

*Making Everything Easier!™*

*Solera Networks Special Edition*

# **Big Data Security**

FOR  
**DUMMIES®**

**Learn to:**

- Harness the power of Big Data to detect advanced threats and targeted attacks
- Collect digital evidence and streamline incident response
- Integrate Big Data Security into your existing security fabric

*Brought to you by*



**Steve Piper, CISSP**



# About Solera Networks

Solera Networks is a leading Big Data Security intelligence and analytics company. Its award-winning DeepSee platform levels the battlefield against advanced targeted attacks and malware, and gives security professionals clear and concise answers to the toughest security questions. Solera DeepSee is powered by next-generation deep-packet inspection and indexing technologies, full-packet capture, malware analysis, and real-time security intelligence and analytics capabilities.

Global 2000 enterprises, cloud service providers, and government agencies rely on Solera for real-time situational awareness, security incident response, cyberthreat detection, data loss monitoring and analysis, organization policy compliance and security assurance — allowing them to respond quickly and intelligently to advanced threats and attacks, while protecting critical information assets, minimizing exposure and loss, and reducing business liabilities.

For more information, visit [www.soleranetworks.com](http://www.soleranetworks.com).

# ***Big Data Security*** FOR **DUMMIES®**

SOLERA NETWORKS SPECIAL EDITION

**by Steve Piper, CISSP**



WILEY

John Wiley & Sons, Inc.

## Big Data Security For Dummies®, Solera Networks Special Edition

Published by  
John Wiley & Sons, Inc.  
111 River St.  
Hoboken, NJ 07030-5774  
[www.wiley.com](http://www.wiley.com)

Copyright © 2013 by John Wiley & Sons, Inc., Hoboken, New Jersey

Published by John Wiley & Sons, Inc., Hoboken, New Jersey

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the Publisher. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permissions>.

**Trademarks:** Wiley, the Wiley logo, For Dummies, the Dummies Man logo, A Reference for the Rest of Us!, The Dummies Way, Dummies.com, Making Everything Easier, and related trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries, and may not be used without written permission. Solera Networks and the Solera Networks logo are trademarks or registered trademarks of Solera Networks Inc. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc., is not associated with any product or vendor mentioned in this book.

**LIMIT OF LIABILITY/DISCLAIMER OF WARRANTY: THE PUBLISHER AND THE AUTHOR MAKE NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE ACCURACY OR COMPLETENESS OF THE CONTENTS OF THIS WORK AND SPECIFICALLY DISCLAIM ALL WARRANTIES, INCLUDING WITHOUT LIMITATION WARRANTIES OF FITNESS FOR A PARTICULAR PURPOSE. NO WARRANTY MAY BE CREATED OR EXTENDED BY SALES OR PROMOTIONAL MATERIALS. THE ADVICE AND STRATEGIES CONTAINED HEREIN MAY NOT BE SUITABLE FOR EVERY SITUATION. THIS WORK IS SOLD WITH THE UNDERSTANDING THAT THE PUBLISHER IS NOT ENGAGED IN RENDERING LEGAL, ACCOUNTING, OR OTHER PROFESSIONAL SERVICES. IF PROFESSIONAL ASSISTANCE IS REQUIRED, THE SERVICES OF A COMPETENT PROFESSIONAL PERSON SHOULD BE SOUGHT. NEITHER THE PUBLISHER NOR THE AUTHOR SHALL BE LIABLE FOR DAMAGES ARISING HEREFROM. THE FACT THAT AN ORGANIZATION OR WEBSITE IS REFERRED TO IN THIS WORK AS A CITATION AND/OR A POTENTIAL SOURCE OF FURTHER INFORMATION DOES NOT MEAN THAT THE AUTHOR OR THE PUBLISHER ENDORSES THE INFORMATION THE ORGANIZATION OR WEBSITE MAY PROVIDE OR RECOMMENDATIONS IT MAY MAKE. FURTHER, READERS SHOULD BE AWARE THAT INTERNET WEBSITES LISTED IN THIS WORK MAY HAVE CHANGED OR DISAPPEARED BETWEEN WHEN THIS WORK WAS WRITTEN AND WHEN IT IS READ.**

For general information on our other products and services, or how to create a custom For Dummies book for your business or organization, please contact our Business Development Department in the U.S. at 877-409-4177, contact [info@dummies.biz](mailto:info@dummies.biz), or visit [www.wiley.com/go/custompub](http://www.wiley.com/go/custompub). For information about licensing the For Dummies brand for products or services, contact [BrandedRights&Licenses@Wiley.com](mailto:BrandedRights&Licenses@Wiley.com).

ISBN: 978-1-118-51727-7

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

## Publisher's Acknowledgments

We're proud of this book and of the people who worked on it. For details on how to create a custom book for your business or organization, contact [info@dummies.biz](mailto:info@dummies.biz) or visit [www.wiley.com/go/custompub](http://www.wiley.com/go/custompub). For details on licensing the brand for products or services, contact [BrandedRights&Licenses@Wiley.com](mailto:BrandedRights&Licenses@Wiley.com).

Some of the people who helped bring this book to market include the following:

### ***Acquisitions, Editorial, and Vertical Websites***

**Development Editor:** Kathy Simpson

**Project Editor:** Jennifer Bingham

**Editorial Manager:** Rev Mengle

**Business Development Representative:**  
Kimberley Schumacker

**Custom Publishing Project Specialist:**  
Michael Sullivan

### ***Composition Services***

**Project Coordinator:** Kristie Rees

**Layout and Graphics:** Carrie A. Cesavice,  
Christin Swinford

**Proofreader:** Dwight Ramsey

**Special help from Solera Networks:**  
John Vecchi, Bret Jordan, Alan Hall,  
Armen Sargsyan, Andrew Brandt,  
Davin Baker, Ajay Uggirala, Joe Levy

---

### **Publishing and Editorial for Technology Dummies**

**Richard Swadley**, Vice President and Executive Group Publisher

**Andy Cummings**, Vice President and Publisher

**Mary Bednarek**, Executive Director, Acquisitions

**Mary C. Corder**, Editorial Director

### **Publishing and Editorial for Consumer Dummies**

**Kathleen Nebenhaus**, Vice President and Executive Publisher

### **Composition Services**

**Debbie Stailey**, Director of Composition Services

### **Business Development**

**Lisa Coleman**, Director, New Market and Brand Development



# Table of Contents

<b>Introduction</b> .....	<b>1</b>
How This Book Is Organized .....	1
Icons Used in This Book.....	2
<b>Chapter 1: Scanning the Cyberthreat Landscape</b> .....	<b>3</b>
What's the Risk? .....	3
Who's at Risk? .....	4
Trending Cyberthreats .....	5
Shifting attackers .....	5
Broader reach, bigger risk.....	7
Modern malware .....	9
The Cost of Failure.....	14
<b>Chapter 2: Why Traditional Security Isn't Enough</b> ....	<b>15</b>
Basic Threat Detection Techniques .....	16
Traditional Security Defenses .....	17
Endpoint security .....	17
Intrusion prevention systems .....	18
Next-generation firewalls.....	18
Secure e-mail gateways .....	18
Secure web gateways .....	19
Data loss prevention .....	19
Network behavior analysis .....	20
Advanced malware protection.....	20
Security information and event management .....	21
New Offenses versus Traditional Defenses .....	21
Hand-carried mobile devices .....	21
Unknown and zero-day threats .....	21
Encrypted threats .....	22
Lack of context and forensics .....	22
<b>Chapter 3: Meet Big Data Security</b> .....	<b>23</b>
What Is Big Data? .....	23
Internal Big Data sources.....	24
External Big Data sources.....	24
What Is Big Data Security? .....	25
Limited-visibility solutions .....	25
Full-visibility solutions .....	26

Introducing Security Intelligence and Analytics .....	26
How SIA works .....	27
Form factors in SIA .....	27
Common features of SIA solutions .....	29
Advanced features of SIA solutions.....	31
Deploying SIA.....	33
Why size really matters .....	33
Leveraging SPAN ports and TAPs.....	34
Supporting a distributed architecture .....	35
Removing blind spots through SSL decryption .....	35
<b>Chapter 4: Use Cases for Big Data Security . . . . .</b>	<b>37</b>
Situational Awareness .....	38
Cyberthreat Detection.....	39
Before attack .....	39
During attack.....	40
After attack.....	40
Security Incident Response .....	40
Recovering from an incident .....	42
Data Loss Monitoring and Analysis .....	44
Organizational Policy Compliance Verification .....	44
Security Assurance.....	46
<b>Chapter 5: Integrating Big Data Security . . . . .</b>	<b>49</b>
SIEM Integration.....	50
IPS Integration .....	52
NGFW Integration.....	53
Advanced Malware Protection Integration.....	53
Universal Connectors .....	55
<b>Chapter 6: Ten Buying Criteria for Big Data Security . . .</b>	<b>57</b>
Hardware Flexibility .....	58
Ease of Use.....	58
Full-Packet Capture.....	59
Deep Packet Inspection.....	59
Enterprise Performance and Scalability .....	59
Virtual Platform Visibility .....	60
Content Reconstruction .....	60
SSL Decryption .....	61
Extensive Third-Party Integration.....	61
Responsive Customer Support.....	62
<b>Glossary .....</b>	<b>63</b>

# Introduction



**T**hese days, the skill of modern hackers has eclipsed the defensive capabilities of traditional cybersecurity tools. Major breaches are consistently occurring in the networks of the most technically savvy enterprises and government agencies — and even in security vendors' own networks.

Over the years, network security teams have implemented a wide range of “set-it-and-forget-it” tools that attempt to block threats based on signatures and traffic behaviors. Although these tools provide some defensive cover, they may also provide a false sense of security because high-profile organizations continue to suffer from new and ingenious attacks.

Security-conscious organizations are turning to Big Data Security as the newest weapon in their cybercrime arsenals. By collecting all available digital evidence — including raw packets, flow data, and files — organizations can uncover advanced targeted attacks traditional security defenses sometimes miss. These organizations are learning to use internal data sources they never knew existed and to extend the value of known data sources by integrating their Big Data Security solutions into their existing security fabric.

If you're in charge of securing your organization's network or responding to security incidents, this book is for you.

## *How This Book Is Organized*

This book is organized so that you don't have to read it cover-to-cover, front-to-back. You can skip around and read just the chapters that interest you:

- ✓ **Chapter 1, “Scanning the Cyberthreat Landscape,”** reviews recent statistics on enterprise network breaches and discusses the changing personas of successful hackers. It also explores modern malware, advanced penetration techniques, and the growing costs of data breaches for enterprises.

- ✓ **Chapter 2, “Why Traditional Security Isn’t Enough,”** covers signature-based and signatureless defenses, as well as blacklist, whitelist, and behavioral-profiling approaches to security. It provides an overview of major categories of traditional security defenses and describes why none of them are foolproof.
- ✓ In **Chapter 3, “Meet Big Data Security,”** I get to the heart of the matter by defining Big Data Security and contrasting limited-visibility offerings with full-visibility platforms called security intelligence and analytics (SIA). I end the chapter by describing form factors, features, and deployment strategies for Big Data Security solutions.
- ✓ **Chapter 4, “Use Cases for Big Data Security,”** identifies the six most common use cases for Big Data Security and then describes how each can improve your network security posture. I then offer real-world examples of how organizations have leveraged Big Data Security to solve network security challenges.
- ✓ In **Chapter 5, “Integrating Big Data Security,”** I describe the benefits of integrating Big Data Security into your existing security fabric and the role of universal connectors when vendor-supplied integration isn’t available.
- ✓ In **Chapter 6, “Ten Buying Criteria for Big Data Security,”** I describe what to look for — and what to avoid — when evaluating leading Big Data Security solutions.

## Icons Used in This Book



This book uses the following icons to indicate special content.

You won’t want to forget the information in these paragraphs.



These paragraphs provide practical advice that will help you craft a better strategy, whether you’re planning a purchase or setting up your software.



Look out! When you see this icon, it’s time to pay attention. You’ll find important cautionary information you won’t want to miss.



Maybe you’re one of those highly detailed people who really need to grasp all the nuts and bolts, even the most techie parts. If so, these tidbits are right up your alley.

## Chapter 1

---

# Scanning the Cyberthreat Landscape

---

### *In This Chapter*

- ▶ Reviewing the latest trends in cyberthreats
  - ▶ Exploring modern malware and advanced hacking techniques
  - ▶ Dissecting the costs of a successful data breach
- 

**T**oday with cyberthreats growing more sophisticated every day, hacking for kicks is a thing of the past. Nowadays, hackers can be well funded (or even state-sponsored), highly motivated, and trained in advanced techniques to evade even the best-of-breed security defenses.

Before I explain why traditional cybersecurity defenses simply can't keep up (see Chapter 2), and why leading IT security teams are turning to Big Data Security for answers (see Chapter 3), I discuss the cyberthreats that all commercial and government organizations face — and the costs of failing to detect them.

