Palo Alto Networks® 2nd Edition

Next-Generation Firewalls

FOR DUMMIES®

A Wiley Brand

Learn to:

- Differentiate between "good" and "bad" applications

- Identify evasive techniques used by applications and modern threats

- Implement effective application and network controls

Brought to you by

paloalto NETWORKS®

Lawrence C. Miller, CISSP

Palo Alto Networks

Palo Alto Networks is leading a new era in cybersecurity by securing thousands of enterprise, government, and service provider networks from cyber threats and protecting our digital way of life. The next-generation platform uses an innovative traffic classification engine that identifies network traffic by application, user, and content.

The Palo Alto Networks next-generation security platform is built on four main principles:

1. **Natively integrated** technologies that support open communication, orchestration, and visibility;

2. **Automation** of protection creation and reprogramming of the security posture across network, endpoint and cloud environments;

3. **Extensibility** that allows for protection of customers as they expand and market requirements change; and

4. **Threat intelligence sharing** to minimize the spread of attacks by providing protection based on comprehensive global threat data.

The next generation security platform offers superior protection against the sophistication of modern attacks, can reduce the total cost of ownership for organizations by simplifying their security infrastructure, and eliminates the need for multiple, stand-alone security appliances and software products.

Find out more at **www.paloaltonetworks.com**

Next-Generation Firewalls

FOR DUMMIES®
A Wiley Brand

Palo Alto Networks 2nd Edition

by Lawrence C. Miller, CISSP

FOR DUMMIES®
A Wiley Brand

Next-Generation Firewalls For Dummies®, Palo Alto Networks 2nd Edition

Publisher's Acknowledgments

Table of Contents

Introduction

*W*ith new threats growing in number and sophistication, organizations are finding that traditional security products and approaches are less and less capable of adequately protecting their networks against today's advanced attacks.

The rapid evolution of applications, IT infrastructure, and the threat landscape has resulted in a loss of visibility and control for organizations attempting to safely enable and protect their business, customers, and users.

Despite their best efforts to restore visibility and control, and regain the advantage in protecting their networks and information, most organizations remain stymied. In a security market that is largely lacking technological innovation and, thus, full of repackaged and rebranded traditional security products, many organizations turn to an increasing number of single-purpose security devices that still fail to fully address today's security challenges. Even when these security devices are consolidated in an all-in-one appliance, they're often poorly coordinated, falling far short of providing comprehensive security and threat prevention.

The result is inefficiency and complexity — characteristics that are never desirable in any solution. More important, though, disparate and poorly coordinated security devices are ineffective and result in weak security. In a world where applications, infrastructure, and threats are sophisticated and dynamic, having a grab bag of devices performing various security functions that don't integrate with each other results in gaping and dangerous security "holes."

Instead, an entirely new and innovative approach to network security is needed — an approach that works with the latest applications and infrastructure trends, along with the ability to recognize and stop today's most advanced threats. The cornerstone of this approach is the next-generation firewall (NGFW)!

About This Book

This book examines the evolution of network security (Chapter 1), the rapid growth of applications and their associated risks and threats (Chapter 2), the shortcomings of traditional firewalls and products based on them (Chapter 3), the advanced capabilities found in NGFWs (Chapter 4), how to deploy NGFWs (Chapter 5), and how to select the best NGFW for your organization's security challenges (Chapter 6).

Icons Used in This Book

Throughout this book, I occasionally use icons to call attention to important information that is particularly worth noting:

This icon points out information or a concept that may well be worth committing to your nonvolatile memory, your gray matter, or your noggin — along with anniversaries and birthdays!

You won't find a map of the human genome or the secret to cold fusion here, but if you seek to attain the seventh level of NERD-vana, perk up! This icon explains the jargon beneath the jargon!

Thank you for reading, hope you enjoy the book, please take care of your writers. Seriously, this icon points out helpful suggestions and useful nuggets of information.

This is the stuff your mother warned you about. . . . Well, okay, probably not. But these helpful alerts do offer practical advice to help you avoid potentially costly mistakes.

Beyond the Book

There's only so much I can cover in 72 short pages, so if you find yourself at the end of this book, thinking, "Gosh, this was an amazing book, where can I learn more?," just go to `www.paloaltonetworks.com`.

Chapter 1

Understanding the Evolution of Network Security

*J*ust as antivirus software has been a cornerstone of PC security since the early days of the Internet, firewalls have been the cornerstone of network security.

Today's application and threat landscape renders traditional port-based firewalls largely ineffective at protecting corporate networks and information. Applications are the conduit through which everything flows — a vector for our business and personal lives — along with their associated benefits and risks. Such risks include new and emerging threats, data compromise, and noncompliance.

This chapter explains how traditional firewall technology works, why products based on this legacy approach can't meet today's application and threat challenges, and how data compromise and compliance issues are defining network security and the need for better firewalls.

Why Legacy Firewalls Are No Longer Effective

A firewall, at its most basic level, controls traffic flow between network segments. A simple example might be traffic control between a trusted network (such as a corporate LAN) and an untrusted or public network (such as the Internet). Many currently deployed firewalls are still port-based firewalls, or some variation (such as stateful inspection) of this basic type of firewall. These firewalls are popular because they are relatively simple to operate and maintain, are generally inexpensive, have good throughput, and have been the prevalent design for more than two decades.

In the rapid pace of the Internet Age, two decades means the basic technology behind port-based firewalls is medieval. In fact, network security is often likened to the Dark Ages — a network perimeter is analogous to the walls of a castle, with a firewall controlling access — like a drawbridge. And like a drawbridge that is either up or down, a port-based firewall is often limited to just two options for controlling network traffic: allow or block.

TECHNICAL STUFF   Port-based firewalls (and their variants) use source/destination IP addresses and TCP/UDP port information to determine whether a packet should be allowed to pass between networks or network segments. The firewall inspects the first few bytes of the TCP or UDP header in an IP packet to determine the application protocol — for example, SMTP (port 25) and HTTP (port 80).

Most firewalls are configured to allow all traffic originating from the trusted network to pass through to the untrusted network, unless it is explicitly blocked by a rule. For example, the Simple Network Management Protocol (SNMP) might be explicitly blocked to prevent certain network information from being inadvertently transmitted to the Internet. This would be accomplished by blocking UDP ports 161 and 162, regardless of the source or destination IP address.

Static port control is relatively easy. Stateful inspection firewalls address *dynamic* applications that use more than one well-defined port (such as FTP ports 20 and 21). When a computer or server on the trusted network originates a session

with a computer or server on the untrusted network, a connection is established. On stateful packet inspection firewalls, a dynamic rule is temporarily created to allow responses or replies from the computer or server on the untrusted network. Otherwise, return traffic needs to be explicitly permitted, or access rules need to be manually created on the firewall (which usually isn't practical).

All of this works well as long as everyone plays by the rules. Unfortunately, the rules are more like guidelines and not everyone using the Internet is nice!

The Internet often accounts for the majority of traffic traversing an organization's networks. And it's not just web surfing. The Internet has spawned a new generation of applications being accessed by network users for both personal and business use. Many of these applications help improve user and business productivity, while other applications consume large amounts of bandwidth, pose needless security risks, and increase business risk — for example, data leaks and compliance — both of which are addressed in the following sections. And many of these applications incorporate "accessibility" techniques, such as using nonstandard ports, port hopping, and tunneling, to evade traditional port-based firewalls.

IT organizations have tried to compensate for deficiencies in traditional port-based firewalls by surrounding them with proxies, intrusion prevention systems, URL filtering, and other costly and complex devices. But this uncoordinated approach has been largely ineffective in today's application and threat landscape.

Data Compromise Is a Problem

Large scale, public exposures of sensitive or private data are far too common. Numerous examples of accidental and deliberate data loss continue to regularly make nightmare headlines, exposing the compromise of millions of credit card numbers by major retailers, Social Security numbers leaked by government agencies, protected health information (PHI) disclosed by healthcare organizations, and other sensitive information lost by employers in practically every industry. Unfortunately, such incidents are not isolated. In many of these cases, sensitive data was compromised starting with an application that was expressly prohibited by policy but not

adequately enforced with technology, or via an application that was allowed, but also carried a threat that gained a foothold by automatically infecting a computer or fooling a user. Other risks to data include data sabotage (or destruction) and the use of ransomware that encrypts important data, rendering it unusable, unless a hefty ransom is paid (CryptoLocker is an example).

With respect to data loss, data loss prevention (DLP) is sometimes held up as solution. Unfortunately, given the scope, size, and distributed nature of most organizations' datasets, just discovering where sensitive data is and who owns it is an insurmountable challenge. Adding to this challenge, questions regarding access control, reporting, data classification, data at-rest versus data in-transit, desktop and server agents, and encryption abound. As a result, many data loss prevention initiatives within organizations progress slowly and eventually falter. And DLP does nothing to prevent data sabotage (because it was never designed to address this problem).

Controlling the applications that are used to compromise data, either directly or as part of a larger "chain of events" is foundational to securing organizations. Exerting that control at trust boundaries (the network perimeter) is ideal — whether the demarcation point is between inside and outside or internal users and internal resources in the data center. The firewall sits in the perfect location, seeing all traffic traversing different networks and network segments. Unfortunately, legacy port- and protocol-based firewalls can't do anything about any of this — being ignorant of applications, users, and content.

To effectively address data compromise with a firewall solution, organizations should

- ✔ Gain control over the applications on their network — thus limiting the avenues of data loss or compromise

- ✔ Scan the applications they want on their networks, for sensitive or private data, or to detect behaviors in a multi-stage attack designed to steal or sabotage data

- ✔ Understand which users are initiating application transactions and why

- ✔ Implement appropriate control policies and technology to prevent accidental or intentional data loss or compromise

If organizations could control applications and the flow of sensitive or private data in the network, many of the data loss incidents that regularly make the news could be prevented. Unfortunately, legacy security infrastructures, with traditional port-based firewalling as their basis, are ill equipped to provide this functionality.

Compliance Is Not Optional

With a rapidly and ever increasing number of laws and regulations worldwide mandating information security and data protection requirements, organizations everywhere are struggling to attain and maintain compliance. Examples of these regulations include HIPAA, FISMA, FINRA, and GLBA in the United States and the EU Data Protection Act (DPA) in Europe.

Ironically, perhaps the most far-reaching, most effective, and best-known compliance requirement today isn't even a government regulation. The Payment Card Industry Data Security Standard (PCI DSS) was created by the major payment card brands (American Express, MasterCard, Visa, and others) to protect companies, banks, and consumers from identity theft and fraudulent card use. And as economies rely more and more on payment card transactions, the risks of lost cardholder data will only increase, making any effort to protect the data critical — whether compliance driven or otherwise.

PCI DSS is applicable to any business that transmits, processes, or stores payment cards (such as credit cards or debit cards), regardless of the number or amount of transactions processed.

Companies that don't comply can be subject to stiff penalties, including fines of up to $25,000 per month for minor violations, fines of up to $500,000 for violations that result in actual lost or stolen financial data, and loss of card-processing authorization (making it almost impossible for a business to operate).

Although compliance requirements are almost entirely based on information-security best practices, it's important to remember that security and compliance aren't the same thing. Regardless of whether a business is PCI compliant, a data breach can be very costly. According to research conducted

by IBM and the Ponemon Institute, the estimated *per record* cost of a breach (including fines, cleanup, lost opportunities, and other costs) averages $154. Verizon's 2015 Data Breach Investigations Report predicts the expected loss by number of records as $255 per record (for 100 records). Other damages due to a data breach are still more difficult to quantify, such as the damage to a business or brand's reputation, and the true cost to the individual victims.

REMEMBER

Security and compliance are related, but they are not the same thing!

PCI DSS version 3.1 consists of 12 general requirements and more than 200 specific requirements. Of the 12 general requirements, the following specifically address firewall and firewall-related requirements:

- ✔ **Requirement 1:** Install and maintain a firewall configuration to protect cardholder data.
- ✔ **Requirement 5:** Protect all systems against malware and regularly update antivirus software or programs.
- ✔ **Requirement 6:** Develop and maintain secure systems and applications.
- ✔ **Requirement 7:** Restrict access to cardholder data by business need-to-know.
- ✔ **Requirement 10:** Track and monitor all access to network resources and cardholder data.
- ✔ **Appendix D:** To use network segmentation to reduce PCI DSS scope, an entity must isolate systems that store, process, or transmit cardholder data from the rest of the network.

The challenges posed by modern data compromise techniques call for precise application control, as well as visibility and control of the traffic flowing on your network. Unfortunately, traditional port-based firewalls just don't meet this standard.

