Definitive Guide™
to
Next-Generation Threat Protection

Winning the War Against the New Breed of Cyber Attacks

Steve Piper, CISSP

FOREWORD BY:
David DeWalt

Compliments of:

FireEye™

About FireEye

FireEye is the leader in stopping today's new breed of cyber attacks, such as zero-day and APT attacks, that bypass traditional defenses and compromise over 95% of networks. The FireEye platform supplements signature-based firewalls, IPS, anti-virus, and gateways, and provides the world's only cross-enterprise, signature-less protection against Web and email threat vectors as well as malware resident on file shares. It is the industry's only integrated platform that stops attacks across every stage of an attack life cycle, from exploit to exfiltration. Using its patented Virtual Execution technology engine across its platform, FireEye is uniquely able to protect against today's new breed of cyber attacks. FireEye solutions are deployed in over 40 countries and more than 25% of the Fortune 100.

- Ranked #4 on the Deloitte 2012 Technology Fast 500™ North America
- Awarded the Wall Street Journal 2012 Technology Innovation Award
- Inducted into JPMorgan Chase Hall of Innovation

Definitive Guide™
to
Next-Generation Threat Protection

Winning the War Against the New Breed of Cyber Attacks

Steve Piper, CISSP

Foreword by David DeWalt

CYBEREDGE
PRESS

Definitive Guide™ to Next-Generation Threat Protection

Published by:
CyberEdge Group, LLC
1997 Annapolis Exchange Parkway
Suite 300
Annapolis, MD 21401
(800) 327-8711
www.cyber-edge.com

For general information on CyberEdge Group research and marketing consulting services, or to create a custom *Definitive Guide* book for your organization, contact our sales department at 800-327-8711 or info@cyber-edge.com.

ISBN: 978-0-9888233-0-3 (paperback); ISBN: 978-0-9888233-1-0 (eBook)
Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

Publisher's Acknowledgements

CyberEdge Group thanks the following individuals for their respective contributions:

Editor: Susan Shuttleworth
Graphic Design: Debbi Stocco, Christian Brennan
Production Coordinator: Valerie Lowery
Special Help from FireEye: Phil Lin, Lisa Matichak, Brent Remai, David DeWalt

Table of Contents

Foreword

As I've met with national leaders and customers around the world, I've found that there's a great divide between the level of security they need for their networks and the level of security available to them using traditional security tools. That's because the next generation of cyber attacks are already a part of their daily lives, but they're stuck working with traditional security tools that are based on decades-old technology models.

The entire security industry needs a shift in thinking because incremental improvements can't bridge the threat gap created by today's highly adept cybercriminals. I've said very publicly that the current cybersecurity model isn't extensible and requires a fundamentally new approach to security.

That's why I am extremely encouraged after reading this definitive guide. I'm more convinced than ever that we need to be educating each other and acting upon the reality of today's cyber attacks. We are in a cyber "arms race" run by criminal and nation-state organizations with interlocking profit motives and geopolitical agendas. It's getting ugly out there.

I am honored to serve as a member of President Barack Obama's National Security Telecommunications Advisory Committee as well as on the boards of Delta Airlines, Mandiant, and Polycom. From this perspective, we have a unique opportunity to unite the public and private sectors in a common cause. Together, I'm confident we will find innovative solutions to protect our shared critical infrastructure.

It is still staggering to me that cybercriminals and APT actors can break into virtually any network to steal data and disrupt businesses despite the over $20 billion invested in IT security technologies last year!

This guide is all about how to fill this gap in our network defenses to do battle against "today's new breed of cyber attacks," as Steve puts it in this book. The dramatic rise in global cyber incidents shows just how far the threats have escalated and how advanced and intricate these cyber attacks

have become. I urge you to read this guide and share what you learn with your colleagues and peer networking groups. We can't stop this next- generation of cyber attacks without more advanced technologies, better cooperation across industries, and stronger ties between the public and private sectors.

For my part, in joining FireEye, I consider this my renewed pledge to deliver leading-edge platforms to address today's toughest cybersecurity issues. I considered numerous CEO opportunities since McAfee and have watched FireEye on the sidelines for years. I am so excited to join a company with such game-changing technology. And, by sponsoring this guide, my hope is that it will give you a framework to deploy a next-generation threat protection platform that forms the basis of a more resilient, penetration-resistant network.

David DeWalt
CEO, FireEye

Introduction

In recent years, enterprises and government agencies have fallen prey to a myriad of successful cyber attacks of unprecedented sophistication and reach. Despite spending over $20 billion annually on traditional security defenses, organizations find themselves battling a new generation of cyber attacks, such as advanced malware and advanced persistent threats (APTs), that are dynamic and stealthy and extremely successful at compromising today's networks.

If there's any chance of preventing these motivated adversaries from attacking our systems, stealing our data, and harming our critical infrastructure, we've got to think differently. We must realize the limitations of traditional signature-based defenses and leverage new technology to uncover and stop today's new breed of cyber attacks.

Fortunately, there is a solution. Introducing next-generation threat protection (NGTP), an innovative new network security platform proven to help win the war against next-generation threats. If you're charged with securing your organization's network, this is one book you simply can't afford to miss.

Chapters at a Glance

Chapter 1, "Defining Next-Generation Threats," reviews staggering statistics on major data breaches, describes recent high-profile commercial and government cyber attacks, depicts typical costs associated with successful breaches, and contrasts traditional and next-generation cyber attacks.

Chapter 2, "Understanding the Enemy," categorizes three kinds of cyber enemies — cybercriminals, state-sponsored threat actors, and hacktivists — and describes why they are so successful in bypassing traditional security defenses.

Chapter 3, "Anatomy of Advanced Cyber Attacks," defines APTs and reviews high-profile examples that have recently made international headlines. The chapter then

details the potential "ripple effect" of a successful APT on critical infrastructure, explores each of the five stages of the APT life cycle, and provides telltale signs for detecting APTs in your organization.

Chapter 4, "Introducing Next-Generation Threat Protection," gets to the heart of the matter by defining NGTP, describing the characteristics of an ideal NGTP solution, and comparing NGTP to traditional signature-based defenses and sandbox technologies.

Chapter 5, "Next-Generation Threat Protection Explored," expands on Chapter 4 by explaining exactly how NGTP mitigates the new breed of cyber attacks in email messages, Web communications, and files at rest. It explores key features of leading NGTP solutions and describes common ways to integrate them into existing network infrastructure.

Chapter 6, "Selecting the Right NGTP Solution," describes exactly what to look for — and, more importantly, what to avoid — when shopping for an NGTP solution.

Glossary provides handy definitions to key terminology (appearing in *italics*) used throughout this book.

Helpful Icons

Tips provide practical advice that you can apply in your own organization.

When you see this icon, take note as the related content contains key information that you won't want to forget.

Proceed with caution because if you don't it may prove costly to you and your organization.

Content associated with this icon is more technical in nature and is intended for IT practitioners.

Want to learn more? Follow the corresponding URL to discover additional content available on the Web.

Chapter 1

Defining Next-Generation Threats

In this chapter

- Review recent statistics on data breaches
- Explore traditional cyber attacks
- Understanding advanced malware, zero-day and advanced persistent threats

Today's cyber attacks are more sophisticated than ever. In the past year, we've witnessed alarming data breaches of unprecedented complexity and scale, causing every CISO to re-examine the organization's network security posture.

At the same time, the cybercrime industry has completely transformed — from hacking for kicks to cyber attacks for profit, or in some instances, for political gain. Today's *cyber-criminals* are highly trained and incorporate sophisticated attack techniques that are no match for today's inadequate signature-based defenses.

Organizations now face a new breed of cyber attacks. These *multi-vectored* and *multi-staged* threats easily evade traditional security defenses, such as firewalls, intrusion prevention systems (IPS), secure Web and email gateways, and anti-virus platforms.

So, how bad has the problem become? Let's take a look at some recent statistics and references to some of the most heinous cyber attacks of our day.

Staggering Statistics

Several reputable cybersecurity research organizations monitor cyber attack trends against enterprises. Among these is the Verizon RISK (Response, Intelligence, Solutions, and Knowledge) Team, which publishes a widely regarded annual Data Breach Investigations Report.

In 2012, Verizon analyzed 855 data breach incidents that occurred in the prior year resulting in 174 million compromised records. Verizon's analysis yielded some staggering 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)

To download a free copy of the Verizon report, connect to www.verizonbusiness.com.

Also in 2012, FireEye, a leader in next-generation threat protection, published its Advanced Threat Report (1H 2012). According to the report, enterprises are experiencing an average of 643 Web-based malicious events each week effectively penetrating traditional security defenses, such as firewalls, intrusion prevention systems, and anti-virus software. Compared to the same period in 2011, the number of infections per company rose by 225 percent.

To download a free copy of the FireEye report, connect to www.fireeye.com/info-center/.

Despite global increases in information security spending (currently equating to over $20 billion per year for information security products and services), the percentage of data breaches stemming from external hacking is up, attacks incorporating malware are up, and it's still taking weeks to discover major data breaches!

Recent Victims

Unless enterprises and government agencies adopt a new, more sophisticated approach to mitigating next-generation threats, organizations will continue to make headlines in ways they never intended. The following is a sampling of recent high-profile commercial and government cyber attacks that incorporated advanced hacking techniques:

Commercial attacks

- ☑ **Global Payments** (March 2012): Attack dating back to January 2011 in which a hacker exfiltrated information for over 7 million credit cards, costing this credit card processor nearly $85 million and temporary delisting by Visa and MasterCard.
- ☑ **Citigroup** (June 2011): The company disclosed that a cyber attack resulted in the theft of more than 360,000 credit card numbers, of which 3,400 were used to steal more than $2.7 million.
- ☑ **RSA Security** (March 2011): Cyber attackers stole data related to SecurID tokens, rendering them insecure.

Jump to the "RSA Security steps forward to describe their APT attack" sidebar in Chapter 3 for details of this attack.

Government attacks

- ☑ **South Carolina Department of Revenue** (October 2012): A hacker exfiltrated approximately 3.6 million Social Security numbers and 387,000 credit and debit card numbers from an external cyber attack.
- ☑ **U.S. Environmental Protection Agency (EPA)** (August 2012): Social Security numbers, bank routing numbers, and home addresses of more than 5,000 EPA employees were exposed after an employee clicked on a malicious email attachment.
- ☑ **Iran** (May 2012): Flame malware allegedly developed by the United States and Israel was deployed to slow the Iranian nuclear program.

The Cost of Failure

To mitigate both traditional and next-generation threats, IT security organizations must implement a *defense-in-depth* strategy (layers of network and endpoint security defenses). Failing to do so can prove costly. In fact, it can bankrupt a company!

In 2012, Ponemon Institute (www.ponemon.org) published a third annual report titled "2012 Cost of Cyber Crime Study: United States." Upon analyzing the cost of data breaches for 56 U.S.-based enterprises, Ponemon found the average annualized cost of cybercrime for each organization to be $8.9 million, with a range of $1.4 million to $46 million. This is up from $8.4 million (6 percent increase) in 2011. Ponemon also calculated that each organization averages 102 successful cyber attacks per week, up from 72 per week in 2011 (42 percent increase).

To download a free copy of the Ponemon Institute report, connect to www.ponemon.org/library.

Companies victimized by large-scale data breaches face enormous costs, including:

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

When it comes to defending against cyber attacks, the old adage applies — an ounce of prevention is worth a pound of cure. Companies owe it to themselves, their customers, and their stockholders to incorporate next-generation threat protection into their defense-in-depth architecture to stay ahead of today's new breed of cyber attacks.

Today's Threat Landscape

There are dozens of cyber attacks facing today's enterprises and government agencies. I'm going to oversimplify the threat landscape by grouping cyber attacks into two broad categories — traditional threats and next-generation threats.

Traditional threats

The traditional cyber attacks described in this section are "oldies but goodies." But don't underestimate them. Although they can usually be detected by IPS devices, next-generation firewalls (NGFW), and anti-virus software, sometimes newer variants slip through the cracks.

Worms, Trojans, and viruses

A computer *worm* is a stand-alone malware program that replicates itself — typically through vulnerabilities in operating systems — over a network in order to propagate. Worms typically harm networks by consuming bandwidth, but also provide a "lateral" attack vector that may infect supposedly protected internal systems or exfiltrate data. Unlike a computer virus, a worm doesn't append itself to other programs or files.

A *Trojan* (or *Trojan horse*) typically masquerades as a helpful software application, with the ultimate purpose of tricking a user into granting access to a computer. Trojans may self-replicate within the infected system, but cannot propagate to other vulnerable computers on their own; they typically join networks of other infected computers (called botnets; see next section) where they wait to receive further instructions, and into which they submit stolen information. Trojans may be delivered by means of spam email or social media, or may be disguised as a pirated installer for a well-known game or application.

