooth, or other applications and services?
   viii. Will the organization require a specific service provider?
   ix. Will the organization require employees to acquire and install antimalware as a condition for access to corporate information assets and applications? Will the organization provide such antimalware? Will the organization require particular vendors or versions?
   x. What is the organization's policy and process for an employee device that has been infected with malware?
b. Policy considerations should be clearly addressed
   i. The policy should expressly prohibit
      1. Device "jailbreaking," "rooting," or the equivalent.
      2. Making any other modifications to the device hardware and/or OS software beyond routine installation of updates as directly provided by the applicable device maker. Performing such actions or making such unauthorized modifications is essentially an "inside attack" on the device, application, and data security, and should be treated very seriously.
   ii. The policy should be clear on process and timing requirements for reporting lost or stolen devices, changing to a new device, and actions to be taken when an employee leaves the company.

iii. The policy should be clear on whether or not the organization will require the use of whole device password and associated requirements for frequency of change, minimum strength, and so on.
iv. The policy should be clear on whether or not the organization will wipe the whole device and conditions under which the organization would do so (e.g., lost or stolen device, change to the new device, move to a new role, and departure from the company).
v. The policy should clearly state that the organization always reserves the right to wipe the organization information assets and applications and/or the whole device if deemed necessary in the organization's sole discretion to secure company information assets or applications.
vi. The policy should be clear that wiping company information assets and applications may impact other applications and data (e.g., including but not limited to the native address book data).
vii. The policy should disclaim any liability for the loss of noncompany applications or data, whether directly or indirectly resulting from the usage of company applications or data, and/or the wiping of such applications or data, or the whole device.
viii. The user should be encouraged to minimize the risk of losing personal applications and/or information assets.
ix. The policy should be clear on any restrictions on the usage of cameras, browsers, Bluetooth, or other applications and services. The ability to enforce such restrictions may be dependent on the device capabilities, which in turn may become an eligibility consideration.
x. The policy should be clear on any requirements for the use of antimalware (including specific vendors or versions as applicable) and process and timing requirements for reporting any suspected instances of malware infection.

4. Acceptable use considerations
   a. Questions to be answered
      i. What is the organization's policy regarding the use of a device by users other than the employee?
      ii. Will the organization provide Intranet access to non-company-owned devices?
      iii. Will the organization require employees to conform to acceptable use guidelines on all Internet usage, even if not enabled through corporate infrastructure and/or for personal reasons (e.g., as a condition of receiving stipend, reimbursement, or access to company applications or data)?
   b. Policy considerations should be clearly addressed
      i. The policy should be clear on whether the device is enabled for corporate applications and information asset access may be used or loaned to other users (e.g., if a personal mobile device has a separate password and the whole device password is not used, it may be acceptable for the company end user to allow someone else to temporarily use the device, as the use of the device does not require the company end user to first "unlock" the device that enables access to corporate data or applications).
      ii. If the organization provides Intranet access (e.g., through a mobile virtual private network [VPN] client), the policy should be clear that the company's acceptable use guidelines for desktop/laptop browser usage will apply to any usage of Intranet and/or Internet access that is enabled through the use of a mobile VPN infrastructure.

iii. Many companies will not apply acceptable use policies to usage not enabled through corporate infrastructure—if the organization chooses to do so (e.g., as the condition for receiving stipend or reimbursement), then the policy should be clear on this.

5. End-user support
   a. Questions to be answered
      i. Will the organization provide any end-user support for employee-owned devices?
      ii. If so, for what applications, services, or scenarios (e.g., lost or stolen device)?
   b. Policy considerations should be clearly addressed
      i. The policy should be clear on what support, if any, will be provided and
         1. Explicitly for which applications, services, and scenarios
         2. Any "self-service" actions that must first be taken before requesting support
         3. Process and/or tools for requesting support (e.g., submitting trouble ticket vs. calling)
      ii. Many companies will opt for a support policy that is expressly limited to the employee-owned mobile device and applications and requires that the users first attempt to resolve routine issues via "self-service" mechanisms (e.g., always contacting the carrier for billing issues, contacting the carrier first if not able to connect, and resetting own password via the self-service portal).
6. Policy violations
   a. Questions to be answered
      i. What should happen if the user violates the policy?
      ii. Should different violations be treated differently (e.g., eligibility vs. security vs. acceptable use)?
   b. Policy considerations should be clearly addressed
      i. The policy should be clear on the consequences of policy violation and any differences from one policy or type of policy to the next.
   c. Unauthorized access to be clearly addressed
      i. On receipt of a monthly bill, employees should immediately check the call detail record section of the bill for any indication of unauthorized calls. The discovery of any such calls should be immediately reported to whom internally?

Chapter 11

Information Assurance

*Open Research Questions and Future Directions**

Seth J. Kinnett

Contents

If the fields of computer science and technology analysis could be considered in their adolescence, truly, the discipline of information assurance (IA) is in its infancy. As recently as 2005, trade journals included articles filled with discussion surrounding the "pressing need to formalize information security as a profession" (Wyatt 2005). Compare this discussion to a long-since established field such as medicine. We would not expect to see medical journals discussing the need for accreditation and a greater quantification of professionalism. This presents the IA field with decidedly greater opportunities to chart its course for the future. Those professionals working in the field today have the chance to shape the future direction of their profession and their own careers. As with any field that is young, the body of research available will be relatively small. It is useful for both professionals and academics to understand not only what research is available, but also to understand what issues are still ripe for exploration. The purpose of this chapter is to highlight open research questions and to forecast the probable future direction of this rapidly emerging and important field.

First, it is important to establish some definitions along with an understanding of how IA interacts with the government, business, and society. To the former, we will define IA as "the set of measures that protect and defend information and information systems by ensuring their availability, integrity, authentication, confidentiality, and non-repudiation" (Cegielski 2008). From a contextual perspective, it is clear that "information security is essential to the proper working of government and, indeed, society as a whole is dependent on critical IT infrastructures" (Wyatt

* The views and opinions stated herein solely represent the views and opinions of the author and do not necessarily reflect the views and/or opinions of his employer and/or his colleagues.

2005). In the United Kingdom in 2005, an institute was proposed to aid in the development of consistent standards and to promote the best practices for IA. This broad scope of influence is evident in the "broad backing of government, industry and academia. . . that will give the institute credibility" (Wyatt 2005). Another article, published in 2007, highlights the fallout of this initiative 2 years later. While it does not focus on the proposed creation of an institute, it supports the idea that there have yet to be clearly defined best practices across all interested parties in this field.

The National Information Assurance Strategy, launched in June 2007 in the United Kingdom, had three main goals:

- To make the central and the local government able to deliver better services through the appropriate use of information technology (IT)
- To strengthen United Kingdom's national security by protecting information and systems at the risk of compromise
- To enhance United Kingdom's economic and social well-being as government, business, and citizens realize the full benefits of IT (Grant 2007)

It would be difficult to argue the importance of these goals. The real issues arise in the discrepancy around how these goals should be realized. "Ross Andersen, professor of security engineering at Cambridge University's Computer Laboratory, said the National Information Assurance Strategy harked back to the mid-1980s. 'It is full of consultant-speak, recycling tired old ideas. It is mind-candy.' Andersen said attempts to make individuals responsible for information assurance and security recalled similar provisions for patient health information. 'All that did was make everyone run for cover,' he said" (Grant 2007). The key point to understand from this dialogue is that IA is not just about technology. It is just as much of a cultural issue as it is a technical one. As in the implementation of information security controls, technical controls are only one facet of the larger strategy. Locks on doors and user training are often just as effective as a computer password. In other words, "it's not just a matter of getting the next patch right or getting the best firewall or the best screening tool. It's a much bigger issue" (McCormick 2007). A successful IA strategy must be broad in scope and must include the human component. According to David Porter, a consultant with Detica, "a neglected area [in the UK government strategy] is how you get people to actually take ownership of the information they hold or generate. This is a soft cultural change and nothing to do with technology" (Grant 2007). While the initiative had a broad input, clearly, there are some growing pains. To further understand the key issues facing the IA field, we now turn to an academic view. We will now explore the diverse skill sets needed for success.

Sarbanes–Oxley (SOX) was arguably the single most influential piece of legislation exploring the evolving field of IA. These professionals found themselves in high demand. Their skills suddenly needed to be applied to a very particular set of problems—verifying the relevancy of controls surrounding financial data, documents, and disclosure. Previously, "an auditor conducting a review of a company's income statement or balance sheet [attempted] to assess whether or not the presentation of the compiled accounting data accurately [reflected] the business's financial position" (Cegielski 2008). After SOX, the auditor gained the additional requirement of having to "opine on [the] effectiveness of the internal controls of the computer-based information systems from which those financial statements were generated. Simply stated, an opinion on the internal controls of the computer-based information systems of a financial statement audit client was an attestation that many accounting firms were ill prepared to render" (Cegielski 2008). Just as a financial accountant must have expertise (or access to expertise) in IA, so much that the IA professional must have some knowledge of finance to render judgments accurately. This highlights

the importance of versatility along with the understanding that IA is always tied to key business issues. The savvy IA professional most possesses business acumen to be successful. Determining the appropriate curriculum to educate the future generations of IA professionals is a key question that is critical to the successful evolution of the discipline.

In Casey Cegielski's essay on IA curriculum, he explores the notion that the knowledge that an information security professional must have is the ability to analyze controls. In particular, the professional must be able to evaluate controls relevant to a specific business process or objective. "An internal control is a process designed to provide reasonable assurance regarding the achievement of objectives in the following areas of business function: (1) effectiveness and efficiency of operations, (2) reliability of financial reporting, and (3) compliance with applicable laws and regulations" (Cegielski 2008). Cegielski takes a novel approach to IA education, arguing that it ought to be a component of the curriculum for those bound for other professional careers such as accounting. He suggests that "the traditional educational model for professional accountancy offered by many colleges and universities around the nation is inadequate to address the current demands within the profession for technology-based knowledge and skills" (Cegielski 2008). In his study of the key skills required to be successful as an IA professional, Cegielski compiles a number of surveys and ultimately arrives at 18 skills that are applicable to the associate, manager, or partner levels of IA. These include project management skills, basic accounting knowledge, financial statement auditing knowledge, database knowledge, spreadsheet application skill, oral communication skills, written communication skills, time management skills, business process knowledge, quantitative analysis skills, professional etiquette skills, system analysis/documentation knowledge, business strategy planning knowledge, internal controls knowledge, legal/regulatory knowledge, computer network skills, information security knowledge, and interviewing skills (Cegielski 2008). Note how explicit "information security knowledge" is only one of the 18 components. Perhaps one of the largest open questions is: is Cegielski correct? Have students undertaken his curriculum and subsequently displayed quantifiable results? Answering this question will allow us to refine Cegielski's hypothesis accordingly.

Businesses by their nature are keen to support that which presents the potential of profit. They would subsequently be expected to demand quantifiable results from IA initiatives. "Quantifying the ROI for any program is important because it is one indicator of the degree to which a program contributes to the parent organization's strategic plan. It can help prioritize investments. ROI can be used to help quantify an individual's or team's job performance, which can support annual performance appraisal evaluation rating levels" (Tichenor 2007). The research conducted by Charley Tichenor of the Defense Security Corporation Agency explores the return on investment (ROI) of IA programs. Mr. Tichenor focuses on financial ROI in his study but notes that ROI can be measured in four key categories: financial, customer satisfaction, improvement of internal processes, and investment in learning and growth. A survey of the related literature reveals little in the way of quantified IA ROI in the latter three categories. One essay notes that "there is room for future research to improve the ROIA Model to address the ROI of non-financial benefits" (Tichenor 2007).

One practical business scenario involves the efficacy of seals* used to verify website security. Alexei Nikitkov of Brock University writes that "academic research finds that [information assurance] seals potentially meet some of the most acute consumer concerns, but that consumers have inadequate understandings about the seals, and low regard for them" (Nikitkov 2006). Note that the seals purportedly do address the actual consumer concerns about web security but consumers view them in low regard. Nikitkov's research indicates that seals are essentially useless.

* For example, WebTrust, TRUSTe, and VISA.

> For example, Mauldlin and Arunachalam (2002) find that consumers perceive no difference between three providers of web assurance, WebTrust, TRUSTe, and VISA. These researchers find that all seals equally impact customers' intent to purchase even though each seal addresses different dimensions of information risk. Houston and Taylor (1999) and Kovar et al. (2000) find that seals induce false expectations that products will be of higher quality. Kaplan and Nieschwietz (2003) find that consumers form these expectations whether the e-business uses the existing (e.g., WebTrust) fictitious seal (Nikitkov 2006).

The failed implantation of web assurance seals begs us to return to the matter of education. Eugene Spafford, a professor at Purdue University, does little to mask his disappointment of some IA professionals. He suggests that there is too much patchwork in lieu of investing in strategic architecture and high-level planning. Even large organizations that could deploy effective mechanisms choose not to do so. Their decision is the result of an abundance of low cost and convenient alternatives.

He notes that the availability and familiarity of a few common artifacts hassled us to deploy them (or variants) everywhere, even to unsuitable environments. By analogy, what if everything in the society was constructed of bricks because they are cheap, common, and easy to use? Imagine not only homes built of bricks, but everything else from the space shuttle to submarines to medical equipment. Thankfully, other fields have better sense and they choose appropriate tools for important tasks (Spafford 2009).

In other words, slapping a seal on a website does not even feel like a strategic solution. The evidence shows the public does not get it, does not buy it, and does not acknowledge it as credible. Yet, the answer is not obvious; it is not just more technology.

"Thirty years ago, we had computer security problems. And it was a whole different set of technologies. And 30 years from now, we'll have computer security problems. And that's a whole different set of technologies. Clearly, there is more to it than the technology. And we have to address it that way" (McCormick 2007). McCormick's position places more responsibility upon people and the process than the latest tech innovation. Spafford also discusses his concerns with decision makers such as C-level executive, IT personnel, and business users who may well be experts in their respective fields but are not necessarily versed in IA. He argues that these individuals have the wrong attitude and make poor decisions as a result. "Security is an enabler, not a disabler." Too many people view security as a set of rules and constraints that say do not do this, do not do that, you cannot conduct this transaction. Security is an enabler. It says, you can go out and do business with confidence (McCormick 2007). When asked how professionals can be more proactive when designing systems even in the face of conflicting goals from business users. How can we actually apply these best practices? Spafford argues that

> It's a process that you have to continually revisit and iterate on. It's not a point solution ever. And we have too many companies, and too many consultants, who are selling it that way—saying, "I'll make your system secure, if only you do this." And that's simply wrong. It's like leading a healthy lifestyle. You aren't going to be healthy the rest of your life just because you got a vaccination against measles (McCormick 2007).

In addition to adapting the proper attitude and approach toward IA, Spafford suggests we are not asking the right questions. Even seasoned professionals can get pulled into the reactive,

put-a-patch-on-it mentality. Spafford outlines a number of what he deems to be misleading questions:

> How do I secure my commodity operating system against all threats? How do I protect my system with an intrusion-detection system, data loss and prevention tools, firewalls and other techniques? How do I find coding flaws in the system I am using so I can patch them? How do we build multilevel secure systems? Each of these questions implies that it can be answered in a positive, meaningful manner. That is not necessarily the case (Spafford 2009).

One specific example is the continued overuse of the C-programming language, a language that allows buffer overflows and requires arduous and unnecessarily laborious memory management. We have languages that could accomplish many of our goals without the risks inherent in C. But C is ubiquitous, well understood by many programmers, and a frequently taught language in computer science courses. Spafford's brick analogy presents itself directly: just because something is abundant and cheap does not mean it is the best choice. Combined with the notion that security exists to slow down progress and the environment in which IA professionals operate appears even more nebulous to navigate.

According to Spafford, "there is no perfect security in any real system—hardware fails, people make mistakes, and attacks outside our expectations may defeat our protection mechanisms. If an attacker is sufficiently motivated and has enough resources (including time), every system can be defeated in some manner" (Spafford 2009). One possible conclusion: sometimes the building must be torn down rather than mended. As we noted with web assurance seals, patchwork not only yields poor results, but has also been proven to be counterproductive to a firm's credibility.

We have examined the curriculum, the relationship between technologies and cultural issues, as well as a survey of critical publications pertaining to IA. To the latter, we have shown that the future of IA requires attention to several key areas: education, strategic planning, and a rigorous, interrogative posture toward addressing flaws in policy, strategy, and execution. The successful assurance of our information is paramount to the government, business, and cultural stability. As a result, we must vigorously search for ways to improve and innovate. Academics and professionals alike face significant challenges. But every challenge presents an opportunity. And the opportunities for the adept are many and varied.

References

Cegielski, C.G. 2008. Toward the development of an interdisciplinary information assurance curriculum: Knowledge domains and skill sets required of information assurance professionals. *Decision Sciences Journal of Innovative Education*, pp. 29–47.

Grant, I. 2007, September 4. Experts slam UK information assurance strategy. *Computer Weekly*, p. 10.

Houston, R.W. and Taylor, G.K. 1999, Consumer perceptions of CPA WebTrustSM assurances: Evidence of an expectation gap. *International Journal of Auditing*, 3, pp. 89–105.

Kaplan, S.E. and Nieschwietz, R.J. 2003. A web assurance services model of trust for B2C ecommerce. *International Journal of Accounting Information Systems*, 4, pp. 95–114.

Kovar, S.E., Burke, K.G., and Kovar, B.R. 2000. Consumer responses to the CPAWebTrustSM assurance. *Journal of Information Systems*, pp. 1–35.

Mauldin, E. and Arunachalam, V. 2002. An experimental examination of alternative forms of web assurance for business-to-consumer e-commerce. *Journal of Information Systems* (Supplement), pp. 33–54.

McCormick, J. 2007, July. Security: A business enabler, not disabler. *Baseline*, pp. 41–42.
Nikitkov, A. 2006. Information assurance seals: How they impact consumer purchasing behavior. *Journal of Information Systems*, pp. 1–17.
Spafford, E.H. 2009. Answering the wrong questions is no answer. *Communications of the ACM*, pp. 22–24.
Tichenor, C. 2007. Education and training: A model to quantify the return on investment of information assurance. *The DISAM Journal*, pp. 125–134.
Wyatt, B. 2005, November 15. The case for professionalism in IT security. *Computer Weekly*, p. 78.

Security Awareness Training

Chapter 12

Protecting Us from Us

Human Firewall Vulnerability Assessments

Ken M. Shaurette and Tom Schleppenbach

Contents

Organizations spend millions each year to defend and protect private data from compromise. Intrusion prevention and detection systems and technology have become very sophisticated, and hacking attempts have become equally advanced. There is a continuous war being waged as the average user plugs along, using Facebook, online banking, and Internet searches for genealogy information or whatever hobby of interest.

Even with all the technology deployed and policies and procedures written, a consistent weak link to data privacy continues to be "Us," people who are busy, easily distracted, get tired, or simply forget and do something foolish that leaves our personal information or that of customers at risk. The control closest to these data is often both the best place to implement good security and the weakest link. That protection is the human firewall.

The attack technique that works best against the human control and has fooled many of us at some point is called social engineering. Why does this attack technique, which tricks innocent people into revealing information that they typically know, should not continue to work? Just about anyone with an e-mail account has seen these types of attacks known as phishing, pharming, or combinations that use both phone and e-mail called vishing (voice and phish combined).

Many organizations have policies in place instructing employees not to click on links in e-mail, not to accept or respond to e-mail from unsolicited people, or not to enter any information into untrusted or insecure web links.

The Story

All this talk of social engineering reminds me of a story (an adventure) that I had with my kids a few years ago.

In a quest to stay in shape, the author rides a road bike, often taking 15–20 mile journeys several days a week during the warm months in Wisconsin. One year, a decision was made to put the training to a challenge; so, an entry was made in a 100-mile race that involved several hills. There was no expectation of winning, but finishing and maybe not being the last one to come across the finish line might be nice.

Planning is everything with this sort of thing, and this is where the first error (vulnerability) was made. After filling out the entry form and paying the associated fees, it was identified that on the weekend of the race that the author's spouse had to work. That suddenly shifted the responsibility for taking care of the young children from the spouse to the author.

After a little research, it was determined that near the race starting point, there was a campground and fortunately, there was also a youth who was of babysitting age. Thinking that this could be a fun adventure for the kids, it was decided to camp out; this author could compete in the race and a memorable weekend could be had by all. Memorable—ouch that might turn out to be an understatement.

The day before the race, the camping started out well with everything going pretty much as planned. The camping was at a Wisconsin state park, the kids were having a blast, and the preparation could be started for the 100-mile race the following day. The next morning, instructions were given to the oldest child on dos and don'ts, and then off the author went to the race. The race was certainly a challenge, but the goals were accomplished as intended. The race was finished and the author did not finish in last place.

Back at the campsite, everything was still going pretty well; the kids did everything they were suppose to, and it was time to just relax. Just as evening approached, it started raining; so, we opened the back of our vehicle, used the liftgate to sit under, and returned our gear to the vehicle to avoid anything getting wet. The author as could be expected was a bit tired and fatigued from the race and decided to turn in early, but not before instructing the oldest to keep things in order and have everyone turn in at a reasonable time.

Sometime in the middle of the evening, there was noise outside our tent, Raccoons! The author got up and shoed them off, noticing that the cooler with our food was not secured and the liftgate on our vehicle was still open. He placed a couple of heavy logs on the cooler to secure it and after a quick inspection of the vehicle, the liftgate was shut.

As morning approached, a scratching noise could be heard outside. The author once again stepped outside the tent only to find he had accidentally closed the liftgate on the vehicle when a small raccoon was still inside. It must have been hidden under a seat and went unnoticed before the liftgate was shut. At first, the kids thought this was kind of funny and admittedly I thought so myself. However, the small raccoon started to panic and began running all over the vehicle. The keys for the vehicle happened to be on the front seat and sure enough, the raccoon stepped on the lock switch. Now, we were all locked out of the vehicle with a panicked raccoon inside our locked vehicle. By now, the baby raccoon had begun urinating all over everything on the inside of the

vehicle. It is uncertain if it was the young raccoon's nerves or the large bag of candy that it had eaten that had given him the rather soft bowel movements.

The cell phone was also in the vehicle; there were no spare keys and the raccoon had full control of our vehicle. We decided to make a trip to the park ranger station to explain what had happened and ask if he could place a call to the author's spouse to provide help. Think for a moment; it would not take that long, but did you ever try to explain something this unusual to a Wisconsin Department of Natural Resources Park Ranger? With that slight look of disbelief, the park ranger looked at me quizzically and said "you have a small raccoon locked in your vehicle with the keys in it? How did this happen again?" Trust me, you cannot make this stuff up; it is one of those things that just happens. Finally, the park ranger allowed us to use the phone and now, we had to explain the whole story again to a slightly skeptical spouse. Our vehicle was equipped with "On-Star" technology so she called them and just imagine what they asked. Yup, you guessed it they wanted to hear the whole story about the amazing candy-high baby raccoon locked in the car. The support people at On-Star verified the location of the vehicle and were able to unlock it for us, and certainly laughing and telling everyone in customer support all about it.

So we all went back to the vehicle, opened it up fully, expecting this poor scared baby raccoon to come busting out of the car running for the hills. Well, that did not happen. We spent more than an hour coaxing him with a stick to get him to finally come out from under the seat and vacate the vehicle. It took about another two hours just to clean up everything so that we could go inside the vehicle. For such a small animal, he made one big mess.

People live busy lives, working, raising families, doing chores around the house, and taking care of the kids when the spouse has to work—even when we plan to ride in a long bike race. This can leave us vulnerable where we are distracted and tired and maybe we do not sleep too well. The next thing we know is we are at work and we get an e-mail from "Jim" (noise outside the tent) or it comes from one of our credit card merchant asking questions. In a weak moment, we go ahead, answer the questions (quickly check the car, close the doors), and in a split second, we have enabled a breach of personal information or we have put the entire organization's private and protected customer information at risk.

. . . So what can we do to Protect Us from Us?

Implement the Human Firewall

You have responded to industry standards and regulations such as the payment card industry's (PCI) data security standard (DSS) that calls for regular vulnerability assessments. Your external penetration testing shows that your organization has implemented reasonably well all the bells and whistles related to protecting the perimeter from the Internet. The network firewalls, intrusion prevention, and regularly scheduled repeat testing are all in place that makes you feel your security program world class. You have invested in the best technology that your budget can buy. Did you forget to invest in the link closest to your data, your employees?

Testing the human firewall for vulnerability is fast becoming popular. This testing is also often referred to as social engineering. Even with all the information security awareness training that organizations do, these vulnerability tests have a very high success rate and are a good payback. Mark Chapman, the founder of Phisline.com, is quoted as saying, "Companies need more than just education. Organizations that use a strong information security program along with strategic campaigns to test their people (the human firewall) have begun to mature their risk management.

They can gain valuable insights into the vulnerabilities of the people, going beyond just the process and technology protecting your critical information assets."

Too often, awareness programs are conducted once a year for a few hours rather than an organization putting together a proactive ongoing program that delivers awareness messages throughout the year. An organization using the human firewall vulnerability assessment with e-mail from social engineering campaigns cannot only test the effectiveness of the organization's information security awareness program, but can also use the knowledge gained to share with the employees to make the message even more real. The human vulnerability assessment could include the following activities.

Public Information Gathering

With the popular "targeted attacks" that occur today and are often more successful, an attacker does not just randomly begin sending out e-mails. Like most thieves, the target is researched to gather more information. Thieves will look for your personal habits such as when you go to work or come home. They find targets that appear most profitable to attack because of the expensive clothes or jewelry they wear, or the car they drive. Cyber thieves and attackers will use public resources and gather public information about the organization and organization's employees. This is a basic form of "Google hacking," which by using Internet search engines, social networking sites, and various other data-gathering methods, information that is in the public domain about an organization or the organization's employees and customers can be gathered. This information can be very helpful in making social engineering attacks more successful. For example, sending a message that comes from the organization's information technology (IT) manager by name may have a greater impact than just a letter with an anonymous signature from an unknown name. While a lot of information may be public, an organization and its employees should be aware of the kind and amount of data that can be obtained about them or about the organization that may be useful to an attacker.

The key to testing the "human firewall" is the use of skills for establishing a trust relationship, typically between individuals who have previously never met. The act of social engineering is designed to attempt to obtain otherwise secure data by convincing an individual into revealing nonpublic information. This can be done by masquerading as a privileged or an authorized employee, or using other means to gain a person's trust. In performing social engineering, a common thread often occurs whereby an individual is tricked, cajoled, awed, persuaded, or otherwise convinced that providing the requested information is an appropriate course of action. Social engineering is often based on trickery and misleading activities that encourage employees to release information that may be of value to assist in other attack vectors or simply to access customer data directly such as account numbers and passwords, often using a computer system. For example, an employee in an enterprise may be tricked into revealing his or someone's password for access to a sensitive application.

Types of Social Engineering to Test Employee Security Awareness

The objective of using e-mail to test the human firewall for vulnerability can be to test the effectiveness of the security awareness program as well as to test the organization's technology for effectively blocking the "spam" type of message often related to a mass e-mail attack. As noted earlier, the attacker may gather business intelligence publicly available on the Internet. Some of this Internet information in and of itself may not be confidential, but can be very revealing about

the organization and is often valuable when conducting more intrusive levels of social engineering or other penetration attacks.

The e-mail of a social engineering campaign can use these data from the public information gathering to build a targeted attack that will even further increase its potential for success. Along with providing an organization a better idea of their data footprint on the Internet, the e-mail type of the social engineering attack will evaluate the response of the organization employees to a specially crafted e-mail attack. The activities can be designed to determine how susceptible personnel are to the e-mail method of information reconnaissance. It just takes a single moment when one lets his guard down, forgets to do something, or is so tired and something gets missed. Often, people are just so busy receiving an e-mail that at a glance looks legitimate that they feel they can just do what it says quickly and get it off their to-do list.

Social engineering takes advantage of human behavior, will catch you when you least expect, maybe when you are tired or rushed, or just want to do the best thing you can. Like taking care of the kids in the story while still accomplishing your goal of getting the work done, finishing the race, and not in the last place. That high-tech firewall (the vehicle) you just paid lots of money to install would not provide much value (protecting your gear) if the doors are left open. In the case of technology, your users could be getting tricked into clicking on a malicious link, leaving too much information on social networking sites, or in other places of the public domain. They download malicious software because an e-mail told them that they would win money or their computer needed a scan to find a new virus.

Social engineering is essentially the art of exploiting the human psychology using trickery, rather than by breaking in or using technical tools. It is not as sophisticated as technical hacking, but it can be very devious. For example, instead of looking for vulnerabilities in your network, a social engineer might call an employee and pose as an IT support person to trick the employee into divulging his password. Another very successful technique is to send an e-mail that appears to come from another trusted employee asking that they must test a new company website, maybe an all-employee survey, and the site is really a fake that captures a person's passwords along with deploying a small piece of malware to capture future keystrokes.

The concepts and techniques used in social engineering have been around as long as there have been scam artists of any sort. The now infamous hacker "Kevin Mitnick" was not a highly technical specialist; he understood people and preyed on them using his socializing and scamming skills. Social engineering has proven to be a very successful way for a criminal to "get inside" your organization. Once a social engineer has a trusted employee's password, he can simply log in and snoop around for sensitive data. He can even masquerade as that employee to gain the trust of other employees often gaining access to confidential account information and lots of company data.

The common theme of social engineering threats is that they prey on human weakness and not the technical side of our computing environments. While not new, in tougher times as we are in today, attackers seek out people's weaknesses (such as the fact I was tired in my story), or consider an employee's inherent trusting nature and desire to be helpful or their greed to exploit these vulnerabilities. A Federal Bureau of Investigation (FBI) friend, who will remain nameless, once said many years ago, "Many viruses would go nowhere if it weren't for male stupidity!" Remember some of the old virus signatures that counted on a human for them to spread? They carried names such as anna.kornikova.bat or nakedwife.exe.

The organization's security program must include technical education (e.g., training for configuration of the firewall) for the technician and security awareness (e.g., knowledge of policy, safe computing practices, acceptable use, and incident response) for everyone. While improving and maintaining technology remains crucial, organizations must continue to better prepare

and equip their staff to handle confidential information. Let us call it installing a more effective "human firewall." The staff will better accept change management and be less resistant to controls and documentation. Awareness is a powerful defense for protecting networks and also prevents humans from being exploited. It can make them feel like they are a valuable and important part of the company as well.

Security awareness must go beyond a simple annual workshop to present security policy and procedures. The security program must do more than simply inform users not to click on an attachment. Educating the end users properly is often the only way to protect against many of today's targeted attacks. Teach the users how to handle the data. We call it Ken's Golden Security Rule: "Treat all data you work with like it is data about yourself or one of your family members. Handle the data you work with like it is your own personal data." When employees have an attitude that the data they work with are personal, they are much more likely to understand and respect controls that are in place. They will use improved common sense and the best controls possible, which may simply be paying attention to the organization policy and procedures to protect the data better. You want every person to reach the point where information security is part of their DNA. They are not following security practices; they are doing their normal job and by the nature of how they perform their duties, it is done securely. Technicians do not build a secure network, they configure the network, and normal configuration includes making sure that this is done in a secure manner; the same concept goes for writing programs, developing websites, and virtually any computing job.

There are even expectations now such as with Internet banking to provide better education and awareness to a bank's customers (FFIEC Supplement to Authentication in an Internet Banking Environment FIL-50-2011). Even your customers have become your responsibility and another link in the human firewall to protect their own data and the reputation of the bank. Educating them protects them and you.

Organizations continue to ask the question: where does our responsibility end in the education of employees against threats? This is a difficult question in today's world; the mobile workforce requires access to key business functions and data to perform a job. In reality, the technology is just not there yet to protect the data effectively down to the personal smart phone level. It is getting better and by the time this book is published, there will likely be improved solutions. So, in some cases, a technology solution may not be available to protect the data without the human firewall implemented with security awareness training. A look at enterprise risks and the organization's key risk indicators within the business may help answer that question. First define the level of risk and then the probability of data loss without education. A few of the risk and key risk indicators may include:

- How large is the mobile workforce? Define the overall risk of employees not being trained properly.
- What data access is really required remotely? Deploy new technology to maximize efficiency, integrity, and safeguarding the data if available.
- Remote and mobile solutions available to the workforce, effectively deploying and utilizing technology.
- Continuous change in the available functions and features in a handheld device. Do you put controls in place that make a device less functional and as a result, less valuable?
- Manage the turnover rate. Instability of the workforce leads to ever-decreasing productivity as skill sets are never fully developed and/or trained to protect the data.
- Adhere to all key regulations governing the organization of data. If the data are compromised, the consequences may include fines, penalities, regulatory interventions, and possibly additional regulations in the future.

(Note: it is the opinion of the author that we would have less regulation if we as organizations would have considered information security rather than just profit margin in the first place.)

Some Human Firewall Testing Results

Information security awareness testing exercises using e-mail (phishing campaigns) and telephone social engineering tests for vulnerabilities of the human firewall are a couple of the more common assessments. How much research and care is done in crafting the e-mail or conducting the phone campaign will dictate their success rate. In most cases, organizations have effectively trained their staff against some attacks; the telephone is less successful than it used to be, while e-mail tests tend to still have too high of a success rate. However, e-mail attacks that have been in the public domain have become less successful, while targeted attacks that use information about their company or the individual target are still quite successful. Remember that all it really takes is one, the right one, and your company is owned by the attacker.

In e-mail phishing, an attacker or tester would set up a spoofed website and then send spoofed e-mails to a sample of employees selected by the organization. The objective is to lure the employee to the spoofed website to click on the embedded web links or divulge inappropriate information. This may be done using a variety of techniques such as masquerading as a member of the company, posing as a member of the IT staff attempting to solicit "help" for an ongoing test, or posing as a member of human resources (HR) attempting to get the users to sign up for a "prize."

Many organizations struggle to define the parameters of performing the actual security awareness exercise largely because the approach looks so legitimate that there is a fear of a high success rate. The key is to use the exercise as a tool to provide more effective training to the employee base to teach the staff why the exercise was successful, point out some of the tricks attackers use, and what employees can be aware of. There is a caution that must be exercised when testing. You cannot use tests that make employees so paranoid that business processes are suddenly no longer being performed effectively, such as an employee refusing inquiries from the internal staff when the information is really needed.

The other less effective social engineering technique is telephone pretexting. In telephone pretexting, an attacker or a tester would call the targeted employees. In many cases, employee information is freely available using the key information searches on the Internet or through social media sites. The objective is to obtain sufficient information from the individuals called to gain unauthorized confidential information or even system access. The effective attackers or the tester has an ability to pose as someone else that an employee may trust. They are good actors. Typically, the longer an attacker can stay on the phone with an individual, the more trust is gained and information gathered. While still somewhat effective, it seems that people have become more sensitive to the interruption of a phone call and are less likely to give out information. Employees should be trained to identify when a caller is trying to gain confidential information without being properly vetted or identified.

Security Awareness Training High-Level Outline

In today's industry, most organizations have begun to provide some type of security awareness training. Many larger organizations have built very elaborate programs and have implemented regular testing with the human firewall vulnerability assessments that we have discussed in this

chapter. At a high level, these programs should have the following basic principles to provide training to the staff to increase the awareness of security and data privacy:

1. When an employee is first hired, an awareness of all company policy and procedures and their roles and responsibilities is important, and often, this is the first signature for them to acknowledge the receipt and the opportunity to understand and ask questions about information security.
2. Mandatory annual training, which should also include a curriculum and planning along with keeping a roster with employees, again signing an acknowledgment that they have received training and copies of the policy, and given the opportunity to ask questions.
3. Ongoing security alerts and messages are delivered to the employees throughout the year, maybe in some cases even based on their job roles. The message may consist of posters in the cafeteria and the hallway, e-mails about recently discovered scams, and alerts of attacks that are taking place in the real world.
4. Periodic meetings are held with personnel when there are security concerns, fraud alerts, or new procedures related to security and privacy. Including in an employee's performance review the responsibility to protect confidential information is a great technique. Performance reviews and messages delivered at staff meetings can be effective.
5. Create a regular information security section in the company newsletter. Provide something similar to customers to increase their awareness.
6. Establish a quarterly security memo, maybe even as frequent as monthly hints and tips to all employees regarding personal and workplace safety. Be careful not to make the message cause them to become immune to your message. Crying wolf too often got all the sheep killed.

Conclusion

It is our belief that we will always be faced with vulnerabilities surrounding the human firewall, and attackers will always be looking for new and easy ways to attack our organizations. Our best chance to be the most successful in protecting company information is to prepare the appropriate mix of employee awareness with technology and continuously look for ways to increase the maturity of the information security program. When we finally make information security part of every person and the DNA of the organization, the success realized by attacks against the human firewall will become less successful.

DOMAIN

4

APPLICATION DEVELOPMENT SECURITY

Application Issues

Chapter 13

Service-Oriented Architecture

Walter B. Williams

Contents

What Is a Service-Oriented Architecture?

This chapter is an enhanced version of a chapter with the same title published in the 2012 edition of the *Information Security Management Handbook*. Over the past few years, there has been a renewed interest in attacking Web Services, with new tools created that permit the proper penetration testing of application Web Services. The original text of the chapter is included for completeness, especially for those who do not have the 2012 edition and those who want a comprehensive overview of a service-oriented architecture (SOA), how to analyze it for vulnerabilities, and how to defend it from attack.

To properly understand what an SOA is, it is helpful to understand where it sits as an architecture and the problem that an SOA tries to solve. An SOA, first, is a software architecture. Software architectures are the structures of software, their elements, and the relationship between them. SOA was developed by applying the concept of software architecture to solve the problem of how to connect disparate systems in a way that they could function together in a systematic manner.

Each application in an SOA is treated as if it can only perform one specialized function or as if it were an element in the larger software. This specialized function is called a service. Thus, each service functions as an element of the SOA, where the structure and the defined relationship between the elements must transcend the application and often the corporate boundaries.

Distributed Computing and Services

SOAs are not the only approach to integrate disparate computer applications into a larger unified system. A message-oriented model (MOM) was developed with the idea of distributing the components of a system among the existing and emerging applications and platforms. Messages were used to connect these systems, implemented through specialized platforms called message queues. Message queues are very reliable, but required not only a system to manage them but also a staff to maintain the messaging system. Most importantly, support for the message queues had to be built into all the components of the application, restricting the implementation of a system built on an MOM architecture to the internal needs of a single corporate infrastructure.

An Architecture Based upon Services

As businesses realized that other companies' applications provided a better solution to meet their requirements than their own internally developed systems, they realized that they could provide better products to their customers if they found a way to leverage the capabilities of the expertise of other corporations for components that were not a part of the core competency of their own company.

As an example, a travel company might recognize that MapQuest's maps or Google Maps were vastly superior to their own. Since their home-grown map application was not how they brought in new business, but still provided a desired component to their customers, it would benefit everyone if there was a way to leverage the capabilities of the other company's superior product.

SOAs abstract the diverse applications, protocols, systems, and data into four key concepts:

- Application front end
- Service
- Service repository
- Service bus

An application front end is the owner of the business process the application provides and other services that it can use. A service is an implementation that provides business, logic and data, and a service contract that specifies functionality, usage, constraints that must be observed by any client, and an interface that exposes the functionality. A service repository stores the service contract of the individual services. A service bus interconnects the application front ends and the services.

There are various technologies that can be applied in the implementation of an SOA. Some of these are more appropriate for internal enterprise-specific projects and others that may be applied to any project with any scope. Precisely because of the common architecture, there are issues that transcend the individual technologies. These must be considered when establishing a security architecture appropriate for the protection of the business and its objectives as supported by the specific implementation of an SOA.

Process Integrity

Data integrity is the key factor in the success of many SOAs, as the data are transferred from service to service. However, the integrity of the data is not sufficient to guarantee that the service returns the correct results due to the complexity of business processes that span multiple systems and often involve multiple corporations. Process integrity involves not only the integrity of the assets, but also their utility throughout the architecture.

The principles of entity, domain, and referential integrity are borrowed, where appropriate, from relational databases. Entity integrity requires that each row in a table can be uniquely identified. Domain integrity requires that certain data can be within a valid range, such as the date of purchase of an item not being in the future or before the date of which the item was first placed on sale. Referential integrity refers to the validity of the relationship between different data sets, preserving as an example the names of the residents (Figure 13.1) in relationship to their appropriate residence (Figure 13.2).

Where the data must be processed across multiple systems through the use of their services, there is a risk of inconsistencies that impact the validity of both the data and the use of the data in all services. There might be technical failures, business exceptions, and special cases that impact

Figure 13.1 Web 2.0 and SOA (Web Services).

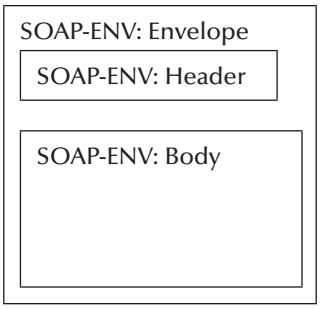


Figure 13.2 SOAP structure.

the integrity of the process. Since the process is not centrally controlled, the impact of the failure of any particular component may be high.

There have been a number of techniques applied to solve this problem, each with their own merits and issues. The most common approach is to apply logging and tracing. This is similar to the use of transaction logs in a transactional system, allowing for recovery to a valid state in the event of a failure. The problem is that it is difficult for logging and tracing to resolve issues that relate to logical processes that span the multiple systems involved in an SOA.

Online transaction processing systems were developed to enable large numbers of users to manipulate shared data concurrently. Such systems are based upon the concept of transactions on a unit of work that transforms the data from one state to another. A unit of work is presumed to be atomic, or indivisible, consistent (moving from one consistent state to another), in isolation where no process is visible to others, and durable where committed updates are permanent. Such systems depend upon a central control mechanism that resolves conflicts. Such a central control mechanism is often unavailable to an SOA that leverages services from multiple organizations.

The two-phase commit protocol (2PC) was developed to allow online transactional processing to span multiple systems. A transaction coordinator is implemented as part of a transaction monitor. This enforces that in the first phase of the processing, all relevant locks have been acquired and that the state of the data has been properly captured. Dependent upon the results of this examination, the transaction coordinator informs the participating systems if they should roll the transaction forward or backward. These systems "vote" on how to handle these data and a single abort vote will cause the entire transaction to be rolled back.

All these mechanisms of tracking changes to the data are predicated on some assumptions that often do not apply in an SOA. One is that it is possible to ensure the isolation to these data; another is that the transaction is short term. Neither can be assumed in an SOA, where the various services may be entirely ignorant of access and the use of data by other services, and transactions are often long lived. SOAs are also often implemented on discontinuous networks and none of the above mechanisms are designed to operate under such conditions.

Two techniques scale well to address the issues of process integrity. The first is that of persistent queues with transactional steps. Persistent queues, which follow these data, can guarantee the consistency of the individual steps of the process, where errors are handled by a de-queue and the error is returned with these data. Such systems depend heavily upon the presence of a message queuing system and are more often implemented in the internal SOAs where such systems are present.

The second is transactional chains and compensation. Complex workflows are created through individual process steps (transaction chains), where failures are dealt with compensating

transactions that logically undo a prior transaction. Each transaction is made visible to each service, so that the data may be made available to a compensating control.

To permit these workflows to transcend the boundaries and controls of a single corporation, the Business Process Execution Language (BPEL) was developed. It is based on Business Process Markup Language (BPML), an extension of Extensible Markup Language (XML) as a formal metalanguage for modeling business processes that provide an abstract execution model for describing collaborations and transactions. BPEL allows the capability to

- Describe the logic of business processes through the composition of services
- Compose larger processes from smaller processes
- Handle synchronous and asynchronous operations
- Invoke services in series or in parallel
- Selectively compensate completed activities in the case of failure
- Maintain interruptible long-term transactional systems
- Resume interrupted or failed activities
- Route the incoming messages to the appropriate service
- Correlate requests within and across business processes
- Schedule activities based on predefined times
- Define the order of execution
- Handle both the message and time-related events

With BPEL, business processes can be described in two distinct ways:

1. As executable business processes wherein the exact details of the process are defined and follow the orchestration paradigm.
2. As abstract business processes or the public message exchange. These are not executable and follow the choreography paradigm.

In the orchestration paradigm, a central process, which can be another service, takes control and coordinates the execution of different operations in the services involved in the operation.

The choreography paradigm does not rely upon a central process, but is a peer-based system where each service knows when to execute its operations and with what other services to interact.

Enterprise Service Bus

An enterprise service bus is the technical infrastructure of the distributed environment. It is composed of an XML-based communications protocol with a message-oriented middleware core to perform the actual message delivery. There are a variety of message bus frameworks in common use. Some, such as Enterprise Java Beans within the J2EE specification and Microsoft's .NET, are based on the capabilities of an application architecture. Others rely upon either message queues or object-oriented communication infrastructures such as Common Object Request Broker Architecture (CORBA). In practice, a successful enterprise service bus is not a single product, no matter how flexible or how many communication protocols it supports, but one that supports accessing services on a meta level, which can leverage the capacities of all application architectures, allowing .net, Enterprise Java Beans, and other diverse applications to function within a single business process.

Web Services and SOA: An Alternative Service Bus

Unlike an internal SOA, a Web Services-based SOA cannot rely upon a single monolithic service bus. To this end, Web Services are based on slightly different principles than a traditional SOA. Each service needs to be reusable, stateless, autonomous, abstract, discoverable, and loosely coupled. Instead of a form of service bus, you have a service integration layer, which operates as a logical or a virtual service bus.

The services engage with this service integration layer and with each other through a formal contract that defines the terms of information exchanged and provides supplemental service description. Since services need to be discoverable, they make themselves known through a service provider. The services also need to know which service to call and thus they will have a service requester. These roles can be and often will be reversed, as the role of the service changes within the larger workflow from the client to the server. There may be more than one service provider through which a workflow must pass before it arrives at its ultimate destination; these are called intermediary services. Intermediary services may or may not do more than discover the next step in the workflow, depending upon the nature of the service and the contract it has as a service provider.

Web Services tend to be broken down into one of the set of roles:

- Utility service
- Business service
- Controller service
- Proxy service
- Wrapper service
- Coordination service
- Process service

The nature of the service offered, how to engage it, and the results to be expected are all defined in a specialized XML document present on the service provider. This document will be written in the Web Services Description Language (WSDL). The WSDL functions as the integration layer of the web service, providing the basis for other services to discover how to engage the particular service.

Some implementations of Web Services will register themselves in a central registry of services using a specification called the Universal Description, Discovery, and Integration (UDDI). UDDI repositories provide a market place of generic services and are often hosted by major corporations.

Many protocols can be used and are used to communicate between the various web services over Transmission Control Protocol (TCP)/Internet Protocol (IP). The most common is the Simple Object Access Protocol (SOAP). SOAP provides a standard message format that consists of an XML document capable of hosting remote procedure call (RPC) and document-centric data. SOAP can be easily leveraged by both synchronous and asynchronous data exchange models. SOAP, as a raw protocol set, does not define a preset language, allowing the application designer to create a language specific to the architecture.

An alternative to SOAP-based Web Services is the representative state transfer (ReST)-based web services. ReST leverages the existing vocabulary of the Hypertext Transfer Protocol (HTTP) or other robust application layer protocols with the existing vocabularies. SOAs based on ReST are easier to implement, but are less flexible.

With second-generation Web Services, or Web 2.0, a limited vocabulary was developed to provide a common framework for common constructs that all business services rely upon, such as business process or workflow, security, reliability, policies, and attachments. These standards, managed by the Organization for the Advancement of Structured Information Standards (OASIS), are called the WS or Web Services standards. The most common of these protocols are defined below.

Web Services Description Language

WSDL, currently in its second revision, serves to define a Web Service to service discoverers. It does this in two ways: one abstract and the other concrete. In the abstract, a WSDL will describe the messages the service sends and receives, typically using XML schema. It defines operations associating message-exchange patterns with one or more messages. Message-exchange patterns are the sequence and cardinality of messages sent and received. At the concrete level, a WSDL will specify bindings of transport and wire format details for an interface, and associates a network address with a binding.

Simple Object Access Protocol

SOAP is probably the most commonly used protocol within an SOA because of its versatility. It has a very simple structure to it, being composed of an envelope, header, and body.

The envelope is a construct that defines the overall framework for interpreting an SOAP message, in essence defining the vocabulary, who should deal with it in whole or in part, and what parts if any are mandatory.

The header carries a representation of a resource that is needed to process the SOAP message, but that cannot be obtained through the uniform resource identifier (URI) for the resource carried within the message.

The header and the envelope allow SOAP to provide a flexible and custom workflow implementation for the SOA. However, this flexibility comes at a cost of lower performance as each SOAP message must be parsed so that the vocabulary may be learnt and its instructions may be followed appropriately by the service.

Representative State Transfer

ReST is used by SOAs where performance considerations outweigh flexibility and security. ReST depends upon six fundamental architectural assumptions in the use of HTTP and other protocols:

Clients are separated from servers by a uniform interface, with clients remaining unconcerned with data storage and servers unconcerned with a user interface. This preserves the portability of clients and the scalability of servers.

No client information is stored on the server between requests, providing a stateless environment. The server can be stateful, as needed, providing reliability and scalability.

Clients may cache information from the server and may reuse this in preference to fresh data from servers.

Clients cannot determine if they are connected directly to the terminal service provider or to an intermediary service provider, allowing scalability and load balancing.

Clients can have their code extended on a temporary basis by servers. Examples include JavaScript run within a browser or Java applets.

Clients and servers have a uniform interface consisting of the identification of resources, usually through URI, manipulations of these resources, self-descriptive messages, most often through multipurpose Internet mail extensions (MIME) types, and the use of hypermedia as the engine of application state.

Distributed Component Object Model

Distributed Component Object Model (DCOM) is a set of remote procedural call libraries designed for the Microsoft Windows operating system. DCOM is often used within an enterprise-specific SOA, and is an alternative to SOAP or ReST.

Common Object Request Broker Architecture

CORBA is a platform-agnostic method of remote procedural calls with predefined mappings into many common languages designed to allow different applications written in different languages to interchange instructions and data. CORBA is often used within an enterprise-specific SOA, and is an alternative to SOAP or ReST.

Data Distribution Service

Data distribution service (DDS) was developed to provide a mechanism to distribute the data in a publish/subscribe model. It is an alternative to SOAP or ReST, but is rarely used in modern implementations.

Windows Communication Framework

Windows Communication Framework (WCF) is a Microsoft-specific technology designed for building web services using Microsoft products. Until recently, Microsoft technology did not support the key web services technologies such as Security Assertion Markup Language (SAML), and relied upon WCF to provide proprietary alternatives. It is a replacement for DCOM and is an alternative to SOAP and ReST.

WS-Coordination

WS-Coordination is a defined set of instructions to coordinate the behavior of various web services within the SOA. It is used to maintain process integrity and functionality. WS-Coordination can manage any of the protocols typically used to call services and provide them with the required information.

WS-Transaction

WS-Transaction is a defined set of instructions to handle atomic (individual) transactions and business transactions for long-term operations. WS-Transaction is an alternative to BPEL.

BPEL for Web Services

BPEL as has been a very successful language for defining the workflow in SOA specific to enterprises. BPEL–WS (also written as WS–BPEL) is just an extension of BPEL for web services.

WS-Security

WS-Security is a suite of standards that provide the security layer of a Web 2.0 SOA. The suite consists of WS-Policy, SAML, XML signature, and XML encryption. Although WS-Trust, WS-Authorization, WS-Secure conversation, and WS-Privacy also exist, they are less common (Figure 13.3).

Security Assertion Markup Language

SAML provides the authentication service for web services-based SOA. It is not mandatory, and not supported by all platforms and vendors. There are major compatibility issues between the different versions of SAML, where products that support SAML 2.0 will also not likely support SAML 1.1. SAML is commonly used to extend trust boundaries, allowing for federated identity. SAML assertions will contain the information necessary to provide both authentication and authorization.

.NET Passport

.NET Passport is a Microsoft propriety alternative to SAML, which was not supported in Microsoft products until recently.

WS-DigitalSignature	WS-Encryption	XMLKMS (Key management service)
WS-SecureConversation	WS-Federation (SAML, .NET passport)	WS-Authorization
WS-Policy	WS-Trust	WS-Privacy
WS-Security		
SOAP		
SSL/TLS		

Figure 13.3 WS-Security sack.

XML Encryption

XML encryption provides XML with the structure to encrypt the whole of a separate XML document, part of an XML document, or the attachment to an XML document. The most common tags are

- Encrypted data
- Encrypted data schema
- Encrypted type
- Encryption method
- Encryption properties, which will likely contain CipherData, CipherValue, and/or CipherReference
- KeyInfo
- Encrypted key
- Reference list (where multiple items have been encrypted with the same key)

XML Digital Signature

It provides the use of digital signatures in the signing of the whole, part, or attachments to XML documents. XML digital signatures can leverage certificates issued by a public key infrastructure.

WS-Policy

WS-Policy is often considered to be a component of WS-Security as it is used to communicate the security policy of the web service much like the WSDL that is used to communicate the rules of engagement for the service. Like the WSDL, it defines those rules that must be followed and those that are optional, but in a hierarchical manner wherein some or all the child objects must be satisfied to comply with the defined policy.

Attacking SOAs

Despite very impressive security controls to guarantee the integrity of the data, the utility of the process, the confidentiality, and the availability of the information, there are numerous ways to successfully attack even the best-protected web service.

Many web services are not authenticated properly; often, authorization is not checked after authentication. Most importantly, the service logic itself may be used in ways that the developers never intended.

To footprint a web service, often, you have to go no further than the UDDI and the WSDL. The UDDI will provide the location of the specific service and the WSDL provides, in unprotected detail, how to engage the service. WSDL can often be discovered by simply appending "?wsdl" to the end of a uniform resource locator (URL). You look in the WSDL for open methods and resources that are unprotected. Service tags, easily found with a <service.*?> regex pattern, define the name of the service, and how to engage it (Figure 13.4).

While any web browser can read and parse the syntax of a WSDL file, using a browser to read the syntax manually is an inefficient way to look for vulnerabilities. There are a variety of free tools

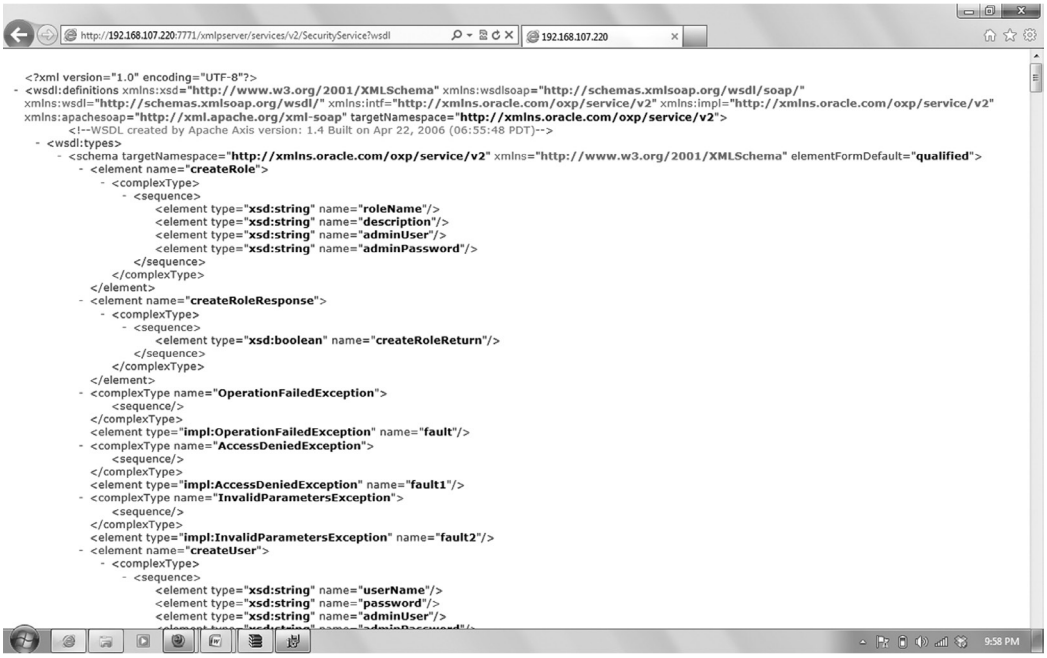


Figure 13.4 Service tags define the name of the service and how to engage it.

that can be leveraged to parse and test web services. One of the most useful tools is WS Digger by McAfee Foundstone. You simply load the URL for the WSDL into the tool (Figure 13.5). Simply selecting the service you wish to exploit will provide details on these data expected and will allow for manual testing of the service's response to the data (Figure 13.6). WS Digger provides the vulnerability analyst or penetration tester with not only the ability to parse out any control, but also to feed the tool with input data for testing.

WSKnight is another WSDL enumeration tool that can make finding vulnerabilities easy. The circle in Figure 13.7 shows a service call in a WSDL that does not require authentication.

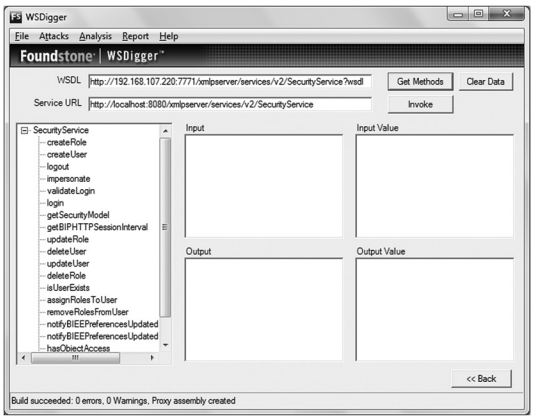


Figure 13.5 One of the most useful tools to parse and test Web Services in WS Digger.

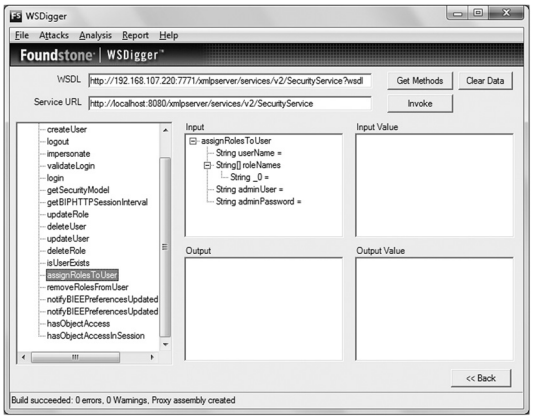


Figure 13.6 Selecting a service to exploit will provide details on data expected and allow for manual testing of the service's response to the data.

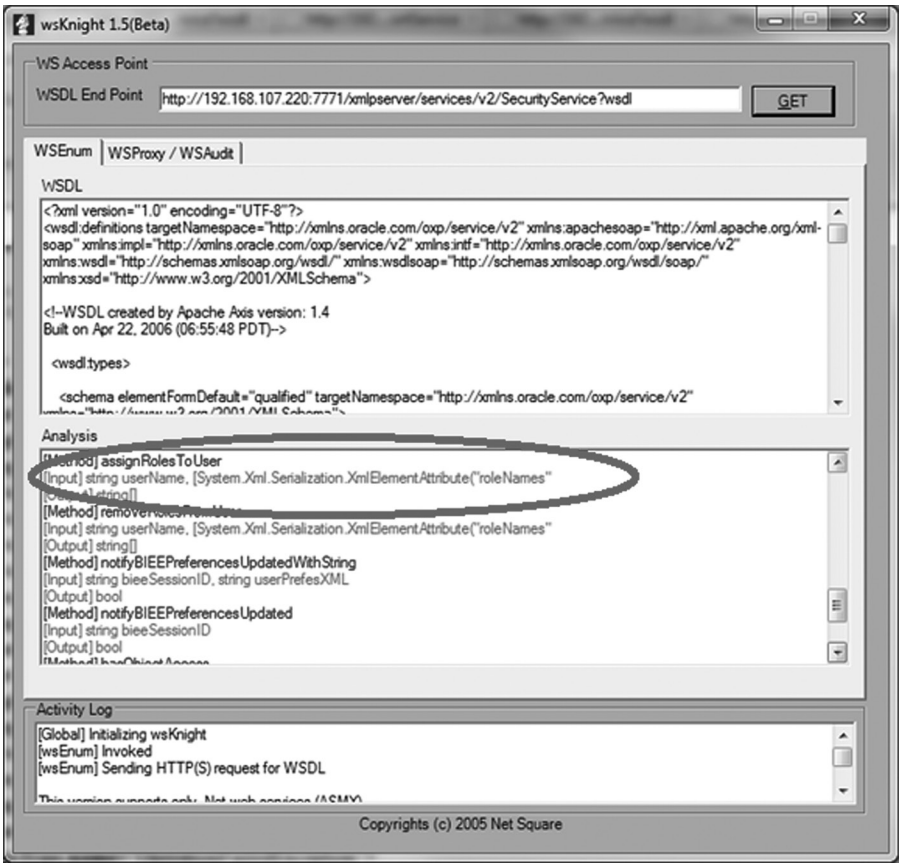


Figure 13.7 WSKnight is a WSDL enumeration tool that can make finding vulnerabilities easy.