## *What's the Risk?*

Several reputable organizations monitor cyberthreat trends and the effects of those threats on organizations. Among these is the Verizon RISK (Response, Intelligence, Solutions and Knowledge) Team, which publishes a highly regarded annual Data Breach Investigations Report. (To download a free copy of the report, connect to [www.verizonbusiness.com](http://www.verizonbusiness.com).) In 2012, Verizon analyzed 855 data-breach incidents that occurred in the prior year, which resulted in 174 million compromised records; this analysis yielded some interesting statistics:

- ✔ 98 percent of the incidents stemmed from external agents (up 6 percent from the prior year).
- ✔ 85 percent took weeks to discover (up 6 percent).
- ✔ 81 percent involved some form of hacking (up 31 percent).
- ✔ 69 percent incorporated malware (up 20 percent).

## Who's at Risk?

This question has a simple, disturbing answer: Everyone. Every enterprise and government agency is a target for cyberattacks — right now, at this very moment. What's worse, most analysts estimate that at least 95 percent of large organizations are already infected. Maybe even yours.

Successful cyberattacks make headlines every day around the globe. Following are just a few high-profile attacks that occurred in 2011 and 2012 alone.

### *Commercial:*

- ✔ **Saudi Aramco** (August 2012): Hackers spread Shamoon virus to the company's network by causing data on 55,000 PCs to be erased and replacing each file with an image of a burning American flag.
- ✔ **Yahoo!** (July 2012): More than 400,000 user accounts were stolen via a cyberattack.
- ✔ **LinkedIn** (June 2012): Hackers stole a file that contained more than 6 million hashed passwords, and several hundred thousand passwords were revealed.
- ✔ **Citigroup** (June 2011): The company disclosed that a hacker stole more than 360,000 credit-card numbers, of which 3,400 were used to steal more than \$2.7 million.
- ✔ **Sony** (April 2011): The hacktivist group LulzSec claimed they had stolen 100 million PlayStation Network credentials, but later were found to have stolen 77 million. (I discuss this group and other "Hacktivists," later in this chapter.)
- ✔ **RSA Security** (March 2011): Hackers stole data related to SecurID tokens, rendering them insecure.

*Government:*

- ✓ **South Carolina Department of Revenue** (October 2012): Approximately 3.6 million Social Security numbers and 387,000 credit and debit card numbers were exposed in a cyberattack.
- ✓ **U.S. Environmental Protection Agency** (August 2012): A computer security breach resulted in the exposure of the Social Security numbers, bank routing numbers, and home addresses of more than 5,000 employees.
- ✓ **Israel** (June 2012): Several government websites were unavailable a day after the *Washington Post* published a report claiming that the United States and Israel created Flame.
- ✓ **Iran** (May 2012): Flame malware is alleged to have been developed by the United States and Israel to slow the Iranian nuclear program. Flame was described as being 20 times more complicated than the earlier malware program Stuxnet.
- ✓ **NATO** (July 2011): Hackers stole 12,000 usernames and passwords.
- ✓ **U.S. Central Intelligence Agency** (June 2011): LulzSec temporarily brought down all public CIA websites.

## Trending Cyberthreats

When it comes to cyberthreats, the only constant is change. In this section, I review some of the most common trends in the ever-changing cyberthreat landscape.

### *Shifting attackers*

Hacking has changed dramatically over the past half century. In the 1970s and 1980s, *phone phreaking* (unauthorized manipulation of telephone switching equipment primarily to place free long-distance phone calls) was the craze. The 1983 movie *WarGames* introduced the general public to computer hacking, and the legend of hackers as cyberheroes was born.

In the 1990s, widespread Internet access provided opportunities to a new generation of hackers, cracking software copy-protection, and later, defacing websites primarily for bragging rights.

Today, you can classify the motives that drive most hackers into three general categories: cybercriminals, state-sponsored hackers, and hacktivists. I discuss all three types in the following sections.

### ***Cybercriminals***

*Cybercriminals'* driving motive to hack is profit. They might hack to steal credit-card numbers in bulk (sometimes by the millions) to sell in underground markets or run operations to steal Facebook, Twitter, and/or e-mail account credentials (which are also quite profitable). Still others develop hacking tools, such as Trojans, and offer services to prospective criminal entrepreneurs, contributing to a black economy with an estimated size in the billions of dollars.

Cybercriminals bent on espionage might break into corporate networks to gain access to proprietary data, which they could then hold for ransom or sell to the victim's competitors. Still others create *malware* (a word created by shortening and combining *malicious software*; code designed to perform actions on a computer system that work against the best interests of its owner) to exploit vulnerabilities from within.

One of the most infamous cybercriminals was Albert Gonzalez. Gonzalez was convicted in 2010 of hacking into the databases of Heartland Payment Systems. He stole more than 170 million credit-card numbers over the course of two years. (That's equivalent to a little more than half the population of the United States!) He was sentenced to 20 years in prison — the stiffest sentence imposed on a cybercriminal to date.

### ***State-sponsored hackers***

Perhaps the most notable shift in the hacking community over the past decade has been the emergence of *state-sponsored hackers*. These hackers are employed by nations to break into government and/or commercial computer systems in other nations to achieve political objectives.

The acts committed (or allegedly committed) by one nation against another are often referred to as *cyberwarfare* (or *electronic warfare*), committed by so-called *cyberwarriors*. Cyberwarfare pertains to stealing data, committing espionage, and crashing computer systems, but it can also result in loss of life. A state-sponsored hacker could, in theory, penetrate the network security defenses of a nuclear power plant causing a widespread nuclear meltdown.

### ***Hacktivists***

*Hactivism* is the use of computers as a means of protest or to promote political ends. Unlike state-sponsored hackers, *hacktivists* serve themselves by leveraging hacking techniques to deface public websites, redirect traffic away from them, crash them (through so-called web sit-ins), or steal confidential data. Two high-profile hacktivist groups are Anonymous and LulzSec.



Hacktivists often employ denial-of-service attacks to temporarily disrupt public websites. To learn more about this hacking technique, jump to the “Denial-of-service attacks” section later in this chapter.

## ***Broader reach, bigger risk***

Every host on your network is vulnerable to an attack. I don’t care how secure you think your network is: If a human can access it via the network, any host can be compromised. And every host *can* be compromised. Think of your network as being one large attack surface. The larger and more geographically dispersed it gets, the easier it is to penetrate.

Aside from the natural growth of your network due to the expanding nature of your business or government agency, several networking trends are causing your network’s attack surface to expand at a rapid pace.

### ***Social media***

The recent tidal wave of social media has created new opportunities — and new risks — for business and government agencies. Popular social media sites such as Facebook, Twitter, and Google+ introduce risks that IT departments must evaluate and mitigate.

Social media sites continue to be conduits for malware. A well-intentioned user may check out her Facebook page during lunch, click a link (posted by a trusted friend), and infect her computer with malware that is so new, no antivirus can yet detect it. Or an ill-intentioned employee might use Facebook to sneak confidential data out of the company.

## ***Virtualization***

Mainstream adoption of virtualization platforms (such as VirtualBox, VMware, and Xen) has reached critical mass. Nearly every large organization leverages virtualization to host mission-critical applications in the data center.

Virtualization, however, poses a few inherent risks that don't apply to physical hosts:

- ✔ IT can't natively inspect traffic between virtual machines (VMs) without the use of a specialized tool.
- ✔ Many VMs go unprotected because IT hasn't yet budgeted for virtual security protection.
- ✔ New virtual hosts are frequently pushed into production without the knowledge (or approval) of IT security — a problem commonly known as *VM sprawl*.

## ***Cloud computing***

The explosion of *cloud computing* (applications delivered as services over a computer network; also known as *Software as a Service*, or *SaaS*) has caused new concerns for IT security. Whether applications are deployed via a public cloud, a private cloud, or a hybrid, data can be breached just as easily through a cloud architecture as it can through a traditional computer network unless proper security measures are taken.



When working with a cloud vendor, be sure to ask about its IT security defenses. Although the onus of security certainly falls on the vendor's shoulders, assuming it has implemented appropriate security measures can prove costly.

## ***Mobile devices***

Many people consider mobile devices, such as smartphones and tablets, to be the next frontier for cyberattacks. This trend coupled with the BYOD (Bring Your Own Device [to work]) movement causes immense concerns for IT. It is

already a challenge for IT to secure the devices that the organization owns, and even harder and more complicated to secure privately owned devices connecting to the network.

As a result, attacks on mobile devices — especially those targeting the Android and iOS operating systems — are growing rapidly in number and sophistication.

## *Modern malware*

As I note earlier in this chapter, Verizon reported that malware was behind 69 percent of the data breaches they investigated in 2011 — a 20 percent rise above the preceding year's levels. Every IT security organization should be concerned about malware, so in this section I explore the most common types of malware on computer networks today.

### *Worms and Trojans*

A computer *worm* is a stand-alone malware program that replicates itself over a network in order to propagate. Worms typically harm networks by consuming bandwidth, but also provide a “lateral” attack vector that may result in infecting supposedly protected internal systems or by exfiltrating data. Unlike a computer virus, a worm doesn't append itself to other programs; worms typically exploit vulnerabilities in operating systems to replicate and spread.

A *Trojan* (or *Trojan horse*) masquerades as a legitimate file or helpful program, with the ultimate purpose of granting a criminal unauthorized access to a computer. Trojans may self-replicate within the infected system, but can't propagate to other vulnerable computers on their own; they typically join networks of other infected computers (called *botnets*) where they wait to receive further instructions, and into which they submit stolen information. Trojans may be delivered by means of spam e-mail or drive-by downloads. Or they may be deliberately disguised as a pirated installer for a well-known game or popular application hosted on peer-to-peer file sharing networks in an effort to reach target computers.

### *Spyware*

*Spyware* collects information about users without their knowledge. As its name suggests, the presence of spyware is typically hidden from the user, so it can be difficult to detect.

This type of malware can collect almost any type of data, including web surfing habits, user logins, and credit-card numbers. In the case of *adware*, a subset of spyware, the software tracks your web browsing or other computer activity and then produces targeted ads, the origin of which can be difficult to determine.

## ***Buffer overflows and SQL injections***

Two commonly used hacking techniques that exploit vulnerabilities in web-based applications are buffer overflows and SQL injection attacks.

A *buffer overflow* (or *buffer overrun*) is an anomaly in which a program writes more data into a memory buffer than the buffer is designed to hold, some of which spills into adjacent memory, causing the application to perform incorrectly or even crash. Buffer overflows are commonly triggered by hacker inputs that are designed to execute code or alter the way that the program operates.

An *SQL injection* attacks databases through a website or web-based application. The attacker submits SQL statements into a web form in an attempt to get the website (or web application) to pass the rogue SQL command to the database. If successful, an SQL injection attack can reveal database content (including credit card and Social Security numbers, passwords, or other confidential data) to the attacker.



The most reliable way to mitigate buffer overflows and SQL injection attacks is to patch the vulnerability at the application level. In many cases, however, this fix can't be performed on legacy code. In such cases, an intrusion prevention system (IPS) or web application firewall (WAF) can help.

## ***Botnets***

A *botnet* is a collection of compromised Internet-connected computers on which malware is running. Each compromised device is called a *bot* (or *zombie*), and the human controlling a botnet is called the *bot herder* (or *botmaster*). Command and control of a botnet typically involves web servers operated for the specific purpose, although some older bots use Internet relay chat (IRC) to control the botnet. Bots may also be used to commit denial-of-service attacks (which I discuss later in this chapter), relay spam, and/or download additional malware to the infected host computer.

### ***Polymorphic malware***

*Polymorphic malware* comprises cyberthreats that change form in order to evade detection. Evolution of the malicious code can occur in a variety of ways, such as through filename changes, the addition of *padding* (strings of varying lengths), and/or the use of compression, packers, and encryption. Although the appearance of the code varies with each mutation, its core functionality remains the same. By some accounts, more than 90 percent of today's malware employs some form of polymorphism as a tactic to evade detection.



*Detonation* is a modern-day strategy for defeating polymorphic obfuscation. The practitioner intentionally executes the suspected file in a safe environment (called a *sandbox*) and then employs behavioral analysis to determine a file's degree of maliciousness.

### ***Denial-of-service attacks***

A *denial-of-service (DoS) attack* is an attempt to make a network resource unavailable to its intended users, either temporarily or indefinitely, by inundating the resource with external communication requests. Common DoS attacks involve flooding servers with large volumes of malformed requests or excessive traffic (such as the archaic smurf attacks or ping floods). Today's DoS attacks target the application layer, such as the web server, using botnets, or tools like SlowLoris or LOIC (Low Orbit Ion Cannon). Perpetrators of DoS attacks typically target sites or services hosted on high-profile web servers, often for political gain.

A *distributed denial-of-service (DDoS) attack* is identical to a DoS attack except that the flood of communication requests comes from botnets, often on tens of thousands of hosts, all programmed to attack a target host at precisely the same time.

These threats are reasonably easy to identify with signature-based defenses (see Chapter 2), but plenty of other attacks are far more difficult to detect. For details, keep reading.