A computer *virus* is malicious code ranging in severity from mildly annoying to completely devastating. It attaches itself to a program or file enabling it to spread from one computer to another, leaving infections as it travels. However, unlike a worm, a virus can't travel without human action.

Spyware and botnets

Spyware is software that covertly gathers user information through an Internet connection without the user's knowledge, usually for advertising purposes (called *Adware*, which displays pop-up ads), but sometimes to steal confidential information such as usernames, passwords, and credit card numbers. Spyware applications are typically bundled as a hidden component of shareware or freeware programs downloaded from the Internet. Once installed, the spyware monitors user activity and then covertly transmits that information in the background to someone else.

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 (called *command-and-control* or *CnC* servers) operated for the specific purpose of controlling bots, though some older botnets are directed by the bot herder using Internet Relay Chat (IRC). Bots are often used to commit *denial-of-service attacks,* relay spam, store stolen data, and/or download additional malware to the infected host computer.

Social engineering attacks

Social engineering attacks — such as phishing and baiting — are extremely common. As I discuss in Chapter 3, these attacks, when successful, can lead to much broader, more-sophisticated cyber attacks.

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 email communication. After clicking on a (seemingly innocent) hyperlink, the user is directed 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* targets a specific person or persons within an organization. Attackers will often gather personal information about their target ahead of time to increase the probability of their success.

☑ *Whaling* is directed specifically toward senior executives and other high-profile targets within an organization.

Baiting occurs when a criminal casually drops a USB thumb drive or CD-ROM in a parking lot or cyber café. This drive or disc is labeled with words such as "executive compensation" or "company confidential" to pique the interest of whoever finds it. When the victim accesses the media, it installs malware on his or her computer.

Buffer overflows and SQL injections

Two commonly used techniques that exploit vulnerabilities are buffer overflows and SQL injection attacks.

A *buffer overflow* is a cyber attack where the hacker writes more data into a memory buffer than the buffer is designed to hold. Some of this data spills into adjacent memory, causing the desktop or Web-based application to execute arbitrary code with escalated privileges or even crash. Buffer overflows are commonly triggered by hacker inputs or by malicious files/Web objects 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 Web application to pass the rogue SQL command to the database. A successful SQL injection attack can reveal database content (such as credit card and Social Security numbers, passwords, and more) to the attacker.

Next-generation threats

Traditional signature-based security defenses — including IPS, NGFW, and anti-virus products — are mainly designed to detect known threats. But today, it's the unknown threats that are making the biggest headlines.

This section details the most dangerous cyber attacks facing enterprises and government agencies today. In Chapter 2, I describe why traditional security defenses are inadequate for detecting and preventing them.

Zero-day threats

A *zero-day threat* is a cyber attack on a publicly unknown operating system or application vulnerability, so named because the attack was launched on (or increasingly before) "day zero" of public awareness of the vulnerability — and, in many instances, before the vendor was even aware. (Although in some instances, the vendor is already aware of the vulnerability, but hasn't disclosed it publicly because the vulnerability hasn't yet been patched.)

Zero-day attacks are extremely effective because they can go undetected for long periods (usually several months but sometimes a couple of years), and when they are finally identified "in the wild," patching the vulnerability still takes days or even weeks.

Advanced persistent threats

Advanced persistent threats (*APTs*) (also known as *advanced targeted attacks*, or *ATAs*) are sophisticated network attacks in which an unauthorized person gains access to a network and stays undetected for a long period of time. The intention of an APT is to steal data rather than to cause damage to the network. APTs target organizations in sectors with high-value information, such as credit card processors, government agencies, and the financial services industry.

APTs often use spear phishing (see prior "Phishing and baiting attacks" section) for gaining initial network entry. Once an initial host has been compromised, the APT proceeds using a *slow-and-low* strategy to evade detection.

Chapter 3 is dedicated to the topic of APTs and describes how sophisticated cybercriminals have used slow-and-low tactics in some of the largest data breaches on record.

Although I've proposed a definition for APT in this book, definitions across the information security industry vary. Some say APTs are only committed by nation-states (such as China and Russia) for political motivations, reserving the term ATA for financially motivated attacks. Others use the term to define any "sophisticated" cyber attack, regardless of methodology. However, I believe my definition depicts the majority view of information security professionals.

Polymorphic threats

A *polymorphic threat* is a cyber attack — such as a virus, worm, spyware, or Trojan — that constantly changes ("morphs"), making it nearly impossible to detect using signature-based defenses. Evolution of polymorphic threats can occur in a variety of ways, such as filename changes and compression (file size).

Although the appearance of the code within a polymorphic threat changes with each "mutation," the essential function usually remains the same. For example, a spyware program intended to act as a *keylogger* (unauthorized malware that records keystrokes) will continue to perform that function even though its signature has changed.

The evolution of polymorphic threats has made the jobs of IT security professionals much more difficult. Vendors that manufacture signature-based security products must constantly create and distribute new threat signatures (a very expensive and time-consuming proposition, I might add), while enterprises and government agencies — often with thousands of hosts (especially Microsoft Windows hosts; see "Why Windows is so prone to cyber attacks" sidebar) to protect — are constantly deploying the signatures their security vendors produce. It's a vicious cycle, that is always well behind the cybercriminals, with no end in sight.

Blended threats

A *blended threat* is a cyber attack that combines elements of multiple types of malware and usually employs multiple attack vectors (varying paths and targets of attack) to increase the severity of damage and the speed of contagion. Nimda, CodeRed, and Conficker are a few well-known examples of blended threats.

A blended threat typically includes:

- ☑ Multiple means of propagation
- ☑ Exploitation of operating system and/or application vulnerabilities
- ☑ The intent to cause harm to network hosts

Blended threats are widely considered by information security professionals to be the worst risk to network security since the inception of viruses, as most blended threats require no human intervention to propagate.

Why Windows is so prone to cyber attacks

When you read about high-profile data breaches in the trade press — at least those committed by Internet-borne threats, rather than involving physically stolen laptops or USB thumb drives — virtually all of them result from a cyber attack against a Microsoft Windows host.

So, does that mean that Windows hosts are more prone to cyber attacks? IT security pundits think so and generally offer two explanations — both of which I deem as perfectly valid.

The first explanation merely relates to the near-monopolistic desktop operating system market share that Microsoft enjoys, especially in business environments. Although analyst estimations vary, most show that approximately 9 out of 10 end-user computing devices feature a Windows operating system. Thus, when sophisticated hackers are developing complex worms, Trojans, botnets, and spear-phishing attacks, Windows is clearly the target of choice.

The second explanation has more of a "technical" slant. Windows, as well as Microsoft DOS before it, was designed to be a single-user operating system — frankly, with security as an afterthought. In other words, Windows was designed to let the user have free rein over the entire operating system. However, years later — starting with Windows NT — Windows was modified to support multiple user logins. But rather than re-architecting Windows from scratch as a multi-user operating system, Microsoft chose to preserve compatibility with programs designed for older Windows versions.

In doing so, Microsoft left Windows full of holes (vulnerabilities) for hackers to exploit. So many, in fact, that the second Tuesday of every month is known as "Patch Tuesday," when Microsoft releases new patches (through Microsoft security bulletins) mainly to address vulnerabilities within Windows operating systems. Although Patch Tuesday began in 2004, it's still in full operation today.

Does this mean that non-Windows operating systems are completely safe from viruses, malware, and other cyber attacks? Of course, not. But since Mac OS X and Linux were designed from the ground up with security in mind, these platforms are far less susceptible.

Chapter 2

Understanding the Enemy

In this chapter

- Categorize three kinds of cyber enemies
- Learn how attackers bypass traditional security defenses

"If you know the enemy and know yourself, you need not fear the result of a hundred battles. If you know yourself but not the enemy, for every victory gained you will also suffer a defeat."

— Sun Tzu, *The Art of War*

"Know thy enemy," an important theme from Sun Tzu's infamous *The Art of War* manuscript, certainly ties nicely to the war against cyber attacks (and now, *cyber terrorism*) today. Before we dive into the anatomy of advanced persistent threats (in Chapter 3) and how to defend against them (in Chapter 4), let's spend a little time getting to know the enemies, including understanding their motivations and why they are successful.

Who is the Enemy?

The face of cyber attackers has changed dramatically with the passing of each decade. In the 1970s and 1980s, phone phreaking (unauthorized manipulation of telephone switching equipment primarily to place free long-distance phone calls) was the craze. In 1983, the movie WarGames, starring a young Matthew Broderick, introduced the general public to computer hacking by way of modem, and the legend of hackers as cyber heroes was born.

The 1990s brought widespread Internet adoption, including the emergence of the World Wide Web. Hackers back then defaced public websites primarily for bragging rights — up until the turn of the century.

Hacking has now transformed into a multi-billion-dollar industry. Long gone are the days of hacking just for kicks. Today, there are essentially three types of cyber attackers that enterprises and government agencies must contend with—cybercriminals, state-sponsored threat actors, and hacktivists.

Cybercriminals

Simply put, cybercriminals are individuals who hack for profit. In most instances, they break into company networks in an attempt to steal credit card numbers (sometimes numbering in the tens or even hundreds of thousands) and sell them on the open market. Although not quite as profitable, Facebook, Twitter, and email account credentials sell for a pretty penny, too.

One of the most notorious cybercriminals ever convicted is Albert Gonzalez. In 2010, Gonzalez was convicted of hacking into the databases of a regional credit card payment processing company, stealing more than 170 million credit card numbers over two years. He was sentenced to 20 years in prison — the stiffest sentence imposed on a cybercriminal to date.

State-sponsored threat actors

Arguably the most notable shift in the hacking community within the last decade has been the emergence of state-sponsored threat actors. These are individuals employed by a government (not necessarily their own government) to penetrate commercial and/or government computer systems in other countries to compromise data, sabotage computer systems, or even commit cyber warfare.

China and Russia are among the countries most often cited for recruiting state-sponsored threat actors. But they are not alone. The following are well-known examples of cyber attacks allegedly perpetrated by nation-states, including the United States:

- ☑ Iran is accused of cyber attacks against U.S. banks and oil companies in Saudi Arabia and Qatar (2012).
- ☑ The United States and Israel are accused of creating Flame malware targeting Iran, Syria, and others (2012).
- ☑ China cracks RSA SecurID tokens (2011).
- ☑ The United States and Israel are accused of launching the Stuxnet worm targeting a uranium enrichment facility in Iran (2010).
- ☑ China attacks Google (dubbed "Operation Aurora") to access Gmail accounts of Chinese human rights activists; the same attack targets Adobe, Juniper, Dow Chemical, Northrop Grumman, and others (2009).
- ☑ China steals blueprints for America's new joint strike fighter planes, the F-35 and F-22 (2009).
- ☑ Russia attacks the websites of Estonia's parliament, ministries, banks, and newspapers amid relocation of the Bronze Soldier of Tallinn memorial (2007).

Iran recently launched an ambitious $1 billion governmental program to boost national cyber capabilities. Experts believe that although China's and Russia's cyber war capabilities are vastly superior to Iran's, politically speaking, Iran is significantly more likely to target U.S. cyber infrastructure in light of the ongoing international impasse over Iran's nuclear program.

Hacktivists

Hacktivism is the use of digital tools in pursuit of political ends. Unlike cybercriminals who are motivated by money, hacktivists are motivated by political ideology. Typical cyber attacks committed by hacktivists include website defacements, redirects, information theft, and virtual sit-ins through distributed denial-of-service (DDoS) attacks (overwhelming websites through hundreds or thousands of simultaneous and repetitive connections).

Some hacktivists have joined together to collectively target victims. In 2011, LulzSec claimed responsibility for several

high-profile cyber attacks, including multiple attacks against Sony and the crashing of the U.S. Central Intelligence Agency (CIA) website. In 2012, Anonymous claimed responsibility for crashing several Israeli government websites following Israel's aerial strikes on Gaza.

Cyber mercenaries for hire

In an August 2012 "Defense Dossier" report, the American Foreign Policy Council (AFPC; www.afpc.org) asserted that Russia is allegedly subcontracting some of its state-sponsored cyber attack work to cybercriminals, including members of the once-renowned Russian Business Network (RBN). Until its apparent demise in 2008, RBN was involved in just about every cybercriminal scheme imaginable — phishing, malware, DDoS attacks, and more.