And although preventing data compromise is always a good idea, compliance often makes it a mandate.

Chapter 2

Defining the Application and Threat Landscape

*N*etwork security used to be relatively simple — everything was more or less black and white — either clearly bad or clearly good. Business applications constituted good traffic that should be allowed, while pretty much everything else constituted bad traffic that should be blocked.

Problems with this approach today include the fact that applications have become

✔ Increasingly "gray" — classifying types of applications as good or bad is not a straightforward exercise.

✔ More difficult to accurately identify based on traditional port and protocol assignments.

✔ The predominant vector of today's cybercriminals and threat developers who use applications as unwitting carriers of malicious payloads.

This chapter explores the evolving application and threat landscape, the blurring distinction between user and business applications, and the strategic nature of many of these applications (and their associated risks) for businesses today.

Applications Are Not All Good or All Bad

Over the past decade, the application landscape has changed dramatically for organizations. Corporate productivity applications have been joined by a plethora of personal and consumer-oriented applications that are often available as Software as a Service (SaaS) or web-based applications. This convergence of corporate infrastructures and personal technologies is being driven by two popular and important trends — *consumerization* and *bring your own device* (BYOD).

The process of consumerization occurs as users increasingly find personal technology and applications that are more powerful or capable, more convenient, less expensive, quicker to install, and easier to use than corporate IT solutions. These user-centric "lifestyle" applications and technologies enable individuals to improve their personal efficiency, handle their nonwork affairs, and maintain online personas, among other things.

Catering to this demand, technology vendors and developers enjoy vast economies of scale and the pervasive benefits of viral marketing.

Closely related to consumerization is BYOD — an increasingly popular trend in which organizations permit their employees to use their own personal devices, primarily smartphones and tablets, for work-related purposes. More often than not, the same applications used for social interaction on personal devices are being used for work-related purposes. And as the boundary between work and their personal lives becomes less distinct, users are practically demanding that these same tools be available to them in their workplaces.

The rapid adoption of many popular SaaS and mobile applications is often driven by users, not by IT. The ease with which they can be accessed, combined with the fact that today's knowledge workers are accustomed to using them, points toward a continuation of the consumerization trend and a growing "shadow" IT culture in which individuals and departments use both sanctioned and unsanctioned applications.

The applications driven by consumerization combine with those supported by IT, resulting in a wide variety of application types in organizations today. Examples of these applications include

- ✔ Collaboration and cloud storage tools such as Box, Dropbox, Google Docs, iCloud, Microsoft Office 365/OneDrive
- ✔ Web-based email such as Gmail, Outlook.com, and Yahoo! Mail
- ✔ Content management tools such as SharePoint
- ✔ Customer relationship management (CRM) portals such as Salesforce and SugarCRM
- ✔ Social networks such as Facebook and LinkedIn
- ✔ Web publishing tools such as YouTube
- ✔ Unified messaging tools such as Skype and Vidyo
- ✔ Posting tools such as Twitter
- ✔ Anonymizers and proxies such as Tor and UltraSurf
- ✔ Remote access tools such as Ammyy, LogMeIn, Remote Desktop Protocol (RDP), and TeamViewer

WARNING!

The use of anonymizers and proxies on any network should be considered risky and suspect.

WARNING!

Remote access tools can be both good and bad. They are valuable productivity tools for IT administrators and support technicians, but also prone to exploit by attackers in order to control systems.

To appreciate how rapidly these applications, both sanctioned and unsanctioned, have proliferated the corporate network, consider that the Palo Alto Networks' Spring 2015 *Application Usage and Threat Report* found that SaaS-based application usage increased 46 percent on customer networks between 2012 and 2015. Cloud-based storage and web-based email accounted for the overwhelming majority of these applications — 40.7 percent and 38 percent, respectively. With more than 40 percent of unknown malware threats and exploits still being delivered by email, and the inherent risk of data leakage due to sensitive data potentially being uploaded to cloud-based storage, the risks to organizations cannot be ignored.

Unsure of how to leverage these trends in their business processes, many organizations either implicitly allow these SaaS-based and mobile applications simply by ignoring their use in the workplace, or explicitly prohibit their use, but are then unable to effectively enforce such policies with traditional firewalls and security technologies. Neither of these two approaches is ideal, and both incur inherent risks for the organization. In addition to lost productivity, adverse issues for the organization include

- ✔ Creating a "shadow IT" subculture of back-channel or underground workflow processes that are critical to the business's operations, but are known only to a few users and fully dependent on personal technologies and applications

- ✔ Introducing new risks to the entire networking and computing infrastructure, due to the presence of unknown, and, therefore, unaddressed and unpatched, vulnerabilities, as well as threats that target normal application and user behavior — whether a vulnerability exists in the application or not

- ✔ Being exposed to noncompliance penalties for organizations that are subject to regulatory requirements such as HIPAA, FINRA, and PCI DSS

- ✔ Having employees circumvent controls with external proxies, encrypted tunnels, and remote desktop applications, making it difficult, if not impossible, for security and risk managers to see the risks they're attempting to manage

The challenge is not only the growing diversity of the applications, but also the inability to clearly and consistently classify them as good or bad. Although many are clearly good (low risk, high reward), and others are clearly bad (high risk, low reward), most are somewhere in between. Moreover, the end of the spectrum that these applications fall on can vary from one scenario to the next and from user to user or from session to session.

For example, using a social networking application to share product documentation with a prospective customer would be "good" (medium risk, high reward), while using the same application to forward details of an upcoming release to a

"friends list" that includes employees of a competitor would be "not so good" (high risk, no reward).

Indeed, many organizations now use a variety of social networking applications to support a wide range of legitimate business functions, such as recruiting, research and development, marketing, and customer support — and many are even inclined to allow the use of lifestyle applications, to some extent, as a way to provide an "employee friendly" work environment and improve morale.

Many companies are also seeing significant benefits from the use of these applications and technologies, including an increased ability to share ideas, more rapid access to knowledge experts, and a reduction in travel, operations, and communications costs. Today's network security solutions, therefore, must be able not only to distinguish one type of application from the next, but also to account for other contextual variables surrounding its use (for example, file transfers and URL access) and to vary the resulting action that will be taken accordingly.

Enabling Facebook usage while protecting the business

Facebook is rapidly extending its influence from the personal world to the corporate world, as employees use these applications to get their jobs done. At the same time, many organizations are looking at the more than 1.5 billion Facebook users as an opportunity to conduct research, execute targeted marketing, gather product feedback, and increase awareness. The end result is that Facebook (and other social networks, such as LinkedIn) can help organizations improve their bottom line.

However, formally enabling the use of Facebook introduces several challenges to organizations. Many organizations are unaware of how heavily Facebook is being used, or for what purpose. In most cases, policies governing specific usage are nonexistent or unenforceable. Finally, users tend to be too trusting, operating in a "click now, think later" mentality, which introduces significant security risks.

Like any application that is brought into the enterprise by end-users,

(continued)

blindly allowing Facebook may result in propagation of threats, loss of data, and damage to the corporate reputation. Blindly blocking Facebook is also an inappropriate response because it may play an important role in the business and may force users to find alternative means of accessing it (such as proxies, circumvention tools, and others).

Organizations should follow a systematic process to develop, enable, and enforce appropriate Facebook usage policies while simultaneously protecting network resources:

1. **Find out who's using Facebook.** There are many cases in which there may already be a "corporate" Facebook presence established by marketing or sales, so it's critical that IT determine which social networking applications are in use, who is using them, and the associated business objectives.

2. **Develop a corporate Facebook policy.** When Facebook usage patterns are determined, organizations should engage in discussions regarding what should and should not be posted. Educating users on the security risks associated with Facebook is another important element to encouraging usage for business purposes.

3. **Use technology to monitor and enforce policy.** The outcome of each of these discussions should be documented with an explanation of how IT will apply security policies to safely and securely enable use of Facebook within enterprise environments.

Applications: "I'm Not a Number!"

Although "distinguishing one type of application from the next" sounds simple, it really isn't — for a number of reasons. In order to maximize their accessibility and use, many applications are designed from the outset to use standard ports, such as TCP ports 80 (HTTP) and 443 (HTTPS), that are commonly allowed through legacy port-based firewalls that see applications as little more than a number. You've heard the expression, "If all you have is a hammer, everything looks like

a nail," right? Well, to a port-based firewall, increasingly every application looks like HTTP or HTTPS!

Other applications use a variety of techniques in an attempt to run anywhere, at any time. Common techniques include the following:

- ✔ **Port hopping,** where ports/protocols are randomly shifted over the course of a session
- ✔ **Use of nonstandard ports,** such as running Yahoo! Messenger over TCP port 80 (HTTP) instead of the standard TCP port for Yahoo! Messenger (5050)
- ✔ **Tunneling within commonly used services,** such as when file-sharing and messaging applications like Telegram run over HTTP
- ✔ **Hiding within SSL encryption,** which masks the application traffic, for example, over TCP port 443 (HTTPS)

WARNING!

These techniques are also used by attackers for malicious purposes to evade detection by port-based firewalls.

At the same time, enterprise users are increasingly embracing SaaS and web-based applications and services such as Salesforce.com, WebEx, and Google Apps — which often initiate in a browser but then switch to more client-server-like behavior (rich client, proprietary transactions, and others).

The result of the shift to SaaS and web-based applications is that HTTP and HTTPS now account for approximately two-thirds of all organizational traffic. This is not a problem, per se, but it does exacerbate an inherent weakness of traditional security infrastructure. Specifically, the wide variety of higher-order applications riding on top of HTTP and HTTPS — whether or not they actually serve a legitimate business purpose — are practically indistinguishable for older port-based firewalls. The negative impact of organizations further losing control over their network communications is clear and underlines the fact that the application landscape has evolved dramatically.

Cloud-based SaaS applications:
I can't see clearly now

Organizations are adopting SaaS-based application services at a breakneck pace. These applications continue to redefine the network perimeter, providing critical functionality and increasing productivity, but at the same time introducing potential new security and data risks if not properly controlled.

In most organizations that use SaaS applications, users are provided access to a specific list of services the organization has deemed acceptable or suitable for business purposes. Given the high number of unique SaaS applications and the large percentage of usage observed in Palo Alto Networks' Spring 2015

Application Usage and Threat Report (see the figure), it's likely that many users aren't strictly complying with such usage policies, and are instead using nonsanctioned SaaS applications at work. This further increases the risk of data leakage to organizations, due to the lack of visibility from regular logs or notifications from unauthorized SaaS applications, as well as additional risk of intermeshing users' personal and work emails, which may create situations where a user's personal email account is attacked and the attacker is then able to steal data or compromise the user's work email account.
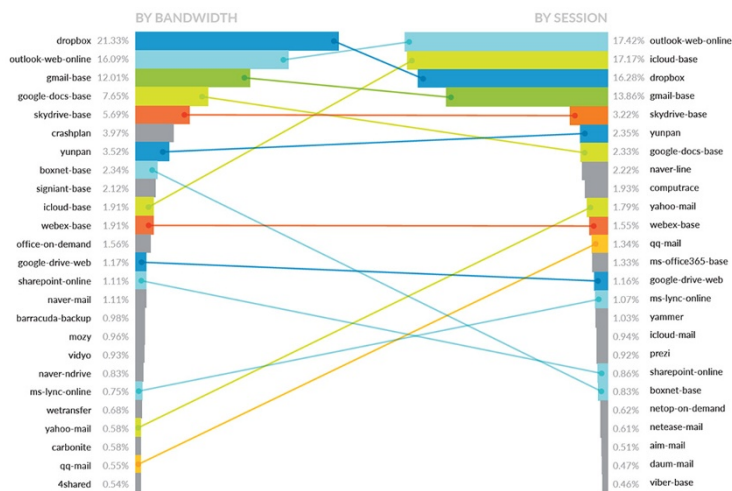
Source: Palo Alto Networks' Spring 2015 Application Usage and Threat Report

Threats Are Coming Along for the Ride

The increasing prevalence of application-layer attacks is yet another disturbing trend. Email and web browsers are still the main attack vectors today, with malicious content either attached or downloaded as an executable or macro-based file. The malicious use of remote access applications is another significant attack vector. Threats that directly target applications can pass right through the majority of enterprise defenses, which have historically been built to provide network-layer protection. Threat developers exploit the same methods (described in the previous section) to infiltrate networks that application developers utilize to promote ease of use and widespread adoption, such as tunneling within applications.

