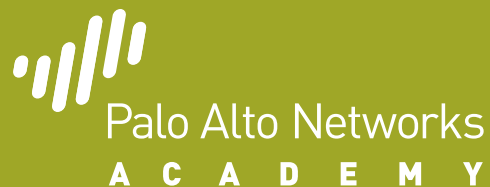


# CYBERSECURITY SURVIVAL GUIDE

## Principles & Best Practices

January 2015

Lawrence C. Miller, CISSP  
Shoba Trivadi — Editor



### Advisory Panel:

Matthew Ancelin, CISSP – Chief Contributor & Reviewer

Judith Backel, CISSP – Reviewer

Jay Mackey, CNSI, CNSE 4.1, CISSP, CEH – Reviewer

# Contents

Foreword	1
Introduction <i>Principles and Best Practices</i>	4
1. <b>Cybersecurity Landscape &amp; Threats</b> <i>Modern Applications</i>	7
2. <b>Cybersecurity Countermeasures</b> <i>Traditional &amp; Next-Generation Countermeasures</i>	50
3. <b>Cybersecurity Best Practices &amp; Principles</b> <i>Enterprise Security Design Elements</i>	85
4. <b>Cybersecurity Solutions from Palo Alto Networks</b> <i>Next-Generation Firewall Technologies</i>	145
Knowledge Check Answers	178
Acknowledgements	183



# Foreword

Welcome to the first book from Palo Alto Networks Academy written specifically for colleges. At Palo Alto Networks, Inc., we believe that the knowledge of enterprise security and cybersecurity will in turn secure our future. This book is designed to be a primer on fundamental enterprise security concepts.

In the context of the modern Cybersecurity landscape, the next-generation enterprise security platform is a pivotal countermeasure to today's advanced threats. 'Next-generation' implies that new methods of processing and securing network traffic are being used to provide visibility and control over traffic, applications, and threats. 'Enterprise security' deals with threat protection for large and complex organizations; while cybersecurity scales the vast landscape of the Internet riddled with vulnerabilities and viruses.

Palo Alto Networks brings a long lineage of industry knowledge to the design of its firewalls and network security products. In reaching students of today and tomorrow, we hope to enhance their knowledge, their skills,

and their understanding of this landscape and thereby spark an interest on their part to architect secure networks and data centers

To help proliferate and continue research in the area of enterprise security, we share this book with the academic world where study takes precedence over product knowledge where young minds are shaped with the overreaching philosophies of designing a better world and where the hope to inspire and change continues to thrive.

We also retain statistics and findings that cover research over a period of time. In the interest of study, we leave them here without feeling the compulsion to update them from time to time. In today's internet era, we hope that current numbers are derived by the students themselves and add to the exhilarating discover of trends and traction at any given point in time.

Palo Alto Networks Academy's hope is that the knowledge compiled in this book will help students 'safely enable' their environment as well as build secure networks. To meet that end, this book includes relevant points from the Dummies books from Palo Alto Networks and Wiley Publishing for quick reference.

We hope that students of the next generation glean useful information on cybersecurity from this book. Surviving cybersecurity threats means exposure to designing and securing security measures using next-generation solutions, which this book also introduces.

*Shoba Trivadi*

*PALO ALTO NETWORKS ACADEMY*

**Palo Alto Networks, Inc.** [www.paloaltonetworks.com](http://www.paloaltonetworks.com)

© 2014 – 2015 Palo Alto Networks-All rights reserved.

Palo Alto Networks, PAN-OS, and Panorama are trademarks of Palo Alto Networks, Inc.

All other trademarks are the property of their respective owners.

# Introduction

## Principles and Best Practices

Today's application and threat landscape has evolved and become more complex and dangerous than ever before. Risks to the enterprise include new and emerging threats, data leakage, and regulatory non-compliance.

Industry trends such as Bring Your Own Device (BYOD), cloud computing, consumerization, software-defined networking/security/storage/etc., virtual data centers and others, further complicate modern network security challenges. As a result, many basic tenets of network security—traditional concepts such as defense-in-depth and perimeter security—must also evolve to address these challenges.