According to the AFPC report, there are two reasons why Russia subcontracts work to cybercriminals — or as I call them, *cyber mercenaries*. First, it's extremely cost effective, as these cyber mercenaries make money on the side when not employed by the state. And second, even after extensive cyber forensics, their cyber attacks can't be traced back to government computers. This is particularly confusing to many Westerners who cannot imagine a government so intertwined with a criminal element.

Russia is not the only nation-state to partner with cybercriminals. Network security expert FireEye (www.fireeye.com), a leader in next-generation threat protection, discovered evidence to support a common theory that China is partnering with cybercriminals to purchase access to already-infected machines as a more streamlined way to infiltrate targeted organizations.

In 2011, FireEye researchers spotted an APT class of malware associated with Ghostnet, a large-scale cyber-spying operation with a command-and-control infrastructure based in China, on a machine also carrying mainstream malware associated with cybercriminal operations.

A coincidence? I don't think so.

How the Enemy Succeeds

Now that you have a solid understanding of the three main types of cyber attackers — cybercriminals, state-sponsored threat actors, and hacktivists — I'll discuss why they're so successful.

Bypassing signature-based defenses

Traditional network and endpoint security products — such as intrusion prevention systems (IPS), next-generation firewalls (NGFW), secure Web gateways, and anti-virus solutions — rely on pattern-matching signatures (sometimes called rules or filters) to detect known cyber attacks and, in some instances, unknown attacks targeting known vulnerabilities.

These security defenses are extremely effective at detecting traditional, known cyber attacks such as worms, Trojans, spyware, botnets, and basic computer viruses. But as I discuss in Chapter 1, they're completely inept at detecting today's new breed of cyber attacks, such as zero-day, targeted attacks, polymorphic malware, blended attacks, and APTs. In fact, in most cases, today's new breed of cyber attacks pass through traditional security defenses as if they weren't even there! That's simply because no signature exists to detect the advanced tactics used in the first stage of an overall attack that ultimately gives cyber attackers free rein within the network.

Don't get me wrong. Traditional signature-based defenses are critical components of a well-balanced defense-in-depth strategy. I'm simply asserting that they're not enough to defend against today's new breed of cyber attacks that cut across communication channels (e.g., Web and email) and take place over multiple stages.

Bypassing anomaly-based defenses

Better IPS and network behavior analysis (NBA) solutions incorporate anomaly-based detection methods to help uncover sophisticated cyber attacks. They work by aggregating flow records (e.g., NetFlow, sFlow, cFlow) from network routers and switches and baselining "normal" network traffic over the course of days or even weeks. Once a baseline has been

established, network anomalies can be detected, such as a host sending exorbitant amounts of data outside the organization or an end-user computing device communicating directly with other end-user computing devices.

Although anomaly-based security defenses can detect certain events caused by next-generation threats, they are largely unsuccessful because they're frequently prone to false positives (misclassifying good traffic as bad). And they're also prone to false negatives (misclassifying bad traffic as good) due to the "slow and low" nature of advanced persistent threats.

Chapter 3

Anatomy of Advanced Cyber Attacks

In this chapter

- Define, in detail, advanced persistent threats (APTs)
- Review high-profile APTs making international headlines
- Understand the life cycle of APT attacks

In Chapter 1, I discuss the differences between traditional cyber attacks and today's new breed of cyber attacks. In Chapter 2, I discuss why next-generation threats are so well equipped to bypass traditional security defenses. Now I'd like to discuss the category of advanced cyber attacks that is, by far, generating the most headlines. I'm talking, of course, about advanced persistent threats, or APTs.

In this chapter, I expand upon the definition of APT provided in Chapter 1. I detail some of the biggest headlines APTs have generated in recent years and then discuss their damaging impact on victimized enterprises and government agencies. I conclude the chapter by describing the APT attack life cycle and provide a list of telltale signs to help you determine whether your network has been compromised by an APT attack.

APTs in Depth

In Chapter 1, I define an APT as "a sophisticated network attack in which an unauthorized person gains access to a network and stays undetected for a long period of time." Although this is quite true, it's only part of the story. APTs are unlike any cyber attack seen before.

The term "advanced persistent threat" was actually created by information security analysts in the U.S. Air Force in 2006. It describes three aspects of the attackers, including their profile, intent, and structure:

- ☑ **Advanced:** The attacker is an expert in cyber-intrusion methods and is capable of crafting custom exploits and tools.
- ☑ **Persistent:** The attacker has a long-term objective and will persistently work to achieve it without detection and without regard for time.
- ☑ **Threat:** The attacker is organized, funded, well trained, and highly motivated.

APTs are widely considered the most dangerous type of cyber attack today. Cybercriminals who employ APTs are a different breed. They're experts at "flying below the radar" to avoid detection as they exfiltrate highly sensitive data from enterprises and government agencies.

Unfortunately, most organizations don't know they've been compromised by an APT attack until it's too late. According to the same 2012 Verizon Data Breach Investigations report referenced in Chapter 1, 59 percent of surveyed organizations that experienced major data breaches in 2011 were notified of the breach by a law enforcement agency!

What APTs are not

As important as it is to understand what APTs *are*, it's equally important to understand what they *are not*. An APT is not a single piece of malware, or even a collection of malware. It is not a single activity and it is never launched without a specific target or objective in mind.

APTs are well coordinated, extended campaigns — whether motivated by financial gain, personal politics, or national interests — intended to achieve an objective against a specific target. As you'll soon discover (see the "APT Attack Life Cycle" section later in this chapter), APTs incorporate multiple cyber attack techniques and take place over several stages to form a single coordinated attack.

Three myths of APT attacks

APTs are among the hottest topics of discussion in the information security world today. Unfortunately, there is almost as much misinformation out there about APTs as there is accurate information. Let's take a moment to reflect on three common myths about advanced persistent threats.

Myth #1: Only specific industries are targets for APTs.

A common misconception is that only large organizations in specific industries are targets of APTs. We know this is false based on the headlines alone (see "APTs in the News" section). APTs have been reported across a broad spectrum of industries, including government, financial services, telecommunications, energy, transportation, and even information security, as demonstrated by the attack against RSA Security in 2011.

Myth #2: APTs target critical endpoints only.

A second myth about APTs is that the perpetrators are targeting high-profile, mission-critical endpoints only, and that end-user devices (laptops and desktops) rarely come into play. This notion is a complete fallacy and is actually the opposite of reality. In virtually all instances, the initial point of entry for an APT is an end-user computing device compromised by a spear-phishing attack, Trojan, or other form of malware.

Myth #3: APTs can be addressed by traditional security defenses.

Virtually every information security vendor claims at least some ability to detect, and in some instances prevent, APT attacks. The truth of the matter is that very few can. Traditional security defenses that incorporate threat-detection signatures (such as IPS, NGFW, and anti-virus solutions) are virtually blind to zero-day attacks and polymorphic threats. Relying on traditional security defenses alone is like showing up to a gunfight with a pocketknife. You simply don't stand a chance.

APTs in the News

These days, it seems like a week can't go by without news of a major data breach at a company, university, or government agency. The following are descriptions of the most newsworthy APT attacks in each of the last four years.

To stay on top of major data breaches affecting commercial and government organizations, I highly recommend the *SC Magazine* Data Breach Blog at www.scmagazine.com/the-data-breach-blog/section/1263/.

Flame (2012)

Flame, also known as Flamer, sKyWiper, and Skywiper, is an APT that was identified in May 2012 by the MAHER Center of Iranian National CERT, Kaspersky Lab, and the Budapest University of Technology and Economics. Kaspersky Lab was asked by the United Nations International Telecommunications Union to investigate reports of a virus affecting Iranian Oil Ministry computers.

Computer experts consider Flame the cause of an attack in April 2012 that caused Iranian officials to disconnect their oil terminals from the Internet. It is now widely asserted that the United States and Israel jointly developed the Flame malware to collect intelligence in preparation for cyber-sabotage aimed at slowing Iran's ability to develop a nuclear weapon.

After the initial exploit stage, Flame begins a complex set of operations including calling back to its command-and-control servers to download other malware modules. When fully deployed, Flame is an uncharacteristically large program for malware at 20 megabytes in size — about 20 to 30 times larger than a typical computer virus. It is widely regarded as the most sophisticated malware ever created. Experts believe Flame, which was designed to masquerade as a routine Microsoft software update, was created to secretly map and monitor Iran's computer networks, sending back a steady stream of intelligence to prepare for a cyber warfare campaign.

RSA SecurID Attack (2011)

In March 2011, RSA Security (a division of EMC) disclosed that it had been victimized by an APT, causing it to notify its SecurID two-factor authentication customers and advise them to swap out their (compromised) token devices. In the months following, reports of data breaches caused, in part, by compromised SecurID tokens began to surface. Most notably, Lockheed Martin released a statement admitting that its network was breached by "sophisticated adversaries," but the

company said no assets were compromised. Some security experts, however, are skeptical as to whether the nation's largest defense contract is being completely forthcoming about a breach on which President Obama was reportedly personally briefed.

Soon after RSA Security disclosed the attack to the public, a company official posted a blog providing intricate details about how the APT attack was perpetrated over several stages. To learn more, read the "RSA Security steps forward to describe its APT attack" sidebar later in this chapter.

In EMC's 10-Q filing, it was disclosed that the APT attack against RSA Security cost the company $81.3 million to replace SecurID tokens, monitor customers, harden internal systems, and handle fallout from the security breach.

Stuxnet (2010)

Stuxnet is a highly sophisticated computer worm discovered in June 2010 that was believed to be in place for over a year and used in conjunction with an APT attack against Iranian uranium enrichment infrastructure. In the first stage, Stuxnet initially spread by exploiting a Microsoft Windows vulnerability and then spread laterally in the network to ultimately reach targeted Siemens industrial software and equipment causing it to malfunction. Although this is not the first time that hackers have targeted industrial systems, it is the first documented case of malware to include a programmable logic controller (PLC) rootkit.

Siemens stated that the worm has not caused any damage to its customers, but the Iran nuclear facility procured embargoed Siemens equipment secretly, which was damaged by Stuxnet during the attack. Interestingly, Stuxnet's multiple spreading mechanisms caused it to eventually escape from the Iranian facility and to infect energy giant Chevron. However, company officials said that Stuxnet identified Chevron as an innocent target and was programmed to withhold its damaging payload, thus ending the attack life cycle. As a result, it caused no damage to Chevron's systems and the company was able to remove it.

Experts have found evidence within the Stuxnet source code linking the APT attack to the United States and

Israel, although officials from both countries have denied the accusation.

Operation Aurora (2009)

Operation Aurora was a high-profile APT attack that began in mid-2009 and continued through December 2009. It was first publicly disclosed by Google in January 2010 in a blog post indicating that the attack originated from China and that it targeted the Gmail accounts of Chinese human rights activists. Dozens more organizations, including Yahoo, Symantec, Northrop Grumman, Morgan Stanley, and Dow Chemical, were also targeted by this attack.

Two days following the attack, McAfee reported the attackers had exploited a zero-day vulnerability in Microsoft Internet Explorer and dubbed the attack "Operation Aurora." Once a victim's system was compromised, the next stage of the attack consisted of a backdoor connection that masqueraded as an SSL connection to command-and-control servers running in Illinois, Texas, and Taiwan, including machines that were running under stolen Rackspace customer accounts. The victim's machine then began its lateral search for sources of intellectual property, specifically the contents of source code repositories.

The Ripple Effect of a National APT Attack

In Chapter 1 (in "The Cost of Failure" section), I itemize common costs that companies face when subjected to a large-scale data breach, including forensics costs, regulatory fines, and lost revenue. But what if APT threat actors decided to target something a little more strategic than corporate data?

Imagine, if you will, a coordinated APT attack against power companies in a large region of the United States — say, the Northeast. Let's further imagine that the perpetrators of the attack were successful in compromising the SCADA (supervisory control and data acquisition) systems that control a multi-state power grid, knocking out power for days or even weeks. Can you imagine the "ripple effect" such an attack might cause?

- ☑ Electric power grids crash
- ☑ Gas stations can't pump fuel
- ☑ ATMs can't dispense cash
- ☑ Grocery stores are depleted
- ☑ Hospitals and emergency services can't keep up

Think such an attack is impossible? Think again. According to a 2012 report titled "Terrorism and the Electric Power Delivery System" from the National Research Council, a successful cyber attack on a regional power grid would make Hurricane Sandy look like nothing. Internet-delivered malware designed to destroy control systems could black out large regions of the nation for weeks or months causing widespread civil unrest. According to the report, damage from such an attack would cost many billions of dollars more than the destruction caused by Hurricane Sandy against the East Coast in 2012.