The evasion techniques built into these and many other modern applications are being leveraged to provide threats with "free passage" into enterprise networks. So, it's no surprise that more than 80 percent of all new malware and intrusion attempts are exploiting weaknesses in applications, as opposed to weaknesses in networking components and services. Together with the implicit trust that users place in their applications, all these factors combine to create a "perfect storm." The motivation for hackers has also shifted — from gaining notoriety to political activism, espionage, and making money. The name of the game today is information theft. Consequently, it's no longer in a hacker's best interests to devise threats that are "noisy" or that are relatively benign. To be successful, a thief must be fast or stealthy — or both.

For those hackers who favor speed over sophistication — speed of initial threat generation, speed of modification, and speed of propagation — the goal is to develop, launch, and quickly spread new threats immediately on the heels of the disclosure of a new vulnerability. The resulting zero-day and near-zero-day exploits then have an increased likelihood of success because reactive countermeasures, such as patching and those tools that rely on threat signatures (such as antivirus software and intrusion prevention), are unable to keep up — at least during the early phases of a new attack.

The (re)rise of macro malware

In 1995, the first macro-based malware, WM/Concept, was unleashed upon the public and began the initial wave of macro-based malware targeting Microsoft Word and Excel. Twenty years later, cybercriminals have rediscovered macro-based malware and are once again using it as another tool in their arsenal.

What is a macro?

Macros were originally developed for the Microsoft Office suite as a way to automate repetitive tasks or share tasks between different users. The system was designed so that a user could use a simple feature to record a repetitive task, which would then automatically be transcribed into Visual Basic for Applications (VBA). The macro automated the task, and the VBA code could then be shared with other users.

Unfortunately, macros were not designed with security in mind; functionality was the main goal, and macros allowed users to be more productive by speeding up repetitive tasks. Although the intentions of macros in Microsoft Office documents were altruistic, the unfortunate side effect was the creation of an easy-to-use and effective vehicle for malicious code.

The most famous and well-known macro-based malware was the Melissa virus in 1999. It was distributed within a Word document that would gather the first 50 entries from a user's address book and then mail a copy of the macro-infected Word document to each entry via Microsoft Outlook. When the recipients opened the document, the cycle would continue ad nauseam. Due to the overwhelming number of infected systems attempting to send out emails, the Melissa virus placed many major email servers into a denial-of-service state.

Where are we now?

In response to the Melissa virus and other macro malware, Microsoft put multiple mitigations in place to prevent the spread of macro-based malware. In Office 2003, only digitally signed macros could be run by default. In Office 2007, the letter *m* was appended to the usual Office file extensions (`.docxm`, `.xlsxm`, `.pptxm`) to signify that the file contained a macro. Finally, in Office 2013, macros were simply turned off by default, showing users a notification if a macro was embedded in the document they had opened. The actions taken by Microsoft significantly reduced macro-based malware infections and, in turn, reduced the popularity of macro-based malware usage by cybercriminals.

No good deed goes unpunished, however. In the last decade, a new generation of users who have never used macros or are even aware of what they are due to the dormancy of macros in general has emerged.

Users have a tendency to have a singular goal in mind, which is to accomplish the given task at hand. This causes them to ignore warnings or pop-up messages indicating potential danger because, to them, these buttons and dialog boxes are simple barriers to their productivity. The lack of awareness and a focus on getting to the user's desired content or task has led to a sudden resurgence in the usage of macro-based malware as users unwittingly enable macros in Office documents more and more often.

The two most observed malware families delivered via macro abuse are the Dridex and Dyre malware families. More than 10 percent of all malicious activity identified in Palo Alto Networks' Spring 2015 *Application Usage and Threat Report* involved these two malware families:

✔ **Dridex:** Dridex is a banking Trojan descended from the GameOver Zeus family of malware. Its functions are extremely similar to the well-known GameOver Zeus variants such as Cridex, targeting online banking credentials and containing configurations to mimic logins for financial institutions. Dridex differs from its malware relatives, however, in the fact that it utilizes macro-embedded Office documents to load itself onto potential victim hosts, where it then begins harvesting banking credentials. Well over 99 percent of Dridex sessions were delivered over various email protocols or web browsing.

✔ **Dyre/Upatre:** Upatre is the name of the malware downloader, generally delivered via a macro-based malware Office document, which then retrieves Dyre (Dyreza), a banking Trojan similar in function to GameOver Zeus and its variants. In addition, Upatre utilizes the Microsoft Outlook email client to send itself out to additional victims, effectively worming its way across the Internet. As with Dridex, more than 99 percent of these sessions were over various email protocols or web browsing.

This speed-based approach is facilitated in large part by the widespread availability of threat development websites, toolkits, and frameworks. Unfortunately, another by-product of these resources is the ability to easily and rapidly convert "known" threats into "unknown" threats — at least from the perspective of signature-based countermeasures. This transformation can be accomplished either by making a minor tweak to the code of a threat, or by adding entirely new propagation and exploit mechanisms, thereby creating what is commonly referred to as a *blended threat*.

Adversary dossiers

Although Sun Tzu never surfed the Internet, his maxim to "Know thy enemy" is no less germane. Critical to understanding your organization's risk posture is to understand the adversarial groups that may attempt to breach your network.

Focusing on the Tools, Tactics, and Procedures (TTPs) employed by adversaries provides security teams with actionable intelligence. Once you understand indicators of compromise (IOCs), such as command and control infrastructure, malware deployed, or methods of initial compromise, you can build preventive controls to stop them at every point in the attack life cycle.

Gaining context around the adversary will also allow security teams to prioritize their response efforts. For instance, organizations in government sectors will be more concerned by cyberespionage activity, versus a financial services organization interested in cybercrime targeting financial gain.

Three major threat actor groups are profiled here to help security teams understand if they could be targeted by these groups, and how to reduce their risk of being successfully breached.

Carbanak

✔ **Known aliases:** Anunuk

✔ **Origin:** Russia and Ukraine nexus

✔ **Motivation:** Financial gain and some evidence of cyberespionage

✔ **Summary:** Responsible for theft from various financial institutions. See later sidebar for details.

Sandworm

✔ **Known aliases:** Quedagh

✔ **Origin:** Russia nexus

✔ **Motivation:** Cyberespionage

✔ **Summary:** First disclosed publicly in October 2014. Known attacks began in December 2013 and continued into 2014. Attributed to using the BlackEnergy Trojan to execute espionage activity against industrial control system (ICS) environments.

✔ **Targeted regions/industries:** Europe. Telecommunications, energy, government, U.S.-based education institutions, ICS environments.

✔ **Tactics and tools deployed:** Utilizes two variants of the BlackEnergy Trojan. Has utilized a zero-day vulnerability (CVE-2014-4114) in the past for malware delivery. Also uses spear-phishing attacks. Developed custom plug-in modules for BlackEnergy Trojan, which allows for remote access, network traversal, keylogging, credential harvesting, network capturing, and screen capturing.

Shellcrew

✔ **Known aliases:** Shell Crew, Deep Panda, Axiom, Group 72

✔ **Origin:** China nexus

✔ **Motivation:** Cyberespionage

✔ **Summary:** A technically sophisticated and likely well-funded, state-sponsored APT group. Has used multiple zero days and can move very quickly once inside a network. Known to layer different malware throughout a network to maintain persistence, as well as harvest and use legitimate network and remote access credentials.

✔ **Targeted regions/industries:** Healthcare, government, manufacturing, defense, aerospace, industrial, pro-democracy NGOs, energy, telecommunications, academic institutions, journalists, think-tanks, media. Specific companies affected include Premera Blue Cross, Anthem, OPM, Bit9, RSA.

✔ **Tactics and tools deployed:** Utilizes doppelganger command and control domains to dupe users and obfuscate activity. Heavy usage of spear-phishing and watering hole attacks to deliver malware or harvest credentials. Has been known to use zero-day vulnerabilities. Has also been known to use legitimate network administration tools with legitimate compromised credentials. Malware families known to be used include Poison Ivy, Gh0st, Derusbi, Scanbox, Sakula, Zxshell, Zox family, and PlugX.

Many of today's threats are built to run covertly on networks and systems, quietly collecting sensitive or personal data, and going undetected for as long as possible. This approach helps to preserve the value of the stolen data and enables repeated use of the same exploits and attack vectors. As a result, threats have become increasingly sophisticated. Rootkits, for example, have become more prevalent. These kernel-level exploits effectively mask the presence of other types of malware, enabling them to persistently pursue the nefarious tasks they were designed to accomplish (such as intercepting keystrokes).

Targeted attacks and advanced persistent threats (APTs) (see the sidebar "Carbanak: The great bank robbery") against specific organizations or individuals are another major concern. In this case, hackers often develop customized attack mechanisms to take advantage of the specific equipment, systems, applications, configurations, and even personnel employed

in a specific organization or at a given location, and quietly collect sensitive data over extended periods.

The increasing speed and sophistication of threats emphasize the need for proactive countermeasures with extensive visibility and control at the application layer of the network computing stack.

Carbanak: The great bank robbery

Carbanak is one of the latest examples of a targeted attack that began in August 2013 and is currently still active. The attackers have sent spear-phishing emails with malicious CPL attachments or Word documents exploiting known vulnerabilities. Once an initial system has been compromised, additional reconnaissance is performed to identify ATMs, financial accounts, or other areas where money can be transferred for eventual extraction. Each raid has lasted two to four months. To date, the attackers have targeted up to 100 financial institutions, causing aggregated losses estimated at $1 billion.

Chapter 3

Recognizing the Challenges of Legacy Security Infrastructures

· ·

In This Chapter

▶ Inspecting weaknesses in legacy port-based firewalls

▶ Examining the shortcomings of intrusion prevention

▶ Addressing device sprawl

· ·

*A*s the application and threat landscape has quickly evolved, the impact within many organizations is that IT has lost control. The inability of their existing security infrastructure to effectively distinguish good or desirable applications from those that are bad or unwanted forces most IT shops to take an inflexible and untenable "all-or-nothing" approach to security, in which they do one of the following:

✔ Take a permissive stance — an approach that ensures the accessibility of important applications, but also allows unwanted applications and threats on the corporate network.

✔ Just say "no" in order to maintain a high state of security, but at the risk of limiting business agility and productivity, alienating users and business units, and creating an underground subculture of backdoor processes to circumvent security controls.

Instead, IT needs the capability to exert granular control and provide in-depth protection down to the level of individual applications, in order to confidently say "yes" to legitimate

requests from the business and its end-users. Unfortunately, traditional network security infrastructures have failed to keep pace and are unable to provide this functionality.

In this chapter, you find out how the new application and threat landscape has challenged these legacy security devices, particularly firewalls, beyond their capability to effectively protect today's networks.

Whatever Happened to the Firewall?

Have you noticed that nobody gets excited about a firewall anymore? There was a time when the firewall was the single most important security device in your network. So, what happened?

The answer is a bit of a cliché, but the Internet has changed everything! Years ago, most firewalls did a pretty good job of controlling traffic in and out of corporate networks. That's because application traffic was generally well behaved. Email would typically flow through port 25, FTP was assigned to port 20, and the whole "web surfing" was hanging, uhhh, port 80. Everybody played by the rules that "ports + protocols = applications" and the firewall had everything under control. Blocking a port meant blocking an application. Nice and simple.

Unfortunately, the Internet has never really been nice and simple. And that is truer today than ever before. Today, the Internet often accounts for 70 percent or more of the traffic on your corporate network. And it's not just port 80 web surfing. Typically, 20 percent to 30 percent of it is encrypted SSL traffic on port 443. Even worse, a plethora of new Internet applications insist on making their own rules. They wrap themselves in other protocols, sneak through ports that don't belong to them, and bury themselves inside SSL/TLS tunnels. In short, they just don't play fair.

All these applications carry some inherent risk to your organization. And they play host to clever new threats that can slip through your firewall undetected. Meanwhile, your firewall just sits there like nothing is wrong because it's still playing by rules that don't exist anymore!