### ***Zero-day attacks***

A *zero-day attack* is an attack on an unknown operating system or application vulnerability, so named because the attack occurs on “day zero” of awareness of the vulnerability. Thus,

the developers of the operating system or application have had zero days to patch the vulnerability. Zero-day attacks are extremely effective, because they often go undetected for long periods, and when they are finally detected, repairing the vulnerability still takes days or even weeks.



IPS devices equipped with vulnerability-based signatures (rather than exploit-based signatures) are best equipped to mitigate zero-day threats, because IPS vendors may release new signatures before the vendor has even published its patch. But until a new vulnerability has been identified, every IPS is 100 percent defenseless against corresponding zero-day attacks.

## ***Social engineering attacks***

Social engineering attacks have been around for decades. In many cases, criminals have found that it's easier to trick someone into clicking a link, opening a file, or giving you their password than it is to crack into that system.

A few types of computer-based social engineering attacks are increasingly common:

✓ **Phishing:** *Phishing* is an attempt to acquire information (and, indirectly, money) such as usernames, passwords, credit-card information, and Social Security numbers by masquerading as a trustworthy entity in an electronic communication, such as e-mail. Phishing is most often carried out by e-mail *spoofing* — directing users to enter personal details on a fake website that looks and feels almost identical to the legitimate one.

Phishing can be specialized, as follows:

- **Spear phishing:** *Spear phishing* targets a specific person or company. Attackers may gather personal information about their target ahead of time to increase the probability of their success.
- **Whaling:** *Whaling* is directed specifically toward senior executives and other high-profile targets within an organization.

✓ **Baiting:** In a *baiting* attack, a criminal casually drops a USB thumb drive or CD-ROM in a parking lot or cybercafe. This drive or disc is labeled with some intriguing title like Executive Salaries. When the victim accesses the drive or disc, it installs malware on his or her computer.

### *APTs and advanced targeted attacks*

*Advanced persistent threats (APTs)* are sophisticated cyberthreats that leverage Internet-enabled espionage or attacks using a variety of intelligence-gathering techniques to access sensitive information. Other recognized attack vectors include infected media, supply chain compromise, and social engineering. Individuals, such as an individual hacker, aren't usually referred to as an APT because they rarely have the resources to be both advanced and persistent even if they're intent on gaining access to, or attacking, a specific target.



*Advanced targeted attacks (ATAs)* define sophisticated tactics and techniques involving reconnaissance of an attack target, as well as the attack methods employed by the attackers, to penetrate the target network. APT malware is designed to remain undetected for extended periods of time.

ATAs, by their very nature, involve far more effort and research on the part of the attacker than (for example) a common scattershot spam campaign, and take place over a protracted period of time. ATAs employ both social engineering and technological attacks typically targeted against specific (usually privileged) users within networks, and may rely on a combination of previously stolen private information and public-source knowledge about a business or government entity to identify those users as potential entry points into the targeted network.

Attackers' motives may fall into any of the three general categories described earlier in this chapter, but are most commonly associated with government or industrial espionage for political or economic gain. The types of information a criminal targets in an ATA may not be easily identified or quantified after the fact — as, for example, a database of credit card numbers. Hacktivists have also been known to engage in long-term attacks against perceived enemies or groups hostile to their interests.



APTs and ATAs are extremely difficult to detect because they use *low and slow* tactics to avoid detection. The attacker may gain access to the network by means of social engineering techniques to deliver malware to users of targeted systems. Once installed, the attacker will commonly attempt to map the network from the inside to identify other potentially vulnerable hosts. After such hosts are identified, attackers attempt to access these systems, capture information over an extended

period of time, and remain undetected. Stolen information sent back to the criminal operation may be batched into small chunks, delivered at random time intervals — sometimes over an encrypted channel.

To counter the threat of APTs and ATAs, a new generation of IT security products designed to detect and mitigate these highly sophisticated attacks has begun to emerge. Fortunately, Big Data Security solutions are at the forefront.

## *The Cost of Failure*

As you can well imagine, the cost of a data breach can be immense. In some cases, it can kill a company. Companies that have been victimized by a large-scale data breach also face enormous costs in the following categories:

- ✓ Investigation and forensics costs
- ✓ Customer and partner costs
- ✓ Public relations costs
- ✓ Lost revenue due to damaged reputation
- ✓ Regulatory fines
- ✓ Civil claims and legal fees

According to a study published in 2012 by Ponemon Institute ([www.ponemon.org](http://www.ponemon.org)) in which 400 individuals were interviewed from 49 U.S.-based organizations that experienced actual data breaches (in 2011), the average total organizational cost per data breach was \$5.5 million. That equates to \$194 per compromised record!

Enterprises can't afford to miss out on the considerable data protection offered by Big Data Security solutions. Simply put, it's a no-brainer.

## Chapter 2

---

# Why Traditional Security Isn't Enough

.....

### *In This Chapter*

- ▶ Understanding basic threat detection techniques
  - ▶ Reviewing today's traditional security defenses
  - ▶ Realizing why traditional security defenses sometimes fail
- .....

Some security pundits suggest that the network perimeter is dead. Well, I say that's a gross exaggeration. As long as hosts connect to the Internet through a common gateway, you'll always have a network perimeter, and you'll always have to defend it. To give these pundits credit, however, IT departments don't always invest enough resources in monitoring internal threats that bypass perimeter defenses — threats that come through mobile devices, portable media, 3G/4G connections, and more.

Because no cybersecurity solution is foolproof (no matter what any vendor says), the best approach to security is *defense in depth*. This approach involves applying layers of best-of-breed endpoint security, network security, and Big Data Security (more about this in Chapter 3) solutions to detect, prevent — and in some instances, analyze — ongoing cyberthreats.

In this chapter, I want to make sure that you're grounded in the traditional cybersecurity defenses that should comprise your defense-in-depth strategy, along with an understanding of how they detect threats. But more importantly, I want you to understand why these traditional security defenses sometimes fail and why more are needed to keep up with today's sophisticated attacks.

## Basic Threat Detection Techniques

Security products that are designed to detect (and sometimes block) cyberthreats incorporate one or more of the following basic threat detection techniques: threat signatures, blacklisting, whitelisting, and behavioral profiling. To explain these concepts, I use a real-world analogy with air travel:

- ✔ **Threat signatures:** Before your checked baggage can board an airplane, it is scanned using a highly sensitive explosives-detection system that looks quite similar to the CT scanners you'd find at your local hospital. This machine is equipped with sophisticated pattern-matching software designed to detect many different explosive devices. In the IT world, this is analogous to signature-based antivirus clients and network intrusion prevention systems inspecting network traffic looking for known malware and exploits.
- ✔ **Blacklisting:** The U.S. Transportation Security Administration maintains a “no-fly” list to help keep known terrorists off airplanes. In IT, this is known as *blacklisting* — wherein communications with known-bad Internet hosts are flagged for investigation or blocked.
- ✔ **Whitelisting:** U.S. Customs and Border Protection has a Global Entry program that enables frequent international travelers (Americans and foreigners from select countries) to receive expedited clearance through airport immigration upon arriving in the United States. In the IT security world, this is known as *whitelisting* when communications from specific Internet hosts are approved in advance and free from more detailed inspection.
- ✔ **Behavioral profiling:** Airport law enforcement personnel are trained to observe and question travelers who act suspiciously — even after they've passed through airport security. In IT security, *behavioral profiling* is when tools attempt to detect anomalies from a baseline of “normal” network traffic (as in the spread of malware).

Now that you understand the primary ways that security products detect threats, read on to review the traditional security defenses that incorporate these threat detection techniques to form your defense-in-depth strategy.

## *Traditional Security Defenses*

The traditional endpoint and network security defenses described in this section are commonly used in today's enterprise and government networks.

### *Endpoint security*

*Endpoint security* enables a computing device to assume at least some responsibility for its own security. Endpoint security systems work on a client/server model: Client software is installed on each network device, and a central server (or gateway) monitors and/or manages the client software installed on the devices.

Following are a few familiar examples of endpoint security:

- ✓ Virus, malware, and spyware prevention
- ✓ Personal firewalls
- ✓ Spam filtering
- ✓ URL filtering
- ✓ Application controls
- ✓ File integrity monitoring

Vendors of these solutions include AVG, ESET, Kaspersky Lab, McAfee, Microsoft, Panda, Sophos, Sourcefire, Symantec, Trend Micro, and Tripwire.



The remaining security products described in this section all fall under the category of network security devices.

## *Intrusion prevention systems*

An *intrusion prevention system* (IPS) monitors network traffic for malicious activity. If an IPS is configured for *inline* operation (that is, if it's in the direct path of flowing traffic), it can block threats. If, however, the IPS is configured for *passive* operation (that is, it merely inspects copied traffic), it's capable only of providing alerts when it detects threats.



The latter mode of operation is also referred to as *intrusion detection system* (IDS) mode. Today, IDS is a mode of operation within an IPS rather than a unique product offering.

IPS solutions incorporate thousands of signatures to detect both known and unknown threats to operating systems and applications. Better IPS offerings also incorporate anomaly-based detection methods (see “Network behavior analysis,” later in this chapter) and stateful protocol analysis.

Vendors include Cisco, HP (Tipping Point), IBM, Juniper, McAfee, and Sourcefire.

## *Next-generation firewalls*

A *next-generation firewall* (NGFW) is a multifunction security appliance that incorporates firewall, IPS, and application control processes into a unified network security platform. Many NGFW solutions also offer URL filtering subscriptions to restrict web traffic.

Enterprises are turning to NGFWs to reduce network security costs — and, perhaps more important, to implement full-featured application control, as described in the nearby sidebar “Not your father’s IPS.”

Vendors include Check Point, Dell (SonicWALL), Fortinet, McAfee, Palo Alto Networks, and Sourcefire.

## *Secure e-mail gateways*

A *secure e-mail gateway* (SEG) monitors inbound e-mail traffic for spam, viruses, malware, and other threats. It also monitors outbound e-mail traffic for sensitive data such as credit-card, Social Security, and bank-account numbers.

## Not your father's IPS

In 2011, IT research firm Gartner defined a new category of IPS solutions that go beyond traditional first-generation IPS capabilities. This new category, called *next-generation network intrusion prevention system (NGIPS)*, incorporates application, contextual, and content awareness capabilities that enable IT security to detect more threats with lower operating expenses.

Better NGIPS offerings correlate threats against endpoint intelligence to reduce the number of actionable

security events and to automate IPS tuning (modifying the IPS detection policy). These solutions also incorporate user identity tracking to identify targeted hosts with more than just an IP address.

Some NGIPS products even offer full-featured application control, enabling organizations to reduce their network's attack surface (see Chapter 1) by allowing access only to safe, IT-approved Internet-facing applications.

Vendors include Cisco, McAfee, Microsoft, Proofpoint, and Symantec.

## Secure web gateways

A *secure web gateway (SWG)* monitors inbound web traffic for malware and other cyberthreats; it also performs URL filtering on outbound traffic to restrict web browsing to safe, IT-approved websites. Better SWG solutions offer granular policy-based control of web-based applications, such as instant messaging, multiple-player games, peer-to-peer applications, and Voice over IP (VoIP) applications.

Vendors include Blue Coat, Cisco, McAfee, and Websense.

## Data loss prevention

*Data loss prevention (DLP)* (also called *data leakage prevention*) is designed to detect (and, in some cases, prevent) potential breaches by inspecting data in use, in motion, and at rest. DLP solutions can be configured to search for two types of data:

- ✓ **Described:** Data formatted with a fixed schema, such as credit-card and Social Security numbers
- ✓ **Registered:** Data such as a specific document, spreadsheet, block of text, or database record

Vendors include CA, Fidelis, McAfee, RSA, Symantec, Verdasys, and Websense.

## *Network behavior analysis*

*Network behavior analysis* (NBA — formerly network behavioral anomaly detection, or NBAD) is a sophisticated network security system that baselines normal network traffic to detect anomalies, such as advanced persistent threats and advanced targeted attacks (see Chapter 1). It does this by processing network flow records (such as NetFlow, cFlow, jFlow, sFlow, and IPFIX) generated by routers and switches and then applying sophisticated algorithms to those records to uncover potential threats.

Vendors include Arbor Networks and Lancope.

## *Advanced malware protection*

A new category of IT security products detects advanced new threats that slip through signature-based security defenses. Although the industry has no standard name for this category yet, it's most often referred to as *advanced malware protection*, or sometimes *next-generation threat protection*.

An advanced malware protection system uses sophisticated algorithms to detect suspicious files and routes them to a *sandbox* (a virtual machine configured to emulate the target environment) for detonation and monitoring. If the file proves to be malicious, the system blocks subsequent malware instances and communications.

Vendors include Cuckoo, Damballa, FireEye, Norman, and Sourcefire.