To access the National Research Council report, connect to www.nap.edu/catalog.php?record_id=12050

APT Attack Life Cycle

The anatomy of advanced persistent threats varies just as widely as the victims they target. However, cybersecurity experts researching APTs over the past five years have unveiled a fairly consistent attack life cycle consisting of five distinct stages:

- ☑ **Stage 1:** Initial intrusion through system exploitation
- ☑ **Stage 2:** Malware is installed on compromised system
- ☑ **Stage 3:** Outbound connection is initiated
- ☑ **Stage 4:** Attacker spreads laterally
- ☑ **Stage 5:** Compromised data is extracted

Let's now explore each of these five APT life cycle stages in more detail.

Stage 1: Initial intrusion through system exploitation

System exploitation is the first stage of an APT attack to compromise a system in the targeted organization. By successfully detecting when a system exploitation attempt is underway, identification and mitigation of the APT attack is much more straightforward. If your defenses cannot detect the initial system exploitation, mitigating the APT attack becomes more complicated because the attacker has now successfully compromised the endpoint, can disrupt endpoint security measures, and hide his actions as malware spreads within the network and calls back out of the network.

System exploits are typically delivered through the Web (remote exploit) or through email (local exploit) as an attachment. The exploit code is embedded within a Web object (e.g., JavaScript, JPG) or file (e.g., XLS, PDF) to compromise the vulnerable OS or application enabling an attacker to run code, such as connect-back shellcode to call back to CnC servers and download more malware.

In the attack against RSA Security in 2011, an employee was tricked into opening an email with the subject of "2011 Recruitment plan.xls," which included a malicious Microsoft Excel spreadsheet attachment that successfully exploited the system using a zero-day Adobe Flash vulnerability. (For more details, read the sidebar titled "RSA Security steps forward to describe its APT attack" later in this chapter.)

System exploit code is developed by attackers to corrupt memory or cause a buffer overflow condition within the vulnerable OS or application enabling arbitrary code execution. In the case of local exploits, oftentimes social engineering is used to initiate the necessary user interaction needed to complete the infection. In the case of remote exploits, such as Web drive-by downloads, no user interaction is required beyond visiting the Web page.

Stage 2: Malware is installed on compromised system

Once a victim system is exploited, arbitrary code is executed enabling malware to be installed on the compromised system. Visiting a Web page or a simple double-click of the mouse is all it takes for the user's system to become compromised and infected with the malware payload.

Not all spear phishing emails originating from an APT threat actor contain attachments. Many contain hyperlinks that, when clicked on by the user, open a Web browser (or sometimes another application, such as Adobe Reader, Microsoft Word, or Microsoft Excel). Each link is then redirected to a hidden address with a base64-encoding key. The hidden address refers to a *dropsite*, which assesses the browser for known vulnerabilities and returns a Trojan downloader. Upon execution, the downloader conveys a base64-encoded instruction to a different dropsite from which a Trojan (malware) is delivered.

Stage 3: Outbound connection is initiated

The malware installed during the prior stage often contains a remote administration tool, or RAT. Once up and running, the RAT "phones home" by initiating an outbound connection, often an SSL-encrypted channel, between the infected computer and a CnC server operated by the APT threat actor. APT threat actors go to this trouble to establish outbound callbacks to bypass traditional and next-generation firewalls, which allow session traffic to flow bi-directionally if initiated from within the trusted network.

Once the RAT has successfully connected to the CnC server, the attacker has full control over the compromised host. Future instructions from the attacker are conveyed to the RAT through one of two means — either the CnC server connects to the RAT or vice versa. The latter is usually preferred as a host initiating an external connection from within the network is far less suspicious.

Stage 4: Attacker spreads laterally

It's highly unlikely that the initially breached end-user computing device contains strategic data. So the APT attacker

must spread laterally through the network to search for hosts operated by IT administrators (in an effort to steal administrative credentials) and high-value servers and databases containing sensitive data — the ultimate target of the APT attack. This is how Flame operated.

Lateral movement does not necessarily involve the use of malware or tools other than those already supplied by the compromised host operating system, such as command shells, NetBIOS commands, VNC, Windows Terminal Services, or other similar tools used by network administrators to service remote hosts. Once the ultimate target has been identified and adequate logon credentials are possessed, the attacker's hard work and determination begin to pay off.

Stage 5: Compromised data is extracted

In this stage of the network breach, the APT attacker has three obstacles to contend with. First, transferring all of the target data at once (targeted data is often quantified in gigabytes) could trigger a flow-based anomaly alert (if NBA technology is used; see Chapter 2) due to an unusually high volume of traffic initiated by the targeted server or database. Second, the attacker needs to ensure that the host receiving the data can't be linked back to him (or her). And third, transferring data as plain text could trigger an alert from a data leakage prevention (DLP) system. Let's explore how experienced APT threat actors overcome all three obstacles.

To overcome the first obstacle, a savvy APT attacker will exfiltrate data from the target server or database in "chunks" — perhaps in increments of 50-100 megabytes. One strategy is to group files or records together into compressed, password-protected RAR files.

Some RAR files can be parts of multi-volume sequences, enabling the attacker to split a large quantity of data into volumes. Each RAR file would have an extension to depict the number of the volume, such as part1.rar (the first volume), part2.rar, part3.rar, and so on.

The second obstacle is a little more challenging. The attacker wants to get the data offsite as soon as possible, but can't risk sending it to a host that can be traced back to the attacker. To

overcome this challenge, the attacker might select for a staging area a virtual host that is hosted by a cloud-based service provider. That way the host can be instantly destroyed after the data has been extracted.

The third and final obstacle in this phase can be accomplished by encrypting each RAR file before it is transferred (often via FTP) to the staging host. Most RAR files support strong AES 128-bit encryption, which is more than sufficient.

Attacker covers his tracks, remaining undetected

If an enterprise or government agency has any hope of detecting an APT on its own, it's far more likely to happen while the attack is still in progress. This is because most APT attackers are extremely good at covering their tracks.

The following are tactics that APT attackers employ during and after the attack to minimize the risk of detection:

- ☑ Planting malware to distract the IT security staff and keep them busy doing other things.
- ☑ Spreading to network file shares, which are relatively unprotected and only completely wiped in extreme circumstances.
- ☑ Deleting the compressed files after they've been extracted from the staging server.
- ☑ Deleting the staging server if it's hosted in the cloud or taking it offline if under control by the attacker.
- ☑ Uninstalling malware at the initial point of entry.

RSA Security steps forward to describe its APT attack

Following RSA Security's March 2011 data breach (described earlier in this chapter; See "APTs in the News" section), the company posted details in its corporate blog describing exactly how the attack occurred.

According to a company official, the attack started with a spear phishing attack that targeted specific company employees possibly identified through social media sites. In this case, the attacker sent two different spear phishing emails over a two-day period to two small groups of employees with a subject of "2011 Recruitment plan.xls" and a Microsoft Excel spreadsheet attachment.

The email was crafted well enough to trick one of the employees into retrieving the email from their junk mail (spam) folder and then double-clicking on the attached Excel file. Unbeknownst to the user, the spreadsheet contained a zero-day exploit that installed a RAT through an Adobe Flash vulnerability. Once the RAT was in place, it initiated an outbound connection and the attacker gained full control of the user's machine.

As the initially compromised PC was not a strategic asset, the attacker's next tactic was to move laterally inside the network by compromising additional hosts. He first harvested access credentials from the first compromised PC, including credentials to a domain admin account. The attacker then performed privilege account escalation on non-administrative users on other systems. He repeated this process until he stumbled across a high-value target — a computer operated by an IT server administrator.

Soon after, the attacker located highly sensitive servers (allegedly containing top-secret SecurID two-factor authentication algorithms), compromised them, and established access to staging servers at key aggregation points to get ready for extraction. Then the attacker went into the servers of interest, removed data, and moved it to the staging servers where the data was aggregated, compressed, and encrypted for extraction.

Finally, the attacker used FTP to transfer many password-protected RAR files from the RSA file server to an outside staging server at an external, compromised machine at a hosting provider. The files were subsequently pulled by the attacker and removed from the external compromised host to remove any traces of the attack.

On a personal note, I applaud RSA Security for coming forward with precise details of this APT attack. By learning these intricate details, organizations can gain insights from RSA's misfortune and perhaps implement new strategies and technologies — including next-generation threat protection (see Chapter 4) — to help mitigate the real risk of APT attacks.

Telltale Signs of an APT Attack

Although APTs are extremely difficult to detect, the following is a list of common telltale signs that your organization may have been compromised by an APT.

- ☑ Finding system exploit code embedded in email attachments or delivered via Web pages.
- ☑ Increase in elevated logons late at night.
- ☑ Outbound connections to known CnC servers.
- ☑ Finding widespread backdoor Trojans on endpoints and/or network file shares.
- ☑ Large, unexpected flows of data from within the network — from server to server, server to client, client to server, or network to network.
- ☑ Discovering large (I'm talking gigabytes, not megabytes) chunks of data appearing in places where that data should not exist.

Be especially wary if you find compressed data in formats not normally used by your organization.

- ☑ Abnormal SSL-encrypted network communications.
- ☑ Windows Application Event Log entries of anti-virus and firewall stop and restart commands.

A major reason why organizations fail to identify APT attacks is because their security devices are only (or mainly) configured to examine inbound traffic at the perimeter. Acquiring and/or configuring security solutions to inspect outbound traffic significantly improve your chances of detecting APTs and other cyber attacks.

I hope you never have to face cleaning up from an APT attack. If you do, it will be one of the most challenging things you've ever had to do in your information security career. Prevention and early detection — through the use of next-generation threat prevention technology — is the best way to minimize the potential of being victimized by an APT attack. To learn more about this innovative new category of network security technology, turn to Chapter 4.

Financial services CSO counts on FireEye to close its IT security gap

Recently, the chief security officer (CSO) of a large multinational financial services firm — a member of the S&P 500 with nearly 10,000 employees — completed an assessment of his organization's network security defenses and found a gap that needed to be filled, and fast.

The gap the CSO discovered pertained to defending against a new generation of IT security threats that traditional signature-based security defenses simply cannot detect. These threats include zero-day attacks, polymorphic malware, blended threats, and the most damaging of them all, APTs.

The CSO instructed his team to evaluate all available advanced threat protection solutions on the market to determine which ones are best equipped to detect and prevent this new class of nefarious attacks. After researching more than a half-dozen offerings, he and his team narrowed shortlist down to two vendors, one of which was FireEye.

The CSO decided to evaluate the two competing solutions concurrently. He tested the FireEye Web Malware Protection System (MPS) appliance against a comparable appliance from a competing vendor. Both boxes were configured for inline operation and both monitored identical perimeter traffic. The two solutions were tested side-by-side for a period of six weeks.

The results of the dual evaluation proved to be conclusive. The FireEye appliance found two to three times more legitimate threats than the competing solution, with zero false positives. Although the competing appliance generated more alerts than the FireEye MPS, the IT security team was able to prove that it was because the competing box generated numerous false positives.

Choosing FireEye over the competition was an easy decision. Not only does FireEye offer the best advanced threat protection available for inspecting inbound traffic, but its callback filter (see Chapter 4) makes it easy to detect outbound connections to CnC hosts for malware hand-carried into the organization on laptops or other mobile devices.

Several months later, the company has been well protected from today's new breed of cyber attacks with attempts detected and prevented on an almost-daily basis. Although life offers no guarantees, this company's CSO is now sleeping much better at night.

Chapter 4

Introducing Next-Generation Threat Protection

In this chapter

- Visualize an ideal solution for mitigating the new breed of cyber attacks
- Define next-generation threat protection and review its key components
- Compare next-generation threat protection to traditional signature-based defenses and sandbox technologies

Today's corporations, universities, and government agencies are experiencing unprecedented cyber attack activity — both in number and sophistication. In a never-ending game of cat and mouse, the cat currently has the upper hand. And unless your organization is prepared, you may be its next victim.

In prior chapters, I hope you've gained an appreciation for how serious today's next-generation threats are and why traditional security defenses are helpless to stop them. Now it's time to unveil a new category of network security defense — which up until just recently didn't even exist. I'm talking, of course, about next-generation threat protection.

I begin this chapter by discussing — almost daydreaming about — what's really needed to combat today's most sophisticated cyber attacks. Then I segue into next-generation threat protection, starting out with a definition followed by details of its key components and features.

We've got a lot of ground to cover in this chapter. Let's start by discussing what the world really needs to stay ahead of next-generation threats.

What the World Really Needs

In a perfect world, there would be no cyber attacks. There would be no such things as malware, Trojans, or APTs. And businesses, universities, and government agencies wouldn't need to spend over $20 billion per year to stop them.