Port-based firewalls have poor vision

Because they're deployed in-line at critical network junctions, firewalls see all traffic and, therefore, are the ideal resource to provide granular access control. The problem, however, is that most firewalls are "far-sighted." They can see the general shape of things, but not the finer details of what is actually happening. This is because they operate by inferring the application-layer service that a given stream of traffic is associated with, based on the port number used in the packet's header, and they only look at the first packet in a session to determine the type of traffic being processed, typically to improve performance. They rely on a convention — not a requirement — that a given port corresponds to a given service (for example, TCP port 80 corresponds to HTTP). As such, they're also incapable of distinguishing between different applications that use the same port/service (see Figure 3-1).
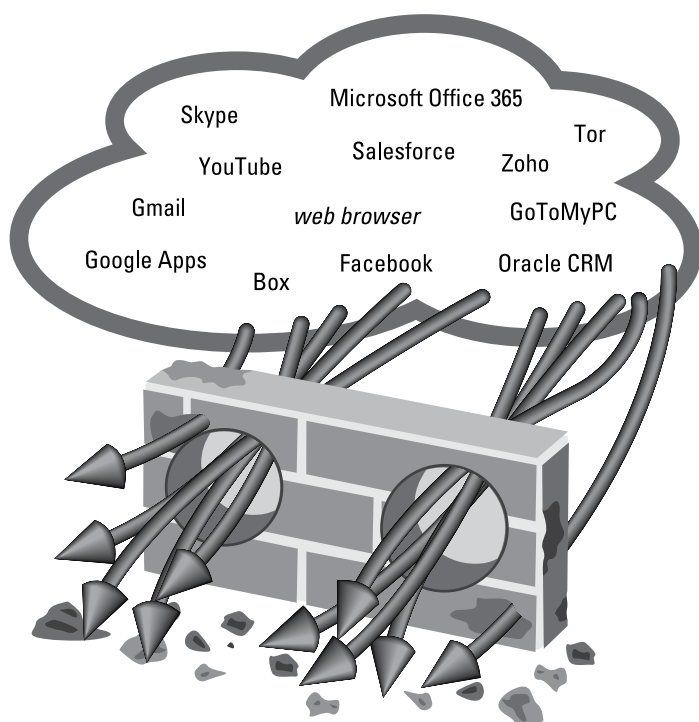
Figure 3-1: Port-based firewalls can't see or control applications.

The net result is that traditional, "port-based" firewalls have basically gone blind. Besides being unable to account for common evasion techniques such as port hopping, protocol tunneling, and the use of nonstandard ports, these firewalls simply lack the visibility and intelligence to discern which network traffic

✔ Corresponds to applications that serve a legitimate business purpose

✔ Corresponds to applications that can serve a legitimate business purpose but, in a given instance, are being used for unsanctioned activities

✔ Should be blocked because it includes malware or other types of threats, even though it corresponds to legitimate business activities

On top of everything else, their control model is typically too coarse-grained. Said firewalls can either block or allow traffic, but offer little variation in between to craft a more appropriate response for all the "gray" applications that enterprises would ultimately like to support — for example, by allowing certain functions or file transfers within an application but not others, allowing but also applying traffic-shaping policies, allowing but scanning for threats or confidential data, or allowing based on users, groups, or time of day.

It doesn't really help matters that the most common steps taken to address the inadequacies of traditional firewalls have, for all intents and purposes, been completely unsuccessful.

Bolt-on functionality is fundamentally flawed

Many purveyors of traditional firewalls have attempted to correct the far-sighted nature of their products by incorporating deep packet inspection (DPI) capabilities. On the surface, adding a measure of application-layer visibility and control in this manner appears to be a reasonable approach. However, the boost in security effectiveness that can be achieved in most cases is only incremental because the additional capability is being "bolted on," and the foundation it's being bolted onto is weak to begin with. In other words, the new functionality is added on rather than integrated, and the port-based

firewall, with its complete lack of application awareness, is still used for initial classification of all traffic. The problems and limitations this leads to include

- ✔ **Applications that should not be on the network are allowed onto the network.**

- ✔ **Not everything that should be inspected necessarily gets inspected.** Because the firewall is unable to accurately classify application traffic, deciding which sessions to pass along to the DPI engine becomes a hit-or-miss proposition.

- ✔ **Security posture gets limited.** The bolted-on application classification ability often doesn't get shared with later enforcement capabilities (for example, file transfer control). This makes it impossible for those enforcement options to be precisely applied "per application."

- ✔ **Policy management gets convoluted.** Rules on how to handle individual applications essentially get "nested" within the DPI portion of the product — which itself is engaged as part of a higher/outer-level access control policy.

- ✔ **Inadequate performance forces compromises to be made.** Inefficient use of system resources and CPU and memory intensive application-layer functionality can put considerable strain on the underlying platform. To account for this situation, administrators can only implement advanced filtering capabilities selectively.

Firewall "helpers" don't help

Over the years, enterprises have also tried to compensate for their firewalls' deficiencies by implementing a range of supplementary security solutions, often in the form of stand-alone appliances. Intrusion prevention systems, antivirus gateways, web filtering products, and application-specific solutions — such as a dedicated platform for instant messaging security — are just a handful of the more popular choices. Unfortunately, the outcome is disappointingly similar to that of the DPI approach, with an additional twist.

Not everything that should get inspected does because these firewall helpers either can't see all the traffic, rely on the same

port- and protocol-based classification scheme that has failed the legacy firewall, or provide coverage only for a limited set of applications. Policy management is an even greater problem given that access control rules and inspection requirements are spread among several consoles and involve multiple policy models. And performance is still an issue as well, at least in terms of having a relatively high aggregate latency.

Then comes the kicker: device sprawl. As one "solution" after another is added to the network, the device count, degree of complexity, and total cost of ownership all continue to rise. Capital costs for the products themselves and all the supporting infrastructure that is required are joined by a substantial collection of recurring operational expenditures, including support/maintenance contracts, content subscriptions, and facilities costs (power, cooling, and floor space) — not to mention an array of "soft" costs such as those pertaining to IT productivity, training, and vendor management. The result is an unwieldy, ineffective, and costly endeavor that is simply not sustainable.

Traditional IPS Is a Poor Match for Today's Threats

Intrusion prevention systems (IPSs) detect and block attacks focused on vulnerabilities that exist in systems and applications. Unlike intrusion detection systems (IDSs), which focus only on alerting, IPS systems are intended to be deployed in-line to actively block attacks as they're detected.

One of the core capabilities of an IPS is the ability to decode protocols to more accurately apply signatures. This allows IPS signatures to be applied to very specific portions of traffic, thereby reducing the percentage of false positives that were often experienced with signature-only systems. It's important to note that most IPS offerings will use port and protocol as the first pass of traffic classification, which, given the evasive characteristics of today's applications, may lead to an erroneous identification of the application. And because IPS systems are focused mainly on attacks, they're typically deployed in

conjunction with a firewall as a separate appliance or as a combination firewall and IPS.

IPS is designed to stop threats using a "find it and kill it" approach. It isn't designed to control applications. But even for stopping threats, IPS has its flaws.

Given the new application and threat landscape, organizations are also reexamining traditional IPSs. The major IPS vendors are struggling to differentiate across several basic elements of IPSs:

- ✔ **Server and data center protection:** There are only a handful of detection and prevention techniques, and most IPS products support them all. These techniques include protocol anomaly detection, stateful pattern matching, statistical anomaly detection, heuristic analysis, blocking of invalid or malformed packets, and IP defragmentation and TCP reassembly (for anti-evasion). Most IPS vendors also use vulnerability-facing signatures (as opposed to exploit-facing signatures) and turn off server-to-client protection to improve performance.

- ✔ **Research and support:** This comes down to how much actual research vendors are doing, and how quickly they can respond to help enterprises protect against new attacks and vulnerabilities. Much is made of the efforts of the research teams of IPS vendors, and while there certainly are differences, much of the research is outsourced to a few industry research stalwarts. The other aspect is critical — regardless of who does the research: Can the vendor deliver timely updates to protect customers from new and emerging threats?

- ✔ **Performance:** Organizations are clearly sensitized to IPS performance issues. The introduction of traffic/application latency and bandwidth/performance are major concerns that cause enterprises to deploy "out-of-band" IPS. Clearly, being able to keep up with enterprise expectations for throughput and latency is top of mind for many customers.

As defenses mature, however, attackers evolve. Given that intrusion detection and prevention systems, like firewalls, are based on legacy techniques that are relatively well understood, new attacks are able to exploit well-known weak spots, including

✔ **Application-borne threats:** Threat developers are using applications, both as targets and as transmission vectors. Applications provide fertile ground for both methods. Some application-borne threats are well understood (for example, many of the threats that move across social networks); others are not. Regardless, attackers find it far easier to piggyback on applications and start their attack with the client. For example, Koobface is a worm that targets users on social media sites such as Facebook and Skype, and is used to steal personal information such as passwords and banking information. CryptoLocker is ransomware that is often propagated via email and encrypts certain files on a victim's computer and requires payment of a ransom (usually in bitcoins) to decrypt the files.

✔ **Encrypted threat vectors:** The other important technique that threats employ is encryption. Although security researchers have warned for years that encryption can be used by various threats, encrypted attacks still need a conduit — enter user-centric applications. Users are easily duped into clicking on encrypted links (too many users think that HTTPS means "safe"), which can send encrypted threats sailing through enterprise defenses. This is increasingly simple on social networks, where the level of trust is extremely high. The other closely related vector is obfuscation via compression — traditional IPSs can't decompress, and thus can't scan compressed content.

A word on data leaks

Some of the biggest information-security news stories over the past few years involve the leaking of confidential or sensitive organizational data via applications (for example, U.S. government agencies and contractors, pharmaceuticals, and retailers). In most cases, the applications that the data leaked across were expressly forbidden — unfortunately, their policies couldn't be enforced with traditional firewalls and IPSs, or alerts (that required manual response) were lost in a sea of information. Given these high-profile security breaches, it's no wonder that organizations are starting to look for a better solution to help protect against such embarrassing incidents.

A common theme here is the level of control needed to prevent these newer threats — controlling applications and content, decrypting SSL/TLS, unzipping content to look for threats — all of which goes well beyond what an IPS traditionally does. A major limitation of an IPS, despite all the work to transition from IDSs, is that IPS remains a negative security model and is architected as such. Put more simply, an IPS relies on a "find it and kill it" model — which doesn't work very well for the types of control necessary to deal with many of these new threats that move over applications. Nor does it lend itself to an architecture and platform capable of decrypting and classifying all traffic.

A positive security model operates by expressly allowing all communications that are known to be benign, appropriate, or necessary, and excluding everything else. A negative security model operates by seeking to classify only undesirable communications and content, and employing countermeasures for those that are known to be bad.

UTM Only Makes What Is Broken Cheaper

Unified threat management (UTM) devices are another approach to modern security challenges that are nonetheless based on traditional techniques. UTM solutions were born as security vendors began bolting intrusion prevention and antivirus add-ons to their stateful firewalls in an effort to reduce the cost of deployment. UTM products don't perform their functions any better than stand-alone devices. Instead, they provide convenience to the customer by consolidating multiple functions into one device. Unfortunately, UTMs have a reputation for being inaccurate, hard to manage, and performing poorly when services are enabled, relegating them to environments where the value of device consolidation outweighs the downside of lost functionality, manageability, or performance.

The primary advantage of the UTM solution is that it typically does a reasonable job of addressing the issues associated with device sprawl. Instead of having all the "helper" countermeasures deployed as separate devices, with UTM they all come in one physical package.

But so what? The result is really no different than the bolted-on approach and, therefore, exhibits the same deficiencies. Inadequate application classification and resulting blind spots in the inspections that are performed remain as fundamental problems, while performance and policy management issues are compounded even further based on having to account for multiple additional countermeasures instead of just one.

It's Time for a Truly Integrated Approach

Traditional port-based firewalls really don't provide value anymore — not in a world where network boundaries are disintegrating and Internet applications are exploding.

But you already know that, which is why you've been forced to make up for their glaring deficiencies with more specialized appliances — intrusion prevention systems, proxies, antivirus, anti-spyware, URL filtering, and more. Sure, these tools add some incremental value, but it's getting harder to justify their additional cost and complexity — especially during challenging economic times.

More security appliances don't necessarily mean a more secure environment. In fact, the complexity and inconsistency associated with such an approach can actually be a detriment to your organization's security strategy.

Clearly, such a strategy doesn't scale. More important, none of these additional products give you the visibility and control you need over the applications running on your network.

It's time to address the core problem. It's time to make the firewall the visibility and enforcement point for a modern security platform. After all, firewalls deployed at key locations in the network are really the best way to gain visibility and control over what enters and leaves the network.

Chapter 4

Solving the Problem with Next-Generation Firewalls

*N*etwork security in most enterprises is fragmented and broken, exposing them to unwanted business risks and ever-rising costs. Traditional network security solutions have failed to keep pace with changes to applications, threats, and the networking landscape. Furthermore, the remedies put forth to compensate for their deficiencies have, for the most part, proven ineffective. It's time to reinvent network security.

This chapter is about the next-generation firewall (NGFW): what it is, what it isn't, and how it can benefit your organization.

The Next-Generation Firewall

To restore the firewall as the cornerstone of enterprise network security, NGFWs "fix the problem at its core." Starting with a blank slate, NGFWs classify traffic by the application's identity in order to enable visibility and control of all types of applications — including web applications, SaaS, and legacy — running on organizational networks.