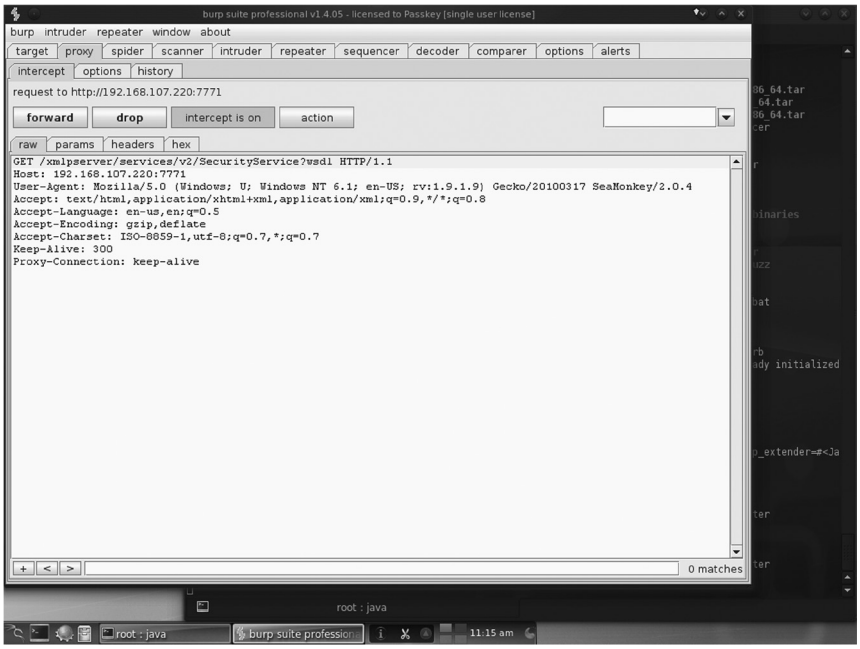


Figure 13.8 Buby is a free BurpSuite plug-in.

Buby is a free BurpSuite plug-in written by Ken Johnson that can not only enumerate a WSDL and manually test it for input handling, but can also scan the file for common vulnerabilities. Buby is a Ruby extension that relies on the jruby, ruby gems, savon, and nokogiri (Figure 13.8).

Unlike with other Web Services testing tools, to test using BurpSuite, you configure BurpSuite as a proxy and direct the application of traffic through BurpSuite. Once you have manipulated the application, the first step is to enumerate the WSDL. Then, using the application logic and WSDL as a source, you tell BurpSuite to form an SOAP request. The request will contain the elements used to test the SOAP request for common vulnerabilities. As BurpSuite is an attack proxy, you can also inject codes such as structured query language (SQL) or XML statements.

If the Web Service is using Document Object Model (DOM), the DOM parser will read the entire XML into the memory before processing. One can overload a DOM processor using complex XML structures and large envelopes. Once the DOM processor is overloaded, it is often possible to run the arbitrary code or take down the service entirely.

Another technique that can be used to elevate privileges is to poison the XML with simple API for XML (SAX) parsing. One can place a tag inside the logic of another tag and overwrite the data without authorization. As an example:

<AccountInformation>
<AccountNumber>XJ12M</AccountNumber><Privileges>15</Privileges> <Account Number>XJ12M</AccountNumber>

If later in the same XML, the following is written, the privileges will be overwritten to become 1:

<AccountNumber>XJ12M</AccountNumber><Privileges>1</Privileges> <Account Number>XJ12M</AccountNumber>

Just as in URL manipulation, the parameters of an SOAP packet can be tampered with if they are not signed. Metacharacters can be injected into the parameters and they break the services logic. Where data in the XML are not signed, one can often cause a denial of service within a service provider by mistyping these data expected in a tag.

Just as with any web application, SQL injection and Lightweight Directory Access Protocol (LDAP) injection may reveal the information of services exposed to the architecture but not designed to provide the information you request through the injection attack. If the underlying database or directory service does not properly check authorization, price lists, and account information, even password hashes and x509v3 private keys may be provided.

When Web Services leverage file systems and the underlying operating system, the SOAP packets can manipulate the web service to providing file data or executing the code in the context of the account under which the web service is running.

In the absence of proper account lockout policies, Web Services accounts can be brute forced by using an SOAP message that contains account after account. Proper logging can record this, but may not alert you to the attack unless the log is monitored.

Web Services messages can have their sessions hijacked when the session is maintained with either cookies or information in the SOAP header.

Defending SOAs

Fortunately, Web Services have many excellent and robust controls that can be applied to diligently and proactively protect the service, data, and business process.

There are three layers to defend: the service container, the service, and the messages between the services.

Secure Sockets Layer (SSL)/Transport Layer Security (TLS) provides excellent and well-understood transport layer security, which can be easily deployed in a Web Service. Unfortunately, if the WS-Routing protocols are used in the Web Services, SSL/TLS will not work.

In the absence of SSL/TLS, or even while using an encrypted transport, XML encryption and XML signature can be leveraged to sign and encrypt the data within the SOAP packets. XML signature contains four major items, although two of the four are optional:

1. Pointers to the object to be signed
2. The actual signature
3. The key (or means to discover the key) (optional)
4. Options tag (optional)

The item to be signed can be internal to the XML document itself, can be an attachment, such as a binary object, or can be another XML document. There are three classifications of XML signatures:

1. Enveloping
2. Enveloped
3. Detached

Enveloping signatures wrap the data that are being signed within the tags. Enveloped signatures refer to a parent XML element through a reference contained within the tag. Detached signatures obviously refer to objects external to the XML which are signed.

XML encryption works much like that of SSL/TLS, where the recipient's public key is used to encrypt the symmetric key used to actually encrypt and decrypt the message. Upon receipt, the private key is then used to decrypt the encryption key, which is then used to decrypt the encrypted portion of the XML. This is both the strength and the weakness of XML encryption. The data confidentiality is maintained, but unless the portion of the XML containing the XML encryption is signed, it is a trivial thing to replace the encrypted data with other encrypted data, and because a public key is known (being public), the recipient will be able to decrypt the strongly substituted encrypted data and process the related workflows.

Depending upon the service, it may not be advisable to expose it to the entire Internet. Access control at the IP address level can provide a measure of assurance. While IP addresses can be spoofed, the permitted IP address has to be known in advance by the attacker. In the absence of that information, even the normally unprotected WSDL file will be unavailable.

To provide for both authentication and the basis of authorization, all web services support multiple authentication frameworks. One that is unique to web services and provides the support for distributed administration is the SAML. SAML allows company A to provide authorization to accounts created and maintained by company B, and a mechanism for services providing these accounts to be made available to the members of company A without having to provide company A with the passwords of their members or customers.

As with every application, the Web Service must be configured to handle errors and exceptions gracefully so that no information regarding the technical details of the error are passed to the service caller, but are logged so that the support staff can debug and resolve the issue.

Auditing SOAs

The auditing of an SOA will involve inspecting the WSDL and WS-Policy for each service to make certain that only those functions desired for the workflow can be called, that all services that provide sensitive functionality or operate on the data that remain confidential require authentication, that data are encrypted and that encrypted data are signed. The auditor must also verify that authentication and authorization are logged by the service, and that logs can record who did what, when, and with what authority.

Since the SOA is a complex multiplatform cross-enterprise unity of capabilities in the service of business objectives, properly auditing any SOA for compliance with internal or external standards will take time and diligence. However, no shortcuts should be taken as a flaw caught in an audit may prevent the business from compromise by a criminal who found a service without authentication or sensitive data that remained unencrypted.

Further Reading

Chappell, D. 2004. *Enterprise Service Bus*. O'Reilly Media, Sebastopol, CA.

Erl, T. 2004. *Service-Oriented Architecture: A Field Guide to Integrating XML and Web Services*. Prentice Hall, Upper Saddle River, NJ.

Eston, T., Abraham, J., and Johnson, K. 2011. Don't drop the SOAP: Real world Web Service testing. Black Hat white paper. https://media.blackhat.com/bh-us-11/Johnson/BH_US_11_JohnsonEstonAbraham_Dont_Drop_the_SOAP_WP.pdf

Johnson, K. 2011. Attacking Web Services Pt 1—SOAP, InfoSec Institute. http://resources.infosecinstitute.com/soap-attack-1/

Johnson, K. 2011. Attacking Web Services Pt 2—SOAP, InfoSec Institute. http://resources.infosecinstitute.com/soap-attack-2/

Juric, M. 2006. *Business Process Execution Language for Web Services*. Packt Publishing, Birmingham, UK.

Krafzig, D., Banke, K. and Slama, D. 2005. *Enterprise SOA, Service-Oriented Architecture Best Practices*. Prentice Hall, Upper Saddle River, NJ.

Rosenberg, J. and Remy, D. 2004. *Securing Web Services with WS-Security.* SAMS Publishing, Indianapolis, IN.

Shreeraj, S. 2007. *Hacking Web Services*. Charles River Media, Boston, MA.

Systems Development Controls

Chapter 14

Managing the Security Testing Process

Anthony Meholic

Contents

Overview

Companies of all sizes are facing the daunting challenge of establishing and maintaining their online presence. The explosion of social media and mobile applications has made it almost mandatory that a company provide online services to compete in the "always connected" world. The result is the need to migrate traditional services to websites, web, and mobile applications. The implementation of the remote deposit capture by financial institutions is a prime example of what once used to require a visit to a brick-and-mortar branch. Now, the customer simply takes a picture of the check and sends it via a personal computer (PC) or a smartphone. This newly found freedom comes at a cost, however. Namely, businesses must be in the position to ensure the security of these new services. This is typically done through the use of what is referred to as ethical hacking or security testing. The decision on what type of test, when to test, and who should test typically lies with the information security professional. However, it is possible that someone outside the information security realm may be designated as the manager of security testing. To properly manage this activity, it is crucial that the person making these decisions has the appropriate knowledge to select what is best for any given situation.

Precursors to Security Testing

Before trying to tackle the task of security testing management, it is necessary that the individual be familiar with some of the concepts and terminology associated with these activities. While some of the items are optional, it is usually advisable to include as many of these as possible in the security testing process. At the very least, competent security testing management dictates that the mandatory items are always included in the management of security testing. Understanding these concepts and knowing when and how they should be implemented is the first step in security testing management.

Test scope—The test scope is mandatory for all security testing. This is a vital document that will exactly define what devices, applications, or systems will be tested. The test scope must list the exact Internet Protocol (IP) addresses, domains, uniform resource locators (URLs), applications, or systems that the testing team is authorized to engage while testing. This prevents testing from creeping or accessing devices, applications, URLs, or systems not intended to be a part of the test.

Rules of engagement (ROE)—This is mandatory for all security testing. As the name implies, the ROE details the specific types of testing that is authorized. The ROE is used by the testing team to dictate the tools, methodology, and actions that are authorized for the testing activity. The omission of this document could lead to corruption/loss of data, production devices being damaged, and even legal actions.

Statement of work (SoW)—This is optional but recommended for all instances of security testing. It is submitted by the vendor and lists the specific details regarding the security testing. At a minimum, it should include the fees, charges, or rate for the activity, the exact IP range(s), and/or domains and the date of the test. Any special arrangements such as reimbursement for airfare, hotels, and meals should also be included if these expenses will be incurred.

Test authorization letter—This is optional and depends upon the environment being tested. The letter provides written authorization to the testing team to conduct the tests on the specified device, systems, or applications. This is usually needed only if the testing scope includes devices, systems, or applications that are not directly hosted and/or managed by the host company. By issuing this, the testing team has the documentation to prove that the activity had been authorized and can avoid potential legal problems.

Notice of testing—This is optional. The notice of testing is used to send a message to the entities that monitor the network that testing is about to commence. This alert will forego them sending out alarms when the suspicious activity is detected. There should be a notice of testing sent out prior to and also upon the completion of the testing activities.

Security Testing Management

The first step in managing this process is to have an understanding of the types of tests. The term "security testing" refers to a group of tests consisting of vulnerability assessments, penetration tests, web application tests, and mobile platform tests (MPTs). Each of the various tests requires technical skills, experience, and appropriate tools to be conducted successfully. It is important to note that these components are different for each test. In other words, a person who is very adept at vulnerability assessments (VAs) may not necessarily be equally adept at mobile platform testing. The information security professional needs to understand this difference so that the test selection is appropriate for the desired result.

Types of Testing

VAs are the first type of test to be considered. The VA is the least technical of the tests and can be performed relatively quickly. A VA can be defined as

> The use of automated scanning tools to identify vulnerabilities and then to manually verify results to eliminate false positives and common false negatives.

The results represent a snapshot of the environment and of the *possible* vulnerabilities present. The VA does not delve into any discovered vulnerabilities to assess their scope or veracity. This is simply a listing of vulnerabilities that the tool has determined possible in the tested environment. Manual verification of each discovered item is mandatory so as to rule out false positives. When a vendor is tasked with performing this type of test, the SoW needs to include a section confirming that the manual vulnerability verification will be performed. It is not unusual for unscrupulous testing vendors to simply run the automated scan and present the generated report. This leaves the customer with the onerous task of weeding out the false positives that may be impossible if the resources or technical skills are limited as with most server message blocks (SMBs). The typical VA should be between 40 and 80 h depending upon the complexity of the targets.

Penetration tests (PTs) are the next type to be considered. These have received a copious amount of media attention, and they have been outrageously misrepresented in both film and literature. A PT can be defined as

> Testing to reproduce real-world attack scenarios and to identify *and exploit* vulnerabilities.

The difference between a VA and a PT is the attempt to exploit the discovered vulnerabilities. The PT will start with automated scans of the targets using some of the same tools as in the VA, but the pen tester will take the next step and try to exploit each vulnerability. This is an important exercise because the exploitation of vulnerabilities will reveal the exact scope of the exposure of internal/backend systems. The testing teams conducting a PT must have technical skills to commensurate with the environment being tested. PTs are very labor intensive and will usually require between 80 and 120 h for completion.

Another form of security testing that is getting to be quite frequent is web application testing (WAT). Companies conducting business over the Internet require the use of websites and applications. Prior to putting these into production, a company should have them tested for the possible vulnerabilities or security flaws. Sometimes, this is required by an Internet service provider (ISP) before hosting a site or an application. One hosted website or application that has security flaws could result in the compromise of the ISPs backend system, thereby affecting multiple other customers. The WAT can be defined as

> The testing of a website or web application to discover vulnerabilities that could cause a compromise of data or network infrastructure.

The WAT can be performed using automated tools, manually or a combination of both. The testing teams performing a WAT need specialized skills and tools relating to this unique environment.

Depending upon the intricacy of the website/application, the WAT will usually take as much time as a VA but less than a PT; that is, between 40 and 80 h.

The final type of test is the MPT. The current popularly used platforms are Apple (iOS), Android, Blackberry, and Windows Phone. An MPT can be defined as

> The testing of a mobile device and its underlying operating system to detect security flaws, vulnerabilities, and data leakage.

As smartphones and tablets become more prevalent, there is an ever-increasing need for performing mobile platform security testing. Since there are several varieties of operating systems available on the mobile devices, it is critical to identify which are going to be included in the test scope. A vital aspect that is unique to this environment is that testing a mobile application for one of the operating systems does not translate to the other systems. Each requires a unique set of skills and tools. Since there are various hardware, protocols, and services present for each platform, the testing team performing an MPT must have highly technical skills and specialized tools for this type of test. It is another test that requires a commitment of resources and time. A typical MPT will take between 40 and 80 h per operating system. Similar to the VA and WAT, this is also dependent upon the complexity of the application and the platform.

Once the type of testing has been determined, the details of the test will need to be discussed. One of the most often heard words in the security testing vernacular is the term "black box." All the various types of tests can be performed in several ways. These are black box, white box, and

gray box. Each has its own purpose and can help make a test truly significant. However, as with the type of test, the information security professional must understand the difference between them and the context that each should be used.

"Black box" testing refers to a security test being conducted with minimal (or no) prior information provided to the testing team. The usual test case is to give the test team a single URL, domain, or IP address. This is done to best simulate the environment and conditions of a true external attacker. A black box test is useful to discover how easy or difficult it would be to gain access to internal systems with minimal information. Using this type of testing adds a significant amount of time to the test schedule, and this in turn will be reflected in the cost.

"White box" testing is just the opposite. Prior to conducting the test, the testing team is provided all these data and system information in advance. This is very useful when there is a time constraint placed upon the testing schedule. Since the testing team does not have to do any prolonged reconnaissance of the targets and they know the details of the infrastructure, they can focus on crafting attacks on the specific components. This will result in the test being completed relatively quickly and this would result in a lower cost. The drawback is that it will not replicate the conditions that an outside hacker would encounter.

As the name implies, the "gray box" test is a combination of both. In this test, the testing team will be given some initial information but not enough details to remove the necessity of performing reconnaissance. This strikes a workable compromise for the external attack simulation and reduces time and costs.

Selection of a Testing Team

Whether the testing will be performed by an internal team or outsourced to a vendor, there are certain criteria that must be considered to ensure the team's competency. The first being the experience the team has with the type of test selected. One of the first things to ask a potential security testing provider is to describe the difference between a VA and a PT. Any reputable purveyor of security testing will know these differences and will be able to provide details on how they perform the exploitation phase. Any entity that does not know the correct difference, or in some cases, simply states that they are the same thing, should be avoided.

Once that has been established, it is recommended that the experience of the testing team be reviewed. Items such as certifications, experience, and training for the specific type of test being discussed will help establish the team competency. A testing team may consist of several testers. If so, the lead tester needs to be the senior member of the team and needs to have the skills, training, and experience in this type of test. It is not unusual to have junior members in the testing team. They will usually be used to conduct the automated scanning portion and then shadow the lead tester during the other portions of the test. This allows them to gain experience by watching the lead tester and sometimes under direct supervision to actually perform some of the manual testing. Shadowing the testing team by members of the information security team is also a good way to have some knowledge transferred to junior-level employees. The SoW should include a section detailing the personnel on the testing team, their role, and also a stipulation that all critical portions of the test will be conducted by the lead tester or under his/her direct supervision.

While there is a distinct difference between the various tests, there is also a significant difference in the tools that are used. In interviewing the potential testing teams, one of the items that should be included is a discussion on the tools that are going to be used. Familiarization of the various tools allows the information security professional the ability to gauge the testing team's

competency to a certain degree. For example, if a testing team state that they are going to be using Nmap as the only tool for a web application test, it should raise concerns because that tool is not designed to be a web application tool. Knowing the exact tools being employed is also useful for understanding and preparing for any possible side effects that could result from the testing activity.

A professional security testing team should follow a structured methodology for conducting any type of test. Two requirements for the security testing process are consistency and repeatability. The implementation of a consistent testing methodology ensures that the process will be performed the same way no matter who is actually conducting the test. This consistency is very important when comparing the results from different testing teams. A consistent process also ensures that the results are repeatable. At some point in the remediation effort, it will be necessary to repeat the discovery of the vulnerability. Every person who has ever conducted a test has encountered a push back from the customer stating that "This must be a false positive!" or the ever popular, "I don't believe that." By using the consistent process, the tester can walk the doubters through the exact steps knowing that the results will be the same.

One often-overlooked item in the security testing process is the final report. Vendors are notorious for interspersing pages of the marketing material in the final report. While they need to try to get the name, recognition, and exposure, this can cause frustration trying to sift out the real pertinent details of the report. Before selecting a vendor, request a sample of their test. Most vendors have these readily available or can redact an actual report so that you can evaluate it. The content of the final report should be situated such that an executive summary is located in the very front followed by the detailed report. It is always a good idea to request that the raw data from the testing tools should be included as a part of the final deliverables. The final report should also be available in a variety of formats, such as Word, portable document format (PDF), text, and so on, to accommodate the ease of accessibility. If possible, the information security professional should design the format that works best for his environment and should request that the testing team can use that format. This is very advantageous because the report would not contain any superfluous marketing material and the format would be identical for all of the security testing vendors utilized.

Besides the technical competency, tools, methodology, and reporting issues, the one remaining consideration is that of cost. For most information security professionals, the budget allotted to their endeavors can be very tight. When receiving proposals from vendors, it is not unusual to get a wide range for fees to perform the tests. The old adage of "You get what you pay for" is not always the case. Security testing can be very expensive depending upon the type and scope of the test. There are some large companies that specialize in this type of service. The general rule of thumb is that the larger firms are going to be more expensive. However, with the larger firms, the testing teams will usually have more experience and can handle last-minute requests better than the smaller vendors. The smaller vendors, however, can sometimes provide excellent service at a much lower cost. The going rate for security testing ranges from $200 to $400/h; there are even some higher. Another option is to charge a flat fee for the test. If this is used, the SoW should include a section detailing what is covered under the fee and list any ancillary costs, such as charges for retesting, custom report format, and so on.

In summation, the selection of a security testing team has many components for consideration. A careful deliberation must be given to select the testing vendors that best fit. It is recommended that the information security professional select several test providers and rotate their use. This allows the information security professional to have multiple teams test the same environment. It keeps the testing teams from getting complacent and acts as a validation that the other teams had not missed any vital vulnerability.

Testing Methodology

Components of an Effective Testing Methodology

The primary focus for a testing methodology should be to guarantee proper guidance to the testing team and to make sure that there is consistency in conducting the tests. At the same time, it is important that the methodology retains some flexibility. The art of security testing is very individualized. There are a large number of tools that do the same tasks. Indeed, in most cases, there are several ways to go about getting the same information. The methodology must not place constrictions on the testing team and must force them to work outside their experience or training. The use of the best industry practices is recommended to be used as the basis for a customized methodology. The most common one in practice is the five-phased testing methodology. These are

Phase I—Project planning
Phase II—Information gathering
Phase III—Vulnerability detection
Phase IV—Exploitation and control
Phase V—Analysis and reporting

Testing Methodology Phases

Phase I: Project Planning

The first phase for testing should be project planning. In this phase, the information security professional will set the parameters for this particular test. The type of test—VA, PT, WAT, or MPT—will be the first item to be determined. Once that has been decided, the next item will be whether it should be a black box, white box, or gray box test. As discussed, each has its own advantages and disadvantages. For externally facing systems, it is recommended that at least one black box test should be performed on the environment annually.

Working with the network group, the exact scope will be the next item to be determined. It is imperative that the scope of the test is *explicitly defined* by the way of URL, domain, or IP range. This is especially true if the environment has connectivity to a vendor or a hosting entity. Without strictly defined boundaries, the testing team may overstep the intended scope of the test and may inadvertently conduct unauthorized testing activities resulting in possible legal action.

Also important is the positioning of the testing activity. Conducting an internal test will offer a different set of results as compared to an external test. It is recommended that testing can be alternated between the two so as to have as complete a picture as possible. While the threat from an external source is given, for many years, security metrics have indicated that the most common threat is to sensitive data from inside access. To help evaluate this threat, an internal test is always a good idea.

Another aspect for consideration is whether to conduct authenticated tests. In this type of test, the team is provided several sets of credentials so that the test can be extended to systems/applications requiring access control. The testing team should be supplied a minimum of two sets of credentials for each defined level of access. The purpose is to test and verify that the access privileges associated with each role are appropriate and also to test the viability of privilege escalation. This is a crucial test to help prevent insider compromises. The other focus will be to ensure that there is an account lockout feature. During the test, the testing team will cause an account lockout, unless that safety function has not been enabled. By providing the team with

several account credentials, it will not be necessary for the testing team to call to have an account unlocked, which will save time. It is highly recommended that an authenticated test must be conducted at least once a year.

The next decision that needs to be made is the type of system that is being tested. In most companies, there are several systems used for the development, testing, and staging of application/services prior to migration onto production level devices. Should testing be performed on any of these types of systems, it is imperative that they must be the *exact* mirror images of production. This means that the dataflow, types of devices down to the same models, configuration, and structure must be identical. If there are any discrepancies or differences, then the testing may not be a true indication of the vulnerabilities present on the production system.

In most cases, the quality assurance (QA) or user acceptance test (UAT) systems used are similar but have some differences from production. If testing is still to be conducted, the information security professional must detail the differences and must arrange to have a limited test conducted on the production system that would focus only on the identified areas of difference. This will address any issues that could not be discovered in the QA/UAT environment.

After all these criteria have been established, the next item to be thoroughly detailed is the ROE for the testing team. If this vital piece is left undefined, it is possible that the testing team would conduct tests that would have a deleterious effect on the infrastructure, data availability, and integrity. Testing conducted on a production system could cause irreparable damage to customer data, internal systems, and result in legal actions. One of the first ROEs to be considered is to not include any type of denial of service attacks. If this is a nonproduction system, then this is not as critical because any outage would not affect the regular business functions. Obviously, this is not recommended on production systems.

Another ROE for consideration is the interaction with confidential data. In the best of situation, testing should never be performed on systems using live customer data. The possibility of data compromise is very high and could result in serious issues with customers. A test set of data that mirrors the actual data should always be available and used. If this is not an option, then specific ROEs must be created and communicated to the testing team as to how to proceed. It should be clearly stipulated that the testing may not manipulate these data in any manner. That is, there is no adding, deleting, or modification to any of the database tables. Further, it must also be prohibited to copy these data in any manner. There may be a need for screenshots to be taken as part of the testing process. If so, then the images must either obfuscate the confidential data, which negate the typical use of the screenshot, or the image must be appropriately safeguarded by ensuring that the access to the image is controlled and monitored. After the test, all files containing the confidential data must be deleted from all devices. The most important thing is that all ROEs must be clearly defined and communicated to the testing team.

For some companies, the testing schedule will need to be conducted only during certain hours of the day, that is, during nonbusiness hours. Any time restrictions will need to be determined and communicated to the testing team. When establishing time restrictions for the testing team, a consideration must also be made for having internal support available. If the test is being conducted at 3 a.m. and the team runs into an issue where an account needs to be unlocked or a webserver was inadvertently knocked off-line, then the information security professional must have arranged to have the appropriate level of support available.

The final task in this phase is what is referred to as the "Get Out of Jail Free Card." The official description is a test authorization letter. This is issued to the testing team prior to the initiation of the test so that there is documentation of all the covered activities. The letter should be very specific to the dates, targets, type of test, source location, and testing team members. The letter

should include the contact information for the testing team and the information security professional. If there are specific ROEs, such as no disk-operating system (DoS), then they should also be included.

While not a required part of the testing process, there is actually an option at this time that can be quite beneficial. Whenever any type of testing is being conducted, the intrusion detection (IDS), intrusion prevention (IPS), and network event monitors will most likely detect this suspicious activity; if not, a replacement is strongly advised! This should in turn generate all sorts of alarms and alerts to the appropriate network and security entities. To forestall these types of alerts, the information security professional may want to send out a notice to the IDS/IPS and network monitors that a test is scheduled and the activity is authorized. A typical notice of testing letter would include the exact date range, the exact targets, the originating source IPs of the testing team, and the point of contact for any issues.

However, the performance of testing can also be used as a way to determine the effectiveness and sensitivity of IDS/IPS and network monitoring. It is recommended that at least once a year that the IDS/IPS and monitoring entities *not* be alerted that a test is going to be conducted. When doing this, the testing team should also be requested to conduct the testing in a stealthy mode until the activity has been detected. By doing this, the information security professional receives an indication of the activity threshold that will trigger an alert. Knowing the level of activity that can go undetected is a vital piece of information when determining the level of risk inherent in the system. For example, if the information security professional receives a call within 15 min of the initiation of testing, then that shows a very sensitive monitoring system. But if it takes several hours, or in some cases days, there should be some concern that a lot of unauthorized activity had transpired before the alert was issued. While not a direct part of the test, this can be used to validate the IDS/IPS and monitoring controls.

Phase II: Information Gathering

This phase is used to perform reconnaissance and gather information about the targeted system. The activity in this phase is dependent on the type of test being conducted. If a white box test had been selected, then this phase is not required and the team would move onto phase III. For a gray box test, the testing team will attempt to gather more details on the identified targets. The information that had been provided to the team will be used to extract even more details prior to moving on to the next phase. This greatly shortens the activity required for this phase.

Black box testing requires an extensive amount of time to be spent in this phase. The team will usually start gathering information about the intended target using publicly accessible information. For example, if it is a website, then by simply going to the site and having the browser display the source code of the page reveals a large amount of information. The next task would be to quietly probe the targets for patch levels, open ports, and services running. This information will have a strong influence on the planning of the attack. This type of activity will continue until enough information has been obtained and the team can start to craft the nature of attacks to be used.

Phase III: Vulnerability Detection

The primary focus in this phase will be vulnerability detection. The identified targets will be probed to determine the possible weaknesses and known vulnerabilities. Detection is primarily

accomplished through the use of automated scanning tools. By running these tools, the testing team can efficiently discover the possible vulnerabilities with minimal resources and time. Once the discoveries are listed, the testing team must then manually evaluate each item to eliminate any false positives. Except for a VA test, the testing team would move on to phase IV. This is the only testing phase for a VA and the team would move to phase V.

Phase IV: Exploitation and Control

Phase IV is where the real work for the other type of testing is accomplished. In this phase, the testing team attempts to penetrate the target system(s) security by exploiting the vulnerabilities discovered in phase III. The testing team will attempt to gain access to the data and subvert back-end systems with the ultimate goal to gain control of the target. A critical part of this activity is to attempt to move the source of the attack to a system inside the network perimeter. This is done by the testing team gaining a foothold on one of the targets. It will then attempt to load the testing tools onto the system. In testing jargon, this is referred to as creating a pivot or repositioning. By establishing an internal source as the testing platform, the team is now behind the firewall and most IDS/IPS and network monitoring systems are not configured for the same level of scrutiny for internal traffic. This means that the team can conduct further testing with impunity. It also means that the backend systems are now exposed. One of the items testing teams like to exploit is the trust relationships between internal devices. It is not uncommon to have an implied trust between an internal server and a database (many times, there are hard-coded connection strings from these "protected" devices). This makes it even easier for the testing team to further exploit the system. By being able to successfully establish a pivot or reposition the attack, the testing team has essentially captured the "keys to the kingdom."

In the unfortunate situation where the testing team has been successful with the pivot, the testing team must document all the assets involved. This means that they must provide a detailed listing of all tools installed, files copied, and modifications made to establish the pivot point. Upon completion of the test, they must also document the date and time that each of the list items had been removed or returned to normal. Unless this is closely watched, it would be possible for the testing team to leave remnants on the system. There have been documented cases where these residual items were used by unauthorized entities to conduct actual attacks. This is one of the pitfalls in performing a PT.

When performing an MPT, there are some items that need to be considered. The mobile platform is a unique entity in that it uses a totally different set of protocols, services, and hardware. There are also different ways an application can be brought to the mobile platform. The first is simply using a browser to access a website created for the mobile environment. This is the easiest way to establish a presence in the mobile arena. All the development is done on the server side, and it would be ubiquitous in nature so that each of the operating systems could access this with minimal issues. It should be noted that some of the proprietary browsers on mobile devices (such as those found on the Blackberry Playbook and HP Touchpad) may have problems rendering to some website components. From a testing perspective, this scenario would require performing a web application test combined with the MPT. The WAT would evaluate the web application and the MPT would evaluate the mobile device for security flaws or data leakage. This comprehensive approach to testing would ensure that all aspects of the application would be scrutinized.

The next way a mobile device can access an application is through the installation of a custom application designed specifically for the operating system. The application is usually downloaded

from a central source (i.e., iTunes, Android app store, etc.) and installed on the device. An MPT would have to be performed to check for security issues and data leakage on the device. A request to review the actual source code would be even more thorough in helping to discover any security issues. However, it is unlikely that the source code would be made available to the testing team. In this case, the documentation of any QA, UAT, and security testing performed by the developers would be extremely beneficial.

There is also a third option that is a hybrid of the previous two solutions. In this scenario, a customized graphical user interface (GUI) is developed as an application for a mobile platform. This GUI is just a wrapper for connecting to an established mobile-friendly web application. While there is custom development, the backend processes utilize an established (and hopefully tested) web application and use this to feed the data stream to the customized GUI wrapper. For testing purposes, a WAT combined with an MPT would again be appropriate to ensure a comprehensive test.

The testing of mobile platforms has one other unique item for consideration. Owing to the varying operating systems and devices, testing has to be conducted for each operating system that is being utilized. This means that rather than just a single test, any security testing for a mobile application may have to include testing for as many as four systems (iOS, Android, Blackberry, and Windows Phone 7.5). Also, owing to the difference in the systems, the testing should *not* be performed on emulators (i.e., do not perform an iOS test on an ×86 emulator). While the application may function in a similar manner, the underlying operating system calls, data flow, and communications are not accurately replicated; so, the test would not be relevant to the environment on the actual device. A skilled testing team would be familiar with this concept and would indicate this in their testing plan. This is very important when putting together the SoW since this segregated testing could have a major impact on the testing schedule and costs.

Phase V: Analysis and Reporting

With the testing activities completed, the team begins the analysis and reporting phase. The main activity at this time is for the team to carefully review all the information from the previous phases. All discoveries are examined to eliminate any false positives. Once the results have been confirmed, the information is collated into a draft report. The report should include a detailed technical section that identifies specific security strengths and weaknesses. It should also include details for the vulnerabilities uncovered. The vulnerability reporting section must include a list of the assets where the vulnerability had been discovered, common vulnerabilities exposure (CVE) or continuous comprehensive evaluation (CCE) identifier number, risk level, type and description of vulnerability, and recommendations to resolve the issue. Supporting documentation should also be presented with the draft report. These would include items such as the raw data from the tools, source code for any customized shellcoding, and screenshots. These ancillary documents will permit the information security professional to verify how a vulnerability had been discovered and also provides a corroborating evidence that an exploit had been successful.

The draft report should be reviewed by the information security professional to verify that the enclosed issues are valid and the associated risk value is appropriate for the environment. It is not unusual for a network or a system to have some known vulnerabilities that cannot be resolved for whatever reason. These vulnerabilities need to be noted and a letter of acceptance needs to have been submitted. This letter would state that this particular vulnerability had been noted, and is not able to be resolved at the present time, and that the senior management agrees to accept the risk. In performing the review of the draft report, any known issues can be removed from the report.

The draft report is then sent back to the testing team for editing and the final report is issued. The final report differs slightly from the draft in that an executive summary is included. This is usually presented to the senior management so that the results can be communicated. It is concise in nature and does not include the "nuts and bolts" of the test process. The information security professional should also use this document to oversee the management of the threats and vulnerabilities.

It is recommended that the results of successive tests be compared with prior results. This provides the ability to discover any vulnerability trends. For example, if there have been significant vulnerabilities resulting from improper patching found in successive tests, then an obvious trend would be that there seems to be an inefficient patching process. Likewise, trends showing a decrease in the type of vulnerability would be a good indicator that previous remediation efforts have been successful in addressing the issue.

Know Your Tools

A critical skill for an information security professional is familiarization with security testing and hacking tools. There are a plethora of tools available both for purchase and as an open-source free download. For the most part, these tools can be obtained by anyone with or without training. To make an informed decision on a testing vendor, the information security professional must be able to recognize which tools are appropriate for the testing being considered. Even with the correct tools, it is imperative that the testing team have the training and expertise before being turned loose with some of these tools. When improperly used, they can cause serious damage to the infrastructure or databases.

The tools can be divided into three categories. Each has its own advantages and disadvantages but the ultimate decision for use is left to the actual tester. The distinctions for each category are listed below:

Category 1 Tools

Description

Category 1 tools are commercially licensed products (or services) for performing the various types of testing. The most commonly found tools in this category are the vulnerability scanners. These are used for network, application, and web application automated scanning. There are some products that purport to be "automated penetration testing" tools but in reality, they are just highly advanced vulnerability scanners. They are very good at what they do but the use of these products should in no way be construed as the equivalent of a test that incorporates a manual testing component.

Advantages

These tools are typically very stable so that when used, the tester does not have to worry about the tool crashing (or worse unexpectedly crashing the target). These tools also have technical support should the tester encounter a problem or require assistance with its use. The main advantage is that they also will have regular updates provided. This means that the tools will always be scanning for the most current vulnerabilities.

Disadvantages

It is not unusual for these tools to have very restrictive licensing. They are usually associated with a specific user or with a specific IP and Media Access Control (MAC) address. Others are restricted by domains or IP ranges so that the tool cannot be used outside a company's dedicated IP range. Another issue is that they cannot be readily shared. This is a distinct disadvantage for any internal testing team because it may require the purchase of multiple licenses.

Examples

While this list contains some of the more common tools, it is by no means all inclusive. Some typical category 1 tools are Qualys, CORE Impact, AppScan, WebInspect, and App Detective.

Category 2 Tools

Description

Like category 1 tools, these are also commercially licensed products (or services) for performing the various types of testing. These will also include the automated scanners for the various open systems interconnection (OSI) layers as described with the category 1 tools. The primary differences are with the licensing and cost.

Advantages

These tools have all the advantages of the category 1 tools. Namely, they are typically very stable, have technical support, and have regular updates provided. The notable difference in the licensing is that they are usually not associated with just a single user, IP, or MAC address. Another licensing feature is that some of these will have an option for a server version installation on a centralized device. This allows the tool to be accessed and used by a number of testers without the requirement of installing it on individual devices. For an in-house testing team, this is very beneficial as far as tool management and cost-effectiveness are concerned. These tools are also not as expensive as the category 1 tools.

Disadvantages

While there are many similarities, category 2 tools are usually more specialized in functionality. That is, instead of having a single tool to perform several types of tests, category 2 tools will focus on single components or functions. This may require the purchase of several different category 2 tools.

Examples

While this list contains some of the more common tools, it is by no means all inclusive. Some typical category 2 tools are LC5 Administrator, Canvas, Black Widow, and Sleuth.

Category 3 Tools

Description

Category 3 tools are open-source, freeware, or shareware tools available for downloading from a wide variety of sources. Since they are an open source, there are typically no restrictions as to the

use or the distribution. They are similar to category 2 tools in that they are usually developed for a singular type of testing. However, since there is little to no cost, building a toolbox of these tools would not be an issue.

Advantages

These tools have no licensing restrictions and have little to no cost; so, these can be readily obtained for whatever type of testing is required. These tools can be very effective and produce quality results similar to those of the other categories.

Disadvantages

There are numerous caveats to consider before using these tools. The first is that they *must* be thoroughly tested before using on the corporate network. If a vendor is using these tools, it would be recommended that they confirm that this would not be the first use of the tool or that it had been tested for stability. Since they are open source, it would be advisable to be reviewed for the malicious code that could be detrimental to systems being tested. Again, getting confirmation from the vendor using the tool would be wise. The final caution is that tools downloaded from the Internet should never be trusted unless they have been properly vetted in a secure environment. Vendors using open-source tools should be able to provide documentation of how the tool was procured. The important message is that it is better to be cautious when open-source tools are being used.

Examples

While this list contains some of the more common tools, it is by no means all inclusive. Some typical category 3 tools are Nessus, Nmap, Metasploit, Achilles, and Nikto.

Reporting Requirements

Even if a highly skilled testing team has been selected and if they were able to conduct the requested test without any problems, it is all meaningless unless the testing team can effectively communicate the findings. Writing the test report requires the same level of skill as the actual technical test. Indeed, for many vendors, the report is actually written by a dedicated technical writer. No matter who authors the report, there are certain criteria that must be followed so as to get the most out of these data. The guiding principle should be consistency.

The first tenet for consideration is the establishment of a consistent method for categorization of the vulnerabilities. This is especially important when there are multiple parties performing the testing. Each vendor may have a different way to categorize the same vulnerability. For example, one vendor may list the vulnerability as "SQL Injection" whereas another may list it as "Client-side input validation." Both are legitimate descriptions of the same vulnerability. Another problem is that having a nonconsistent categorization makes the identification of trends very difficult. Trying to collate all the similar vulnerabilities using varied vulnerability categories or descriptions would be very tedious and time consuming.

Another drawback is that when creating the executive summary and reporting to other nontechnical entities, the use of some of the vulnerability categories/descriptions can be difficult. Putting down the statement that "There were three instances of Cross-site Request Forgery" would

mean very little to most people. However, the statement "Improper Input/Data Validation" is much easier to grasp and yet accurately describes the underlying issue.

Creating and using a well-defined vulnerability categorization process is recommended to improve vulnerability management. If this categorization can be performed by the testing team, the final report will be much easier to analyze. It should be discussed with the testing vendor to see if they would be able to use this categorization. If not, then the first thing an information security professional will need to do is to map the discoveries to the established categories.

Using the Open-Web Application Security Project (OWASP) as the guidance, a concise vulnerability categorization can be created. These 10 categories cover almost all types of vulnerabilities encountered during testing. The categories are

Input/data validation
Authentication
Authorization
Configuration management
Sensitive data compromise/leakage
Session management
Cryptography
Parameter manipulation
Exception management
Auditing/logging

The next item to be determined in the report format is the vulnerability risk rating. It is almost a guarantee that no two testing teams will identically rate the same vulnerability. Some vendors have a tendency to rate vulnerabilities a level higher than usual on the premise that the more high-risk vulnerabilities they report, the more competent they appear. Others will stick to the rating from the CVE documentation and will not look at the vulnerability in the context of the environment. When having multiple vendors conduct testing, this lack of consistent rating can greatly influence the ability to properly track trends, prioritize remediation activity, and provide inaccurate data to determine the overall corporate risk. It is vital that the information security professional discusses the method of vulnerability rating that the testing vendor utilizes. In this way, the ratings can be either accepted or, like the categorization, migrated to a consistent rating system.

The development and implementation of a consistent rating methodology would be beneficial. Some vendors may be amenable to actually using the requested rating system but at the very least, it would be possible to extrapolate the reported ratings to the same rating scale. The development of a rating methodology should incorporate at least five criteria. These are

Data classification—Identifies what type of data is involved (i.e., confidential, restricted, and public).
Type of exposure—Identifies whether these data are exposed to the public, Internet, vendors, intranet, or local only.
Financial impact—This should be scaled according to the level of business in the company.
Business impact—This should be determined by the business function involved. A typical indicator is the recovery time objective (RTO) assigned to this particular business function.
Technical skills required—This is an indication of the skills and technical knowledge required to exploit this particular vulnerability.

Once these criteria have been defined, the use of a simple scale of 1–5 for each item will permit the calculation of the rating for the vulnerability. A typical scale for this type of methodology would be low = 5–12, medium = 13–18, and high = 19–25. By using this process for rating all the discovered vulnerabilities would then place them in the appropriate context for the tested environment. This would result in better prioritization of resources for remediation and also the credibility to any trending efforts.

Process Deliverables

One of the most advantageous tasks that an information security professional can do to properly manage security testing is to establish a formalized process with a set of deliverables. In this way, it will be possible to ensure that despite the fact that multiple vendors or testing teams will be sued, the process is consistent and provides proper documentation. The final piece of a well-structured management process is to define the deliverables for each section of the testing process.

As discussed earlier, the first item to address is a scoping document. This is a vital component of the process because it will define the exact target(s), type and nature of the test, and most importantly the ROEs that the testing team must use. This should be reviewed by both parties and the testing team must acknowledge that they understand and will comply with the contents of this document.

The next deliverable should be the testing authorization document (the "Get Out of Jail Free" card). The content should include the date range of the test, testing team members, type of tests authorized, and contact information of the information security professional overseeing the activity. This should never be an option, and prior to any testing, the vendor needs to sign-off on their agreement with the stipulations detailed in this document.

The start test notice is an optional component of the process. In most cases, sending out the start test notice to the IDS/IPS and event monitoring groups is a courtesy to avoid undue alerts. However, if a part of the exercise is to determine the efficacy of these entities, then this document is not required.

Another optional document is the daily status report. In some cases, it may be necessary to receive daily reports on the progress of the testing. This is especially true for large or complex environments. By having the testing team submit a daily test report, it is possible to track what areas are going to be tested next. This in turn permits the normal business functions to be conducted without the fear of interruption if they are not going to be the focus of the testing activity. The drawback to this is that it takes time away from the testing for the team to submit this report.

When the testing activity has been completed, the end test notice should be sent to all the involved parties. This would include the IDS/IPS and event monitoring groups who by this time should have detected the testing activity. It is important to send out as soon as possible because if real unauthorized activity should commence at this time, it may be misconstrued as a part of the test and may not be reported. The distribution of the end test notice will prevent this from occurring.

The last deliverable is the final test report. While this has been covered in detail, this document can be utilized for other purposes. The discovery of vulnerabilities is just the first stage in the vulnerability management process. Simply identifying the vulnerabilities does not enhance a company's security posture. Indeed, this is when the real work starts. The information security professional must take these data from the report and work with the relevant groups to resolve the vulnerabilities.

Tracking open vulnerabilities is essential to ensure that any discovered items are properly resolved. The tracking should include metrics surrounding the security testing process. Items such as a breakdown of vulnerability category, line of business, and risk level will enable vulnerability trending. Using these data, it will be possible to identify which areas are experiencing recurring vulnerabilities. This is a valuable information because the appropriate resources could be allocated to the problem areas. On a broader scale, the tracking of subsequent security testing can also be useful to the information security professional by providing the metrics to indicate the activities of the information security program. This quantifiable presentation can be a valuable tool in demonstrating the return on investment for the information security budget.

Conclusion

The information security professional has been charged with the task of ensuring that quality, efficient, and meaningful testing is performed to identify and resolve vulnerabilities in the corporate environment. With the ever-changing types of technology and accessibility, this task has become quite a daunting endeavor. To successfully manage the security testing process, a full understanding of this space is required.

The information security professional must be able to distinguish the differences between the types of tests and know what type of test is best for the desired results, most importantly, how to determine which testing team or vendor can provide the necessary services. If this is approached in a thorough manner and armed with the correct knowledge, the task can be quite manageable and could result in positive exposure for the information security program.

Chapter 15

Security and Resilience in the Software Development Life Cycle*

Mark S. Merkow and Lakshmikanth Raghavan

Contents

This chapter examines in detail the environment in which software is developed and deployed while applying the enduring principles of software security to help designers and developers better appreciate the whys and hows of secure and resilient software development. After reading this

* From Mark S. Merkow and Lakshmikanth Raghavan, *Secure and Resilient Software Development*, Copyright 2010 Taylor & Francis Group, LLC.

chapter, you will have a deeper understanding of how deliberate practices and attention to security within the development life cycle can improve the processes of developing software and the products produced by the processes.

Resilience and Security Begin from Within

The *only* reliable way to ensure that software is constructed secure and resilient is by integrating a security and resilience mindset and process throughout the entire software development life cycle (SDLC). From the earliest days of software development, studies have shown that the cost of remediating vulnerabilities or flaws in the design are far lower when they are caught and fixed during the early requirements/design phases than after launching the software into production. Therefore, the earlier we integrate security processes into the development life cycle, the cheaper the software development becomes in the long haul.

These security processes are often just "common sense" improvements, and any organization can and should adopt them into its existing environment. There is no one right way to implement these processes—each organization will have to fine tune and customize them for its specific development and operating environments. These process improvements add more accountability and structure into the system too.

Regardless of which software development methodology an organization follows—Waterfall, Agile, Extreme Programming (XP), and so on—these security processes must be present in one form or the other. Even though the development life cycle explained below fits more into custom software development, security-related processes must be included in all life-cycle models meant for product development or line-of-business applications within an enterprise's software development practices.

Figure 15.1 provides a high-level overview of the fundamental security and resilience processes that should be integrated into the various SDLC phases, from requirements gathering to deployment and beyond. Each process yields its own findings and recommendations are prepared to make the appropriate changes to the design, architecture, source code, use of third-party

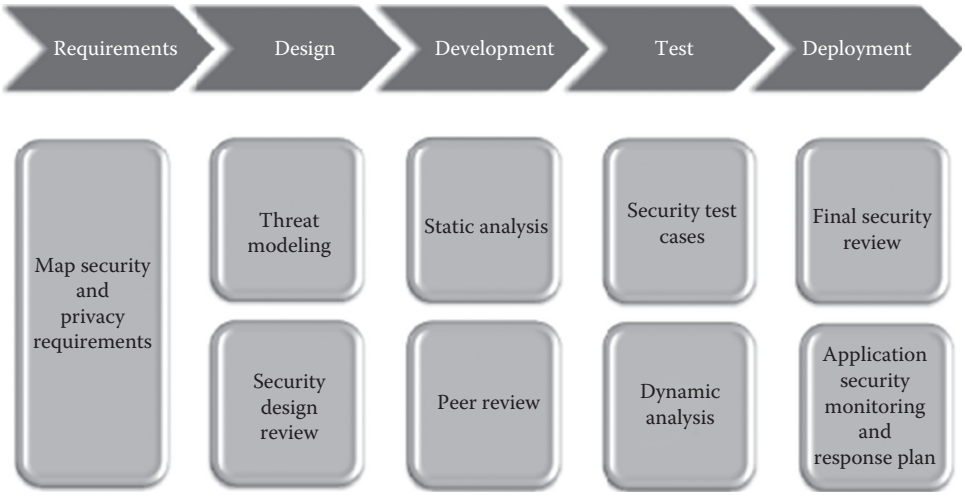


Figure 15.1 Security and the SDLC.

components, deployment configurations, and other considerations to help understand and reduce the risk down to an acceptable level. Here, you will find guidance on practices that you should consider for implementing each phase of development:

- Requirements gathering and analysis
- Systems design and detail designs
- Application coding and reviews
- Testing steps
- Deployment steps

Requirements Gathering and Analysis

The key activities during the requirements gathering and analysis phase are intended to map out and document the nonfunctional requirements (NFRs) for the system under development. It is vital to have these ready before the translation of business requirements into technical requirements begins; designers need to understand the constraints they are expected to face and be prepared to answer the call for security and resilience, as well as other NFRs. To be effective, business systems analysts and systems designers should be sure they are very familiar with the environment in which they are operating, by reviewing and maintaining their knowledge about

- Organizational security policies and standards
- Organizational privacy policy (which may have varying requirements in different places)
- Regulatory requirements (Sarbanes–Oxley, HIPAA, etc.)
- Other relevant industry standards (PCI, DSS, ANSI-X9 for banks, etc.)

The NFRs are then mapped against the critical security and resilience goals of

- Confidentiality and privacy
- Integrity
- Availability
- Nonrepudiation
- Auditing

Finally, these security requirements are prioritized and documented for subsequent phases. See Figure 15.2 for an example of this type of mapping.

Systems Design and Detailed Design

Threat modeling and design reviews are the two major resilience processes that you will encounter during the design phase. There are two classes of vulnerabilities: design-related and implementation-related vulnerabilities. While the latter are very easy to find, the former are very expensive and time consuming to locate and fix if they are not detected early in the SDLC. Security subject-matter experts should be deeply involved with the project during this phase to ensure that no bad design issues creep into the design and architecture of the software or the system.

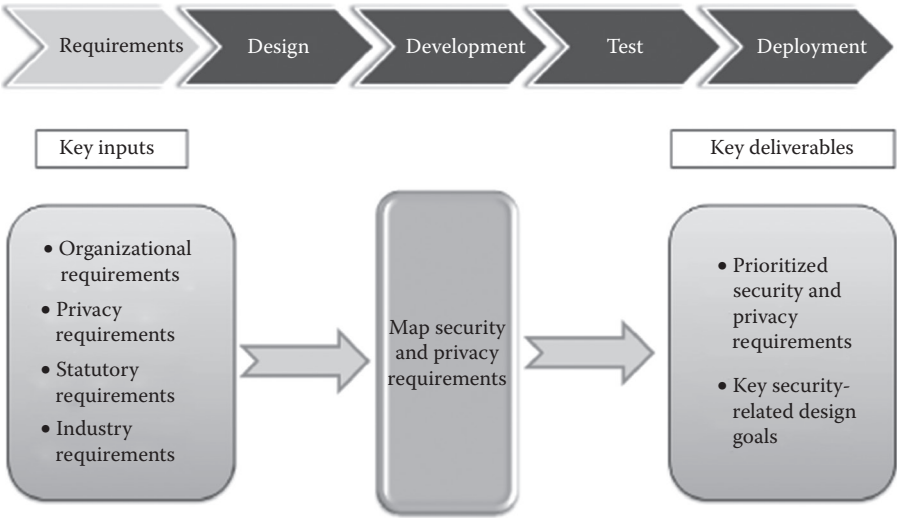


Figure 15.2 Requirements phase.

Detailed threat modeling is an excellent way to determine the technical security posture of an application to be developed or under development. It consists of four key steps:

1. Functional decomposition
2. Categorizing threats
3. Ranking threats
4. Mitigation planning

Functional Decomposition

Functional decomposition is typically performed using data flow diagrams. The key aspect of this step is to understand the boundaries of untrusted and trusted components, which allows for a better understanding of the *attack surface* of an application that an attacker might want to exploit.

Categorizing Threats

Even though the attackers' goals vary, understanding the different types of threat agents and their potential impacts on an organization is a very important activity. **STRIDE** is a framework developed by Microsoft for classifying threats. The different threat categories used are

- *Spoofing of user identity:* An example of identity spoofing is illegally accessing and then using another user's authentication information, such as username and password.
- *Tampering:* Data tampering involves the malicious modification of data. Examples include unauthorized changes made to persistent data, such as that held in a database, and the alteration of data as it flows between two computers over an open network, such as the Internet.

- *Repudiation:* Here, threats are associated with users who deny performing an action without other parties having any way to prove otherwise—for example, a user performs an illegal operation in a system that lacks the ability to trace the prohibited operations. Nonrepudiation refers to the ability of a system to counter repudiation threats. For example, a user who purchases an item might have to sign for the item upon receipt. The vendor can then use the signed receipt as evidence that the user did receive the package.
- *Information disclosure:* Threats in this area involve the exposure of information to individuals who are not supposed to have access to it—for example, the ability of users to read a file to which they were not granted access or the ability of an intruder to read the data in transit between two computers.
- *Denial of service:* Denial of service (DoS) attacks deny the service to valid users—for example, by making a web server temporarily unavailable or unusable. You must protect against certain types of DoS threats simply to improve system availability and reliability.
- *Elevation of privilege:* In this type of threat, a nonprivileged user gains privileged access and thereby has sufficient access to compromise or destroy the entire system. Elevation-of-privilege threats include those in which an attacker has effectively penetrated all system defenses and becomes part of the trusted system itself, a dangerous situation indeed.

Ranking Threats

Ranking potential threats for a software system requires a fair amount of subjective judgment. The level of damage caused by a successful exploit can vary significantly depending on various factors. **DREAD** is a model developed, again by Microsoft, to accomplish the same in a well-organized manner. We arrive at a risk rating by asking the following questions:

- *Damage potential:* How great is the damage if the vulnerability is exploited?
- *Reproducibility:* How easy is it to reproduce the attack?
- *Exploitability:* How easy is it to launch an attack?
- *Affected users:* As a rough percentage, how many users are affected?
- *Discoverability:* How easy is it to find the vulnerability?

Mitigation Planning

With a list of ranked threats, you can document a high-level mitigation plan by mapping them to the potential vulnerabilities in the software system.

Design Reviews

The next activity in this phase is the security design review. A security subject-matter expert, not a member of the core development team, usually carries out the design review with the key objective of ensuring that the design is "secure from the start." These reviews are typically iterative in nature. They start with the high-level design review and then dive deeply into each component or module of the software.

Threat modeling and design reviews can leverage commercial off-the-shelf tools, custom in-house software, or even simple checklists. The personnel must use their best judgment based on the environment, the organizational structure, and the existing processes and practices. See Figure 15.3 for the major steps in the design phase.

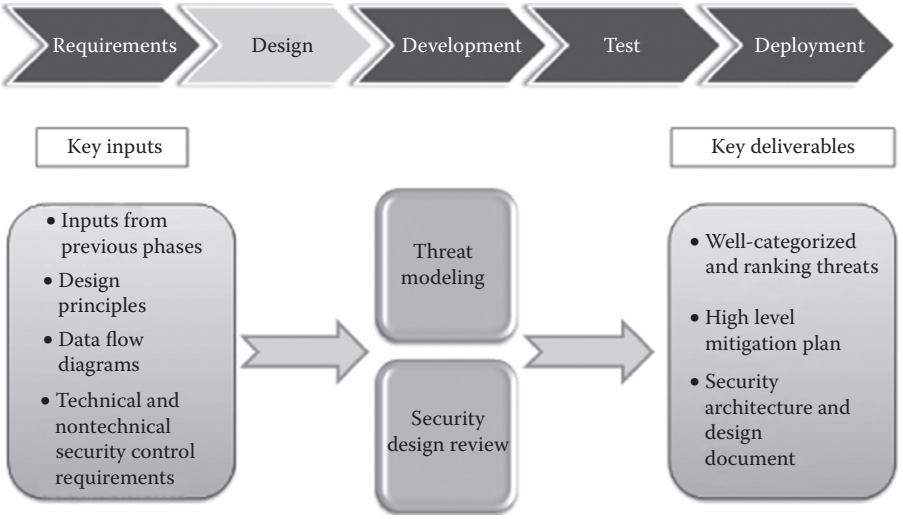


Figure 15.3 Design phase.

Development (Coding) Phase

The activities in the development phase often generate implementation-related vulnerabilities. Static analysis and peer review are the two key processes to mitigate or minimize these vulnerabilities.

Static Analysis

Static analysis involves the use of automated tools to find issues within the source code itself:

- Bug finding (quality perspective)
- Style checks
- Type checks
- Security vulnerability review

Automated security review tools tend to have a high percentage of false positives, but they are very efficient at catching the low-hanging vulnerabilities that plague most application software (lack of input validation, SQL injection, etc.). Static analysis cannot, however, detect all types of vulnerabilities or security policy violations—that is where manual peer review becomes important.

Peer Review

A peer review process is far more time consuming than automated analysis, but it is an excellent control mechanism to ensure the quality and security of the code base. Developers review each other's code and provide feedback to the owners (original coders) of the different modules so that they can make appropriate changes to fix the flaws discovered during the review. Developers can accomplish this with or without the use of specialized tools.

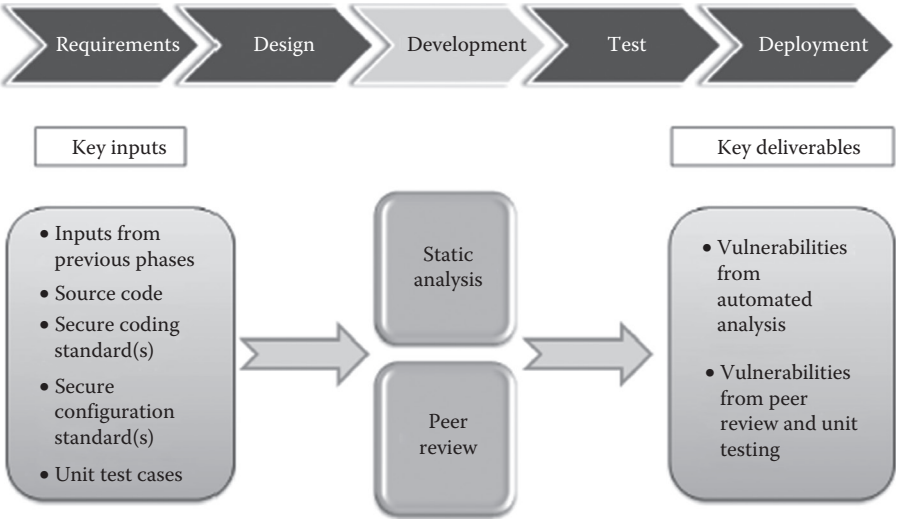


Figure 15.4 Development phase.

Unit Testing

Unit testing is another key process that many organizations fail to perform regularly but is important from a security and resilience perspective. Unit testing helps to prevent bugs and flaws from reaching the testing phase. Developers can validate certain boundary conditions and prevent vulnerabilities such as buffer overflows, integer over- or underflows, and so on within a module or submodule of an application. See Figure 15.4 for a diagram of the security activities in the development phase.

Testing

The test phase is critical for discovering vulnerabilities that were not discovered and fixed earlier. The first step in the test process is to build security test cases. A key input to this process is the systems requirements documentation. The (security) test team uses all the assumptions and business processes captured to create several security test cases. Security testers then use these test cases during the dynamic analysis of the application. The software is loaded and operated in the test environment and tested against each of the test cases. A specialized penetration testing team is often deployed during this process. These manual security reviews are very effective in discovering business logic flaws in the application.

Dynamic analysis also consists of using automated tools to test for security vulnerabilities. Just like static analysis tools, these tools are also very efficient in ensuring "code complete" scanning coverage and catching high-risk vulnerabilities such as cross-site scripting, structured query language (SQL) injection, and so on.