Unfortunately, we don't live in a perfect world. Money and politics are fueling cybercriminals, hacktivists, and state-sponsored threat actors to use every tool at their disposal to break into your organization's network. Since we don't live in a perfect world, let's talk about what the world needs to stay ahead of the bad guys.

Signature-less defenses

Organizations today need to explore a new threat protection model in which their defense-in-depth architecture incorporates a signature-less layer that specifically addresses today's new breed of cyber attacks.

Although traditional security defenses are critical for blocking known cyber attacks, experience has shown that it's the unknown cyber attacks that are most worrisome, and on the rise. And since these zero-day, polymorphic, and APTs are largely unknown and becoming the new norm for successful breaches, the world needs a signature-less solution to stop them.

Protection — not just detection

Before there were intrusion prevention systems (IPS), there were intrusion detection systems (IDS). An IDS, by design, can only detect known threats (or unknown threats targeting known vulnerabilities). As time progressed, organizations demanded that their IDS not only detect but also block cyber attacks. Thus, IPS was born.

In that vein, the world needs an advanced threat protection platform that not only detects the needle in the haystack, but blocks it, too, across all potential entry vectors.

Multi-stage protection architecture

In a perfect world, IT would maintain full control of every computing device on the network. Then you'd only have to worry about cyber attacks originating from outside the network and attempting to penetrate it through the perimeter.

Of course, with mobile computing on the rise and IT being compelled to implement bring your own device (*BYOD*) policies, sometimes cyber attacks are hand-carried right through the office front door. What the world needs is an advanced threat protection solution that not only monitors cyber attacks from the outside in, but the inside out, as well — across all stages as they attempt to call back out or spread laterally through the network. If you can't stop threats from entering through the Web, email, or the office front door, then at least stop them from communicating out and spreading further.

Highly accurate detection engine

As with traditional signature-based defenses, detection accuracy is king. What the world needs to adequately defend against next-generation threats is an advanced threat protection solution that is highly accurate, with no false positives (good files classified as bad) and no false negatives (bad files classified as good).

False positives and false negatives are products of security platforms with poor detection capabilities. False positives are mainly a "nuisance" as they consume valuable security analyst cycles chasing after false alarms. False negatives, on the other hand, can be "company killers" as advanced malware passes right through the network security device completely undetected.

Backed by global threat intelligence

Every cyber attack has a "ground zero" — a single host that is the first target on Earth to ever experience a given cyber attack. What the world needs is a mechanism for allowing advanced threat protection systems to share intelligence, not only within a single organization, but also among different organizations globally.

We may not live in a perfect world. But there is an ideal solution for combating today's most sophisticated attacks.

Introducing next-generation threat protection.

Defining Next-Generation Threat Protection

Next-generation threat protection (NGTP) is a new breed of network security technology specifically designed to identify and defend against today's new breed of cyber attacks. Intended to augment — not replace — traditional security systems, NGTP represents a new layer in the defense-in-depth architecture to form a threat-protection fabric that defends against those cyber attacks that go unnoticed by common signature-based defenses.

NGTP platforms customarily ship on high-performance, purpose-built rackmount appliances. Preferred NGTP vendors offer an integrated platform that inspects email traffic, Web traffic, and files at rest, and shares threat intelligence across those attack vectors.

NGTP platforms are unlike any network security offering on the market. NGTP appliances inspect traffic and/or files looking for thousands of suspicious characteristics, including obfuscation techniques like XOR encoding and other disguising behavior. Sessions are replayed in a (safe) virtual execution environment (think virtual machines, but using a custom-built virtualization engine specifically designed for security analysis) to determine whether the suspicious traffic actually contains malware (more on this in the "How it Works" section later in this chapter).

Call the bomb squad!

I've always been a big fan of using real-world analogies — and, in some instances, clichés — to describe how a given technology functions or the benefits it provides. When it comes to next-generation threat protection, I think I've found a good one.

At first, I considered the "finding a needle in a haystack" analogy. That's certainly fair to use when discussing NGTP solutions, but these days large enterprises and government agencies are targeted with advanced cyber attacks several times each day. So instead of looking for one needle in a haystack of cyber attacks, NGTP is really searching for dozens.

Then I considered a crash test dummy analogy, where the suspicious traffic component (suspected malware) is the dummy and the Microsoft Windows session running in the virtual execution engine (replicating the target environment) is the car. This analogy also isn't bad, but it breaks because the car is going to crash every time regardless of what's inside. Often files examined by an NGTP appliance are simply benign.

After searching long and hard for a better analogy, I think I've found a good one. Ever hear accounts in the news of bomb squads being called to examine a suspicious bag left at the airport or in a busy area like Times Square in New York City? In these instances, the bomb squad sends in a robot to examine the suspicious bag and, if necessary, pick it up and place it in a bomb disposal truck capable of withstanding massive explosions without affecting its surrounding area.

Comparing this analogy to an NGTP solution, the person who called the bomb squad is like the malware detection algorithm. The robot is analogous to the subsystem responsible for redirecting the suspected malware into the bomb disposal truck. And the truck is like the virtual session used to "prod" the suspected malware (the suspicious bag) to determine any potential damaging effects (an explosion), but in a safe and secure environment.

I hope this analogy helps you understand how NGTP technology works and provides an easy way for you to explain it to less-technical colleagues.

Comparison to traditional signature-based defenses

In Chapter 2 (see the section titled "How the Enemy Succeeds"), I describe why and how today's new breed of cyber attacks are able to bypass traditional signature-based security defenses such as firewalls (with threat signatures),

IPS devices, secure email and Web gateways, and anti-virus solutions). But as I hadn't yet introduced NGTP technology, I didn't outright compare these defenses to NGTP solutions. I'll remedy that now.

Table 4-1 provides a summary of how traditional signature-based defenses compare to NGTP solutions. But remember, NGTP solutions are a new signature-less layer in your architecture to augment signature-based defenses.

Protection Comparison	*NGTP Solutions*	*Traditional Defenses*
Detect known malware using IPS-style signatures	✓	✓
Identify attacks within encoded binaries	✓	✗
Replay suspected traffic in a safe virtual environment	✓	✗
Inspect outbound traffic to stop dynamic callback channels	✓	✗
Auto-generate threat intelligence to defend against targeted attacks	✓	✗

Table 4-1: Comparison of NGTP to traditional defenses.

The protections in Table 4-1 will be explored in detail in the "Key Features" section later in this chapter.

Comparison to sandbox technologies

A *sandbox* is essentially a small, self-contained version of a (typically Windows-based) computing environment offering a minimal suite of applications and services. It was originally developed for software developers to test new programming code in a safe, non-production environment. Sandboxing technology was later adopted by information security professionals as a way to *manually* examine suspicious binaries without compromising production systems. The operating system and applications contained in the sandbox (virtual machine) typically match the organization's desktop standard so suspected malware can exploit those same inherent vulnerabilities.

Organizations with multiple desktop configuration standards — as is the case with most large organizations — are particularly at risk when relying on so-called NGTP solutions that incorporate rudimentary sandboxing technology. If a malware-infected file is analyzed within a virtual execution environment equipped with operating systems and/or applications that don't mirror the malware's target environment, then that file may be classified as good, resulting in a potentially serious false-negative condition.

As the name implies, a sandbox serves as a safe environment for "exploding" (see "Call the bomb squad!" sidebar earlier in this chapter) potential malware and examining its intended effects. However, by itself it is not a scalable analysis technique given the number and volume of suspicious objects and file types used to hide exploit code. Also, unfortunately, cybercriminals and APT threat actors can detect whether their malware is being executed in a sandbox environment, and if it is, to quell its damaging payload. For these reasons, traditional sandbox technologies are simply no match for sophisticated threats.

One more word of caution on the subject of sandbox technology. Regardless of what an NGTP vendor calls it, if the sandbox component is hosted by the vendor "in the cloud," security, performance, and privacy considerations come into play (see "Don't get stuck in the cloud" sidebar). Preferred NGTP solutions offer high-performance, dedicated appliances enabling suspected malware to be tested onsite within seconds — rather than minutes or hours — while always preserving the privacy of your data.

Don't get stuck in the cloud

Some NGTP vendors tout their ability to perform virtual malware analysis in the cloud, minimizing the horsepower (CPU, memory, disk) required in their hardware appliances. In this case, the vendor's appliances only inspect a small subset of traffic and redirect suspicious objects to the cloud, rather than examining the object in an on-box virtual testing environment.

This sounds nice, as a cloud-based architecture offloads the analysis, but when you peel back the layers of this onion, there are three problems with this design — security coverage, scalability, and privacy.

Malicious code can be embedded in dozens of different file types and Web objects, but virtual malware analysis in the cloud typically only analyzes two or three file types. Cloud-based NGTP solutions are often easy to bypass simply by embedding malicious code within a file format not analyzed.

Beyond limited security coverage, without accurate exploit detection or pre-filtering heuristics, exporting all binaries or PDFs to the cloud to be placed in an inspection queue (among thousands of other requests) does not scale. This inefficient analysis delays determining whether you've found malware as well as provides attackers an easy bypass tactic. They simply flood the pipe with benign binaries or PDFs.

Lastly, when binaries or PDFs are exported to the cloud for inspection, there is always a risk that a sensitive file could be exported to the cloud as well, unbeknownst to your organization. This is a particular concern for certain European countries with strict privacy laws.

For these reasons, market-leading NGTP vendors offer high-performance, purpose-built appliances that not only inspect traffic and block threats, but also perform virtual malware testing within your own environment.

Key Components

Now that you have a sense for what next-generation threat protection is all about — including how it compares to traditional security defenses and how it differs from ordinary sandboxing technology — let's dive deeper and explore the key components of leading NGTP solutions.

Malware protection system

At the heart of every NGTP solution is the malware protection system (MPS). This component analyzes suspicious object types contained in Web traffic, email messages, or files at rest. It also blocks known threats using MPS-generated threat intelligence to stop inbound threats and unauthorized outbound communications.

The tactics of discerning potential malware in Web traffic and in email messages are different. Some vendors offer a one-size-fits-all MPS that attempts to detect threats in all three mediums (or two, if unable to inspect files at rest). Avoid such suboptimized solutions, as they typically suffer from high false-positive and false-negative rates. Also, avoid so-called NGTP solutions that combine MPS functions with other network security components, such as firewall, IPS, and application control. These solutions typically offer "rudimentary" inspection capabilities limited to just .exe, .pdf, and/or .dll files.

Virtual execution engine

Earlier in this chapter, I discussed the role of the virtual execution engine and how it compares to ordinary sandbox technology. I can't stress enough how critical this component is to the efficacy of a next-generation threat protection system — and to your organization's ability to defend against today's advanced cyber attacks.

A good virtual execution engine can programmatically filter for suspicious objects, profile the target victim, and then execute the suspicious code against the intended OS and application(s) to maximize the chances of detonating the system exploit and accompanying malicious payload. Meanwhile, it should yield virtually no false positives or false negatives. It can mean the difference between success and failure in the war against today's new breed of cyber attacks.

Better NGTP offerings contain virtual execution engines capable of inspecting dozens of file types (rather than just .exe and .dll files), including: asf, com, doc, docx, dll, exe, gif, ico, jpeg, jpg, mov, mp3, mp4, pdf, png, ppsx, ppt, pptx, qt, rtf, swf, tiff, unk, vcf, xls, xlsx, zip... and the list goes on.

Once a suspected threat has been classified as malware by the virtual execution engine, new threat intelligence is automatically created and distributed to your other MPS appliances (if you have a central management system; see next section) and possibly to other organizations around the world (see "Cloud threat intelligence network" section just ahead).

Central management system

Although most NGTP appliances provide a Web-based graphical user interface (GUI) for local administration, better NGTP vendors offer both a local GUI and a separate central management system appliance for centralized management, consolidated threat monitoring, reporting, alerting, and malware intelligence distribution.

Cloud threat intelligence network

Earlier in this chapter, in the "Don't get stuck in the cloud" sidebar, I discussed why performing malware inspection in the cloud is a bad idea for security, scalability, and privacy reasons. But the cloud is an ideal place for hosting one key NGTP component — the *cloud threat intelligence network*.

The cloud threat intelligence network interconnects all MPS appliances — at least those from vendors that offer such a service — to share threat intelligence (malware profiles and callback destinations) for newly discovered cyber attacks.

NGTP vendors that offer a *cloud threat intelligence network* (not all of them do) typically give their customers two options — (1) the ability to receive the threat intelligence and (2) the ability to share and receive threat intelligence. Most organizations choose the latter option as vendors usually offer a discount for sharing intelligence.

Since only metadata from infectious objects is needed to create threat intelligence, organizations need not worry about the potential for sensitive data to leave the network.