This book is divided into the following parts:

- **Part 1: Cybersecurity Landscape and Threats** concentrates on the Cybersecurity landscape and the imminent threats in this area. It also goes into traditional as well as modern counter measures.
- **Part 2: Cybersecurity Next-Generation Countermeasures** presents traditional and next-generation tools and techniques that provide countermeasures.
- **Part 3: Cybersecurity Best Practices and Principles** concentrates on best practices and principles and introduces an in-depth analysis of a specific design element and security method; the zero-trust model. The hope is that it helps students understand the principles of designing a secure network.
- **Part 4: Cybersecurity Solutions from Palo Alto Networks** ends with specific examples from Palo Alto Networks and presents its next-generation firewall technology focussing on App ID, User ID and Content ID.



# 1. Cybersecurity Landscape & Threats

## Modern Applications

The application landscape is changing rapidly and radically and it is no longer black and white, but rather an infinite spectrum of “gray”. Personal technologies and Enterprise 2.0 applications are increasingly being used for work-related purposes as the boundary between our work and personal lives becomes less distinct.

According to research from McKinsey and Company and the Association for Information and Image Management (AIIM), many companies are recognizing significant benefits from the use of Enterprise 2.0 applications and technologies in their organizations including better collaboration, increased knowledge sharing, and reduced expenses (for example, for travel, operations, and communications).<sup>1</sup>

Thus, the enterprise infrastructure (systems, applications, and networks) is rapidly converging with personal technologies and applications, making it practically impossible to define where the Internet begins and the corporate infrastructure ends.

### **Key Concept**

*Enterprise 2.0 applications are defined by Applipedia as “a system of web-based technologies that provide rapid and agile collaboration, information sharing, emergence and integration capabilities in the extended enterprise.”*

## **Securing an Enterprise Network**

Securing an enterprise network used to be a fairly simple exercise — the corporate network was the “trusted” network, the Internet was an “untrusted” network, and a firewall was deployed at the perimeter between the two networks.

Any traffic that originated from the trusted network to the untrusted network was allowed, while any traffic that originated from the untrusted network was blocked. A few exceptions for specific applications, commonly referred to as “holes,” were defined as rules on the firewall.

1. [The Application Usage and Risk Report: An Analysis of End User Application Trends in the Enterprise, Fall Edition 2009, Palo Alto Networks, Sunnyvale, 2009. \(For more recent reports, search for ‘recent reports on applications.’\)](#)

For example:

- port 80 for web traffic
- port 443 for secure web traffic
- port 25 for e-mail, and ports 20 and 21 for FTP

But as with many things, network security is not so simple today. Classifying applications as good or bad—and consequently allowed or blocked—is difficult because many applications can be used for both good and bad purposes.

Applications have also become increasingly evasive, using techniques such as dynamic port hopping and SSL hiding, to slip past legacy port-based (or packet filtering) and stateful inspection firewalls. **Thus, applications (including Malware) have become the predominant attack vector for cybercriminals and threat developers to infiltrate networks and systems.**

### Key Concept

*An attack vector is defined as the path or means through which vulnerability is exploited.*

### Classifying Applications

***It has also become increasingly difficult to classify applications as either good or bad, in a clear and consistent manner.*** Many applications are clearly good (low risk, high reward) or clearly bad (high risk, low reward), but most are somewhere in between—depending on how the application is being used.

For example, many organizations now use social networking applications, such as Facebook, for important business functions such as recruiting, research and development, marketing, and consumer advocacy. However, these same applications can be used to leak sensitive information or cause damage to an organization's public image—whether inadvertently or maliciously.

### **Key Concept**

*Application—A software application is a program or group of programs designed for end users. Applications can be systems software or applications software.*

### **Personal Devices**

Personal devices—such as smartphones and tablets—have become ubiquitous in businesses and organizations everywhere, as employees are being allowed to use these devices and their associated applications for both personal and work-related purposes. This trend is known as *bring your own device* (“BYOD”).

### **BYOD**

BYOD relieves organizations from the cost of providing equipment to employees, but creates a management challenge due to the vast number and type of devices. The organization merely provides the network and network services, but the employee provides the device to connect to the network.