These tests are iterative in nature and result in a list of vulnerabilities that are then ranked for risk and prioritized. The development team then fixes these errors and sends the remediated code back for regression testing. See Figure 15.5 for a diagram of the security steps in the test phase of the SDLC.

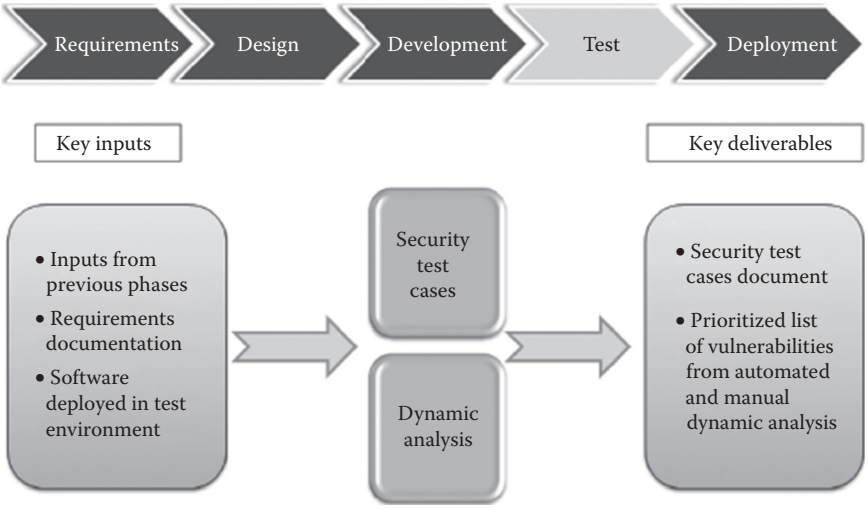


Figure 15.5 Test phase.

Deployment

The deployment phase is the final phase of the SDLC, when the software is installed and configured in the production environment and is made ready for use by its intended audience.

A key part of managing changes is to have a change advisory board (CAB). A CAB offers the multiple perspectives necessary to ensure good decision making. A CAB is an integral part of a defined change management process designed to balance the need for change with the need to minimize the inherent risks. For example, the CAB is responsible for oversight of all changes in the production environment. As such, it fields requests from the management, customers, users, and information technology (IT).

During the deployment phase, security subject-matter experts who may or may not be part of the CAB perform a final security review to ensure that the security risks identified during all the previous phases have been fixed or have a mitigation plan in place. During this phase, the development team coordinates with the release management and production support teams to create an application security monitoring and response plan. The production support team, in conjunction with the network/security operation center, uses this plan during the operation of the application to manage security incidents and engage the appropriate teams for response and remediation.

The ongoing monitoring of the application also includes periodic security testing of the application in production, using manual and automated testing techniques to help assure that new threats and vulnerabilities, due to changes in supporting software or reliant systems, do not affect the security and resilience of the application. See Figure 15.6 for the security activities in the deployment phase of the SDLC.

Security Training

Even though training may not seem to fit directly into any particular SDLC phase, it plays a very important role in improving the overall security and resilience of developed software. Training

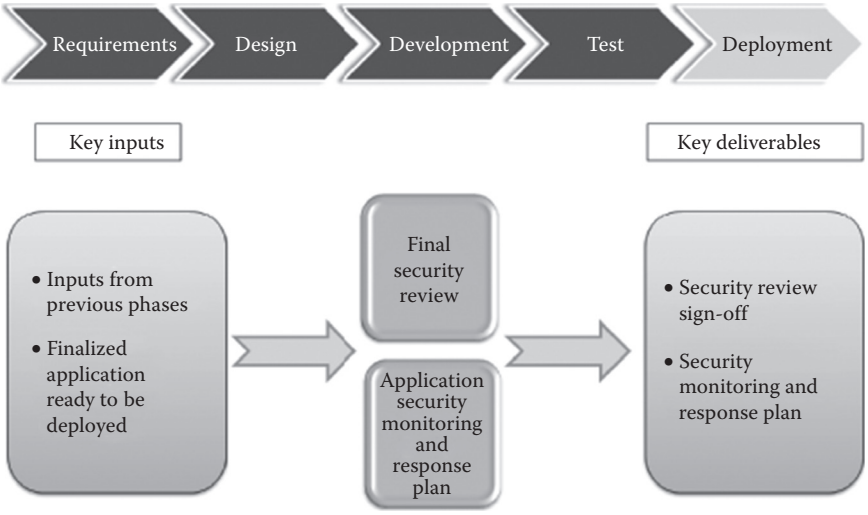


Figure 15.6 Deployment phase.

should be a prerequisite for anyone who has a role anywhere in the software development environment. All developers and other technical members of the software design/development/test teams should undergo security training that explains the responsibilities of their role, establishes the expectations for their part for security and resilience, and provides the best practices and guidance for developing high-quality software.

Summary

Resilient and secure applications do not come about by accident—only through careful planning and deliberate actions can high-quality software emerge from the SDLC. Every phase of the SDLC is rife with iterative activities and steps that require dedicated time and efforts aimed at quality. Applying the principles of resilience and security to the SDLC itself improves the methodology and the outputs of the process at the same time.

DOMAIN 5

CRYPTOGRAPHY

Cryptographic Concepts, Methodologies, and Practices

Chapter 16

Cloud Cryptography

Jeff Stapleton

Contents

Introduction

Cryptography is presumed to play a critical role in securing the cloud, much as it is assumed that cryptography secures the Internet. To discuss cryptography's role in securing the cloud, one must first understand the cloud. The chief technology officer (CTO) for the Central Intelligence Agency (CIA) [1] provided a whimsical definition:

- Cloud, n. A visible mass of vapor, especially one suspended in the sky
- Cloud, v. To darken; obscure; threaten
- Cloud, it. The single most overhyped term in the history of information technology

Despite the rather witty approach, the cloud is essentially outsourcing to a service provider. The National Institute of Standards and Technology (NIST) provides a list of five essential

characteristics of the cloud [2] that distinguishes it from legacy service providers. Each trait affects the overall security and the associated use of cryptographic methods:

- On-demand self-service—the ability for cloud scribers to unilaterally provision computing capabilities in an automated manner is a challenge to establish cryptographic keys securely.
- Broad network access—the availability of cloud services over the network using the standard mechanism for heterogeneous thin or thick client platforms may restrict cryptographic methods to the lowest common denominator.
- Resource pooling—the management of cloud provider resources to service multiple subscribers in a multitenant environment necessitates cryptographic key separation per subscriber.
- Rapid elasticity—the provisioning of cloud provider resources to scale rapidly with demand requires comparative key distribution and key termination processes.
- Measured service—the fee structure for using provider resources will vary; however, offering higher fees for more secure environments may incentivize subscribers to rely on weaker cryptographic methods.

The Cloud Security Alliance (CSA) Security Guideline [3] identifies two primary assets: data and applications. The protection afforded these data should be commensurate with the importance and sensitivity of the information. Information protection includes data confidentiality, integrity, and authenticity. For the purposes of this chapter, nonrepudiation is technically defined as integrity and authenticity provable to a third party, the legal and regulatory issues notwithstanding. Cryptography can provide all the information protection depending on the cryptographic algorithm and methods.

Cryptography

Cryptography can provide data confidentiality and privacy, data integrity and authenticity, and arguably nonrepudiation. The Payment Card Industry Data Security Standard (PCI DSS) [4] recognizes the storage, processing, or transmission of cardholder data. Data storage includes databases, files on hard drives, e-mails on servers, removable media, and even printed copies. Processing is data residing in the memory of laptops, desktops, servers, network appliances, and mainframes. Data transmission includes cables, networks, and wireless connections.

Data Confidentiality and Privacy

For data confidentiality or privacy, encryption can be used to protect the data during transmission, processing, or storage. Processing is when data, such as the executable code or sensitive authentication tokens, are kept encrypted in the memory and are decrypted only when needed. Refer to Figure 16.1. For data in transit (a), data can be encrypted at point A, safely transmitted, and decrypted at point B. For data in storage (b), data can also be encrypted at point C, stored securely, and decrypted at point C. For either transmission or storage, either symmetric or asymmetric keys can be used. The same symmetric key would be used by A to encrypt these data and would be used by B to decrypt these data, whereas for asymmetric keys, the public key would be used by A to encrypt these data and the private key would be used by B to decrypt these data. However, since modulo arithmetic limits these data, size and exponentiation are computationally intensive; asymmetric cryptography is typically used to establish a symmetric key. The asymmetric keys are used periodically to establish symmetric keys and the symmetric keys are used for every transmission or storage access.

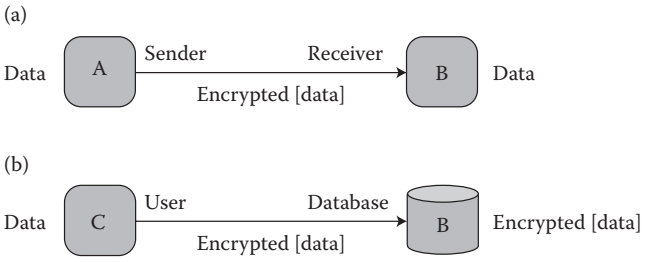


Figure 16.1 Encryption for data confidentiality and privacy.

Data Integrity and Authenticity

For data integrity and authenticity, message authentication codes (MACs) can be used to detect data modification or substitution during transmission, processing, or storage. Refer to Figure 16.2. For data in transit (a), the MAC is generated using a symmetric key and both the data and the MAC are sent from point A to point B, where the MAC is recalculated and compared to the received MAC. If the recalculated MAC and received MAC match, then there is a high probability that these data are unaltered, signifying that their integrity has been preserved. Since only A and B share the symmetric key, no other parties could have generated the MAC; so, the authenticity of these data is also confirmed. For data in storage (b), the MAC can also be used at point C to store these data and MAC, to ensure that these data have not been altered or substituted. Another MAC solution, called a hashed message authentication code (HMAC), is generated using a hash algorithm, symmetric key, and data.

Nonrepudiation

For nonrepudiation, digital signatures (DS) might be used to provide integrity and authenticity provable to a third party. Refer to Figure 16.3. For data in transit (a), similar to an MAC or HMAC, these data and the signature are sent from point A to B where the signature is verified using the corresponding public key. Since the digital signature can only be generated by the private key and theoretically, only point A has access to the private, nonrepudiation that is achievable because neither point B nor any other party could have cryptographically generated the signature. For data in storage (b), the digital signature can be used at point C to store the data and signature to ensure the data has not been altered or substituted.

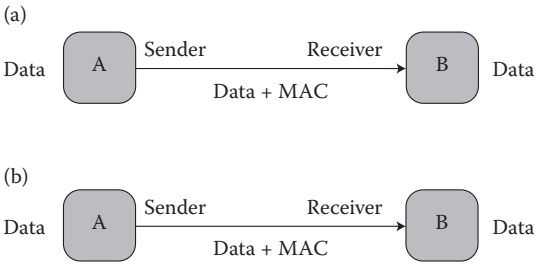


Figure 16.2 Message authentication for integrity and authenticity.

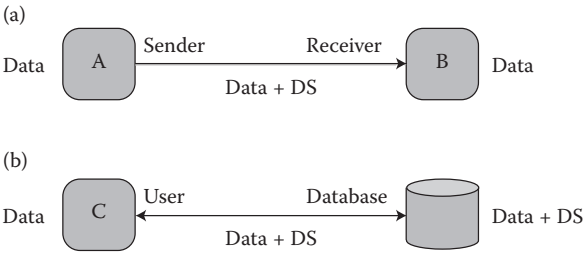


Figure 16.3 Digital signature for nonrepudiation.

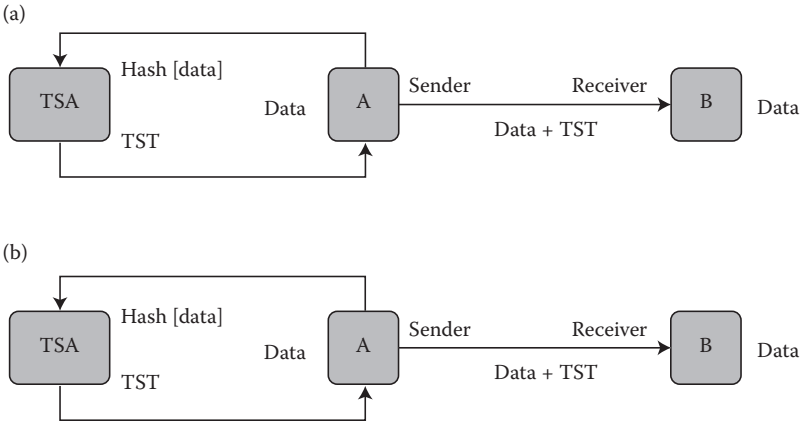


Figure 16.4 Time stamp token for nonrepudiation.

Likewise, for nonrepudiation, X9.95 [12] defines time stamp tokens (TST) that might be used to provide integrity and authenticity provable to a third party with a verifiable time stamp [12]. Refer to Figure 16.4. For data in transit (a), similar to an MAC, HMAC, or digital signature, these data and the TST are sent from point A to B where the TST is verified using the information from the transportation security administration (TSA), such as its public key to validate the TSA digital signature of the TST. Because the TST can only be generated by the TSA, the data integrity is independently verifiable to a trusted time source. For data in storage (b), the TST can be used at point C to store the data and TST to ensure the data has not been altered or substituted. Further, if signed data (i.e., the data shown in Figure 16.4 are really data + DS) are used, then the TST provides integrity over these data and the signature that provides strong nonrepudiation as point B can not only prove that point A generated the digital signature but can also prove when the digital signature was generated.

Cloud Security

For cloud service, the endpoints discussed in the section "Cryptography" can be translated as the cloud subscriber (point A) sharing the data with its cloud provider (point B) and the cloud provider

(point C) storing information locally. Alternatively, the endpoints can be mapped to a cloud provider (point A) sharing the data with its cloud subprovider (point B). Consider the three cloud scenarios:

- Multitenant provider
- Cloud subprovider
- Cloud customer

First, consider the various security services cryptography can provide when a subscriber (X) outsources services (Y) to a cloud provider as shown in Figure 16.5. Encryption can be used to protect the transmission of these data between the subscriber and the provider from authorized disclosure. Encryption might also be used to protect the storage of data by the cloud provider. However, encryption by itself cannot provide integrity or authenticity; so, a combination of an MAC, HMAC, digital signature, or TST should be used in conjunction with data encryption. For data transmission, both parties need to verify that the data received are unaltered (integrity) and that the data come from the legitimate (authenticity) sender. Further, both parties need to verify that the data stored by the cloud provider likewise have integrity. If the provider is merely storing and regurgitating the data to the subscriber such that any data changes are made by the subscriber, the same integrity value (e.g., MAC) used during transmission can also be reused for data storage. However, if the provider is processing these data, then a separate integrity value (or mechanism) must be used for storage different than the integrity value used for data transmission. For example, the data transmission between the subscriber and the provider might be an MAC whereas the provider might use TST for local storage.

Multitenant Provider

Consider a multitenant provider as shown in Figure 16.5. The provider (P) services four separate subscribers such as A, B, C, and X. For data transmission and integrity, different keys need to be used for each communicating pair (P and X), (P and A), (P and B), and (P and C). Note that the provider can use the same asymmetric key pair to establish separate symmetric keys. For example, Secure Sockets Layer (SSL) or transport layer security (TLS) can be used to establish a data encryption key and HMAC key using the public key of the provider. However, many of the SSL and TLS key establishment schemes require both the endpoints to use asymmetric key pairs. Public keys could be exchanged manually; however, the exchange of public key certificates for cloud environments is obviously a more effective method.

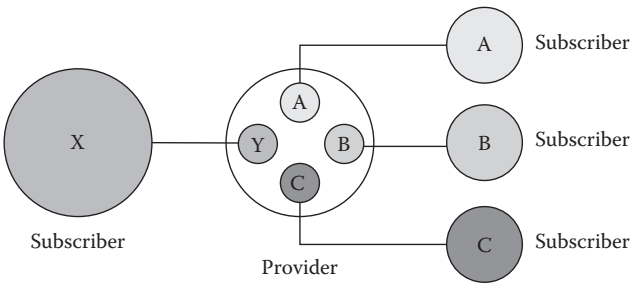


Figure 16.5 Cloud provider.

Further, for the provider to enforce data segregation between each subscriber, separate keys should be used. When separate keys are used for each subscriber, the provider could provide a copy of the keys to each subscriber as part of its data backup and recovery services. For example, if the provider keeps the encrypted backup copies of each subscriber in escrow, then in the event that the provider goes out of business, the subscriber can recover its encrypted data.

Cloud Subprovider

Consider when a cloud provider outsources some or all of its services to another provider as shown in Figure 16.6. Subscriber X relies on the provider for services Y; however, some or all the services may be further off-loaded to a subprovider. Off-loading by a provider to a subprovider might occur for various legitimate reasons:

- The provider might be a reseller of the subprovider services.
- The provider might need additional computing resources to meet the processing demands and availability agreements of its subscribers.
- The provider might benefit from cost savings provided by the subprovider.
- The provider might be an alternate name doing business as (DBA) of the subprovider.

In addition to the issue of whether the subscriber agreement allows, prohibits, or even addresses the provider using subproviders, another set of issues is the use of cryptography and key management.

In this example, the subprovider is also a multitenant environment supporting other subproviders such as subscriber 1, subscriber 2, and subscriber 3. Hence, all the issues for the multitenant provider apply equally well to the subprovider. Another issue is whether the cryptographic keys shared between subscriber X and the provider are likewise shared with the subprovider, or alternatively, the provider and subprovider share a separate set of keys. In Figure 16.5, the provider had multitenants so that if the provider and subprovider share a separate set of keys, the provider might protect all its subscribers' data using the same set of keys, or possibly, the provider and subprovider might use separate keys for each subscriber (X, A, B, and C). Each provider will undoubtedly have different solutions, but whose data are protected by which keys and who has the access to use those keys are a critical key management and security issue.

Another issue is how subscriber X accesses the services Y at the subprovider. The provider might act as a front-end to the subscriber and might communicate with the subprovider on the back-end. However, in today's web services world, a transparent redirect is more often the solution.

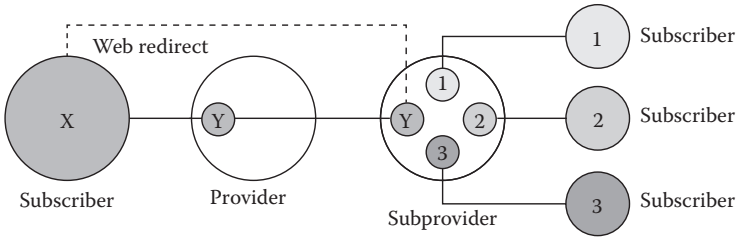


Figure 16.6 Cloud subprovider.

When a redirect occurs, the subscriber is typically unaware that the service resource has changed. This is another challenge for key management. If SSL or TLS is being used, then two separate sessions can be negotiated; one between the subscriber and the provider, and another between the subscriber and the subprovider. If other or additional methods are used, then the keys must be managed securely.

Cloud Customer

Consider when a subscriber is not the end user of the services but rather the subscriber's customers are the end users, as shown in Figure 16.7. Subscriber X offers services Y to customer Z relying on the provider for those services. The cryptographic keys shared between the subscriber and the provider should not be the same keys used by the customer, but one of the issues is whether the customer gets the cryptographic keys from the subscriber or from the provider.

Further, the subscriber most likely has numerous customers ($Z_1 \ldots Z_n$) who might be accessing the same subscriber data (e.g., white papers) or customer-unique data (e.g., pricing quotes, reports). Thus, each customer may need separate cryptographic keys to prevent one customer from accessing another customer's data. For example, if different pricing quotes for customers Z_1 and Z_2 were transmitted using the same encryption key and if Z_1 obtained the encrypted pricing quotes for Z_2, then Z_1 would be able to decrypt the Z_2 data. But, if there were different encryption keys for each customer, this would not be possible, or at least the data leakage would not be the responsibility of the subscriber who used the unique keys per customer as part of its due diligence.

Conclusion

The three cloud scenarios (multitenant provider, cloud subprovider, and cloud customer) have shown the potential complexity of using cryptography and the subsequent intricacy of managing keys for the cloud. Figure 16.8 shows all the three scenarios in one diagram. One of the higher risks of cloud computing is the natural business tendency to implement minimal security due to the cost and difficulty of securing subscribers, providers, subproviders, and customers. Ironically, cryptography as a service [5] already exists in the cloud, such as online certificate authorities, time stamp authorities, and even the global positioning system. Whether or not other cloud provid-

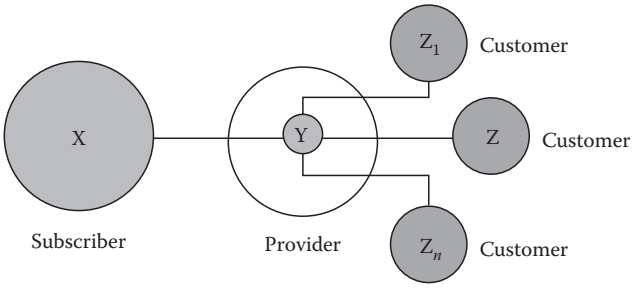


Figure 16.7 Cloud customer.

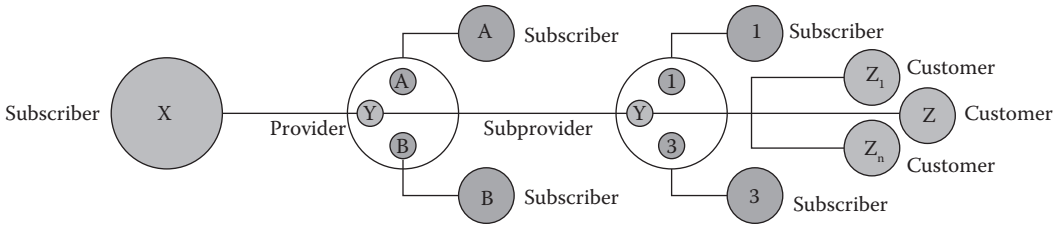


Figure 16.8 Cloud complexities.

ers will embrace cryptography and implement sound key management policy and practices is an unknown quantity at this time.

The Cloud Consumer Advocacy Questionnaire and Information Survey [6] conducted in 2011 recommended the encryption of data for different users to address the comingling of data with other cloud customers. The survey indicated that 74% responded "yes" when asked if the cloud service provider (CSP) adheres to any established governance framework involving data security controls. Of those CSP that adhere to a governance framework, 79% undergo regular third-party audits, and 67% allow their customers to audit the CSP data security controls. For encryption and key management practices, 84% of the CSP provide end-to-end encryption for the data in transit, and 69% provide encryption for the data at rest. The survey noted that although encryption is a point of emphasis with many CSP, the key management implementations are highly dependent on the provider and need to be vetted carefully according to the needs of the tenant.

The Security Guidance for Critical Areas of Focus in Cloud Computing [3] defines 14 domains to evaluate a cloud provider and identify the gaps: (1) cloud computing architecture framework; (2) governance and enterprise risk management; (3) legal issues; (4) compliance and audit management; (5) information management and data security; (6) interoperability and portability; (7) traditional security, business continuity, and disaster recovery; (8) data center operation; (9) incident response; (10) application security; (11) encryption and key management; (12) identity, entitlement, and access management; (13) virtualization; and (14) security as a service.

The Cloud Controls Matrix [7] addresses 11 control areas: (1) compliance; (2) data governance; (3) facility security; (4) human resources; (5) information security; (6) legal; (7) operations management; (8) risk management; (9) release management; (10) resiliency; and (11) security architecture. The compliance controls include independent and third-party audit. The data governance controls include data classification, information leakage, and risk assessments. The information security controls include policy, encryption, and key management. The security architecture controls include data security, application security, and data integrity. The control areas are cross-referenced to numerous other audit and assessment standards, including control objectives for information and related technology (COBIT), Health Insurance Portability and Accountability Act (HIPAA), International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) 27001, NIST Special Publication 800-53, and PCI DSS.

The Consensus Assessments Initiative Questionnaire [8] provides detailed questions that an information security practitioner can use to determine a cloud provider's compliance to the Cloud Controls Matrix [7].

Standards are being developed. For example, NIST is developing further guidelines for the cloud [2,9,10] and the Accredited Standards Committee X9, certified by the American National Standards Institute (ANSI) to develop national standards for the financial services industry, is

developing the X9.125 [11] standard. Whether or not the United States will take the X9.125 work and submit it to ISO for international standardization is unknown at this time.

Cryptography is being developed. For example, format-preserving encryption (FPE) algorithms using modern cryptography are available to encrypt the data in such a manner that preserves the original nature of these data. Given a 16-digit credit card primary account number (PAN), the encrypted result is another 16-digit number, eliminating the need to modify applications to accept the random strings of bits produced by the standard encryption algorithms. Another research area is homomorphic encryption where operations performed on the encrypted data transfer back to the decrypted data. Theoretically, a subscriber could provide the encrypted data to a provider for processing, and then decrypt these data and see the correctly processed result without the provider ever having an access to the original data.

Protocols are needed. For example, a subscriber should be able to validate that a provider is compliant to an agreement by receiving verifiable information. Further, an independent third-party auditor would be able to validate that a provider is compliant to a set of security controls common to multiple subscribers. Another example might be a subscriber's ability to validate the geographical location of its data by the provider's facility for compliance to national laws.

Information security practitioners have a rich set of control objectives and evaluation criteria to perform a risk assessment of any cloud provider prior to engaging services. Lesser tools exist to validate the compliance during service engagement and fewer tools are available to validate post engagements. What is truly needed is a cryptographic architecture common to all cloud providers that subscribers and customers can rely upon, regardless of the cloud model and independent of the application services.

References

1. Hunt, I.A. 2011. Big data, big bets, big opportunities. Central Intelligence Agency, October 2011.
2. NIST Special Publication 800-145, NIST definition of cloud computing, September 2011.
3. Security guidance for critical areas of focus in cloud computing v3.0, Cloud Security Alliance, 2011.
4. Payment Card Industry Data Security Standard v2.0 (PCI DSS), October 2010.
5. Stapleton, J. 2009. Cryptography as a service. *ISSA Journal*, 7(1), 17–21, January 2009.
6. Cloud Consumer Advocacy Questionnaire and Information Survey (CCAQIS), Cloud Data Governance Project, Cloud Security Alliance, 2011.
7. Cloud Controls Matrix (CCM) v1.2, Cloud Security Alliance, 2011.
8. Consensus Assessments Initiative Questionnaire (CAIQ) v1.1, Cloud Security Alliance, 2011.
9. NIST Special Publication 800-146 DRAFT, Cloud computing synopsis and recommendations, May 2011.
10. NIST Special Publication 800-144, Guidelines on security and privacy in public cloud computing, December 2011.
11. ANSI draft X9.125, Cloud services compliance data standard, 2012.
12. ANSI standard X9.95, Trusted time stamp management and security, 2010.

DOMAIN

6

SECURITY ARCHITECTURE AND DESIGN

Principles of Security Models, Architectures, and Evaluation Criteria

Chapter 17

Identity and Access Management Architecture

Jeff Crume

Contents

Introduction

Identity and Access Management (IAM) is a fundamental discipline within the realm of information security. After all, it does little good to encrypt data, block malformed network packets, quarantine malware, and the like if there exists no means to ensure that only authorized users have access to sensitive data. The *principle of least privilege* requires that there be some mechanism to ensure that the only access granted be that which is necessary for each individual user to perform the duties appropriate to their job, as determined by organizational policy. Implicit in that definition are the functional capabilities referred to as "the 4As" of IAM:

- Authentication (Are you who you claim to be?)
- Authorization (Are you allowed to do that?)
- Administration (How do I manage the previous two As?)
- Audit (Did I do the previous 3As correctly?)

Authentication

Authentication is fundamental to the entire process. The system must know, within a reasonable level of certainty, who the user is or security enforcement breaks down. Users are typically authenticated based upon something they:

- Know (e.g., password, PIN, information about the user, etc.)
- Have (e.g., security token, smart card, mobile phone, etc.)
- Are (e.g., a biometric characteristic such as a fingerprint, iris scan, voiceprint, etc.)

Each method has its own set of strengths and weaknesses based upon cost, complexity, usability, manageability, and effectiveness (e.g., likelihood of false positives and false negatives). Potentially any of these approaches can be breached so it is customary to use multiple methods in conjunction, known as *multifactor authentication*, when higher levels of security are required.

Authorization

Being able to correctly identify *who* a user is, though, is not sufficient. The system must also know *what* that user is allowed to do—which data they may access and which services they may invoke. Much attention is given to the process of authenticating users, as well it should be, but perhaps the more difficult task is that of determining if that user is authorized to perform the action they are requesting.

The reason for this is that security policy may range from a very simple case which states that all users within a specific group, for example, customer service agents, are allowed to transfer funds on behalf of a customer. A more complex example might also add conditions which require a more fine-grained approach to authorization.

For instance, customer service agents may only be permitted to transfer funds:

- Up to a prescribed limit within a given 24-hour period
- Only for certain classes of customers to which they have been assigned
- Only when access systems from within the corporate network

- Only during that agent's typical working hours
- Only when strongly authenticated via multiple factors

Clearly, determining if all these conditions have been satisfied will require a level of in-depth analysis that goes well beyond merely verifying the identity claim of the user.

Administration

Provisioning user accounts, managing changes to identity data, and ensuring that accounts are disabled and/or deleted from the system when they are no longer needed (i.e., deprovisioning) requires coordination of a complex set of tasks such as:

- Determining which access rights are appropriate for a given user
- Gathering approvals for those rights
- Ensuring that separation of duty policies have not been violated
- Recertifying that existing privileges are still justified on a periodic basis
- Interfacing with the various operating systems, applications, data bases, web servers, authentication systems, and so forth to actually create or delete accounts and manage privilege levels

For many enterprises, administration tends to be the most expensive and the most error prone of the 4As since it requires numerous manual interventions across different parts of the organization. The inefficiencies introduced by these interactions make this administration task a prime candidate for automation. Introducing the notion of *roles*, which group together common sets of users and privileges, can also help simplify the task by codifying standard processes.

Audit

The 4th A, audit, involves logging actions taken on various systems to ensure that the previous three As have been done correctly. For instance, an effective audit process can identify authentication failures, authorization anomalies, and administrative actions that exceed predetermined thresholds. When these occur, alerts can be sent to notify appropriate personnel to take corrective action as well as longer term trends identified through periodic reporting.

IAM: Typical Practice

Historically, IAM tasks have been performed on a per system basis. Often this is the result of internal organizational issues that make it more expedient to delegate these responsibilities to local administrators and application developers as part of their prescribed job responsibilities. This means that to provision an account on a given system, the administrator for that system is contacted. Often a paper form is submitted to document the request and this form must first be routed to the appropriate approvers to ensure that the request is reasonable and necessary and falls within security policy guidelines.

This means that when a new employee is hired and their job requires accounts on multiple systems, this manual process must be repeated separately for each individual system. This activity can be time consuming and costly as it can typically take on the order of 2 weeks or more to complete and involves the efforts of a dozen or more people to approve and provision the accounts.

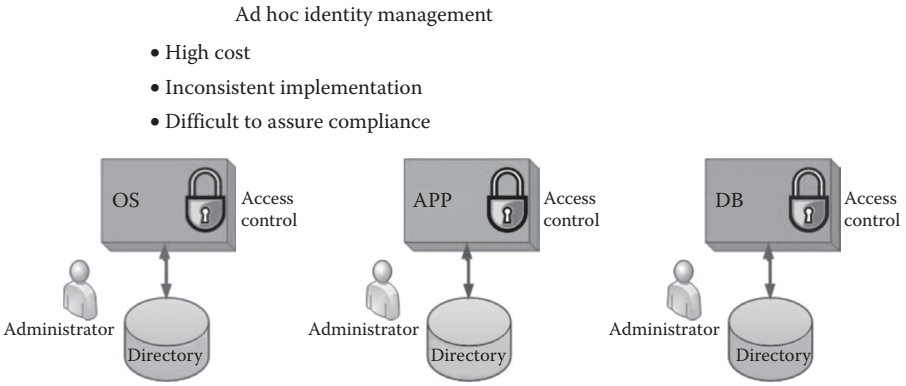


Figure 17.1 An ad hoc approach to identity management typical but not best practice.

By the same token, unraveling the web of access rights and accounts that need to be disabled when an employee is terminated or changes job roles can be equally, if not more, inefficient, and runs the risk of leaving accounts orphaned and ripe for abuse.

This ad hoc approach to identity management, shown in Figure 17.1, is a typical practice but is far from being a best practice. Auditors frequently flag systems handled in this manner due to the inconsistent application of security policy. Business leaders pay the price both in terms of higher overhead costs as well as compliance failures, which can threaten to shut down an organization's livelihood if not addressed promptly.

Alternative Architecture: Integrated Decentralized

A better approach to identity management that provides a greater degree of sharing among the various "silos of identity" involves synchronizing identity data across disparate systems to provide a more comprehensive view. This integrated, yet still distributed, architecture, shown in Figure 17.2, allows for the same degree of autonomous control as the ad hoc approach along with the additional visibility and coordination of relevant identity data.

For instance, directory synchronization technology can be used to ensure that whenever a user's e-mail address changes, that all other identity stores pick up this change as well. This approach also offers the option of providing a higher level identity store, or *enterprise directory*, which reflects identity data gathered from various sources across the organization.

This "bottom up" architecture represents an improvement over its ad hoc counterpart, due to the addition of this ability to integrate what would have otherwise been a series of isolated data stores, and is best suited to organizations which require a highly decentralized approach, often due more to organizational considerations rather than technical ones.

Alternative Architecture: Centralized

Better still is an even more integrated architectural alternative that provides for centralized administration and control, often with some level of decentralized, delegated administration for particularly complex tasks. This centralized approach, shown in Figure 17.3, not only integrates related identity data across the various directories but goes a step further and provides a common

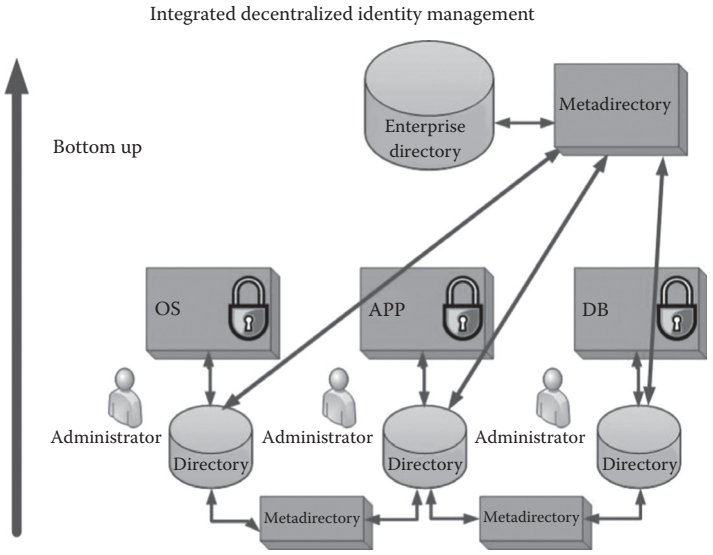


Figure 17.2 An integrated, yet distributed, architecture allows for the same degree of autonomous control as the ad hoc approach along with the additional visibility and coordination of relevant identity data.

administrative interface as well as common authentication and authorization services. These services can then be leveraged and reused across the entire organization to provide:

- More consistent adherence to security policy
- Quicker time frames for account provisioning and change management

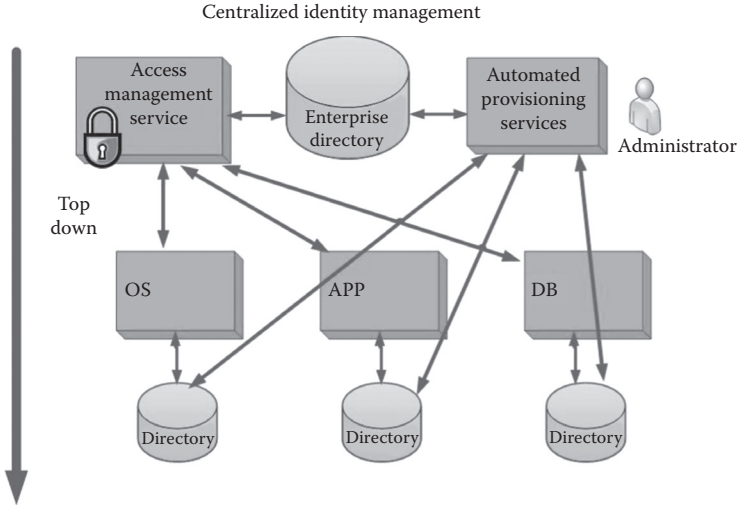


Figure 17.3 A centralized approach not only integrates related identity data across the various directories but goes a step further and provides a common administrative interface as well as authentication and authorization services.

- Better auditing through a common logging point
- Lower costs through economies of scale, automation of repetitive, manual operations, and reduction in administrative personnel (although domain experts will likely still need direct access to systems for more complex, one-off tasks not well suited for automation)

The IAM Ecosystem

From an organizational perspective, IAM functions are best implemented as part of a more holistic, comprehensive system rather than as a set of autonomous, piecemeal tools, processes, and personnel. To achieve this more integrated approach, shown in Figure 17.4, each of the key components of the system must work together. The following is a description of these elements and how they fit into a larger IAM conceptual architecture.

Storing Identities

Identity data is the lifeblood of an identity management system. Some examples of this sort of data include

- User names
- User accounts
- Access control rights
- E-mail addresses
- Phone numbers
- Job roles
- Location
- Department, and so on

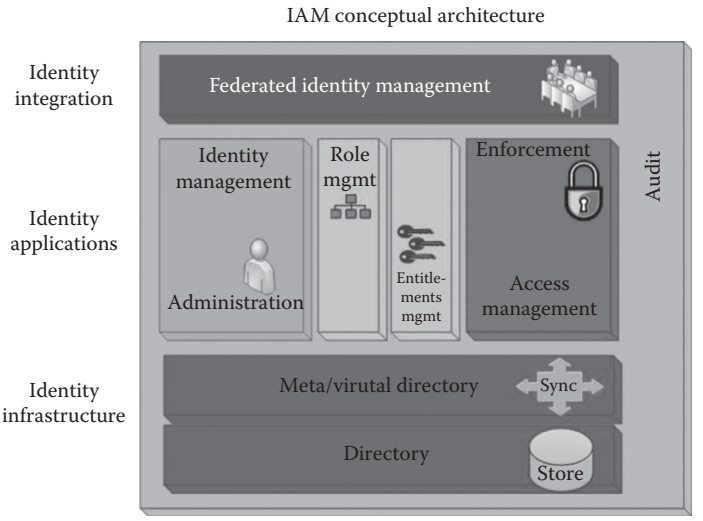


Figure 17.4 From an organizational perspective, IAM are best implemented as part of a more holistic, comprehensive system rather than as a set of autonomous, piecemeal tools, processes, and personnel. To achieve this more integrated approach, each of the key components of the system must work together.

Any IAM system will need the ability to store identity data for future use. Any place that these data are stored is a *directory*. A typical directory implementation will be comprised of

- A store for keeping the data
- A structure for organizing the data
- An interface for accessing the data

Enterprise-class directories often require high-performance directory data access. For this reason, some commercial implementations are built on enterprise-class data base technologies. Often the actual details of the underlying data base are effectively hidden once the directory has been installed leaving only the benefits of the robust data store with the complexity of the supporting technology out of sight.

In addition to having a place to store identity data, there needs to be a systematic, consistent way of organizing the data. The way these data are arranged is called a *schema*. The directory schema provides the overarching definition of which identity data will be kept and how it relates to other identity data.

Finally, there needs to be a mechanism for accessing data in the directory—both reading and writing identity data. The most popular industry standard for this is the Lightweight Directory Access Protocol (LDAP). The LDAP standard describes a set of message formats and data flows for interacting with a directory to read and write entries. By definition, LDAP is not, technically speaking, a directory in and of itself, but rather the means for accessing a directory.

Even though most technicians tend to think of only in terms of LDAP-accessible directories, in fact, most directories within an IT infrastructure vary widely in their construction and interface methods. Identity data may be stored directly in data base tables, proprietary data stores, spreadsheets, or flat files and be accessible via a wide variety of industry standard and proprietary APIs. Schemas may be loosely defined with store and retrieval methods limited in scope and functionality. Nevertheless, all of these varied identity stores need to be considered in the larger IAM architecture to provide complete coverage.

Integrating Identities

Most organizations have numerous directories in place. They may place most of their focus on just a few, key standards-based or mission-critical proprietary deployments, but to achieve the sort of benefits mentioned previously such as improved security, lower cost, higher efficiency, and better compliance, it is necessary to create linkages among all key identity stores. Organizations that lack this integration must deal with duplicate, conflicting identity data, which may be out of date and require manual intervention to resolve.

The two primary approaches, shown in Figure 17.5, designed to deal with these issues are

- Directory virtualization
- Directory synchronization

A *virtual directory*, as its name implies, provides the illusion of a single, integrated data store by maintaining an index of all relevant identity data along with where that data is stored. This way all requests for identity data can be directed to a single logical component, the virtual directory, which then routes the request to the appropriate authoritative source transparently. To improve performance, virtual directories often also include a cache which stores recently accessed data

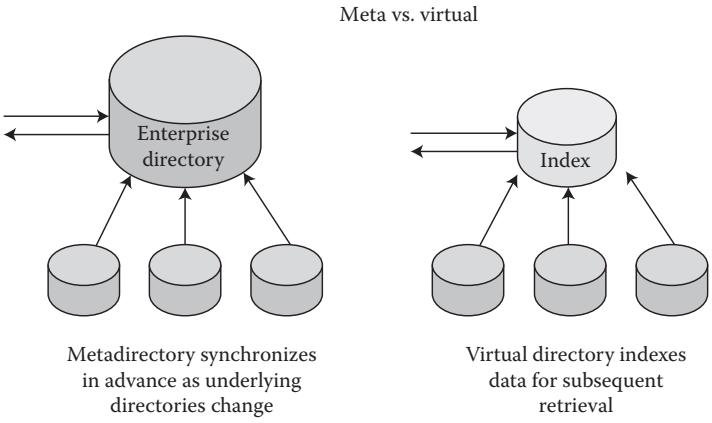


Figure 17.5 To achieve benefits such as improved security, lower cost, higher efficiency, and better compliance, it is necessary to create linkages among all key identity stores. Organizations that lack this integration must then deal with duplicate, conflicting identity data, which may be out of date and require manual intervention to resolve. The two primary approaches designed to deal with these issues are directory virtualization and directory synchronization (metadirectory).

locally so as to avoid the latency that would otherwise result from having to transmit requests to what might be a remotely located authoritative source.

A *metadirectory* implementation typically takes a different approach. Rather than waiting for a request to arrive and then calling out to the appropriate underlying directory, it synchronizes changes to potentially far flung directories at a single *enterprise directory*. Like the virtual directory, this method provides a single point for accessing directory data, but in this case the data has already been fetched in advance and can, therefore, be served up locally without relying on whether it has been previously cached or not.

From a performance perspective, virtual directories tend to be optimized for *write* operations since changes to identity data at a local directory do not require additional processing. On the other hand, metadirectories that synchronize data to a common enterprise directory are better optimized for *read* operations since they do not require additional access to other identity stores.

Within the context of IAM, typically there are far more reads performed than writes. For instance, provisioning a user account would involve writing information into a directory. This would happen once to set up the account and then again on occasion when modifications are necessary. However, the process of authenticating that user and authorizing subsequent requests would typically occur many orders of magnitude more frequently.

Note that while these principles are generally true, there may, in fact, be exceptions since hybrid approaches are not uncommon in real world implementations.

Administering Identities

Directories, along with a mechanism to integrate them, are fundamental components within an IAM architecture. As such, they form the *identity infrastructure* that is needed to store, synchronize, and access identity data. As powerful as these foundational technologies may be, though, they are not sufficient to address the full scope of issues that go along with the 4As of administering,

authenticating, authorizing, and auditing. In addition, there is a need for a set of *identity applications* that build on this identity infrastructure and augment these basic capabilities.

In this context, administration involves an identity lifecycle which begins with *provisioning* (i.e., creating accounts) and ends with *deprovisioning* (i.e., deleting accounts) and may involve any number of modifications along the way. In ad hoc IAM environments these tasks are handled separately by domain-specific experts who are narrowly focused on the concerns of users for a specific system or set of systems. Tools native to that system are typically used to fulfill these requirements with little concern for the larger organizational impact.

A more holistic approach is often better at dealing with IAM issues across a diverse set of heterogeneous systems. Achieving maximum operational efficiency along with consistent adherence to security policy requires the use of tools which can automate and coordinate administrative tasks. An enterprise-class identity management system would likely include many, if not all, of the following capabilities:

- Identity request initiation
- Role management to identify which accounts/privileges need to be created, modified or removed
- Workflow engine to orchestrate the process
- Request approval tracking to obtain and document responses
- Interfaces to managed systems to perform approved actions
- Reconciliation processing to ensure that managed systems remain in compliance
- Self-service user interface for resetting passwords and modifying user information

Each of these functions will be described in the following sections.

Identity Request Initiation

To affect a change of some sort through the identity management system, a request for that change must be initiated. Typically, such changes occur through either of the two methods:

- An automated identity feed from another system
- A manual request via a user interface

The first method is often preferred since it occurs as a natural consequence of some other business process and, therefore, is seen as more responsive to changes occurring within the organization. An example of this would be linking the provisioning process to a human resources (HR) data base, which contains a list of all employees along with key information needed by the identity management system such as

- Employee name
- Job code
- Department
- Location
- Employment status
- Supervisor

As previously noted, any data store containing this sort of identity data should rightly be considered a directory. Therefore, it follows that the connection between this HR "directory" and the

IAM system could be accomplished through the sort of directory synchronization/virtualization technology described earlier. These tools possess the ability to key off predefined triggers in the source directory to initiate subsequent processing in a target directory. In this case, the target is the actual IAM system, itself.

By leveraging this capability, the addition of a new employee into the HR system can serve as the trigger to automatically begin the process of provisioning appropriate accounts and access rights for this new user of the organization's IT systems. By the same token, a termination or job responsibility change entered into the HR system could automatically kick off a request to disable or modify access rights granted to this employee.

While this automated approach is often preferred, there are times when it may not be entirely practical. For some organizations, updates to the HR system may be delayed by days or even weeks even though the need to disable access for a terminated employee may require near real-time processing. In other cases, exceptions arise where a user may need additional access rights that may not be typical for an employee of performing that same job role.

Regardless of the reason, an effective IAM system should also provide a means for manually creating requests for provisioning/deprovisioning. A web page which presents a list of identity management functions and systems to which those functions should be applied is a typical alternative.

In either case, though, whether it is through an automated interface from another system or from a manual request entered by a user, the identity management system can proceed through its normal processing for fulfillment.

Role Management

Just as the overall IAM architecture may evolve into a series of siloed, ad hoc systems, the same can occur when it comes to provisioning accounts. For example, a user requests access to a particular system and that request is granted. Another user does the same and that request is granted, and so forth. Each request is treated as a one-off event with no thought given to the larger impact on long-term management of these accesses once granted, nor how future changes to systems should be handled.

Figure 17.6 illustrates how this purely reactive approach to granting *entitlements*, or permissions for users to access certain resources, can grow into an entangled web of relationships that becomes increasingly difficult to visualize, much less control.

A better approach in many circumstances would be to create groupings which allow for common, consistent treatment of frequently repeated requests. These abstractions are referred to as *roles* within the IAM system and, if designed appropriately, can greatly simplify the task of identity management, especially in large organizations where scalability is critical.

As Figure 17.7 illustrates, users can be grouped together by their job function, location, level, and so on to create a *business role* while entitlements can be grouped together to represent a higher level capability or transaction to form an *application role*. In this example, therefore, granting the ability to let doctors admit patients now requires only a single step of adding the ADMIT application role to the DOCTOR business role. Without roles, a change of this sort would have required identifying all users who are doctors and granting this privilege to each one separately.

Further, if a new lower level entitlement was suddenly required to discharge patients due to some organizational policy change or regulatory compliance requirement, then all users with the ability to discharge patients would have to be identified and their access rights modified. Such a

Without roles: Typical practice

User:
- The entity requesting access to a resource
- Ex: John Smith, AppXYZ

Resource:
- Ex: App, data base, service, and so on

Entitlement:
- A permission to access a particular resource
- Ex: Open table, read record, write record

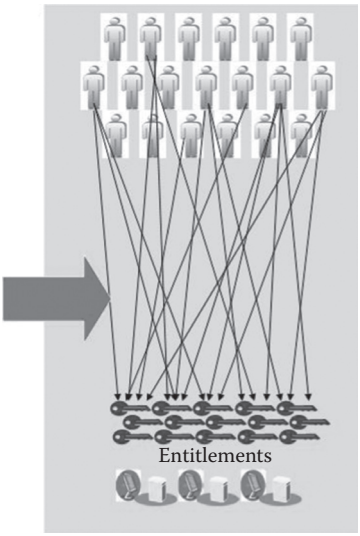


Figure 17.6 A reactive approach to granting *entitlements*, or permissions for users to access certain resources, can grow into an entangled web of relationships that becomes increasingly difficult to visualize, much less control.

With roles: Best practice

User:
- The entity requesting access to a resource
- Ex: John Smith, AppXYZ

Resource:
- Ex: App, data base, service, and so on

Entitlement:
- A permission to access a particular resource
- Ex: Open table, read record, write record

Business role:
- A logical collection of users performing a similar business function
- Ex: Doctor, nurse, lab technician

Application role:
- A logical collection of entitlements needed to perform a particular task
- Ex: Create patient record, admit patient, discharge patient

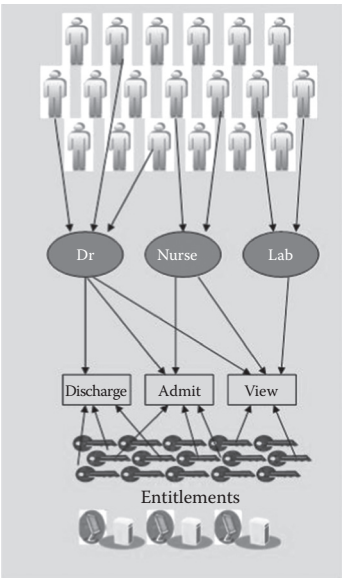


Figure 17.7 Users can be grouped together by their job function, location, level, and so on to create a *business role* while entitlements can be grouped together to represent a higher level capability or transaction to form an *application role*.

process would likely be time consuming and error-prone due to its manual nature. However, with the sort of role management scheme illustrated here, this same change could be accomplished with a single modification of the DISCHARGE application role and the effects would automatically propagate to all users authorized to perform this function.

It is worth noting that many of the first-generation role management systems provided only a single layer of role abstraction. This meant that both user and application function information had to be encapsulated into a single-layered role construct. Problems often arose from the fact that too much information was being packed into each role making it highly specific not easily generalized and applied to larger numbers of users. As a result, the ratio of roles to users, which ideally should be very small since it is this reduction that creates the desired economies of scale, were often lost and roles became brittle and unable to adapt quickly to change. Consequently, many early adopters of role management schemes were unsatisfied with the results.

The sort of second-generation role management capabilities available now do not suffer from these limitations. Not only do they provide the ability to provide multiple abstraction layers for role types (e.g., business role, application roles), but they also allow for hierarchical *child roles* which inherit characteristics from higher level *parent roles*. For instance, in the example considered here, the DOCTOR role might contain all the access rights that all doctors would need. However, some doctors might need additional capabilities due to the nature of their work so child roles could be created for certain specialists such as radiologists, pediatricians, and oncologists to handle their unique requirements. This way a pediatrician would not only get access to their specialty-specific functions but also would automatically inherit the more general entitlements that all members of the DOCTOR role receive.

Choosing the right arrangement of roles for a given organization has also been simplified in recent years with "bottom up" *role mining* tools which examine existing directories looking for patterns within access rights for given groups of users and "top down" *role engineering* tools which facilitate the formulation of new roles based upon knowledge of the organization by an expert.

Approval Processing and Workflow Tracking

Just because a request is received for a new account or entitlement does not necessarily mean it should be granted. Therefore, the IAM system must first obtain concurrence from designated approvers. Often, these approvers organizationally reside within the various lines of business that own the applications and data in question rather than IT personnel who traditionally have been thrust into the role of acting as gate keepers for resources that ultimately belong to others. Simplifying the methods used to gather and record approvals in ways that do not require significant IT skills allows these functions to be delegated to more appropriate personnel who ultimately possess the responsibility for the use of protected resources.

By using electronic notifications such as emails to approvers when they have requests awaiting their action and providing a simple, intuitive web-based interface for them to examine incoming requests and respond with their concurrence or denial, the goal of putting control of data in the hands of the data owners can be achieved.

At the heart of this approval processing, a workflow engine is needed which can map out who the various approvers are and in which sequence their actions are needed. For instance, for a user to be granted access to a particular system, the organization's policy may dictate that either of two designated supervisors plus a representative from the line of business approve. The workflow engine would, therefore, need to follow the prescribed sequence and orchestrate the sending, receiving, and documenting of approvals.

The approval processing component may also be required to allow overrides in cases where exceptions need to be granted and reminders sent out when approvers are late in responding. It may also be desirable to have some level of automated policy enforcement within this component to comply with the *separation of duties* requirements, making sure that conflicting privileges are not granted in the first place.

Interface to Managed Systems

Once all the required accounts have been identified, either through manual specification or automatic role processing, and approvals have been obtained, the IAM system must interface with the various systems on which these accounts will reside to create, modify, or delete the appropriate access rights. Complicating this task is the fact that there are a wide variety of system-specific mechanisms for achieving this. Some systems may simply rely on an underlying directory which supports the LDAP industry standard. Others may require invoking data base functions to write to specific records. Still others may only work with proprietary APIs or command line interfaces. In all cases, the job of the IAM system is to smooth over these irregularities by working out the details and providing a consistent level of function.

To accomplish this, the IAM system can leverage the same sort of directory integration approach described previously. For instance, the same technology that enables a metadirectory to communicate with a wide range of HR systems to provide input to the provisioning process can also be deployed to deliver the output of the effort by exploiting the administrative interface supported by each managed system. In a sense, the metadirectory acts like the conduit which connects any directory to any other directory and it is precisely this generic interface capability that makes it so valuable as both a feeder and manipulator of identity data.

Reconciliation Processing

Much of the value of an IAM derives from its ability to ensure compliance through consistent application of security policy. However, if managed systems are not locked down in such a way that domain administrators can use local interfaces to modify identities on those systems then malicious or inadvertent violations can occur by subverting the approval process. For this reason, it is necessary to implement a *reconciliation* process which can automatically compare the accounts and privileges that *should be* on the various managed systems to those that *actually are* in place. Reconciliation processing can be initiated to run periodically during nonpeak hours as needed and respond to inconsistencies in any of the following ways:

- Notify appropriate personnel of the unauthorized change.
- Automatically disable or delete unapproved accounts.
- Rollback privilege levels that have been increased inappropriately.

When done properly, the reconciliation process functions as an automated audit with corrective actions and serves to "close the provisioning loop" by ensuring ongoing compliance.

User Self-Service

Over time, users may need to modify identity information about themselves. Ideally, these changes would flow automatically in most cases from changes occurring in other sources. An example of

this would be a name change, which would be entered into the HR system and then fed to the IAM system and then ultimately propagated to the various directories across the organization. In other cases, users may need to reset passwords they have forgotten.

A web interface into the IAM system allows users to handle these tasks on their own without requiring a call to the Help Desk or other administrative personnel. The most frequent call to most Help Desks is for password resets. In some cases, this can be 50% or more of all calls and, therefore, represent a majority of the cost of running this service. As a result, empowering users to perform these simple administrative tasks can result in substantial savings. *In many cases, this savings alone can more than offset the cost of implementing the entire IAM system,* a rarity in the realm of IT Security where most expenditures are unable to pay for themselves through actual cost savings.

Enforcing Access Controls

Once a user's accounts have been provisioned, they will then proceed to use those accounts. At that point, the focus shifts from creating and managing access rights to *enforcement* of those privileges. In other words, the emphasis of the IAM system moves from *administration* to *authentication* and *authorization*.

Authenticating Identities

To determine if a request, such as retrieving data or executing a transaction should be allowed, it must first be determined if the user is, indeed, who they claim to be. As mentioned previously, this *authentication* process is based upon something that the user knows, has, or is. For instance, a password or PIN may serve as a bit of secret knowledge that proves the user's identity since only that user knows what it is. Alternatively, the user may be required to present a smart card or enter information such as a constantly changing one time password (OTP) displayed on a security token to prove possession of the authentication device. Still other options involve measuring certain physical characteristics of the user such as their voice, face, iris, fingerprint, etc. to establish that they are authentic. Each of these approaches carries with them important advantages and disadvantages that need to be considered when designing an authentication system.

Something You Know

Knowledge-based authentication has the advantage of being relatively easy to deploy since it requires no special hardware and authentication secrets (e.g., passwords, PINs) can be changed frequently as needed. However, security systems based on this technique can be easily subverted if the secrets they rely on are not well guarded or well chosen. Allowing end users to pick their own passwords typically results in trivial choices that are easy for them to remember but may also be easy for an attacker to guess. Since most users, if left to their own devices, will choose passwords that relate to themselves such as the names of family members, pets, hobbies, and so on, the more an attacker knows about the user, the more likely they are to successfully guess their password. *In general, if a user can remember a password, an attacker can also guess it.*

Security organizations typically attempt to deal with this problem of trivial passwords by creating complexity rules that require the use of alphabetic characters (sometimes both upper and lower case) to be used in conjunction with numeric and special characters to mitigate this risk and also subvert attempts to guess a password through brute force methods, which involve attempting a wide variety of possible combinations.

However, if a user is forced to pick a password that is sufficiently difficult to guess, then the password may also be sufficiently difficult to remember, resulting in the user writing down the password for future reference. This creates a different set of problems since recorded passwords now require physical security protections to prevent disclosure. Unfortunately, most users are not likely to guard these with the necessary care.

An additional security precaution involves requiring that users choose different passwords for each system to minimize risk in the case where a password is compromised. This only multiplies the problem of memorizing passwords essentially guaranteeing that they will be written down.

Complicating matters further is the fact that passwords are often forgotten and need to be reset. In fact, *the better the password from a security standpoint, the more likely it is that it will be forgotten* due to its complex nature. This results in an often unanticipated support cost as users must place a call to a help desk for assistance. For many organizations such calls cost $20 to $30 per incident, erasing any savings that might have been achieved by avoiding the deployment of security tokens or biometric readers.

Single Sign-On

Single sign-on (SSO) tools can greatly reduce the issues introduced by security policies which force users to keep up with multiple, complex, unique passwords, which need to change on a regular basis. Rather than insisting on what ultimately results in an untenable situation that can actually lessen security rather than strengthen it, SSO allows for the automatic generation and maintenance of strong, randomly chosen passwords that are unique to each system while requiring the user to keep track of only one password to unlock the SSO tool, itself.

With SSO, the user simply logs into the tool and then the tool provides the authentication credentials automatically to any system requesting them. Some SSO tools can also automatically generate new, random passwords when prompted by the target system that the current password has expired. The advantage to the user is that the tool hides all the complexity of password generation, storage, and recall so users are less inclined to write passwords down, less inclined to forget the one password they need to remember, and less likely to choose trivial passwords since the tool handles the complex details. This results in improved usability, lower support costs, and better security.

A common objection to SSO approaches, however, is that they effectively put all their eggs in one basket, so to speak. An attacker now only needs to steal one password to get access to all the systems for which the user has accounts. *However, since users tend to choose the same password for the systems they access, this risk already exists, unbeknownst to the security department.* Adding SSO is, therefore, not increasing this risk that already exists. Further, a way to mitigate this exposure would be to require *multifactor authentication*, which combines something the user knows, has, or is, in order to access the SSO tool.

Something You Have

Knowledge may exist in more than one brain simultaneously but a physical token cannot be in two places at once. For this reason, possession-based authentication systems are generally regarded as more secure than knowledge-based systems, although, this is not necessarily always the case. With possession-based systems, the user is required to have some designated security token with them to complete the authentication process. They may need to insert a smart card into a designated

reader, hold a security device near an RFID/NFC reader, or enter a constantly changing one-time password displayed on the device.

The disadvantage to this approach is that the user must keep the token with them, which may be inconvenient, tokens can malfunction or be lost or stolen, additional cost and complexity is required to acquire and deploy tokens and readers and the tokens, themselves, while generally secure, have been known to be compromised through a variety of attacks. For these reasons, possession-based methods are often combined with knowledge-based methods to form a *second-factor* authentication system, which lessens the risk of compromise by requiring an attacker to defeat both controls.