The essential functional requirements for an effective NGFW include the ability to

- ✔ Identify applications regardless of port, protocol, evasive techniques, or SSL encryption before doing anything else
- ✔ Provide visibility of and granular, policy-based control over applications, including individual application functions
- ✔ Accurately identify users and subsequently use identity information as an attribute for policy control
- ✔ Provide real-time protection against a wide array of threats, including those operating at the application layer
- ✔ Integrate, not just combine, traditional firewall and network intrusion prevention capabilities
- ✔ Support in-line deployments with negligible performance degradation

Typical capabilities of traditional firewalls include packet filtering, network- and port-address translation, stateful inspection, and virtual private network (VPN) support. Typical intrusion prevention capabilities include vulnerability- and threat-facing signatures, and heuristics.

The key to NGFWs is the ability to do everything a traditional firewall does with the advanced capabilities that combine innovative identification technologies, high-performance, and additional foundational features to yield an enterprise-class solution.

Application identification

Establishing port and protocol is a first step in application identification but, by itself, it's insufficient. Robust application identification and inspection enables granular control of the flow of sessions through a firewall based on the specific applications that are being used, instead of just relying on the underlying set of often indistinguishable network attributes (see Figure 4-1).

Positive application identification is the traffic classification engine at the heart of NGFWs. It requires a multifaceted approach to determine the identity of applications on the network, regardless of port, protocol, encryption, or evasive tactics. Application identification techniques used in NGFWs (see Figure 4-2) include
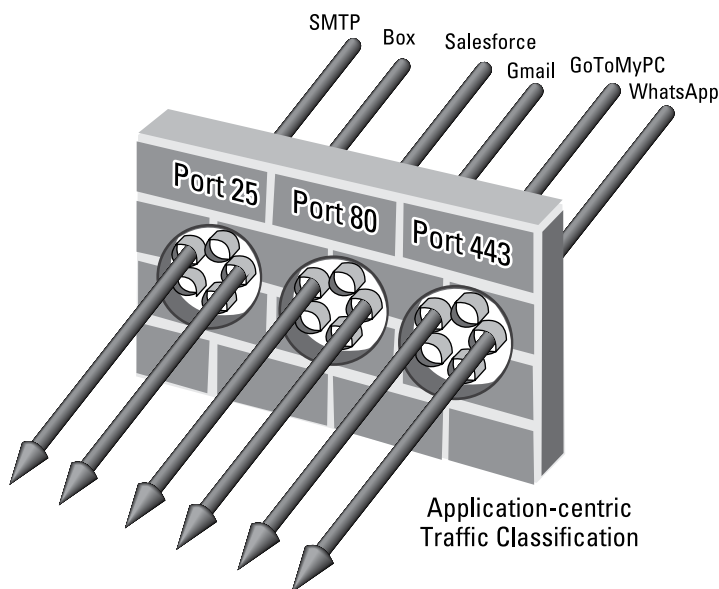
Figure 4-1: Application-centric traffic classification identifies specific applications flowing across the network, irrespective of the port and protocol in use.

✔ **Application protocol detection and decryption:** Determines the application protocol (for example, HTTP) and, if SSL/TLS is in use, decrypts the traffic so that it can be analyzed further. Traffic is re-encrypted after all the NGFW technologies have had an opportunity to operate.

✔ **Application protocol decoding:** Determines whether the initially detected application protocol is the "real one," or if it's being used as a tunnel to hide the actual application (for example, Tor might be inside of HTTPS).

✔ **Application signatures:** Context-based signatures look for unique properties and transaction characteristics to correctly identify the application regardless of the port and protocol being used. This includes the ability to detect specific functions within applications (such as file transfers within SaaS applications).

✔ **Heuristics:** For traffic that eludes identification by signature analysis, heuristic (or behavioral) analyses are applied — enabling identification of any troublesome applications, such as P2P or VoIP tools that use proprietary encryption.
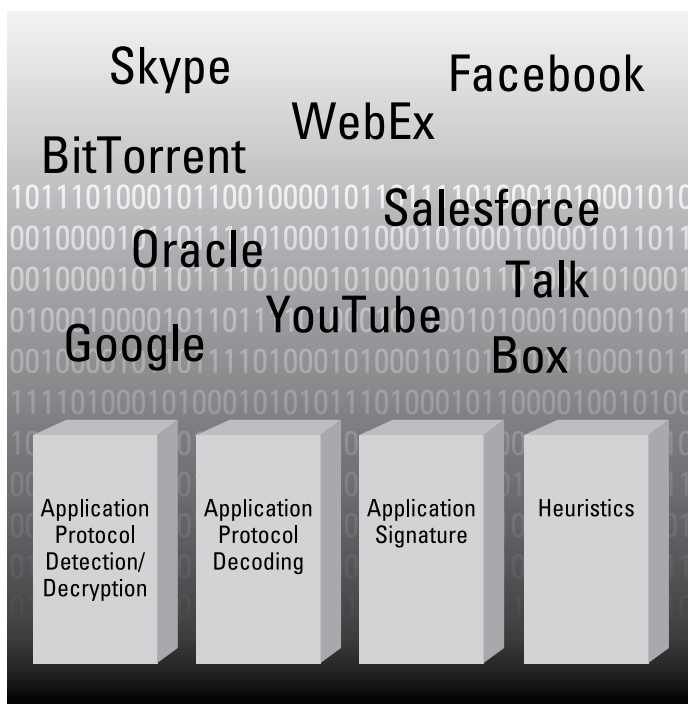
Figure 4-2: NGFW techniques used to identify applications regardless of port, protocol, evasive tactic, or SSL encryption.

> **TIP**
>
> Having the technology to accurately identify applications is important, but understanding the security implications of an application so that an informed policy decision can be made is equally important. Look for an NGFW solution that includes information about each application, and its behaviors and risks, to provide IT administrators with application knowledge such as known vulnerabilities, ability to evade detection, file transfer capabilities, bandwidth consumption, malware transmission, and potential for misuse.

User identification

User identification technology links IP addresses to specific user identities, enabling visibility and control of network activity on a per-user basis. Tight integration with LDAP directories, such as Microsoft Active Directory (AD), supports this objective in two ways:

✔ It regularly verifies and maintains the user-to-IP address relationship using a combination of login monitoring, end-station polling, and captive portal techniques.

✔ It communicates with AD to harvest relevant user information, such as role and group assignments.

These details are then available to

✔ Gain visibility into who specifically is responsible for all application, content, and threat traffic on the network

✔ Enable the use of identity as a variable within access control policies

✔ Facilitate troubleshooting/incident response and reporting

With user identification, IT departments get another powerful mechanism to help control the use of applications in an intelligent manner. For example, a remote access application that would otherwise be blocked because of its risky nature can be enabled for individuals or groups that have a legitimate need to use it, such as IT administrators (see Figure 4-3).
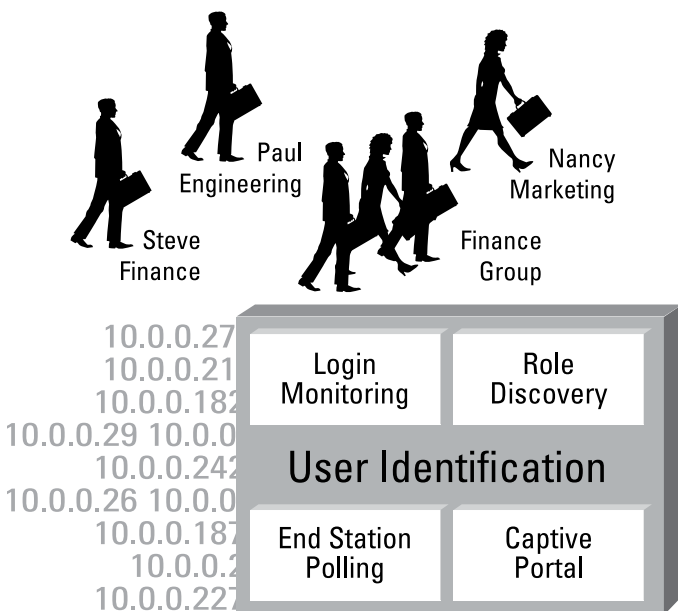
Figure 4-3: User identification integrates enterprise directories for user-based policies, reporting, and forensics.

Content identification

Content identification infuses NGFWs with capabilities previously unheard of in enterprise firewalls:

✔ **Threat prevention:** This component prevents malware and exploits from penetrating the network, regardless of the application traffic on which they ride.

- *Application decoder:* Pre-processes data streams and inspects it for specific threat identifiers.

- *Stream-based malware scanning:* Scanning traffic as soon as the first packets of a file are received — as opposed to waiting until the entire file is in memory — maximizes throughput and minimizes latency.

- *Uniform threat signature format:* Performance is enhanced by avoiding the need to use separate scanning engines for each type of threat. Viruses, command-and-control (C&C) communications, and vulnerability exploits can all be detected in a single pass.

- *Vulnerability attack protection:* Robust routines for traffic normalization and defragmentation are joined by protocol-anomaly, behavior-anomaly, and heuristic detection mechanisms to provide protection from the widest range of both known and unknown threats. This is similar to the functionality provided by IPS devices.

- *Leveraging cloud-based intelligence:* For content that's unknown, the ability to send to a cloud-based security service for rapid analysis and a "verdict" that the firewall can then use.

✔ **URL filtering:** Although not required, URL filtering is another tool sometimes used to classify content. An integrated, on-box URL database allows administrators to monitor and control web surfing activities of employees and guest users. Employed in conjunction with user identification, web usage policies can even be set on a per-user basis, further safeguarding the enterprise from an array of legal, regulatory, and productivity-related risks.

> ✔ **File and data filtering:** Taking advantage of in-depth application inspection, file and data filtering enables enforcement of policies that reduce the risk of unauthorized information transfer, or malware propagation. Capabilities include the ability to block files by their actual type (not based on just their extension), and the ability to control the transfer of sensitive data patterns such as credit card numbers. This complements the granularity of application identification, which for many applications offers the ability to control file transfer within an individual application.

With content identification, IT departments gain the ability to stop threats, reduce inappropriate use of the Internet, and help prevent data leaks — all without having to invest in a pile of additional products that cause appliance sprawl, and that still don't work well because of their lack of integration (see Figure 4-4).
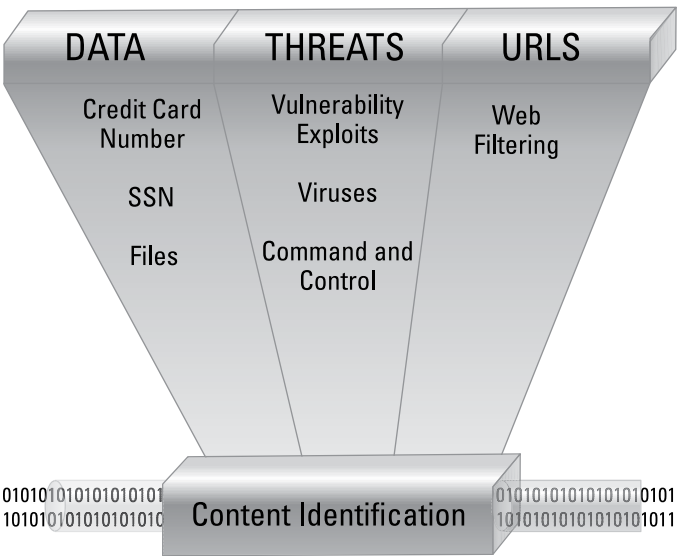
Figure 4-4: Content identification unifies content scanning for threats, confidential data, and URL filtering.

Policy control

Identifying the applications in use (application identification), who is using them (user identification), and what they're using them for (content identification) is an important first step in learning about the traffic traversing the network. Learning what the application does, the ports it uses, its underlying technology, and its behavior is the next step toward making an informed decision about how to treat the application.

When a complete picture of usage is gained, organizations can apply policies with a range of responses that are more fine-grained and appropriate than simply "allow" or "deny" — the only options available in traditional port-based firewalls. This is made possible by the combination of application, user, and content identification, and the positive security model of NGFWs. Traditional port-based firewalls have the security model, but lack intelligence. Other security devices may have some of the intelligence, but not the security model. Examples of policy control options in NGFWs include

- ✔ Allow or deny
- ✔ Allow but scan for exploits, viruses, and other threats
- ✔ Allow based on schedule, users, or groups
- ✔ Decrypt and inspect
- ✔ Apply traffic shaping through QoS
- ✔ Apply policy-based forwarding
- ✔ Allow certain application functions
- ✔ Allow (or prevent) certain types of file transfer
- ✔ Any combination of the aforementioned

High-performance architecture

Having a comprehensive suite of application awareness and content inspection capabilities is of little value if IT administrators are unable to fully engage them due to performance constraints. So, it's important to select an NGFW that is designed from the start to deliver high performance. The issue is not just that these capabilities are inherently resource

intensive. There's also the tremendous traffic volume confronting today's security infrastructure, not to mention the latency sensitivity of many applications. Rated throughput and reasonable latency should be sustainable under heavy loads, even when all application and threat inspection features are engaged simultaneously — which is the ideal configuration from a security perspective.