## Consumerization of IT

A second important trend, known as the *consumerization* of IT, is closely related to the BYOD trend. Consumerization occurs as users find personal applications and technologies that are more powerful or capable, more convenient, less expensive, quicker to install, and easier to use than corporate IT solutions, enabling them to improve personal productivity, handle non-work affairs, and maintain online personas, among other things. Consumerization also relates to a concept known as the “Internet of things” or Internet enablement, which is finding its way into a myriad of devices including refrigerators, televisions, and automobiles.

### Key Concepts

*BYOD—Bring your own device is a term used when personal devices are brought to the workplace. It also refers to BYOP or bring your own phone, BYOT or bring your own technology, BYOP or bring your own PC, etc. Consumerization—This term stands for the encouragement of the consumption of goods and to make products suitable for mass distribution.*

## Web and Enterprise Applications

Web 2.0 or Enterprise 2.0 refers to an evolutionary change in the behavior of web-based applications. Dynamic content, rich multimedia, and interactivity changed the face of the Internet and posed new challenges in securing it. Entirely new platforms, such as Facebook, offer multiple applications delivered through a single presence online to the user.

## Popular Web and Enterprise Applications

Some popular examples of Web 2.0/Enterprise 2.0 applications include:

- Social networks like Facebook
- Unified messaging tools like Skype
- Messaging tools like AOL Instant Messenger (AIM)
- Publishing tools like YouTube
- Productivity tools like Salesforce.com or TurboTax
- Office 365 or Online Microsoft Office Suite
- Blogging tools like Blogger
- Browser-based file sharing tools, such as MegaUpload.com, DropBox, and Google Docs
- Enterprise bookmarking and tagging tools like Cogenz
- RSS tools like NewsGator
- Social bookmarking sites, such as Twitter, Pinterest, and Reddit
- Wikis like Socialtext



### Rapid Adoption of Web and Enterprise Applications

To gain an appreciation for how rapidly Enterprise 2.0 applications have been adopted in enterprises, consider the following (based on an analysis of 347 organizations worldwide):

- Facebook chat overtook Yahoo! IM and AIM in less than 18 months since its inception in April 2008.
- Google Docs increased from 33 percent to 82 percent in the 6 months between March and September 2009.
- Twitter jumped 252 percent in terms of sessions and 775 percent in terms of bandwidth in the 6 months between March and September 2009.<sup>2</sup>

In today's fast-paced internet-driven world, reports and metrics change quickly. Please search under the terms, '*recent adoption of web and enterprise applications*' for current statistics.

### Potential Benefits and Risks

Organizations are often unsure of the potential business benefits—and the inherent risks—of the BYOD and consumerization trends, and therefore either:

- Implicitly allow personal technologies and Enterprise 2.0 applications by simply 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.

2. The Application Usage and Risk Report: An Analysis of End User Application Trends in the Enterprise, Fall Edition 2009, Palo Alto Networks, Sunnyvale, 2009.

## **Adverse Effects of Ineffective Policies**

Whether implicitly allowed (and ignored) or explicitly prohibited (but not enforced), the adverse results of ineffective policies such as these include:

### **Lost Productivity**

Lost productivity as users must either find ways to integrate these unsupported technologies and applications with the enterprise infrastructure (when allowed), or use applications that are unfamiliar to them or less efficient (when prohibited).

### **Disruption of Critical Business Operations**

Potential disruption of critical business operations due to underground or backchannel processes that are used to accomplish specific workflow tasks or to circumvent controls, and are known to only a few users and fully dependent on their use of personal technologies and applications.

### **Exposure to Risks**

Exposure to additional risks for the enterprise due to unknown—and therefore unpatched—vulnerabilities in personal technologies and applications, and a perpetual cat-and-mouse game between employees that circumvent controls (for example, with external proxies, encrypted tunnels, and remote desktop applications) and security teams that manage these risks.

### **Penalties for Non-Compliance**

Penalties for regulatory non-compliance, for example, the U.S. Health Insurance Portability and Accountability Act (HIPAA) in healthcare organizations or the Payment Card Industry's Data Security Standard (PCI DSS) in retail industries.

## Infinite Number of Unknown Technologies

Beyond managing the risks associated with a relatively limited, known set of core applications that are authorized and supported in the enterprise, security managers must now manage the risks associated with a practically infinite number of unknown personal technologies and applications that may be used in the organization.