Something You Are

Biometric-based authentication systems seek to overcome many of the disadvantages of knowledge and possession-based systems by measuring certain physical characteristics of the user and matching them to previously stored data. They provide greater convenience by not requiring the user to keep up with any special security tokens but still carry with them the cost and burden of deploying special readers wherever users will need to authenticate. Since biometric measurements can vary from time to time due to fluctuations in the physical environment (e.g., changes in appearance, sound, environmental conditions, etc.), a certain degree of error must be tolerated or there is a risk that a legitimate user will not be recognized by the system (i.e., false negative/reject). By the same token, if too much variation is allowed, there is a risk that someone who is not the authorized user may be incorrectly identified as a match (i.e., false positive/accept). Striking the right balance of preciseness of match is key to optimal operation of the system. Also, care should be taken that the biometric reader has not been compromised so as to prevent impersonation or replay attacks. As with possession-based systems, biometric-based systems frequently combine multiple authentication factors to lessen the risk of compromise.

Authorizing Identities

Once a user's identity has been established, the IAM system must then determine whether the requested action is permitted by security policy. This authorization task involves two steps: decision and enforcement.

Enforcement tends to be relatively easy since it ultimately involves either allowing a requested action to proceed or blocking it and providing appropriate feedback. The *decision* step may be a great deal more complicated since it must consider all the conditions involved before rendering a permit or denying outcome.

Referring to a previous example, a customer service agent at a bank may be permitted to perform money transfers as a result of being a member of the "Agent" role. However, the bank may have a policy that prohibits junior agents from transferring greater than a set limit or across multiple geographic units or restricts activity to a limited set of accounts. A more senior agent may be afforded a higher transaction limit and be given greater latitude in terms of accounts. There may be additional policies against performing such transaction during certain times of the day or days of the week as well as restrictions on the physical location of the agent performing the transfer.

As a result, a complex set of conditions need to be tested to determine whether the request should be authorized or not. Further still, the criteria may change over time as new company policies or regulatory requirements are implemented.

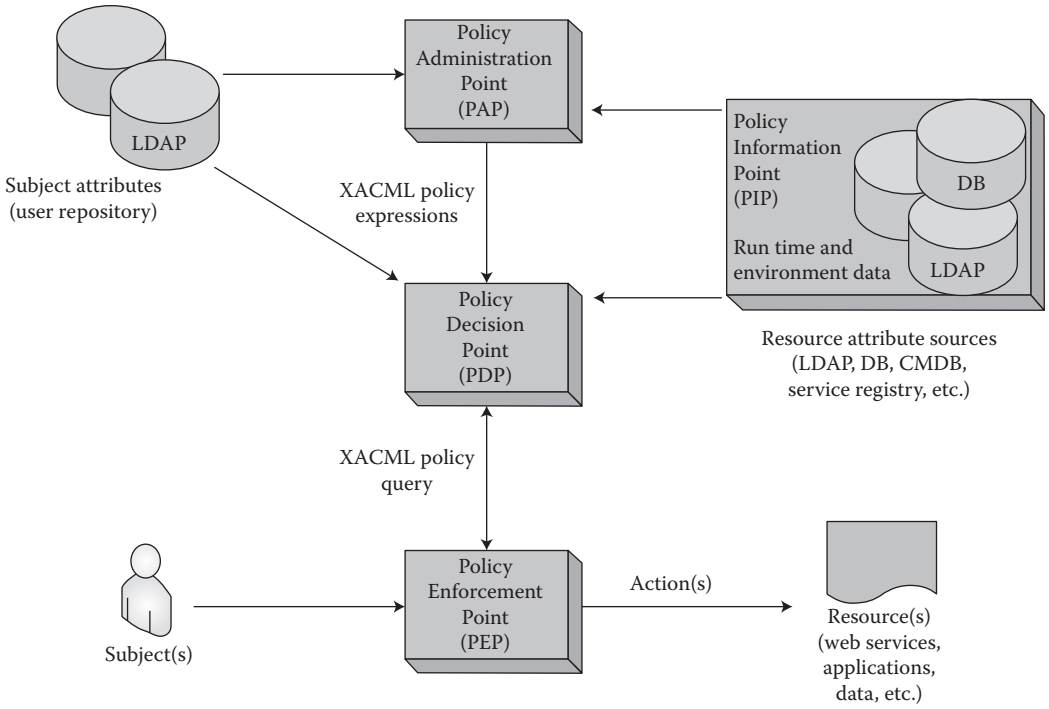


Figure 17.8 A more flexible approach involves encapsulating decision logic such that it may be invoked by multiple modules, which then perform enforcement locally.

Traditionally, application developers have imbedded these checks deep within the various programs involved. The problem with this approach is that changes are difficult to implement since they may involve multiple programming changes in a variety of code segments.

A more flexible approach, shown in Figure 17.8, involves encapsulating decision logic such that it may be invoked by multiple modules, which then perform enforcement locally. The eXtensible Access Control Markup Language (XACML) OASIS industry standard refers to the components which perform these functions as a Policy Decision Point (PDP) and Policy Enforcement Point (PEP), respectively. XACML also provides a protocol which allows a PEP to query a PDP for an access control decision and receive a response which it can enforce.

This modular architecture concentrates the more complex decision processing into a centralized service that:

- Improves consistency of access control policy enforcement across all applications since they rely on the same authorization component.
- Improves flexibility and adaptability to changes in policy by minimizing the number of places that changes must be implemented.
- Minimizes the amount of time spent by application developers in designing and implementing access control logic reusing existing capabilities allowing them to focus less on security and more on functionality.
- Decreases deployment time and costs by establishing a consistent, repeatable security model which leverages a proven infrastructure.

Extending Identities

The components and capabilities described thus far comprise the primary elements necessary to implement an enterprise-wide IAM system. However, there may be cases where identities need to extend beyond the boundaries of the organization to interoperate with the disparate IAM systems of other organizations.

An example of this might involve a case where the employees of a retailer need to interact with various suppliers and customers to place and fulfill orders. In this case, the retailer's IAM system needs to not only provision accounts for its own users but also ensure that its employees have appropriate levels of access to business partner systems. As well, there may be a requirement that some of the retailer's customers be able to access the retailer's systems using identities supplied by their organizations.

To accomplish this in an automated, seamless manner across heterogeneous systems with varying policies implemented using technologies from different vendors, a common set of protocols for *federated identity management* is needed.

In the aforementioned first case, the retailer acts as an *identity provider* for its own employees by creating local accounts and vouching for their identities with industry standard credential which can be forwarded to the supplier's system acting as a *service provider*, which consumes the identity assertion and allows access to desired services. In the latter case, the retailer acts as a service provider by receiving credentials created by the customer's system and granting access to those users who have been verified by their organization serving as their identity provider.

Federation provides a means for exchanging these identity assertions and can be used to automate account creation and update requests originating on the identity provider system and fulfilled by the service provider through a system of *federated provisioning*. The same relationship may apply for authenticating users by translating credentials from the identity provider into a native format, which can be processed by the service provider systems. This way a user from the identity provider system can benefit from a *federated single sign-on* experience across IAM systems from different organizations.

There exist a wide range of evolving federation standards such as SAML (Security Assertions Markup Language), WS-Federation, OpenID, OAuth, and others. Some, like WS-Federation, are able to imbed other standards while others are less interrelated.

It is important to bear in mind, though, that while the technology exists for extending identities across organization boundaries as described, some of the more complicated aspects transcend the technical realm. Resolving issues involving the threshold necessary to establish mutual trust, legal responsibility, when disputes arise, and harmonizing the differences among differing security policies across federation partners are just a few examples of these nontechnical concerns.

Conclusion

IAM systems are fundamental components of a larger security infrastructure. Many organizations struggle to manage identities due to the complexities of trying to administer, authenticate, and authorize users across a diverse set of incompatible, standalone systems. Ad hoc approaches are inadequate to address the larger needs of the organization, which is why a more holistic focus based on an end-to-end IAM architecture is preferred. Establishing and reusing common services for identity storage, integration, provisioning, and access enforcement provides a more flexible, efficient, auditable, and secure environment which can be extended through the use of federation technologies, as needed.

Chapter 18

FedRAMP

Entry or Exit Ramp for Cloud Security?

Debra S. Herrmann

Contents

The U.S. government agencies that plan to use cloud computing products and services are now required to use products and services that have

1. Been through a formal security evaluation by an accredited third-party assessment organization (3PAO)
2. Received a provisional authorization from the FedRAMP Joint Authorization Board (JAB)

This chapter describes the FedRAMP security evaluation process and the roles and responsibilities of cloud service providers (CSPs) and vendors, 3PAOs, system integrators, and federal agencies.

Background

On December 8, 2011, the U.S. Office of Management and Budget (OMB) issued a memo to the chief information officers (CIOs) of all U.S. federal government agencies announcing the formation of the FedRAMP program and stipulating new security policies to which all federal agencies must comply. The FedRAMP program is the culmination of an 18-month joint effort by the OMB, the

National Institute of Standards and Technology (NIST), the Department of Homeland Security (DHS), the Department of Defense (DoD), and the General Services Administration (GSA). DoD, DHS, and GSA are members of the FedRAMP JAB. The name of this new program is appropriate since the intent is to "ramp up" the security posture of federal agencies. The initial focus is to standardize the security requirements and assessment criteria for cloud computing products and services. Over time, the scope of FedRAMP may expand to other areas.

Methodology

The FedRAMP program has incorporated some of the lessons learned from the decade-long experience with the Federal Information Security Management Act (FISMA) and the Common Criteria Testing Labs (CCTLs)—FedRAMP's first cousin—into its methodology.

Earlier, the OMB produced an annual FISMA scorecard. At first, each federal agency was given a letter grade (A through D) to reflect their level of compliance with FISMA requirements; later, a switch was made to color codes (green, yellow, and red). The performance appraisal of an agency's CIO was often tied to their FISMA scorecard. These grades were based on agency self-reporting and an audit of an agency's security certification and authorization (C&A) records. Unfortunately, the quality and rigor applied to the C&A process differed from agency to agency and even project to project within an agency. More than one agency that received an A or green FISMA rating also made the news for major data breaches.

FedRAMP is designed to correct this situation through the use of independent accredited 3PAOs, standardized assessment criteria, and oversight by the FedRAMP JAB. In addition, more emphasis is placed on security assessments, inspections, and testing than auditing.

The security assessments and inspections performed by an accredited 3PAO are similar to a product or type certification. Like the CCTLs, 3PAOs will evaluate cloud computing products and services—not complete end-user systems deployed by a federal agency. In the beginning, the CCTLs received a lot of bad press for taking too long; 2 years was the time frame often cited. The problem was not the CCTLs or the common criteria methodology. The problem was that vendors showed up at the CCTLs with a finished product and expected the CCTLs to "test" their product and make it secure. In reality, the infamous 2 years was spent reverse engineering the missing life cycle documentation to get a product ready for a common criteria evaluation.

As any security engineer worth her salt knows, security engineering is a life cycle engineering endeavor that starts with security requirements. To avoid the same pitfall, the FedRAMP assessment labs are referred to as 3PAOs, not FedRAMP security testing labs. Furthermore, FedRAMP encourages early coordination between CSPs and 3PAOs to ensure that vendors understand their roles and responsibilities before formal security assessments and inspections begin.

On January 8, 2012, the FedRAMP program posted the mandatory technical, management, and operational security controls for the low- and moderate-risk baselines. The high-risk security control baseline will be defined later. All federal agencies must classify their IT systems in accordance with FIPS PUBs 199 and 200 as being low, moderate, or high risk based on predefined confidentiality, integrity, and availability characteristics. The new FedRAMP security requirements are additions and enhancements to the security requirements currently defined in NIST SP 800–53. A few examples of the additional language that has been incorporated are listed below. These new security requirements are intended to mitigate the security and privacy concerns inherent in cloud computing, like those shown in Table 18.1.

Table 18.1 Security and Privacy Concerns Related to Cloud Computing That FedRAMP Is Designed to Mitigate

I. Data Protection
• Possession of personally identifiable information (PII), authentication data, access control lists, access control rights and privileges, sensitive agency data, mission-critical data, and safety-critical data is entrusted to a CSP, their subcontractors, and suppliers
• Different countries have different laws governing requirements for protecting the security and privacy of the data the CSPs possess
• Misappropriation of data with little or no legal remedy, especially in the case of wireless cloud computing
• Externally stored data are subject to corporate espionage, international espionage, and cyber terrorism
• Theft of intellectual property that is in the possession of CSPs
• Limited ability to migrate data, recover data, or confirm data destruction upon contract termination, CSP bankruptcy, or purchase of a CSP by another corporation
• Verifying controlled access to offline archives
• Not knowing for sure where or in what country data are actually stored
• Ease of monitoring cloud communications
• Lack of data segregation, comingled data, inadequate logical separation
• Insider abuse and attacks
• Physical security controls alone are inadequate
II. Secure Operations
• Continuity of service during natural or man-made local, national, or international disasters
• Enforcing adherence to priority of service, service-level agreements (SLAs), and quality-of-service (QoS) requirements
• Lack of security for stored kernels
• Multitenancy where systems, services, and data are shared by unrelated users
• Difficulty or impossibility of accessing security audit logs
• Loss of control over security patch management and upgrades
• No redundancy or diversity to meet reliability, maintainability, or availability (RMA) requirements; if the CSP enterprise is down so is the user community (e.g., if an electric utility goes down, a million people sit in the dark for 3 days with no air conditioning, hot water, or lights, and they cannot do anything about it)
• Common use of open source software by CSPs

continued

Table 18.1 (continued) Security and Privacy Concerns Related to Cloud Computing That FedRAMP Is Designed to Mitigate

• External management of cloud security services, lack of accountability
• External ownership of security management tools and consoles
• Spoofing by posing as legitimate cloud clients by purchasing cloud services for nefarious purposes
• Browser compromise, contamination, and hijacking
• Complex encryption key management issues
• Complex security incident response, notification, and reporting

- *AC-17 remote access:* The networking protocols implemented by a CSP must be approved and accepted by the FedRAMP JAB.
- *AU-11 audit record retention:* Online audit records must be maintained by a CSP for 90 days. Retention of offline audit records must meet the requirements of the U.S. National Archives and Records Administration (NARA).
- *CM-7 least functionality:* A CSP must use the Center for Internet Security guidelines (Level 1) to establish a list of prohibited or restricted functions, ports, protocols, and services. This list in turn must be approved and accepted by the FedRAMP JAB.
- *IA-2 identification and authentication of organizational users:* A CSP must define replay resistant authentication mechanisms, which then must be approved and accepted by the FedRAMP JAB.

The FedRAMP evaluation process is documentation-heavy. For example, the Excel workbooks designed to capture the security inspection evidence and findings for just eight of the security controls produce over 100 pages of documentation.

Roles and Responsibilities

In addition to itself, the FedRAMP methodology defines four distinct roles, as shown in Table 18.2:

- Cloud service provider
- 3PAO
- System integrator
- Federal agency

The CSP is responsible for incorporating FedRAMP security requirements into their product or service and generating the associated security engineering artifacts. To do so, the CSP must first select whether they intend to adhere to and be evaluated against the low- or moderate-security control baseline. This choice is more than a security consideration—it is a strategic business decision. As shown in Table 18.3, if a product receives a moderate-risk provisional authorization, it can be used by federal agencies in both low-risk *and* moderate-risk systems. In contrast, if a product receives a low-risk provisional authorization, it can only be used by federal agencies in a low-risk system. This begs the

Table 18.2 Summary of Organizational Roles and Responsibilities within the FedRAMP Methodology

Role	*Responsibilities*

Cloud service provider	• Develops system security life cycle documentation for the product to be evaluated:
	  • System security plan
	  • Information security policies
	  • User guide
	  • Rules of behavior
	  • IT contingency plan
	  • Configuration management plan
	  • Incident response plan
	  • E-authentication workbook
	  • Privacy threshold analysis
	  • Privacy impact assessment
	  • Supplier declaration of conformity
	• Selects and implements FedRAMP security controls for the low or medium baseline
	• Completes:
	  • Security control tailoring workbook
	  • Security control implementation summary
	• Contracts with an accredited 3PAO to perform an independent assessment
	• Submits 3PAO designation form to FedRAMP
	• Submits items to be inspected to the 3PAO
	• Implements corrective action identified in the PO&AM
	• Maintains provisional authorization through continuous monitoring program
	• Complies with federal requirements for change control and incident reporting

3PAO	• Performs independent security assessments, inspections, and testing of cloud service provider products and systems • Creates security assessment package (security assessment plan, security assessment inspection evidence and findings, security assessment report) based on FedRAMP guidelines and standards • Participates in continuous monitoring activities

continued

Table 18.2 Summary of Organizational Roles and Responsibilities within the FedRAMP Methodology

Role	*Responsibilities*
System integrator	• Ensures that cloud computing products and services have had a 3PAO assessment
Federal agency	• May sponsor a 3PAO assessment of a cloud service provider product • May select a cloud service provider product that has already received a provisional authorization • Uses the FedRAMP methodology • Incorporates provisional authorizations into final system C&A • Ensures that contracts require cloud service providers to comply with FedRAMP requirements, especially maintaining a provisional authorization

question, why would a CSP go to the time and expense to have a 3PAO evaluated any product against the low-risk security control baseline? From a business perspective, it makes infinitely more sense to have all CSP products and services evaluated against the moderate-risk security control baseline.

Part of this exercise involves completing a security control tailoring workbook and generating a security control implementation summary. In addition, the CSP is responsible for generating the following NIST compliant documents:

- System security plan
- Information security policies
- User guide
- Rules of behavior
- IT contingency plan
- Configuration management plan
- Incident response plan
- E-authentication workbook
- Privacy threshold analysis
- Privacy impact assessment

Table 18.3 Use of Provisional Authorizations

Provisional Authorization Applicable to	*Can Be Used in Low-Risk Systems*	*Can Be Used in Moderate-Risk Systems*
Low-risk baseline of security controls	X	
Moderate-risk baseline of security controls	X	X

The CSP is responsible for selecting a 3PAO. Based on the CCTL experience cited earlier, it is highly recommended that a CSP and prospective 3PAO have several preinspection coordination meetings to ensure that each party thoroughly understands their roles and responsibilities, as well as the time table for performing specific activities. Once a decision has been made, the CSP submits a 3PAO designation form to the FedRAMP PMO. Lastly, the CSP submits their product and the associated security engineering artifacts to a 3PAO to start the initial formal security assessment and inspection. At the same time, a supplier's declaration of conformity that attests to the truth of the description and implementation of the security controls documented in the CSP-generated SSP is provided.

The 3PAO conducts the security assessment, inspection, and testing in accordance with FedRAMP standards and guidelines. This process involves producing the security assessment plan, capturing the security assessment evidence and findings in predefined Excel workbooks, and summarizing the results in a security assessment report. The 3PAO sends the complete security assessment package to the CSP, who forwards it to FedRAMP for review and approval.

If successful, the FedRAMP JAB will grant a provisional authorization for this particular version of a cloud computing product or service. The provisional authorization may be dependent on a CSP completing the corrective action identified in a plan of actions and milestones (PO&AM).

Once a provisional authorization is granted, the product or service enters a continuous monitoring phase, as defined in NIST SP 800–137. If significant changes are made to a cloud computing product or service during the continuous monitoring phase, a new security assessment and inspection may be warranted.

The three phases of the FedRAMP methodology and the NIST standards applicable to each are summarized in Figure 18.1.

Conclusion

FedRAMP is an ambitious initiative. Time will tell whether or not it is successful and accomplishes its stated goals. Four primary concerns are raised in this regard.

First, cloud computing is already a fact of life. It is difficult if not impossible to retrofit a security inspection methodology onto systems that are already up and running and were not designed to its specifications. If FedRAMP had been instituted 3 or 4 years ago, it would have a better chance of success.

Second, the FedRAMP methodology shifts a lot of responsibility for producing NIST-compliant security engineering artifacts onto the CSPs who are commercial vendors. It is unlikely that many if any of these commercial vendors have the experience or expertise on hand to do this. As was seen during the implementation of the common criteria methodology, commercial vendors do not have this experience or expertise, nor do they budget the time and resources necessary to produce these security engineering artifacts. Few security engineers who design and develop commercial products know how to develop a protection profile or security target; likewise, they will be unlikely to be able to produce the 14 NIST-compliant documents the FedRAMP methodology assigns to vendors.

The FedRAMP methodology prohibits 3PAOs from developing these 14 documents unless the 3PAO can demonstrate rigorous independence requirements in accordance with ISO/IEC 17020.

Third, the FedRAMP methodology gives federal agencies the ability to opt out if they cannot or do not want to comply. As part of their annual report to the Federal CIO, federal agencies must list all cloud services that the agency determines cannot meet FedRAMP requirements.

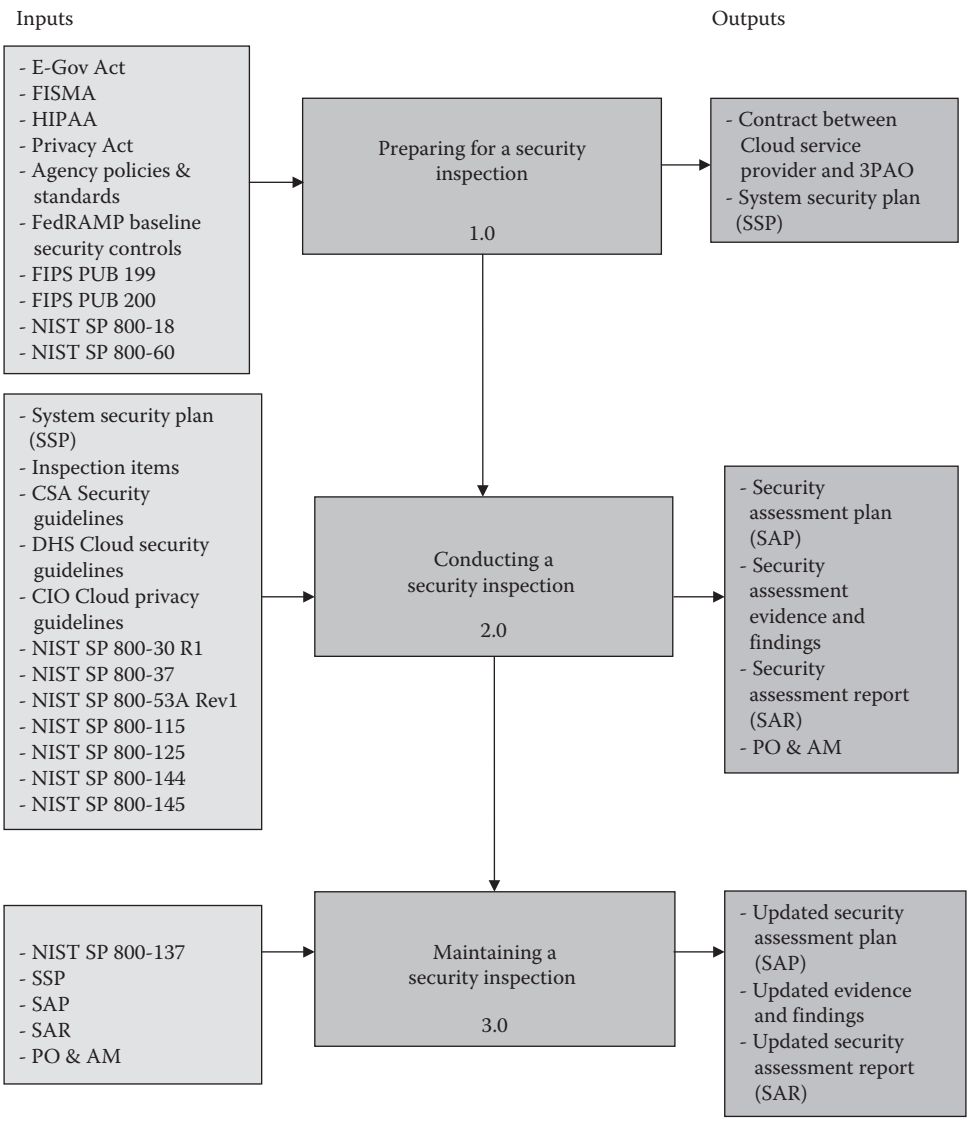


Figure 18.1 FedRAMP security inspection methodology.

Furthermore, agencies have the option of implementing the FedRAMP methodology internally, but not using an accredited 3PAO. Most federal agencies have their hands full trying to comply with mandatory requirements. It is possible that many will opt out.

Fourth, the idea of granting a provisional authorization that can be used by many federal agencies makes sense at a government-wide level. The same is not true on an agency-by-agency basis. In other words, why would agency A have any incentive to go to the time and expense associated with sponsoring an accredited 3PAO evaluation so that agency B can reuse the provisional authorization for free? Federal agencies are only concerned about their own budgets, not that of the entire federal government.

On the other hand, if the FedRAMP roll-out goes well, FedRAMP could become the gold standard for cloud security, in which case the state and local governments and private industry are likely to adapt the FedRAMP methodology as well. The FedRAMP PMO declared a successful interim operational capability (IOC) in June 2012. Full operation (FOC) was achieved in April 2013 and sustaining operation is scheduled to begin in FY14. The first provisional authorization was granted by the JAB in December 2012. The second provisional authorization was granted by the JAB in January 2013. Both provisional authorizations were for moderate risk infrastructure as a service (IaaS) offerings.

Hence, time will tell whether FedRAMP is the entry or exit ramp for cloud security in the federal arena.

Glossary

3PAO	Independent accredited FedRAMP third-party assessment organization
C&A	Security certification and authorization
CIO	Chief information officer
CCTL	Common Criteria Testing Lab
CSA	Cloud Security Alliance
CSP	Cloud service provider
DHS	U.S. Department of Homeland Security
DOD	U.S. Department of Defense
FedRAMP	U.S. Federal Risk Assessment and Management Program
FISMA	Federal Information Security Management Act
GSA	U.S. General Services Administration
JAB	FedRAMP Joint Authorization Board
OMB	U.S. Office of Management and Budget
PO&AM	Plan of actions and milestones that describes specific tasks and timelines for corrective action a CSP must take to receive or maintain a provisional authorization
SAP	Security assessment plan
SAR	Security assessment report
SSP	System security plan

Further Reading

1. Cloud computing from the security perspective, National Cyber Security Division, Department of Homeland Security, 2009.
2. Privacy recommendations for the use of cloud computing by federal departments and agencies, Privacy Committee, Web 2.0/Cloud Computing Subcommittee, Federal CIO Council, 2010.
3. Federal Information Security Management Act (FISMA) (P.L. 107–347, Title III), December 2002.
4. The Privacy Act of 1974, (P.L. 93–579), as codified at 5 U.S.C. § 552a (as Amended).
5. National Institute of Standards and Technology Special Publication 800–144, Guidelines on security and privacy in public cloud computing, (draft) January 2011.
6. National Institute of Standards and Technology Federal Information Processing Standards Publication 199, Standards for security categorization of federal information and information systems, February 2004.
7. National Institute of Standards and Technology Federal Information Processing Standards Publication 200, Minimum security requirements for federal information and information systems, May 2006.

8. National Institute of Standards and Technology Special Publication 800–18, Revision 1, Guide for developing security plans for federal information systems, February 2006.
9. National Institute of Standards and Technology Special Publication 800–30, Risk management guide for information technology systems, July 2002.
10. National Institute of Standards and Technology Special Publication 800–37, Revision 1, Guide for applying the Fisk management framework to federal information systems: A security life cycle approach, February 2010.
11. National Institute of Standards and Technology Special Publication, 800–53, Revision 3, Recommended security controls for federal information systems and organizations, August 2009.
12. National Institute of Standards and Technology Special Publication 800–53A, Revision 1, Guide for assessing the security controls in federal information systems and organizations—Building effective security assessment plans, June 2010.
13. National Institute of Standards and Technology Special Publication 800–39, Managing information security risk; organization, mission, and information system view, March 2011.
14. National Institute of Standards and Technology Special Publication 800–128, Guide for security focused configuration management of information systems, August 2011.
15. National Institute of Standards and Technology Special Publication 800–64, Security considerations in the software development lifecycle, October 2008.

DOMAIN 7

OPERATIONS SECURITY

Concepts

Chapter 19

Data Storage and Network Security*

Greg Schulz

Contents

As information technology (IT) moves farther from the relatively safe and secure confines of data center glasshouses and internal physical networks with interfaces for Wi-Fi mobile and Internet computing, security has become even more important than it was in the past. Cloud, virtual machine (VM), and storage networking with remote access enable flexible access of IT resources by support staff, users, and clients on a local and wide area basis. This flexibility, however, also exposes information resources and data to security threats. This means that any desired increased accessibility must be balanced between data protection and business productivity. As networked storage enables storage and information resources to be accessed over longer distances and outside the safe confines of the data center, more security threats exist and more protection is needed.

Security issues also increase as a result of networking with virtual and physical IT resources and applications or services being delivered. For example, a nonnetworked, stand-alone server and dedicated direct-attached storage with secured physical and logical access is more secure than a server attached to a network with general access. However, the stand-alone server will not have the flexible access of a networked server that is necessary for ease of use. It is this flexible access and ease of use that requires additional security measures. As new enabling technologies, including IP-based networks to facilitate distance, are leveraged, they also enable security threats and attacks. These attacks can occur for political, financial, terrorist, industrial, or sheer entertainment reasons.

Eliminating Blind Spots, Gaps in Coverage, or "Dark Territories"

It is important not to treat all applications, their data and associated infrastructure resources, and associated management the same. Using policies and procedures collectively called infrastructure resource management (IRM) can ensure the proper level of treatment. The security of information and related assets is an important part of IRM, including data management and different levels of protection to meet various threat risks. Business and threat analysis should be used to determine what to encrypt and the applicable level or granularity of encryption to be used. It is also important to eliminate "dark territories," blind spots, or gaps in coverage (see Figure 19.1).

Blind spots or gaps in coverage are not unique to security; enabling an agile, flexible, dynamic, resilient, and converged environment relies on having timely situational awareness of resources and service delivery. Because the focus in this chapter is on logical and physical security of data and information resources on both local and remote bases, the focus of removing dark territories or blind spots is to eliminate gaps in coverage that can result in points of vulnerabilities or threat risks.

When it comes to moving data electronically via a network transfer or by shipping physical media, you may know when and where it left as well as its estimated time of arrival (ETA), but do you know where the data was during transit or while in flight? Do you know who may have had access to it or may have been able to view its content, particularly if it was not encrypted? Can you provide auditable trails or activity logs of where the data moved or deviated from planned routes or paths?

In the transportation industry, terms such as "dark territory" have historically been used by railroads to indicate areas with minimum to no management or control coverage. Other transportation-related terms include "blind spots" or "flying blind" to indicate lack of situational awareness that can result in loss of management control. What these have to do with cloud and virtual data storage networking is that a "dark cloud" can be considered a resource without adequate insight and awareness of who has access to it and what they may be doing with it.

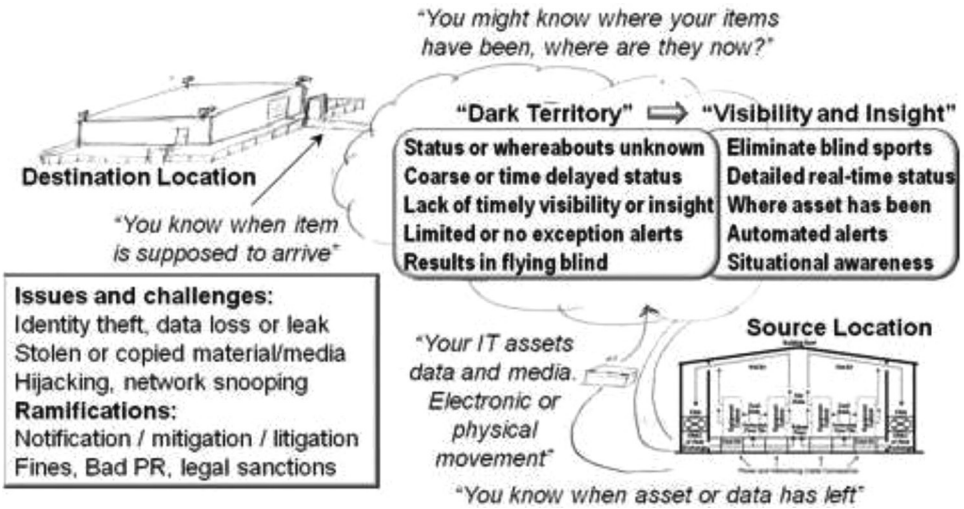


Figure 19.1 Eliminating "dark territory," "dark clouds," and blind spots.

At the top left of Figure 19.1, various technologies and techniques are shown that are used at the source and destination for managing digital assets and media. Also shown are issues and lack of real-time management insight while assets are being moved in blind spots.

For example, data needs to be moved to public and off-site remote private providers. Once data and applications are in use at public or private providers and on premise, what visibility is there into how secure information and associated resources are being kept safe? When information is being moved, is it via electronic means using networks or bulk movement using removable media (Flash SSDs), regular hard disk drives (HDDs), removable hard disk drives (RHDDs), optical CDs, or DVDs, or via magnetic tape? For example, to move a large amount of data initially to a cloud or managed service provider (MSP), a magnetic tape copy of the data may be made to be used for staging at the remote site, where it is then copied to a disk-based solution. What happens to the magnetic tape? Is it stored? Is it destroyed? Who has access to the tape while it is in transit?

Possible areas of "dark territory" or gaps in coverage include

- Public or private clouds that lack visibility into who is accessing resources.
- Shipping containers containing storage systems or media (SSDs, disks, or tapes).
- Lack of leak detection on public and private networking links.
- Physical and logical tracking of where data or storage media are during transit.
- Who has access to eDiscovery, search or data classification tools, and audit logs?
- What physical access and audit logs or trails exist, and how they are preserved?
- Tools including radio-frequency identification (RFID) for tracking assets.
- Physical security and logical or encryption for data in-flight and at rest.
- No video or logs for access to physical resources and facilities.

Security Threat Risks and Challenges

There are many different threat risks (see Figure 19.2) for IT cloud, virtual, and traditional data centers and the systems, applications, and data they support. These risks range from acts of man to acts of nature, and from technology failure to accidental and intended threats. A common belief is that most threat risks are external, when in reality most threats except acts of nature are internal. Firewalls and other barriers can work together to fight attacks from outside, but equally strong protection is necessary against internal threats. Another common security threat risk within most IT networks is inadequate security on "core" systems or applications within an environment. For example, poor password control on enterprise backup/recovery systems, virtualization systems, and management interfaces may be too common instead of being common sense to change.

Threats may be physical or logical, such as a data breach or virus. Different threat risks require multiple rings or layers of defenses for various applications, data, and IT resources, including physical security. The virtual data center relies on both logical and physical security. Logical security includes access controls or user permissions for files, objects, documents, servers, and storage systems along with authentication, authorization, and encryption of data.

Additional common threat risks include

- Logical or physical intrusion from internal and external sources
- Cybercrimes, virus, botnets, spyware, root kits, and denial of service (DoS)
- Theft or malicious damage to data, applications, and resources

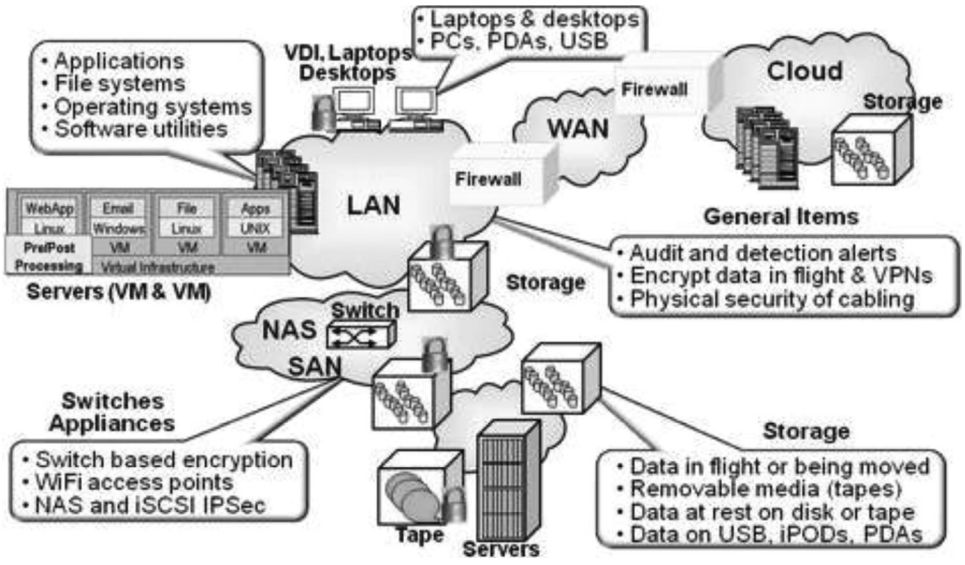


Figure 19.2 Cloud and virtual data storage networking security points of interest.

- Lost, misplaced, or stolen data, or pirated network bandwidth
- Regulatory compliance and information privacy concerns
- Exposure of information or access to IT resources when using public networks
- Internal or external unauthorized eavesdropping or sniffing
- Shift from private physical to virtual and public cloud resources
- Blind spots or dark territory and clouds with loss of visibility or transparency

Another facet of logical security is the virtual or physical destruction of digital information known as digital shredding. For example, when a disk storage system, removable disk or tape cartridge, laptops, or workstations are disposed of, digital shredding ensures that all recorded information has been securely removed. Logical security also includes how storage is allocated and mapped or masked to different servers along with network security, including zoning, routing, and firewalls.

Another challenge with cloud and virtual environments is how various customers' or business functions' applications and data are kept separate in a shared environment. Depending on the level of the shared or multitenant solution combined with specific customer, client, or information services consumer security and regulatory requirements, different levels of isolation and protection may be required. For example, on a shared storage solution, is having different customers or applications provisioned into separate logical units (LUNs) or file systems sufficient? As another example, for more security-focused applications or data, are separate physical or logical networks, servers, and storage required? In addition to multitenant hardware, software, and networks, either on your own premises under your management or via an on-site MSP or external provider, who has access to what, when, where, and for what reasons?

Additional security challenges include

- Subordinated and converged management of shared resources
- Mobile and portable media, PDAs, tablets, and other devices

- Encryption combined with deduplication, compression, and eDiscovery
- Orphaned data, storage, and other devices
- Classifying applications, data, and alignment of service-level objectives (SLOs)
- Growth of unstructured data, ranging from files to voice and video
- Converged networking, compute and storage hardware, software, and stacks
- Number of and diversity of log files to monitor as well as analyze
- International and multilanguage support via tools and personnel
- Automated policy-based provisioning
- Managing vendors and suppliers along with their access or end points

In addition to the above, other challenges and requirements include compliance requirements such as PCI (Payment Card Industry), SARBOX, HIPAA/HITECH, BASIL, and others. Security requirements for cloud, virtual, and data storage networks vary and include jurisdiction of specific regulations, fraud and data leak detection notification, data encryption requirements, auditable event, as well as access and activity logs.

Taking Action to Secure Your Resources

The security of your networks and systems is essential in normal times and crucial during service disruption. DoS attacks have become the new threat, causing disruptions and chaos. Some security issues to be considered include physical and logical security along with encryption of data, virtual private networks (VPNs), and virtual local area networks (VLANs). The security of the network should extend from the core to the remote access sites, whether home, remote office, or a recovery site. Security must be in place between the client and server (or the web), and between servers. Securing the home environment includes restricting work computers or PCs, use of VPNs, virus detection, and, of course, system backup. Security becomes more important the farther away you are from a secured physical environment, particularly in shared environments.

Common security-related IRM activities include

- Authorize and authenticate access
- Encrypt and protect data in-flight and at rest
- Monitor and audit activity or event logs
- Grant or restrict physical and logical access
- Monitor for data leaks and policy compliance

As with many IT technologies and services, there will be different applicable threat risks or issues to protect against, requiring various tiers and rings of protection. The notion of multiple rings or layers of defense is to allow for flexibility and enable worker productivity while providing protection and security of applications and data. A common belief is that applications, data, and IT resources are safe and secure behind company firewalls. The reality is that if a firewall or internal network is compromised, without multiple layers of security protection, additional resources will also be compromised. Consequently, to protect against intrusions by external or internal threats, implementation of multiple protection layers, particularly around network access points, is vital.

There are many things that can be done, ranging from protecting physical facilities and equipment to securing logical software and data. Securing coverage should extend in terms of visibility

and coverage from physical to virtual, from private to public, as well as MSPs. Other things that can be done include preserving segregated administration functions by various technology management groups (servers, operating systems, storage, networking, applications) in a converged, coordinated manner. This means establishing policies and procedures that span technology management domains along with associated visibility or audit tools. Security should also include leveraging encryption, certificates, and tokenization in support of authorization, authentication, and digital rights management.

DOMAIN

9

LEGAL, REGULATIONS, COMPLIANCE, AND INVESTIGATIONS

Information Law

Chapter 20

National Patient Identifier and Patient Privacy in the Digital Era

Tim Godlove and Adrian Ball

Contents

Information cannot be shared among health care providers if systems are not interoperable. The current lack of standardized technology not only prevents integration and interoperability of information systems but also severely limits the ability of electronic health records, often referred to as EHRs, to safely facilitate continuous, informed care across health care settings. In other words, health care information systems need to talk to one another.

Assigning everyone a universal patient identifier, or UPI, would improve a doctor's ability to share information and make it easier for hospitals to differentiate one William Smith from another. However, a universal health identity (ID) system would empower the government and corporations to exploit the single biggest flaw in health care technology today: patients would not have access to who sees, uses, or sells their data.

Overview of EHRs

Until the latter half of the twentieth century, physician and practitioner notes, medical tests, and medical images were written by hand, maintained in a tangible medium, and stored in a physical location. Although there were many technical problems and other issues such as lack of system interface and nonstandard vocabularies, health information stored in an electronic format has been around since the 1960s, predominantly at academic and research medical centers (Brown, 2012). Beginning in the 1960s, patient information has slowly progressed toward being wholly recorded and stored in EHRs. Both President George W. Bush and President Barack Obama pushed for digitizing U.S. health records by 2014. To encourage medical institutions across the country to convert to an EHR system, the Congress earmarked $19.2 billion in incentives in the American Reinvestment and Recovery Act (ARRA).

According to the Health Information Management Systems Society (HIMSS) 2011, EHRs are longitudinal electronic records that contain patient health information generated from multiple clinical care delivery settings. Most EHRs contain two types of data—data scanned into a graphical file and electronically created records that fall into four broad categories: prescription orders, orders for tests, test results, and physician notes (Carter, 2011). The information in EHRs typically includes patient demographic information, clinician treatment notes, presented problems, lists of medications, medical history notes, laboratory and radiology reports, and other health care-related information. The EHR is intended to streamline patient-care workflow and provide access to a complete record of health care by way of interconnection with one or more EHR systems, decision support, quality management, and outcome reporting.

Those who support the move toward a digital medical information access and storage system argue that such a move would result in significant savings to a health system with upwardly spiraling costs. The increased accessibility for clinicians to important health information about the patients they treat will deliver other savings and benefits, including reduction in needless duplication of medical tests and procedures. Furthermore, access for clinicians to a more complete history of the patient's care over time will help decrease medical errors (Kizer, 2007; Miller, 2008). In a nutshell, EHRs are designed to improve the overall efficiency and quality of patient care at a less cost and with fewer errors resulting in injury or death.

Doctors and other health care providers are not the only individuals, groups, or even institutions that maintain EHRs. Insurance companies and third-party payers such as Medicare and Medicaid maintain electronic patient health information. While EHRs must comply with regulations mandated in the Health Insurance Portability and Accountability Act (HIPAA) limiting the access to health care clinicians, other institutions in possession of personal health records are not required to comply with HIPAA, even though these institutions often control the information contained in the personal health record (Roman, 2009; Stead, 2007).

The conversion to the total use of EHRs is not without its critics. The primary objection against the use of EHRs is a concern that the security measures in place will not sufficiently protect the privacy of patient information or will not deal with breaches of security, patient medical identity, theft, and unlawful access. Another concern is the use of medical information by employers, insurance companies, and any other individual or entity wishing to use the information for their benefit and to the detriment of a patient. In addition, concerns exist about the indirect use of medical information without a patient's consent such as for research, provider certification, public reporting, marketing, and other commercial activities (Kizer, 2007). Clinicians have expressed the concern over complete digitization of medical records due to the risk of not having the access to information when needed because of problems with the technology and working with technology experts who do not fully understand the need to protect patient privacy (Kizer, 2007; Miller, 2008).

National Patient Identifier and UPI

Within the health care delivery system (e.g., including physicians, insurance companies, managed care organizations, hospitals, or pharmacies), patients often have several different identifiers (Katish et al., 2011). During their lives, patients move to different locations and access a variety of health care providers who maintain records of the health care provided. The use of a unique identifier is intended to simplify the access to EHRs and thus improve the quality of care, reduce administrative costs, and decrease injury and death caused by medical errors. The purpose of a UPI that each patient would be assigned is to eliminate the use of different identifiers for the same patient across different health care providers (Katish et al., 2011). The UPI would allow for better continuity of care, accurate record keeping, improved follow-up and preventive care, correct and prompt billing and payment for services, decreased waste, and detection of fraud and misuse of patient information.

Unlike a social security number that has been used as a unique individual identifier for a broad range of purposes such as school records, employment, Internal Revenue Service identification, or accessing financial accounts, a UPI would be a unique number for accessing only medical records (HHS, 1998). HIPAA legislation supports the creation of a unique identification system, but such a system has not been implemented because of concerns of patient privacy and security of medical information. In the last 10 years, the adoption of EHRs has expanded considerably (Science Daily, 2008). According to Hillestad et al. (2008), who conducted a study for the RAND Corporation on the use of EHRs and UPI, the cost to create a national identification system would be about $11 billion.

The Healthcare Information Management and Systems Society put a considerable effort into promoting the use of a UPI. In 2011, HIMSS argued that the use of a national identifier would increase the ability to both access and secure health information. In addition, HIMSS made three broad recommendations to the Congress in 2011 in its support of a nationwide patient identifier (HIMSS, 2011). First, HIMSS recommends in its report that the Congress continue supporting the acceptance and implementation of health information technology, arguing that a national patient identifier used in association with EHRs will promote better care, reduce errors, and increase billing and payment efficiency. Second, HIMSS recommends that the Congress continue supporting investment in the Medicare and Medicaid EHR meaningful use inventive programs that will continue the efforts made by both Presidents Bush and Obama to move to a national EHR system. The third recommendation by HIMSS is that the Congress remove any barriers to creating a national health identifier included in the 1999 Omnibus Appropriates Act that prohibits the federal government from using funds toward the creation of a system that uses a national health identifier (HIMSS, 2011). Finally, HIMSS points out in its report that the implementation of a national health identifier is not synonymous with the issuance of a national identity card but is a means of linking a unique EHR with the patient.

Privacy and Security Concern over Using a National ID or UPI

While the use of digitized health records offers the benefit of providing clinicians with easy access to and sharing of medical information with other clinicians, ease of storage, and speed of transmission, EHRs make this information vulnerable to security breaches. Other vulnerabilities are exposed when businesses use private medical information for their own purposes and profit without notification or permission obtained from patients. Although the implementation of EHR has many benefits, one area of considerable concern is the protection of patient privacy and dealing with security breaches and patient medical record theft. The Clinton Administration instituted

federal regulations to safeguard EHR privacy. The Bush and the Obama administrations continue the effort to implement a total shift to EHRs.

Health information stored in EHRs is a valuable commodity to criminals who sell such information in the black market or use it to commit Medicare fraud. Between 2010 and 2011, there was a 97% increase in health data breaches in the United States (Manos, 2012). Since 2009, 19 million people have been affected by health information breaches that occurred across all 50 states (Manos, 2012). Part of the problem is the lack of security that prevents health information from being stored on unencrypted laptops and other portable storage devices (Manos, 2012). In addition, there is little oversight in protecting patient privacy among health care organizations that disclose patient information to their third-party business associates (Manos, 2012).

Arguments Supporting the Use of UPIs

Those who support the implementation of a UPI system argue that such a system would increase the level of security. According to Hillestad et al. (2008), the use of system check codes tied to the UPI would guard against the number of input errors as well as lessen the chance of accessing and entering data into the wrong patient EHR. Further, the use of a UPI would improve the ability to store and retrieve records across different systems and would enhance the quality of care by reducing repetitive and unnecessary procedures.

The proponents of a UPI system point out that without a unique identifier, most health systems and health care providers use a technique called "statistical matching," which uses attributes such as name, birth date, sex, and social security number to retrieve a patient's medical record. However, statistical matching can result in the retrieval of incomplete EHRs about 8% of the time. In addition, statistical matching increases the risk of privacy breaches because a great amount of personal information is used during the records search process. Hillestad et al. (2008) argue that the use of UPI makes the implementation and maintenance of security protocols easier than using the statistical matching system used today.

The President's Council of Advisors on Science and Technology (PCAST) report to the President realizing the full potential of health information technology to improve health care for the American: the path forward of December 2010 does not support a requirement for UPIs or the creation of federal databases of patients' health information (Holdren and Lander, 2010). In addition, the report does not explain a proposed approach to eliminate the need for a UPI. Another major hurdle with EHRs is the incorrect record linkage, for both patient care and research needs. How will the systems know the appropriate linkage between medical records for the same individual, particularly across a myriad of health care organizations, providers, and EHRs that may contain information on the same person? Relying on identity resolution, technologies and probabilistic person-matching algorithms are imperfect, and do not resolve to identify individuals with sufficient certainty to be neither used in health care nor biomedical research (Niland, 2011).

A health information exchange (HIE) can be thought of as a database of databases where the databases are the data sources (e.g., EHRs and enterprise master patient indexes [EMPIs]) exposed by the participating Health Information Organizations (HIOs). The database is the collection of those exposed data sources. On the macro level, HIEs support three sequentially executed database queries: patient-match, available-patient-documents, and patient-documents-retrieve. An HIO first executes the patient-match query to determine if other HIO databases (e.g., EMPIs) contain a matching patient. The patient-match query only returns those matching patients who have agreed to participate in the HIE. An HIO then executes an available-patient-documents query to determine

which documents (e.g., a continuity of care document [CCD]) the matching patient(s) has consented to share with the querying HIO. A patient may have some medical information (e.g., a mental health issue or sexually transmitted disease) they do not want to share with some or all the HIOs participating in the exchange. Finally, an HIO executes a patient-documents-retrieve query to retrieve all or some of the documents the matching patient(s) has consented to share with the querying HIO. E.F. Codd, the inventor of the relational database model that is used to manage the majority of today's electronic data, identified 12 rules for defining a relational database. Rule number two, guaranteed access, stated that all data must be accessible without ambiguity (Codd, 1985). This rule is essentially a restatement of the fundamental requirement for primary keys. Today's statistical matching algorithms violate this rule and can return false-negative and/or false-positive patients in response to a patient-match query. This violation creates data-management issues with negative patient safety, privacy, security, and public trust consequences. HIEs could address these issues by adopting a UPI to support unambiguously identifying patients across the HIE database allowing patient-match queries to return either a unique matching patient or a "no patient found." The assignment of the UPI could be voluntary to address most of the objections that have prevented the adoption of a patient identifier but should be mandatory for patients wanting to participate in an HIE.

Arguments against Issuing UPIs

Senator Tom Coburn of Oklahoma, who is a practicing physician, raised concerns that security measures at this time are not sufficient to protect the information stored in EHRs from cyber intruders or others who would unlawfully obtain and use the personal health information (Sterstein, 2011). President Obama in the ARRA allocated more than $20 billion dollars for use in converting existing paper-based medical records into a digital format and to implement EHRs. Senator Coburn points out that Chinese cyber attackers have breached security measures and obtained sensitive information—technology that could be used to hack into and obtain private medical information worldwide. Generally, the legislation does not include provisions for reporting data breaches involving health information. According to Sterstein (2011), "All but one of the several pending Senate bills that would mandate data-breach notifications excludes health information." Senate Bill 1535 (S-1535) that Richard Blumenthal of Connecticut sponsored does provide the support for disclosing breaches involving the health performance (Sterstein, 2011). In 2011, 73 individuals were indicted for health care fraud and medical identity theft for stealing the identities of doctors and Medicare beneficiaries and using this information to submit $163 million in false billing. In June 2011, an individual who was a defendant in an administrative hearing was able to access sensitive medical records of Arizona state government employees involved in litigation with the defendant (Sterstein, 2011). These are just a few examples that highlight the weaknesses in the current security systems to protect patient identity and health care information.

UPI Implementation

The UPI system could be included within the identity ecosystem (IE) described by the National Strategy for Trusted Identities in Cyberspace (NSTIC) (Baker, 2011). The NSTIC IE is made up of the following:

1. Trust-marked organizations that have met the requirements of the IE as determined by an accreditation authority

2. Individuals engaged in an electronic transaction
3. Subjects of the transaction
4. Identity providers responsible for establishing, maintaining, and securing the subject's identity

In an HIE, the trust-marked organizations are the HIOs that have complied with the overarching set of interoperability standards, risk models, privacy and liability policies, requirements, and accountability mechanisms of the HIE (Baker, 2011). The individuals are the participating HIO's HIE users (e.g., physicians) and the subjects are the patients who have consented to participate in the HIE. The voluntary UPI (VUPI) is the digital identity assigned by the HIE's identity provider. An identity provider would be responsible for establishing, maintaining, and securing the VUPI for those patients wanting to participate in the HIE. The VUPI could enhance patient control over the privacy of their information, improve the quality of medical care and the efficiency of its delivery, reduce medical errors related to misidentification of patients, decrease incidents of health care-related identity theft, and help control health care costs as a result of these impacts (Hieb, 2012).

The American Society for Testing and Materials (ASTM) standards E1714 (a standard guide for properties of a universal health care identifier) and E2553 (a standard guide for the implementation of a voluntary universal health care identification system) could be leveraged to implement a VUPI system. According to the ASTM standards, the VUPI system would not include a central database of patient demographic information. Instead, it would integrate with the HIOs' EMPI that would in turn assign a VUPI (Dimitropoulos, 2009). A production VUPI system based on these ASTM standards became operational in May 2011 as part of a pilot with the Western Health Information Network (Hieb, 2012).

The following example describes the assignment of a VUPI in the context of a for-treatment HIE with four participating HIOs. During a visit to HIO1, the patient (William Smith) agrees to participate in the HIE and requests a VUPI from the HIE's identity provider on behalf of HIO1. The identity provider validates the request and issues a VUPI that is captured in HIO1's EMPI and associated with the patient's HIO1 EHR. In this example, HIO1 associates used the internal patient ID (i.e., 100) with the issued VUPI (i.e., 8562837139.7491742218) within their EMPI. The patient also receives an identification card from the identity provider who documents the issued VUPI. HIO1 also obtains the patient's consent to exchange all, some, and none of his/her EHR with HIO2, HIO3, and HIO4, respectively. The patient subsequently provides his VUPI to HIO2, HIO3, and HIO4, who in turn capture it within their respective EMPIs and associate it with their respective EHR for the patient. HIO2, HIO3, and HIO4 are now able to unambiguously locate the patient's HIO1 EHR, determine which document(s) the patient has consented to share, and retrieve all or some of those document(s) using VUPI-based patient-match, available-patient-documents and patient-document-retrieve queries, respectively. Figure 20.1 is provided to help illustrate this sample of the HIE and VUPI transactions.

Even though each EMPI captures different demographic traits for William Smith, the VUPI allows each HIO to uniquely identify the patient's EHR in one another's EHR. Statistical matching cannot deliver these results. For example, if HIO3's statistical matching algorithm heavily weighted positive matching on a patient's social security number, then its patient-match query of HIO1 would return a false-negative result since HIO1 does not capture the patient's social security number. If HIO2's statistical matching algorithm heavily weighted positive matching on a patient's name and date of birth, then its patient-match query of HIO1 could return a false-positive result since HIO1 could have a different patient named William Smith born on July 11, 1965 (there are over 24,000 William Smiths in the United States [https://names.whitepages.com/

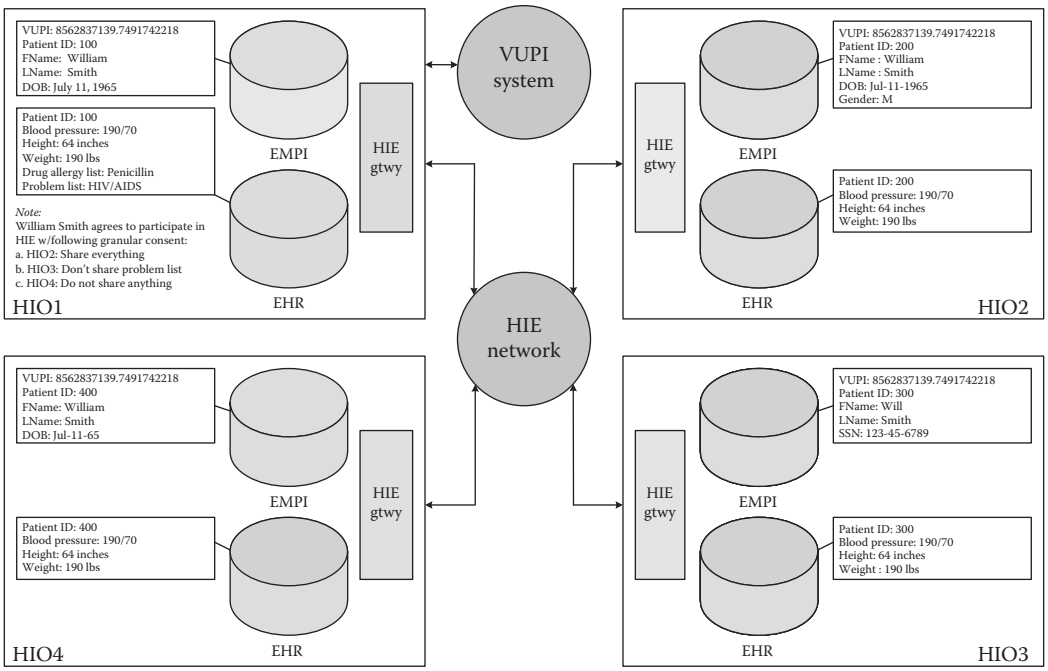


Figure 20.1 VUPI for HIE.

william/smith, 2012]). The risk of similar false-positive results has a higher probability in countries such as Vietnam where about 40% of the population shares the same last name (http://english.vietnamnet.vn/en/society/19435/vietnam-s-nguyen-family-name-most-common-in-the-world.html).

Conclusion

Few health care or insurance companies with access to personal health information give patients a complete control over who can view and use the information in their medical records. The critics of UPIs cite concerns such as the unauthorized use "for profit" and security risks that could result in medical identity theft, falsification of medical records, and the fraudulent use of information to obtain multiple UPIs. The issue is not whether national patient identifiers or UPIs are beneficial. The issue is that not enough progress has been made in developing security systems powerful enough to prevent security breaches, offer oversight for how patient health information is used, and inform policies that provide patients with the right to control who has access to their personal health information and how this information is used. Once these problems are resolved, the use of EHRs and UPIs will be highly beneficial in improving health care.

References

Baker, S. 2011. National strategy for trusted identities in cyberspace, enhancing online choice, efficiency, security, and privacy. Retrieved from http://www.whitehouse.gov/sites/default/files/rss_viewer/NSTICstrategy_041511.pdf.

Brown, C.L. 2012. Health-care data protection and biometric authentication policies: Comparative culture and technology acceptance in China and in the United States. *Review of Policy Research*, 29(1), 141–159.

Retrieved from http://econpapers.repec.org/article/blarevpol/v_3a29_3ay_3a2012_3ai_3a1_3ap_3a141-159.htm

Carter, B. 2011. Electronic medical records: A prescription for increased medical malpractice liability? *Vanderbilt Journal of Entertainment and Technology Law*, 13(2), 385–406. Retrieved from http://www.jetlaw.org/wp-content/uploads/2011/05/Carter-FINAL.pdf

Codd, E. 1985. "Is your DBMS really relational?" and "does your DBMS run by the rules?" *Computer World*, October 14 and October 21.