For traditional security products, especially those with bolted-on capabilities, each high-level security function is performed independently. This multi-pass approach requires low-level packet processing routines to be repeated numerous times. System resources are used inefficiently and significant latency is introduced (see Figure 4-5).
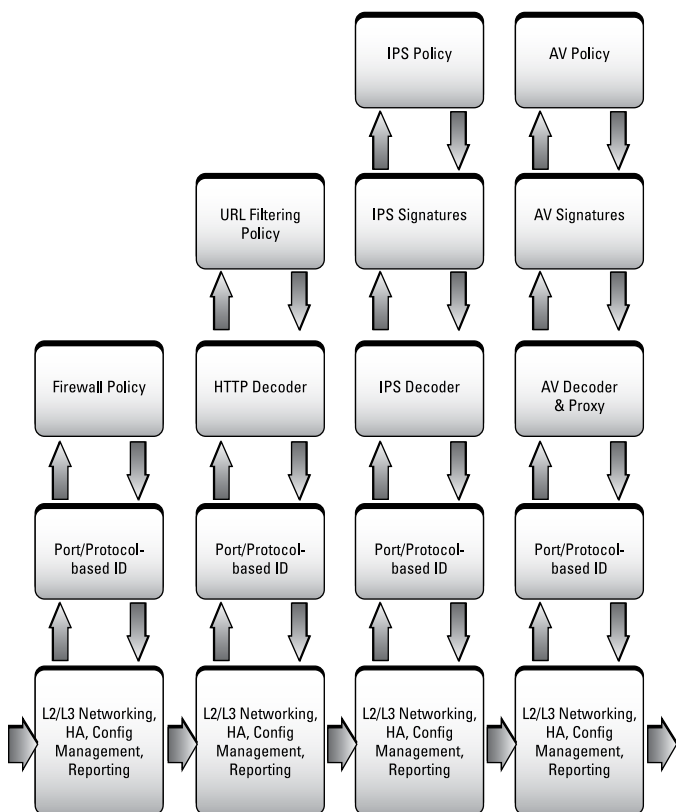
Figure 4-5: Legacy multi-pass architectures.

In contrast, an NGFW that uses a single-pass architecture eliminates repetitive handling of packets, reducing the burden placed on hardware and minimizing latency. Separate data and control planes help provide an enterprise-class solution (see Figure 4-6).
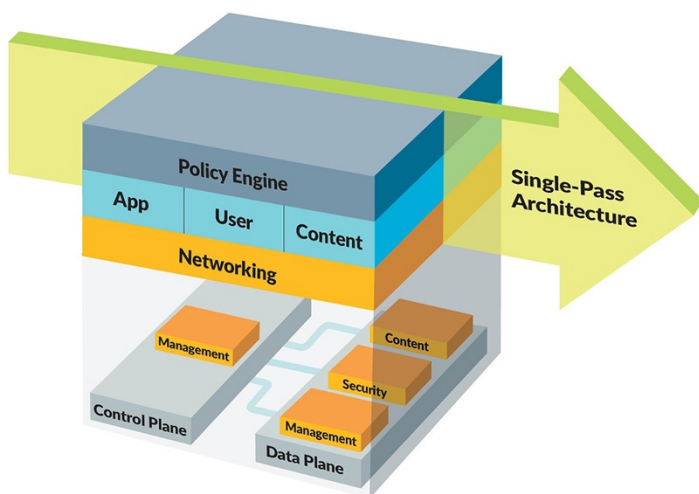
Figure 4-6: Single-pass architecture and separate control and data planes provide performance and availability.

What a Next-Generation Firewall Isn't

There are many network-based security products available that perform functions similar to an NGFW, but they aren't the same thing. Here are some examples:

✔ **Unified threat management (UTM):** UTM appliances host multiple security functions, such as port-based firewall capabilities and basic intrusion prevention. UTM solutions suffer from inferior security capability because not all classification knowledge (for example, the application) is shared with all enforcement options — limiting the flexibility and precision of security policies.

✔ **Proxy-based products:** Proxies (both firewall and caching) sit between source and destination, intercepting traffic and inspecting it by terminating the application session and reinitiating it to the target destination. The proxy establishes the connection with the destination, on behalf of the client, hiding computers on the network behind the proxy. However, only a limited number of applications can be supported because each individual application has to have its own proxy.

✔ **Web application firewalls (WAFs):** A WAF is designed to look at web applications, monitoring them for security issues that may arise due to possible coding errors. WAFs look only at Layer 7, instead of inspecting the entire OSI stack.

REMEMBER

WAFs protect applications, and NGFWs protect networks.

✔ **Vulnerability and patch management:** Vulnerability and patch management solutions scan hosts for known vulnerabilities in software and operating systems, verify that patches and updates are installed, and correct the identified vulnerability. This is not a function of NGFWs.

✔ **Data loss prevention (DLP):** These solutions prevent transmission of data that matches an identified pattern (such as credit card numbers). These solutions are implemented for network functions with no real-time requirements regarding speed and latency.

✔ **Secure web gateways:** These solutions use URL categorization to enforce policies regarding user access to websites and block malware propagated by malicious websites. Compared to NGFWs, these solutions have limited capabilities and are easily circumvented by users.

✔ **Secure messaging gateways:** These include spam filters and IM gateways, and provide antispam and antiphishing protection, antivirus scanning, attachment filtering, content filtering, data loss prevention, and policy compliance and reporting. Unlike NGFWs, these functions are not performed in real-time and are used for applications like email, which is less latency sensitive.

Benefits of Next-Generation Firewalls

NGFWs produce numerous benefits over traditional network security infrastructures and solutions:

- ✔ **Visibility and control:** The enhanced visibility and control provided by NGFWs enable enterprises to focus on business-relevant elements such as applications, users, and content for policy controls, instead of having to rely on nebulous and misleading attributes like ports and protocols, and to better and more thoroughly manage risks and achieve compliance, while providing threat prevention for allowed applications.

- ✔ **Safe enablement:** Achieve comprehensive coverage — by providing a consistent set of protection and enablement capabilities for all users, regardless of their location.

- ✔ **Simplification:** Reduce complexity of the network security and its administration — by obviating the need for numerous stand-alone products. This consolidation reduces hard capital costs, as well as ongoing "hard" operational expenses, such as support, maintenance, and software subscriptions, power and HVAC, and "soft" operational expenses, such as training and management.

- ✔ **IT and business alignment:** Enable IT to confidently say "yes" to the applications needed to best support the business — by giving them the ability to identify and granularly control applications while protecting against a broad array of threats. This includes the ability for IT to add security rules "in stages" — actively investigating traffic that is unknown (based on advanced visibility) and then adding security rules as appropriate.

Chapter 5

Deploying Next-Generation Firewalls

● ●

In This Chapter

▶ Implementing employee, desktop, and network controls

▶ Asking the right questions to help you choose the best solution

▶ Designing your network for optimum performance and security

● ●

*F*ar too often, technical solutions are implemented without considering the implications for an organization's overall security strategy. To avoid this mistake, it's important to ensure that your policies are up to date and the technology solutions you're considering support a comprehensive security strategy. It's also important to have a clear understanding of your organization's requirements.

This chapter describes the different types of controls that must be considered in an organization's security policies and provides specific examples of technical requirements you need to explore as you define your requirements and develop a request for proposal (RFP) for your vendors. Finally, it covers the importance of properly segmenting your network and sensitive data and how to address mobile users.

Safe Enablement through Smart Policies

Enablement is first and foremost about education and knowledge of applications, behavior, risks, and users. In the case of Software as a Service (SaaS) and web applications, the users

have long since decided on the benefits, although there continue to be opportunities for education on the choice of the best application for the job. IT's role is that of an advisor and mentor, advising users about risks and behaviors and guiding them regarding which of the array of available applications might be best at solving their requirements. But enablement is also about raising the awareness of the risks associated with applications. For IT to be relevant, it needs to evaluate and adopt SaaS and web applications wholeheartedly and without prejudice. When that's achieved, IT can successfully educate the users on all the risks associated with the use of those applications.

It comes down to using the right tool for the job and being smart about it. For example, in a heavily regulated environment such as stock trading, the use of instant messaging may be prone to retention and auditability rules. IT's role is to educate the traders on the implications of each of the tools, participate in the development of the use policy, and subsequently monitor and enforce its use. In this example, that policy could prevent the traders from using Facebook chat for instant messaging, but enable MSN for that use instead.

REMEMBER

Governance and its management counterpart work best if they're based on a set of smart corporate policies that are developed by the four major stakeholders in the application landscape: IT, HR, executive management, and the users. Clearly, IT has a role to play, but it can't be the strictly defined role that IT so often plays, nor can IT be lax about its role as the enabler and governor of applications and technology.

TIP

If application controls are going to be implemented and enforced, they should be part of the overarching corporate security policy. As part of the process of implementing an application control policy, IT should make a concerted effort to learn about all applications that are being used in the organization, including SaaS and web applications. This includes embracing them for all their intended purposes and, if needed, proactively installing them or enabling them in a lab environment to see how they act. Peer discussions, message boards, blogs, and developer communities are valuable sources of information.

Employee controls

Most companies have some type of application usage policy, outlining which applications are allowed and which are prohibited. Every employee is expected to understand the contents of this policy and the ramifications of not complying with it, but there are a number of unanswered questions, including the following:

- ✔ Given the increasing number of "bad" applications, how will an employee know which applications are allowed and which are prohibited?

- ✔ How is the list of unapproved applications updated, and who ensures employees know the list has changed?

- ✔ What constitutes a policy violation?

- ✔ What are the ramifications of policy violations — firing or a reprimand?

The development of policy guidelines is often challenging because tension between risk and reward has polarized opinions about what should be allowed and what should be prohibited. At the core of the issue is the fact that the two organizational groups that are typically involved in policy development — IT security and HR — have largely been sidelined during adoption of new technologies. To build a policy for safe use after new technologies and applications have been implemented is no easy task.

Documented employee policies need to be a key piece of the application control puzzle, but employee controls as a stand-alone mechanism will remain largely ineffective for safe enablement of applications.

Desktop controls

Desktop controls present IT departments with significant challenges. Careful consideration should be applied to the granularity of the desktop controls and the impact on employee productivity. As with employee policies, desktop controls are a key piece to the safe enablement of applications in the enterprise and, if used alone, will be ineffective for several reasons.

The drastic step of desktop lockdown to keep users from installing their own applications is a task that is easier said than done:

- ✔ Laptops connecting remotely, Internet downloads, USB drives, and email are all means of installing applications that may or may not be approved.

- ✔ Removing administrative rights completely has proven to be difficult to implement and, in some cases, limits end-user capabilities.

- ✔ USB drives are now capable of running applications, so an application can, in effect, be accessed after the network admission is granted.

Desktop controls can complement the documented employee policies as a means to safely enable applications.

Network controls

At the network level, what is needed is a means to identify all applications and block or control them. By implementing network-level controls, IT is able to minimize the possibility of threats and disruptions stemming from the use of applications.

Several possible control mechanisms can be used at the network level, each of which carries certain drawbacks that reduce their effectiveness:

- ✔ **Stateful firewalls:** Stateful firewalls can be used as a first line of defense, providing coarse filtering of traffic and segmenting the network into different password-protected zones. One drawback to stateful firewalls is that they use protocol and ports to identify and control what gets in and out of the network. This port-centric design is relatively ineffective when faced with applications that hop from port to port until they find an open connection to the network.

- ✔ **Intrusion prevention systems (IPSs):** An IPS added to a firewall deployment enhances the network threat-prevention capability by looking at a subset of traffic and blocking known threats or bad applications. IPS offerings lack the breadth of applications and the performance required to look at all traffic across all ports and, as such, cannot be considered a full solution.

Managing a firewall and IPS combination is usually a cumbersome task, requiring different management interfaces pointed at separate policy tables. Simply put, the current bolt-on solutions don't have the accuracy, policy, or performance to solve today's application visibility and control requirements.