## Repurposed by Malicious Actors

Applications can also be hijacked and repurposed by malicious actors, such as was done with Skype users in Syria during their recent period of civil unrest. The Flame malware discovered in 2012 had components which would reprogram a portion of the Skype application, enabling a third-party to spy on individuals using what otherwise appeared to be a legitimate Skype install.

## Application Hide-and-Seek

Many Enterprise 2.0 applications are designed to circumvent legacy port-based firewalls so that they can be easily installed and used on any device, anywhere and anytime. This is accomplished by dynamically adjusting how these applications communicate using tactics such as

- **Port hopping**, using random, changing ports and protocols (TCP or UDP) during a session.
- **Use of non-standard ports**, such as running Google Talk over TCP port 80 (HTTP) instead of the standard TCP port 5222 for XMPP (Extensible Messaging and Presence Protocol), or running DNS over a port other than port 53.

- **Tunneling within commonly used services**, such as when a peer-to-peer (P2P) file sharing program like BitTorrent or an instant messenger (IM) client like Pidgin is running over HTTP.
- **Hiding within SSL encryption**, which masks the application traffic, for example, over TCP port 443

### Application Usage and Threat Report

The April 2013 *Application Usage and Threat Report* by Palo Alto Networks found that out of 1,317 unique applications analyzed, 26 percent (334) used SSL. Of those 334 applications, 39 applications hop ports, 35 applications use TCP Port 80, and 9 applications use a range of non-standard ports.

### Applications that Use SSL

Figure 1-1 shows the number and type of applications that use SSL.