Now that you're grounded in the basic components of an NGTP solution and you understand how it differs from traditional signature-based defenses and sandbox technology, turn to Chapter 5 to gain deeper insight into how an NGTP solution really works.

Chapter 5

Next-Generation Threat Protection Explored

In this chapter

- Understand how NGTP mitigates next-generation threats in email messages, Web communications, and files at rest
- Explore the key features found in leading NGTP solutions
- Integrate NGTP into your existing network infrastructure

Traditional signature-based security defenses are simply no match for today's new breed of cyber attacks, such as zero-day attacks, polymorphic malware, blended threats, and most importantly, APTs. A new generation of cyber attacks requires an entirely new way of thinking.

Now that you have a sense of what next-generation threat protection is all about (from Chapter 4), let's roll up our sleeves and understand how it works, what features are important, and how an NGTP platform can integrate into your existing IT infrastructure.

How It Works

Let's start out by describing where to place your MPS appliances in relation to how threats enter the organization. Typically, administrators will place MPS appliances as the last line of defense to inspect Web and/or email traffic for threats that have bypassed the firewall and IPS. Web MPS appliances are placed behind secure Web gateways while email MPS appliances should be positioned behind anti-spam and secure email gateways (but in front of the enterprise email server). See Figure 5-1 for a typical Web and email MPS deployment.

MPS appliances should also be deployed in the datacenter to inspect file shares to stop the lateral spread of malware and protect sensitive data.

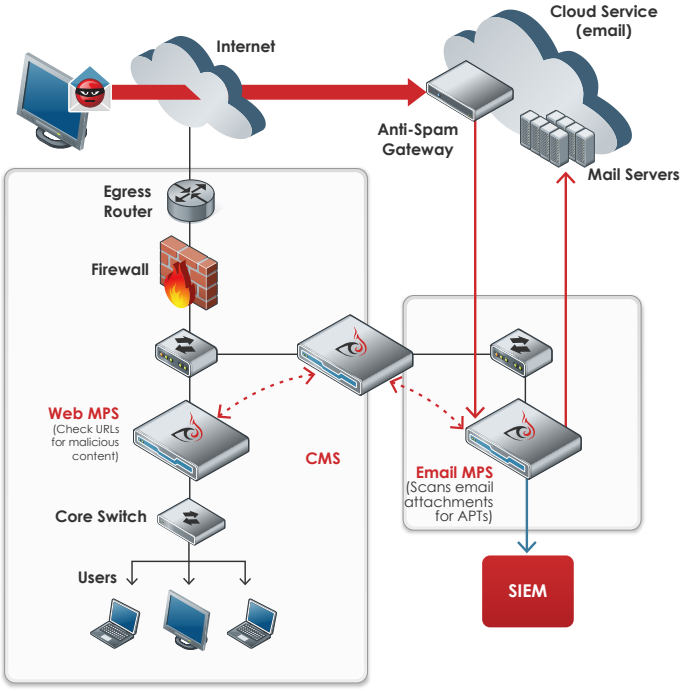


Figure 5-1: Typical NGTP implementation diagram.

APT attacks take place over multiple stages. An MPS platform should be designed to protect at every stage, from initial exploit to data exfiltration, which maximizes your capability to defuse the attack and prevent a breach.

With the MPS platform deployed, the following is a step-by-step depiction of how a typical NGTP system functions:

Step 1: MPS inspects inbound and outbound traffic (and files at rest) looking for known threats, CnC callbacks, spear phishing, and suspicious binaries/Web pages. If a known threat or CnC callback is detected, the connection is blocked and an alert is triggered.

Step 2: For unknown, zero-day attack detection, the MPS identifies a suspicious binary, attachment, or Web page and

replays it in the virtual execution engine (within the same appliance) for analysis. This is a byte-by-byte reconstruction of the identical suspicious flow going to the targeted victim.

When evaluating an NGTP solution, be sure its MPS component is capable of inspecting more than just HTTP traffic because threats use dozens of protocols, such as HTTP, FTP, IRC, custom protocols, and more.

Step 3: Virtual execution engine launches with a specific (patched or unpatched) Microsoft Windows operating system and relevant application(s) (e.g., Microsoft Office, Microsoft Internet Explorer, Adobe Reader) of the host targeted by the suspected threat. The object is replayed and observed for malicious behavior, including corrupting a root file system, attacking an application using a heap spray, registering a new Windows service, or calling back to a known infection URL. If the binary is determined to be benign, the event is logged and the virtual machine is reset.

Step 4: If zero-day malicious activity is confirmed, the virtual machine captures the rest of the attack life cycle. The malware binary is loaded and all malware-generated host activities and network traffic is recorded. Threat intelligence is generated to stop associated callback traffic across the network, a high-priority alert is logged, malware forensics are recorded, and a new malware protection profile is created to block this now known threat.

Because the callback traffic never leaves the actual host's network and the attacker is never notified, the attack has failed and sensitive data remains safe in the organization.

Step 5: The new threat intelligence is forwarded to the central management system (CMS) appliance, where it is distributed to other MPS appliances in the organization. If the organization has subscribed to the cloud threat intelligence network (see Chapter 4), then all participating organizations are instantly protected.

Inline and out-of-band deployments

Organizations have two choices for deploying their email and Web MPS appliances. They can be configured for inline (active) or out-of-band (passive) operation. (File share MPS appliances are always configured for out-of-band operation for

inspecting files at rest and can quarantine malicious file share objects, if so configured.)

Inline deployments allow an organization to block newly identified advanced cyber attacks from calling back to CnC servers to prevent future occurrences of the now known attack. The deployment places the MPS directly in the flow of traffic (just like an IPS or a message transfer agent, or MTA).

MPS appliances can also be configured for out-of-band operation in either *monitor-only mode*, where the MPS engine alerts on detected cyber attacks (just like an IDS), or *TCP RESET mode*, where the MPS submits TCP RESET packets to each session partner to disrupt nefarious TCP sessions.

For out-of-band operations, connect the Web or email MPS to a mirrored switch SPAN port. If no SPAN port is available, a network TAP (from Gigamon, VSS Monitoring, NetOptics, and others) will suffice (or, configure upstream MTAs to send a blind carbon copy to the email MPS).

Many organizations start with out-of-band monitoring to gauge the system's accuracy and stability. Once the organization is comfortable with the NGTP system, the MPS appliances are reconfigured for inline blocking mode.

If you intend to install your MPS appliances for inline operation, be sure to select models that support fail-open connectivity. In the (unlikely) event the appliance was to lose power or otherwise become disabled, traffic would continue to pass through its copper interfaces, rather than bringing the network to a screeching halt. (Fiber interfaces will always pass traffic, regardless of the appliance's state.)

Key Features

So now that you know what NGTP is all about and have some insight into how it works, let's review the key features that typically comprise today's leading NGTP solutions to address every stage of an APT attack life cycle.

NGTP feature sets vary widely. As you review these features, take note of which ones are particularly applicable to your organization. Then refer to Chapter 6 for additional NGTP buying considerations.

Virtual execution of suspicious objects

Signature-less analysis to detect unknown threats is critical. Look for the ability to replay traffic containing suspicious objects — such as Web pages, binaries, and files — in the safety of a virtual execution environment. This is different than forwarding a suspicious file or executable to a sandbox. Replay is the byte-by-byte capture and reconstruction of the traffic flow within a virtual execution environment. A single Web page, for example, is made up of 20 to over 200 different objects served from dozens of different Web locations. Replay technology is the only way to analyze complex Web-based attacks, such as drive-by download attacks.

Don't get trapped into thinking that today's new breed of cyber attacks can only be present in exe or dll files. As I touched on in Chapter 4, malware can be embedded in Web pages and dozens of file types.

The ability of the virtual execution engine to minimize the potential of false positives and false negatives is imperative. A false positive (good file misclassified as bad) could block important content from reaching its destination. A false negative (bad file misclassified as good) could be devastating — especially if a file containing advanced malware pertaining to an APT attack is allowed to pass.

Fast-path blocking

Blocking outbound callbacks and now known inbound threats goes hand-in-hand with signature-less analysis. Although I've spent a fair amount of time describing the limitations of traditional signature-based defenses, I've also acknowledged how important they are to stop known attacks in a defense-in-depth strategy. Look for NGTP solutions that incorporate both signature-based and signature-less techniques to defend efficiently against known and unknown attacks, respectively.

However, in the event your IPS, NGFW, secure Web gateway, anti-spam, or other network security device misses an attack containing commonly known malware, the fast-path blocking capability of an (inline) MPS appliance is designed to instantly block or quarantine the attack.

Once advanced malware successfully penetrates the organization (remember, it could be hand-carried into the office on a mobile computing device), it attempts to call back to the CnC host to either download RAT software (see Chapter 3) or to receive instructions from the attacker. The MPS appliance is equipped to stop sessions connecting to malicious URLs and known-bad IP addresses, or those utilizing custom malware protocols. If the callback filter of the MPS appliance detects an attempt by an internal host to connect to a known external CnC host, the connection is blocked (assuming the MPS is configured for inline operation) and an alert is triggered (see Figure 5-2.)

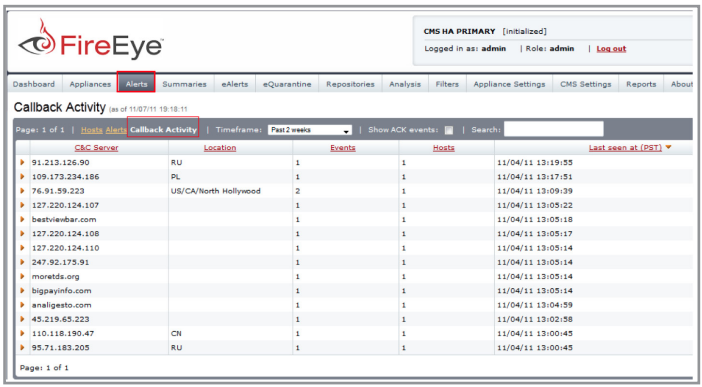


Figure 5-2: Sample FireEye callback activity.

Unlike most traditional security devices, MPS appliances are designed to inspect both inbound and outbound Internet traffic.

Malicious file quarantine

Malicious files, emails, and related attachments detected by the virtual execution engine can be quarantined and stored on the MPS appliance (or the central management console, if available; see next section) for further forensic analysis beyond the forensics provided by the MPS. The files may also be collected as digital evidence by computer crime investigators from the FBI or other law enforcement authorities.

Centralized management

For organizations with three or more MPS appliances, I highly recommend acquiring a centralized management appliance (sometimes called a *central management system*) to centrally monitor and manage your NGTP system via an easy-to-use Web-based interface. Typical tasks performed using a centralized management console include:

- ☑ Aggregate event data from all MPS appliances and process the data into dashboards, reports, and alerts.
- ☑ Centrally aggregate quarantined malicious objects that contain malware.
- ☑ Aggregate and disseminate threat intelligence generated by internal MPS appliances and from the malware protection cloud; upload threat intelligence to the malware protection cloud, if permitted to do so
- ☑ Configure MPS appliance settings and apply them to each MPS appliance individually or in groups.
- ☑ Download and apply software updates to all MPS appliances from one central location.
- ☑ Monitor the performance of all MPS appliances.
- ☑ Export event data to SIEMs (security information and event management), incident management systems, or other external applications.
- ☑ Control user access and administrative privileges.

Some NGTP vendors offer multiple centralized management appliance models to choose from, depending on the quantity of MPS appliances managed and the volume of cyber attack activity.

Malware intelligence sharing

Part of the beauty of an NGTP solution is organizations (automatically) are protected by their locally generated threat intelligence and can choose to share this with other organizations. Through a cloud threat intelligence network (owned and operated by the NGTP vendor), once one organization has detected a brand new threat ("ground zero" for the attack), all other organizations are protected within minutes. Many call this *collective immunity*.

While organizations can defuse APT attacks with a standalone MPS, collective immunity makes their MPS security analysis more efficient by focusing MPS resources on analyzing truly unknown cyber attacks. Plus, by anonymously sharing newly created threat intelligence via the cloud, organizations commonly receive a vendor discount on their annual cloud threat intelligence network subscriptions for their willingness to share threat intelligence.

It's important to understand that at no time will any of your files, or even content contained within those files, ever be sent to the cloud threat intelligence network. Threat protection profiles, for example, only include anonymized data, such as a checksum of the file.

Custom rule support

Leading NGTP systems enable more-advanced users to import custom malware-detection rules created using the YARA rules language. (YARA is a tool designed to help malware researchers identify and classify malware samples.) When an imported YARA rule is triggered by the MPS, the virtual execution engine immediately analyzes associated objects for potential cyber attacks. This is helpful for organizations that are frequently targeted by a specific class of cyber attack.