## ***Security information and event management***

*Security information and event management (SIEM)* systems frequently serve as central points for managing and analyzing events from network security devices across large distributed enterprises and government agencies. A SIEM has two primary responsibilities: aggregating events and logs from network devices and applications and using that intelligence to uncover network problems. SIEMs are useful for uncovering threats within the data they're configured to receive, but that data only represents a fraction of the data available to Big Data Security solutions.

Vendors include ArcSight (HP), LogRhythm, McAfee (Nitro), NetIQ, Q1 Labs (IBM), and Splunk.

## ***New Offenses versus Traditional Defenses***

No cybersecurity solution is foolproof. Here are common reasons why traditional security defenses sometimes fail.

### ***Hand-carried mobile devices***

Perimeter network security devices such as IPS and NGFW are great at blocking known threats in inspected traffic. But every day, people bypass perimeter security defenses the old fashioned way: They carry laptops and mobile devices through the front door. Once connected to the network, malware that an employee unintentionally downloaded over the weekend can spread rapidly through the network.

### ***Unknown and zero-day threats***

Signature-based security products have zero chance of catching zero-day threats. (See Chapter 1 for a refresher on zero-day threats.) Advanced malware protection devices with sandboxing technology have the best shot at catching new

malware, but this technology only inspects certain types of traffic copied from specific portions of the network. By the time you discover new malware, some amount of damage has already been done.

### *Encrypted threats*

Most network security devices are blind to encrypted traffic, and with up to a third of an organization's Internet traffic encrypted, this blind spot is far too big for any organization to live with. To mitigate this vulnerability, traffic must be decrypted before inspection; then the clean traffic should be reencrypted before being sent to its final destination. (Reencryption is especially important for organizations that are required to maintain PCI compliance.)



Some, but not all, network security devices offer onboard SSL decryption capabilities. For those that don't, organization should acquire stand-alone SSL decryption solutions. Stand-alone solutions not only provide plug-and-play functionality, but also offload the computationally heavy task of SSL decryption and key generation, and are worth considering for large environments.

### *Lack of context and forensics*

Network security devices, such as IDSs, are effective only if analysts can discern actionable events from irrelevant ones, and to do that, they need contextual awareness and forensics analysis capabilities. An analyst needs to know whether a threat is applicable to the target host's operating system and/or applications; if so, she needs to perform real-time analysis to obtain answers to the following questions:

- ✓ What data has been compromised?
- ✓ What systems are involved?
- ✓ How was the attack carried out?
- ✓ How can we be sure the attack is over?

How can security analysts answer these questions quickly and improve the effectiveness of their security infrastructure? Read on to discover the power of Big Data Security.

## Chapter 3

# Meet Big Data Security

### *In This Chapter*

- ▶ Defining Big Data in the context of information security
- ▶ Contrasting limited- and full-visibility Big Data Security solutions
- ▶ Exploring the features and benefits of security intelligence and analytics

**T**o succeed in information security, you must come to terms with a universal truth: No matter how hard you try, or how much money your organization is willing to spend, your network *will* be compromised at some point. (In fact, it probably already has been!)

The security defenses that I discuss in Chapter 2 offer at least some protection against a wide variety of threats, but they're simply not enough in today's world. Luckily, you already have what you need to improve your situational awareness, provide context, and make your existing defenses more effective.

I'm talking, of course, about your organization's Big Data.

## *What Is Big Data?*

*Big Data* is a collection of data sets so large and complex that they're awkward to work with when you use traditional database management and analysis tools. Big Data challenges include capturing, storing, searching, sharing, analyzing, and visualizing large data sets.

Big Data isn't unique to information security. It's applicable to a myriad of use cases, including scientific discovery, economic analysis, business intelligence, counterterrorism, and fraud detection. Because this book is about Big Data Security, I limit my interpretation to the realm of information security, starting with common Big Data sources.

### *Internal Big Data sources*

Typical internal sources of Big Data include:

- ✓ All IP traffic flowing across your network, including web traffic, e-mail, and file transfers/attachments
- ✓ Network flow records (such as NetFlow, cFlow, jFlow, and sFlow) from network routers and switches
- ✓ VM-to-VM (virtual machine to virtual machine) IP traffic on VMware, Xen, and other virtualization platforms
- ✓ User account directories, such as Microsoft Active Directory and LDAP
- ✓ Detonation and behavioral analysis result feeds from malware “kill chain” analysis appliances, such as FireEye or Norman Malware Analyzer

### *External Big Data sources*

Typical external sources of Big Data include:

- ✓ Cyberthreat and reputation feeds, such as Emerging Threats, Google Safe Browsing, Malware Domain List, SANS Internet Storm Center, SORBS (Spam and Open-Relay Blocking System), VirusTotal, and other spam or IP address blacklists
- ✓ IP geolocation services, such as Digital Envoy, Geobytes, MaxMind, and Quova
- ✓ Website intelligence services, such as DomainTools, Robtex, and the global domain registry database

Now that you have a foundation of internal and external Big Data sources, you're ready to delve into commercially available solutions that can harness these sources.

# *What Is Big Data Security?*

*Big Data Security* refers to any computer-based solution (combination of hardware and software) that captures and analyzes some or all Big Data sources for the purposes of uncovering and mitigating cyberthreats.

Some commercial Big Data Security solutions incorporate most of the Big Data sources mentioned in the previous section into one highly scalable platform. Solutions of this type — called *full-visibility solutions* — are ideal for uncovering unknown or hard-to-detect cyberthreats and for performing forensic analysis after a cyberattack occurs. Other Big Data Security solutions — called *limited-visibility solutions* — incorporate only some of the Big Data sources and are suited for uncovering known or easy-to-detect threats.

## *Limited-visibility solutions*

Examples of Big Data Security solutions that have a relatively limited view of the network include SIEM and NBA, both of which I explain in this section.

### *Security information and event management*

A reliable *security information and event management (SIEM)* system is critical to the success of any enterprise information security organization. SIEMs complement existing IT security investments — firewalls, intrusion prevention system (IPS) and network access control (NAC) appliances, data loss prevention (DLP) solutions, secure web gateways, endpoint security solutions, and so on — by aggregating and correlating log and event data at a central location.

A SIEM can be classified as a Big Data Security solution, but its visibility is limited to the data that it is fed, such as log, flow, and endpoint event data.

### *Network behavior analysis*

A *network behavior analysis (NBA)* solution also could be classified as a Big Data Security solution, because it aggregates, profiles, and inspects massive numbers of flow records (such as NetFlow, jFlow, sFlow, and cFlow) from network routers and switches.

NBA solutions may also detect cyberthreats based on behavior profiling (see Chapter 2). To detect this type of traffic, these solutions baseline normal network traffic so that they can detect anomalies such as malware propagation or the export of massive amounts of data from a rogue host. They also offer numerous benefits for network operations personnel, including the capability to troubleshoot network outages and performance degradations.

As useful as NBA solutions are, they can inspect only the traffic flow information they're configured to receive. As a result, NBA may not provide visibility into the full traffic payload, including user and application data.

### *Full-visibility solutions*

What if you could have a Big Data Security solution that truly sees everything on your network? Imagine a giant digital video recorder (DVR) that records every TV show on every channel, and stores the recordings for weeks at a time. Now imagine a solution that not only records every packet, flow, and file that traverses your network, but also provides built-in traffic-analysis capabilities.

Well, I'm pleased to tell you that such a solution exists, and it's available to you right now! It's a relatively new category of information security technology called *security intelligence and analytics*.



From this point forward, when I discuss the merits of Big Data Security, I'm referring primarily to security intelligence and analytics, because it offers the greatest breadth and depth of Big Data analysis for information security.

## *Introducing Security Intelligence and Analytics*

*Security intelligence and analytics (SIA)* is what I consider to be the ultimate Big Data Security solution. By capturing every packet, flow, and file that traverses your network, SIA provides an additional layer of defense by detecting threats that are virtually invisible to traditional defenses.

Unlike traditional network security devices that inspect only a portion of your traffic and capture packets only at the point of attack, SIA solutions capture and inspect 100 percent of all network traffic — before, during, and after the attack.

SIA solutions perform several distinct functions, including the following:

- ✓ Situational awareness
- ✓ Cyberthreat detection
- ✓ Security incident response
- ✓ Data loss monitoring and analysis
- ✓ Organizational policy compliance verification
- ✓ Security assurance (think always-on verification of the effectiveness of your other security tools)



Chapter 4 provides detailed explanations of these six SIA use cases.

## *How SIA works*

An SIA system is the computer-network equivalent of a closed-circuit security camera system. It's always on, recording network activity 24 hours per day, 7 days per week. Your SIA system never sleeps.

SIA solutions are designed for the enterprise — capturing and indexing data (including packet header and payload, OSI layers 2 through 7) at wire speed, providing a complete, forensically sound record of all network activity. SIA solutions also have built-in tools to perform real-time or back-in-time analysis of files, applications, flows, or packets. These solutions must have ample storage capacity because they record and store terabytes of data for days, weeks, or even months at a time.

## *Form factors in SIA*

Leading SIA providers offer solutions in three form factors: physical appliances, virtual appliances, and software. These multiple choices enable an organization to select the ideal platform for its networking environment.

### *Physical appliances*

Most organizations choose to deploy SIA solutions on purpose-built, rack-mountable turnkey appliances (see Figure 3-1). These appliances are selected by the vendor to achieve a given level of throughput and performance. IT departments don't need to worry about selecting appropriate hardware or installing and configuring the software, because it incorporates a purpose-built hardened and preinstalled operating system.



**Figure 3-1:** Sample SIA appliance from Solera Networks.

### *Virtual appliances*



To maintain complete visibility of your network, I suggest that you also implement virtual SIA appliances to inspect VM-to-VM (virtual machine to virtual machine) traffic in your virtual network infrastructure (VMware ESX, Citrix XenServer, Windows Hyper-V, and KVM). Otherwise, VM-to-VM traffic results in a blind spot on your network, and you're not leveraging your SIA system to its fullest potential.

### *Software*

Some organizations prefer to do their own selection of the hardware that houses the SIA software. (In fact, some government agencies are required to choose hardware from an approved-products list.) These organizations can opt to deploy SIA software on their own hardware platforms; if so, their IT personnel must install, configure, and maintain the system and ensure that it's optimized for their required performance level.

## Common features of SIA solutions

Not all SIA solutions are created equal, so it's important to assess both the basic and advanced features that are important for your organization. Following are descriptions of some features offered by virtually all SIA solutions.

### Customizable dashboard

The SIA dashboard (see Figure 3-2) is the primary user interface for monitoring the system and investigating potential threats. Better SIA solutions offer a selection of prebuilt dashboards and a library of drag-and-drop widgets (small boxes of data) so that users can customize the dashboard to their liking. A widget may contain summary data within a table, for example, or a pie, bar, or column chart.

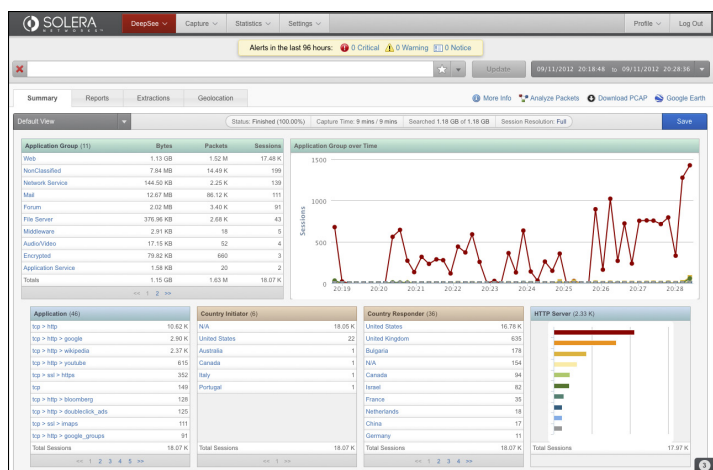


Figure 3-2: Sample SIA dashboard.

### Rules and alerts

Most SIA offerings enable users to configure rules and alerts. A *rule* is an action that can be taken when recorded traffic meets a certain condition, such as routing captured traffic associated with a known-bad site to an advanced malware protection device for further analysis. An *alert* is simply a notification that one of your rules has been triggered. Alerts

are often displayed on the dashboard but may also be sent via e-mail or an SMS text message.

## ***Comprehensive reporting***

Today's SIA platforms feature powerful reporting capabilities and provide a library of easy-to-read report templates that IT departments can customize to meet their needs. SIA solutions commonly feature reports that include:

- ✓ Identification of the network traffic generated by common Web-based applications, such as webmail
- ✓ E-mail sender and receiver addresses, and the subject lines of messages
- ✓ Usernames, nicknames, or accounts on instant messaging, chat, and social media sites (*social personae*)
- ✓ File characteristics, including names, MIME types, content disposition, and transport method
- ✓ IPv4 and IPv6 source/destination addresses and their address space geolocations
- ✓ HTTP server names, referrers, web queries, port numbers, SSL common names, and full URLs

## ***Reputation services and file analysis***

Modern SIA solutions usually provide one or more OSINT (open source intelligence) or commercial reputation and malware threat feeds. By simply right-clicking elements within the browser interface, a security analyst can check the integrity and reputation of any URL, IP address, file, or e-mail address against multiple services at the same time. Examples of reputation, threat feeds, and file analysis services include Google SafeBrowse, Robtex, the SANS Internet Storm Center, VirusTotal, Bit9, Team Cymru, and Norman (see “External Big Data sources,” earlier in this chapter).