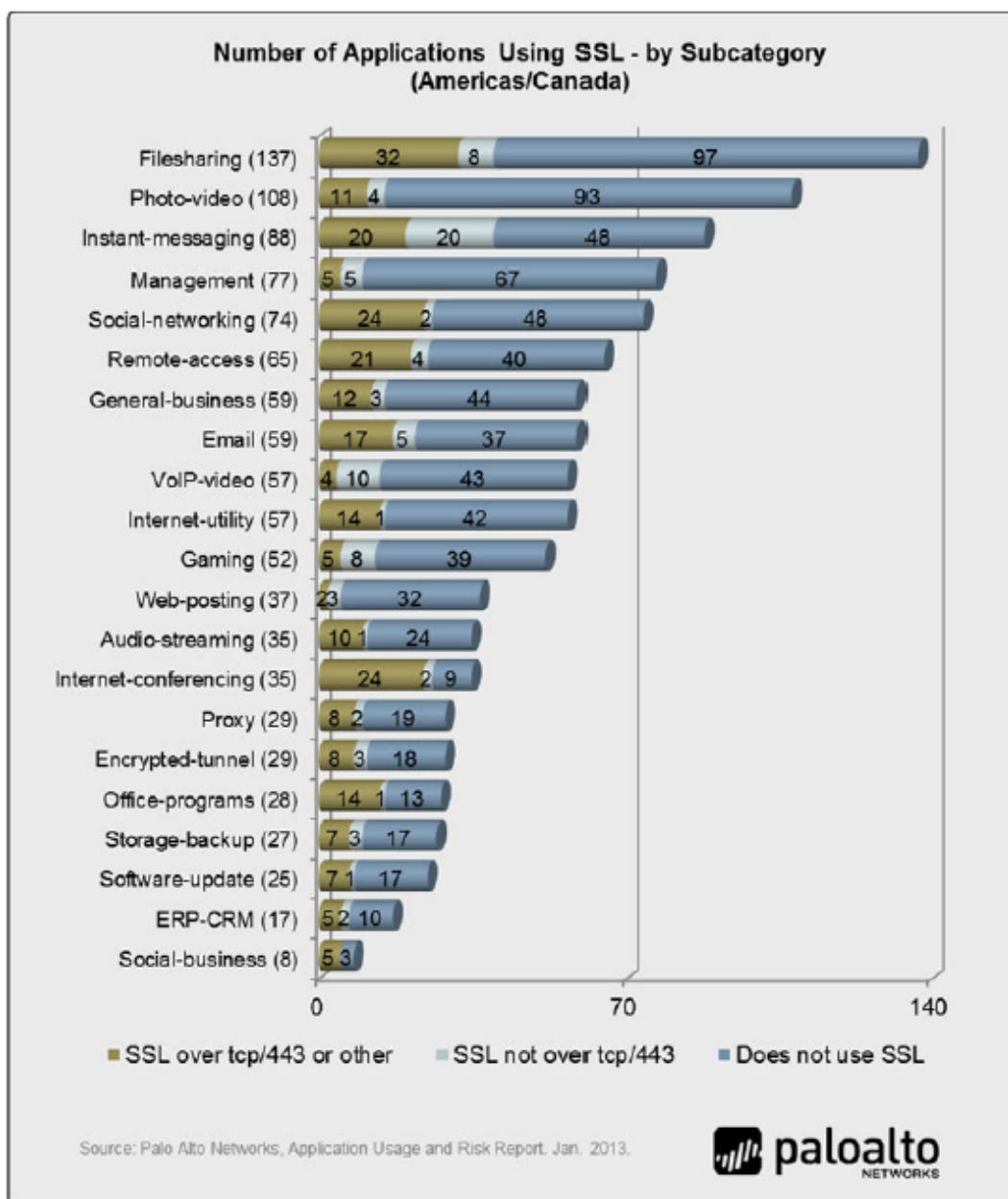


Figure 1-1: Category breakdown of applications capable of using SSL

## Client-Server Applications

Many traditional client-server applications are now designed for Web use and many business applications use these same techniques for ease of operation while minimizing disruptions. For example, RPC and SharePoint use port hopping because it is critical to how the protocol or application (respectively) functions, rather than as a means to evade detection or enhance accessibility.

## HTTP and HTTPS

With the trend toward developing web-enabled and browser-based applications, HTTP and HTTPS have become predominant and now account for approximately two thirds of all enterprise traffic.

Traditional firewalls and other security infrastructure are unable to distinguish whether or not these applications, riding on HTTP and HTTPS, are being used for legitimate business purposes.

### Key Concept

*Remote Procedure Call (RPC) is an inter-process communication (IPC) protocol that enables an application to be run on a different computer or network, rather than the local computer on which it is installed.*

## Cloud-Based Applications

Many enterprises are increasingly using cloud-based applications such as Salesforce.com, WebEx, Office 365, and Google Apps. These applications are typically accessed using a web browser, but then switch to client-server behavior once the application is started. Common applications

that can port-hop include both business and personal use applications such as SharePoint, iTunes, MS RPC, Skype, and TeamViewer (see Figure 1-2).

### Common Applications that can Port Hop

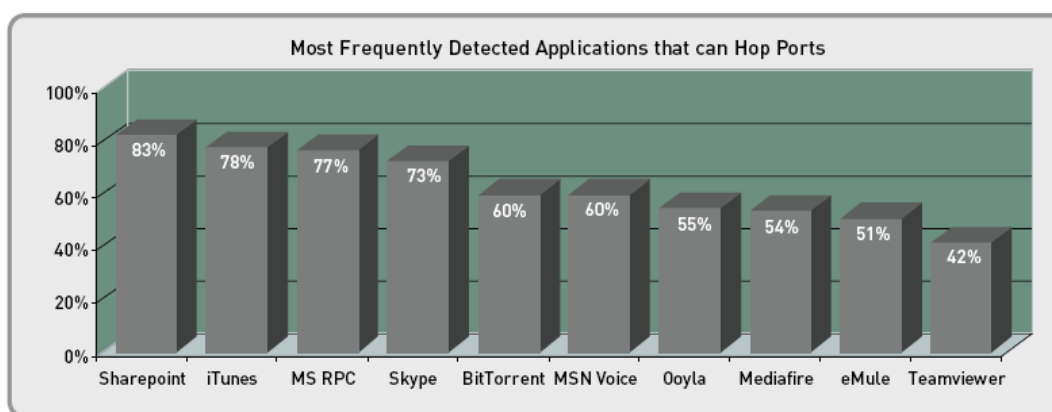


Figure 1-2: Most frequently detected applications across ports

### Threats Are Evolving Too

More ominously, application-layer (Layer 7) attacks are becoming increasingly common.