Dimitropoulos, L.L., PhD, 2009. *Privacy and Security Solutions for Interoperable Health Information Exchange—Perspectives on Patient Matching: Approaches, Findings and Challenges*. Retrieved from healthit.hhs.gov/...0.../PatientMatchingWhite_Paper_Final.pdf

Healthcare Information Management and Systems Society (HIMSS). 2011. *Nationwide Patient Identification Solution Issue* (Online). Retrieved from http://www.himss.org/policy/d/HGRR/201110_Nationwide Patient_ID_SolutionPresentation.pdf

Hieb, B., MD, Chief Technology Officer, Global Patient Identifiers Inc. 2012. *Voluntary Universal Healthcare Identification System* (Online). Retrieved from http://www.gpii.info/system.php

Hillestad, R., Bigelow, J.H., Chaudhry, B., Dryer, P., Greenberg, M.D., Meili, R.C., Ridgely, M.S., Rothenberg, J. and Taylor, R. 2008. Identify crisis: An examination of the costs and benefits of a unique patient identifier for the U.S. health care system. *Rand Health* (Online). Available at http://www.rand.ord/pibs/monographs/MG753

Holdren, J.P. and Lander, E. 2010. *Report to the President Realizing the Full Potential of Health Information Technology to Improve Healthcare for Americans: The Path Forward*. Executive Office of the President, President's Council of Advisors on Science and Technology.

Katish, E., Sondheimer, N., Dullah, P., and Stromberg, S. 2011. Is there an app for that? Electronic health records (ENRS) and a new environment of conflict prevention and resolution. *Law and Contemporary Problems*, 74(31), 31–56. Retrieved from http://scholarship.law.duke.edu/lcp/vol74/iss3/3

Kizer, K. 2007. The adoption of electronic health records: Benefits and challenges. *Annals of Health Law*, 16(2), 323–334. Available at http://www.ncbi.nlm.nih.gov/pubmed/17982826

Manos, D. 2012. Reported health data breaches rose by 97% in 2011, report find. *Healthcare IT News* (Online) Retrieved from http://www.ihealthbeat.org/articles/2012/2/1/health-data-breaches-increased-by-97-in-2011-report-finds.aspx#ixzz1lKlnpwXP

Miller, S.J. 2008. Electronic medical records: How the potential for misuse outweighs the benefits of transferability. *Journal of Health and Biomedical Law*, 4(2), 353–373.

Niland, J.C., PhD, Associate Director and Chair, Information Sciences, City of Hope, HIT Policy Committee & HIT Standards Committee, PCAST Workgroup, February 15, 2011, Panel 3, Population Health Written Testimony.

Roman, L. 2009. Combined EMR, EHR, and PHR manage data for better health. *Drug Store News* (Online). Retrieved from http://fmdarticles.com/p/articles/mi_m3374/is_9_31/ai_n35619257

Science Daily. 2008. *Creating Unique Health ID Numbers Would Improve Health Care Quality, Efficiency, Study Claims* (Online). Retrieved from http://www.rand.org/news/press/2008/10/20.html

Stead, W.W. 2007. Rethinking electronic health records to better achieve quality and safety goal. *Annual Review of Medicine*, 36 (Online). Available at http://arjournals.annualreviews org/doi/abs/10.1146/annurev.med. 58.061705.144942

Sterstein, A. 2011. Coburn: Computerized records will bring on hackers. *National Journal* (Online). Retrieved from http://mobile.nationaljournal.com/healthcare/coburn-computerized-patient-records-will-bring-on-hackers-20111110

U.S. Department of Health and Human Services. 1998. *Unique Health Identifier for Individuals: A White Paper*. Retrieved from http://epic.org/privacy/medical/hhs-id-798.html

Chapter 21

Addressing Social Media Security and Privacy Challenges

Rebecca Herold

Contents

Addressing information security and privacy within business organizations has provided numerous additional challenges with recently introduced technologies (such as big data analytics, the use of personally owned computing devices at work, and cloud services) and comparatively new online habits (such as social networks, the use of always-on location tracking apps, and using the same user IDs for social networks as for work systems) of individuals. Among the many challenges are those that come along with social media use. There are many benefits that can be realized through the use of certain social media sites within businesses, but it is important when planning to take advantage of those benefits to also know and understand the associated risks, both to privacy and to network and information security.

What Is Social Media?

Social media is media that is designed to be disseminated through social interaction, created using highly accessible and scalable publishing techniques. Social media uses Internet and web-based technologies to transform the traditional showing of information into an interactive sharing of information. It supports transforming people from content consumers into content producers. Social media is increasingly used for more types of interactions, and is playing significant roles in social, political, and other types of causes.[*] It is increasingly being depended upon as a source of information, significantly so within the business industry, even though many, and indeed most, sites are far from trustworthy with regard to the accuracy of the information.

A few examples of the commonly used types of social media sites include

- Blogs such as TypePad, WordPress, and so on
- Collaboration sites, such as wikis (e.g., Wikipedia, Delicious) and social news (e.g., Digg)
- Livecasting and meeting sites such as Skype, Livestream, and so on
- Microblogs such as Twitter
- Photography- and art-sharing sites such as Photobucket, Flickr, Picasa, VineMe, and so on
- Presentation-sharing sites, such as Scribd, Slideshare, and so on
- Product review sites such as Epinions.com, MouthShut.com, and so on
- People review sites such as RateMDs.com, Healthgrades.com, and so on
- Social networks such as Facebook, LinkedIn, Google+, Pinterest, and so on
- Video-sharing sites such as YouTube, Vimeo, and so on
- Virtual worlds such as Second Life, Maple Story, and so on

Benefits

When businesses are determining policies for social media use, not only must the risks be considered, but the business leaders must also acknowledge that there are potential benefits to the organization for using social media sites. When used in a thoughtful and risk-mitigating way, social media sites can help to improve your business and business services. The key is to establish social media policies and supporting procedures that not only mitigate the associated risks, but at the same time also support the appropriate business uses. Business organizations, and their business associates and other contracted workers, are using social media to provide beneficial services in many areas, such as the following:

[*] Cahr, D. What is "social media" anyway? Legally Social, 2011. http://www.legallysocial.com/2011/02/what-is-social-media-anyway/

- Customer service
- Knowledge sharing and collaboration
- Patient health education
- Customer awareness
- Learning
- Marketing
- New contacts
- News/world events
- Patient care
- Research
- Crisis management

The information security and privacy personnel must work with marketing, human resources (HR), and other areas of the organization, which are responsible for the listed activities to determine both the risks and the benefits.

Risks

As with any technology, along with the good, there is always the harmful. While there are many benefits, there are also many risks and dangers with social media use, most of which can negatively impact all types of businesses. The following sections list some of the most common damages that have already occurred, some of them many times.

Using Social Media Apps

There are increasingly more apps* being created for social media sites every day. Apps introduce even more risks largely because they are yet another party, a third party, taking information from social media sites and the site users, often in ways the app users do not realize, and then using that information in ways that the associated individuals may not like or want. For example, some currently popular apps and the associated incidents include:

- Foursquare automatic posts have led to physical altercations, break-ins, and other types of crime as a result of people making their posts public instead of restricting them to just those they really want to see them.
- When Spotify came out, a lot of people were embarrassed to have all their music choices showing on the publicly accessible ticker on Facebook. Many had others (such as their friends or family members) who were getting on their computers and purposefully listening to songs that would be displayed as coming from the actual computer owner just as a prank to embarrass them.
- FarmVille and Texas Hold'em reportedly sent Facebook user information to at least 25 advertising and data firms.†
- Until July of 2012, Instagram had a privacy vulnerability that exposed private photos to anyone without requiring authorization.‡

---

* "Apps" is a shortened common term that has evolved from "application software."

† Vamosi, R., Protect your online privacy (without reading all the fine print). *PCWorld*, March 30, 2011. http://www.pcworld.com/businesscenter/article/221104/protect_your_online_privacy_without_reading_all_the_fine_print.html.

‡ Ragan, S., Instagram patches privacy vulnerability that exposed private photos. *Security Week*, July 12, 2012. http://www.securityweek.com/instagram-patches-privacy-vulnerability-exposed-private-photos.

In August 2012, there were more than 13,000 health, fitness, and medical apps in existence,[*] and more apps were emerging at an increasingly quickening pace. While some of these are targeted at patients, a large number of them were created primarily for doctors and/or nurses, many of which are meant to be used to prescribe treatment in various ways. Apps are being used to treat health problems such as "diabetes, cardiology, rheumatoid arthritis, and physical therapy—and allow doctors to prescribe apps to their patients." Physicians can also communicate with their patients using apps.[†]

To properly address information security and privacy risks related to having business workers using social media apps, it is important to determine the following:

a. What apps are your workers using? Document them.
b. Are they using apps that are collecting business information, or could be, unbeknownst to your employees?
c. Are they using apps that are collecting customer information?
d. Do they even know?

BYOD Issues

It is much harder to address privacy and information security risks when your personnel are using their own personally owned computing devices, commonly called BYOD (short for "bring your own device") risks. There are some important ways that you can address these risks, though. Your policies can cover what personnel can and cannot do with information about your business, your employees, your patients, and your customers. This includes the types of information that should not be shared on social media sites. It is appropriate that you direct your workers not to post information about work that would negatively impact work. And that they must not post information about coworkers. If you make the policies about your customers, patients, personnel, and business information assets, then they are applicable even when your workers are using their own devices, and outside of your facilities. You need to make sure that you clearly state this within your policies. Some of the basic rules you should create for social media activities include:

- Do not post about work.
- Do not post about coworkers.
- Do not post about customers, patients, or other individuals associated with your business.
- Do not sync or share files between personally owned computers and the organization's computers/systems.

Posting Photos and Videos

Most businesses have great concerns about having photos and videos taken in their facilities. Hospitals are one such type of business where photos and videos taken within their facilities brings great privacy and compliance risks. They need to address the types of information about customers and patients, in photos, videos, and even comments that their workers can post to social networking sites. If they do not, they may have to deal with a privacy breach. For example, there have been

[*] Scher, D., Five signs the medical apps industry is maturing. *MedCity News*, Aug. 8, 2012. http://medcitynews.com/2012/08/five-signs-the-medical-apps-industry-is-maturing/

[†] Brustein, J., Coming next: Using an app as prescribed. *New York Times*, Aug. 20, 2012. http://www.nytimes.com/2012/08/20/technology/coming-next-doctors-prescribing-apps-to-patients.html?pagewanted=all

multiple Health Insurance Portability and Accountability Act (HIPAA) violations resulting from the disclosure of patient information online.* However, if a patient wants to take their own photos while in the hospital and then put them online, that is not something that can really be controlled, unless they start including others in their posts. Other types of organizations, in other industries, face similar challenges. These situations need to be addressed. Your social media policies that cover the posting of photos, videos, and other types of images need to include directives that include

- Posting about the workplace, patients, customers, and coworkers
- Posting personal photos, recordings, and videos
- Posting patient, visitor, customer, and consumer images
- Obtaining consent when others are in the images

Common Risks and Scams

Businesses also need to address the many types of risks that social networks present to their organizations, to their customers and patients, and to their own personnel. Here is a list of some of the most common exploits that online fraudsters and criminals use within social networking sites. Not only do these present risks to businesses, their customers and patients, their employees, and their contracted workers' information, along with the company's reputation and legal liabilities, but they can also cause direct harm in a variety of ways to the workers, customers, and patients in their personal lives.

- Clickjacking
- Denial of service (DoS) attacks
- Fake donation sites
- Hackers
- Key loggers
- Malicious links
- Phishing
- Social engineering
- Spam
- Spear phishing
- Spoofing
- Viruses, trojans, worms, and other malware

Eleven Topics to Cover within Social Media Policies

There are at least 11 topics businesses need to cover within social media policies to help protect their business, staff, patients, visitors, and customers from the risks presented by social media use.

Appropriate Use

Be sure to clearly define the appropriate use of social networks from all possible locations and devices. The policies should address the information (patient, business, customer, and personnel)

* For example, see Green & Associates, Social media HIPAA violations on the rise, *FindLaw*, June 15, 2012. http://knowledgebase.findlaw.com/kb/2012/May/629328.html

that can and cannot be used or posted when at the office, when away from the office, and when using personally owned computing devices. The key is to make the policy about the information. Generally, businesses cannot control how people conduct their personal lives away from the office, or how they use their own computers, but they can specify how patient information, business information, customer information, employee information, and other company information assets can and cannot be used or shared, no matter the location or device ownership.

Business policies should describe the appropriate use of social networks (Facebook, LinkedIn, YouTube, and Twitter in particular)

- From the company's networks
- From the company-owned computing devices
- From networks using personally owned computing devices
- From staff-owned computing devices and/or networks
- From public computers/networks

Blogging

Blogging is pervasive. At the end of 2011, there were more than 181 million blogs worldwide, five times more than there were in 2006.* It is likely that a large portion of business workers also blog. Businesses need to make sure that they have policies that explicitly address blogging expectations and requirements. These are too often left unaddressed. Such policies should detail the types of information about the business, organization, hospital, clinic, patients, customers, visitors, workers, and intellectual property that can and cannot be included within blog posts. Such policies need to make clear that references to others, not just by name but also by description, should not be included in blog posts.

Wikis

There should be policies for wikis as well. Specify the wikis that are approved for posting business and other information, and those that should not be used. Some organizations have posted protected health information (PHI) and other types of personal information to wikis, in violation of not only multiple legal requirements but also in violation of their own company's privacy policies. Typically, this was done not realizing that what they did put the sensitive information out, basically for the world to see.

Information Not to Post

Business policies need to define the information that is considered to be "personal information." No matter what you want to label it, use whatever is appropriate to your own organization. But clearly specify the types of information that should not be posted to any website, such as PHI under HIPAA, nonpublic personal information (NPPI) under the Gramm–Leach–Bliley Act, and the more generic personally identifiable information (PII) and sensitive personal information (SPI) that are used extensively throughout the many worldwide data protection laws, as well as in the

* NM Incite. Buzz in the blogosphere: Millions more bloggers and blog readers, March 8, 2012. http://nmincite.com/buzz-in-the-blogosphere-millions-more-bloggers-and-blog-readers

breach notice laws of at least 50 U.S. states and territories.* Also describe the types of confidential business information that should never be posted online.

Marketing

There are seemingly unlimited ways in which businesses, hospitals, and clinics are using social media sites for marketing purposes. Be sure to create policies that clearly outline how the sites can and cannot be used for marketing. This includes specifying the persons, positions, or departments that are authorized to do marketing on the sites, the types of information that should not be posted in marketing activities, and very importantly, the types of information that should not be collected from social networking sites and then subsequently used for marketing. Too many organizations are taking what they find online and putting it into their marketing communications without obtaining any consent from the associated individuals. It is important to understand that just because information is posted for the public to see on a social networking site does not mean that anyone can take that information and use it in any way they want. The marketing requirements and guidelines should include the following:

- Positions and departments authorized to post
- Types of information acceptable to post
- Types of information that should not be posted
- Directives against taking personal information found on online sites to use for marketing or other types of business activities

Security Controls

Businesses need to implement security technologies, physical and administrative controls, and make their workers aware of the security risks involved in using social networking sites. Appropriate security must be implemented to help keep the threats from social media sites from damaging business activities, customers, patients, business activities, and information. These controls should include, at a minimum, antimalware, firewalls, spam prevention, and data leak prevention (DLP) tools that are kept regularly updated to protect against new social network threats.

Time Spent on Social Media Sites

While it is not feasible in most organizations to simply prohibit all social media use while at work, it is reasonable and recommended that businesses establish the parameters within which workers can, and cannot, use social media sites while at work. Here are some general guidelines to specify

- Time spent on social networks during work hours
- Directives to not use social media sites while with customers or patients
- Requirements for using social media sites at work only for short period
- Requirements that social media use should be restricted to only during breaks

---

* See the full list, along with links to the actual text of the laws, at NCLS, State security breach notification laws, http://www.ncsl.org/issues-research/telecom/security-breach-notification-laws.aspx.

Linking with Others

The topic of communications with coworkers, clients, patients, customers, and others related to the business needs to be appropriately and clearly addressed. Businesses need to be sure they do not overstep what is reasonable in restrictions of what personnel do on the sites. By keeping social media policies about the business and associated information, organizations will have better policies than if they try to restrict activities in general. When it comes to linking on LinkedIn, friending on Facebook, and so on, with customers, patients, and coworkers, provide the following guidance:

- Do not ask for passwords from workers to their social media sites.*
- Specify that only authorized personnel can participate from social media accounts established to represent the organization.
- Do not link, friend, and so on from personal accounts that list the employer.

It is also becoming more common for organizations to not allow their staff to friend, link, or otherwise be connected to their customers or patients.

Posting Photos and Videos

While businesses cannot tell workers what they generally can and cannot post on their personal sites with regard to photos and videos, businesses can, and should, have policies that address what can and cannot be posted regarding business, patient, organization, customer, coworker, and other similar types of information.

With regard to patients and customers who want to take photos and videos with coworkers, doctors, nurses, and other staff

- Ask that they only post images that include the staff with the staff's knowledge
- Ask that they do not include others within their images

With regard to staff postings, include the following in social media policies:

- No posting of patient, visitor, or customer images unless approved by the area responsible for privacy compliance or with a written consent of the patient, visitor, or customer
- No posting of images showing facility entries or other staff unless approved by the area responsible for privacy compliance

Reacting to Posts

Establish policies that cover how workers should and should not respond to posts they see online that are related to the organization, coworkers, business, patients, customers, and so on. Typically, workers should not respond themselves; that could create some legal problems and liabilities for their organization if they are seen as representing the views of the organization. Instead, have policies and

* For more information about requiring employee passwords, see Herold, R., 6 Good reasons NOT to ask for Facebook passwords, *Privacy Professor*, March 23, 2012. http://privacyguidance.com/blog/2012/03/23/6-good-reasons-not-to-ask-for-facebook-passwords

procedures requiring personnel and contracted staff to report such posts to the appropriate area in the organization.

Some general directives to consider including in social media policies about reacting to posts about customers and staff include:

- Do not respond directly to negative posts.
- Report the negative posts to the public relations (PR) office.
- Do not argue, defame, or otherwise act negatively in communications with others online.

Donor Searches

There is a growing trend to use social networking sites for health-related activities, such as for organ donations. Health care organizations should establish policies that detail the appropriate ways in which such posts should be made. Include the following:

- Only authorized personnel can post messages for such searches.
- Only authorized personnel can post replies to posts offering organs.

Summary

All types of organizations must address information security and privacy issues related to social media use. Organizations in highly regulated industries will have some additional types of unique issues to address. For example, health care providers will need to protect patient information and maintain HIPAA compliance. The first step to successfully controlling such use, and preventing breaches and other problems, is to

1. Establish comprehensive social media policies
2. Have each department establish supporting procedures that will help them meet compliance with the policies
3. Provide regular training and ongoing awareness communications about this topic

To realize the many benefits of social media sites, businesses must be sure to also know and understand the associated risks, both to privacy and to network and information security.

Investigations

Chapter 22

What Is Digital Forensics and What Should You Know about It?*

Greg Gogolin

Contents

Digital forensics is the application of scientific principles to the process of discovering information from a digital device. A form of digital forensics has been around nearly as early as the time computers were invented, but forensic capabilities have witnessed many advances in the past few years as digital forensic processes have matured and needs have become more prevalent. Digital forensics can involve nearly any digital device, not just computers, although technology often evolves faster than forensic capabilities do. Some of the common areas in which digital forensics is used include computers, printers, cell phones, mobile devices, global positioning systems (GPS), and storage media. The less common areas include automobile systems, appliances, office equipment, and other programmable devices.

Forensic Science

The precise date when forensic science began is unclear as there are many different fields in which forensic science can be applied. Certainly, people have been trying to determine how people died

for thousands of years. In the Chinese book *Hsi Duan Yu* (*The Washing Away of Wrongs*), which appeared around 1248, the author details methods to distinguish the effects of different ways of dying; for example, death by drowning as opposed to death by strangulation. Nearly 700 years later, the first crime laboratory was established in the United States by the Los Angeles Sheriff Department in 1930. Howard Schmidt, who served as an advisor to President George W. Bush and President Barack Obama, is credited with establishing the first U.S. government digital forensics laboratory. Although forensic science has been evolving for many centuries, digital forensics is a relatively new development.

For something to be considered a science, it has to study, describe, and investigate phenomena in its field. A key aspect of this is that new knowledge generated by the study and investigation has to be repeatable. A peer-review process is often followed, including within a laboratory and the publication process. A complex investigation can have many opportunities for error or misinterpretation, and the review process helps reduce the instances of error. In the digital forensics field, tools and techniques are often reviewed, but it is not uncommon for the findings that are presented in court cases to be the work of a single investigator and therefore unverified.

In many situations, digital forensics does not have a scientific rigor behind it, which is present in other forensic areas such as wet laboratories. Part of the reason is that digital forensics is a relatively new science and another reason is that digital technology progresses at such a rapid rate that the digital forensic processes tend to lag the pace of technological innovation. Figure 22.1 is a flowchart representation of the scientific process.

There are three aspects of the scientific process that I want to highlight. The first is to clearly define the question or purpose of the research. The second is to define a hypothesis. A hypothesis is a potential explanation for a phenomenon. A digital forensic investigator often needs to develop

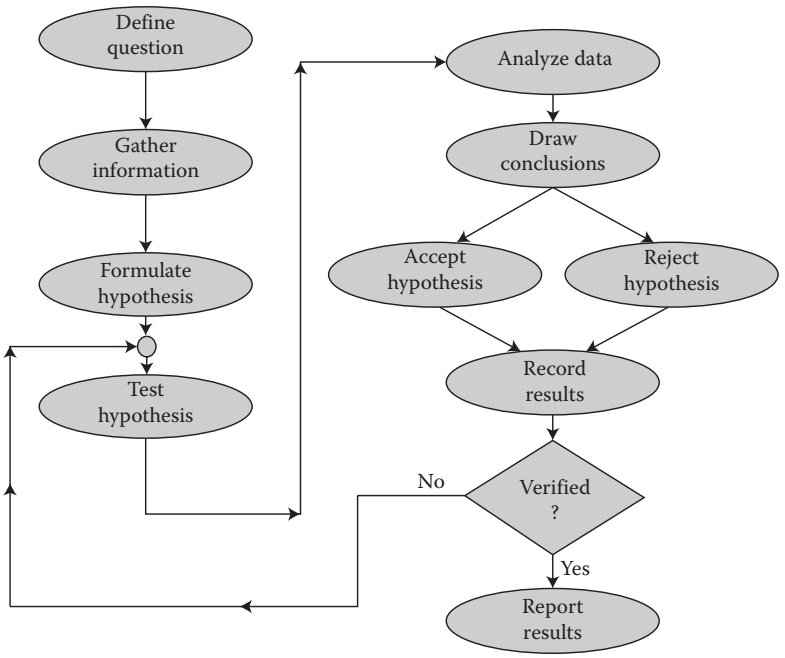


Figure 22.1 Scientific process.

a hypothesis to explain what happened on a computer and what it was used for. The third aspect that I want to emphasize is that many discussions of the scientific process overlook the verification of results.

Too often this is not done, and improper results are reported. Digital forensic cases can be life changing for many individuals, and every effort must be made to ensure that the findings of the investigation are accurate. I do not want to discount the other steps in the scientific process, but I wanted to emphasize these three aspects, in particular, the verification of results.

Digital forensics is not limited to criminal investigation. It can be used to solve problems in a corporate setting such as recovering lost files and reconstructing information from damaged equipment and also to test for changes to devices that are subject to a stimulus. Malware and botnet research are other areas that use digital forensics, particularly when trying to determine impacts. An example would be to use forensic processes to establish the baseline state of a device, introduce the stimulus, and then compare the resulting state with the baseline.

What Does It Take to Be a Digital Forensic Investigator?

Digital forensic investigators need skills and interests in a variety of areas. The first question I ask someone who is considering this field is if they like puzzles. When investigating a case, you may not know any details other than that something has happened. So, if someone needs to be shown or told what has occurred, they may not be a good fit for this field. Sometimes, cases come to an investigator's attention with instructions to find out what the computer or device user was doing. At times, this may be an open request, whereas at other times, it is within a specific time frame.

Many cases follow a similar pattern, and a methodology can help along with a consistent investigation. However, many times, the investigator needs to improvise an approach as there is not always a clear way to do things. This can be the result of new technology in which a methodology has not been developed, due to cost issues, or simply because it is the first time the investigator has encountered that situation. The point is that an investigator needs to be someone who can figure things out and not rely solely on others to do so. Another important characteristic is the ability to handle frustration because investigative tools and software do not always function without their challenges. This can be a fairly common occurrence when dealing with cell phones and small devices. I have had many students who stop their investigation in class at the first sign of difficulty rather than working through the challenges. They do not even try to find an insight into their difficulty from the web or the help system provided with the tool. Someone needs to be persistent and creative to be a successful investigator.

Another critical aspect of being a forensic investigator is the ability to keep your mouth closed. Case specifics usually require some level of confidentiality, and this must be maintained. Similarly, if someone is looking to enter the field as a private investigator or law enforcement professional, lack of a criminal record may be mandatory. Within a corporate setting, investigators may not need to be licensed, but they do need to maintain a high degree of integrity within the context of the corporation.

I have gone through many smartphones and computers covertly to determine the degree of an employee's misconduct. The result is that I know what personnel changes are likely to occur before anyone else.

Irrespective of whether the environment is corporate, law enforcement, or as a private investigator, a background check is likely to occur. Particularly, in law enforcement and private investigator licensing, fingerprint registration is likely to be a requirement. Private investigators also need bonding or liability insurance. Most states require that private investigators have the experience

before becoming licensed; so, students who are fresh out of college may find that they need to work for someone else under their license before becoming individually licensed.

The work itself seems to follow a sine wave rather than a consistent flow. Cases often explode into multiple devices and locations, which can mean long and inconsistent hours. After-hours investigation may be the rule for some cases, and often, this may be at a distant site. What does seem to be the rule is that cases appear when they are not expected, and it is a good practice to be ready. For example, computer forensics investigations usually include taking forensic images of the computers under investigation. Typically, this means taking a forensic image of the storage devices. The location where the images are being copied to should be forensically prepared in advance. For example, if a computer has a 1-TB hard drive, the forensic image could be taken on another 1-TB hard drive. This forensic image hard drive should not just be a new hard drive that is in an unopened box from a retail store because it is unknown what may be already stored on that drive. New hard drives commonly come with utilities and other programs preinstalled. A hard drive should be completely erased and reformatted before it is used. Experienced investigators often wipe a hard drive and then overwrite the entire drive with a hex character. This process takes time, and when time is of the essence, preparing forensic storage media in advance can save considerable time.

Let us complete the thought on forensically prepared storage media. The purpose of forensically prepared storage media is that it allows the investigator to testify that the only information contained on the forensic image drive is the one from the suspect computer and that there is no evidence of contamination. Anything that is not part of the forensic image would be the hex character that was due to the wiping process. Hash algorithms, such as MD5 and SHA-1, are also used to verify that an exact copy has been taken.

What Are the Trends and Challenges in Digital Forensics?

One trend that is occurring is standardization and licensing of forensic investigators. In several states, it is a felony to perform digital forensic services without being licensed as a private investigator. The requirements for licensure vary; so, the appropriate state agency should be consulted. It is important that individuals must understand how the state within which they reside defines digital forensics so that they do not find themselves in a legal situation.

A recent study that I conducted of Michigan law enforcement indicated a rapid increase in the number of cases requiring digital forensic services and a significant shortage of forensic investigators. It appears that if most criminal cases do not already have a digital aspect, they will have in the near future. Surveillance devices and the prevalence of things such as cell phones are one of the main reasons for the rise in digital aspects. Social media, where planning and bragging of exploits occur, almost guarantees that there will be digital artifacts.

Cell phone, mobile device ownership, and usage surpassed five billion in 2010. This provides for the capture of information with embedded devices such as cameras, as well as for the exchange of information verbally and through technologies such as text messaging and e-mail. The integration of smartphones into the World Wide Web and social media provides for a rich and extensive number of artifacts that may be of interest to a digital forensic investigator.

Most cell phones and camera-equipped mobile devices have GPS capabilities. This allows the camera to incorporate the GPS coordinates of the location where a picture was taken into the picture file header. The analysis of pictures and videos, whether from cell phones or surveillance systems, is an explosive area of growth. The images often need to be enhanced to obtain the necessary level of detail, and this expertise with graphics is a skill that requires a combination of training and technology that is not commonly available.

The movement toward a cloud environment is changing the digital forensics world. Cloud computing uses computing and storage resources from a pool. The pooled resources may be contracted from a third-party cloud provider. The third party often shares the pooled resources among many organizations. A key advantage of cloud architecture is that if a contracting organization needs extra resources for a short period, they can contract the resources from the cloud provider. When they no longer need the resources, they simply revert to a smaller amount of resources. This saves the organization money by eliminating the need to purchase the equipment that they only need for a short period.

Cloud computing is like a food-catering service. If you want to throw a big party, it makes sense to rent seating, rather than to buy several tables and chairs that you will not need after the party. When the party is over, the seating is cleaned and goes back to the owner, who is then free to rent it out to someone else who needs it. When the next renter receives the seating, it is in a clean state and there is no evidence that you had the seating. This illustrates the challenge with cloud computing for a forensic investigator. When you no longer need the resources that are contracted, the resource is contracted for another purpose. This can create multiple situations of interest in terms of forensics.

Scenario 1: If the data on the cloud storage device are deleted but not wiped when it is repurposed, there is a potential for reconstruction of and access to the data that were previously on the device by someone who should not have the access.

Scenario 2: If the data on the cloud storage are deleted and wiped when it is repurposed, the recovery of the data or artifacts is not possible—especially if the storage resource is already in use by another entity.

The cloud environment also means that artifacts may no longer be contained on a single device, which can complicate discovering where the artifact originated. The very nature of the cloud, where resources can be used for a short time and then can be repurposed, means that artifacts can exist one moment and then disappear the next moment. Someone perpetrating a crime could contract the cloud resources similar to what was previously outlined in scenario 2. Once the perpetrator is finished with the resources, the evidence is wiped and the resources are used by someone else, likely completely destroying the trail of evidence. This type of situation has actually been occurring for many years in the form of botnets and similar technologies. The advances in cloud computing allow taking nefarious activity to a whole new level, which makes advancements in forensics that is more critical.

Antiforensics is attempting to hide, destroy, or alter artifacts to prevent their reconstruction by forensic analysis. Using antiforensic techniques can make forensic reconstruction difficult or even impossible. Many tools are created to provide security for individuals who use intense algorithms and techniques. There is little to prevent the use of these tools in activities of less-than-honorable intent, which may frustrate forensic efforts. Even when methods have been developed to address these tools, the time and computer-processing power required to defeat the tools and techniques are often prohibitive.

Another aspect that is not so much a trend but rather evidence that the world is becoming a smaller place is the internationalization that is enabled through the ease of connectivity. The language and cultural implications, as well as the ability of criminal perpetrators to be in different countries at the time of their transgressions, give rise to a multitude of challenges for investigators. Very few technical degree programs require foreign language fluency, which can hamstring an investigation from the start. Foreign languages and dialects present a number of challenges in understanding the communication such as e-mail, as well as the extra difficulty that slang and

idiomatic expressions introduce. In part, owing to the proliferation of digital devices and Internet use and network connectivity, these challenges will likely only increase.

The worldwide nature of connectivity presents other challenges. There are jurisdictional and cooperation issues to deal with. Even within the United States, a case can be interpreted as a felony in one county but can be interpreted as a misdemeanor in another. State lines have not even been crossed! Crimes that cross state lines or international borders can be much more difficult to address. Perpetrators know which countries are likely to cooperate in a situation and they plan accordingly. Servers are often hosted in countries that have lax laws or enforcement against a particular activity. Illegal online gambling servers are commonly hosted in Latin American countries.

To complicate things even further, the location of servers can quickly change with the use of virtual technology. Virtualization allows one physical computer to run multiple logical computers within it. A rough example is when someone sets up multiple accounts on their home personal computer. Little Johnny signs on to his account, with his own screen background and configuration characteristics. Stephanie signs on to the same computer using her account and uses her own configuration, which is independent of Little Johnny's. Virtualization takes this further in that the operating system can be installed multiple times in different locations on one computer, and the multiple installations can be run simultaneously and independently. Virtualization has existed in a corporate computer environment for decades, but recent advancements have reduced the cost and complexity so that it can be used by a very broad range of individuals. These virtualization capabilities help to support the movement to a cloud architecture. An individual can create a virtual machine and can potentially move or copy it between computers—and countries—within minutes. The implications of this in a digital forensics environment are dramatic.

Computer forensics often focuses on storage media such as a computer hard drive. Hard drives are rapidly increasing in capacity and other storage technologies are evolving. Flash memory and solid-state drives (SSD) are likely to replace hard drives in the near future because of their superior access speeds. Bell and Boddington (2010) found that SSD can confound the current digital forensic techniques. Similarly, Wei et al. (2011) found that the traditional hard drive sanitation techniques are ineffective on SSD. Computer technology will continue to evolve and these examples illustrate that a forensic examiner must evolve as the paradigms within which they operate evolve.

The last trend that I want to discuss—although there are many more trends that could be touched on—is simply the explosion of data and storage capabilities. An individual can purchase a terabyte of hard drive storage for <$50. This puts the ability to store multiple terabytes of data within the reach of an incredible number of people. It also means that corporations can create dramatically large databases and data warehouses that were previously inconceivable. It can take days to perform a keyword search on a 1-TB hard drive using forensic tools. Extrapolate that out to a corporate environment where the storage capacity is far greater, and it is possible to find that it is impossible to keyword searches on all the storage media using current techniques.

One challenge is that it may not be able to "freeze" all the data in all the systems at the current state. The data are always changing and being modified by multiple systems and users. It is kind of like telling the world to stop rotating while measurements are taken. Further, the data may reside in many locations. The point is that digital forensics is a far different environment in a corporate setting than it is when looking at an individual's private computer. But with the advancements in computer technology, some of the challenges with corporate digital forensics are appearing in the environment of an individual's private computer.

Resources Available to Digital Forensic Investigators

Before I describe some of the tools and technologies, I would like to describe some of the organizations and other resources that are available. Formal training is one of the most important considerations. In a legal situation such as a court case, the competency of an investigator is going to be one of the first things evaluated. If the investigator has little or no current relevant training or certification, the outcome of the case may be in doubt. Most vendors provide training on their tools, and many conferences and information security organizations provide training opportunities.

In many instances, online resources are the most valuable. Knowledge bases and virtual communities can be invaluable in working through problems that an investigator encounters. Vendors often provide a searchable bulletin board that is frequented by users of the vendor's products. Open-source tools often have very passionate support groups. Social networks can be valuable in obtaining contacts that may help provide insight into issues. Organizations such as the International Information Systems Security Certification Consortium (ISC)^2 provide training, testing, and certification in the areas that compliment digital forensics. The federal government has a variety of programs such as InfraGard and opportunities through the Department of Homeland Security that can be very beneficial. Search.org is one of the many organizations that I have found that provide free online resources that I use frequently.

Many universities have research groups that specialize in digital forensics that may be willing to help. Which university or research group is appropriate to consult depends on the type of the case or situation the investigator is working on. For example, if the case involves an incident response, Carnegie Mellon's computer emergency response team (CERT) could prove to be a valuable resource. But if the incident is something about the recovery of the data from the damaged media, contacting a different university may be appropriate. It is also important to realize that the top research universities are not the only game in town. Researchers at smaller universities may actually have broader exposure and more field experience than those at universities where research is a primary goal.

A good way to wade through the maze is to find a third-party reference such as the National Security Agency (NSA). The NSA designates universities as centers of excellence if they meet certain criteria, and most universities that have a significant presence in digital forensics have gone through this certification process. Additionally, the NSA certification is in different areas; so, this can further assist in finding potential resources.

Many tools and technologies are available to assist forensic examiners in addressing the challenges described earlier. There are commercially available tools as well as open-source tools. Some tools are very specialized and focus on doing one particular thing well, whereas other tools attempt to address a much broader perspective. There is often a leapfrog situation occurring where a new version of a particular tool surpasses its competitors. A short time later, a competitor may introduce a new version and it becomes the leader until the next leapfrog situation. For that reason, I will not try to rank particular tools and technologies or provide detailed instructions on how to perform a specific operation with a tool. However, I am a firm believer in leaving bread crumbs for myself and anyone else to follow. So, I will not hesitate to describe my experiences with tools. It is my hope that I can point out the potholes that I have encountered so that you can avoid them or at least prepare for them.

Conclusion

Digital forensics is a rapidly advancing field that has many challenges and crosswinds. The opportunities are endless, but they are not for the faint of heart. Frustration is a common partner; so,

the ability and mentality to press on through is a key characteristic an investigator should have. Someone who needs to be shown how to do everything may want to rethink his career options. A can-do attitude is essential, but the investigator does not need to go it alone. A variety of resources are available to assist, and most of the investigators who have worked through the learning curve to achieve competence are more than eager to help others do the same. Usually, they had others to lean on; so, once you reach a level of expertise with the assistance of others, do not forget to return the favor.

Chapter 23

eDiscovery*

David G. Hill

Contents

Recall that governance, as one of the three pillars in the governance, risk management, and compliance (GRC) model, concerns itself with the processes and systems that ensure the proper accountability for the conduct of an enterprise's business. Data governance, as part of information technology (IT) governance within the overall governance framework, is required to ensure the preservation, availability, confidentiality, and usability of an enterprise's data. These are all mandates that relate to data protection. And a major responsibility of data governance is to help with civil litigation.

Understand that the importance of eDiscovery is going to increase as more and more businesses realize the implications of those changes to the Federal Rules of Civil Procedure (FRCP) as a focal point, and as litigation demands continue to mount. However, recognizing the importance of eDiscovery is only the first step. Doing eDiscovery right is the next step.

Civil litigation is already a heavy burden on many businesses and that burden is only likely to grow. Doing eDiscovery right is a tough balancing act between the financial risks of lost lawsuits as well as court-imposed financial penalties in the case of a failure to do things right and trying to keep the expenses of doing eDiscovery within as reasonable bounds as possible. Doing eDiscovery right is not easy because eDiscovery is much more than the simple search of electronically stored information (ESI). Rather, eDiscovery is a set of processes that can be integrated with evolving technologies to serve the purpose of managing ESI for civil litigation.

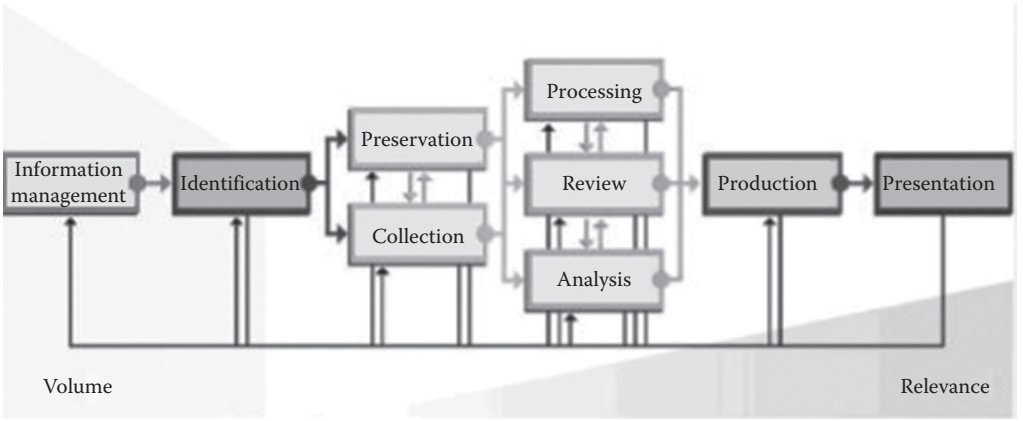


Figure 23.1 Electronic discovery reference model.

Fortunately, an existing framework covers the steps in eDiscovery. This very useful framework is the electronic discovery reference model (EDRM) (Figure 23.1), which is part of the ongoing valuable work of the EDRM group (www.edrm.net). The EDRM can serve as a useful base for reference in exploring the concepts of eDiscovery. The EDRM group has created specific focus groups that constantly address the relevance of the model and how organizations can pass the information in a standard way between the various steps—for instance, when they choose to outsource a part of the process. The individual organization may interpret the sequence of events or the process differently, but the EDRM serves as a useful basis for comparison.

Information Management: Getting eDiscovery Off on the Right Foot

Recall that information management manages the content and decision-making relationships of information as it moves through the life cycle of a specific business process or cross-functional workflow. In this case, the business process/workflow is eDiscovery. Given the nature of this type of workflow, which touches multiple different areas of business information, the general approach lends itself to a more proactive rather than a reactive approach.

Treating an eDiscovery event as a one-off ad hoc incident is likely to be very costly and invariably does not achieve the desired results. Trying to complete all the set of complex tasks that doing eDiscovery right entails is hard to do in a fire-drill hurry-up mode. That is especially true when there is a steep learning curve in trying to put together the complex process. The result is that an eDiscovery process may complete within the time constraints allowed, but it is not done well and creates an obvious litigation risk.

One reason is that all the time is spent in gathering the information; so, no time is left to evaluate it properly. Or the process may result in unacceptable delays. Or inadequate preparation may lead to adverse inferences or spoliation claims. Finally, the lack of processing around collected information often leads to lengthy and unnecessary content review that causes further delay. Whether it is done on time or done late, the net result may be unnecessarily high legal costs, higher-than-estimated internal IT costs (such as for more storage than necessary), and legal penalties for not

complying on time or not complying correctly with the rules. On top of that, since an enterprise is not likely to put its best case forward, it may lose a case that it might have otherwise had won. And that costs money, since civil litigation is about money—who wins and who loses.

Consequently, incumbent on enterprises is the responsibility to plan in advance what needs to be done for eDiscovery. That requires establishing the necessary policies, processes, procedures, and practices ahead of time.

eDiscovery Information Management Process

Assume that the hide-the-head-in-the-sand, *tabula rasa*, approach of welcoming each new civil litigation event without any real learning from past events is not a real process model for eDiscovery. Then, consider two possible models for managing the repeatable information management process for eDiscovery. Call one the informal model and the other the formal model. A business may actually be somewhere on a continuum between the two extremes as far as process methodology, maturity, and complexity are concerned, but for the sake of discussion, the two models can be treated as discrete extremes.

In the informal model, experience, knowledge, and working relationships among the key constituencies (such as members of the general counsel's office and the IT organization) all play an important role. Organizational learning takes place (at least to some extent), so that what is learned from one civil litigation event can be applied to the next event (insofar as is possible).

The basic processes for doing the key tasks are put in place, such as conducting litigation holds. The key data sources, such as e-mail systems, may be recognized as a common source for litigation holds. The key constituencies develop guidelines and a general understanding of what needs to be done and then apply them on a case-by-case basis.

In the formal model, a rigorous process management program, including both a records management program and a data mapping program, need to be put in place. In addition, these tie into the data retention management process. An organization has to defend its position on the exclusion of records that have expired as part of the everyday policies of the business. The good news is that a records management program may already exist, so the foundations are in place. In general, however, a lot of time, effort, and money have to be expended to make eDiscovery a rigorous repeatable process.

It is no wonder, then, that the informal model predominates. The eDiscovery jobs get done and the apparent out-of-pocket costs are less than with the formal model. But how efficient is the informal model? The cost of eDiscovery is already high in many cases and is likely to grow higher. Moreover, can businesses win more cases and reduce the financial awards in the cases they lose? That also represents costs.

The formal model is also not a panacea. How much the costs of eDiscovery can be cut is likely to be very uncertain, and eDiscovery is still likely to be expensive (although hopefully not as much so). Whether the win/loss ratio in civil litigation can be improved may simply be speculation and conjecture.

Still the road less taken should be examined to see what it brings to the table. Two examples will serve to illustrate some of what needs to be done: records management and data mapping.

Records Management: Back to the Future

Traditionally, records management is dealt with paper records, not electronic records. However, eDiscovery is only about eRecords; so, a records management program also has to encompass

eRecords if it is to be of any assistance to the eDiscovery process. Among the objectives that a records management program brings to the table are classification, retention management, guaranteed document authenticity, and the ability to establish litigation holds—all previously identified as the key issues.

Putting together a robust records management program requires a team of knowledge workers who can carry out the complex and time-consuming task. The data stewards include business unit personnel who are familiar with the different types of eRecords in the business, how they are used, and what value they have in the organization. The specialist data stewards include lawyers and compliance personnel who are familiar with the legal requirements for specific types of eRecords, such as retention policy and confidentiality. The data custodians include IT personnel who understand the types of eRecords that the company generates or receives, who understand the IT capabilities of the company, and who should know where eRecords are stored and how to retrieve them.

But wait—there is more. The team must also include records management specialists. These are the people who manage the records management process. Note that records management is a business function, not an IT function; so, the records management specialists are a type of data steward, not a data custodian. The records management team should be a part of the data governance approach for the business.

Data Mapping: Carrying Out the Data Knowledge Imperative

Although data mapping can be logically put in the identification node of the EDRM model, the concept of data mapping is generic to all eDiscovery requests and not just to particular requests. Therefore, for the purposes of discussion, data mapping will be examined as a part of the information management node.

Recall the data knowledge objective that was added to the list of data protection objectives to accommodate governance. A data map that gives a complete and accurate picture of a company's data sources is essential to achieve that objective for governance and is an essential requirement for a formal eDiscovery information management model.

The data mapping challenge can be overwhelming. The first challenge is the discovery process of simply trying to understand what is available, and that includes trying to understand from a variety of perspectives:

- Structured, semistructured, and unstructured information in all its various forms and permutations, including databases and e-mail archives
- Active changeable production information and active archive information for legacy systems
- Data stored on business systems—direct attached storage (DAS), network-attached storage (NAS), and storage area networks (SANs)
- Data stored on desktop computers and mobile devices, including laptop computers, personal digital assistants (PDAs), and cell phones
- Data stored on both nonremovable media (such as on disk arrays) and removable media (tape, external disk drives, and flash memory drives)
- Data protection information locally on the disk and the tape
- Disaster recovery information, either offsite at a company site or at a third-party site

And that is just the beginning. There are other issues with trying to pin down what information is available, where it is, and how to access it. Information is dynamic in nature in that it can

move around from place to place—and perhaps may do so unpredictably. All information is not under the central control of an IT organization. Information is often in fragmented application silos, where information is isolated from one another.

Finally, knowing just what type of information is where is not enough. Further granularity is necessary. Knowing that there are word-processing documents on a system does not distinguish the value of them from a business-use perspective, nor does it classify them from, say, a sensitivity perspective. On top of everything else, the process of data mapping is likely to be expensive. No wonder, then, that when faced with the mind-boggling challenge of data mapping, most enterprises would like to take a pass.

Unfortunately, taking a pass is becoming less and less an option, as data knowledge is becoming more and more a "have to do" requirement that may very well become mandatory, and data mapping is a means to accomplish that requirement. Now, eDiscovery alone is not necessarily the reason, although governance requirements are a major impetus. However, when the need for compliance is added, further weight is given to the need for data mapping. For example, classifying information for sensitivity purposes may be difficult, but it is necessary to satisfy data privacy requirements. And the third pillar of the GRC framework also has to be taken into account: risk management. Where are the risks and rewards in not doing things properly? Moreover, information is an asset to a business, and businesses use it for other purposes than transactions of one type or another. Mine the value.

Frankly, data mapping is never likely to be complete or perfect, and it does not have to be. Some type of triage approach needs to be taken to determine priorities. The methodologies of information assurance and information risk management may play a big role in determining the risk/rewards that have to be taken into account to assign priorities. Naturally, the low-hanging fruits, such as legacy systems and e-mail systems, are targets for data mapping, but even there, a formal analysis process is likely to be important in helping to determine the granularity of what needs to be done. And not developing a data map is becoming a less-than-viable option.

Again a Return to the Need for Data Governance

Note that data knowledge is not the only data protection objective that a formal eDiscovery information management model can help with. Data preservation now has a data quality attribute for ensuring the completeness, accuracy, and consistency of information. The data auditability objective requires the authentication of information for reporting and evidentiary purposes.

However, a formal eDiscovery model does not have to be—and should not be—developed in isolation. The concept fits nicely into the data governance function. Using the data governance concept within the GRC frameworks creates leverage (funds used can serve multiple purposes simultaneously, except for specific extensions for particular needs) and synergy (combined actions can yield greater benefits for a smaller overall investment).

Project teams that are working on somewhat similar requirements separately will not only spend more money, they will also have to reinvent the wheel in many cases, and they may come up with inconsistencies that have to be resolved at an additional cost. So, once again, data governance plays an essential role in data protection, this time eDiscovery.

Chapter 24

Overview of the Steps of the Electronic Discovery Reference Model*

David G. Hill

Contents

The specific steps of the EDRM model are roughly divided into information collection, information analysis, and information delivery steps for carrying out a particular eDiscovery event process. In summary form, these steps are as follows:

- Information collection
  - *Identification*—locates all the information that may be used in a pending or prospective legal proceeding
  - *Preservation*—protects the necessary information against deletion or alteration that would result in spoliation
  - *Collection*—gathers the information that will be used in the electronic discovery process
- Information analysis
  - *Processing*—reduces the volume of information to only that necessary for legal processing and then converts, as appropriate, into a more manageable format for the review and analysis steps
  - *Review*—examines the information to determine what is relevant to the matter at hand and what can be excluded as privileged information
  - *Analysis*—evaluates a collection of electronic discovery materials from which the relevant summary information can be determined
- Information delivery
  - *Production*—delivers information that is required and relevant to a legal proceeding in the proper format and with the use of the appropriate means of delivery
  - *Presentation*—occurs when electronically discovered information is displayed at proceedings related to the case

The workflow steps are not necessarily linear from left to right. The iteration of the steps may be necessary to refine the process. For example, suppose that during the course of the process, some additional electronically stored information (ESI) is now thought to be relevant. The steps of identification, preservation, and collection have to be repeated for that information.

The process can be seen as more cyclic, given the repetitive nature of requests from legal to information technology (IT) for eDiscovery information. As an enterprise moves to a more proactive model, IT and legal can come together to enable a more seamless cyclic model that reduces the need for repetitive activity. That takes place through the proactive management of ESI content across the whole process, regardless of its role.

Identification

Identification is the discovery process that locates all the information that can be conceivably used in a pending or prospective legal proceeding. As such, the identification step is the data mapping step. If the data mapping has not already been done as a result of an established and ongoing formal data mapping process, this is the step where it has to be done.

This step, however, is more than just data mapping. A litigation response plan also has to be put together. As a part of that process, the key witnesses and data stewards, who have the administrative control of the data, have to be identified. In addition, a meet-and-confer meeting with representatives from both sides of the dispute should take place to define the terms and the scope of discovery based on what is reasonable effort.

In the actual EDRM model, the term *data custodian* is used for the person who has the administrative control of a document or electronic file. That is a reasonable definition in that this person

is the source for information. However, the definition conflicts with the previous usage in this chapter. We shall substitute *data steward* (an agent who administers the data on behalf of the data owner) for *data custodian*, to distinguish the meaning of data custodian as one who physically guards the data (and is essential from a preservation perspective).

The key witnesses and data stewards help to identify what information is relevant to a particular litigation. If a data map is already available, this becomes a subset of the overall map. If a data map is not already available, a targeted data map process has to be followed to find only the information relevant for this particular situation.

Preservation

Preservation results in saving information that may be relevant in a contested matter, whether that is a litigation or a government investigation. A company has the affirmative duty to preserve that information and to produce it as necessary to an adverse party, even though the information may be detrimental to a company's legal position.

Preservation may come after collection, depending on the methods of identification, collection, and processing. The traditional approaches identify documents and then preserve them, collect them, and then process and review them. Newer technology enables the proactive identification and collection of all data across the enterprise, which can be subsequently searched, processed, and placed on preservation, depending on the required responsiveness and scope.

Preservation may result in an enterprise retaining a lot more data than is absolutely necessary. This could become a major problem for organizations that do not take a proactive approach. Without a prescriptive approach to legal hold, backed by a process of authenticity, many organizations face large amounts, if not all, of their corporate data being retained indefinitely due to a preservation order. Without the application of a formal deletion policy, the cost and management implications are significant, especially as the volumes of data grow.

However, prior to the full implementation of the preservation process, a meet-and-confer meeting between the adversarial parties (as mandated by the FRCP) should take place (if it has not already, as discussed earlier). (Some information may have had to have been put on litigation hold in reasonable anticipation of requirements beforehand.) The meeting is an attempt to reach an agreement on the scope and responsibilities related to the discovery process.

The meet-and-confer meeting may or may not take place before one of the parties issues a preservation letter to the other party. The preservation letter is designed to request only the information that might be relevant or important to one's case. Asking for everything but the kitchen sink, even though much of the information may be patently irrelevant, can be considered a bad-faith litigation tactic by the courts. However, honest disagreement may arise. If so, a written counteroffer can be sent, or, if a counteroffer has not already taken place, a meet-and-confer meeting can be suggested.

Litigation Hold

A litigation hold letter has to be distributed to everyone, including key witnesses, data stewards, and data custodians, who are part of the preservation process. All recipients have to formally acknowledge (by signing a certification) that the recipient not only has read the hold, but also understands the obligations that the hold contains and will fulfill those obligations.

When information is put on litigation hold, the normal data retention policies that might enable the deletion (i.e., destruction) of the data have to be halted. Recall the earlier discussion that

this information is subject to chain-of-custody management to establish the causal time history of events that affect the information so that it can be considered to be authentic and therefore eligible to be used as an evidence. That work is necessary to avoid any possible spoliation of data claims.

The data on hold are not just a point-in-time copy of the information at that time, that is, a historical copy of the data. The data on hold are also all the new relevant data that are created during the time that the litigation hold is in place. People who are involved in the preservation process should be given clear instructions on how to preserve the new data.

One of the issues is whether the data are in a production copy of the data or in a data protection copy of the data. If they are on data protection copies that reside on the backup tape, the question arises as to how to manage the process when a rotational scheme is used that involves periodic recycling to a scratch pool of the oldest tapes in the rotational scheme. There are a number of options, but essentially, some tapes have to be identified as relevant to the litigation hold process and have to be taken out of circulation (at least until they have been copied) so that they cannot be recycled.

Another issue is metadata. Metadata is data about the data that accompany the data and are used for tracking, understanding the history of, or managing the data. The recipients of the litigation hold letter should be given specific instructions about the preservation of metadata that is associated with the relevant material as well as the data themselves. This is especially important when ESI is produced (i.e., given) to the requesting party in native format. Native format requires that the application that uses the data accompany the data. For example, to understand a spreadsheet, the underlying formula for a cell must be known, as well as the actual value that was calculated in a specific instance of the spreadsheet. Understanding the metadata is even more important when trying to understand a database application.

Note that a process also has to be in place to enable the data to be taken off the litigation hold and normal business processes (including data retention policies) to resume.

Winnowing Process

A number of factors contribute to the cost of eDiscovery, but the amount of data that has to move through each step in the EDRM model is at or near the top of the list. The more data are involved, the more time of more people is spent and the more support infrastructure, such as software tools and storage, is needed. Without managed processing or winnowing, the full collection set, regardless of relevance, is passed on to legal, which then incurs wasted time and costs associated with obviously reviewing the irrelevant material. Thus, it is no wonder that attempting to limit the scope of what has to be preserved is important.

Trying to reduce the amount of information at each step, consistent with good practices, is a major determinant in being able to keep the costs of eDiscovery within some kind of sensible bound. Winnow down the amount of data from what might be relevant, and so has to be considered as such at the preservation step, to what is really relevant, which is what is delivered in the production step. That difference may be substantial. The difference between what information needs to be preserved and what information needs to be produced could be two orders of magnitude (i.e., 100 times) less.

Winnowing down the amount of data that might be relevant at the preservation step to what is really relevant in the production step is essential for controlling the overall cost of the eDiscovery process.

Note that the amount of the data and the relevance of the data are somewhat inversely proportional to each other, that is, as the amount of data falls, the relevance of the remaining data increases.

Starting off at the preservation step by minimizing the amount of data that has to be moved on to subsequent steps can make the subsequent steps faster and less costly.

Collection

In the collection step, the information that will be used in the electronic discovery process is gathered. Note that not only the content of the data but also the activity of the user may have to be taken into account.

Auditability, Completeness, and Accuracy Are Essential

The data auditability objective must receive paramount attention to establish the chain of custody that is necessary to satisfy the authentication requirements to be able to use any collected ESI and its associated metadata as an evidence. Part of the security procedures should be to identify the privileged work product so that it is not a part of the other data that are collected or produced.

Whoever acts as the collection agent needs to be able to prohibit unauthorized access to the data as well as be able to track all attempts to access the data as part of the requirements for establishing the chain-of-custody process.

Ensuring the completeness and the accuracy of the collection can be a challenge. For example, transformations to the metadata of a file may change during a file's lifetime, either as a result of an action by an end user or automatically by an operating system or other software, such as encryption or migration of the file. These changes may make it difficult to determine that a file was actually created, modified, or viewed by a particular person. Consequently, the processes for determining how to create a complete and accurate collection are important.

Collection Process

Data are typically collected from a piece of storage media—either fixed, such as a hard disk on a laptop or storage array, or portable, such as a magnetic tape cartridge or a flash memory drive—or over a network, such as from a third-party service supplier's storage at a remote site.

Two questions arise: how to collect the data, and where to store them. They are interrelated. One approach is to collect the data by freezing them in place. This is possible in an active archive if the archive management software can handle the process, such as guaranteeing the data retention of the information. However, no guarantee exists that all the data are on an active archive.

However, if the data are not already in an active archive and are available on fixed storage, the collection process in essence creates an archive, since the data collected are by definition fixed—alterations or deletions are not acceptable. This archive could be on the tape and collected via a process called the supervised tape archive process. Alternatively, everything might be written to the disk on a stand-alone governance appliance. The advantage of this method is that the appliance has server capability to run the application software that can present the data in a useful manner. However, the tape method can still be used because copies can be made for both off-site and on-site processing and analysis.

Portable media can be handled in one of the two ways. One is simply to impound the media and store them in a physically secure facility. However, the media might be needed for some business purpose, such as a backup tape. Also, no mechanisms exist for ensuring the chain of custody

because someone who obtained unauthorized access to the supposedly secure facility could alter the data, say, on a flash memory drive. Moreover, processing and analyzing the data may be difficult, as they have to be done piecemeal. So, a catch, copy, and release strategy is more appropriate. *Catch* means to acquire the piece of media, *copy* means to faithfully duplicate the data and secure them for chain-of-custody purposes, and *release* means to return the original piece of media for its originally intended purpose.

For backup tapes, native-environment restoration requires the original backup/restore software to be used in copying (which is the purpose of this particular restoration) the data to an eDiscovery-process-managed piece of media, that is, a piece of media that is chain-of-custody- and auditing-compliant. Nonnative extraction is an alternative method that is typically used by third-party vendors who specialize in backup tape processing. This approach is seen as faster and less expensive than native-environment restoration.

Getting the data back from a third-party vendor depends on the use of the data, but the third party may not allow an on-site visit. If the data are a backup copy, then a native-environment restoration will work because that process has already been put in place. For an archive or an active changeable production database (such as when using a software-as-a-service application), the collection may require some help on the part of the third-party service supplier.

Processing

The processing step attempts to cull the volume of information before the review and analysis steps start. The decisions made in the prior steps of the life cycle—identification, preservation, and collection—somewhat determine the requirements and activities that have to be undertaken during the processing step. That is, the types and amounts of data preserved and collected as well as the time frames for the production step have already been determined.

The processing step gets the data ready for review. The agreement is necessary on both what data need to be processed and what should be the input and output format of the data. These attributes shape the scope of the processing effort, which, in turn, affects the time frame as well as the cost of processing and reviewing the ESI.

Processing Methods

The automated processing of datasets to cut them to an easier-to-use subset reduces the cost of review because attorneys will not have to review what is not there. Overall, the technical approaches or processes that are used to reduce a large amount of data to a much smaller set are called *data culling*.

One technology that is helpful in this process is deduplication, such as the single-instancing data reduction technique to get rid of multiple copies of the same file. Contrast the traditional and new technologies in this area. eDiscovery technologies that have to "grab" documents for preservation, collection, and processing need to deduplicate to drive the efficiency from that point on. Technologies that already have data management under control and are deduplicating the content at the source are more efficient through all the steps of EDRM. The latter approach is obviously about being more proactive (and is an illustration of how a particular data protection technology may be useful for more than one purpose).

File-level filtering using selected metadata criteria, such as selected time stamps that are associated with the file, can cut down the size of the datasets to be reviewed. Moreover, the collection process may grab a lot of unnecessary files—say, when copying a whole disk. The various types of

files, including those of an operating system, may be safely exorcised from the files that need to be reviewed as part of the litigation process.

Of course, an electronic search is a familiar way to filter the data. The specific words that are likely to be relevant to the matter at hand, whether they are in the text of a document itself or in the metadata, can be used in the search process. Or the process may be as simple as finding the names of key individuals on the "from" or "to" lines of e-mails.

The search may employ the familiar full Boolean logic in a search engine. This type of search allows more than one key word to be used in the search process (using search operators such as AND, OR, and NOT). Proximity operators can be used to determine words that are close to one another in a document, which is useful in a context search.