## ***Query favorites***

Just as web browsers enable you to save favorites, modern SIA applications enable users to save custom search queries (sometimes called *filters*) for future use. Users can quickly execute favorite search queries to quickly detect threats, malware, or suspicious traffic.

## Metadata retention

In addition to analyzing several days' worth of full packet data, security and incident responders often want to perform long-term analysis of network traffic to evaluate anomalous or suspicious behavior. Unfortunately, storing a full year's worth of packet data is rarely realistic, given the amount of data involved.

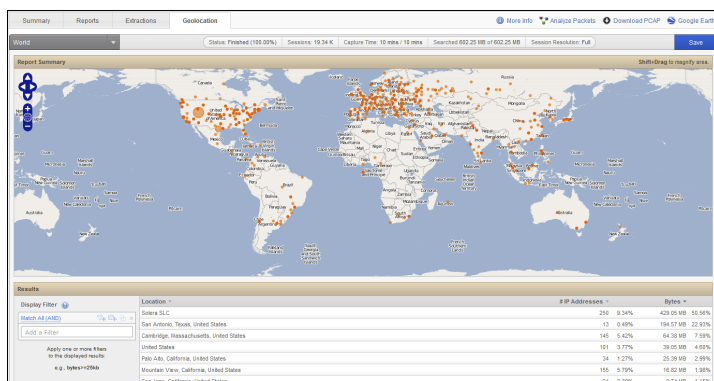
SIA systems, however, provide metadata retention, enabling security analysts to devote a portion of storage to full-packet capture and another portion to metadata. This feature allows analysts to optimize their systems to retain an appropriate amount of both full-packet data — perhaps a week's or month's worth — while still allocating enough space for a year's worth of network metadata.

## Advanced features of SIA solutions

The preceding features are available in most SIA offerings, but you'll find the following only in more advanced SIA solutions.

### Geolocation

*Geolocation* is the practice of assessing the real-world geographic location of an Internet-connected computer or device. Preferred SIA solutions offer geolocation integration, which enables users to view the origin, destination, and flow of network traffic (see Figure 3-3).



**Figure 3-3:** Geolocation view of external traffic sources.

Geolocation enables analysts to identify patterns and concentrations of traffic traveling to and from unusual or unexpected locations (such as countries where you have no business dealings). Users can zoom in on specific paths and flag IP addresses, locations, or even countries that appear to be suspicious. Some SIA solutions even allow users to import data directly into Google Earth!

### ***Root-cause exploration***

When a suspected network intrusion occurs, time is of the essence. To accelerate network forensic tasks, preferred SIA solutions offer a capability commonly referred to as *root-cause exploration*. This feature automates the identification of an actual session or file that caused a security incident. By enumerating these events, the feature helps analysts quickly identify the source of an infection or network breach, thereby reducing time to resolution drastically.

### ***File reconstruction***

Many SIA solutions allow users to reconstruct original documents (such as Microsoft Word documents or PDF files), images (such as JPG files), or executables that traversed the network. Better solutions have additional reconstruction capabilities, such as e-mail, chat messages, web pages, and file transfers over tunneled protocols. Full-event reconstruction is possible because every packet is recorded, classified, and indexed. This is still a massive amount of data to sift through and only the better SIA solutions can perform reconstruction at real-time speeds. This capability enables incident responders to act as soon as the threat is detected, rather than having to wait until the attack is over to respond.



File reconstruction is also very useful following a security breach to determine what, if any, confidential information has been extracted from the network. When faced with having to publicly report a breach, having this information can save organizations millions of dollars. What if only 10 database records were lost and not all 10,000,000?

### ***Web, e-mail, and chat reconstruction***

Some sophisticated SIA solutions enable analysts to view web pages exactly as users originally saw them. Analysts can also review instant-message (IM) and e-mail conversations in their original form for clues to the source of a security event.



Many cyberthreats begin with a compromised website, a malicious file, or a link in an e-mail or IM. When you evaluate SIA solutions, consider reconstruction (of web pages, e-mails, and chat messages) to be a must-have feature.

### ***Third-party integration***

No IT security solution should work in a vacuum. Rather, security solutions should work in concert to simplify monitoring, increase effectiveness, and reduce total cost of ownership. Better SIA solutions offer application programming interfaces (APIs) to integrate with popular SIEM, IPS, next-generation firewall (NGFW), advanced malware protection, and unified threat management (UTM) solutions. (For details on these information security solutions, see Chapter 2.)

Some SIA providers also offer a universal-connector client application that integrates with popular web browsers. If you see something suspicious while you're using a traditional IT security product, you can click on the universal-connector sidebar to instantly view the network activity you want to investigate within your SIA application. Universal connectors save analysts time and effort by making it easy to conduct quick analysis of suspicious indicators.

## ***Deploying SIA***

After you choose the best possible SIA solution for your organization, with just the right combination of standard and advanced features, you face a new task: deployment. This section gives you some pointers.

### ***Why size really matters***

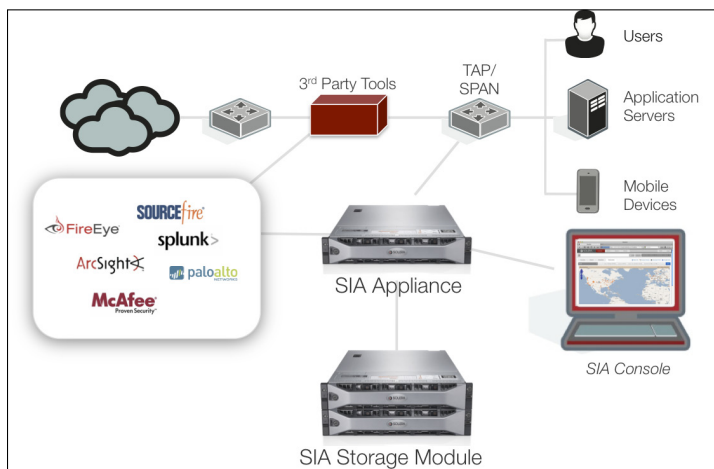
Whether you're deploying your SIA solution on purpose-built appliances (as most organizations do) or on your own hardware platform, make sure that you've got enough horsepower (CPU, memory, and storage) to accommodate your organization's requirements for sustained throughput and archiving. If your hardware falls short, some element of operation could be compromised, and you might not have all the necessary data for real-time analysis.

## Leveraging SPAN ports and TAPs

A common way for your SIA appliance to gain full network visibility is to connect it to the *SPAN ports* on your network switches. SPAN ports replicate all traffic flowing through a switch, typically for the benefit of network security and performance tools. However, enabling SPAN ports can negatively impact the performance of highly utilized switches.

If your SPAN ports have already been allocated to other monitoring tools — a problem commonly referred to as *SPAN-port contention* — or your switches are operating at capacity, you can use a network TAP to tap into your network (typically, between a router and a switch or between two switches) and replicate traffic to your SIA appliance.

Figure 3-4 shows a typical SIA deployment, featuring an SIA appliance connected to the network through a network TAP. This diagram also features an additional SIA storage array — an option commonly available from preferred SIA vendors.

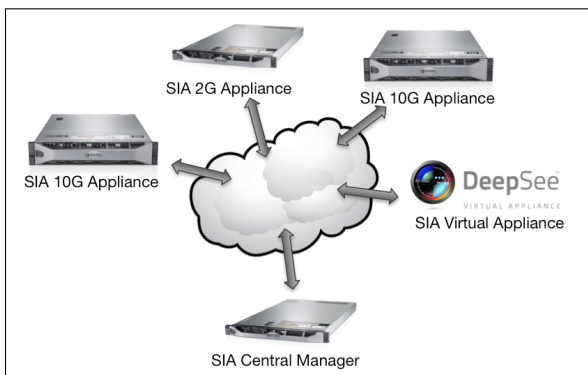


**Figure 3-4:** Typical SIA deployment architecture.

## *Supporting a distributed architecture*

If your organization is large and/or geographically dispersed, you may require a distributed SIA architecture. In this case, you can use a central SIA console to manage multiple SIA appliances (plus virtual SIA appliances and/or SIA servers) to aggregate views and reports across many SIA systems.

Figure 3-5 depicts a distributed SIA environment in which multiple SIA systems are managed by a centralized console.



**Figure 3-5:** Distributed SIA architecture.

## *Removing blind spots through SSL decryption*



Most network security devices are blind to Secure Sockets Layer (SSL)-encrypted traffic. SIA solutions are no different. The rising number of cyberattacks tunneled through SSL means that no organization can afford to stay blind.

Top-tier SIA vendors typically provide a solution for decrypting SSL traffic prior to collection and indexing by the appliance. The solution is often a stand-alone SSL decryption appliance that is equipped with the organization's SSL keys and certificates so encrypted traffic can be decrypted prior to being stored and indexed by the SIA appliance.

## A security analyst's worst nightmare

Imagine that you're director of network security at a Fortune 500 financial-services company. Your home phone rings at 3 a.m., rousing you from the calm of a deep, warm sleep. As you fumble for the phone, you look at caller ID and realize that your boss — the company's chief information security officer (CISO) — is calling. Your heart begins to race.

It turns out that your boss just received a call from the company's vice president of public relations. Overnight, a prominent hacking group infiltrated your company's network and exfiltrated data pertaining to more than 100,000 credit-card accounts. The group alerted the *Washington Post* and *The Wall Street Journal*, which could make this incident tomorrow's front-page news.

Your CISO gives you until 7 a.m. to answer the following questions:

- ✓ Did a network breach truly occur?
- ✓ If so, how did the hackers get in?
- ✓ What systems were affected?

✓ What, if any, data was stolen?

Oh, yes — he also wants to know how you're going to make sure that this never happens again.

You hang up the phone; throw on some clothes; grab your car keys and your computer; and head off to work, leaving tire tracks in your driveway. When you get to the office, however, your security products show no signs of intrusion. How can you be 100 percent sure that your network wasn't compromised?

This scenario may seem extreme, but situations like it have taken place countless times in recent years. Just ask the folks at Citigroup or Heartland Payment Systems — or, for that matter, the CIA!

Security intelligence and analytics is the ultimate Big Data Security solution because it's always on, it's always watching, it knows everything — and it can save your hide in a scenario just like this.

## Chapter 4

# Use Cases for Big Data Security

### *In This Chapter*

- ▶ Identifying the most common use cases for Big Data Security
- ▶ Increasing your network security posture
- ▶ Seeing how real-world organizations benefit

**I**f you've read the first three chapters of this book, you probably have a better sense of why traditional security defenses fail — or at least aren't foolproof — and of how Big Data Security can help. Now I peel back the layers of the onion a little further to show you the various use cases for Big Data Security.

Enterprises and government agencies rely on Big Data Security for six primary tasks:

- ✓ Situational awareness
- ✓ Cyberthreat detection
- ✓ Security incident response
- ✓ Data loss monitoring and analysis
- ✓ Organizational policy compliance verification
- ✓ Security assurance

In this chapter, I explore each task in its own section.

## Situational Awareness

To describe what I mean by *situational awareness*, I'll borrow an analogy from Martin Roesch, a true cybersecurity visionary who created Snort (a popular open-source intrusion detection technology) and founded Sourcefire. I first heard Marty tell this (paraphrased) story nearly a decade ago, and it still rings true today:

Imagine that you're a bank teller sitting behind the teller window on a slow Tuesday morning. The lobby is empty, and your manager — the only other person working in the bank — just stepped away to use the ladies' room. The front entrance to the bank is just around the corner, not visible from where you're sitting. But you happen to have a black-and-white television monitor positioned nearby, displaying a live feed from the front-door security camera.

You hear the front door swing open, so you look down at that television screen and see a man walking in dressed in a ski mask and thick black hoodie carrying something in his right hand — but you can't quite make out what it is. A panic button just below the counter where you're sitting can signal the police. Do you press it?

If you were to ask Marty whether he'd press the panic button, he'd say, "It depends." If you're a bank teller in Fargo, North Dakota, for example, and it's a cold, rainy day in February, half your customers may be dressed like this, carrying what you now realize is an umbrella. Pressing the panic button unnecessarily in this situation might cost you a customer — and possibly your job. On the other hand, if you're working in Miami, Florida, on a hot, sunny day in July, a man wearing a ski mask should make you more than a little nervous.

This story illustrates how situational awareness works in day-to-day life: It's about applying data from your environment to a situation at hand to help you understand its context. Or to put it in information security terms, situational awareness helps you determine which security events are "real" and which ones are inconsequential.