✔ **Proxy solutions:** Proxy solutions are another means of traffic control, but they look at a limited set of applications or protocols and, as such, only see a partial set of the traffic that needs to be monitored. So an application will merely see a port blocked by a proxy and hop over to the next one that is open. By design, proxies need to mimic the application they're trying to control, so they struggle with updates to existing applications as well as development of proxies for new applications. A final issue that plagues proxy solutions is throughput performance brought on by how the proxy terminates the application, and then forwards it on to its destination.

The challenge with all these network controls is that they don't have the ability to identify all applications; they look at only a portion of the traffic and suffer from performance issues.

Defining Your Requirements and Developing a Request for Proposal

After creating or updating your organization's security policies, it's time to define your organization's requirements for a next-generation firewall (NGFW) solution. At a very high level, this includes doing your due diligence on the vendors you're considering. You should be asking questions about your potential vendors, such as the following:

✔ What is the company's vision and how well does it execute on that vision? Does it address the current threat landscape, as well as support the IT architectures desired (for example, virtualization and cloud)?

✔ How innovative is the company? How well-regarded is the company — in terms of both technology and its dedication to customer success?

- ✔ What is the company's culture?

- ✔ What is its development process? What is its quality assurance process?

- ✔ What is the size and financial condition of the company?

- ✔ Is the company a potential acquisition target? If so, is it more likely to be acquired in order to quickly gain an edge because of its innovation and proprietary technology, or to kill off a competitor?

- ✔ How large is its installed customer base?

- ✔ Does it have other customers (perhaps even competitors) that are in a similar industry as your own organization?

- ✔ Does it have any reference accounts or customer success stories to share?

Next, define your organization's technical requirements. Fortunately, you don't necessarily have to reinvent the wheel here. Begin by taking a look at your organization's security policies (see the previous section) to see what capabilities will be needed in order to implement and support those policies.

There are also plenty of examples of firewall and network security requirements practically everywhere. In fact, most regulatory compliance requirements relating to data protection are based on information security best practices. Even if your organization isn't subject to any of these regulations, using them for guidance isn't necessarily a bad thing. For example, the Payment Card Industry Data Security Standard (PCI DSS), which is applicable to *every* organization that processes a credit or debit card, defines several firewall requirements, all of which can easily be modified and incorporated into a formal RFP for your organization.

Drilling down into specific feature requirements, your RFP should address several requirements, covered in the following sections.

Application identification

Describe how the gateway will accurately identify applications and the mechanisms used to classify applications:

- ✔ Is identification based on IPS or DPI technology? If so, how are accuracy, completeness, and performance issues addressed when scanning network traffic?

- ✔ How is the traffic classification mechanism differentiated from other vendors?

- ✔ How are unknown applications handled?

- ✔ Are custom application signatures supported?

- ✔ How is SSL/TLS-encrypted traffic identified, inspected, and controlled?

- ✔ How do the SSL controls delineate between personal traffic (such as banking, shopping, and health) and nonpersonal traffic?

- ✔ Are applications identified in a business-relevant way, and what is the process for updating the application database (for example, software upgrade or dynamic update)?

- ✔ If a new application is needed, what is the process for adding it to the device?

- ✔ Can an end-user submit an application for identification and analysis and/or define custom applications?

- ✔ Does the product support URL filtering? Describe the URL filtering database. Is the database located on the device or on another device?

- ✔ Describe/list any other security functions that can leverage the application information collected, including drill-down details and user visibility features.

Application policy control

Describe the process for implementing policy-based application controls, all application policy control parameters (such as user, IP address, date/time), and how they can be used:

- ✔ Can policy controls be implemented for all applications identified?

- ✔ Can policy controls be implemented for specific users or groups?

- ✔ Can all policy controls be based on the application identified (for example, file blocking)?

✔ How are remote access environments supported (for example, Citrix and Terminal Services)?

✔ Can the solution perform traditional firewall-based access controls?

✔ Can policy controls be implemented from a single management interface?

✔ Are users warned when they attempt to access a URL or application that violates policy?

Threat prevention

Describe the intrusion prevention features and antivirus engine:

✔ List the types of threats that can be blocked. List the file types that can be blocked.

✔ Is data filtering supported?

✔ Can the threat prevention engine scan inside SSL/TLS-encrypted traffic? Compressed traffic?

✔ Can the use of strong cipher and encryption protocol versions be enforced?

✔ Can combinations of events (indicators of compromise, or IOCs) be recognized so that security operators can be notified of possible issues (for example, compromised endpoints)?

Management

Describe the management capabilities and visibility tools that enable a clear picture of the traffic on the network:

✔ Does device management require a separate server or device?

✔ Are application policy controls, firewall policy controls, and threat prevention features all enabled from the same policy editor?

✔ What tools provide a summary view of the applications, threats, and URLs on the network?

✔ Describe any log visualization tools.

✔ Are reporting tools available to understand how the network is being used and to highlight changes in network usage?

✔ Describe the logging and reporting capabilities of the solution.

✔ Describe how management access is ensured when the device is under heavy traffic load.

✔ Are any central management tools available?

Networking

Describe the network integration and implementation capabilities:

✔ Describe any Layer 2 or Layer 3 capabilities.

✔ Are 802.1q VLANs supported? What is the VLAN capacity?

✔ Is dynamic routing supported (for example, OSPF, BGP, and RIP)?

✔ Is equal-cost multi-path routing (ECMP) supported for performance and reliability?

✔ Describe any QoS or traffic shaping features.

✔ Is IPv6 supported?

✔ Are IPsec VPNs supported? SSL VPNs?

✔ What deployment options are available (for example, in-line, tap, passive)?

✔ Describe any high availability (HA) capabilities.

Hardware

Is the solution software-based, an OEM server, or a purpose-built appliance? Describe the architecture along with any performance implications.

IT solution

Does the NGFW need to integrate with other systems? This can be other security subsystems (for example, SIEM) or a larger IT architecture (for example, cloud and SDN). What APIs and features are present to support the integration?

Deployment Flexibility Matters

It's important to design your network to maximize perfor-
mance and efficiency. Properly deploying an NGFW in the
most optimal location or locations on your network is no less
important. Segmentation is a key concept in the proper design
of networks and deployment of firewalls. Although there are
many different ways to segment a network, NGFWs bring a
unique combination of hardware- and software-related seg-
mentation capabilities that enable organizations to isolate key
sections of their network, such as a data center.

The concept of security zones, which for purposes of isolat-
ing sensitive data or critical network infrastructure (again, for
example, a data center), are roughly equivalent to network
segments (see Figure 5-1). A security zone is a logical container
for physical interfaces, VLANs, a range of IP addresses, or a
combination thereof. Interfaces that are added to each security
zone can be configured in Layer 2, Layer 3, or a mixed mode,
thereby enabling deployment in a wide range of network envi-
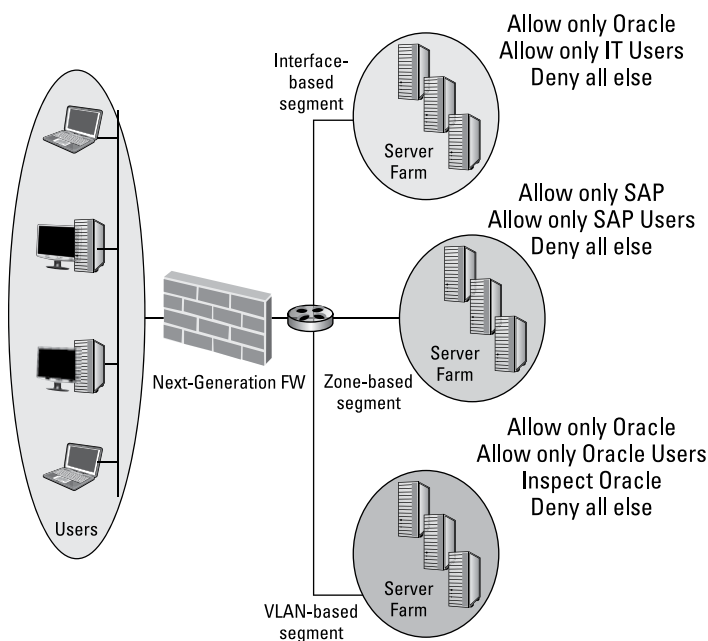ronments without requiring network topology modifications.

Figure 5-1: Network segmentation and security zones.

Many different technologies can be used to segment the network, but when looking at segmentation as a way to isolate the sensitive data or critical infrastructure, several key requirements need to be taken into account:

✔ **Flexibility:** Segmenting the network for security purposes may sometimes require the modification of the network architecture, a task that most companies will avoid if at all possible. The ability to segment a network using IP address ranges, VLANs, physical interfaces, or a combination thereof, is paramount.

✔ **Policy-based security:** Policies must be based on the identity of users and the applications in use — not just IP addresses, ports, and protocols. Without knowing and controlling exactly who (users) and what (applications and content) has access within a segment, sensitive data may be exposed to applications and users that can easily bypass controls based on IP addresses, ports, and protocols.

✔ **Performance:** Segmentation means applying in-depth security policies in a network location that is typically business-critical, high-volume traffic. This means it's critical that the solution delivering the secure segment operate at high speeds with very high session rates and minimal latency.

Addressing Mobile and Remote Users

Another technical limitation for traditional firewalls is providing visibility and control for users who are mobile or remote, beyond the perimeter established by enterprise firewalls. The challenge for NGFWs in this case is to deliver a solution that provides the same degree of protection and application enablement received by users on the local network without having to manage a completely independent set of policies.

Another major challenge is to avoid the limitations and disadvantages associated with the current crop of solutions in this area, including the following:

- ✔ **Cloud or CPE-based proxies:** Associated web services and products typically focus on a narrow traffic stream (for example, port 80/HTTP only), can have a limited set of services/countermeasures (such as URL or malware filtering only), and — because they rely on a proxy architecture — often have to allow many applications to bypass their filters in order to avoid breaking them.

- ✔ **Backhaul via VPN technology:** Whether it's IPsec or SSL-based makes little difference. There is an inevitable bump in latency as client traffic is directed back to one of a few central sites where the VPN gateways are typically located. Of even greater concern, however, is the lack of application visibility and control of the head-end devices that are subsequently used to identify and filter this traffic.

In comparison, a solution that relies on a persistent client that can be installed on demand provides a better alternative. Like the VPN-based approach, remote traffic is sent over a secure tunnel. The difference in this case is that the connection is automatically made to the nearest NGFW — whether it's deployed at one of an organization's hub facilities, out in a regional or branch office location, or as part of a public/private cloud implementation. The latency impact is thus minimized, and the user's session is protected and controlled by the full portfolio of application-, user-, and content-oriented identification and inspection technologies — exactly as if the user were operating on the local network instead of remotely. The net result is an easy-to-implement solution that provides remote and mobile users with the same degree of application enablement and protection as their in-office counterparts.

Chapter 6

Ten Evaluation Criteria for Next-Generation Firewalls

● ●●

*T*his chapter gives you a few answers to look for from the vendors you're considering, once you've developed your request for proposal (RFP).

Note: If you haven't yet developed an RFP to define your next-generation firewall requirements, go to Chapter 5 — go directly to Chapter 5, do not pass Go, do not collect $200!

Identify Applications, Not Ports

Identifying an application as soon as the firewall sees it, irrespective of port, protocol, SSL/TLS encryption, or other evasive tactics, provides the greatest amount of application knowledge and the best opportunity for policy control.

Finally, it is important that the next-generation firewall (NGFW) has an extensive library of application signatures installed on the device, in order to avoid any latency issues that may occur with a hosted or "in-the-cloud" database. The library should be regularly updated with new, business-relevant application signatures from the vendor or through a subscription service, and signature updates should be automated (if desired).

Application identification is at the core of traffic classification on NGFWs. It is intelligent, scalable, and extensible, and always on — across all ports and on all traffic. If this isn't true, it isn't an NGFW — and, more important, its security capability will be limited.

Identify Users, Not IP Addresses

Seamless integration with enterprise directory services (such as Active Directory, LDAP, and eDirectory) enables administrators to tie network activity to users and groups, not just IP addresses. When used with application- and content-identification technologies, IT organizations can leverage user and group information for visibility; policy creation; forensic investigation; and reporting on application, threat, web surfing, and data transfer activity.

User identification helps address the challenge of using IP addresses to monitor and control the activity of specific users — something that was once fairly simple, but has become difficult as enterprises moved to an Internet-centric model.

Compounding the visibility problem is an increasingly mobile enterprise, where employees access the network from virtually anywhere around the world, internal wireless networks reassign IP addresses as users move from zone to zone, and network users are not always company employees. The result is that the IP address is now an inadequate mechanism for monitoring and controlling user activity.