A newer method of searching is called *concept searching*, which is used to identify the content that is conceptually similar to the search terms. No standard method exists for doing concept searching. Currently, concept searching is not an accepted means for eliminating the data from a collected set, probably because a concept search cannot sufficiently conclude the presence or absence of the data that may serve as the evidence. However, concept searching may be very helpful in defining and refining search terms in the processing step and in helping to navigate the data in the review step.

Other Processing Considerations

A key question is whether or not to convert the data for review. One way of converting is to convert to quasi-paper format, such as portable document format (PDF) or tagged image file format (TIFF). However, the conversion process is expensive, and most of the large amount of reviewed material is likely to be deemed irrelevant to the matter at hand anyway; so, an initial review of the documents in native form (i.e., opened with their native application) may be a better first step. Then, if it is necessary, convert only nonprivileged information that appears relevant in this step (obviously, nonrelevant ESI can be excluded from further review). Note that a quality control process, both automatic and manual, needs to be in place to ensure auditability. Reporting, especially when a third-party service supplier is involved, is part of the control process.

One example can illustrate the difficulties. If e-mail is converted into TIFF or PDF format, a lot of the embedded metadata is lost. The internal e-mail distribution list at the time that the e-mail was sent would be part of that lost metadata. So, the original ESI (native-format e-mail) has higher evidential integrity because it would show who was on that distribution list.

Processing documents for electronic discovery can be very expensive. The process of converting and indexing data into a common searchable and usable format can be a very complex and difficult activity. Depending on the situation, the process may be a very labor-intensive specialized activity.

No wonder that innumerable options exist to process the data, from internally with an investment in software tools and infrastructure to the use of outside services.

Review

Once the data have been processed, they are ready for review. The review process determines what documents are responsive. A *responsive* document is one that meets the established parameters of the document request that led to the search process in the first place. Response documents then have to be produced, which means that these documents have to be delivered to other parties in

the legal matter in appropriate forms through the use of appropriate delivery mechanisms. Note that the word *document* is used broadly to include not only any file produced by a software application, such as word-processing documents, but also e-mails, databases, spreadsheets, and graphic files. The review process also excludes documents that are seen to be privileged, such as attorney work product and certain client–attorney communications, from having to be disclosed.

The scope and objectives of the review have to be determined. Reviewing each and every piece of documentation may very well be infeasible; so, limiting the scope of the review through the use of carefully selected technology or other means is likely to be necessary. The key issues need to be documented, and a clear distinction needs to be made between issues of fact and issues of law. The review team needs to know what it should be looking for in the documentation.

Choosing between In-House and Online Litigation Tool Support Technologies

Often, a business may not have all the skill sets or technical capabilities in-house to carry out the review process in its totality. Therefore, the business has to turn to an outside third-party vendor to fill in the gaps that it cannot provide internally. In fact, trying to build a litigation support system internally is probably not a good idea.

Two basic options exist for vendor selection: in-house or online. An in-house review is conducted using an application that is executed and maintained on an internal network. An online review is performed over a network, such as the Internet, to a site hosted by a third-party vendor.

An online tool gives less direct control but increased flexibility. If an in-house tool is selected, a business will want to leverage its investment over multiple legal matters. However, using more than one online vendor over time enables the business to have the flexibility of being able to select the right mix of review functions and features to be able to address each particular legal matter in the best way possible. That approach takes more evaluation time, does not leverage the learning curve in using one set of tools (whether in-house or out of house), and raises the overall cost.

A combination solution may be the answer. Some vendors have partnered with others to provide a combination of in-house solution and an online solution or an online solution that involves different tools to meet different needs. One approach is to have on-premises eDiscovery archiving, capture, and data management solutions that integrate with hosted (out-of-house) case management solutions for the support of eDiscovery review and process analytics across multiple outside counsels. The passing of preserved and processed content for further review and chain-of-custody analysis with the resultant work product information finding its way back into the in-house environment are the features of this approach.

The review process then has to be carried out by the lead attorney and the review team. Productivity and quality control metrics need to be in place to help manage the process.

Analysis

Analysis is the process of sifting through a collection of electronic documents and other materials to find context and content that are important for the legal matter at hand. Analysis helps to find the key patterns and topics within the ESI, identify important people, discover the specific vocabulary and jargon, and target individual documents. Effective analysis requires a blend of good

technology and techniques. The goal is not only to obtain the key information quickly and easily, but also to do so in a less costly and more accurate manner than could be done using an exhaustive manual review. Note that analysis is not a separate step in the EDRM model, but rather, it is part of another step—the review step.

Analysis uses the body of documents that have been output through the processing step as its input. Therefore, ESI has had to be put into an analyzable form, such as being indexed, as a precursor to the search process.

Sample Types of Analytical Tools

A number of analytical tools are available. Search (as previously discussed) in its various incarnations is a primary analytical tool. Clustering technology groups together items of ESI on the same topic, such as documents including e-mails. However, clustering is not an exact approach; so, it may include items that should not be included and exclude items that should not be excluded.

Guided navigation is an analysis tool for examining a collection set or the result of a query, where the data items have been categorized in a number of ways. This is a drill-down approach (which has been employed in business intelligence tools for years) in which one can start at a high level and then drill down to successive levels of details (from category, find the key person; from the key person, find another person with whom the first person has exchanged e-mails; from the e-mails exchanged between the two people, find the e-mails most relevant to the legal matter at hand). This helps with not only gaining an overall understanding of what is going on, but also to be able to find the specific issues of importance.

Visualization tools create a visual output to help create a better understanding of the relationships among the items available for analysis, using the famous principle that a picture is worth a thousand words. One type of visualization is a social network analysis that shows the most important people in an analysis and the communication links between them. A context group structure is another technique. This technique depicts the actions that are taken in an electronic e-mail discussion over time, such as replies and forwards of e-mail.

Topical cluster analysis visually presents the interrelationships and the internal structure among groups of documents and messages. These are just some of the tools, and all their strengths, limitations, and applicability in a particular situation have to examined before a decision can be made on whether or not to use any of them.

Production

According to the Federal Rules of Civil Procedure, "a responding party must produce the information in a form or forms in which it is ordinarily maintained or in a form or forms that are reasonably usable." The four basic choices are paper, quasi-paper, quasi-native form, and native. Paper is, of course, self-explanatory.

Quasi-paper is a little trickier. Quasi-paper is an electronic version of paper, such as an immutable PDF of TIFF files. The only difference is that some metadata may be incorporated with the text.

Quasi-native means producing ESI electronically in a format that can be read by an application other than the native application that was used to create the information in the first place.

Native form means the document is produced using the file extension for the information of the application that created it. For example, a spreadsheet file would be in a particular file format with a particular file extension and would require a copy of the spreadsheet program to be able to

read the file. Keep in mind that an original copy should be preserved for production and a working copy should be used for review purposes, as any interaction is likely to change the metadata associated with the document. Typically, the application has no means for freezing any changes to the metadata of a document as part of a litigation hold.

Produced documents should be run through a quality control process prior to release. Managing the redaction process may be a problem. *Redaction* is the process of removing privileged information from documents prior to producing them. Although rule changes allow the ability to retract privileged information if it has been inadvertently produced, it is best not to produce the privileged information if at all possible.

The production process has to be carefully managed for the ESI that it sends to another party as well as for the produced ESI that it receives. Checking for completeness, making sure that the ESI is safely stored, and making sure that only authorized people have access are all essential.

Presentation

The respective attorneys have to decide what ESI they will display at various events during a legal matter, such as depositions, hearings, or trials. They need to be equipped with the technology that enables this to be done, such as a computer with a native application on it, a spreadsheet application, and an overhead projector, which can display images from the computer to all the people at a legal event. The selection of a relatively small set of ESI that has reached this stage for presentation is correct based on the best judgment of the legal team for each party involved in the legal matter.

Chapter 25

Cell Phone Protocols and Operating Systems*

Eamon P. Doherty

Contents

The cell phone originally was a simple telecommunication device that was only used to make calls. The operating system was originally an embedded system with a simple interface that allowed people to conveniently operate the hardware, namely the phone's keypad, to make or receive calls. From the time the cell phone was invented until the early 1990s, it was only for making and receiving calls. Phones such as the Motorola D520 that was announced in 1998 were part of a widely accepted trend of new cell phones that allowed for SMS messages. The operating systems had to become more complex as consumers now wished to synchronize their computers and their cell phones to update large phone books with points of contacts for both calling and sending SMS messages.

* From Eamon P. Doherty, *Digital Forensics for Handheld Devices*, Copyright 2013 Taylor & Francis Group, LLC.

In the early 2000s, there were many sophisticated cell phone operating systems for phones such as the Microsoft CE versions, Palm OS, Android, Symbian, Apple's iOS, and others. These phones acted like small computers and each their own following of people who used software development kits (SDKs) to create applications. By 2011, some forensic tools such as Mobil Edit divided smartphone operating systems into the following groups: Apple, Blackberry, Android, and Windows 6.5. However, this is not a complete list because there is the Symbian operating system which is used on smartphones all over the world. Other cell phones might have the Palm operating system which has been around since 1996 when it was first introduced on personal digital assistants (PDAs).

Many cell phone forensic practitioners and students have noticed that the cell phone has become a small-handheld computer with the ability to send and receive email, surf the web, watch movies, and do many activities that they could only do previously on their desktop computer. There is an operating systems concept called "feature migration," which means that the capabilities of large computers move to mid-sized computers over time and then eventually to handheld devices. If one wishes to know what trends will eventually evolve on handheld devices, one needs to look at the operating systems and activities of large powerful desktop systems. The sixth edition of *Operating Systems* by Siberschatz, Galvin, and Gagne is an excellent source for gaining more understanding of feature migration as well as the theoretical aspects of how small-scale and large-scale operating systems work.

If one reads the user agreements that accompany mobile devices, it appears that legally replacing operating systems with newer ones are not possible. This could be the reason why so many people frequently purchase new cell phones and mobile devices with newer operating systems. It is also generally well known that older operating systems offer less protection than new ones because hackers and others learn and publish online the vulnerabilities of older operating systems. Preston Gralla, a writer of a book on malware, says that keeping your operating system, applications, and security software updated increases security. Preston Gralla mentions one cell phone virus in his book, but in 2011, it is generally well known and highly publicized that there are over 1000 viruses or strains on cell phones and smartphones. Viruses were once something only computer users worried about, but now they have migrated to cell phones.

Cell Phone Operating Systems: Finding the ESN and IMEI

If one navigates through the cell phone operating system, there is always a subsection about system hardware models, device serial numbers, electronic serial numbers (ESNs), IMEI (international mobile electronic identifier) numbers, and the version of the operating system one is using. In Windows Mobile 6.0, it is easy to get to the name, address, phone, and email of the owner. Just go to settings, owner information, and the details appear. If an examiner wishes to get the device serial number, go to settings, system, identity, and advanced details. It is very important for anyone interested in cell phone forensics to learn about the ESN. The ESN is a unique identifier that the manufacturer of the cell phone embeds in it. Each time a person makes a call, a signal is transmitted and is picked up by a cell phone tower that provides support for that phone. A mobile switching office can check the ESN and any other data embedded in the signal associated with that phone number against a network database to see if it is valid and part of the network. If the data are correct, then the call is allowed to proceed through the telecommunication network and also logged for billing. The call may also be routed through other networks in other countries if it is an international call where someone may answer it.

The IMEI is 15 digits long and numeric. It includes numbers 0–9 and does not use the hexadecimal number system. Some dual-SIM (Subscriber Identity Module) cell phones such as the SciPhone have two different IMEI numbers in the phone. The IMEI is found in GSM (Global System for Mobile Communications) phones and is used worldwide to help prevent cell phone fraud and theft of service to the cell phone service provider. The IMEI also has a bar code next to the 15-digit number. If a person scratches out the IMEI and bar code to avoid identification, it is possible that enough of a bar code is left to scan. If one does not have a bar code scanner, an investigator may bring the cell phone to the library and ask the librarian to scan it where they scan library cards. Most modern supermarkets also have a bar code reader at the checkout counter where one should be able to see the results.

The IMEI is found by removing the cell phone's plastic back cover and battery. In New Jersey, some municipal law enforcement officers who are continuing education students have told me in an academic setting that if they wish to obtain the IMEI number without the consent of the suspect, a search warrant is needed to remove the back plastic cover from the cell phone. This was because the IMEI is not in plain view, and the phone is considered a closed container. The Fourth Amendment protects U.S. citizens from unreasonable search and seizure. The same is true about a SIM card inside a phone. However, since I am not a lawyer or legal authority, it is best to check with your state's Attorney General's office.

The SIM is an acronym for the subscriber information module found within the cell phone. It contains the information about the phone subscriber, the cell phone number, and a code that lets them use the cell phone network. SIMs can often be taken out of one cell phone and put in another. The second cell phone can often be powered on and used almost immediately. I performed an experiment with a vintage Blackberry model 6230 that had an LCD (liquid-crystal display) monochrome screen and a newer Blackberry model 7290 with a color screen. I took the SIM out of the first phone, placed it in the second phone, and then successfully placed a call to a colleague. The SIM was needed to make the call and the IMEI or the ESN was sent to the network too. It may be possible for a law enforcement officer with a subpoena to get the information about the call that was placed. It may also be possible for the law enforcement officer to know which phone was used to place a call if the telecommunication provider keeps information about the ESN that is embedded in the signals. There are logs that the service providers keep about calls, which may prove important in certain cases. In the United States, telecommunication service providers are required to assist law enforcement officers in lawful investigations where a search warrant was issued. The United States federal legislation is known as CALEA and became effective in 1994. It was passed for matters of national security as codified at 47 United States Code (U.S.C.) §§ 1001–1002.

Bryan Miller wrote an article for the *New York Times* on July 20, 1995 about phone cloning and being vulnerable at chokepoints such as the airports and near the George Washington Bridge in New Jersey. One possible way that the phone cloning can happen is if people would park a car on the side of a road near crowded chokepoints such as a place where highways would converge to one wide area with toll booths. The George Washington Bridge is a crowded place at any time, day or night, as cars pass to or from New York City. Phone cloners are people who may work as part of a ring. One person might sit in a car with a wireless sniffer and collect mobile identifier numbers (MINs), ESNs, and possibly subscriber names. This person might get so many pieces of data with the sniffer that a laptop is needed to organize and create records in a spreadsheet for each caller that was eavesdropped on.

Policemen have told me that when phone cloning goes on, the person with the spreadsheet would later sell the spreadsheet to another person. That second party would create SIM cards with

the stolen information and configure the phones to have the same ESN. These phones were then sold to people to make expensive out-of-country calls. The owner of the real original SIM card was the one who was billed. This activity was illegal and once common in New Jersey until the New Jersey State Police ran a successful campaign to stop it.

There is another interesting discussion of an illegal cell phone exchange racket in India. Phones were being cloned for making calls between India and Dubai. Cloning is when the real cell phone number of a person and the cell phone's ESN or the ESN is put on another phone and then the real owner gets the bills for the calls.

Cell Phone Operating Systems and Protocols: Synchronization

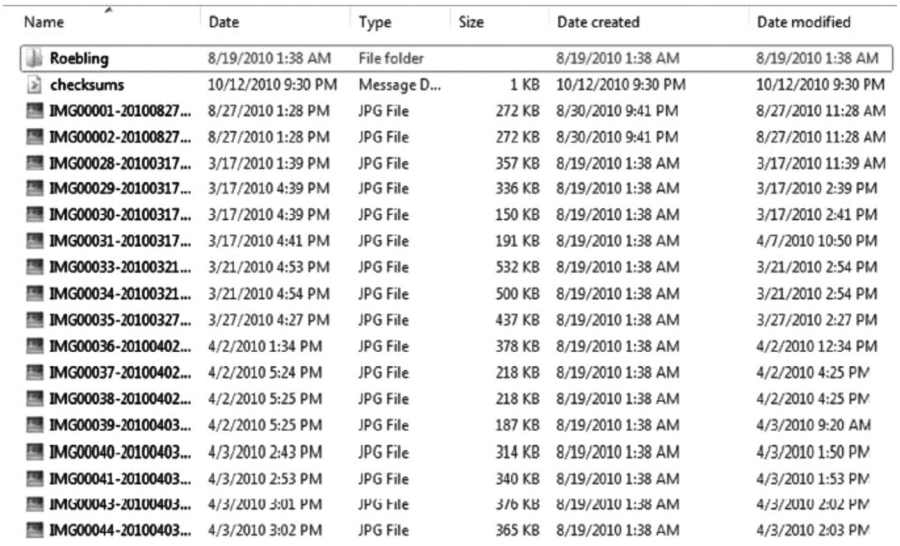
When people speak of cell phone synchronization, they could mean two things. The first meaning could be to make sure that the files in the desktop computer are the same as the ones in the cell phone. Many people add contacts to their address book and wish the latest copy of the address book to be the same on both the desktop and the cell phone. Many people synchronize their cell phone with their desktop at the end of a workday and put all the latest pictures, address books, and documents on the desktop with all the changes. The second type of synchronization has to do with updating the time and date of the cell phone with the present time zone one is in. An example of this type of synchronization would be if someone drove through Eastern Indiana to Western Indiana where the time zone changes from Eastern Standard Time (EST) to Central Time (CT). The following paragraph will discuss a real example of this time and date synchronization.

Some cell phones such as the Blackberry Bold 9700 came preconfigured to synchronize with the time and date of the locality where it is. The date and time is checked against the network and if there is a difference between the phone and the network, the operating system loads the date and time from the local telecommunication network. Consider this example: if one flew from Japan to New York and the flight only took 11 h due to a tailwind, the cell phone will go back in time 1 h as soon as the cell phone is powered back on in New York.

Some cell phones such as the Blackberry Curve will often need to be manually reconfigured when time zones are crossed because they do not update automatically. The question of automatic date and time updating becomes important to the investigator because files and logs have time stamps. Files have a creation time and date as well as a modification time and date (see Figure 25.1). If someone is a suspect in a harassment investigation, does the time and date of certain activities coordinate with the events on the phone? Fortunately, many cell phone investigative tools give the current date time of the acquisition and the date and time on the cell phone. Many of the cell phone tools such as Susteen Secure view show the MD5 hash values for the files in addition to the date and time. If one seized the system files from the cell phone and compared the MD5 hash files with the known correct MD5 hash values of the operating system for that phone, one could find possible tampered files from malware.

Cell Phone Differences Worldwide

It is generally known that European cell phones are approximately 6 months ahead of American cell phones in terms of storage capacity, features, and in the number of megapixels on the digital camera. The cell phones in South Korea, Japan, and Hong Kong are generally considered to be 1 year ahead of those in the United States. To illustrate an example of this advancement



Name	Date	Type	Size	Date created	Date modified
Roebling	8/19/2010 1:38 AM	File folder		8/19/2010 1:38 AM	8/19/2010 1:38 AM
checksums	10/12/2010 9:30 PM	Message D...	1 KB	10/12/2010 9:30 PM	10/12/2010 9:30 PM
IMG00001-20100827...	8/27/2010 1:28 PM	JPG File	272 KB	8/30/2010 9:41 PM	8/27/2010 11:28 AM
IMG00002-20100827...	8/27/2010 1:28 PM	JPG File	272 KB	8/30/2010 9:41 PM	8/27/2010 11:28 AM
IMG00028-20100317...	3/17/2010 1:39 PM	JPG File	357 KB	8/19/2010 1:38 AM	3/17/2010 11:39 AM
IMG00029-20100317...	3/17/2010 4:39 PM	JPG File	336 KB	8/19/2010 1:38 AM	3/17/2010 2:39 PM
IMG00030-20100317...	3/17/2010 4:39 PM	JPG File	150 KB	8/19/2010 1:38 AM	3/17/2010 2:41 PM
IMG00031-20100317...	3/17/2010 4:41 PM	JPG File	191 KB	8/19/2010 1:38 AM	4/7/2010 10:50 PM
IMG00033-20100321...	3/21/2010 4:53 PM	JPG File	532 KB	8/19/2010 1:38 AM	3/21/2010 2:54 PM
IMG00034-20100321...	3/21/2010 4:54 PM	JPG File	500 KB	8/19/2010 1:38 AM	3/21/2010 2:54 PM
IMG00035-20100327...	3/27/2010 4:27 PM	JPG File	437 KB	8/19/2010 1:38 AM	3/27/2010 2:27 PM
IMG00036-20100402...	4/2/2010 1:34 PM	JPG File	378 KB	8/19/2010 1:38 AM	4/2/2010 12:34 PM
IMG00037-20100402...	4/2/2010 5:24 PM	JPG File	218 KB	8/19/2010 1:38 AM	4/2/2010 4:25 PM
IMG00038-20100402...	4/2/2010 5:25 PM	JPG File	218 KB	8/19/2010 1:38 AM	4/2/2010 4:25 PM
IMG00039-20100403...	4/2/2010 5:25 PM	JPG File	187 KB	8/19/2010 1:38 AM	4/3/2010 9:20 AM
IMG00040-20100403...	4/3/2010 2:43 PM	JPG File	314 KB	8/19/2010 1:38 AM	4/3/2010 1:50 PM
IMG00041-20100403...	4/3/2010 2:53 PM	JPG File	340 KB	8/19/2010 1:38 AM	4/3/2010 1:53 PM
IMG00043-20100403...	4/3/2010 3:01 PM	JPG File	376 KB	8/19/2010 1:38 AM	4/3/2010 2:02 PM
IMG00044-20100403...	4/3/2010 3:02 PM	JPG File	365 KB	8/19/2010 1:38 AM	4/3/2010 2:03 PM

Figure 25.1 The date and time stamps on Blackberry Bold 9700 cell phone pictures.

concept, please consider this example that William Bulkeley, a writer for the *Wall Street Journal*, gave on February 8, 2007. Ten megapixel cameras were for sale in Europe and Asia while only 3.2 megapixel cameras were available in the United States. The problem with cell phones with advanced hardware is that they often contain newer versions of operating systems than American cell phones. Newer foreign phones most often cannot be examined by most of the state-of-the art cell phone forensics software available in the United States.

I addressed this concern with cell phone forensic tool vendors and law enforcement officials at the September 2010, High Tech Crimes Investigation Association Conference (HTCIA) in Atlanta, Georgia. People whom I spoke to from the law enforcement community expressed concern that their present cell phone investigative tools could not allow them to investigate new models of foreign phones. This could be a problem if there is a need to search a cell phone in a timely manner. The phone would need to be sent to a Regional Computer Forensic Laboratory, known as an RCFL. A case manager would then interview the officer about the case and assign it a priority. If there were other many high-priority cases, then the phone could not be examined in a timely manner, thus allowing the chance for the recovery of digital evidence to be lost.

Mobil Edit was one tool that was displayed at the 2010 HTCIA conference in Atlanta that provided support for a range of foreign phones. Many digital device investigators said that they would like to see their present forensic software vendors also include support for foreign phones. They told me that the support would be best provided in an upgrade of their present forensic software because limited budgets often prevent them from purchasing additional tools.

Many of the foreign phones with larger-capacity storage take longer periods of time to examine because there is more material to examine. With more processing power, higher-resolution screens, increased storage, and more available bandwidth, people are surfing the net, creating and sending video, making phone calls, and sending as well as receiving email. This volume of digital media becomes an increased burden on the examiner. Since the certified computer examiner's (CCE) time is expensive, many law offices just ask the CCE to image the phone and pass the image of the cell phone to a paralegal to examine and then create a report on it.

Imaging is a term that means to copy all allocated and unallocated parts of storage. Solomon, Barett, and Broom, three authorities on digital forensics, also refer to the image as a forensic duplicate. The forensic duplicate starts at byte zero and goes to the last byte of the storage device. The duplicate or image contains parts of files, parts of sessions, and many pieces of important forensic evidence that can help the CCE get a clearer understanding of the case and create a report. An image is different than a backup because a backup only includes files that are presently listed in the file allocation table. A backup does not include existing files that were marked for deletion.

Many law offices that were using paralegals rather than CCEs to save expenses also found that the volume of material to sift through was still too voluminous. This also made the use of paralegals too expensive. Now, many law firms have resorted to hiring document readers who may have little legal training. This becomes a problem because they may miss some pertinent evidence because of their limited training and knowledge of law.

However, who can manually read all these documents, emails, watch the videos, and completely report on all relevant items in the phone? Now there is expensive e-discovery software that organizes the contents of an acquisition, produces charts of associations, and helps sift through the enormous amount of material. E-discovery tools can help save time and labor but since e-discovery tools are very expensive, there are some who feel that the number of law firms or private investigators that can perform large-scale quality digital forensics investigations will dwindle to a few large companies with deep pockets.

Cell Phone Differences Worldwide: Various Bands

Some people use the term "tri-band" with regard to various types of cell phones. Tri-band means that the cell phone uses three very different sets of frequencies. The reason that every country does not use the same frequency is due to different parts of the world using 9 The Cell Phone different frequencies for television, radio, aircraft, and cell phones. A band plan for a country may show what frequencies are used for aircraft, cell phones, and other broadcasts. The band plan may also show acronyms such as GHz or MHz so it is important to first obtain the electromagnetic spectrum to learn what signal frequencies are microwaves, VHF, AM, FM, shortwave, radar, x-ray, infrared, and ultraviolet light. The electromagnetic spectrum gives one an idea what type of signals exist and how they are classified. Shortley and Williams have an easy-to-read electromagnetic spectrum chart in their book *Principles of College Physics*.

The cell phone networks use not only different frequencies but also different protocols such as GSM or CDMA. It is advisable for the student of cell phone forensics to start his or her education by looking at some cell phones that are in use in his or her locality and learning what frequencies and protocols are used there. Some computer science students who attended Fairleigh Dickinson University in 2003 were from Hong Kong and had what was known as tri-band phones. These cell phones could be used in Hong Kong, the United States, and in the United Kingdom. These phones were a necessity for students who often flew to these countries to visit relatives, go to school, or assist in their family's international business.

This paragraph will illustrate a real example of a quad-band phone that utilizes four bands. In 2010, the quad-band cell phone such as the SciPhone mode i68 ++ became popular on eBay. It is made in China and is sold for approximately US$50. It uses a GSM quad-band protocol, which includes the 850, 900, 1800, and 1900 MHz frequencies. The interesting concept about frequencies is that the higher the frequency, the shorter the wavelength. Low frequencies have

longer wavelengths. The shorter waves do not travel as far as the long waves. Wave propagation is the term used with how far a wave travels. Jerry Wilson, a writer of a physics book, describes the mathematical equation that explains frequency and wavelength as

Speed of light/frequency = wavelength

Cell phones that utilize higher frequencies must use more power than cell phones that operate on lower frequencies, or there must be cell phone towers and repeaters in closer proximity to retransmit the signal. Shannon's law of information basically states that the higher the frequency, the more the bandwidth, and the more information that can be sent. Higher frequencies are often preferred for cell phones with high-resolution digital cameras where high-bandwidth video needs to be sent. *Newton's Telecom Dictionary* says that Shannon's law is "a theorem defining the theoretical maximum at which error-free transmission can be transmitted over a bandwidth limited channel in the presence of noise." *Newton's Telecom Dictionary* states that the theorem works out to approximately 10 bits per hertz in analog circuits. However, in 2011, there are many ways of getting more information on a signal such as using digital circuits and quadra angle signaling. For more information on signaling, frequency, and bandwidth, Roger Freeman's *Fundamentals of Telecommunication* would be a good source of further information.

A cell phone examiner should be able to Google search the brand and model of a phone to see the protocols of that cell phone and learn which geographic regions utilize that protocol. Can a cell phone that utilizes certain protocols and frequencies make a call in the area where the alleged threat took place? If a person has a vintage monoband American cell phone that uses TDMA (time division multiple access), it probably should be ruled out as a possible phone that was used to make a threatening call in the United Kingdom in 2011. 10 Digital Forensics for Handheld Devices the terms "MHz," "kHz," and "GHz" are often used without much explanation. A hertz or cycle is a sine wave. A series of 1000 cycles is a kilocycle or now what is commonly referred to as a kilohertz. It is based on $10 \times 10 \times 10$. It is different than a kilobyte, which is 2 to the 10th power, or 1024. A MHz is also based on the power of 10 and not 2. A million hertz or MHz is a million cycles whereas a megabyte (MB) is more than 1 million bytes. An MB is 1,048,576 bytes or 2 to the 20th power. Shortley and Williams give advanced explanations of signaling and may be useful for the cell phone investigator who needs to understand the physics behind the technology.

Many cell phones have very high-quality miniature microphones. The digital signal also allows a high fidelity of audio to be transmitted across the phone networks as compared to analog circuits. That is why the cell phone is being used as an electronic listening device by stalkers. There are numerous news stories of cell phone stalking victims that are available on the Internet, which are worth watching to learn about the extent these devices can be used to frighten the people. For those unfamiliar with the case, certain family members would often receive calls on the cell phone commenting on clothes they wore or activities that they were involved in. The specificity of the comments showed that they were being watched.

The mechanics of this stalking will be explained now for the purpose of educating cell phone forensic examiners. A person may download spytones for the cell phone because they do not ring. Then he or she would set the phone to autoanswer and hands-free calling. This allows the phone to be called and used as an electronic eavesdropping device to covertly collect video and audio. A cell phone investigator should also consider the time, date, call logs, frequencies used, and the nature of the case to determine if the phone was also unknowingly used for stalking someone or spying on them.

Cell Phone Internal and External Storage

Cell phones have a variety of internal and external storage configurations. Some Blackberry Bold models sold in 2010 in the United States for example have a 2 GB internal storage card as well as a considerable internal storage capacity. The Blackberry Bold 9900, sold in India in 2011, is one example of a device that was reported to have 8 GB of internal storage. The increased storage capacity is the result of research and development that increases the number of transistors within a chip. Here, we see an example of both Moore's law and the availability of advanced cell phones outside the United States.

It is also a good idea to review the units of storage if one is going to examine cell phones. The unit of storage encountered in digital devices is the byte. Early examples of cell phones used kilobytes of RAM. The kilobyte (kB) is actually 1024 bytes but some people will round it off to be 1000. Megabyte (MB) is often rounded off to 1 million bytes but is actually 1024×1024 bytes. Gigabytes (GB) are usually rounded off by people to mean 1 billion bytes but are actually $1024 \times 1024 \times 1024$ bytes. These little differences in theory and practice are good to know if one has to appear in court and must answer questions from the opposing counsel.

A cell phone will have random access memory known as RAM. This RAM is volatile and means that if the power is lost, the contents of RAM are lost. However, some operating systems may use a swap file and the contents of RAM may be found in a swap file. This is why knowing the model of phone and some details about the operating system becomes so important in an investigation. There are often a variety of external cards such as micro-SD cards that can be found in cell phones or nearby in the possession of a suspect. External cards are important because they may hold important evidence that could be known as exculpatory data, or data that proves someone innocent. Investigators need to look for both types of evidence that can prove someone guilty or innocent. In Figure 25.2, one can see the MD5 hashes of some files of simulated exculpatory evidence showing that the identical-looking pictures were different.

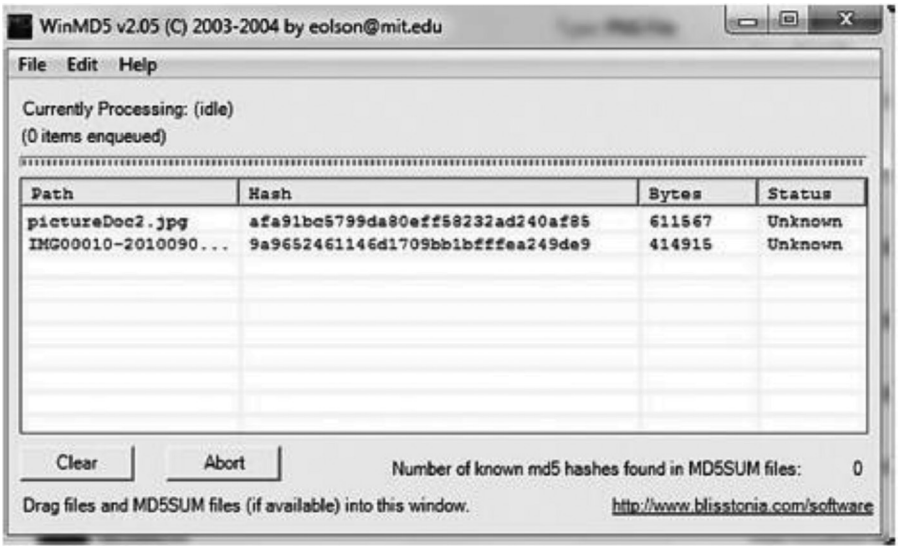


Figure 25.2 Different MD5 hashes of identical-looking pictures.

If one performs an investigation of a cell phone, he or she should make sure that the e-forensic tool can be operated to perform a seizure of the internal memory, internal storage, external cards, and a SIM card. A SIM card is mainly for identifying a caller initiating a call on a telecommunications network but its secondary purpose is its ability to be a storage device for a person's address book. The address book or books of the cell phone owner are not necessarily saved in the SIM but it may be optionally saved there. Because it is optional, it is also important to check it for data too. The acronym SIM means subscriber identity module. Since the capacity of SIM cards has increased, a new tertiary purpose is to use it as a storage area for saved text messages. It is important to remember that the SIM card holds the phone number and identifier of the subscriber at the minimum, but could also possibly store address books and text messages. The SIM card has traditionally worked with GSM digital networks but recently there are SIM cards in China that can also work with CDMA protocol phones.

In Figure 25.3, we see a phone on a camera stand waiting to be examined. The dual-SIM card phone means the cell phone simultaneously holds two different SIM cards. Each SIM card has its own identity and telephone number. It is not uncommon for people to have more than one SIM card for legitimate purposes. A person may wish to have an identity for work and an identity for his or her personal life, in order to keep one's business life and personal life separate. Each card may be in a different name and even have a different service provider such as Verizon and T-Mobile. Some people have two SIMs because of the lack of quality coverage by any one company where they live.

Cell phone service provider companies provide varying qualities of service for the same geographical areas because some service providers may have standard antennas on a pole along a highway while others may have a high gain antenna on top of a water tower. Some people may also have a SIM card with a number for the United States and one for 12 Digital Forensics for Handheld Devices another country such as England. When people go to another country,

Figure 25.3 A cell phone being examined using a camera and stand.

they often prefer to have a local phone number and a card from the service provider of that country. The calls are much cheaper when it is considered a local call. There are also some models of cell phones that hold three SIM cards such as the Intex IN5030 that is available in the Far East. This type of cell phone can be used to hold one personal identity and perhaps two business identities.

The important consideration when doing an examination of a cell phone is to make sure that one uses a separate research machine that goes online and looks up the specifications of that model of cell phone. The examination machine stays offline to avoid malware and connects directly to the cell phone to collect and collate the evidence. The research machine is connected to the Internet and can be used for purposes such as surfing the web to learn if a certain cell phone model supports multiple SIM cards and external memory cards. The examiner should systematically check the internal memory, external cards, and all the SIM cards for possible evidence. Lastly, it is important to consider that a person may have an illegitimate business that is done with a completely different identity on a removable SIM. The examiner should look for evidence of wear in the other SIM card slots and also examine the call logs of the phone and operating system files to see what identities were used in that phone.

Internal Cards: Sim Cards/Locked and Unlocking

Many times people have SIM cards, which bind that SIM card and GSM phone to a certain network. The Sony Ericsson T610 is one example of a cell phone that will give an error message to put the original authorized SIM card back if someone replaces it with an unauthorized SIM card. This appears to be a technical security measure built into certain cell phones that is done to make sure that a person who received an expensive cell phone as part of a promotion cannot dissolve their business relationship and go to a cell phone service provider who charges less money. Many people have complained about the locked SIM cards. A group of programmers created programs that can be found online that "unlock" your phone and remove the restriction of the cell phone to only accept a particular SIM. This allows a person to go to another cell phone service provider.

Many people use the term "unlocking a phone" synonymously with the term "jail breaking a phone." However, the two terms are very different. A cell phone investigator may hear the term "jail broke phone" from someone in a chat room or as part of an investigation. The term is often confused with unlocked. Unlocked means that the security feature within a GSM cell phone was reconfigured to accept another SIM card other than the one originally sold with the device and provided by the cell phone service provider. The term "jail broke phone" means that the cell phone was reconfigured by a special application to allow programs that are unapproved by the cell phone manufacturer for that phone.

If one has a jail broke Apple iPhone, for example, he or she may be able to download and run applications that may easily allow the cell phone camera to zoom in on a subject as a picture is being taken. Another example of an unapproved application available online for jail broke cell phones may conveniently allow the user to choose fake GPS coordinates and embed them in the phone. Suppose a person is at a casino in Atlantic City but wants to tell the boss at his or her company that they are in New York City at a conference. The person may take a picture of the Atlantic City conference room but use the unapproved application to embed the GPS coordinates of the conference where they are supposed to be in New York City.

The Need for a Faraday Bag

Cell phones are devices that connect to telecommunication networks through wireless signals. Some cell phones can also connect to Bluetooth devices or wireless networks. If the security features are not enabled on the cell phone, then it may be possible to connect to the phone and alter, delete, or add digital evidence to the phone. It is very important that the digital evidence be preserved from the time of seizure until it is presented as evidence in court. If evidence is suspected of being tampered with, it could be ruled as inadmissible in court. Therefore, it is important for CCEs to preserve digital evidence by using a Faraday bag and noting its usage on the chain of evidence form. Smith and Bace, two authors of a forensic testimony book, discuss the importance of preserving evidence and protecting the integrity of digital evidence.

Faraday bags look very similar to antistatic bags. The difference is that the antistatic bag prevents damage to the device from small electrical charges that have built up and are discharged from static electricity, but it does not protect the device from outside connectivity. Static bags are obtained commonly when purchasing electronic equipment such as a wireless weather station, computer memory chips, or an EZPass transponder for the car that makes it convenient to pay tolls. The Faraday bag is based on the concept of a Faraday cage. The Faraday cage is an enclosure that prevents outside signals from penetrating the cell phone or examination equipment. The Faraday bag is made with materials that block wireless signals from entering the bag, thus protecting the integrity of the device in the bag from outside influences.

The Faraday bag will not prevent the device from internal data alteration by items such as logic bombs. A logic bomb is set to go off if certain conditions are met. If a person was supposed to simultaneously press a set of keys daily to keep a destructive program from running on the cell phone, this would be one example of a logic bomb. The phone that was seized from someone may be protected from outside control of hackers with the use of a Faraday bag, but the phone may be victim to a logic bomb if certain conditions are not met while the phone is in possession of the CCE.

I was once teaching a class about cell phone forensics to a class of visiting cybercrime students from Kyungnam University from South Korea. One of the students asked if another type of metallic bag such as an aluminum foil bag could be used in an emergency situation if no Faraday bag was available. To demonstrate an answer to the student's question, I placed various cell phones in aluminum foil bags and asked students to call the phone. The signal was blocked. The phones also did not appear as wireless device icons on the student's laptop display. The lesson learned was that aluminum foil bags give some protection from connectivity but a proper Faraday bag is best.

I explained that the effectiveness of the aluminum foil bags was not known and that could offer an unwanted line of questioning in court. Each tool and methodology used in the collection, preservation, and examination of digital evidence should be able to withstand the Frye Test. The Frye Test helps ensure that the tools and methodologies used to gather, process, and examine evidence in an investigation are accepted as general practice by authorities in that field.

Investigative Computer and Precautions to Take

It has already been stated in this book that it is very important to preserve the integrity of the digital evidence obtained from the telephone from the time it was seized until it is presented in court. It is important for many digital evidence incident response team members to use a checklist

so that they know that they did not forget anything. Once the evidence is obtained, a chain of custody form should be filled out. Each time the evidence is copied, processed, or transported, it should be documented on the chain of custody form. If others receive a copy of the evidence for prosecution or defense purposes, they too should sign for it.

When the examiner is ready to investigate the phone, he or she may have a checklist to make sure that the examination machine is ready. This computer that is known as the examination computer can be a laptop or a desktop. The main requirement is that it has at least a Pentium 90 for processing speed and enough RAM to operate the cell phone forensic software. There must also be enough available storage for the contents of the seized phone. This may be a challenge with an older computer but fortunately old computers with Windows 98 had USB ports and this allows USB external storage devices to be used. The examination computer should have a current, properly licensed copy of the examination software used for the phone.

Since this examination computer is used to download the data from the phone and then prepare it in a readable format for examination, it is best to check in the operating system that the USB port is working properly. In Windows 7, one can go to Control Panel, Device Manager, Universal Serial Control Bus, and then double click. It would also be a good idea to close all unnecessary programs.

This examination computer should also be protected from unwanted outside connectivity. This means that the wireless ports, infrared ports, Bluetooth ports, modem port, and Ethernet port should all be disabled. This examination computer may also be checked for viruses by running a current version of a properly licensed antivirus program to remove the possibility of a virus altering the data. A properly licensed program for antispyware should also be run. A scrubbing program such as CCleaner that deletes all unallocated hard drive space and clears temp files should be run too.

The examiner may also wish to prepare a Faraday cage or a stronghold tent such as the one from the Paraben Corporation so that no signals will escape or penetrate the area where the cell phone examination is being conducted. The examiner may elect to sit in the stronghold tent when conducting the examination so that the evidence is not tainted by outside signals from hackers or people who wish to compromise the examination. Each person that enters and leaves the examination facility should sign in and sign out. If an examiner retires, transfers, or quits, the cylinders for the locks should be rekeyed. It is recommended that each examiner has a badge with one's name, organization, weight, and height. If a person has a dramatic change in appearance due to weight or facial hair change, then a new picture should be placed on the badge. Cameras should also be placed at the entrances and exits so that visitors, examiners, and illegal entries are documented on film. If one contacts ASIS International, they have a bookstore that sells the Certified Protection Professional (CPP) set of study guides that address these physical security best practices.

Precautions: Examining Phone—High-Profile Case

If it is a high-profile case, then it may be important to make sure that the examination room is in a Tempest facility with no possibility of eavesdropping. This means that only filtered power is used so that the wiring for the outlets cannot be used for transmitting data from the examination room. Heating and air conditioning ducts should have a grate in them so that someone cannot climb in the facility and eavesdrop on the investigation or compromise the data.

The walls of the examination facility should be like the ones at the New Jersey RCFL that contain a mesh wire system so that a criminal cannot easily breach the security of a wall. There should

also be no false ceiling or raised floors where people can easily hide eavesdropping equipment or crawl into. The walls should have copper so that signals cannot escape. A guard and close-circuit television are also necessary. It is also suggested that the examination of the digital device should not be conducted in a room where there is a flat roof above. This is because criminals can use a Sawzall or reciprocating saw to cut a hole in the roof after hours and get to the evidence. The NIST has a set of guidelines that discuss the security needed for Tempest facilities.

The digital examiner should also make sure that his or her credentials are current and not out of date before performing the examination on the cell phone. If he or she took online instruction, attended a workshop, or got any continuing education, it should be noted on the resume in case the examiner's credentials are questioned in court. Before the examination begins, the examiner should question all possible weaknesses or conditions that could cause doubt with a jury if the case should ever go to court.

Precautions: Protecting Equipment from Static Electricity

Cell phones are electromechanical devices with operating systems and can be affected by low batteries, humidity, temperature, and other complex environmental factors. Sometimes, examination machines have intermittent hardware failures on USB ports and hard drives. The operating systems may also get corrupted. It is therefore important that cell phone examiners have more than one forensic tool and examination station because of all the complex variables that must work together in order for a cell phone to be seized and the electronic evidence to be collated by the examination software. I have demonstrated the same cell phone on the same examination machine numerous times with varying results. The cell phone may be acquired on the first time but many times it takes three or four attempts before the process is successful and complete. It is not anyone's fault but there are so many factors concerning the environmental factors. The discharge of static electricity in the winter months is a big concern. That is why it is important to wear a grounding strap, use an antistatic mat, or use a static potential equalizer known as a "Static Buster" as pictured in Figure 25.4. If none of these options are possible, at least touch a metal object before touching the phone or computer.

Figure 25.4 Static potential equalizer—Static Buster.

CCEs often tell me that it can take one or many times to acquire the evidence on the cell phone because of the correct parameters on the variables of the phone and examination machine that must be within acceptable boundaries for the process to work. Some cell phone examiners that use powerful desktops to run the forensic cell phone software will connect the desktop to an uninterruptible power supply because the power often intermittently goes out in the summer due to storms. Brown outs from too many air conditioners running on hot days is another reason for intermittent power losses. Even a short bit of power loss will cause the desktop computer to reset and reboot, thus ruining the data acquisition from a cell phone with high-capacity storage in it. Therefore, many cell phone examiners will have two examination machines with uninterruptible power supplies, and a licensed copy of BitPim, Device Seizure, Susteen Secure View, and Mobil Edit. Others will have a Cellebrite and a small portable storage device for incident response in the field. Because of the number of things that must be right, having a wide range of tools is considered a must.

Susteen Secure View can now incorporate the data of its e-forensic competitors in its reports so now juries and judges do not have to read a multitude of reports for each device. Cloke, Goldsmith, and Bennis, experts in workplace conflict and resolution, give numerous examples that one might consider parables, which appear to teach that more can be gained from cooperation than competition. This lesson has appeared to be internalized by Susteen, the maker of Secure View. Many forensic examiners have said that they have also bought Secure View so that they can incorporate the data from other forensic tools such as Device Seizure of the Cellebrite UFED products in their main report.

Putting It All Together: Cell Phone Hardware

We reviewed many aspects of the cell phone in this chapter and will look at a dissected one in Figure 25.5. There is the case front and back case along with a separate latch for the battery. There is also a coiled-up antenna. In the 1990s, many phones had telescoping metal antennas. There should also be a speaker that allows the person to hear the conversation. There is also a screen that may be an LCD in the 1990s phones or a color display in the twenty-first century phones. There may be holes that allow combination headphones and microphone sets to be plugged in. There is a full keyboard on modern cell phones but cell phones of the 1990s have a dual tone multifrequency (DTMF) board, which only allows number calling, on/off, and dial number.

Figure 25.5 The internals of the cell phone.

All cell phones have a speaker or built-in microphone for speaking into. There is also a printed circuit board, which holds all the components and has a bus system on it for data travel. Many people are familiar with large printed circuit boards in computers, which are called "motherboards." There is usually a metallic hinge that holds a SIM card. The SIM can be used for storage and also has phone subscriber information that is used for identification and billing purposes. There may also be internal storage built in the cell phone which could be volatile but probably is not. There should also be a large microprocessor which acts like a CPU in a computer.

Then there is a radio frequency and power section. Some college students have simplified the concept of cell phone hardware and say that it is half radio/transmitter and half computer phone. Newer phones may also have connectors for external cards such as SD cards and receptacles for cables to connect with computers. It is important to understand the basic hardware so that one knows where to look for evidence. There is the internal chipset with the phone's identity, internal cards, external cards, the SIM, internal memory, and perhaps other storage or computing devices that are connected to the phone.

Major Categories of Computer Crime

Chapter 26

Hacktivism

The Whats, Whys, and Wherefores

Chris Hare

Contents

Since the inception of the commercial Internet in the early 1990s, more and more of our daily lives have gone digital. We shop, do research, look for information, go to school, stay in touch with family and work, all of it online. Today's generation is more digital savvy than any previous generation, and it has changed our social culture dramatically.

The digital age has made it easier for people with a message to get it out to the mass population. Criminals now conduct their crimes online, with significant impact on the global economy, and on the individual victim as well. No longer is it necessary to buy magazine space, television time, or commit a major crime like a bombing to communicate a message.

Hacktivism is about the technology, but it is also less about the technology. That seems like a contradiction, but the technology is merely the means to an end. It is the cross-pollination of political, religious, psychological, philosophical, and sociological ideas that lead the hacktivist to their next target and campaign.

What Is Hacktivism?

Simply put, hacktivism is hacking in its most malicious sense for political messaging. Hacktivists are individual computer hackers or groups who are "aimed at promoting a specific social or political cause" (*New Oxford Dictionary*). This means we can say that hacktivists are skilled computer users, who are trying to gain support for a particular cause. Hacking, in the context of the hacktivist, meets both definitions of a hacker:

- A person who uses computers to gain unauthorized access to data.
- An enthusiastic and skillful computer programmer or user (*New Oxford Dictionary*).

Hacktivists may not be specifically after the target's data, but they are interested in disrupting the target's business or online presence. They do this using their computers and particular programming skills to accomplish their goals.

Hacktivists are, for the most part, hackers. Gabriella Coleman, an anthropologist at McGill University in Canada, has been studying the hacker culture for several years. Coleman (2010) describes it this way: "A hacker is a technologist with a love for computing and a 'hack' is a clever technical solution arrived through a non-obvious means."

An exceedingly simple definition to some, but it drives home the point that "hacking" does not have to be illegal, immoral, or unethical. As a society, however, we hear the term "hacker" and we cringe. Consequently, hacktivism has a strong negative connotation. Part of that negative connotation is a direct result of the damage caused by these individuals, and there are countless examples. Additionally, the widespread media attention, and in some cases exaggeration of the event and its impact, glamorize the hackers and hacktivists adding more fuel to the fire.

Some might argue that Robert Tappan Morris, who unleashed the first computer worm back in November 1988, was a hacktivist. While his motive was ostensibly to measure the size of the Internet at that time, he made use of specific software vulnerabilities to spread the worm. In the end, Morris' intentions were obliterated by the massive denial of service attack affecting email delivery around the world and the more important message regarding the vulnerabilities in commonly used software.

While hacktivism has been gaining more and more attention in the popular press, it is by no means new. Activists have been with society for as long as there have been people—they have a message or objective and they will use any means available to them, whether it means standing on a street corner or using the Internet.

Hacktivism dates back to the early 1990s, when virus writers and hackers began to move to politically oriented messages or targets. *The New York Times* reported:

> Hackers from China and Taiwan dueled against one another. Anti-nuclear activists defaced an Indian government site after the government in New Delhi conducted nuclear tests. Hackers rallied in support of the Zapatistas, an insurgent group in southern Mexico that even in the pre-Twitter days used the Internet to drum up overseas support (Sengupta, 2012).

With so many people already connected to the Internet—the latest statistics indicate more than 2.2 billion or about a third of the planet's total population (Internet Population, 2012)—not only is it easy for hacktivists to gain attention to their cause, but any significant disruption in the Internet fabric will result in media attention, thereby furthering their cause.

Hacktivism allows its participants to convey their message against their specific target without the need for traveling, expensive media resources, and without risking any physical harm to themselves (Jevans, 2011). While their intent may not be to harm anyone else, the repercussions of their actions cost many organizations in both the recovery of their information assets, their web sites, or even their reputation. No matter how you slice it, hacktivists get what they are after as a result of their activities: media publicity.

While many of the hacktivists actions involve using the technology to "attack" their target by defacing the target's web site, stealing information and so on, it does not have to be that way. As described earlier, hacktivism is about getting your message out and driving change through the use of technology to communicate and implement that change. Hackers, and by inference hacktivism, has a strong subculture, with "complicated ethical codes" (Coleman, 2010) and a strong, albeit not isolated, emphasis on "freedom, meritocracy, privacy and free speech" (Coleman, 2010).

Anthropologist Gabriella Coleman has spent years studying the hack mentality and specifically the Anonymous group. She describes them as tricksters, who both "scare the daylights out of people" and are sometimes playful. Either way, they are difficult to grasp from a sociological perspective (Sengupta, 2012).

In many cases, such as Anonymous, the hacktivists do not know who their counterparts are. This anonymity provides protection for each other, but it also makes it hard for them to know if anyone has turned on them from within their ranks (Sengupta, 2012).

What Is Digital Activism?

Digital activism does not differ much in definition from hacktivism. Digital activism is defined as, "The practice of using digital technology to increase the effectiveness of a social or political change campaign" (MacManus, 2010). Mary Joyce is one of the pre-dominant digital activists and launched a project in 2010 known as the Meta-Activism Project, which is more of a digital think tank to study digital activism. Joyce commented in her interview with MacManus (2010) that very few examples of digital activism have actually been a success.

Activists determine if there has been a success in their event based upon achieving their goal. For example, if the goal is to overturn an allegedly fraudulent election result through demonstrations and that goal is not achieved, then the activists would declare the event a failure whether or not social media or other digital technologies were used to organize it.

The basic approach between hacktivism and digital activism appears to be the same, that being to impact a social or political campaign. That is where the similarity stops. Digital activism

is about using the various activist approaches and implementing them using digital technology to more rapidly communicate and organize events where interested parties can converge to convey the message to the identified target, be it political, cultural, or corporate. Hacktivists do not typically operate in the confines of a group, and use the technology not to convey information about an event but as the basis for the event itself.

Electronic Civil Disobedience

But are hacktivism and digital activism the way of the future? Stephan Wray in Marketou (2001) commented that "As hackers become politicized and as activists become computerized, we are going to see an increase in the number of cyber-activists who engage in what will become more widely known as electronic civil disobedience (Marketou, 2001)."

Electronic civil disobedience is defined by Wikipedia as "Electronic civil disobedience, also known as ECD or cyber civil disobedience, can refer to any type of civil disobedience in which the participants use information technology to carry out their actions. Electronic civil disobedience often involves computers and the Internet and may also be known as hacktivism" (Wikipedia, 2012a).

This definition makes ECD similar to hacktivism, although people involved in ECD are not interested in hiding their actual identity, while it appears those involved in hacktivism are keenly interested in hiding their identity to avoid legal repercussion under various computer crime laws.

One primary commonality between hacking, hacktivism, digital activism, and ECD is that they all have their own unique terminology allowing the members of those particular cultures to communicate with each other and to some degree obscure that communication and its specific meaning to the rest of the society.

Why Do Hacktivists Do What They Do?

Politics exists in all things, whether we like it or not. Many would argue that technology has increased the impact of politics. With the number of people and varied opinions of its users, the Internet has had a political impact across the globe. While the Internet has provided the opportunity for global political discussion on things like "Access, technological determinism, encryption, co-modification, software, intellectual property, the public sphere, decentralization, anarchy" (Marketou, 2001), the Internet is also the realm of "no-copyright, plagiarism, confusion and exchange" (Marketou, 2001). This makes the Internet a landscape for political discussion, debate, communication, and ideology development.

The vast growth of services like Facebook, which if it were a country would be the third largest in the world, provides an excellent example of the speed at which technology has impacted our lives. Services such as blogs and social media provide a forum where any individual can document and share their ideas, regardless of the degree of their acceptance among their friends, colleagues, geographical region, or society in general.

There are many reasons these so-called "bring the facts to the masses" groups like Anonymous and LulzSec do what they do. Many of their activities are to communicate a highly specific message, such as the attacks against the Church of Scientology, religion in general, corporate behavior, politics, lack of computer and/or information security, and just for the fame and glory of it all (Jevans, 2011).

In some cases, hacktivists may be just looking for an outlet to express their own frustrations with the current state of society, without the fear of any repercussions, because after all, they

are smarter than the rest of the population. In March 2012, a member of Anonymous, Jeremy Hammond, was arrested for his part in a series of hacktivist activities. His mother, Rose Collins, supports the position that they are smarter than the rest of the population in one respect, and not in others: "He [Hammond] has a 168 IQ, but has no wisdom" (Ranson, 2012).

The hacktivist groups today are more focused on their message and getting the public relations or media attention associated with their exploits. The role hacking actually plays is less of an issue, although every Internet user and many others feel the effects of their activities through many means. Hacking is a tool—a means to an end—for the hacktivist. Because they have the intelligence and technical ability like Jeremy Hammond just makes it easier for them to accomplish their online objectives.

But hacktivists have been around for years. Many of them are highly astute programmers or technology savvy individuals who use the available technology to satisfy their own unique perspectives. Some of this mentality arose from the concept of free software or the GNU General Public License, also known as copyleft (Bluestein, 2012). But what was more important according to Bluestein (2012) is that the open source software movement, from which software like Linux arose, has no political associations. There is almost a sense of being outside the traditional notion of intellectual property law because much of this software is protected by the GNU General Public License and not traditional copyright or intellectual property law (Bluestein, 2012).

The notion of software not protected by the traditional copyright and intellectual property protection had a major appeal to the hacker and hacktivist alike because both prefer to consider themselves outside the scope of traditional law, or unbound by it. They have their own ideals about what society should be like—software should be free, as espoused by the GNU Project. The GNU Project did not exactly intend for software to be free—only for it—software—to provide the purchaser the ability to do what they want with it, even if they had to pay for it (Stallman, n.d.). The GNU Project believes that the purchaser should have the right to expand upon the software, change it and augment it to meet their unique requirements. What the GNU Project was actually trying to say was aptly defined in their definition of free software:

> "Free software" means software that respects users' freedom and community. Roughly, the users have the freedom to run, copy, distribute, study, change and improve the software. With these freedoms, the users (both individually and collectively) control the program and what it does for them (GNU, n.d.).

This in essence is akin to the right to free speech, but not the right to a free television or beer.

The question then is, what does this—the notion of free software—have to do with hacktivists? Since hacktivists are out to communicate a message, often a political one or one which they deem to be oppositionally defiant to their way of thinking, they will use open source software as a tool to help them achieve their goals. This does not make open source software bad, but it emphasizes how any technology savvy person can take either commercially built or open source software and use it to achieve a potentially illegal end.

However, because of the GNU project and many programmers building applications under the protection of the GNU General Public License, the hacktivists have access to the source code to modify it and change its operation to suit their particular needs. Commercially produced software does not make its source code available, as that is deemed to be intellectual property, which means hacktivists and hackers alike cannot make effective use of it to achieve their goal.

When you think about the changes in technology, even over the past 40 years, the pace at which the global community received information has been astounding. During the Vietnam

War, film had to be transported back to the United States. With the changes in technology by the time of the Gulf War, news information about the war came to US viewers in real time. Now with the global Internet, the speed at which information is available through Facebook, Twitter, and even the ubiquitous web page, every Internet user has the ability to send out a message aligned with their particular cause or complaint of the day in minutes, if not seconds.

These changes in information delivery and availability make it easier for information security professionals, researchers, and the general public to get access to the information they need, whether it be data on the latest software vulnerability, political candidates or where to buy house windows. These same changes, however, have it made it easier for hackers and online criminals to access and steal information, deface web sites or launch targeted attacks against networks, companies and organizations, and even countries.

The "Arab Spring" events in the Arabian countries in 2011 should be testimony enough to the value of online services and Internet-based communication. It is reported that more than 86% of the activities on Facebook by Tunisia members were solely for the purpose of organizing actions and spreading information throughout the world and the country about the events. The lesson is that citizen news reporting and social mobilization was enhanced by the "instant" communication methods available today (Miercoles, 2012).

Specifically, the Arab Spring movement, while not hacktivism in the sense of the definition presented earlier in this chapter, is an example of digital activism. As reported in the Arab Social Media report published by the Dubai School of Government, many of the demonstrations and events in the various Arab countries were actually organized on Facebook (Dubai, 2011).

The technology and legal changes over the last 10 years have encouraged Coleman and other interested parties to examine hackers and hacktivism; to examine the methods used to refashion age-old political concerns. The impact of the hacker and hacktivism modus operandi have reached far beyond the information security and technology realm. The reasons they do what they do extend far into psychology, philosophy, and sociology. Organizations can take every precaution to protect themselves from the possibility of a hacker attack or falling victim to an offended hacktivist. Good programming practices, proper implementation of a multilayered defense and human awareness all have great impacts.

Major Hacktivists

While there are and always have been activists for a wide range of topics, there are not that many groups that call themselves "hacktivists." Certainly, there are a wide range of "hackers" currently using the Internet, but there are a lot more of them who do not engage in the hard-core antics of the more media-visible groups.

Anonymous

In 2008, a group of individuals got together and started "pranking" the Church of Scientology, which had stated they owned copyright and trademark over sacred church scriptures. Pranking is a term used by Anonymous to describe the activities they engage in with a target. Eventually, these hackers, or as we call them now as hacktivists, ganged up on the Church of Scientology because it challenged their belief system, that being the church represents itself as a "super-proprietary secret system" (Bluestein, 2012). Eventually, the group moved out from just the Church of Scientology to other targets as well.

Anonymous' history dates well back before 2008, into 2003 when a chat forum named 4chan was created by Christopher Poole. It was this early chat forum where people were recruited to participate in online "pranks." From here, the basis of what is now known as Anonymous was formed (Knafo, 2012a).

Anonymous is making use of technology heavily used by today's younger generation, such as Twitter and YouTube. Because there is no single point of reference—any of the Anonymous members could create a YouTube account and post video, there is no easy way to measure the number of videos or tweets which exist in YouTube or that are sent everyday using services like Twitter and Facebook, although Anonymous has called for shutting down Facebook (Brown, 2012).

Time reported Anonymous as one of the 100 Most Influential People In the world for 2012 (Gellman Barton, 2012). Well, Anonymous as we know is not a person, but a group of unknown individuals taking shelter behind the mask.