One of the most significant, yet overlooked, benefits of Big Data Security is that it provides powerful situational awareness that enables you to make well-informed decisions about potential cyberthreats. With Big Data Security, you know

exactly which applications, files, sessions, and users are traversing your network, exactly what data is coming and going, and to and from where. Big Data Security enables you to make better-informed decisions, so you know exactly when to press the panic button.

## Cyberthreat Detection

At the start of Chapter 2, I discuss the merits of a defense-in-depth strategy. Well, Big Data Security plays an important role in that strategy, especially when it comes to detecting advanced cyberthreats.

Zero-day malware, stealth botnets, and APTs have dominated the headlines in recent times. The trends are clear: Hackers are no longer motivated just by the prospect of fame or the thrill of vandalism, but now focus on economic benefit and even information warfare. Targeted attacks, engineered and carried out by sophisticated criminal hackers, jeopardize the economic welfare of virtually every organization. These types of attacks are designed to be difficult for traditional network security tools to detect.



APTs and ATAs take the slow-and-low tactic to evade traditional defenses.

So how can Big Data Security mitigate advanced cyberthreats? It can help at all three stages of the cyberthreat continuum: before, during, and after the attack.

### *Before attack*

Before an attack, Big Data Security can help you:

- ✓ Gain situational awareness (covered earlier in this chapter) by familiarizing you with the types of traffic on your network so that you can better recognize out-of-the-ordinary communications.
- ✓ Reduce your network's attack surface by identifying applications, communications, and operating systems that aren't approved for use in your organization (more on this in the "Organizational Policy Compliance Verification" section).

## *During attack*

During an attack, Big Data Security can help you:

- ✓ Detect the threat by identifying anomalous communications, such as an internal host connecting to an outside host for unusually long periods, an internal host transmitting an abnormally large amount of data, or an end-user host (desktop or laptop) communicating directly with other end-user hosts rather than servers.
- ✓ Identify *rogue hosts* (computers that are planted inside the organization for nefarious reasons) that are clearly outside the operating system and/or application parameters set by your IT department.
- ✓ Quarantine the threat by identifying other hosts that may have been compromised.

## *After attack*

After the attack, Big Data Security can help you:

- ✓ Verify whether a breach really occurred and whether any lingering threats need to be remediated.
- ✓ Identify who attacked you and what data may have been extracted.
- ✓ Determine the scope and extent of the breach.
- ✓ Determine ground zero for the attack.
- ✓ Understand exactly how the breach happened so you can ensure that it doesn't happen again.



As I discuss in Chapter 2, traditional security defenses simply aren't equipped to mitigate advanced cyberthreats. Big Data Security is your secret weapon for detecting these advanced threats, so use it — and use it often.

## *Security Incident Response*

Incident response is arguably the most popular (and perhaps most obvious) use case for Big Data Security. When your traditional security defenses trigger an alert, or you discover

a cyberthreat through your Big Data Security solution, you need to respond — *fast*. Every second that ticks away could mean more data loss and/or another host being compromised. Time is most certainly of the essence.

## Loading the ultimate incident-response weapon: CSIRT

Virtually every Global 2000 enterprise and large government agency has a team of cybersecurity professionals called CSIRT (pronounced “see-sirt”), which stands for Computer Security Incident Response Team. Although the complete list of CSIRT responsibilities varies by organization, one task common to all CSIRT teams is incident response.

When a member of the organization’s IT department (often, someone from the help desk) calls the CSIRT hotline, a CSIRT incident handler is dispatched within minutes to investigate the situation. When that person arrives and confirms the cyberthreat, he or she attempts to answer the five dreaded questions of incident response:

- ✓ Who did this to us?
- ✓ How did they do it?
- ✓ What systems and data were affected?
- ✓ Can we be sure that the incident is over?
- ✓ Can we be sure that it won’t happen again?

If you use traditional security tools, answering these questions is challenging, because you have access to only a subset of the information you

need. Sure, log data aggregated by a SIEM is helpful for piecing the puzzle together, but you can’t derive packet payload from log files.

Think of reviewing a SIEM report like reviewing your phone bill. You can certainly see when (long distance) calls were made, including the duration of each call and the number that was dialed. But there’s no way to confirm who was actually speaking and what the parties were discussing. Implementing Big Data Security is like (legally) tapping into the phone line to record everything that was said — 24 hours per day, 7 days per week.

A Big Data Security solution, built on a security intelligence and analytics (SIA; see Chapter 3), records every packet that traverses the network. CSIRT incident handlers have access to every packet captured, which can provide definitive evidence of an attack; this evidence allows members of the CSIRT team to make accurate conclusions rather than educated guesses.

Today, a CSIRT team without Big Data Security is like a team of air-traffic controllers equipped with just binoculars. If you’re not using Big Data Security to its fullest potential, you’re just asking for trouble.

When a security event is identified by an IPS, NGFW, or malware detection appliance, the event that triggered the alert usually doesn't represent your network's first contact with this threat. To ensure that the incident doesn't continue and can't happen again, it's important to determine the root cause of a security event (see the "Recovering from an incident" section).

## *Recovering from an incident*

The effects of an incident are often related to the scope of the systems and data that was accessed by an attack. Here are a few ways that Big Data Security can help you recover from these effects:



- ✓ **Tracking the source:** Most traditional network security tools can't determine the scope of an incident on your network; they only alert you to the moment in time when the threat was detected. By contrast, Big Data Security solutions allow you to track the pathway of a threat by fingerprinting files and searching all network data. You can see how target systems responded to threats and discover what was compromised. You can definitively determine the full picture of the incident on your network.

With better Big Data Security solutions, you can start with the event that triggered an incident and then travel backward in time — across machines, applications, and people — to get the full context of what happened. This capability makes it fast and easy to discover the initial vulnerability and source of the incident.

- ✓ **Monitoring all network connections:** After an incident has occurred on your network, how can you be sure that it's over? Many organizations that suffer high-profile breaches deal with the lingering effects of those breaches for months or even years.

With Big Data Security, however, IT security personnel have real-time situational awareness, so they know whether an incident is over, whether attackers are still present on the network, and whether any machines are still compromised. By monitoring the connections among machines and out to the Internet, Big Data Security makes it possible to prove that your network is secure.

## Big Data Security schools a university on incident response

A major private university recently evaluated several leading Big Data Security solutions to help secure its network, which serves 35,000 students and 10,000 faculty and staff members. As you can imagine, a large organization with a complex campus-wide network creates many possibilities for hackers to cause havoc. One incident in particular sparked the university's search for a Big Data Security solution.

Two years ago, a student installed keylogger software on multiple systems in the computer lab in an attempt to steal student credentials and forward the data off-campus through encrypted channels. Fortunately, a bright young computer science major discovered the keylogger software, leading to multiple university security analysts having to scour the drives to find out who was logged in when the software was installed. After analyzing every computer hard drive in the lab, the analysts finally identified the offending student, who was expelled soon

after — but not before expending several weeks of effort.

Knowing that there had to be a better way, the university's chief information security officer (CISO) turned to Big Data Security for the answer. He invited multiple leading vendors to make presentations, narrowed his short list to two, and then conducted on-site evaluations. In the end, he acquired DeepSee Appliances from Solera Networks ([www.soleranetworks.com](http://www.soleranetworks.com)). He chose Solera Networks because it was the only Big Data Security vendor that offered full-packet capture with real-time file reconstruction and the ability to uncover the root cause of cyberthreats quickly.

Several months later, another industrious computer science student installed yet another keylogger application in another computer lab. But the university had its new Big Data Security system in place, so finding the culprit this time took hours instead of weeks.

- **Collecting forensic evidence:** Finally, if you're lucky enough to identify the culprit (or culprits) by name, your Big Data Security solution can assist law-enforcement computer forensics officers by collecting digital evidence that may be used to prosecute the alleged perpetrator(s) in a court of law.

## Data Loss Monitoring and Analysis

Data breaches are occurring at an alarming rate — more than one per day in 2011. The average cost of a data breach for U.S.-based organizations was \$5.5 million in 2011, according to Ponemon Institute's 2012 report on the topic (see Chapter 1). Cleaning up after a breach is often the most expensive part of the process, and it's directly related to the type of information and number of records exposed.

If you don't know what was exposed, however, the cleanup job gets even more difficult — and even more expensive. Unfortunately, most organizations find themselves in this predicament. Log tools alone can't explain exactly what happened. In cases related to credit-card theft, organizations may be financially exposed to the maximum possible scope of a breach. Knowing exactly what happened can save an organization millions of dollars.

With Big Data Security at your disposal, you can protect yourself in the following ways:

- ✓ Monitor and record all files leaving your organization through web, e-mail, and instant messaging
- ✓ Monitor queries to your SQL database and relate them to the source endpoint
- ✓ Construct policies that extract, recreate, and take policy-based action on potentially sensitive files leaving the organization (for example, forwarding such files to a DLP system for analysis)
- ✓ Record all traffic to and from your critical data stores (like a video camera in a bank) and replay events if a breach occurs

## Organizational Policy Compliance Verification

To IT security professionals, the word *compliance* is most often associated with some industry or government regulation,

such as PCI or FISMA. But the term also relates to organizations' internal policies that regulate the use of computers and the Internet. These policies are commonly referred to as *acceptable-use policies (AUPs)*.

Some organizations combine multiple specific policies into one large AUP; others maintain a separate AUP for each topic. Regardless, AUPs often cover the following topics related to computers and the use of computer networks:

- ✓ Internet and social media
- ✓ E-mail and instant messaging
- ✓ Operating systems and applications
- ✓ Laptops and mobile devices
- ✓ Personal computing devices

Employees may wonder why their organization's IT department is so restrictive when it comes to computer use and Internet access. Is the chief information officer simply mean? Does the chief executive officer want to make sure her employees are working every moment? Although I can't personally comment on the personalities or work habits of your organization's senior executives, I *can* tell you that the primary reason for implementing AUPs is to achieve one objective: reducing the network's surface area of attack.

Every computing device is a target for hackers. Most operating systems and applications have at least one exploitable vulnerability. Vendors know about many of these vulnerabilities and may already have patched them, but some of them may be zero-day vulnerabilities, which I discuss in Chapter 1. The more freedom organizations grant their employees to select and customize operating systems, applications, and computing devices and to use the Internet, the less secure the organization's IT infrastructure will be.

How can Big Data Security help you monitor and enforce AUPs? Easy. Your Big Data Security solution sees everything, so you have a real-time catalog of communications and applications in use on the network. You also can see when users violate AUPs by visiting unauthorized websites, downloading unauthorized content (such as pirated music and movies), and posting inappropriate content on social media websites during work hours from work computers.

## Big Data Security for your health

A privately held U.S. consumer-products company recently sought a Big Data Security solution that would not only aid in security investigations, but also help maintain compliance with internal AUPs.

This company employs approximately 1,500 employees worldwide and has annual sales of approximately \$500 million. It offers consumer products that promote healthy living, including personal-care products, vitamins and supplements, and energy drinks.

Recognizing the importance of enforcing AUPs to reduce the company's surface area of attack, the longtime CIO evaluated multiple Big Data Security solutions but ultimately selected DeepSee Appliances from Solera Networks ([www.soleranetworks.com](http://www.soleranetworks.com)). He chose the Solera solution because it could identify dozens

of popular applications and also reconstruct content (web pages, chat messages, Microsoft Office files, PDF documents, images, and so on) in real time.

The company selected an appliance model capable of storing and indexing data up to 10Gbps with enough capacity to store three weeks' worth of traffic. The appliance was connected to the network through a TAP device that enabled the company to capture all traffic flowing into and out of the network.

Now, rather than just hoping that users are complying with company AUPs, the CIO can monitor the network for compliance and respond to policy violations, only occasionally engaging with the company's human resources department about violations that, frankly, aren't appropriate to discuss in this book.

Sure, you can realize definite employee productivity gains by enforcing AUPs. At the end of the day, however, by monitoring and enforcing your AUPs, you're reducing your network's attack surface, which results in reduced risk and fewer successful cyberattacks.

## Security Assurance

In the days following a zero-day attack, vendors of signature-based network security tools (see Chapter 2) scramble to publish new signatures that defend against the new threat. But a dangerous window of vulnerability is wide open between the time when the zero-day attack is discovered in the wild

and the time when your network security devices are updated with the new signatures. How can you be sure that your organization wasn't victimized by this zero-day attack?

Big Data Security gives you *security assurance* by enabling you to replay captured traffic during this dangerous window through security devices once they're equipped with updated signatures. This helps to determine whether your network was previously victimized by the zero-day attack. If a security event related to the new signature (or signatures) is triggered, you can leverage your Big Data Security solution to determine the scope of the attack and to help quarantine its effects.

When you're replaying captured traffic for inspection by your network security devices (IDS, IPS, or NGFW) after they're updated with new signatures, you'll want to work in a nonproduction environment, such as a lab, to ensure there is no contamination. Also, purchasing redundant network security devices for the purpose of security assurance is highly recommended. But if budget is a concern, using freely available open-source solutions may be a viable alternative.