Look for the following techniques in NGFWs to verify and maintain the user–to–IP address relationship and accurately identify users:

- **Login monitoring:** Login activity is monitored to correlate an IP address to user and group info when a user logs in to the domain.

- **End-station polling:** Each active PC is polled to verify IP address information to maintain accurate mapping when users move around the network without reauthenticating to the domain.

- **Captive portal:** Associates a user and an IP address in cases where hosts are not part of the domain via a web page–based authentication form.

- **Ease of deployment:** User identification should be performed without impacting critical infrastructure. Some solutions require an agent to be installed on every domain controller in the organization, which can impact performance and significantly complicate deployment.

Identify Content, Not Packets

With employees using any application they desire and surfing the web with impunity, it's no wonder that enterprises struggle to protect the network from threat activity. The first step in regaining control over the threat activity is to identify and control applications to reduce the unwanted or bad application activity — commonly used as threat vectors. Next, policies to control content can be implemented to complement the application usage control policies.

Content identification capabilities in an NGFW should include the following:

✔ **Threat prevention:** Look for innovative features to address changes in the threat landscape and prevent application vulnerabilities, spyware, and viruses from penetrating the network. Examples of such features include application decoders that take streams of application data that have been reassembled and parsed, and inspect the stream for specific threat identifiers, as well as uniform threat engines and signature formats to detect and block a wide range of malware (such as viruses, C&C, and vulnerability exploits) in a single pass instead of using a separate set of scanning engines and signatures for each type of threat.

✔ **Stream-based virus scanning:** This technique begins scanning as soon as the first packets of a file are received as opposed to waiting until the entire file is loaded into memory to begin scanning. Stream-based virus scanning minimizes performance and latency issues by receiving, scanning, and sending traffic to its intended destination immediately without having to buffer and then scan the file.

✔ **Vulnerability attack protection:** Application vulnerability prevention is enabled using a set of intrusion prevention system (IPS) features to block known and unknown network and application-layer vulnerability exploits, buffer overflows, denial-of-service (DoS) attacks, and port scans from compromising and damaging enterprise information resources. IPS mechanisms include the following:

• Protocol decoders and anomaly detection

• Stateful pattern matching

- Statistical anomaly detection

- Heuristic-based analysis

- Block invalid or malformed packets

- IP defragmentation and TCP reassembly

- Custom vulnerability and C&C signatures

Traffic is normalized to eliminate invalid and malformed packets, while TCP reassembly and IP defragmentation is performed to ensure the utmost accuracy and protection despite any attack evasion techniques.

✔ **URL filtering:** The URL filtering database should be on-box to reduce latency issues associated with hosted databases. Customization features should include the ability to create custom URL categories and to create granular policies for specific groups and users that can allow, block, or warn and then allow access to websites.

✔ **File and data filtering:** Data filtering enables administrators to implement policies that reduce the risks associated with the transfer of unauthorized files/data.

- *File blocking by type:* Control the flow of a wide range of file types by looking deep within the payload to identify the file type (as opposed to looking only at the file extension). File blocking capability should have the flexibility to vary by application.

- *Data filtering:* Control the transfer of sensitive data patterns such as credit card and social security numbers in application content or attachments.

- *File transfer function control:* Control the file transfer functionality within an individual application, allowing application use yet preventing undesired inbound or outbound file transfer.

✔ **Unknown content analysis:** For content elements that are unknown, and that could potentially represent a threat (for example, unknown URLs and files), integration with a malware analysis environment (for example, threat sandbox) allows the NGFW to dynamically learn about new threats.

Visibility

NGFWs give IT administrators actionable data presented in an effective manner — the ability to quickly and easily view specific, detailed application, user, and content information is invaluable.

In particular, visibility at this level allows administrators to continually analyze and refine their security posture.

Control

A robust NGFW provides granular application usage control policies, such as any combination of the following:

✔ Allow or deny

✔ Allow certain application functions and apply traffic shaping

✔ Allow certain file transfers (for example, by type or per application)

✔ Allow but scan

✔ Decrypt and inspect

✔ Allow for certain users or groups

Performance

In-line NGFWs must perform advanced network security functions that are computationally intensive — and they must do so in real-time while introducing little or no latency. An NGFW needs to be capable of handling high throughputs using a variety of implementation options (ranging from purpose-built hardware appliances to horizontally scalable virtual appliances). Ideally, to ensure availability of management and packet processing, the management plane and control plane should be separate.

Flexibility

Networking flexibility helps ensure compatibility with virtually any organization's computing environment. Enabling implementation without the need for redesign or reconfiguration depends on supporting a wide range of networking features and options, such as the following:

- ✔ 802.1q and port-based VLANs
- ✔ Trunked ports
- ✔ Equal-cost multi-path routing (ECMP)
- ✔ Transparent mode
- ✔ Dynamic routing protocols (such as OSPF and BGP)
- ✔ IPv6 support
- ✔ IPsec and SSL VPN support
- ✔ High-capacity interfaces and multiple, mixed modes (such as tap, Layer 1, Layer 2, and Layer 3)

Flexibility to integrate with a variety of IT architectures (such as cloud, SDN, and VDI) is also an important consideration. APIs, as well as features that maintain security in dynamic, on-demand environments, are key enablers.

Reliability

Reliability helps ensure nonstop operations and entails features such as the following:

- ✔ Active-passive and/or active-active failover
- ✔ State and configuration synchronization
- ✔ Redundant components (such as dual power supplies)

Scalability

Scalability is primarily dependent on having solid management capabilities (including centralized device and policy management, and synchronization among devices) and high-performance architectures and implementations. Both physical and virtual implementations are important to cover the full range of customer use cases.

Manageability

Manageability is an important characteristic to look for in an NGFW. A sophisticated solution that is too difficult to administer and maintain will inevitably fail to achieve maximum effectiveness and even risks being deployed in an incorrect — and insecure — manner. Important management capabilities include the following:

- ✔ Local and remote management
- ✔ Centralized management
- ✔ Role-based administration
- ✔ Automatic signature updates
- ✔ Real-time monitoring of device status and security events
- ✔ Robust logging and customizable reporting

Glossary

advanced persistent threat (APT): A sustained Internet-borne attack, usually perpetrated by a group with significant resources, such as organized crime or a nation-state.

antivirus (AV): Antivirus (or antimalware) software.

BitTorrent: A peer-to-peer (P2P) file-sharing communications protocol that distributes large amounts of data widely without the original distributor incurring the costs of hardware, hosting, and bandwidth resources. Instead, each user supplies pieces of the data to newer recipients, reducing the cost and burden on any given individual source.

CPL: The filename extension for Control Panel items in Microsoft Windows.

customer-premises equipment (CPE): Service provider equipment that is located on the customer's premises.

equal-cost multi-path (ECMP) routing: A routing method of packet forwarding via multiple paths for enhanced performance and reliability.

Federal Information Security Management Act (FISMA): A legislative mandate that establishes a framework for U.S. government agencies to protect government information, operations, and assets.

Financial Industry Regulatory Authority (FINRA): An independent not-for-profit organization responsible for ensuring the U.S. securities industry operates fairly and honestly.

Gramm-Leach-Bliley Act (GLBA): Establishes information security requirements for the financial industry, including governance for the collection, disclosure, and protection of personally identifiable information (PII). Also known as the Financial Services Modernization Act of 1999.

Health Insurance Portability and Accountability Act (HIPAA): U.S. legislation passed in 1996 that, among other things, protects the confidentiality and privacy of protected health information (PHI).

high availability (HA): A system or component that is designed for maximum reliability with redundant components and no single points of failure.

Hypertext Transfer Protocol (HTTP): The primary communication protocol of the Internet.

Hypertext Transfer Protocol over SSL/TLS (HTTPS): A secure communication protocol widely used on the Internet.

instant messenger (IM): A type of real-time online chat over the Internet.

Internet Protocol Security (IPsec): A protocol suite for protecting communications over IP networks using authentication and encryption.

intrusion prevention system (IPS): A security appliance or software that detects and prevents known vulnerability exploits.

Koobface: Koobface is a worm that tricks Facebook users into downloading and installing a fake update of the Adobe Flash player. Among other things, Koobface attempts to collect sensitive information such as credit card numbers from an infected PC.

Lightweight Directory Access Protocol (LDAP): An open standards–based protocol for accessing and maintaining distributed directory services.

Microsoft Remote Procedure Call (MS-RPC): An interprocess communications protocol used on Microsoft Windows networks for client and server software communications.

Open Systems Interconnection (OSI) model: The seven-layer reference model for networks. The layers are Physical, Data Link, Network, Transport, Session, Presentation, and Application.

Payment Card Industry Data Security Standard (PCI DSS): A proprietary information security standard mandated for organizations that handle American Express, Discover, JCB, MasterCard, or Visa payment cards.

peer-to-peer (P2P): A distributed application architecture that enables sharing between nodes.

Quality of Service (QoS): A measure of the overall performance of a network, typically including availability, bit rate, delay, error rate, jitter, and throughput.

Secure Shell (SSH): A set of standards and an associated network protocol that allows establishing a secure channel between a local and a remote computer.

Secure Sockets Layer/Transport Layer Security (SSL/TLS): A transport layer protocol that provides session-based encryption and authentication for secure communication between clients and servers on the Internet.

Security Incident and Event Management (SIEM): Security technology that provides real-time analysis of network security alerts.

Server Message Block (SMB): An application-layer protocol also known as Common Internet File System (CIFS).

Simple Mail Transfer Protocol (SMTP): The Internet standard for email using TCP port 25 (by default).

Skype: An application that allows users to make telephone calls over the Internet. Additional features include instant messaging, file transfer, and video conferencing.

Software as a Service (SaaS): A category of cloud computing services in which the customer is provided access to a hosted application that is maintained by the service provider.

Software Defined Networking (SDN): An approach to networking that uses virtualization to abstract higher-level network services from underlying physical hardware.

stateful inspection: Maintains the status of active connections through the firewall to dynamically allow inbound replies to outbound connections. Also known as dynamic packet filtering.

TeamViewer: Provides remote control of PCs over the Internet, allowing a user to instantly take control over a computer anywhere on the Internet, even through firewalls.

Tor: A system that enables users to communicate anonymously over the Internet.

Transmission Control Protocol (TCP): A connection-oriented protocol responsible for establishing a connection between two hosts and guaranteeing the delivery of data and packets in the correct order.

UltraSurf: Implements a proxy with complete transparency and a high level of encryption that enables users to browse any website freely. It's used heavily in countries with Internet censorship.

uniform resource locator (URL): An Internet address in the form `http://www.example.com`.

User Datagram Protocol (UDP): A connectionless-oriented protocol often used for time-sensitive, low-latency communications that don't require guaranteed delivery.

virtual desktop infrastructure (VDI): A desktop operating system environment (OSE) that is virtualized and hosted on a centralized server.

Virtual Local Area Network (VLAN): A LAN segment that is partitioned by broadcast domain at Layer 2 (Data Link) of the OSI model, typically configured on a switch or router.

virtual private network (VPN): An encrypted tunnel that extends a private network over a public network (such as the Internet).

Voice over Internet Protocol (VoIP): Technology that enables voice communications over IP.

zlib: A software library used for data compression.

THIS COULD BE

THE END

OF BREACHES

Discover the power of Palo Alto Networks Next-Generation
Prevention Platform. End-to-end cybersecurity for any business.

paloalto
NETWORKS®

See where it all stops: **go.paloaltonetworks.com/TheEnd**

Regain control of the applications and users on your network!

Traditional firewalls haven't changed much over the past 20 years and can no longer protect your network. That's because they were never designed to control all of the evasive, port-hopping, and encrypted Internet and cloud-based applications that are so common today. You've added intrusion prevention, proxies, antivirus, URL filtering, and much more — but an uncoordinated mix of functions is no match for today's security challenges.

- *The ever expanding application landscape — and how it creates new risks and challenges for your organization*

- *Why traditional firewalls are ineffective against today's threats — and why quick fixes and add-on capabilities don't work*

- *What a next-generation firewall is — what it isn't, and why you need one (or more)*

- *How to get the most out of your firewall — by creating effective policies, asking the right questions, and segmenting your network for optimum performance*

- *Discover advanced features and capabilities — that make next-generation firewalls a powerful solution to protect and control what runs on your network*

Lawrence C. Miller, CISSP, has worked in information security for more than 20 years. He is the coauthor of *CISSP For Dummies* and a dozen other titles. He is also a Palo Alto Networks customer and liked it so much he bought the company — well, he's not that rich (yet) — but he did write this book!

FOR DUMMIES®
A Wiley Brand

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley's ebook EULA.