As a group, their victims vary widely including

- The Vatican
- Banks
- Entertainment companies
- FBI
- CIA
- Stratfor (a security firm)
- The San Francisco Bay Area Rapid Transit (BART) System
- Church of Scientology
- The United States Department of Justice
- CBS
- Recording Industry Association of America (RIAA)
- Motion Picture Association of America
- Spanish Police Departments

The list of victims of this nefarious organization ranges across religious and political boundaries.

To provide a look into Anonymous, one needs only to examine the communication sent to a Spanish Police department after arresting three people who were allegedly involved in the PlayStation Network attacks (Plafke, 2011). After the three suspects were arrested, Anonymous sent this communication to the Spanish Police (Plafke, 2011):

> Greetings Spanish Government:
>
> We know you have heard of us; We are Anonymous. It has come to our attention that you deemed it necessary to arrest three of our fellow anons, . . . which you claim to be the leaders of Anonymous and for their participation in DDoS attacks against various websites . . .
>
> First and foremost, DDoSing is an act of peaceful protest on the Internet. The activity is no different than sitting peacefully in front of a shop denying entry. Just as is the case with traditional forms of protest . . .
>
> Regardless of how many times you are told, you refuse to understand. There are no leaders of Anonymous. Anonymous is not based on personal distinction . . .
>
> Arresting somebody for taking part in a DDoS attack is exactly like arresting somebody for attending a peaceful demonstration in their hometown. Anonymous believes this right to peacefully protest is one of the fundamental pillars of any democracy . . .

> You have not detained three participants of Anonymous. We have no members and we are not a group of any kind. You have, however, detained three civilians expressing themselves . . .
>
> You are providing us with the fuel, but now you must expect the fire.
> Awaiting your action.
> Anonymous.
> We are Legion.
> We do not forgive your attacks on freedom.
> We do not forget your ignorance.
> Expect Revolution.
> Expect us.

A study of the linguistics involved in the communication would be an interesting analysis. However, the important message in this communication is their stance on a distributed denial of service (DDoS) attack. The victim of the DDoS would most certainly not consider it the same as a demonstration. Having a crowd of demonstrators outside the office building is very different from having the organization's online presence, which may be the means for generating income, impacted by denying access to it, or defacing it.

For information security professionals, it is easy to see that if Anonymous' view is that a DDoS is not only an acceptable behavior for Internet users and should not be punished, even though there may be direct consequences from the Computer Fraud and Abuse Act of 1986 (in the United States), then it is highly probable they will resort to other, more offensive and "dangerous" acts.

Saki Knafo of the *Huffington Post* has written several articles about Anonymous. In a discussion with an individual from Anonymous known only as "Phoenix" after the January 2012 attack on the Department of Justice, Knafo asked "Phoenix" why they did it. "Phoenix" responded to Knafo explaining:

> "You've heard Anons say before that this is a war," he said. "A full scale information war. That's not mere propaganda, many regard that as a perfectly accurate description. And the stake at play is, simply, 'Who will control access to information? Everyone or a small subset?'"
>
> In case it wasn't clear, he then labeled that subset: "The government."

Anonymous hasn't always been at the forefront of the media by causing widespread Internet "attacks," or something they prefer to call pranks or "raids." In their beginning, Anonymous spent a lot of time "trolling." "At its most basic level, trolling is about humiliating people who seem to take themselves too seriously or pretend to be something they're not: 11-year-old girls, corporate executives, whoever. The troll jabs at them until they jab back, exposing their vulnerabilities, then jabs at those weak spots until they do something rash and truly embarrass themselves" (Knafo, 2012a).

LulzSec

LulzSec was founded after a group of Anonymous "members" decided it was necessary to put some of "lulz" or laugh, back into their pranks. However, they also "targeted police departments and other familiar foes, along with seemingly inoffensive companies" (Knafo, 2012b). LulzSec formed as a spin-off after a series of internal struggles within Anonymous in 2011.

The exact nature of the internal struggles or reasons for the break from the Anonymous ranks is not clearly known. However, LulzSec, or Lulz Security, clearly became a predominant force during the latter half of 2011 (Brito, 2011). LulzSec, according to Brito (2011), do not seem to be motivated by financial gain, nor do they appear to be state-sponsored spies. This means we can lump them into the same column as Anonymous—that of hacktivists.

However, unlike the later philosophy of Anonymous, LulzSec appears to be a throwback to the early days, when the emphasis was on doing things "for the lulz" (Brito, 2011). While the group's web site is no longer on line, reports illustrate that it followed similar trends. The web site reportedly had the theme from the TV show "Love Boat," where the group had posted modified lyrics calling it the "Lulz Boat" (Brito, 2011). Unlike the "preachy manifestos" (Brito, 2011) issued by Anonymous, LulzSec issued comical press releases (Brito, 2011).

The "lulz of it" can be demonstrated in LulzSec compromise of the Black & Berg Cybersecurity Consulting firm's web site. While it apparently took little time for LulzSec to compromise the site and put a picture of their mascot on it, they refused the reward, saying they did it "for the lulz" (IBTIMES, 2011).

LulzSec did many things "for the lulz." That is the claim made by the group, although it was challenged by various people. It is on the one hand easy to see from some of their activities that they might be doing it "for the lulz," and hard to understand where the "lulz" is in some of their others.

Among its victims are

- Black & Berg CyberSecurity Consulting (by request)
- PBS
- United States Senate web site
- Unveillance
- Sony Pictures
- Bethesda Softworks
- Nintendo
- EVE Online
- *Escapist Magazine*
- Minecraft
- Riot Games
- CIA
- United Kingdom's Serious Organized Crimes Agency
- Arizona State Police
- Fox Broadcasting Company
- News International
- Infraguard Atlanta
- United Kingdom's National Health Service
- Westboro Baptist Church

LulzSec did eventually start to "attack" Anonymous' commonly accessed sites. What followed was a not-so-civil barrage of denial of service attacks from both sides, leaving the rest of the Internet users stuck in the middle (Lynley, 2011). Despite their accomplishments, the arrest of members of LulzSec likely led to the decision to disband the group in June 2011.

But even in their final hours, LulzSec released hundreds if not thousands of usernames and passwords for a wide variety of web sites (Cluley, 2011). The moral is that many organizations with web sites are still not taking the proper care of usernames and passwords for user accounts on their

web sites. As Cluley (2011) points out, there are likely people happy to see LulzSec gone, not only from the perspective of the illegality of their actions, but because they would not be a target of the group's activities.

As of June 2012, legal action against the members of LulzSec is still ongoing.

WikiLeaks

The exact nature surrounding the creation of this organization has been difficult to uncover. There are rumors that the organization was initially created by dissident Chinese students among others including mathematicians, journalists, and start-up company technologists (Whittaker, 2011). None of the founding individuals have ever been confirmed. Exactly how the most visible face of WikiLeaks came to be the spokesperson is also unknown. The WikiLeaks domain name was registered in October 2006, with Julian Assange being the visible face of the organization since 2007 (Wikipedia, 2012b).

Julian Assange himself has been convicted of hacking (Whittaker, 2011), and so the logical progression from hacker to hacktivist makes some sense. However, Assange has portrayed the work done by WikiLeaks as "Scientific journalism" (Bland, 2010). In her article "Telling the Story of WikiLeaks," Hannah Gurman (2011) describes Bill Keller of the *New York Times* as referring to WikiLeaks as a group of "propagandists and not journalists."

Keller's perspective is justified by his explanation that journalists take information from a vast number of sources and assemble it into a narrative (Gurman, 2011) for the public to read and form opinions from. WikiLeaks makes no such effort. They take the information which is received from the various whistleblowers who have provided documentation and simply make that information available to people without analysis or narrative.

WikiLeaks, as an organization, publishes these vast quantities of data because they have self-proclaimed themselves as the source providing "information, transparency and debate" (Whittaker, 2011), originally against governments in the Middle East. That focus was short lived, as the group asked any who had information to provide it to "expose unethical behavior in their governments and corporations" (Whittaker, 2011).

Since its inception, WikiLeaks has published thousands of documents from various corporations and governments, some of which have had a profound effect on foreign relations between countries. The organization has relied upon individuals who had access to documents to provide them to WikiLeaks for publication. In some cases, information was obtained and subsequently published using illegal activities such as hacking into or gaining unauthorized access to networks, computer systems, and data. In these instances, the term "whistleblower," which has been used to provide the basis for protective legislation in the case of corporate or government wrongdoing, has been stretched to fit these illegal activities.

Stratfor Global Intelligence

The hacktivist group Anonymous may have attacked the security firm Stratfor, but WikiLeaks published the files Anonymous obtained. This is likely not the first collaboration between groups in this manner, but it was certainly well publicized. WikiLeaks typically obtains its data through the cooperation of a whistleblower, but this was not the case in this high profile incident.

Stratfor (www.stratfor.com) is a United States-based intelligence gathering firm with more than 300,000 subscribers to their service (Ball, 2012). Their emphasis is geopolitical intelligence

and analysis focusing on international affairs using traditional media, open source monitoring, and human intelligence (HUMINT) gathering to collect the information they use in their analysis (Stratfor, n.d.). The company had to launch a massive campaign regarding the incident, including an open message to Stratfor's current and future clients regarding the incident. Stratfor's president, George Friedman also explains in the announcement that some of the messages which were released could be fraudulent or have been altered. One such example are email communications indicating he had resigned (Hacking News, n.d.).

While most employees feel there is some measure of "privacy" in email communications between the sender and recipient, this event is a perfect example of how that privacy can be violated and why every email user should bear in mind that the recipient may not be the only one who actually reads the message.

Many of WikiLeaks targets have been governments or government agencies. Stratfor was neither, and the exact nature of any loss experienced by the company in terms of real money, costs associated with rebuilding systems, lost customers, and so on are not available.

Like Anonymous and LulzSec, WikiLeaks has published documents from, and affecting a vast list of organizations and governments including

- United States Army
- Swiss Bank Julius Baer
- United States Government
- United States Department of Defense

Many of the documents leaked were associated with some of these victims, but dealt with specific subject matter areas including

- The United States War on Terror
- Iraqi War
- Afghanistan War
- Peru Oil Scandal
- Icelandic Financial Crisis
- United States Diplomatic Documents

How Has Social Media Helped the Hacktivist?

Social media services like Facebook, Twitter, and YouTube have provided valuable resources for hacktivists to convey information between each other or to assist in the communication of actual events for people to participate in. However, in the case of Anonymous specifically, they still relied heavily upon Internet Relay Chat (IRC) as their communication method.

One reason for still using IRC is the ease with which they could hide their identity. Users can still use Facebook to create false identities, but since Facebook is a company, it could still be ordered by law enforcement to provide what information they have on suspected user accounts. Since IRC has no concept of a user account, and an IRC server can be operated by anyone with a connection to the Internet, it is much harder, if not impossible to determine which server a user connected to and thereby force the server operator to provide information.

Twitter did become popular with Anonymous and LulzSec as a method for communicating with its followers and "drumming up" support for various raids or pranks. Making the leap from a

legal perspective from a Twitter "follower" to a participant would be difficult for law enforcement, not to mention extremely time consuming.

Just looking for one video on YouTube related to Anonymous provided a list of 20 additional video clips associated to the group—and that is just a start. Twenty does not seem like very many, however, YouTube does not provide a comprehensive filter to search through the 126,000 video clips associated with the word "Anonymous." That could include keywords assigned to the video, the word Anonymous in the title, or the video posted by a user with the term "anonymous" in their user name.

Topics range from specific messages to groups of individuals, governments, and churches, to examples of their exploits. The possibilities are endless, and it cannot be easily identified exactly how many YouTube videos exist that are actually related to the group.

The Impact of the Hacktivist

Activism, regardless of the specific label attached to it, is primarily trying to bring awareness to a particular cause. Perhaps someone has been unjustly tried and convicted—or the particular society thinks they have. Take a look at Nelson Mandela in South Africa. Mandela spent 27 years in prison for his beliefs and challenging the government to end apartheid. The rest of the world knew about it through the activism movement and spreading the word through various news sources.

Had Mandela been imprisoned today, the public outcry and resulting hacktivism and hacking events would have been significant. The speed at which the rest of the world knew about it would have been almost real time. Regardless of the event, technology has sped up our ability to communicate globally. It used to be that information like this chapter would have only been available when you bought the book. However, with technology changes making it easier for anyone to publish a book, the ability to spread a new idea, belief system, ideology, or how to make a bomb is more accessible to everyone.

Groups like Anonymous target government agencies on a regular basis because the government and notion of government in any form challenge their anarchist philosophy. Consequently, Anonymous members are active on sites like YouTube, where they can bring attention to their actions, and gain support for them. Unfortunately, since Anonymous members hide their identities, determining even a good approximation of the number of Anonymous posted clips on YouTube is pretty much impossible.

Do organizations feel an impact of hacktivism? Absolutely. The FBI felt the sting as recently as January 2012, when Anonymous found information about a conference call between the FBI and Scotland Yard to discuss issues of hacking, Anonymous, and other topics (Gardham, 2012). In this case, potentially legally sensitive information was discussed, and then published by Anonymous using YouTube for everyone to hear.

Additionally, there has been extensive fallout from the WikiLeaks publication of the United States diplomatic cables. The initial publication of the first group of cables occurred in five countries: Spain, France, Germany, the United Kingdom, and the United States (Wikipedia, 2012b). None of those initial cables were classified as top secret or higher, but they were damaging enough. Comments ranged from critiques of countries hosting US Embassies, comments regarding foreign diplomats or government leaders, threat assessments from around the world, and dealings between various countries (Wikipedia, 2012b). Some of the revelations in those cables have the opportunity of significant outcomes (Calabresi, 2012), but many of them are of no specific consequence.

Is My Organization a Target?

Jon Cilley at Bit9, a global leader in advanced threat detection specializing in detecting and preventing advanced cyber threats, thinks it is not likely that many organizations would ever become a target of groups like Anonymous. However, in their recent survey conducted in 2012, their survey respondents felt that Anonymous was the top threat.

Cilley takes the position that it is "more likely to have your data stolen from a cyber criminal or nation-state than the hacktivist group" (Cilley, 2012). This is likely the case when you look at the number of publicized Anonymous events and the actual amount of data stolen. Things just do not correlate.

Here are Cilley's (2012) reasons why most organizations would not be targets of groups like Anonymous:

1. Positive business practices and proactive stances—Anonymous pays attention to the news and tends to follow and act upon news reports dealing with topics like privacy, consumer protection, web, and revoked web services.
2. You're not popular enough—Anonymous' members want headlines. If taking down your web site or stealing your information won't result in massive amounts of media attention, then it is likely they will pass you over.
3. Your private data can't be sensationalized—There is a lot of work that goes into a hack, and if the general population, let alone the media, don't care about the stolen information, meaning it doesn't contain stuff worth reporting on the news, then Anonymous likely won't pay attention.
4. There's always a bigger fish—Even if you meet the first three criteria that still doesn't mean you will suffer an attack. The lack of structure in the groups means its actions are very "linear"—they typically deal with only one target at a time. Consequently, by the time they are in a position to get around to your organization, something else may have popped up and you are forgotten about.

However, you still may not be out of the woods. Things change. People join and leave Anonymous, and something your organization does may provoke one or more of them to decide you are important enough. More importantly, however, is that hacking and stealing intellectual property has been going on for more years than Anonymous has been around. It is well known that some countries have active corporate espionage programs to acquire research and development information and gain an advantage over the competition.

With the current trend to store more and more of our information, both personal and corporate, in the "cloud," that just makes it easier—the hackers need only compromise the cloud once, instead of each company to acquire the information (Cilley, 2012).

Implementing a Hacktivism Protection Plan

Unfortunately for the information security professional, there is no specific antihacktivist protection plan available defining what must be implemented to protect the information systems within an organization. There are some things information security departments can do to protect the network, computer systems, and information associated with that organization.

Some elements of that information security plan include

- Corporate Policies
- Application Development Standards and Practices

- Systems Development Life Cycle
- Code Reuse
- Automated and Manual Code Analysis
- Automated and Manual Application Testing
- Application Authentication Standards
- Computer and Network Security Requirements

This is certainly not an all-inclusive list of the controls an organization can use to protect its information; however, these are some of the more relevant topics to this discussion.

Corporate Policies

Every organization needs a set of policies, procedures, standards, and guidelines to establish the baseline manner in which business is conducted. These policies include employee ethics, code of conduct, electronic information handling, information security, network security, and so on. One of the most important corporate policies in the context of this discussion is data classification.

Data classification policies establish the rules for how information is identified and the limitations of using and sharing that information based upon its classification. As part of the data classification policy, the organization establishes a consistent series of data classification types such as top-secret, secret, confidential, and so on. These data classification types become the identifiers used to classify an object, such as an email, a database, a printer, and thereby associate a set of rules with that object.

The data classification policy specifies how information is labeled or marked to indicate its classification. Database designs must include the data classification type for each column of data, so the application can appropriately prevent information which is higher than an individual's classification from being displayed. This is commonly referred to as "need to know," but is also called the Principle of Least Privilege. If the user does not need access to the information to perform their assigned job functions, then do not give them access to that information.

The scope or magnitude of risk for a given information asset is influenced by its classification type. The higher the data classification, the more tightly that piece of information must be controlled. Organizations must be careful not to over-classify documents. For example, if everything the organization produces is classified as top-secret, then either very few people in the organization have access to the information, or extreme measures are taken to protect that information.

Over-classification is a major concern especially with recent publications from WikiLeaks. Hackers and groups like WikiLeaks not only want the quantity of data, but they want data which will be worth something to them either financially, through media attention, or by bringing a wave of embarrassment down upon the targeted organization.

Calabrese (2012) identified significant increases in the number of classified documents produced by the United States government year over year. The more classified documents an organization produces, the more opportunities exist for individuals who may have appropriate and authorized access to intentionally or unintentionally leak that information to outside sources.

As part of defining the information classification types, the organization must decide what the community of interest is for each type. As the level of sensitivity increases, the data classification policy helps establish limits on who the information can be communicated to. If your organization has a thousand employees and every document is classified top secret, would that be considered reasonable? The more reasonable approach would be to establish all of the documents initially as

confidential, and then limit access to specific pieces of information based upon their importance to the organization.

For example, the public relations department may create and publish news releases announcing intended product directions or new products, but they would not release the associated business strategy documents, financials, or critical product details as those would be classified and access to them restricted. Some information which may be more critical to a business might include marketing plans, product development documentation, and financial strategies, all of which may be of interest to your competitors and affect the viability of the business.

Does having a data classification policy eliminate the likelihood of your information being taken, or provided to organizations like WikiLeaks? On its own, no it does not. The data classification policy is only part of a comprehensive information security program that defines how the data are classified, but requires applications and data repositories to understand the relationship between the user requesting the data and the data itself to determine if they should have access to it. If the user is only allowed to be working with confidential documents, then do not give them access to the top-secret five and 10-year business plans. Many of the WikiLeaks document publications could have been minimized if the "whistleblowers" had not already had access to the vast amount of information they leaked.

Application Development

Developing an application for use by Internet-based users, which is commonly deployed using a Web server and web browser interface, has a certain set of challenges for today's organization. The application must be easy to use, provide access to the information the user is requesting, and protect the access to that information by restricting it to only authorized individuals.

The application must be capable of dealing with bad input from users, and a whole litany of various attack methods applicable to the web application. When an organization performs its own application development, there are a number of controls that can help reduce the opportunity for vulnerabilities to be introduced into the code. These controls include

- Adoption of a Systems Development Life Cycle
- Establishing Code Requirements and Standards
- Implementing a Code Library for code reuse
- Performing both automated and manual code review and analysis

Having a consistent and structured systems development lifecycle (SDLC) can improve the quality of the application code, reduce application vulnerabilities, and improve time to production by decreasing the development, coding, testing, and correction cycle. By reducing the opportunity for vulnerabilities to exist in the application code, the organization can thereby reduce the opportunity for hackers to use the application and successfully gain unauthorized access to the organization's information.

There are several different system development life cycles in widespread use. The organization can adapt one or more of them to meet their specific needs. Aside from the structure associated with the SDLC regarding the application design, coding, implementation, and so on, the SDLC also provides a series of "gates" or points where the application must demonstrate conformity to the defined development requirements.

During those "gates," where it is determined if the application development process should move on to the next phase, the organization can include a review of the security controls implemented in

the application and the associated process designs for using the application. If the controls previous identified have not been implemented, then the application does not get to move forward in the development cycle. If it is determined that a new control must be added because of an event which affects the application security, then that control must be added into the development cycle.

While the SDLC provides a high degree of structure, many smaller organizations do not want to follow one because they do incur cost. They can slow down development time because of the emphasis on identifying and meeting requirements. The problem is that the cost of fixing a bug after the application is in production is much higher than fixing it during the initial application development cycle. While the smaller organization may boast about being able to move the application into production faster, they may also be exposing themselves to vulnerabilities or ways around the system because of the lack of planning, coding, implementation errors, or misconfigurations in the production systems.

Application Code Development Standards

Coding standards are important to a development organization because they define the code structure and language constructs which are allowed and those which cannot be used, along with strict design and testing procedures. The objective of these standards is to ensure that programmers in the organization produce code which is consistent in style, level of comments inserted, documentation produced to support the function, and sign off of the acceptance of the code.

Once the application code is reviewed, which includes comprehensive test scripts to verify the operation of the code, then the application code can be signed off by the supervising software engineer for use in the application. At that point, the decision can be made to add the piece of code into the organization's code repository, allowing for it to be reused in a different project later.

The code development standards should include elements such as coding practices on a per application language basis, strict design and testing procedures, code reviews, and the use of both automated and manual testing tools such as application and code scanning and analysis, although many of these things serve a larger purpose than just the development of the application code.

Having a defined set of code practices for your application development serves several functions. They establish a common set of requirements that programmers must follow when developing the application. This includes the use of comments in the code, specific language functions, and how specific problems must be solved and implemented within the application. For example, the coding practice would require use of input validation routines for any data which is supplied by the user before it is processed by the application and stored in any form of data repository. Additionally, code development standards could prescribe specifically how the input validation issue will be addressed. Once the input validation code is written and approved, it can be signed off and added to a code repository to be used again when someone needs an input validation routine.

Code Reuse

Code reuse is a technique, which is often used to prevent past problems from occurring in new applications. In a code reuse situation, programmers can make use of previously approved functions, such as input validation or connecting to a database, from a set of standardized functions of the organization. In this case, the functions will have previously been reviewed and evaluated to ensure application code is secure, and as efficient as possible.

A side benefit from code reuse is shortening development time. Because past problems have already been solved, the programmer does not need to spend the time solving the same problem again. They can search the code repository, find the function, and pull it into their application code. If the function needs modification, it can be modified, tested, reviewed, and updated in the code repository for future use.

Application Code Review, Testing, and Analysis

As part of the testing requirements, the functions and applications must undergo a code review from a team of experienced programmers in that specific application programming language. The team of reviewers is established so the programmer cannot review and approve their own code as being efficient, free from bugs or defects and vulnerabilities.

Additionally, automated and manual code analysis can identify problems in the code before it is placed into production and correct it while still in development. Sometimes, code changes later can have a ripple effect through the rest of the application code and impact the entire application. These are costly errors to correct later. Performing this analysis and verification work provides the opportunity for the development team to correct the coding deficiencies and reduce the opportunity for vulnerabilities in the production application.

Automated code analysis tools can perform both static and dynamic code analysis of the application. Static analysis involves a review of the actual application code and only the code. Data inputs and outputs are not evaluated. Dynamic evaluation occurs when the application is running and allows the testing tool to see what data are being moved into and out of a function, if data are being correctly checked, and so forth.

The application testing and analysis do not just stop at the application code. If the application is web based, then using web-based analysis tools to "walk" through the various features and items in the application can also identify errors where the user could manipulate information in the URL to see a different piece of information or retrieve configuration files used by the application. Implementing the web application to protect the "innards" of the application while still allowing it to function within the restricted capabilities of the Web server permissions can be difficult.

However, automated testing tools, whether used for code analysis or application vulnerability analysis, have their weaknesses. Just like computer-based dictation systems, which do not have the ability to understand context, automated analysis tools cannot understand the context in a series of application language instructions. Consequently, they can identify problems where no problem actually exists, commonly known as a false positive. Therefore, all results from automated analysis tools must be carefully reviewed to determine the actual issues, which must be corrected before the application goes into production and that are false positives and can be safely ignored.

Gaining unauthorized access to the information belonging to the organization is a proven method used by hackers such as Anonymous and LulzSec to obtain information which they have either published, sold, or used for their own gain. Therefore, by enhancing the organization's ability to develop secure applications, it is possible to reduce the opportunity for unauthorized access to your application and therefore your information.

Application Authentication Standards

Ever since computers were first developed, the problem of preventing unauthorized access or restricting users to a specific set of commands has existed. Users create bad passwords: this is a

known and well-documented fact. Entire companies, like RSA, have been built around trying to find better, more secure methods of authenticating a user to a system.

More often than not, web sites are hacked because someone used a bad password or the password file was easily retrieved from the application by using the Web server. Password should never be stored in clear text in an application. They must always be stored using encryption, even if that makes the process of registering the user, authenticating the user, and recovering a lost password more difficult. These issues are well understood not only by the hackers but by the information security community and there is no excuse for it to continue to occur.

By developing a standard for application authentication, the organization is providing the requirements for a programmer to implement an authentication system. Ideally, there would be one authentication system the application verifies the username and password provided by the user with. Because the authentication system is external from the application, the application itself has no knowledge of the user names and passwords—only the authentication system does. The centralized authentication model also allows more controls to be included to protect the passwords in storage, verify them over an encrypted link, and prevent the Web server from being able to "see" the password file and provide it to a malicious user.

Additionally, the process for recovering a lost password can be centralized, along with changing a password and specific password length and complexity requirements. Password management is an area many hackers take advantage of in their attempt to gain access to the application.

Centralized authentication systems are not without their risks though. By using a centralized system, the hacker need only to compromise the authentication system to potentially be able to get a list of usernames, encrypted passwords, and so on. That being said, more effort can be placed on protecting the authentication inside the private corporate network and leave the publicly accessible application outside the perimeter. There are a great many concerns when designing applications which must cross the public/private network boundary, which are beyond the scope of this chapter.

Computer and Network Security

The organization should put some effort into making the computer systems running their applications and the networks used to transport their data as secure as possible. There are many ways to do this, such as firewalls, limiting bandwidth, restricting protocol usage, and the like. The network and computer security controls can be evaluated using a variety of analysis tools such as host or network intrusion detection systems port scanner, and so forth. Like anything else, the control implements should form a layer, where each layer must be compromised in order to get to the next higher layer.

Protecting today's computer systems and networks is just as challenging as it was 30 years ago. In today's organizations, IT security teams must not only deal with corporately owned computing assets, but they also must deal with computer systems and mobile devices. These noncorporate computing systems create new challenges because of the associated risks they introduce with not having the corporately provided security controls.

Implementing a system analysis tool which evaluates a computer system for approved antivirus and other security software when attached to the network can ensure that corporately owned computer systems are protected and meet the minimum requirements established by policy. Many organizations have established strict policies which prohibit the use of individually owned computer systems on the corporate network, although some organizations choose to allow these devices.

The actual security of the network and computer systems, or lack thereof, has also been a significant factor in the success of hacktivist attacks against organizations. Even if you take all the

steps to develop security in your application, the data can still be compromised if your computer system or network is not properly secured.

Conclusion

Major hacktivism groups like Anonymous are structured for problems. With Anonymous specifically, there is no leader. There is no single person or group of people within the organization who set the tone of the message decides on who their next victim will be. They are a group of individuals acting as individuals under the collective umbrella of the organization. This means what is one man's cause does not mean it is another's cause within the same group.

Even though some of these groups, Anonymous and WikiLeaks in particular, may have trouble ahead, the 2012 Verizon Data Breach Investigation Report cites hacktivists like Anonymous as responsible for 58% of the data breaches occurring in 2011 (Verizon, 2012).

And while this chapter was being completed, The Associated Press published a report that WikiLeaks was in the process of distributing more than 2 million emails, allegedly from Syrian government email accounts (9News, 2012).

Unfortunately with Anonymous, the media attention has exacerbated the situation by calling a lot of attention to the group, which further fuels their drive for bigger and more important targets. Dan Brown in his Bit9 blog raised an important question about hacktivist groups, "A new day for Anonymous is rising and it's a day when the group itself may be unable to control its own message" (Brown, 2012). This statement strongly suggests that the leaderless group may start to have trouble at some point with delivering the message, because there is no cohesive perspective. While many people may disagree with WikiLeaks founder Julian Assange's perspective on acquiring and publishing secret government documents, WikiLeaks does have a focus, unlike some hacktivist groups.

Should services like YouTube, Facebook, and the like be held responsible for removing potentially sensitive or illegally obtained information when it has been identified to them? While many Internet users might like that to happen to protect the safety of the people who defend our rights, freedoms and the law, these services may be slow to respond without some form of protection because of the potential for a challenge to free speech, as exists in the United States. The Middle East and China, for example, have both felt the effects of trying to censor the information available to their Internet users. The need for protection to make the removal of this illegal content is significant to protect the service from any potential legal action. However, that protection does not extend to the impacts from the various hacking or hacktivism groups.

While hacking and hacktivism pose a threat to information security and protecting information that needs protecting, hacktivism is as much a sociological phenomenon as it is a legal one. Changing the minds of the hacktivists is not going to happen. Perhaps, we need to bring awareness to the rest of the population into the real impact of these groups, the economic impact, the threat to our various nations and do so without the media hype that glamorizes their activities.

Unless the reward for their activity is removed, hacktivism will unfortunately always be there.

References

9News. 2012. WikiLeaks has data from 2.4 million Syrian emails. *9News*. Published July 5, 2012. Retrieved online July 5, 2012 from http://www.9news.com/news/world/276094/347/WikiLeaks-has-data-from-24-million-Syrian-emails

Ball, J. 2012. WikiLeaks publishes Stratfor emails linked to Anonymous attack. *The Guardian*. Published February 26, 2012. Retrieved online June 28, 2012 from http://www.guardian.co.uk/media/2012/feb/27/wikileaks-publishes-stratfor-emails-anonymous

Bland, S. 2010. Julian Assange: The hacker who created WikiLeaks. *The Christian Science Monitor*. Published July 25, 2010. Retrieved online July 3, 2012 from http://www.csmonitor.com/USA/Military/2010/0726/Julian-Assange-the-hacker-who-created-WikiLeaks

Bluestein, A. 2012. Gabriella Coleman: Helping hackers infiltrate academia. *Fast Company*. Published June 21, 2012. Retrieved online June 26, 2012 from http://www.fastcompany.com/1841040/anthropologist-studies-hacking-and-digital-activism

Brito, J. 2011. 'We do it for the Lulz': What makes LulzSec tick? *TIME*. Published June 17, 2011. Retrieved online July 1, 2012 from http://techland.time.com/2011/06/17/we-do-it-for-the-lulz-what-makes-lulzsec-tick/

Brown, D. 2012. Anonymous: Why the media is getting it wrong. *Bit9*. Published February 9, 2012. Retrieved online June 24, 2012 from https://www.bit9.com/blog/2012/02/09/anonymous-why-the-media-is-getting-it-wrong/

Calabresi, M. 2012. WikiLeaks' war on secrecy: Truth's consequences. *TIME*. Published December 2, 2010. Retrieved online July 2, 2012 from http://www.time.com/time/magazine/article/0,9171,2034488,00.html

Cilley, J. 2012. Survey rates Anonymous top threat: Here's 4 reasons why Anonymous won't hack you [INFOGRAPHIC]. *Bit9*. Published April 25, 2012. Retrieved online June 27, 2012 from https://www.bit9.com/blog/2012/04/25/4-reasons-why-anonymous-wont-hack-you/

Cluley, G. 2011. The end of LulzSec? Hacking group says it is disbanding, after 50 days of attacks. *Nakedsecurity*. Published June 25, 2011. Retrieved online June 29, 2012 from http://nakedsecurity.sophos.com/2011/06/26/the-end-of-lulzsec-hacking-group-says-it-is-disbanding-after-50-days-of-attacks/

Coleman, G. 2010. The anthropology of hackers. *The Atlantic*. Published September 2010. Retrieved online June 29, 2012 from http://www.theatlantic.com/technology/archive/2010/09/the-anthropology-of-hackers/63308/

Dubai. 2011. Arab social media report. *Dubai School of Government*. Published May 2011. Retrieved online June 26, 2012 from http://jrnetsolserver.shorensteincente.netdna-cdn.com/wp-content/uploads/2011/08/DSG_Arab_Social_Media_Report_No_2.pdf

Gardham, D. 2012. "Anonymous" hackers intercept conversation between FBI and Scotland Yard on how to deal with hackers. *The Telegraph*. Published February 3, 2012. Retrieved online on June 28, 2012 from http://www.telegraph.co.uk/technology/news/9059580/Anonymous-hackers-intercept-conversation-between-FBI-and-Scotland-Yard-on-how-to-deal-with-hackers.html

Gellman, B. 2012. The 100 most influential people in the world 2012. *Time*. Published April 18, 2012. Retrieved online on June 24, 2012 from http://www.time.com/time/specials/packages/article/0,28804,2111975_2111976_2112122,00.html

Gurman, H. 2011. Telling the story of WikiLeaks. *Foreign Policy in Focus*. Published February 28, 2011. Retrieved online July 3, 2012 from http://www.fpif.org/articles/telling_the_story_of_wikileaks

Hacking News. n.d. Subscriber info. *Stratfor Global Intelligence*. n.d. Retrieved online June 29, 2012 from http://www.stratfor.com/hacking-news

Hacktivism. n.d. Hacktivism: Dangerous new social conscience retrieved online April 11, 2012 from http://www.itweb.co.za/index.php?option=com_content&view=article&id=52863

IBTIME. 2011. LulzSec wins hacking competition, refuses $10 k award. *International Business Times*. Published June 8, 2011. Retrieved online July 1, 2012 from http://www.ibtimes.com/articles/159446/20110608/lulzsec-hacking-competition-black-berg-cybersecurity.htm

Internet Population. n.d. Internet world stats. Retrieved online June 18, 2012 from http://www.internetworldstats.com/stats.htm

Jevans, D. 2011. Lulzsec, Anonymous and why hacktivism is going to keep increasing. *Privacy and Identify Theft Blog*. Published June 27, 2011. Retrieved online June 18, 2012 from http://blog.ironkey.com/?p=1285

Knafo, S. 2012a. Anonymous and the war over the Internet. *The Huffington Post*. Published January 30, 2012. Retrieved online June 30, 2012 from http://www.huffingtonpost.com/2012/01/30/anonymous-internet-war_n_1233977.html

Knafo, S. 2012b. Anonymous and the war over the Internet (Part II). *The Huffington Post*. Published January 30, 2012. Retrieved online June 30, 2012 from http://www.huffingtonpost.com/2012/01/31/anonymous-war-over-internet_n_1237058.html

Lynley, M. 2011. Hit the deck: LulzSec and Anonymous start trading blows. *VentureBeat*. Published June 15, 2011. Retrieved online July 2, 2012 from http://venturebeat.com/2011/06/15/lulzsec-anonymous-civil-war/

MacManus, R. 2010. Digital activism: An interview with Mary Joyce. *Read. Write. Web*. Published March 8, 2010. Retrieved online June 25, 2012 from http://www.readwriteweb.com/archives/digital_activism_an_interview_with_mary_joyce.php

Marketou, J. 2001. Hacking vs hacktivism: Sniffing the net. *Mujeres en Red. El periódico feminista*. Published August 2001. Retrieved online June 27, 2012 from http://www.nodo50.org/mujeresred/spip.php?article1535

Miercoles. 2012. War reporting: From newsreels to real-time war. *International Journalism*. Published February 8, 2012. Retrieved online June 26, 2012 from http://martajiminternationaljournalism.blogspot.com/2012/02/war-media-coverage-has-been-changing.html

Plafke, J. 2011. Anonymous hacks Spanish police's website after Spanish police arrest 3 Anonymous members. *Geekosystem*. Published June 13, 2011. Retrieved online June 30, 2012 from http://www.geekosystem.com/anonymous-hacks-spanish-police-website/.

Ranson, J. 2012. Hacktivist's mom has a few questions Anonymous. *TownHall.Com*. Published March 15, 2012. Retrieved online June 24, 2012 from http://www.freerepublic.com/focus/news/2859730/posts

Sengupta, S. 2012. The soul of the new hacktivist. *The New York Times*. Published March 17, 2012. Retrieved online, June 6, 2012 from http://www.nytimes.com/2012/03/18/sunday-review/the-soul-of-the-new-hacktivist.html

Stallman, R. n.d. Why software should not have owners. *The GNU Project*. n.d. Retrieved online June 26, 2012 from http://www.gnu.org/philosophy/why-free.html

Stratfor. n.d. About us. *Stratfor Global Intelligence*. n.d. Retrieved online June 29, 2012 from http://www.stratfor.com/about-us.

Verizon. 2012. The 2012 data breach investigations report. *Verizon Business*. Published April 2012. Retrieved online July 2, 2012 from http://www.verizonbusiness.com/resources/reports/rp_data-breach-investigations-report-2012_en_xg.pdf

Whittaker, Z. 2011. WikiLeaks: A brief history pre-2010. *ZDNet*. Published June 20, 2011. Retrieved online July 2, 2012 from http://www.zdnet.com/blog/igeneration/wikileaks-a-brief-history-pre-2010/10750

Wikipedia. 2012a. Electronic civil disobedience. *Wikipedia*. Published May 11, 2012. Retrieved online June 27, 2012 from http://en.wikipedia.org/wiki/Electronic_civil_disobedience

Wikipedia. 2012b. WikiLeaks. *Wikipedia*. Publish date unknown. Last updated, July 1, 2012. Retrieved online July 2, 2012 from http://en.wikipedia.org/wiki/WikiLeaks

Compliance

Chapter 27

PCI Compliance*

Tyler Justin Speed

Contents

Virtually every organization that has something to sell allows patrons to use a credit card to make a payment or provide a donation. Without the capacity to process these credit cards, companies are crippled and cannot function in a profitable manner. Because of this fact, organizations are often entrusted with thousands of credit card numbers, as well as the customer data associated with

those numbers, such as names, phone numbers, addresses, and even (sometimes) social security numbers.

A multitude of security concerns surround the issues of processing and storing these cardholder data. Many people assume only the card numbers themselves must be protected to provide proper security; in fact, all of the personal data provided by the cardholder should be protected from unauthorized viewers. This requires thorough network security measures such as firewalls, DMZ (demilitarized zone) configurations for customer databases, security plans, and so forth. Not only are all of these rules a good idea, they are now required by a consortium of the big players in the credit card industry.

How can an IS professional determine whether his specific organization, department, or network is following industry standards? Where are these standards found? Recently, these questions were all answered by the Payment Card Industry Security Standards Council (PCI SSC), in the form of the Payment Card Industry Data Security Standard (PCI DSS). The PCI DSS provides a sort of framework within which organizations can build network security to a level that provides a solid level of security for cardholder data.

PCI Compliance

PCI compliance refers to an organization's adherence to the regulations set forth by MasterCard, Visa, Diner's Club, American Express, and Discover. Although the rules outlined by these different credit card companies were created in response to the ever-growing need for technological controls, the measures required for securing credit card data are sound principles for guarding any data. Because most organizations need to comply with these rules anyway, and because they are great security considerations regardless, it is important to be at least familiar with PCI compliance and what is required for different types of organizations.

Goal of PCI DSS

The goal of PCI compliance is to protect customer information from fraud and theft through the implementation of commonsense applications of security and data storage. Because PCI is primarily concerned with protecting cardholder data, the documentation for PCI compliance often refers to the "cardholder data environment." This broad term simply refers to any computer server or other aspect of the network wherein cardholder data are stored. Keep in mind that credit card companies provide protection to cardholders for fraud and stolen card numbers. It therefore behooves credit card issuers such as Visa and MasterCard to require organizations to protect cardholder data; again, just because PCI compliance is concerned with providing thorough protection for cardholder information, the rules provided for protection are an excellent framework for use in the pursuit of total network protection.

Who Must Adhere to PCI Compliance?

The bottom line is that if your organization accepts credit cards for payment of any services or products, it must comply with the rules outlined in PCI DSS to maintain PCI compliance. If your organization has Internet-facing IP addresses that collect or transmit cardholder information, your network must have a security scan performed annually. Any path from the Internet into your cardholder data environment is subject to a security scan, or audit. Even if your organization does

not have such IP addresses or apparent paths, less apparent pathways almost certainly exist within your network infrastructure. For instance, a savvy cyber criminal could use your organization's e-mail functions, a virtual private network through which employees have external access to an organization's internal network, or any other possible path to gain unauthorized access to your customers' cardholder information. Be sure to include in a security overview all possible routes to the cardholder data environment, and include a list of all measures taken to protect those paths.

Who Is Authorized to Perform PCI Security Scans?

The Qualified Security Assessor (QSA) companies that have been approved by the PCI Security Council to have certain, certified employees of the organization validate the organization's adherence to PCI DSS. The organization in question does not certify these employees, but the employees must be approved and certified through the PCI Security Standards Council. The certification of both the QSA companies and their QSA qualified employees must be renewed on an annual basis. The PCI Security Council provides self-assessments to companies desiring to know if they are in compliance with PCI DSS, but to be certified as in compliance, the organization in question must receive validation from either an in-house QSA employee or a third-party QSA-certified individual.

The PCI Security Council maintains an up-to-date list of QSAs on the web at https://www.pcisecuritystandards.org/qsa_lookup/index.html. In addition to the Self-Assessment Questionnaires, the website provides users with many other informative PCI documents, including PCI FAQs, QSA requirements, PCI DSS overviews, and more.

The Five Levels of PCI Compliance

The PCI Security Council recognizes the fact that not all organizations have the same level of exposure to customer cardholder information. For instance, some companies may only have all their customer cardholder information handled by a third-party vendor. Others may have a point-of-sale (POS) machine through which the organization swipes a customer card directly connected to the Internet, with no sensitive cardholder information stored on company servers. Still yet, other organizations may not only physically handle the actual credit cards at the POS, but also maintain an up-to-date database of all cardholder information, in which hundreds of thousands of cardholder data records are stored.

It is clear to see that varying levels of security are required for different organizations depending on how many cards the organization handles, and what the organization does with the information, such as storage, processing, and transmission. On the basis of these different levels of involvement with cardholder data, the PCI Security Council has come up with five distinct levels of compliance.

Level 1 Compliance

This level of PCI DSS compliance refers to organizations that handle all cardholder data through a third-party vendor. In these cases, the card is not actually physically present for the transaction. For instance, mail orders, telephone orders, and e-commerce are all transactions wherein the card is not physically handled by an organization taking payment through a credit card. Since the organization is relying on a third party to process the card number and keep it in a database off site, the organization does not need to follow any network-specific security measures for Level 1 Compliance.

If your organization uses a third-party vendor to process all the cardholder data, the organization must take proper steps to ensure the company handling that information is appropriately certified by the PCI Security Council. In other words, just because the organization pays a company to do the work, the organization has not completely relegated the responsibility of security to the card processing company. Never, never, never forget that the bottom line, in regard to security of the information your organization handles, is up to your organization. Be vigilant in your pursuit of security. Do not leave it to chance!

Level 2 Compliance

Level 2 of the PCI DSS compliance is in reference to organizations whose only source of obtaining cardholder information is through the use of a credit card imprint machine. Although most customers only see these now-antiquated machines pulled out of some dusty and forgotten drawer when electricity is unavailable, or when a computer system crashes at a retail outlet, there are still some organizations whose sole method of storing cardholder data is on these physical receipts. Clearly, the requirements for digital security will be much less stringent for a company falling into this category; however, even though the network security needs will be diminished, the requirement for keeping the physical records safe from unauthorized eyes is not. Proper, secure file storage is required to keep unauthorized people from gaining access to cardholder data, including all of the steps necessary to maintain completely protected files.

Level 3 Compliance

The third level of PCI DSS requirements is identical to the requirements of Level 2. This is because as in Level 2, there are no digitally stored records. Level 3 refers to those organizations whose POS machine is a direct-dial machine, using a telephone line. These machines do not communicate with the organization's network, and therefore do not create any digitally stored cardholder data records to protect through network security, just as with Level 2. Instead, these machines create printed receipts for processing with bank deposits. Like the requirements for proper security as listed in Level 2, compliance with Level 3 does not mean an organization can relax security measures, but must instead strengthen the physical security protecting the stored printed receipts.

Level 4 Compliance

The fourth level of PCI DSS compliance, in regard to cardholder data processing, transmission, and storage, has to do with organizations that process cards through the use of a POS directly connected to the Internet. Although the POS sends data via the Internet to the bank, the cardholder data are not actually stored on company servers. Even though the data are not stored on servers, digital security is paramount, because cardholder information will be processed and transmitted through the virtual wires of the web. In this level, proper encryption of data, and secure firewalls and network security measures are required to properly comply with PCI Security Council measures.

Level 5 Compliance

For all organizations not falling within the definitions of the other four levels, Level 5 requirements for compliance are required. These requirements are the most stringent of all levels, and are outlined below. Level 5 compliance assumes an organization is taking and processing credit cards

internally. In addition, Level 5 compliance is for organizations that store credit cards on internally protected databases. Furthermore, Level 5 compliance assumes the organization's network has multiple points of Internet access. Therefore, the requirements attempt to cover all of the digital bases, as it were.

PCI DSS Overview

Twelve key requirements, listed under six different categories, are included in the rules for PCI compliance. These requirements are listed below. Notice that the requirements are somewhat open-ended, although specific enough for an IT specialist to have confidence that she/he has properly implemented the necessary controls and processes to be in compliance. PCI compliance is a massive subject, and this chapter only gives an overview of the requirements. However, keep in mind that there are minutiae and subrequirements that are not listed, due to the massive scale of the subject. This chapter is only meant to be a simple introduction, not a replacement for a thorough study of PCI compliance and how it directly affects and relates to your specific network.

Category 1: Protect and Maintain a Secure Network

The category to "protect and maintain a secure network" may seem a bit too vague, but as we show in the requirements listed below, what is necessary to meet the needs of this category has mostly to do with firewalls and router settings. To begin with, the network must have some sort of secure firewall installed to protect cardholder data. The requirement does not specify whether the data need to be protected from internal or external issues, just that the data need to be properly protected. This is important to keep in mind, as many networks have no internal firewalls protecting critical data from internal attacks. Be sure to consider protection from internal attacks when designing your network and providing data security for your databases.

Requirement 1.1—Install and Maintain a Firewall Configuration to Protect Cardholder Data. Installation of a firewall is a straightforward concept, but what is required for proper and thorough maintenance? The maintenance portion of this requirement is really more concerned with awareness of the current state of the network configuration, and proper adherence to strictly documented requirements. For instance, a current diagram of network connections, topology, and systems should always be kept on file. Any changes performed to the network or the systems therein must always be clearly and completely documented in the record.

Not only should the record be documented, but so should the processes directing the maintenance of the network. For instance, firewalls are required at specific points in the network, such as at every point where the internal network has a connection with the Internet. The process of how these connections and firewall settings are reviewed and maintained must be documented, as well as the business justifications for the use of each protocol, port, and so forth. For instance, it is not enough just to document that port X is open because Christine and Jackie like to play Doom during their lunch breaks. Unless a port is specifically required for legitimate business purposes, or a certain protocol such as HTTP, SMTP, FTP, or VPN is necessary for an organization to function, then it should be turned off or otherwise blocked from use.

Even once a business justification is determined, simply having the protocol or port documented is not enough for proper PCI compliance. Indeed, the organization must not only list what protocols are in use (and why), but also what security measures are in place to protect against intrusions. For instance, let us take the most commonly used protocol, HyperText Transfer Protocol

(HTTP). HTTP is used for accessing and sending web pages on the Internet. Justifying the use of this protocol is not a problem, as any organization desiring relevance must have some sort of access to or presence on the Internet. The problem is that not every website on the Internet is necessary to access. In fact, your organization will undoubtedly want to block access to the majority of the Internet for a large percentage of your users. It would be useless if Jackie and Christine's computers had the X port blocked so that they could not play Doom during work hours, only to have them able to access Doom through an online website.

The answer to this quandary is a properly configured firewall. Through either packet filtering or proxy filtering, the communication between Christine and Jackie's computers and the Doom server will be cut short at the chokepoint of the firewall. The bottom line is that by keeping activity limited to those actions required by daily operation of the organization, while establishing and maintaining secure procedures and methods (such as having securely configured firewalls) for those activities, the critical data, such as cardholder information, are properly protected.

Requirement 1.2—Build a Firewall Configuration That Restricts Connections between Untrusted Networks and Any System Components in the Cardholder Data Environment. This requirement is very similar to 1.1, except that it specifically speaks about filtering and restricting data to and from the cardholder data environment, with additional requirements for router synchronization and specific configuration settings. Synching routers is a common practice, and can be accomplished by using certain network time protocols, such as NTP. Synchronizing routers means each router has been synched to the same exact date and time. This is important for proper troubleshooting and diagnostics. Ideally, each system, including switches, routers, computers, and all components of the network will be synched. This will allow technicians to more accurately determine when systems crashed, and what the possible causes of certain traffic and performance issues are.

Requirement 1.3—Prohibit Direct Public Access between the Internet and Any System in the Cardholder Data Environment. The creation of a demarcation zone, or DMZ, is necessary for secure separation of internal networks and the Internet. DMZs are also sometimes referred to as a perimeter network. The DMZ is a sort of subnetwork that sits between the internal secure network and the outside, less secure network such as the Internet. The DMZ is usually created by implementing two firewalls on either side (internal side and external side) of a point in the network. For instance, there is a data point through which all data must flow to get out of the internal network. This is the internal side of the DMZ, and the router placed at that location should not allow unauthorized traffic to pass through, which could access unauthorized software or websites on the Internet. Such access could cause harm to the services located in the DMZ, or even on the more sensitive internal network. Likewise, on the internal side of the DMZ, there should be a firewall protecting the DMZ and internal network from the incoming traffic from the Internet.

Both firewalls are protecting from traffic originating from their respective sides, however, they should also be configured as though the other firewall did not exist. That is to say, just because an external firewall exists on the external side of a DMZ, the internal firewall should not just protect the DMZ from outgoing traffic originating from the internal network, but also from traffic attempting to access the internal network. It is also important to make certain that the two firewalls are not configured the same.

In fact, it is wise to configure both firewalls differently with the intent of providing a layered approach to protecting the internal network. In other words, just because an attacker breaches the first firewall with certain tricks and tools, he will not necessarily be successful in his attempts to break into a dissimilarly configured firewall on the other side of the DMZ. Dissimilar configuration schemes running on the DMZ firewalls mean attackers will need a whole new set of skills to break through an internal firewall. On the other hand, if both firewalls are similarly configured,

all somebody has to do is to get through the first firewall, and the keys to the internal network are his.

The services needed for accessing the larger, less secure network are also usually contained in the DMZ. For instance, domain name servers (DNS), mail servers, file transfer protocol (FTP) servers, and VoIP servers are usually located in the DMZ. Notice that only servers are located in the DMZ, as no individual node (user computer) should ever be placed outside the protected environment of the internal network. Also notice that database servers, such as those used for storing employee, customer, and cardholder information, are never ever located in the DMZ. This is again to protect the data contained within the databases from the less secure, external Internet.

In addition to the protection provided by the creation of a well-designed DMZ, PCI requires organizations to protect the internal infrastructure of a network by hiding internal IP addresses. This is accomplished through the use of what is called "IP masquerading." Essentially, IP masquerading is a process that allows internally connected computers without specific IP addresses to access the Internet via another computer. This is commonly practiced in most work environments, and allows system administrators to monitor traffic going in and out of that single point.

Requirement 1.4—Install Personal Firewall Software on Any Mobile or Employee-Owned Computers with Direct Connectivity to the Internet (Laptops Used by Employees, etc.) That Are Used to Access the Organization's Network. Most corporations employ individuals who require access to the company website through mobile devices such as laptops and personal cell phones. Because these devices are not always attached to the network, they must be specially protected from malware and attackers. This is done through specifically installing software firewalls on each device to ensure the organization's network is protected when the equipment is installed on the network.

For instance, the computer that this book was written on is a personal laptop. I use it to build websites, write e-mails, create small programs, access social networking sites, play online games, and so forth during my personal time. The company I work for, Electronics International, allows me to use this computer during work hours. What I do with my laptop on my own time is my business, but it could severely affect the security of Electronics International's business once I hook into the company network. Because I am aware of the dangers of plugging a personal laptop into an organization's network, I have taken great pains to protect company data. I have installed a firewall on my computer, created backups of all company-related files, and installed multiple antivirus and antispyware applications on my laptop. I also never place customer, employee, or cardholder information on my laptop, just in case all of the above measures fail, inasmuch as no security measure is perfect.

Category 2: Protect Cardholder Data

Of course, the goal of all the PCI compliance issues is ultimately to provide for the protection of cardholder data. Category 1 was focused on network security in general, and Category 2 gets more specific and places requirements on systems instead of the network as a whole.

Requirement 2.1—Always Change Vendor-Supplied Defaults Before Installing a System on the Network. This requirement states that any installed network equipment should not use the vendor default settings. For instance, many commercial off-the-shelf (COTS) routers use the name "admin" as the default name, and "password" as the password. This type of default setting is true for virtually all routers and their default settings. It would require far more time during manufacturing and installation of the router to have a unique name and password assigned to each router. Anyway, the manufacturer assumes that the first thing an installer will do is change the name and password, just as this requirement states.

The "name" of a router is referred to as the service set identifier, or SSID. If you are in an Internet café, or at home looking for your wireless router, this is the name of your router as seen on your computer. There will be multiple SSIDs available if there are multiple routers broadcasting in the same area. So if somebody asks about the default SSID for a router, you can bet it is probably the model name of the router.

What makes these default settings especially easy for cyber criminals to locate on the Internet is the fact that the default name of the router *is* usually the model of router, such as "Linksys WRTG54." All the cyber criminal needs to do is intelligently search for routers with default names and he will have easy access to them. If somebody has not taken the time to change the default SSID, the chances are good he has also left the router's password the same as well. Some cyber criminals will literally drive around neighborhoods looking for an SSID with a default name. Once they find one, they will try to hack into the system. Many times, they simply want free access to the Internet, although they may also try to access any computer attached to the network. The solution to this issue is far too easy and cost effective not to be implemented. Just log in, change the password, and implement encryption. The bottom line is that a network router should not be allowed online without encryption, a new username, SSID, and password.

Requirement 2.2—Develop Configuration Standards for All System Components That Address All Known Security Issues and Are in Keeping with Industry-Accepted Methods for System Hardening. The basics of this requirement boil down to eliminating unnecessary protocols and services that could be compromised. In other words, if a protocol or service is not being used, disable it. Also, do not just eliminate unnecessary protocols, but protect those protocols and services being used by providing dedicated servers and associated protection for the ones in use. For instance, if it is possible, do not have an e-mail server to also provide domain name resolution. This is a high-horsepower approach to protecting the DMZ, as it requires multiple servers, so smaller companies may not be able to implement this step. Keep in mind that these are PCI requirements, which may not necessarily be the most cost-effective solution for overall security in your particular organization. A PCI Compliance QSA may be able to help your organization meet the requirements of PCI compliance and be in harmony with your organization's budgetary and labor constraints.

Requirement 2.3—Encrypt All Nonconsole Administrative Access. Nonconsole access refers to any remote administrative access. This just means the administrator is not sitting at a computer directly attached to the internal network. This requirement is one of the most important, because it assumes administrators may need to access organization computers through the Internet while not physically present. Before transmission of access codes, passwords, or other login data, the connection must be secure and encrypted. It would be easy for an attacker to grab an unencrypted data stream from the Internet and use it to gain unlimited access to an organization's network. The bottom line for this requirement is to make certain critical and secure information is not simply shot out into the Internet without being properly hidden by an encryption scheme.

Requirement 2.4—Shared Hosting Providers Must Protect Each Entity's Hosted Environment and Cardholder Data. Many companies simply do not have the in-house talent or resources to host their own website, and even some larger companies use online databases such as customer resource management (CRM) applications as part of their daily operations. Although online services often provide viable, cost-effective solutions for smaller companies, they introduce a whole new set of considerations in regard to security. Just because the provider is not in-house does not mean he does not have to abide by in-house rules. This requirement simply states that if an organization is using a shared hosting provider, that provider must maintain unique security measures for each website or service hosted.

For instance, if a provider hosts 100 websites, it cannot simply provide a firewall generic to all 100, but must delineate each service to provide proper protection as outlined in the PCI requirements. This is not because PCI wants to make life more difficult or complex for the provider or the entity using the provider's service, but to protect cardholder data from all sources of potential threat. A firewall may protect the 100 websites from the external Internet, but it does not protect them from one another. Therefore, when choosing a host to provide service for your website or servers, make certain they have protection in place that will meet the requirements of your organization's security plan, including separate firewalling between vendors being hosted by their company.

Category 3: Maintain a Vulnerability Management Program

After securing the networks, including setting up firewalls, switches, and other network devices according to PCI requirements, it is important to recognize and manage new and existing threats to the network. A vulnerability management program will help accomplish this goal.

Requirement 3.1—Keep Cardholder Data Storage to a Minimum. It used to be that many companies would keep all customer-related, legal, and financial records for up to 7 years for tax purposes. Instead of the old method of holding onto cardholder data, Requirement 3.1 states as far as cardholder information is concerned, get rid of it as soon as your organization has no use for it. If your organization has a 90-day charge-back guarantee, there is no need to keep the credit card data for more than 90 days. Even at that, make sure most of the card numbers are encrypted and that customer service employees can only view the last 4 digits for verification purposes.

In short, make certain your organization only holds onto cardholder data for the length of time required by industry regulations, necessity, and law. After that, properly destroy that information. If it is on paper, have a certified paper disposal company take the shredded files for removal. More likely, the data will be on disk. PCI requirements include provision to dispose of outdated cardholder information at least once a quarter. This process should be at least partly automated, so that it relies on nobody's memory to go through and eliminate the data. In other words, a program of some sort should be set up to take care of the elimination. This could be done by moving outdated files to a storage location once they have expired, where a program will automatically delete them at the quarter's end. Physically stored files should likewise be moved at certain intervals, and then shredded and disposed of properly at the end of each quarter.

Requirement 3.2—Do Not Store Sensitive Authentication Information after Authentication. Once a card is processed, the information needed to complete the transaction in the first place (i.e., card number, cardholder information, service code, primary account number [PAN], etc.) should only be stored as required by business. Oftentimes, this means that only the cardholder's name, address, phone number, and the last four digits of the card are stored in association with purchase orders and invoices. If the information is extraneous to the requirements of daily business, do not store it. If a cardholder needs to change an order, he can always provide customer service with the credit card number again.

Requirement 3.3—Mask PAN When Displayed. Whenever the primary account number is displayed on the screen of a user's computer, only the last four or six digits should be readable. The other six to eight digits should be properly encrypted so as to protect the card number from being deciphered. Keep in mind that stricter requirements of industry or company policy will override this requirement. For instance, if your organization's policy requires only the last two or three digits be in plaintext, then the stricter organization requirement will supersede PCI Requirement 3.3.

Requirement 3.4—Render PAN, at Minimum, Unreadable Wherever Stored. There are several methods that can be employed to render a file unreadable at a glance, such as truncation, encryption, and hash-marking; however, this requirement simply requires one of them to be utilized in protecting the cardholder information from being easily read by users. Truncation simply "cuts" the number off at a certain placeholder, such that only the last four to six digits are seen, eliminating the possibility that somebody could duplicate the card number. Encryption and hash-marking replace the actual digits of the account number with a coded alphanumeric symbol, so that the card number is illegible by normal standards. Whatever your organization decides upon, be sure customer card numbers are not readily viewable at a glance.

Category 4: Implement Strong Access Control Measures

Access control measures refer to controls that ensure only specific authorized users are allowed access to particular areas of data storage. When a user logs onto a protected network system, his credentials are authenticated through the use of various methods of authentication. Each user's account is assigned a certain level of access, which allows the user to access files with a matching level of security or classification. Once he has been authenticated and authorized to access the system, the user does not have to restate his login credentials or password until he attempts to access another system, or until he tries to access a level of authorized files higher than the one for which he is logged on.

Access control lists (ACLs) are stored in a database and contain all of the access control information, such as login names, passwords, authority levels, and expiration dates of login information for each user. Every user listed within the ACL should be given a unique user ID, to avoid duplication and confusion of access levels and user login information. ACLs should be maintained by administrators to ensure only those users who should have access are approved. For instance, when an employee is terminated from working at the organization, her authority to access the system should be eliminated from the ACL database.