For more information on YARA, connect to http://code.google.com/p/yara-project/.

AV-suite integration

Preferred NGTP solutions can integrate with popular antivirus (AV) suites (see Figure 5-3), such as McAfee, Symantec, Sophos, and more. By integrating the NGTP solution with an AV suite, each malicious object can be further analyzed to determine if the AV platform was able to detect the malware stopped by the MPS. This enables organizations to more efficiently prioritize incident response follow-ups.

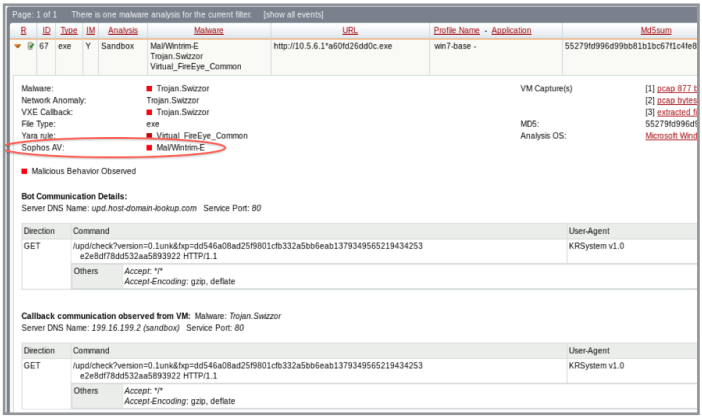


Figure 5-3: Sample anti-virus platform integration from FireEye.

Role-based access controls

Most NGTP systems provide multiple user roles to ensure that administrative privileges are only granted to IT personnel who require them to do their jobs. Common NGTP user roles include:

- ☑ **System administrator** – full administrative control over the entire NGTP deployment.
- ☑ **Regional administrator** – administrative control over one or more MPS appliances.
- ☑ **Security analyst** – access to dashboards and reports only; no ability to modify or delete event data or modify system settings.

Dashboard

The NGTP dashboard (see Figure 5-4) is the primary interface used by security analysts to monitor the security state of the network and the workload of the organization's MPS appliances. Dashboards are customarily accessed via Web browsers and are easy to interpret. Better dashboards offer the ability to "drill down" within event data to reveal details of cyber attacks to help the security analyst determine next steps.

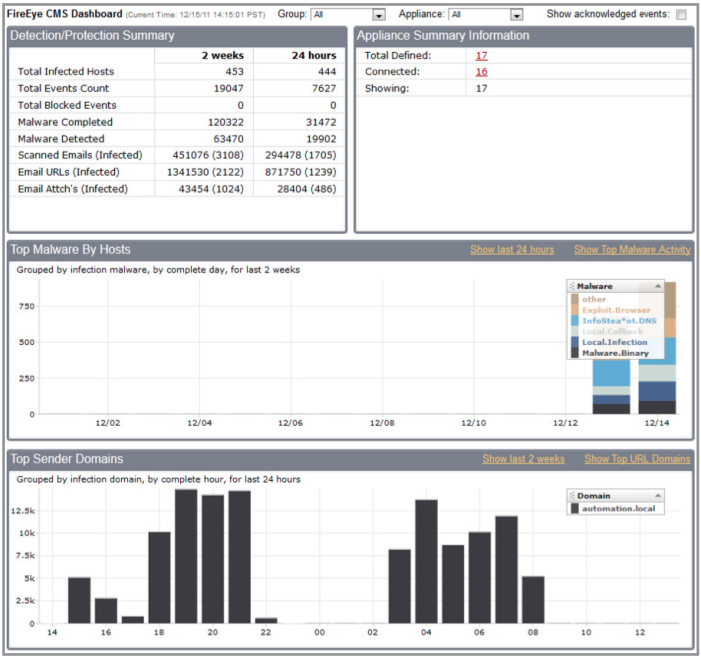


Figure 5-4: Sample dashboard from FireEye.

Reports

Today's NGTP solutions provide powerful and convenient ways to search for and report on specific types of cyber attacks by name or type. Organizations can view event summaries such as top infected hosts and top malware and callback events, including geolocation details. Some NGTP solutions even provide the capability to display security event data in Google Earth!

Reports can be generated by NGTP users on the fly or they can be automatically created at specified time intervals (daily, weekly, monthly) by the centralized management console.

Trending reports, in particular, can help demonstrate progress in reducing the number of compromised systems.

Alerts

Alerts help keep security analysts abreast of potential cyber breaches. Alert notifications can typically be sent via SMTP, SNMP, syslog, and HTTP POST. Alerts are also displayed within the Web-based interface of the centralized management system appliance (see Figure 5-5).

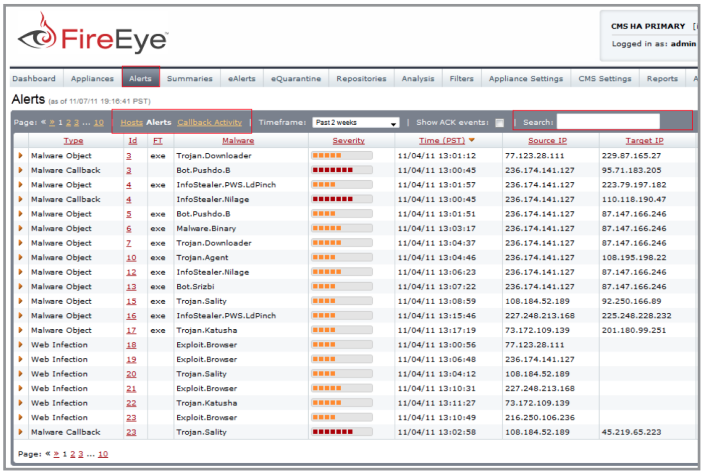


Figure 5-5: Sample alerts summary from FireEye.

Upon detection of a new cyber attack with an associated callback, the MPS registers one or more high-severity alerts. While the MPS appliance (if configured for inline operation) is capable of severing callback communications to defuse the attack, it's important to follow-up and remediate the host compromised by that attack.

Integrating NGTP into Your Existing IT Infrastructure

No information security system should ever operate in a vacuum. Security products should work in concert with one another — and the network infrastructure supporting them — to provide IT with greater context about the environment it's protecting and, ultimately, reduce the risk of successful cyber attacks.

The purpose of this section is to describe how NGTP systems integrate with three commonly requested IT platforms, starting with SIEM.

SIEM

A SIEM (security information and event management) is one of the most commonly requested platforms for NGTP integration, especially in out-of-band deployments. And it's no surprise since the entire purpose of a SIEM is to aggregate security events from across the organization and correlate them (using dozens of pre-built and custom correlation rules) to uncover hidden cyber attacks.

Security events can be exported in real-time streams to SIEM platforms in syslog, Common Event Format (CEF), and vendor-proprietary formats that offer more attack details for deeper analysis.

Organizations also frequently integrate NGTP systems with log management products. Like a SIEM, a log manager aggregates security events (through syslog data), but unlike a SIEM, it (typically) can't correlate data. Log managers are often chosen to satisfy a regulatory compliance mandate (such as PCI DSS), but also provide a convenient means for aggregating log data.

Sample vendors include: HP ArcSight, IBM Q1 Labs, LogRhythm, McAfee, RSA, and Splunk.

Security intelligence and analytics

Security intelligence and analytics (SIA), also known as network forensics, captures every single packet that traverses the network for a variety of purposes, including:

- ☑ Security incident response (forensics)
- ☑ Cyber attack detection
- ☑ Data loss monitoring and analysis

Once an NGTP system has classified a new form of malware, the analyst can employ "big data" analysis techniques and query the SIA database to determine the context in which the

host was compromised and to identify other hosts potentially compromised by the same attack.

Some SIA vendors provide a universal connector interface that plugs right into the Web browser so that the NGTP user can click on an IP addresses straight from the central management console to initiate a query into the SIA database, speeding the process of incident response.

Sample vendors include: NetWitness (RSA), Niksun, and Solera Networks.

Incident management

Incident management (or ticketing) platforms have been around for years. They are commonly used by internal IT and help desk staff to track and manage IT incidents. An incident could be as simple as resolving a help desk call or as complex as terminating an APT.

Organizations often wish to feed NGTP alerts into their existing incident management platform. This is accomplished by forwarding specially formatted SMTP alerts from the NGTP central management system to the incident management system or by parsing the XML format alerts to conform to existing incident alert templates.

Sample vendors include: BMC Remedy, Numara Software, and RSA Archer.

National laboratory experiments with next-generation threat protection

Some say, "Ignorance is bliss." Well, it's not for the CSO of one U.S.-based national laboratory tasked with advancing scientific discoveries in the disciplines of energy, the environment, and national security. On a daily basis, this laboratory handles a huge portfolio of national secrets and sensitive data, making it a prized target for highly motivated and sophisticated cybercriminals.

To guard against potential data breaches, the laboratory deployed a comprehensive range of enterprise-class security devices, including firewall, IPS, and AV solutions. But one day, the laboratory's veteran CSO read an article in an information security journal about an organization about the same size as his that was being devastated by an APT. He also learned how APTs operate and why traditional security defenses are no match.

After consulting his team of experienced security practitioners, he learned about a relatively new category of network security defense — next-generation threat protection. He also learned that one vendor, in particular, had more mindshare in detecting advanced threats than all other NGTP vendors combined — FireEye (www.fireeye.com).

Later that day, a member of the team contacted FireEye to schedule a meeting, which soon after resulted in an onsite evaluation. Taking just one day to implement a pilot to monitor network traffic, the FireEye Web MPS appliance showed immediate positive results. Within hours, alerts were generated by malicious code that went completely undetected by the lab's existing security defenses.

Weeks later, the FireEye MPS appliance went into full inline production. The appliance's fast-path blocking capability stops known inbound attacks and malware callbacks, while its powerful virtual execution engine accurately detects unknown cyber attacks.

The laboratory's CSO can now rest much easier knowing that he won't be reading his laboratory's name in his favorite information security journal anytime soon.

Chapter 6

Selecting the Right NGTP Solution

In this chapter

- Learn what to avoid when evaluating NGTP solutions
- Compile your list of NGTP buying criteria

There are literally dozens of cybersecurity vendors out there, right now, touting their abilities to detect and block advanced cyber attacks, including zero-day attacks, polymorphic threats, and APTs. Although each of these vendors may be able to detect and block *known* attacks, very few are capable of blocking *unknown* attacks — especially when they're targeting unknown (at least to the general public) operating system or application vulnerabilities.

It's important to know what to look for — and, perhaps more importantly, what to avoid — when shopping for an NGTP system. Let's start with the latter.

What to Avoid

The following is a list of things to avoid when evaluating NGTP solutions:

Avoid detection-only solutions. The best NGTP offerings support both inline and out-of-band modes of operation. Many organizations start with out-of-band to gain an initial level of comfort and then graduate to an inline NGTP configuration to block known threats, malware callbacks, and recurrences of newly discovered malware. Also, when evaluating MPS appliances with copper interfaces, be sure they support fail-open connectivity for inline deployments.

Avoid sandbox-based offerings. So-called NGTP solutions that incorporate legacy sandbox technology are easily outsmarted by sophisticated threat actors. These guys design malware to detect the presence of traditional sandbox technology, suppressing the payload of a malware-infected file to avoid detection.

Avoid cloud-based malware analysis. Every NGTP solution should leverage the power and scalability of the cloud. But preferred NGTP solutions leverage the cloud for intelligence sharing — not for malware analysis, as is the case with multi-function network security solutions offering rudimentary NGTP capabilities. By incorporating the virtual execution engine into the MPS appliance, malware analysis is done right at your site, drastically improving malware-identification coverage, scalability, and ensuring confidential files never leave the network.

Avoid all-in-one MPS appliances. For optimal detection of today's new breed of cyber attacks, separate purpose-built MPS appliances should be acquired for email, Web, and file share protection — but be sure they are integrated to share intelligence.

Now that you know exactly what not to look for, it's time to start compiling your shopping list of NGTP characteristics that are important for any security-minded organization. This next section will help.

Important Buying Criteria

Regardless of an organization's size or industry, the following NGTP buying criteria should be front-of-mind to every enterprise and government agency.

Some of the aforementioned buying criteria have already been discussed. Consider their corresponding descriptions to be concise recaps. But for more-detailed explanations, flip back to Chapters 3 and 4 for a quick refresher.