## Chapter 5

# Integrating Big Data Security

### *In This Chapter*

- ▶ Integrating Big Data Security into your existing security infrastructure
- ▶ Seeing real-world examples of integration
- ▶ Exploring universal connections

No IT security solution should ever operate in a vacuum. This rule applies to every endpoint and network security product on the market — and especially to Big Data Security. The era of static security products is over! Long live context-aware security and situational awareness!

You have many reasons to integrate Big Data Security into your existing security fabric. Here are a few of the most important things Big Data Security does:

- ✓ Delivers *context-aware* security — providing deeper identity, application, content, reputation, vulnerability, and threat context to IT security teams at the point when a security and/or policy enforcement decision is made
- ✓ Provides situational awareness to help you gain unprecedented visibility into files, applications, and flows
- ✓ Accelerates incident response when time is of the essence
- ✓ Reduces risk by detecting APTs and ATAs that might otherwise go unnoticed
- ✓ Mitigates security threats by verifying that a cyberattack is truly over

In this chapter, I provide some real-world examples to get you started.

## ***SIEM Integration***

In Chapter 2, I briefly discuss the roles that a SIEM plays in larger companies and government agencies. Specifically, I'm referring to two roles:

- ✓ **Log aggregation**, which allows you to query all your log data in one place
- ✓ **Correlation**, which allows you to leverage prebuilt and custom rules to correlate security events

SIEMs are particularly useful for gluing together disparate pieces of information to understand the meaning behind security related events. SIEMs work kind of like the jury in a criminal trial:

1. The jury listens to testimony by individual witnesses.

*In other words, the SIEM collects data from various security systems, such as IPS, antivirus, and data loss prevention (DLP) systems.*

2. The jury weighs all the evidence.

*That is, the SIEM correlates the collected data.*

3. The jury determines innocence or guilt.

*Meaning the SIEM reports on the presence or absence of a cyberattack.*

For a SIEM user to weigh all the evidence about a potential cyberthreat, she must consult her Big Data Security solution by investigating host traffic specifically related to the suspected threat (source and destination IP addresses) at the exact moment of the alleged attack (date and time). She could certainly log into her Big Data Security console and construct a new query, but by leveraging the integration between the Big Data Security solution and her SIEM, she can initiate that query directly from the SIEM console (see Figure 5-1), thereby saving valuable time and effort.

## Gartner weighs in on Big Data and context-aware security

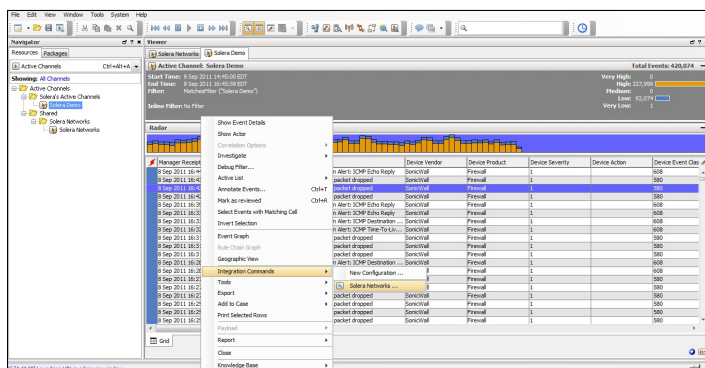
Gartner ([www.gartner.com](http://www.gartner.com)) is a leading IT research and advisory firm serving IT leaders in more than 12,000 organizations. In a recent research brief titled “Information Security Is Becoming a Big Data Analytics Problem,” Gartner Vice President and Distinguished Analyst, Neil MacDonald, discusses the role of big data analytics within information security.

“Information security is becoming a big data analytics problem,” says Mr. MacDonald, “where massive amounts of data will be correlated, analyzed, and mined for meaningful patterns. Investments in additional tools, processes, and skills will be required.”

Mr. MacDonald continues by discussing how context-aware security can improve an organization's security monitoring. “The results of

security analytics can be improved by linking additional context information. There are many sources of context-aware security-related information that can be gathered to supplement monitoring data and improve models. This context includes not only environmental context, such as time of day and location, but also application, identity, and content awareness. Increased context from all layers will increase the amount of richness of the data and improve the ability of big data analytics to discern meaningful patterns.”

Mr. MacDonald predicts, “The amount of data analyzed by enterprise information security organizations will double every year through 2016. By 2016, 40% of enterprises will actively analyze at least 10 terabytes of data for information security intelligence, up from less than 3% in 2011.”



**Figure 5-1: Big Data Security integration with ArcSight ESM.**



Integrating your Big Data Security solution into the consoles of your mission-critical security products isn't just a matter of convenience. When investigating a critical, high-impact security event, every ticking second matters.

## IPS Integration

As you may recall from Chapter 2, an IPS can be placed in inline IPS mode, meaning that it can actually block cyberthreats, and/or it can be placed in passive IDS mode, meaning that it can only provide alerts about detected threats. Over the past decade, more organizations have become comfortable placing IPS appliances inline because advancements in performance and detection technology have yielded far fewer *false positives* (misclassifying good traffic as bad).

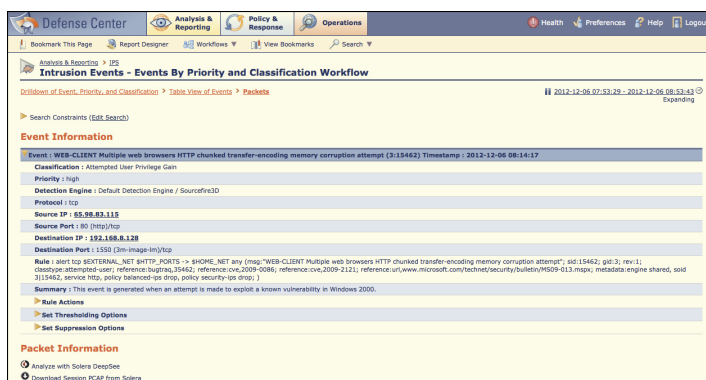


Even when an IPS is placed inline, however, not all signatures in a typical IPS detection policy are configured to block. Certain vulnerability-based signatures may be more prone to misfires (yielding false positives) than more clear-cut exploit-based signatures, which trigger only on detection of known exploits.

When an IPS signature fires, triggering an intrusion event, security analysts must investigate whether the cyberthreat is relevant to the operating system or application that it's intended to attack. (Next-generation IPS solutions automate this task.) Then, if the threat is relevant, analysts must determine whether the attack was successful.

To accomplish these objectives, a security analyst, upon reviewing an intrusion event, must initiate a query into the Big Data Security solution so that he can analyze the traffic that corresponds to the event — before, during, and after the suspected attack was detected. Specifically, he needs to analyze packet captures (PCAPs) associated with the source and destination IP addresses involved in the intrusion event at the precise date and time the event was registered.

This user could connect to the Big Data Security solution's console manually and then configure a new query, but he'd be losing precious time when every second counts. A better approach would be to connect to the Big Data Security solution directly from the IPS's management console and review the details of the intrusion event (see Figure 5-2).



**Figure 5-2:** Big Data Security integration with Sourcefire IPS.

By leveraging APIs from both your Big Data Security provider and your IPS vendor, you can streamline security analysts' workflow, which saves time, saves effort, and reduces the impact of successful cyberattacks.

## NGFW Integration

As I discuss in Chapter 2, an NGFW is a multifunction security appliance equipped with firewall, IPS, and application control technologies. When it comes to detecting cyberthreats, the IPS component of an NGFW operates quite similarly to a stand-alone IPS. Users can launch preconfigured queries directly from the NGFW console through corresponding NGFW and Big Data Security integrations.

## Advanced Malware Protection Integration

Along with Big Data Security, advanced malware protection technology is among the top new innovations in cybersecurity technology over the past decade. Its capability to analyze files in a virtual sandbox environment makes this technology particularly useful for detecting unknown and zero-day threats.

## Big Data Security plays major defense

A major U.S. government defense contractor was recently victimized by a series of zero-day attacks. Although the company had already deployed a series of best-of-breed network and endpoint security solutions, its security analysts found it extremely difficult to determine the sources and effects of these attacks, which cost more than \$5,000 per investigated incident in labor costs alone!

One of the network security administrators had read an article about Big Data Security and wondered whether such a solution might help his organization. He persuaded his boss to invite a few vendors on site.

After learning about the power of Big Data Security, the company conducted on-site evaluations with two competing vendors. It ultimately selected DeepSee Appliances from Solera Networks ([www.solera-networks.com](http://www.solera-networks.com)). The Solera DeepSee solution offered full-packet capture, was able to perform at wire speed, provided a powerful reporting engine, and was compatible with the company's existing network IPS appliances from Sourcefire.

Soon after deploying the new DeepSee Appliances, the network security team realized that its network was being victimized by Google image-search poisoning. In this type of attack, criminals intentionally link keyword-rich image-search queries

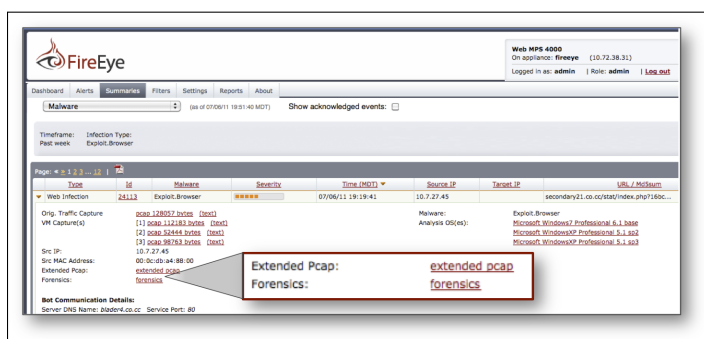
to malware or sites that host malicious code. When the unsuspecting user visits the site to download one of these images, the web server generates and delivers a unique malware executable for the victim computer, thereby creating a zero-day threat that a typical network IPS is unlikely to detect. When the threat was discovered, the team quarantined infected hosts and created custom IPS rules that equipped its Sourcefire IPS to detect repeated infections.

The company realized a side benefit from deploying a Sourcefire-compatible Big Data Security solution: Whenever a high-impact intrusion event registered on the Sourcefire management console, analysts could click a link inside the Sourcefire interface to launch an immediate forensics query on its Solera DeepSee appliances. This feature dramatically accelerated the company's incident-response process, saving valuable time and money.

After several months of running the Big Data Security solution in production, the chief information security officer (CISO) estimated that the Solera Networks investment saved the company approximately \$4,000 per investigated incident, reducing labor costs by 80 percent and saving hundreds of thousands of dollars each year. In addition, the network is far more secure now — a benefit that's priceless to any CISO.

Although advanced malware protection devices are configured for inline deployment, they can block known threats only in near real time, just like an IPS or NGFW appliance. If an advanced malware protection device confirms a previously unknown threat, it automatically creates a new signature for future threat prevention — but at that point, it's too late to prevent that first attack. That's where Big Data Security comes in.

By leveraging your Big Data Security solution's integration with your advanced malware protection device, you can launch preconfigured network forensics queries straight from the malware protection system's console (see Figure 5-3).



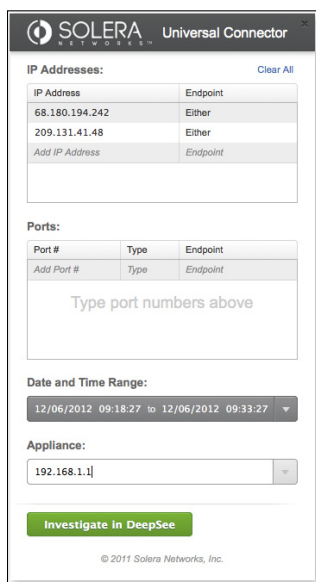
**Figure 5-3:** Big Data Security integration with FireEye.

Just like IPS and NGFW console integration, advanced malware protection console integration saves security analysts considerable time and effort.

## Universal Connectors

Each of the Big Data Security integration examples mentioned in the preceding sections involves the use of APIs from your Big Data Security provider and network security vendors. But what if your network security vendor doesn't offer such an API? Are you simply out of luck?

Not necessarily. Some leading Big Data Security vendors provide what's commonly called a *universal connector*, a small piece of software that installs directly in your web browser (such as Mozilla Firefox) as a plug-in. When you install this plug-in and launch the browser, a small sidebar like the one shown in Figure 5-4 appears. While you stay on the web page that's hosting your network security device's console, you can send queries about any event to your Big Data Security solution for further analysis.



The screenshot shows a sidebar titled "SOLERA Universal Connector". It contains several sections for configuring search queries:

- IP Addresses:** A table with columns "IP Address" and "Endpoint". It lists two IP addresses: 68.180.194.242 and 209.131.41.48, both with "Either" as the endpoint. Below the table are links to "Add IP Address" and "Endpoint", and a "Clear All" link.
- Ports:** A table with columns "Port #", "Type", and "Endpoint". It has a link to "Add Port #". Below the table is a text input field with the placeholder "Type port numbers above".
- Date and Time Range:** A dropdown menu showing the range "12/06/2012 09:18:27 to 12/06/2012 09:33:27".
- Appliance:** A dropdown menu showing the value "192.168.1.1".
- Investigate in DeepSee:** A green button.
- Footer:** Copyright notice "© 2011 Solera Networks, Inc."