In addition to the ACL and the digital protection provided therein, access to cardholder data must also be physically restricted. For instance, even if an organization stores all cardholder data on secure database servers in a properly configured DMZ, any physical receipts, printouts, or imprints with cardholder data must be properly protected through the use of secured rooms and locking cabinets. Make sure the level of physical measures implemented is commensurate with the level of criticality of the information being protected. Cardholder information does not have to be protected on a level equal to Fort Knox, but it should be kept in locking metal cabinets that have been approved for such storage.

The bottom line when considering access control is to take all reasonable steps to ensure only those people with authorization are allowed access to cardholder information. Essentially, this is the primary concern of all PCI compliance measures. Everyone, from the casual snooper to the hardened cyber criminal should be dissuaded from attempting to gain access to the secured files. This does not mean spending millions of dollars on state-of-the-art technology, but instead means using sensible approaches to security, such as encryption, physical locks, appropriately configured firewalls, and so forth.

Category 5: Regularly Monitor and Test Networks

One of the best methods for locating issues in networks is through constant monitoring of network traffic. With a good network monitor, many issues can be discovered and fixed. After fixing

issues, the network can then be retested to determine if the fix was appropriate or if further tweaking is necessary. For instance, network monitoring can assist network administrators with locating such issues as

- *Network bottlenecks:* Bottlenecks affect the throughput performance of networks. Network performance is directly related to gaining immediate access to data on the networks. Access has everything to do with data availability, which is part of the CIA triad of information assurance.
- *Unauthorized software:* Unauthorized software can be malicious in nature or nonmalicious. Malicious software packages are immediate and clear dangers to your network security. Network monitors should be able to quickly detect and eliminate such software. Even the so-called nonmalicious unauthorized software should be immediately deleted, and the responsible individual should be spoken to, written up, or, if the incident was severe enough, fired. The organization should take a stance that there is no such thing as nonmalicious unauthorized software. Screensavers, font packages, and other nonmalicious software are sometimes just a package in which malicious software resides. This type of software is probably okay, but probably not okay is a good enough measure of security for any organization.
- *Testing traffic throughput speeds of specific portions of the network:* Throughput speeds affect access to data, and sluggish overworked networks could be indicative of larger issues occurring. Massive amounts of data flowing through the network without apparent cause usually point to the existence of malware communicating to sources outside the network.
- *Troubleshooting network issues:* Network monitors are great for troubleshooting. Because troubleshooting often involves removing, replacing, and testing individual systems, network monitors allow technicians to isolate suspected systems by monitoring their effect on the network as a whole.
- *Determining origins of incoming and outgoing malware:* By monitoring network packet traffic, technicians can pinpoint the origins of malware. Once IP addresses of malware are identified, administrators can block any traffic to and from those addresses.
- *Determining which users are currently on the network, and in what functions their workstations are engaged:* Another issue to keep in mind when considering network monitoring is data logging. Data logging provides a history of what traffic has flowed through the gateway to and from your network. If somebody has launched a virus onto your network, you will have a much greater chance of locating the culprit and isolating the infected systems if you have a history of network traffic on which to look back. Also, during legal cases involving criminal network activity, having a log to back up claims against disgruntled employees, competitors, or other cyber criminals adds weight to the validity of your case.

Category 6: Maintain an Information Security Policy

Having an information security policy is a key part of thoroughly protecting any network. Without a policy dictating how to handle specific issues, security is essentially an afterthought. The security plan should be detailed enough to deal with particular security concepts and still broad enough to cover all network functions. For example, a standard security plan would include the following:

- Specific network topology, including physical layout, number of users, number of servers, operating systems used, and so forth

- Division of authority for specific security-related tasks
- Router settings and configuration
- DMZ configuration for perimeter routers and firewalls
- Computer security response teams
- Security awareness training
- Computer incident postmortem procedures
- Computer forensics methodologies and law enforcement

Regardless of what level of PCI DSS compliance you are required to meet, your organization would benefit from a solid network security plan. Even the most robust security measures cannot ever provide an absolute guarantee of protection for your network or the data contained therein; however, the legal ramifications, not to mention the maintenance and security headaches associated with a lack of such a security plan, make it all but necessary for even the smallest networks and organizations to create and maintain some sort of security plan.

A Good Place to Start

PCI provides a good overview of network security, in plain English. The PCI requirements provide a good baseline of security for most organizations, even if the organizations do not take credit cards from customers. Regardless of whether these rules apply to your organization, they are worth researching and getting to know.

Chapter 28

HIPAA/HITECH Compliance Overview*

John J. Trinckes, Jr.

Contents

Interrelationship among Regulations, Policies, Standards, Procedures, and Guidelines

There is sometimes a misconception that regulations, policies, procedures, standards, and guidelines are interchangeable or synonymous with one another. This could not be further from the truth. To understand their differences, these terms need to be fully defined as they relate to compliance. These terms will also be defined as it relates to the Health Insurance Portability and Accountability Act (HIPAA)/Health Information Technology for Economic and Clinical Health (HITECH) Act subject matter.

* From John J. Trinckes, Jr., *The Definitive Guide to Complying with the HIPAA/HITECH Privacy and Security Rules*, Copyright 2013 Taylor & Francis Group, LLC.

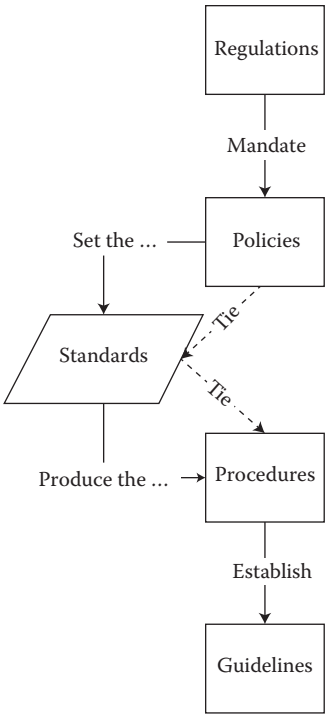


Figure 28.1 Interrelationship of regulations, policies, standards, procedures, and guidelines.

HIPAA/HITECH defines regulations that are mandated by law. These regulations must be implemented and compliance must be met, or there could be severe consequences. These *regulations* form the basis for the covered entities' policies. *Policies* are the intentions of management to come into compliance with the regulations. Policies are high-level requirements that are documented and approved by management to direct employees in the process of complying with the stated objectives. Standards are set by policies that help produce the *procedures* that will be followed to carry out the objectives of the policies. *Standards* attempt to tie the procedures with their policies. Procedures are more detailed than policies and normally provide step-by-step instructions to follow in complying with the policy. Normally, there is one policy statement and several procedures on what the covered entity should do to carry out the policy. Once the procedures have been developed, guidelines are usually established. *Guidelines* are common practices that are followed by employees of a covered entity and are normally the real-life practices that are in place as established by a given procedure (see Figure 28.1).

Reasonable Safeguards

As discussed earlier, a covered entity must have in place appropriate administrative, technical, and physical safeguards that protect against the unauthorized use or disclosure of protected health information. This does not mean that the covered entity's safeguards will guarantee that such use or disclosure will not happen, but rather the potential risk of such activities is acceptable. Of course, the safeguards implemented will vary from one covered entity to

another based upon several factors, such as the covered entity's size or nature of work. It is imperative that every covered entity conducts a risk assessment to determine what safeguards should be implemented based on their distinct requirements. A covered entity may be limited by resources, such as finances or administration; however, there are several examples of safeguards that do not require a lot of effort. Related to administrative controls, providing security awareness training to employees can provide a huge return by increasing the security posture of the covered entity.

One technical control that is probably already in place is the use of passwords on computer systems or programs that store protected health information. A physical control that could be implemented immediately is the practice of speaking quietly when discussing a patient's condition or not using a patient's name when walking through public areas.

Centers for Medicare and Medicaid Services Compliance Review

During 2009, five covered entities were reviewed for compliance with the Security Rule by the Office of E-Health Standards and Services (OESS) of the Centers for Medicare & Medicaid Services (CMS). Historically, these reviews were conducted after complaints were "filed against entities" (FAE). However, covered entities were also reviewed during this year that had no complaints filed. The reason to take a little time to discuss the results of this review is that it gives the covered entity some insight as to areas of compliance that need improvements.

It appears that the following particular areas were focused on (Centers for Medicare & Medicaid Services [CMS] Office of E-Health Standards and Services [OESS] 2009)

- Risk analysis and management
- Security training
- Physical security of facilities and mobile devices
- Off-site access and use of EPHI (electronic protected health information) from remote locations
- Storage of EPHI on portable devices and media
- Disposal of equipment containing EPHI
- Business associate agreements and contracts
- Data encryption
- Virus protection
- Technical safeguards in place to protect EPHI
- Monitoring of access to EPHI

Although the sample size for this review was rather small (only five individual covered entities were reviewed), CMS indicated the following areas of concern over all of them:

1. Risk analysis
2. Currency and adequacy of policies and procedures
3. Security training
4. Business associate agreements

Three other areas (workforce clearance, workstation security, and encryption) appeared to be issues from previous year (i.e., 2008) reviews, but CMS did not note these issues in the current

review. Through the next several sections, areas of concern along with common pitfalls of the covered entities that CMS noted will be discussed.

Risk Analysis

The 2009 HIPAA Security Compliance Review indicated that covered entities are not performing the required risk assessments. The risk assessments are not reviewed on an annual basis or they were outdated based upon significant changes in the covered entities' environment. It appears that the covered entities did not understand the process of conducting a risk assessment. The assessments reviewed were not formally documented or fully developed to cover all necessary steps in the risk assessment process. For instance, the risk assessments appeared to have skipped the first step in identifying systems that store, process, or transmit electronic protected health information. These risk assessments failed to have a complete inventory of systems and locations where these systems are stored.

To mitigate these shortcomings, it is recommended that risk assessments be completed every 18 months or upon a significant change in the technical environment. A significant change in the environment can be brought on by introducing new systems such as electronic medical records solutions. Significantly upgrading existing systems or disposing of retired systems can alter the technical environment of the covered entity. Physically moving electronic assets to different locations can require additional physical controls be implemented. Any reorganization of the covered entity's management or introducing new service offerings can require a change in the environment of the covered entity.

The risk assessments need to address all risks of the covered entity and establish an effective risk assessment process. This risk assessment process includes identifying where data are stored and what systems are considered critical. Identifying threats and vulnerabilities to these systems along with analyzing the controls implemented to protect these systems can be determined by conducting vulnerability/penetration tests of these systems by experienced individuals.

Exploiting these vulnerabilities will provide insight into the impact on the covered entity and assist in assessing the level of risks. By conducting these types of tests or exercises, the covered entity can determine additional controls that should be implemented. As previously discussed, the risk assessment should be formally documented and shared with executive level management to determine corrective actions for any deficiencies identified.

Currency and Adequacy of Policies and Procedures

According to the CMS, out of the five covered entities reviewed, only one had adequate policies and procedures implemented. Conditions of the review indicated that covered entities only had a few policies and procedures in place and most did not address the HIPAA/HITECH Security Standards and Implementation Specifications. In addition, actual guidelines followed were not consistent with the documented procedures or procedures did not follow the documented policies.

One of the recommended solutions is for covered entities to develop formally documented policies that are approved by management and reviewed on a periodic basis. In addition, the workforce member that is responsible for developing these policies and procedures must be one of the covered entity's designated HIPAA security officers. The process for reviewing policies and procedures should include the following:

- Identify the management personnel responsible for the review
- The ability for workforce members to obtain the most recent version of the policies and procedures

- Assess against current operational and regulatory requirements
- Update as necessary
- Document that the review was conducted
- Communicate any updates to the other workforce members as necessary

Developing a standard policy and procedure format will enable consistency across relevant business units. To disseminate policies and procedures, the covered entity should look at a centrally managed repository solution that could automatically notify workforce members when updates occur. For example, this could be performed through an intranet site to inform workforce members of changes or through another type of shared portal that can notify workforce members of updates to policies and procedures. Posting changes in common areas could also provide notification to workforce members of updates. Security awareness training should reiterate any updates to policies and procedures and how these updates are made public to workforce members along with expectations of workforce members to keep abreast of these changes.

In regard to conducting periodic evaluations of policies and procedures, the individuals conducting these reviews should not be the same individuals responsible for the processes under review. These individuals should also have expertise or a reasonable level of competence when conducting these assessments. There are several methods that can be utilized to carry out these reviews such as walkthroughs of the process, interviews of workforce members, assessment of results, or actual testing of controls to determine effectiveness of the policy or procedure in place.

For larger covered entities, the internal audit team may be utilized to conduct the review as long as they have an appropriate level of separations of duties and competencies to conduct such audits. For smaller covered entities or in larger entities that may not have the expertise in-house, an independent third-party service provider should be contracted to assist the covered entity in determining their level of compliance.

Security Training

CMS determined that covered entities do not have policies and procedures in place to address the HIPAA/HITECH Security Rule provisions for security awareness training. For those covered entities that did have policies and procedures in place, they inadequately address the Security Rule requirements. When training was conducted, covered entities did not document or retain evidence of the training as required. In addition, security awareness training was not conducted prior to workforce members being granted access to systems containing electronic protected health information and refresher training was not provided on a regular basis.

Recommendations for these deficiencies are probably self-explanatory. However, as a point of reference, the covered entity needs to develop formal documented security awareness training policies that require new workforce members to receive training prior to gaining access to electronic protected health information. Security awareness training should be provided to all workforce members as a refresher on an annual basis and training material should be updated at least on an annual basis or when necessary. Furthermore, any identified threats through the risk assessment process should be incorporated into the security awareness training. One example of such risk that may be identified is that posed by workforce members working remotely to access electronic protected health information. Finally, attendance of security awareness training needs to be documented, retained, and there should be predetermined sanctions for any workforce member failing to complete this mandatory training or managers that fail to have their members provided with this training.

Business Associate Agreements

There appear to be many deficiencies noted by CMS when it comes to the agreements between business associates and the covered entity. First, the covered entities reviewed had business associates, but there were no business associate agreements (BAAs) between them. Second, there may have been a BAA, but it was not signed by both parties as required. Finally, the BAAs did not address certain requirements dictated by the regulation such as addressing the HIPAA/HITECH Security Rule, developing a comprehensive risk management program, reporting vulnerabilities, reporting breaches, performing activities, and the right of the covered entity to perform an audit on the business associate or require corrections to any deficiencies discovered during the assessment.

It is recommended that covered entities develop a comprehensive process to define the requirements and selection criteria for a business associate. A covered entity should focus on the business processes of the business associate and vet these categories accordingly. A covered entity should develop a standardized template and process of review to document the completion, date, and signatures of both the covered entity and business associate entering into the agreement. The review of the agreements should be conducted at least annually and the procedures should be standardized to document the preamble, body, terms/conditions, and penalties of all business associate agreements.

A contractual document should be developed to describe the business relationship, the services provided, the flow of HIPAA/HITECH Security Rule Standards and Implementation Specifications, and the flow of the specification for the Minimum Necessary Rule regarding electronic protected health information. The business associate contract should have

- A start and end date
- Full service description
- Delivery terms and conditions
- Delivery specifications
- Requirements for conducting a periodic risk assessment and reporting results to the covered entity
- Provision for covered entity to conduct audits on the business associate
- Any other provision regarding vulnerabilities discovered by the business associate's risk assessment to mitigate such risks as applicable

This contractual template should be attached to the business associate template.

HIPAA/HITECH Privacy and Security Audit Program

To ensure compliance with the HIPAA/HITECH Privacy Rule, Security Rule, and the Breach Notification Standards, the Department of Health and Human Services (HHS) is required under Section 13411 of the HITECH Act to conduct periodic audits of covered entities and business associates. The Office for Civil Rights (OCR) established a pilot program that would perform these mandated audits on up to 150 covered entities between November 2011 and December 2012.

The program's objectives are to identify compliance opportunities, best practices, and new risks or vulnerabilities that were not discovered through OCR's complaint investigation and

compliance reviews. This program is a new part of OCR's overall health information privacy and security compliance program.

There were three steps in the process of this pilot audit program. First, starting around July 2011, there was the development of the audit protocol. Just as discussed earlier, this was the test plan development phase. To make sure the testing protocols will work, a small sample of about 20 audits would be conducted. Covered entities were selected, notified, and the audits were performed utilizing the developed testing protocols in phase 2. A review was conducted and changes were made, as necessary, to the testing protocols so that they could be implemented in a standardized manner to the rest of the covered entities chosen. Phase 3 would consist of completing the remaining audits by December 2012.

Of course, every covered entity and business associate could be eligible for an audit. In this first round of auditing, only covered entities such as health services, health plans, and healthcare clearinghouses were chosen. OCR is responsible for selecting the entities that would provide a broad and diverse assessment base. Related to the enforcement authority of the OCR, the covered entity should comply with the audit and cooperate fully with the auditor throughout this process.

As of this writing, KPMG was selected as the OCR auditor for this pilot program.

The audit will include a site visit along with an audit report. Utilizing an interview and observation process, the auditor will determine compliance with the privacy and security standards. The auditor will share the results with the covered entity allowing the covered entity to respond to any findings. A final report, with issues identified along with resolution actions of the covered entity, will be submitted to OCR. The audit could take up to 30 business days to complete. Once the covered entity receives the notification letter that it was chosen for the audit, it has a limited number of days (i.e., 10 business days) to supply all requested documentation.

This notification will commence between 30 and 90 days prior to the expected site visit. Onsite work could take between 3 and 10 business days with a draft of the audit report to follow around 20 business days thereafter. The covered entity will have an opportunity to respond to the findings within 10 business days and the final audit report will be submitted to OCR within 30 business days after the responses are received from the covered entity. These audits are primarily utilized for compliance improvement; however, serious violations could come under compliance review by OCR to address these issues.

Lamkin suggests that the final audit report will include

- Covered entity's name and description
- Methodology and timeframe
- Best practice observations
- Other related documentation such as data, interview notes, and checklists
- A listing of the following for each finding
  - *Condition*—The evidence to back up any notation of noncompliance
  - *Criteria*—Citation of the potential violation of the HIPAA/HITECH
- Privacy or Security Rule the finding presents
  - *Cause*—Supporting documentation to substantiate the reason a finding exists
  - *Effect*—The risk presented by the finding
  - Recommendations to mitigate the finding
  - Any corrective actions taken by the covered entity
- Conclusion

- Corrective action plan
- Recommendations to HHS; that is, continued corrective action or future oversight recommendation

SAS 70/SSAE 16 Debate

The American Institute of Certified Public Accountants (AICPA) developed the Statement on Auditing Standards number 70 (SAS 70) to focus on controls around internal financial reporting. Since data centers and co-locations (COLO) companies house systems that maintained financial reporting applications, users of these companies needed an objective opinion about these data centers. These users started to use the SAS 70 as a requirement before they would utilize the data center (or COLO) companies. The data center owners went out and conducted these audits, but it was not long before marketing got involved and claimed that their businesses were "SAS 70 certified" to validate their data centers.

Unfortunately, the SAS 70 had no objective criteria. Some audit reports may have as little or as many control objectives as the operators of the data center wanted to include on the report. In addition, these companies may claim to be audited even if they did not pass the audit. "The end result is that an SAS 70 audit means nothing without reading the details of the audit report" (Klein 2012). French Caldwell, research vice president at Gartner, concurs with this point by saying, "SAS 70 is basically an expensive auditing process to support compliance with financial reporting rules like the Sarbanes–Oxley Act (SOX). Chief Information Security Officers (CISOs), compliance and risk managers, vendor managers, procurement professionals, and others involved in the purchase or sale of IT services and software need to recognize that SAS 70 is not a security, continuity or privacy compliance standard."

In an attempt to fix some of these issues with the SAS 70 audit, AICPA created a new standard known as the Statement on Standards for Attestation Engagements No. 16 (SSAE 16). The SSAE 16 now requires "the auditor to obtain a written assertion from management regarding the design and operating effectiveness of the controls being reviewed." The company can still choose its own controls, but as long as management attests to the fact that they follow these controls, they can claim to be SSAE 16 audited. It is still up to the report reader to decide the worth of the audit.

Because the SSAE 16 still focused on internal financial audits, the AICPA developed the Service Organization Controls 2 (SOC 2) audit specifically for data centers. To make things a little more confusing, however, they also developed the following:

- Service Organization Controls 1 (SOC1) (also known as SSAE 16)—Types 1 and 2 that can be delivered from a SSAE 16 audit
- Service Organization Controls 2 (SOC 2)—Types 1 and 2 but can use up to five different control objectives as follows:
  - Security
  - Availability
  - Processing integrity
  - Confidentiality
  - Privacy of systems/information
- Service Organization Controls 3 (SOC 3)—Only audit that has a public seal that provides the same level of assurance as SOC 2 but does not provide a detailed description of tests performed. SOC 3 is intended for general release.

As a special note of reference, some companies are now claiming to be "SSAE 16 SOC 2 Certified"; however, there is no such certification available. According to Klein, "As long as users only look for the SSAE 16 audit checkbox, operators will be tempted to use the least rigorous audit criteria to simply pass the audit." This is synonymous with claiming to be "Certified HIPAA Compliant"; there is no such certification available as well. Per the HHS, "There is no standard or implementation specification that requires a covered entity to 'certify' compliance."

Although 45 CFR § 164.308(a)(8) requires a covered entity to perform "a periodic technical and nontechnical evaluation that establishes the extent to which an entity's security policies and procedures meet the security requirements," this evaluation can be performed by either internal or external organizations that provide evaluations or "certification" services. "It is important to note that HHS does not endorse or otherwise recognize private organizations' 'certifications' regarding the Security Rule, and such certifications do not absolve covered entities of their legal obligations under the Security Rule." Just because a covered entity conducted "certification" by an external organization does not mean that HHS will not subsequently find a security violation for which the covered entity will be held responsible.

As Caldwell states, "To ensure that vendor controls are effective for security, privacy compliance and vendor risk management, SAS 70, its successor Statements on Standards for Attestation Engagements (SSAE) 16, and other national audit standards equivalents should be supplemented with self-assessments and agreed upon audit procedures." Some of these other national audit standards include the following:

- Internal Organization for Standardization (ISO) standard certifications
- BITS Shared Assessments—Provided by a consortium of service providers, their customers, audit firms, and other third-party assessors
- SysTrust and WebTrust—Sponsored by AICPA and performed by qualified CPA auditors
- AT Section 101—Sponsored by AICPA and performed by qualified CPA auditors but a more flexible attestation procedure

Corporate Governance

Compliance along with information security is an enterprise-wide, corporate governance issue. Major decisions related to the way that compliance is handled in an organization need to be made at the highest level. With limits in budgets and resources, some covered entities have resorted to implementing a multitier approach to governance. This approach includes the following:

- *An executive-level steering committee*—Normally chaired by the chief information security officer. This committee is responsible for providing the overall broad strategy and commitment to information security efforts including compliance-related matters. This committee will establish budget limitations, goals, priorities, and actions that should be carried out by participants.
- *Advisory groups*—These groups could be organized by projects or functions within the covered entity. These teams can provide detailed insights that can then be reported or recommended back to the steering committee.
- *Subcommittees*—These subcommittees should include representation from across all areas of the covered entity.

Along with setting up an organizational structure, certain rules should be applied. For instance, certain proposals should come from certain responsible parties within the covered entity. These proposals require sponsorship from executive management and each one should have a business case to justify the need. Reviews should be conducted on the effect the proposals may have on existing systems. Finally, the process developed needs to be adhered to and there should be no opportunity for bypassing by the decision makers.

Morrissey adds some rules that should be implemented to handle the corporate governance process:

1. *Chain the committees*—The chairperson for one committee should be on the committee of the next upper level. For instance, the chairperson of one of the subcommittees should be on the executive-level committee that will make the decisions that the subcommittee is handling.
2. *Set authority*—Clear lines of authority should be drawn so that each committee is aware of the level of authority they have over certain decisions.
3. *Make time worthwhile*—Do not have meetings just to have them. For instance, if meetings are only providing status reports, e-mail these reports to responsible workforce members rather than having a meeting. Use meetings to make important decisions and discuss matters that could not be handled in other ways.
4. *Use governance accordingly*—Some covered entities are not at the level to implement this process. Know when the likelihood of success for such a method is appropriate.
5. *Leaders should take a stand*—Individuals in charge of the committees need to have the proper authority and can take a stand. If real change will take place, these leaders must be able to articulate their decisions and have the proper authority to carry them out.

Summary

This chapter explains the differences between regulations, policies, procedures, standards, and guidelines. Many individuals believe that these terms are interchangeable, but they have very different and specific meanings. An understanding of these differences is necessary to better comprehend the overall compliance process. Safeguards are the result of analyzing the necessary resources that must be implemented to adequately satisfy compliance. These safeguards must be reasonably implemented based on several factors, including financial, technical, and personnel resources available to the covered entity. Specific areas of concern were noted through the Centers for Medicare & Medicaid Services' compliance review. These issues were published so that other covered entities can benefit from this insight to strengthen their own compliance efforts.

As part of the HITECH enforcement requirements, the Office for Civil Rights hired KPMG to conduct audits on 150 covered entities by the end of 2012. Although this audit is intended to assist in compliance efforts, severe violations may come under additional investigations. Civil penalties can be severe and covered entities should take the appropriate actions to limit their risks of liability.

This chapter ended with a discussion on SAS 70 and SSAE 16 reporting. A lot of specific information was provided on the audits themselves, but one of the most important tips is that the reports, no matter what type, should be read and a determination should be made as to how well the organization under audit is conducting business.

Information Security Management Handbook: Comprehensive Table of Contents

Domain 1 Access Control

Title	Volume						
	1	*2*	*3*	*4*	*5*	*6*	*7*
1.1 Access Control Techniques							
A Look at RFID Security, Ben Rothke	x						
New Emerging Information Security Technologies and Solutions, Tara Chand	x						
Sensitive or Critical Data Access Controls, Mollie E. Krehnke and David Krehnke	x						
An Introduction to Role-Based Access Control, Ian Clark	x						
Smart Cards, Jim Tiller	x						
A Guide to Evaluating Tokens, Joseph T. Hootman	x						
Controlling FTP: Providing Secured Data Transfers, Chris Hare	x						
Authentication Tokens, Paul A. Henry		x					
Authentication and the Role of Tokens, Jeff Davis		x					

continued

Domain 1 (continued) Access Control

Domain 1 (continued) Access Control

Title	Volume 1	2	3	4	5	6	7
1.4 Access Control Methodologies and Implementation							
Centralized Authentication Services (RADIUS, TACACS, DIAMETER), Bill Stackpole	x						
An Introduction to Secure Remote Access, Christina M. Bird	x						
1.5 Methods of Attack							
Hacker Tools and Techniques, Ed Skoudis	x						
A New Breed of Hacker Tools and Defenses, Ed Skoudis	x						
Breaking News: The Latest Hacker Attacks and Defenses, Ed Skoudis	x						
Counter-Economic Espionage, Craig A. Schiller	x						
Rootkits: The Ultimate Malware Threat, E. Eugene Schultz and Edward Ray		x					
Security Weaknesses of System and Application Interfaces Used to Process Sensitive Information, Sean Price			x				
1.6 Monitoring and Penetration Testing							
Insight into Intrusion Prevention Systems, Gildas Deograt-Lumy	x						
Penetration Testing, Stephen D. Fried	x						

Domain 2 Telecommunications and Network Security

Title	Volume 1	2	3	4	5	6	7
2.1 Communications and Network Security							
Network Security Utilizing an Adaptable Protocol Framework, Robby Fussell	x						

continued

Domain 2 (continued) Telecommunications and Network Security

	Volume						
Title	*1*	*2*	*3*	*4*	*5*	*6*	*7*
The *Five W's and Designing a Secure, Identity-Based, Self-Defending Network (5W Network),* Samuel W. Chun	x						
Maintaining Network Security: Availability via Intelligent Agents, Robby Fussell	x						
PBX Firewalls: Closing the Back Door, William A. Yarberry, Jr.	x						
Network Security Overview, Bonnie A. Goins and Christopher A. Pilewski	x						
Putting Security in the Transport: TLS, Chris Hare	x						
WLAN Security Update, Franjo Majstor	x						
Understanding SSL, Chris Hare	x						
Packet Sniffers and Network Monitors, James S. Tiller and Bryan D. Fish	x						
Secured Connections to External Networks, Steven F. Blanding	x						
Security and Network Technologies, Chris Hare	x						
Wired and Wireless Physical Layer Security Issues, James Trulove	x						
Network Router Security, Steven F. Blanding	x						
What's Not So Simple about SNMP? Chris Hare	x						
Network and Telecommunications Media: Security from the Ground Up, Samuel Chun	x						
Security and the Physical Network Layer, Matthew J. Decker	x						

Domain 2 (continued) Telecommunications and Network Security

Title	Volume						
	1	2	3	4	5	6	7
Wireless LAN Security Challenge, Frandinata Halim and Gildas Deograt	x						
ISO/OSI and TCP/IP Network Model Characteristics, George G. McBride	x						
Facsimile Security, Ben Rothke		x					
Mobile Data Security, George McBride			x				
Integrated Security through Open Standards: A Path to Enhanced Network Visibility, David O'Berry			x				
Adaptive Threats and Defenses, Sean Price				x			
Achieving Global Information Systems Transformation (GIST) through Standards: Foundations for Standards-Based Network Visibility via IF-MAP and Beyond, David O'Berry				x			
A Primer on Demystifying U.S. Government Networks, Samuel W. Chun				x			
IF-MAP as a Standard for Security Data Interchange, David O'Berry					x		
Securing the Grid, Terry Komperda							x
2.2 Internet, Intranet, Extranet Security							
VoIP Security Issues, Anthony Bruno	x						
An Examination of Firewall Architectures, Paul A. Henry	x						
Voice over WLAN, Bill Lipiczky	x						
Spam Wars: How to Deal with Junk E-Mail, Al Bredenberg	x						
Secure Web Services: Holes and Fillers, Lynda L. McGhie	x						

continued

Domain 2 (continued) Telecommunications and Network Security

Title	*Volume*						
	1	*2*	*3*	*4*	*5*	*6*	*7*
IPSec Virtual Private Networks, James S. Tiller	x						
Internet Security: Securing the Perimeter, Douglas G. Conorich	x						
Application-Layer Security Protocols for Networks, Bill Stackpole	x						
Application Layer: Next Level of Security, Keith Pasley	x						
Security of Communication Protocols and Services, William Hugh Murray	x						
An Introduction to IPSec, Bill Stackpole	x						
VPN Deployment and Evaluation Strategy, Keith Pasley	x						
Comparing Firewall Technologies, Per Thorsheim	x						
Cookies and Web Bugs: What They Are and How They Work Together, William T. Harding, Anita J. Reed, and Robert L. Gray	x						
Security for Broadband Internet Access Users, James Trulove	x						
Network Content Filtering and Leak Prevention, Georges J. Jahchan		x					
Web Application Firewalls, Georges J. Jahchan			x				
Understating the Ramifications of IPv6, Foster Henderson					x		
E-Mail Security, Terence Fernandes						x	
2.3 E-Mail Security							
Instant Messaging Security Issues, William Hugh Murray	x						

Domain 2 (continued) Telecommunications and Network Security

Title	*Volume*						
	1	*2*	*3*	*4*	*5*	*6*	*7*
2.4 Secure Voice Communications							
Voice Security, Chris Hare	x						
Secure Voice Communications, Valene Skerpac	x						
2.5 Network Attacks and Countermeasures							
Deep Packet Inspection Technologies, Anderson Ramos	x						
Wireless Penetration Testing: Case Study and Countermeasures, Christopher Pilewski	x						
Auditing the Telephony System: Defenses against Communications Security Breaches and Toll Fraud, William A. Yarberry, Jr.	x						
Insecurity by Proxy, Micah Silverman	x						
Wireless Security, Charles R. Hudson and Chris R. Cunningham	x						
Packet Sniffers: Use and Misuse, Steve A. Rodgers	x						
ISPs and Denial-of-Service Attacks, K. Narayanaswamy	x						
The Ocean Is Full of Phish, Todd Fitzgerald		x					
Botnets, Robert M. Slade			x				
Antispam: Bayesian Filtering, Georges J. Jahchan				x			
Managing Security in Virtual Environments, E. Eugene Schultz and Edward Ray					x		
Attacks in Mobile Environments, Noureddine Boudriga							x

continued

Domain 3 Information Security and Risk Management

Domain 3 (continued) Information Security and Risk Management

Title	Volume						
	1	2	3	4	5	6	7
Understanding Information Security Management Systems, Tom Carlson		x					
Security Management, Ken Buszta	x						
It Is All about Control, Chris Hare	x						
Collaborating Information Security and Privacy to Create Effective Awareness and Training, Rebecca Herold			x				
Security Information and Event Management (SIEM) Technology, E. Eugene Schultz			x				
Managing Mobile Device Security, E. Eugene Schultz and Gal Shpantzer				x			
Establishing an Information Security Program for Local Government, Robert Pittman				x			
Do Your Business Associate Security and Privacy Programs Live Up to HIPAA and HITECH Requirements? Rebecca Herold					x		
Organization Culture Awareness Will Cultivate Your Information Security Program, Robert Pittman					x		
Appreciating Organizational Behavior and Institutions to Solidify Your Information Security Program, Robert K. Pittman, Jr.						x	
Security in the Cloud, Sandy Bacik							x
Getting the Best Out of Information Security Projects, Todd Fitzgerald							x
Mobility and Its Impact on Enterprise Security, Prashanth Venkatesh and Balaji Raghunathan							x

continued

Domain 3 (continued) Information Security and Risk Management

	Volume						
Title	*1*	*2*	*3*	*4*	*5*	*6*	*7*
An Introduction to Digital Rights Management, Ashutosh Saxena and Ravi Sankar Veerubhotla							x
Information Security on the Cheap, Beau Woods							x
Organizational Behavior (Including Institutions) Can Cultivate Your Information Security Program, Robert K. Pittman, Jr.							x
Metrics for Monitoring, Sandy Bacik							x
3.2 Change Control Management							
Patch Management 101: It Just Makes Good Sense! Lynda McGhie	x						
Security Patch Management Process, Felicia M. Nicastro	x						
Configuration Management: Charting the Course for the Organization, Mollie E. Krehnke and David C. Krehnke	x						
3.3 Data Classification							
Understanding Information Risk Management, Tom Carlson and Nick Halvorson				x			
Information Classification: A Corporate Implementation Guide, Jim Appleyard	x						
Ownership and Custody of Data, William Hugh Murray	x						
Developing and Conducting a Security Test and Evaluation, Sean M. Price	x						
Enterprise Security Management, George McBride	x						

Domain 3 (continued) Information Security and Risk Management

Title	Volume						
	1	2	3	4	5	6	7
A Matter of Trust, Ray Kaplan	x						
Trust Governance in a Web Services World, Daniel D. Houser	x						
3.4 Risk Management							
The Role of Information Security in the Enterprise Risk Management Structure, Carl Jackson and Mark Carey	x						
Technology Convergence and Security: A Simplified Risk Management Model, Ken M. Shaurette	x						
Using Quasi-Intelligence Resources to Protect the Enterprise, Craig A. Schiller		x					
Information Risk Management: A Process Approach to Risk Diagnosis and Treatment, Nick Halvorson		x					
Department-Level Transformation, R. Scott McCoy		x					
Setting Priorities in Your Security Program, Derek Schatz		x					
Why and How Assessment of Organization Culture Shapes Security Strategies, Don Saracco		x					
Information Security Risk Assessment, Samantha Thomas Cruz	x						
Risk Management and Analysis, Kevin Henry	x						
New Trends in Information Risk Management, Brett Regan Young	x						

continued

Domain 3 (continued) Information Security and Risk Management

Title	Volume						
	1	*2*	*3*	*4*	*5*	*6*	*7*
Cyber-Risk Management: Technical and Insurance Controls for Enterprise-Level Security, Carol A. Siegel, Ty R. Sagalow, and Paul Serritella	x						
A Look Ahead, Samantha Thomas		x					
The Insider Threat: A View from the Outside, Todd Fitzgerald			x				
Pod Slurping, Ben Rothke			x				
The USB (Universal Security Burden) Nightmare: Pod-Slurping and Other High Storage Capacity Portable Device Vulnerabilities, Kenneth F. Belva			x				
Diary of a Security Assessment: "Put That in Your Pipe and Smoke It!" Ken M. Shaurette			x				
Role-Based Information Security Governance: Avoiding the Company Oil Slick, Todd Fitzgerald					x		
Social Networking Security Exposure, Sandy Bacik					x		
Social Networking, Social Media, and Web 2.0 Security Risks, Robert M. Slade					x		
Applying Adult Education Principles to Security Awareness Programs, Chris Hare					x		
The Information Security Auditors Have Arrived, Now What? Todd Fitzgerald						x	
Continuous Monitoring: Extremely Valuable to Deploy Within Reason, Foster J. Henderson and Mark A. Podracky						x	
Social Networking, Sandy Bacik						x	

Domain 3 (continued) Information Security and Risk Management

Title	Volume						
	1	2	3	4	5	6	7
Insider Threat Defense, Sandy Bacik						x	
Risk Management in Public Key Certificate Applications, Alex Golod						x	
Server Virtualization: Information Security Considerations, Thomas A. Johnson						x	
3.5 Policies, Standards, Procedures, and Guidelines							
Committee of Sponsoring Organizations (COSO), Mignona Cote	x						
Toward Enforcing Security Policy: Encouraging Personal Accountability for Corporate Information Security Policy, John O. Wylder	x						
The Security Policy Life Cycle: Functions and Responsibilities, Patrick D. Howard	x						
People, Processes, and Technology: A Winning Combination, Felicia M. Nicastro	x						
Building an Effective Privacy Program, Rebecca Herold	x						
Establishing an E-Mail Retention Policy: Preventing Potential Legal Nightmares, Stephen Fried	x						
Ten Steps to Effective Web-Based Security Policy Development and Distribution, Todd Fitzgerald	x						
Roles and Responsibilities of the Information Systems Security Officer, Carl Burney	x						

continued

Domain 3 (continued) Information Security and Risk Management

Domain 3 (continued) Information Security and Risk Management

Title	Volume						
	1	2	3	4	5	6	7
3.6 Security Awareness Training							
Measuring Information Security and Privacy Training and Awareness Effectiveness, Rebecca Herold				x			
Change That Attitude: The ABCs of a Persuasive Security Awareness Program, Sam Chun	x						
Maintaining Management's Commitment, William Tompkins	x						
Making Security Awareness Happen, Susan D. Hansche	x						
Beyond Information Security Awareness Training: It Is Time to Change the Culture, Stan Stahl	x						
Protecting Us from Us: Human Firewall Vulnerability Assessments, Ken M. Shaurette and Tom Schleppenbach							x
3.7 Security Management Planning							
The Outsourcing of IT: Seeing the Big Picture, Foster Henderson				x			
Overview of an IT Corporate Security Organization, Jeff Davis	x						
Make Security Part of Your Company's DNA, Ken M. Shaurette	x						
Building an Effective and Winning Security Team, Lynda McGhie	x						
When Trust Goes beyond the Border: Moving Your Development Work Offshore, Stephen Fried	x						
Maintaining Information Security during Downsizing, Thomas J. Bray	x						

continued

Domain 3 (continued) Information Security and Risk Management

Title	Volume						
	1	*2*	*3*	*4*	*5*	*6*	*7*
The Business Case for Information Security: *Selling Management on the Protection of Vital Secrets and Products,* Sanford Sherizen	x						
How to Work with a Managed Security Service Provider, Laurie Hill McQuillan	x						
Considerations for Outsourcing Security, Michael J. Corby	x						
Achieving NERC Compliance: A Compliance Review, Bonnie Goins Pilewski and Christopher A. Pilewski			x				
Controlling the Emerging Data Dilemma: Building Policy for Unstructured Data Access, Anne Shultz					x		
Governance and Risk Management within the Context of Information Security, James C. Murphy					x		
Improving Enterprise Security through Predictive Analysis, Chris Hare					x		
Security Requirements Analysis, Sean M. Price						x	
CERT Resilience Management Model: An Overview, Bonnie A. Goins Pilewski and Christopher Pilewski						x	
Managing Bluetooth Security, E. Eugene Schultz, Matthew W. A. Pemble, and Wendy Goucher						x	
3.8 Ethics							
The Ethical and Legal Concerns of Spyware, Janice C. Sipior, Burke T. Ward, and Georgina R. Roselli	x						

Domain 3 (continued) Information Security and Risk Management

Title	Volume 1	2	3	4	5	6	7
Ethics and the Internet, Micki Krause	x						
Computer Ethics, Peter S. Tippett	x						
3.9 Employment Policies and Practices							
Slash and Burn: In Times of Recession, Do Not Let Emotions Drive Business Decisions, Anonymous						x	
A "Zero Trust" Model for Security, Ken Shaurette and Thomas J. Schleppenbach						x	

Domain 4 Application Development Security

Title	Volume 1	2	3	4	5	6	7
4.1 Application Issues							
Application Service Provider Security: Ensuring a Secure Relationship for the Client and the ASP, Stephen D. Fried	x						
Stack-Based Buffer Overflows, Jonathan S. Held	x						
Web Application Security, Mandy Andress	x						
Security for XML and Other Metadata Languages, William Hugh Murray	x						
XML and Information Security, Samuel C. McClintock	x						
Application Security, Walter S. Kobus, Jr.	x						
Covert Channels, Anton Chuvakin	x						

continued

Domain 4 (continued) Application Development Security

Title	Volume						
	1	*2*	*3*	*4*	*5*	*6*	*7*
Security as a Value Enhancer in Application Systems Development, Lowell Bruce McCulley	x						
Open Source versus Closed Source, Ed Skoudis	x						
A Look at Java Security, Ben Rothke	x						
Neural Networks and Information Assurance Uses, Sean M. Price		x					
Information Technology Infrastructure Library and Security Management Overview, David McPhee		x					
Adaptation: A Concept for Next-Generation Security Application Development, Robby S. Fussell		x					
Quantum Computing: Implications for Security, Robert M. Slade		x					
Mashup Security, Mano Paul			x				
Format String Vulnerabilities, Mano Paul			x				
Service-Oriented Architecture, Walter B. Williams							x
4.2 Databases and Data Warehousing							
Reflections on Database Integrity, William Hugh Murray	x						
Digital Signatures in Relational Database Applications, Mike R. Prevost	x						
Security and Privacy for Data Warehouses: Opportunity or Threat? David Bonewell, Karen Gibbs, and Adriaan Veldhuisen	x						

Domain 4 (continued) Application Development Security

Title	Volume						
	1	2	3	4	5	6	7
4.3 Systems Development Controls							
Building and Assessing Security in the Software Development Lifecycle, George G. McBride	x						
Avoiding Buffer Overflow Attacks, Sean Price	x						
Secure Development Life Cycle, Kevin Henry	x						
System Development Security Methodology, Ian Lim and Ioana V. Bazawan	x						
Software Engineering Institute Capability Maturity Mode, Matt Nelson	x						
Enterprise Security Architecture, William Hugh Murray	x						
Certification and Accreditation Methodology, Mollie E. Krehnke and David C. Krehnke	x						
System Development Security Methodology, Ian Lim and Ioana V. Carastan	x						
Methods of Auditing Applications, David C. Rice and Graham Bucholz	x						
Data Loss Prevention Program, Powell Hamilton				x			
Data Reliability: Trusted Time Stamps, Jeff Stapleton				x			
Security in the .NET Framework, James D. Murray				x			
The Effectiveness of Access Management Reviews, Chris Hare					x		
Securing SaaS Applications: A Cloud Security Perspective for Application Providers, Pradnyesh Rane					x		

continued

Domain 4 (continued) Application Development Security

Title	*Volume*						
	1	*2*	*3*	*4*	*5*	*6*	*7*
Attacking RFID Systems, Pedro Peris-Lopez, Julio Cesar Hernandez-Castro, Juan M. Estevez-Tapiador, and Arturo Ribagorda					x		
Application Whitelisting, Georges J. Jahchan						x	
Design of Information Security for Large System Development Projects, James C. Murphy						x	
Building Application Security Testing into the Software Development Life Cycle, Sandy Bacik						x	
Managing the Security Testing Process, Anthony Meholic							x
Security and Resilience in the Software Development Life Cycle, Mark S. Merkow and Lakshmikanth Raghavan							x
4.4 Malicious Code							
Fast Scanning Worms, Paul A. Henry			x				
Organized Crime and Malware, Michael Pike			x				
Net-Based Malware Detection: A Comparison with Intrusion Detection Models, Robert M. Slade		x					
Malware and Computer Viruses, Robert M. Slade	x						
An Introduction to Hostile Code and Its Control, Jay Heiser		x					
A Look at Java Security, Ben Rothke	x						
Twenty-Five (or Forty) Years of Malware History, Robert M. Slade						x	

Domain 4 (continued) Application Development Security

Title	Volume 1	2	3	4	5	6	7
4.5 Methods of Attack							
Hacking Methods, Georges J. Jahchan	x						
Enabling Safer Deployment of Internet Mobile Code Technologies, Ron Moritz	x						

Domain 5 Cryptography

Title	Volume 1	2	3	4	5	6	7
5.1 Use of Cryptography							
Auditing Cryptography: Assessing System Security, Steve Stanek	x						
Three New Models for the Application of Cryptography, Jay Heiser		x					
5.2 Cryptographic Concepts, Methodologies, and Practices							
Cryptographic Transitions, Ralph Spencer Poore	x						
Blind Detection of Steganographic Content in Digital Images Using Cellular Automata, Sasan Hamidi	x						
An Overview of Quantum Cryptography, Ben Rothke	x						
Elliptic Curve Cryptography: Delivering High-Performance Security for E-Commerce and Communications, Paul Lambert	x						
Cryptographic Key Management Concepts, Ralph Spencer Poore	x						
Message Authentication, James S. Tiller	x						

continued

Domain 5 (continued) Cryptography

Title	*Volume*						
	1	*2*	*3*	*4*	*5*	*6*	*7*
Fundamentals of Cryptography and Encryption, Ronald A. Gove	x						
Steganography: The Art of Hiding Messages, Mark Edmead	x						
An Introduction to Cryptography, Javek Ikbal	x						
Hash Algorithms: From Message Digests to Signatures, Keith Pasley	x						
A Look at the Advanced Encryption Standard (AES), Ben Rothke	x						
Message Digest, Ralph Spencer Poore			x				
Quantum Computing: The Rise of the Machine, Robby Fussell			x				
Cryptography: A Unifying Principle in Compliance Programs, Ralph Spencer Poore				x			
Cryptography: Mathematics vs. Engineering, Ralph Spencer Poore					x		
Cryptographic Message Syntax, Jeff Stapleton					x		
Format Preserving Encryption, Ralph Spencer Poore						x	
Elliptic Curve Cryptosystems, Jeff Stapleton						x	
Pirating the Ultimate Killer App: Hacking Military Unmanned Aerial Vehicles, Sean P. McBride						x	
Cloud Cryptography, Jeff Stapleton							x
5.3 Private Key Algorithms							
Principles and Applications of Cryptographic Key Management, William Hugh Murray	x						

Domain 5 (continued) Cryptography

Title	Volume						
	1	2	3	4	5	6	7
5.4 Public Key Infrastructure (PKI)							
Preserving Public Key Hierarchy, Geoffrey C. Grabow	x						
PKI Registration, Alex Golod	x						
Encryption Key Management in Large-Scale Network Deployments, Franjo Majstor and Guy Vancollie		x					
5.5 System Architecture for Implementing Cryptographic Functions							
Implementing Kerberos in Distributed Systems, Joe Kovara and Ray Kaplan	x						
5.6 Methods of Attack							
Methods of Attacking and Defending Cryptosystems, Joost Houwen	x						

Domain 6 Security Architecture and Design

Title	Volume						
	1	2	3	4	5	6	7
6.1 Principles of Computer and Network Organizations, Architectures, and Designs							
Enterprise Assurance: A Framework Explored, Bonnie A. Goins	x						
Creating a Secure Architecture, Christopher A. Pilewski and Bonnie A. Goins	x						
Common Models for Architecting an Enterprise Security Capability, Matthew J. Decker	x						
The Reality of Virtual Computing, Chris Hare	x						

continued

Domain 6 (continued) Security Architecture and Design

	Volume						
Title	*1*	*2*	*3*	*4*	*5*	*6*	*7*
Service-Oriented Architecture and Web Services Security, Glenn J. Cater		x					
Analysis of Covert Channels, Ralph Spencer Poore		x					
Security Architecture of Biological Cells: An Example of Defense in Depth, Kenneth J. Knapp and R. Franklin Morris, Jr.		x					
ISO Standards Draft Content, Scott Erkonen		x					
Security Frameworks, Robert M. Slade		x					
Information Flow and Covert Channels, Sean Price			x				
Securing Data at Rest: From Smartphones to Tapes Defining Data at Rest, Sam Chun and Leo Kahng			x				
Best Practices in Virtualization Security, Shanit Gupta				x			
Everything New Is Old Again, Robert M. Slade				x			
An Introduction to Virtualization Security, Paul Henry					x		
Service-Oriented Architecture, Walter B. Williams						x	
Cloud Security, Terry Komperda						x	
Enterprise Zones of Trust, Sandy Bacik						x	
6.2 Principles of Security Models, Architectures, and Evaluation Criteria							
Formulating an Enterprise Information Security Architecture, Mollie E. Krehnke and David C. Krehnke	x						
Security *Architecture and Models,* Foster J. Henderson and Kellina M. Craig-Henderson	x						

Domain 6 (continued) Security Architecture and Design

Title	Volume						
	1	2	3	4	5	6	7
The Common Criteria for IT Security Evaluation, Debra S. Herrmann	x						
Identity and Access Management Architecture, Jeff Crume							x
FedRAMP℠: Entry or Exit Ramp for Cloud Security? Debra S. Herrmann							x
6.3 Common Flaws and Security Issues: System Architecture and Design							
Common System Design Flaws and Security Issues, William Hugh Murray	x						

Domain 7 Operations Security

Title	Volume						
	1	2	3	4	5	6	7
7.1 Concepts							
Security Considerations in Distributed Computing: A Grid Security Overview, Sasan Hamidi	x						
Managing Unmanaged Systems, Bill Stackpole and Man Nguyen	x						
Storage Area Networks Security Protocols and Mechanisms, Franjo Majstor	x						
Operations: The Center of Support and Control, Kevin Henry	x						
Why Today's Security Technologies Are So Inadequate: History, Implications, and New Approaches, Steven Hofmeyr	x						
Operations Security and Controls, Patricia A.P. Fisher	x						
Data Storage and Network Security, Greg Schulz							x

continued

Domain 7 (continued) Operations Security

Title	*Volume*						
	1	*2*	*3*	*4*	*5*	*6*	*7*
7.2 Resource Protection Requirements							
The Nebulous Zero Day, Rob Slade	x						
Understanding Service Level Agreements, Gilbert Held	x						
Physical Access Control, Dan M. Bowers	x						
7.3 Auditing							
Auditing the Electronic Commerce Environment, Chris Hare	x						
7.4 Intrusion Detection							
Improving Network-Level Security through *Real-Time Monitoring and Intrusion Detection,* Chris Hare		x					
Intelligent Intrusion Analysis: How Thinking *Machines Can Recognize Computer Intrusions,* Bryan D. Fish	x						
7.5 Operations Controls							
Directory Security, Ken Buszta	x						
Patch Management 101: It Just Makes Good Sense! Lynda McGhie		x					
Security Patch Management: The Process, Felicia M. Nicastro		x					
Validating Tape Backups, Sandy Bacik			x				
A Brief Summary of Warfare and Commercial Entities, Rob Shein				x			
Information Destruction Requirements and Techniques, Ben Rothke				x			
Warfare and Security: Deterrence and Dissuasion in the Cyber Era, Samuel Chun					x		

Domain 7 (continued) Operations Security

Title	Volume						
	1	2	3	4	5	6	7
Configuration, Change, and Release Management, Sean M. Price					x		
Tape Backup Considerations, Sandy Bacik					x		
Productivity vs. Security, Sandy Bacik					x		
Complex Event Processing for Automated Security Event Analysis, Rob Shein						x	
Records Management, Sandy Bacik						x	

Domain 8 Business Continuity and Disaster Recovery Planning

Title	Volume						
	1	2	3	4	5	6	7
8.1 Business Continuity Planning							
Developing Realistic Continuity Planning Process Metrics, Carl B. Jackson	x						
Building Maintenance Processes for Business Continuity Plans, Ken Doughty	x						
Identifying Critical Business Functions, Bonnie A. Goins	x						
Selecting the Right Business Continuity Strategy, Ken Doughty	x						
Contingency Planning Best Practices and Program Maturity, Timothy R. Stacey	x						
Reengineering the Business Continuity Planning Process, Carl B. Jackson	x						

continued

Domain 8 (continued) Business Continuity and Disaster Recovery Planning

Domain 8 (continued) Business Continuity and Disaster Recovery Planning

	Volume						
Title	*1*	*2*	*3*	*4*	*5*	*6*	*7*
Business Continuity Planning: A Collaborative Approach, Kevin Henry	x						
8.3 Elements of Business Continuity Planning							
The Business Impact Assessment Process, Carl B. Jackson	x						

Domain 9 Legal, Regulations, Compliance, and Investigations

	Volume						
Title	*1*	*2*	*3*	*4*	*5*	*6*	*7*
9.1 Information Law							
Sarbanes–Oxley Compliance: A Technology Practitioner's Guide, Bonnie A. Goins	x						
Health Insurance Portability and Accountability Act Security Rule, Lynda L. McGhie	x						
Jurisdictional Issues in Global Transmissions, Ralph Spencer Poore	x						
An Emerging Information Security Minimum Standard of Due Care, Robert Braun and Stan Stahl	x						
ISPs and Accountability, Lee Imrey		x					
The Case for Privacy, Michael J. Corby	x						
Liability for Lax Computer Security in DDoS Attacks, Dorsey Morrow	x						
Compliance Assurance: Taming the Beast, Todd Fitzgerald		x					
The Cost of Risk: An Examination of Risk Assessment and Information Security in the Financial Industry, Seth Kinnett					x		

continued

Domain 9 (continued) Legal, Regulations, Compliance, and Investigations

	Volume						
Title	*1*	*2*	*3*	*4*	*5*	*6*	*7*
The Cost of Risk: An Examination of Risk Assessment and Information Security in the Financial Industry, Seth Kinnett					x		
Data Security and Privacy Legislation, Salahuddin Kamran					x		
National Patient Identifier and Patient Privacy in the Digital Era, Tim Godlove and Adrian Ball							x
Addressing Social Media Security and Privacy Challenges, Rebecca Herold							x
9.2 Investigations							
Operational Forensics, Michael J. Corby	x						
Computer Crime Investigation and Computer *Forensics,* Thomas Welch	x						
What Happened? Kelly J. Kuchta	x						
What Is Digital Forensics and What Should You Know about It? Greg Gogolin							x
eDiscovery, David G. Hill							x
Overview of the Steps of the Electronic Discovery Reference Model, David G. Hill							x
Cell Phone Protocols and Operating Systems, Eamon P. Doherty							x
9.3 Major Categories of Computer Crime							
Potential Cyber Terrorist Attacks, Chris Hare	x						
The Evolution of the Sploit, Ed Skoudis	x						
Computer Crime, Christopher A. Pilewski	x						

Domain 9 (continued) Legal, Regulations, Compliance, and Investigations

	Volume						
Title	*1*	*2*	*3*	*4*	*5*	*6*	*7*
Phishing: A New Twist to an Old Game, Stephen D.Fried	x						
It's All about Power: Information Warfare Tactics by Terrorists, Activists, and Miscreants, Gerald L. Kovacich, Andy Jones, and Perry G. Luzwick	x						
Bluesnarfing, Mano Paul			x				
Cyberstalking, Micki Krause Nozaki				x			
Managing Advanced Persistent Threats, E. Eugene Schultz and Cuc Du						x	
Hacktivism: The Whats, Whys, and Wherefores, Chris Hare							x
9.4 Incident Handling							
Social Engineering: The Human Factor in Information Assurance, Marcus K. Rogers	x						
Privacy Breach Incident Response, Rebecca Herold	x						
Security Event Management, Glenn Cater	x						
DCSA: A Practical Approach to Digital Crime Scene Analysis, Marcus K. Rogers	x						
What a Computer Security Professional Needs to Know about E-Discovery and Digital Forensics, Larry R. Leibrock	x						
How to Begin a Non-Liturgical Forensic Examination, Carol Stucki	x						
Honeypot Essentials, Anton Chuvakin	x						
Managing the Response to a Computer Security Incident, Michael Vangelos	x						

continued

Domain 9 (continued) Legal, Regulations, Compliance, and Investigations

	Volume						
Title	*1*	*2*	*3*	*4*	*5*	*6*	*7*
Cyber-Crime: Response, Investigation, and Prosecution, Thomas Akin	x						
Enterprise Incident Response and Digital Evidence Management and Handling, Marcus K. Rogers		x					
Security Information Management Myths and Facts, Sasan Hamidi		x					
Virtualization and Digital Investigations, Marcus K. Rogers and Sean C. Leshney			x				
Is Software Write Blocking a Viable Alternative to Hardware Write Blocking in Computer Forensics? Paul A. Henry				x			
Discovery of Electronically Stored Information, Salahuddin Kamran					x		
Virtualization Forensics, Paul A. Henry						x	
9.5 Compliance							
PCI Compliance, Tyler Justin Speed							x
HIPAA/HITECH Compliance Overview, John J. Trinckes, Jr.							x

Domain 10 Physical (Environmental) Security

	Volume						
Title	*1*	*2*	*3*	*4*	*5*	*6*	*7*
10.1 Elements of Physical Security							
Perimeter Security, R. Scott McCoy	x						
Melding Physical Security and Traditional Information Systems Security, Kevin Henry	x						

Domain 10 (continued) Physical (Environmental) Security

	Volume						
Title	*1*	*2*	*3*	*4*	*5*	*6*	*7*
Physical Security for Mission-Critical Facilities and Data Centers, Gerald Bowman	x						
Physical Security: A Foundation for Information Security, Christopher Steinke	x						
Physical Security: Controlled Access and Layered Defense, Bruce R. Matthews	x						
Computing Facility Physical Security, Alan Brusewitz	x						
Closed-Circuit Television and Video Surveillance, David Litzau	x						
Mantraps and Turnstiles, R. Scott McCoy		x					
Halon Fire Suppression Systems, Chris Hare			x				
Crime Prevention through Environmental Design, Mollie Krehnke			x				
Data Center Site Selection and Facility Design Considerations, Sandy Bacik			x				
Protection of Sensitive Data, Sandy Bacik				x			
Water Leakage and Flooding, Sandy Bacik				x			
Site Selection and Facility Design Considerations, Sandy Bacik				x			
An Overview of IP-Based Video Surveillance, Leo Kahng				x			
The Layered Defense Model and Perimeter Intrusion Detection, Leo Kahng					x		
Terrorism an Overview, Frank Bolz, Jr., Kenneth J. Dudonis, and David P. Schulz						x	

continued

Domain 10 (continued) Physical (Environmental) Security

	Volume						
Title	*1*	*2*	*3*	*4*	*5*	*6*	*7*
10.2 Technical Controls							
Types of Information Security Controls, Harold F. Tipton	x						
Countermeasure Goals and Strategies, Thomas L. Norman						x	
10.3 Environment and Life Safety							
Workplace Violence: Event Characteristics and Prevention, George Richards	x						
Physical Security: The Threat after September 11, 2001, Jaymes Williams	x						

Information Security Management Handbook

Sixth Edition • Volume 7

Edited by
Richard O'Hanley · James S. Tiller

Updated annually, the **Information Security Management Handbook, Sixth Edition, Volume 7** is the most comprehensive and up-to-date reference available on information security and assurance. Bringing together the knowledge, skills, techniques, and tools required of IT security professionals, it facilitates the up-to-date understanding required to stay one step ahead of evolving threats, standards, and regulations.

Reporting on the latest developments in information security and recent changes to the (ISC)$^{2®}$ CISSP® Common Body of Knowledge (CBK®), this volume features 27 new chapters on topics such as BYOD, IT consumerization, smart grids, security, and privacy.

- Covers the fundamental knowledge, skills, techniques, and tools required by IT security professionals
- Updates its bestselling predecessors with new developments in information security and the (ISC)^2 CISSP CBK
- Provides valuable insights from leaders in the field on the theory and practice of computer security technology
- Facilitates the comprehensive and up-to-date understanding you need to stay fully informed

The ubiquitous nature of computers and networks will always provide the opportunity and means to do harm. This edition updates its popular predecessors with the information you need to address the vulnerabilities created by recent innovations such as cloud computing, mobile banking, digital wallets, and near-field communications. This handbook is also available on CD.

CRC Press
Taylor & Francis Group
an informa business
www.crcpress.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

K16337

ISBN: 978-1-4665-6749-8
90000
9 781466 567498