Integrated NGTP platform for Web, email, and file inspection

Although I've raised this topic more than once, it's worth mentioning it again. To thwart an APT attack, it is critical to have integrated protections across the common entry points for malware: Web, email, and files. The best NGTP platforms incorporate purpose-built MPS appliances (with uniquely different heuristics and algorithms) for detecting malware embedded within email messages, Web traffic, and files at rest, while correlating the findings to stop the APT attack across the enterprise.

Although you may save a few bucks in the short run by purchasing an MPS appliance from a vendor that claims coverage for two or three of these mediums, in the long run, it's simply not worth the risk to invest in a partial solution.

Monitors ingress and egress traffic

Typical NGTP systems monitor *ingress* (inbound) traffic from websites and email messages to identify suspicious binaries that may contain advanced malware. Most NGTP systems do not monitor *egress* (outbound) traffic, and then, surprisingly, some NGTP *only* monitor egress traffic. By monitoring both ingress and egress traffic, the MPS appliance can detect both inbound malware and corresponding outbound callback attempts.

Monitoring both ingress and egress traffic is an important capability found only in leading NGTP solutions. It provides an additional layer of defense — especially for potentially infected mobile devices hand-carried through the office front door.

Inspects broad range of file types

You might be shocked to learn that some rudimentary NGTP solutions are only capable of uncovering malware in unencrypted exe and dll files. The fact of the matter is that malware can be embedded in dozens of object types. This includes something as simple as an XOR-encoded binary, in the case of Operation Aurora, or a Microsoft Excel file, which triggered

a zero-day Flash exploit in the attack against RSA Security (see "RSA Security steps forward to describe its APT attack" sidebar in Chapter 3).

Hybrid document exploits highlight the need for broad network security coverage. In the case of the RSA Security breach, the Excel spreadsheet did not attack Microsoft Excel, but rather triggered an exploit against a separate application altogether.

A good NGTP solution applies sophisticated heuristics and malware-detection algorithms to uncover advanced malware in Web pages as well as across dozens of file types, including com, doc, docx, gif, jpg, mov, mp3, mp4, pdf, png, ppt, pptx, swf, tiff, xls, xlsx, zip, and many more.

Solution for manual malware analysis

The virtual execution engine contained within the MPS appliance is frequently analyzing (even remotely) suspicious files for advanced malware. In doing so, it provides security analysts with forensic details about the exploit, including the vulnerability exploited to create a buffer overflow condition, attempts to escalate privileges within Windows, and the call-back coordinates used to exfiltrated data.

Experienced incident responders sometimes prefer to analyze malware by hand to gain a full 360-degree view of the attack, from the initial exploit and malware execution to follow-on binary download attempts. To satisfy this hunger for rich forensic data, leading NGTP providers offer a stand-alone malware analysis system (MAS), which is typically packaged on a convenient rackmount appliance.

The MAS should incorporate instrumented, automatically configured virtual machines equipped with various versions of Microsoft Windows and a number of software applications, such as Microsoft Office and Adobe Reader. This environment lets analysts drill into suspicious (or known-infected) binaries to gain a deep understanding of the intent and targets of the cyber attackers, without the overhead of creating and maintaining a range of custom test environments.

No false positives or false negatives

False positives and false negatives resulting from poor NGTP detection can prove costly for any organization. A false positive (good file classified as bad) could mean blocking a business-critical file from reaching its destination, resulting in lost time and lost revenue. A false negative (bad file classified as good) is even worse as a file infected with malware is allowed to proceed to its final destination without further cause for analysis. At this point, I don't think I need to explain what this could mean.

To say that even the best NGTP system on the market will never render a false positive or a false negative is, perhaps, a bit of a stretch. But such occurrences should be very few and far between.

To assess the detection quality of an NGTP solution on your shortlist, put it through its paces by performing an onsite evaluation. If you're evaluating two competing solutions, test them both at the same time using the same production traffic (in passive, out-of-band mode) and compare their results.

Support for custom rules

As discussed in Chapter 5, NGTP administrators sometimes wish to import custom byte-level rules created using the YARA rules language to trigger analysis of all matched objects for threats specific to an organization. (This is akin to creating custom signatures for a network IPS.)

When evaluating competing NGTP offerings, consider the extensibility of the solution and its ability to support custom malware detection rules.

Intuitive user interface

It makes no difference how powerful or feature-rich a security application is. If it's too difficult to use, it's unlikely to be adopted by an IT organization — at least on a large scale. NGTP solutions are no exception.

Unlike security solutions that require tuning or creation of policy rule sets, such as a traditional firewall or IPS appliance, an NGTP system is a far more automated solution.

However, it's important that the dashboard be simple and easy to use, and constructing reports and alerts should not require a PhD in astrophysics.

Responsive customer support

Selecting an NGTP vendor is just as important as selecting an NGTP product — if not more so. High-quality technical support is frequently reported by enterprises and government agencies as a top decision criterion for selecting any IT system.

Be sure to assess the quality of a vendor's customer support service prior to making your purchasing decision. To do so, reach out to the vendor's technical support department at least twice during the evaluation phase, rather than the SE assigned to your account. Even if you don't have any problems with your evaluation, make something up. Perhaps ask a general question about the product's functionality. When doing so, gauge the experience and responsiveness of the tech support representative and note how long it took you to reach a human being. If it took 30 minutes to speak with a representative, and that person was unable to answer a simple configuration question, then it may be time to move onto a competing offering.

Glossary

advanced persistent threat (APT): A sophisticated cyber attack that employs advanced stealth techniques to remain undetected for extended periods of time.

advanced targeted attack (ATA): Another name for advanced persistent threat.

baiting: A social-engineering attack in which physical media (such as a USB flash drive) containing malware is deliberately left in proximity to a targeted organization.

blended threat: A cyber attack incorporating a combination of attacks against different vulnerabilities.

bot: An infected computer (or endpoint) centrally controlled by a command and control (CnC) server.

buffer overflow attack: An attack accomplished by placing more data into the buffer than it is configured to hold which ends up enabling the attacker to run custom code (oftentimes with the escalated privileges granted to the vulnerable application or network service).

BYOD (bring your own device): An organizational policy of employees bringing personally owned devices to their place of work to access the organization's data.

central management system (CMS): A rackmount appliance responsible for monitoring and managing MPS appliances within an NGTP environment.

cloud threat intelligence network: An Internet-based service managed by an NGTP vendor to distribute (and receive) cyber attack intelligence to (and from) its customers' MPS appliances.

CnC (command-and-control) server: A server operated by a cybercriminal to provide instructions to bots.

cybercriminal: A hacker illegally stealing data from another computer for personal financial gain.

cyberterrorism: The use of Internet-based attacks in terrorist activities, including acts of deliberate, large-scale disruption of computer networks.

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

data leakage prevention (DLP): A system designed to detect potential data loss based on patterns (such as social security numbers) in a timely manner.

defense-in-depth strategy: 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) attack: A cyber attack intended to disrupt or disable a targeted host by flooding it with benign communication requests from a single host.

egress traffic: Computer network traffic flowing from inside the network to hosts outside the network.

fail open: The ability of copper interfaces on a network appliance to maintain connectivity to prevent network disruption upon appliance power loss or disruption.

false negative: Misclassifying a file containing malware as benign.

false positive: Misclassifying a benign file as containing malware.

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

inline mode: Placement of a network appliance directly in the line of network traffic enabling it to block cyber attacks.

ingress traffic: Computer network traffic flowing from outside the network to hosts within the network.

intrusion detection system (IDS): An out-of-band signature-based security device that monitors network traffic and alerts upon detecting known cyber attacks.

intrusion protection system (IPS): An inline (active) signature-based security device that monitors network traffic and blocks known cyber attacks upon detection.

keylogger: An application that records keystrokes on a computer usually unbeknownst to the user.

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*.

malware analysis system (MAS): Appliance equipped with virtual execution engine that enables users to manually inspect objects suspected of containing malware.

malware protection system (MPS): A rackmount appliance responsible for detecting suspicious network objects and forwarding them to the virtual execution engine (which it also hosts) for signature-less analysis.

multi-staged: A cyber attack incorporating multiple types of malware designed to be launched at different phases of an advanced cyber attack.

multi-vectored: A cyber attack designed to target multiple target hosts within the same organization using multiple attack techniques.

next-generation threat protection (NGTP): Software installed on purpose-built, rackmount appliances that is designed to detect and block today's new breed of cyber attacks.

next-generation threats: Today's new breed of cyber attacks not easily detected by signature-based security defenses. Examples include polymorphic malware, zero-day threats, and APTs.

out-of-band mode: The mode of operation of a network appliance that enables it to analyze traffic copied from a network TAP or switch SPAN port.

phishing: The act of sending an email to a user falsely claiming to be a legitimate entity in an attempt to scam the user into surrendering private information, such as credit card and Social Security numbers.

polymorphic threat: Malware that changes its signature (binary pattern) every time it replicates in order to evade detection by a security device or application.

RAT (remote administration tool): Software that provides the hacker with a backdoor into the infected system to snoop or take control of the host.

sandbox: A software application designed to analyze suspicious binaries in the safety of a virtual machine, although often evaded by sophisticated cyberattackers.

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

SQL injection attack: A form of attack on a database-driven Web application in which the attacker executes unauthorized SQL commands to exploit insecure code.

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

state-sponsored threat actor: A cybercriminal employed by a nation-state to conduct cyber attacks against enemies of the state for politically motivated purposes.

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

virtual execution engine: A component on an MPS appliance that is responsible for signature-less analysis of suspicious objects in the safety of a virtual machine.

whaling: A cyber attack directed specifically at senior executives and other high-profile targets within businesses.

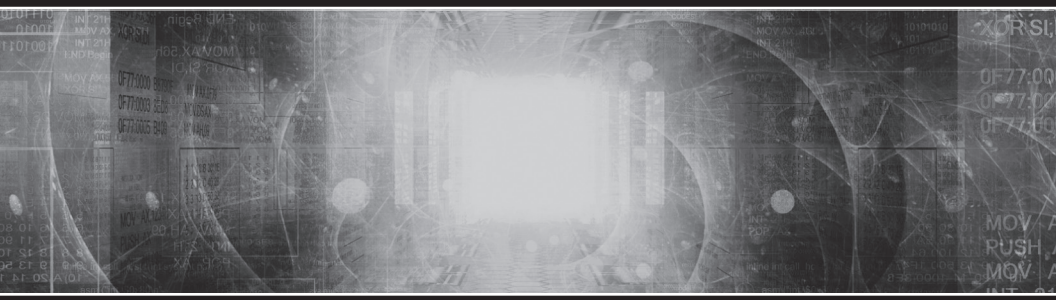
worm: A form of malware that exploits network vulnerabilities to propagate itself to other computers.

zero-day threat: A cyber attack against an unknown (or unreported) operating system or application vulnerability.

THE NEW BREED OF CYBER ATTACKS

HAS PENETRATED 95% OF ALL NETWORKS.

THINK YOU'RE IN THE 5%?



You may think your existing security defenses prevent today's new breed of cyber attacks from entering your network and stealing your data. They don't. Today's cyber attacks easily evade traditional and next-generation firewalls, IPS, AV, and gateways.

FireEye is your best defense. Put a stop to today's new breed of cyber attacks with next-generation threat protection. Visit us today at **www.FireEye.com** and let us help you close the hole in your network.

FireEye™

Fighting a losing battle against today's new breed of cyber attacks? Learn to defeat your cyber enemies through next-generation threat protection (NGTP).

Despite spending over $20 billion annually on traditional security defenses, organizations find themselves battling a new breed of cyber attacks, with zero-day exploits, polymorphic malware, and advanced persistent threats (APTs) leading the charge. If there's a chance of defeating today's well-funded, highly motivated adversaries, we've got to think differently. If you're charged with securing your organization's network, this is one book you simply can't afford to miss.

- Defining next-generation threats — contrast traditional cyber attacks with a new breed of next-generation threats
- Understanding the enemy — explore three kinds of cyber enemies and how they defeat traditional security defenses
- Anatomy of advanced cyber attacks — learn the five stages of the advanced attack life cycle and discover telltale signs for detecting APTs
- Introducing NGTP — review key components of NGTP solutions and compare NGTP to traditional security defenses
- NGTP explored — learn how to mitigate next-generation threats in email messages, Web communications, and files at rest
- Selecting the right NGTP solution — know exactly what to look for, and what to avoid, when evaluating NGTP solutions

About the Author

Steve Piper is an information security veteran with over 20 years of high-tech experience. A freelance writer and consultant, Steve has authored numerous books on information security, network infrastructure, and Big Data. He holds a CISSP security certification from ISC^2 and bachelor of science and MBA degrees from George Mason University. Learn more at www.stevepiper.com.

CYBEREDGE PRESS

Not for resale

ISBN 978-0-9888233-0-3

9 780988 823303