**Figure 5-4:** Big Data Security universal connector from Solera Networks.



Although this technique is by no means as seamless as integrating access to your Big Data Security solution into the console of your network security solution (by using corresponding APIs), a universal connector is the next-best thing, because it simplifies the process of configuring search queries in your Big Data Security solution.

## Chapter 6

# Ten Buying Criteria for Big Data Security

### *In This Chapter*

- ▶ Avoiding the pitfalls of inferior Big Data Security offerings
- ▶ Creating a checklist of important buying criteria
- ▶ Understanding what to look for when evaluating solutions

**N**ot all Big Data Security solutions are created equal. The performance, scalability, and capabilities of solutions vary considerably, as do the quality of the vendors.

In this chapter, I provide ten buying criteria to consider when evaluating Big Data Security solutions. First, though, here are a few pitfalls to avoid:

- ✓ Solutions that sample packets rather than capture every packet that traverses your network
- ✓ Solutions that can't keep up with the speed of your network
- ✓ Solutions that don't provide the flexibility of software-based and virtual appliance options (thereby locking you in to a single vendor for hardware, software, and storage)
- ✓ Solutions that don't integrate with your existing security infrastructure
- ✓ Solutions that take a rocket scientist to learn how to configure and a packet scientist to use
- ✓ Vendors without a track record of success across enterprise, defense and agency, and public-sector market segments

Now that you know what *not* to look for, you're ready to review the attributes of a Big Data Security solution that you *should* look for.

## Hardware Flexibility

Although most organizations prefer to acquire Big Data Security solutions that incorporate purpose-built hardware appliances, some prefer the flexibility of selecting their own hardware platform, perhaps to accommodate larger storage arrays than what the vendor can provide or to simply use existing hardware to reduce costs. Also, some government agencies (especially those in the military) require all software to be installed on hardware from an approved-products list.

By selecting a Big Data Security vendor that offers both hardware and software versions, you make sure that you aren't tied down to proprietary hardware and maintain deployment flexibility based on your business requirements.

## Ease of Use

No matter how powerful, scalable, and feature-rich a Big Data Security solution is, it's practically worthless if users can't figure out how to use it. Be sure to consider only Big Data Security solutions that feature easy-to-use dashboards, reports (see Figure 6-1), and alerts, and those that make it simple to locate and analyze traffic flows of interest.



**Figure 6-1:** Sample Big Data Security report charts.

## Full-Packet Capture

Some aspiring Big Data Security solutions capture only sample traffic, because the hardware doesn't have the horsepower to perform full-packet capture at speeds up to 10Gbps. Even some that do have enough power still perform only statistical sampling of data to speed report generation.



Although packet sampling can improve performance, the risk of missing important packets for accurate and reliable threat detection is too substantial. It's like buying a video surveillance system that only takes still photos every ten seconds! Thus, you should only consider solutions that capture and analyze packets in their entirety.

## Deep Packet Inspection



A high-quality Big Data Security solution should be capable of performing deep packet inspection at all seven layers of the Open Systems Interconnection (OSI) model. Most important, users should be able to view and analyze their data with the full context afforded by such attributes as username and application name.

Accuracy of data categorization is of utmost importance so that users can quickly find the data they're searching for when time is critical instead of doing packet analysis to identify users and applications. Many organizations have policies that allow browsing Facebook from the office but not posting messages or sending e-mail within that application. Only the best Big Data Security solutions can differentiate these activities within a single application like Facebook.

## Enterprise Performance and Scalability

A Big Data Security solution could have every possible feature under the sun, but if it can't collect and analyze data at the speed of your network, you're out of luck.

In addition, your solution must scale with your organization as it grows (or decides to monitor more network segments with its Big Data Security solution). To achieve the enterprise-class scalability that you require, you may need to acquire multiple Big Data Security appliances. If you do so, be sure to select a solution with a central management console that can aggregate data from your underlying Big Data Security appliances, providing aggregated views to make it easy to construct dashboards, reports, and alerts.

## *Virtual Platform Visibility*

Physical Big Data Security appliances can't capture and analyze VM-to-VM (virtual machine to virtual machine) traffic, so you want to select a Big Data Security vendor platform with virtual appliance software that integrate directly into your virtualization platform's virtual switch. Data captured by the Big Data Security virtual appliance should be accessible from the central management console for centralized aggregation, monitoring, reporting, alerting, and analysis by existing security tools in your physical network infrastructure.

## *Content Reconstruction*

In Chapter 3, I describe how Big Data Security solutions reconstruct content so that analysts can view that content in its original form — anything from documents and images to chat messages and e-mails. Select a Big Data Security solution that offers full-featured content reconstruction to maximize the effectiveness of your staff's investigation and visibility efforts.



Once content has been reconstructed, better Big Data Security solutions provide the ability to configure policies to automatically redirect that content to a third-party network security device, such as a next-generation IPS appliance, DLP appliance, or an advanced malware protection appliance for further analysis.

## SSL Decryption

Every Big Data Security solution — heck, any network security solution — is blind when it comes to inspecting traffic encrypted with Secure Sockets Layer (SSL) technology (see Chapter 3 for more on the topic).



Do yourself a favor: If you don't already have a stand-alone SSL decryption solution (see Figure 6-2) for decrypting SSL traffic before storage and analysis, make sure that your Big Data Security vendor has one available. Otherwise, you may miss advanced cyberthreats embedded within SSL-protected communications.



**Figure 6-2:** Netronome SSL Inspector.

## Extensive Third-Party Integration

As I discuss in the last chapter, you should integrate your Big Data Security solution into your existing security infrastructure whenever possible to accelerate investigations and detect cyberthreats. Common Big Data Security integrations include:

- ✓ Security information and event management (SIEM)
- ✓ Intrusion prevention system (IPS)
- ✓ Next-generation firewall (NGFW)
- ✓ Advanced malware protection

For more on this, see Chapter 5.



You want to select a Big Data Security provider that offers an integration of its solution into popular network security platforms. At the very least, select a vendor that offers a universal connector, which makes it easy to submit queries to your Big Data Security solution through your network security product's web-based console.

## *Responsive Customer Support*

Selecting a Big Data Security vendor is just as important as selecting a Big Data Security product, if not more so. Be sure to select a vendor that is 100 percent focused on security instead of offering Big Data Security as a byproduct of network forensics or network performance monitoring.



Also, find an excuse to contact the customer-support department on multiple occasions during the evaluation phase. Consider how quickly the vendor responds to each phone call and/or e-mail inquiry and whether the issue was resolved to your satisfaction.

# Glossary



**acceptable-use policy (AUP):** A set of internal rules that restricts the way in which end users may use company-owned computer, network, and Internet resources.

**advanced persistent threat (APT):** A sophisticated cyberattack that employs advanced stealth techniques to remain undetected for extended periods of time, usually targeting a government or commercially owned entity for the purposes of espionage or long-term reconnaissance.

**advanced targeted attack (ATA):** A sophisticated cyberattack against an entity. This type of attack uses knowledge about the victim organization to target the individuals or devices within the target network that could yield the most fruitful results for the criminals.

**baiting:** A social-engineering attack in which physical media (such as CD-ROMs or USB flash drives) containing malware are deliberately left in proximity to a targeted organization's facilities, where they may be found and later installed by curious victims.

**Big Data:** A collection of data sets so large and complex that it becomes awkward to work with in traditional database management and analysis tools.

**Big Data Security:** A computer-based solution that captures and stores some or all of an organization's Big Data sources that are relevant to information security for the purposes of uncovering and mitigating cyberthreats.

**bot:** An infected computer (or endpoint) centrally controlled by another computer. See also *botnet*.

**botnet:** A network of Internet-connected computers with breached security defenses that a malicious third party may partly or completely control.

**buffer overflow:** A cyberthreat that exploits a vulnerability in an application in a specific way. The hacker intentionally overruns the buffer's boundary causing the application to pass undesirable commands directly to the operating system.

**cyberwarfare:** Politically motivated hacking to conduct sabotage and/or espionage against a nation state.

**data loss prevention (DLP):** A solution designed to detect, and in some cases prevent, potential data breaches by monitoring data in use, in transit, and at rest.

**defense in depth:** A strategy of installing a series of cybersecurity defenses so that a threat missed by one layer of security may be caught by another.

**denial of service (DoS):** A cyberthreat intended to disrupt or disable a targeted host by flooding it with benign communication requests from a single host.

**distributed denial of service (DDoS):** A DoS attack originating from numerous hosts, typically (though not always) associated with a botnet.

**hacktivism:** The use of computers and computer networks as a means to protest and/or promote political ends.

**intrusion detection system (IDS):** A passive device or software application that monitors network traffic and provides alerts about detected cyberthreats.

**intrusion protection system (IPS):** An active (inline) device or software application that monitors network traffic and blocks cyberthreats upon detection.

**malware:** Malicious software (such as a computer virus, worm, or Trojan) created to disrupt computer operation, gather sensitive information, or gain access to private computer systems. See also *spyware*, *Trojan*, and *worm*.

**network behavior analysis (NBA):** A cybersecurity solution that continuously monitors flow data from routers and switches, such as NetFlow and other flow standards, to detect anomalous network behavior.

**next-generation firewall (NGFW):** A multifunction security device, typically marketed to medium-to-large enterprises, that bundles firewall, IPS, application control, and (optionally) URL filtering technologies into one platform.

**phishing:** An attempt to acquire personal information (such as usernames, passwords, and credit-card details) by masquerading as a trustworthy entity.

**polymorphic malware:** Malware that modifies its own code, rendering it more difficult for some signature-based antimalware programs to detect.

**security information and event management (SIEM):** A solution that aggregates and correlates log data from network security and network infrastructure devices to provide analysis of security events.

**spear phishing:** A phishing attempt directed toward a specific organization or person(s) within that organization.

**spyware:** A type of malware that collects information about users, with or without their knowledge.

**SQL injection:** A technique used to attack databases through a website or web-based application. Portions of SQL statements are included in a web-form entry field in an attempt to get the website (or web application) to pass a newly formed rogue SQL command to the database.

**Trojan:** A type of malware that masquerades as a legitimate file or helpful application with the ultimate purpose of granting a hacker unauthorized access to a computer.

**worm:** A form of malware that exploits vulnerabilities in operating system or network protocols to propagate copies of itself to other computers connected to the same network or to USB mass-storage devices connected to the infected PC.

**zero-day attack:** A cyberattack on an unknown operating system or application vulnerability. The attack occurs on *day zero* of awareness of the vulnerability, when neither a patch nor a threat-detection signature exists.



# Security That **Matters**

Our advanced threat detection and security intelligence solutions help protect enterprises and government agencies from those who want to do them harm. And, by doing that, we are helping to protect our way of life...

[www.soleranetworks.com](http://www.soleranetworks.com)



These materials are the copyright of John Wiley & Sons, Inc.  
and any dissemination, distribution, or unauthorized use is strictly prohibited.

**Solera Networks would like to  
thank its sponsors**



# Leverage Big Data Security to uncover advanced cyberthreats and streamline incident response

If you're charged with securing your organization's network or responding to security incidents, this book is for you. Security-conscious organizations are turning to Big Data Security as the newest weapon to fight cybercrime, collect digital evidence, and uncover advanced cyberthreats that traditional security defenses miss.

- **Starting with the basics** — take a look at recent statistics on enterprise network breaches
- **Understanding traditional security** — review traditional security defenses and discover why they sometimes fail
- **Defining Big Data Security** — explore Big Data Security form factors, features, and deployment strategies
- **Reviewing use cases** — review common use cases for improving network visibility and strengthening your security posture
- **Integrating Big Data Security** — understand how to integrate Big Data Security into your existing security fabric
- **Establishing buying criteria** — know what to look for (and what to avoid) when evaluating Big Data Security solutions

[www.soleranetworks.com](http://www.soleranetworks.com)

**Steve Piper** is a cybersecurity veteran with over 20 years of IT experience. A freelance writer and consultant, Steve is the author of *Intrusion Prevention Systems For Dummies* and *Network Packet Brokers For Dummies*. Steve has achieved a CISSP security certification from ISC<sup>2</sup> and BS and MBA degrees from George Mason University. Learn more at [www.stevepiper.com](http://www.stevepiper.com).



**Open the book and find:**

- Lists of common internal and external Big Data sources
- Network diagrams depicting typical deployment strategies
- Strategies for detecting advanced threats and targeted attacks
- Real-world examples of Big Data Security implementations
- Ten buying criteria for evaluating Big Data Security solutions

**Go to Dummies.com®**  
for videos, step-by-step examples,  
how-to articles, or to shop!

For Dummies®  
A Branded Imprint of



ISBN: 978-1-118-51727-7  
Not for resale