### Traditional Security Measures

These are built to provide network-layer (Layer 3) protection and thus these application-layer attacks go largely undetected in the enterprise.

### Threat Developers

Threat developers are using many of the same techniques as application developers.

### Port Hopping

Port hopping, use of non-standard ports, tunneling, and SSL hiding can be used to promote ease of use in legitimate consumer applications, as well as to evade network defenses for malicious purposes.

Thus, the vast majority of all new malware and attacks exploit application vulnerabilities rather than network vulnerabilities.

### Disguised Threats

The widespread availability of threat development websites, toolkits, and frameworks makes it relatively easy to create new threats.

- Hackers can quickly modify existing threats so that “known” threats become “unknown” threats to traditional signature-based countermeasures.
- A hacker can modify the code in an existing threat, or add new propagation and exploit mechanisms to create a *blended threat*.

Today’s threats are increasingly sophisticated and designed to run undetected on networks and systems for extended periods of time, collecting sensitive or personal data. Targeted attacks against specific organizations or individuals are also on the rise.

### High-Profile Cases

Some high-profile examples include credit card thefts, targeted attacks in the energy sector, and nation and state-sponsored espionage.



### **Credit Card Theft**

In December 2013, the Target department store chain revealed that they had experienced a major data breach involving over 70 million customer credit card records, making it the largest credit card data breach to date. The breach was found to have occurred through a supply-chain attack, where the network of a heating and air conditioning contractor had remote access into Target's network. The HVAC company's network was the point of entry into Target's network and was reported as the source of stolen user credentials used in the subsequent attack on Target and their PCI network.

### **Energy Sector**

From October of 2012 to May of 2013 a spike in targeted attacks aimed at the US Energy sector were reported to the ICS-CERT (Industrial Control Systems Cyber Emergency Response Team). While none were breached, gas compressor stations across the Midwest and Plains regions experienced brute-force attempts against their process control networks. Energy sector business networks also saw an increase in breach attempts.

### **Nation and State-Sponsored Espionage**

May of 2014, the United States Department of Justice indicted five Chinese nationals on thirty-one counts, including conspiring to commit computer fraud and abuse, Accessing a protected computer with authorization, transmitting a program, information, code or command with the intent to cause damage, economic espionage, trade secret theft, and others. This comes after reports such as APT1 and Putter Panda, which point to a Chinese military unit specifically targeting economic and military targets globally.

## **Customized Attacks**

In a targeted attack, 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.

According to Verizon's 2013 Data Breach Investigations Report, 92 percent of data breaches resulted from external actors, including organized crime, state-affiliated groups, activists, former employees, and other unaffiliated or otherwise unknown hackers.

## **Stealthy Tools**

Hackers have also become more sinister, motivated by greed and the prospect of making lots of money for information theft. As such, hackers now favor fast, stealthy tools to ply their trade.

As soon as a new vulnerability is discovered, hackers are looking for ways to quickly build, deploy, spread, and modify new threats.

This speed increases the likelihood of a successful attack against an enterprise when a zero-day or near zero-day exploit is discovered because traditional security measures—such as anti-virus and intrusion prevention—rely on reactive threat signatures and updates.

Today's hackers are true cybercriminals, often with the backing of rogue nation-states, organized criminal groups, or radical political activists.

As such, these cybercriminals have the necessary resources to launch sustained attacks against large enterprises and governments, have great technical depth and focus, and are well funded.

## **Asset Theft**

Intellectual property, such as proprietary designs, chemical compositions, and other inventions represent the lifeblood of innovation and competitive advantage. In the hands of a cybercriminal, these become very lucrative and rapidly marketable to competitors and other interested parties. Meanwhile, their theft erodes the profitability of the victim companies and threatens the nation-state's economic position.

## **State Sponsored Attempts**

Nation-state sponsored organizations often have military and/or strategic goals such as the ability to disable or destroy critical infrastructure—power grids, water supplies, transportation systems, and industrial systems, or defend against the same. The Center for Strategic and International Studies reported in 2011 that 33 nations include cyberwarfare in their military planning and organization.

## **The Modern Attack Strategy**

The modern attack strategy has evolved into a coordinated, multi-step process (see Figure 1-3).

## The Multi-Step Process in an Attack Strategy

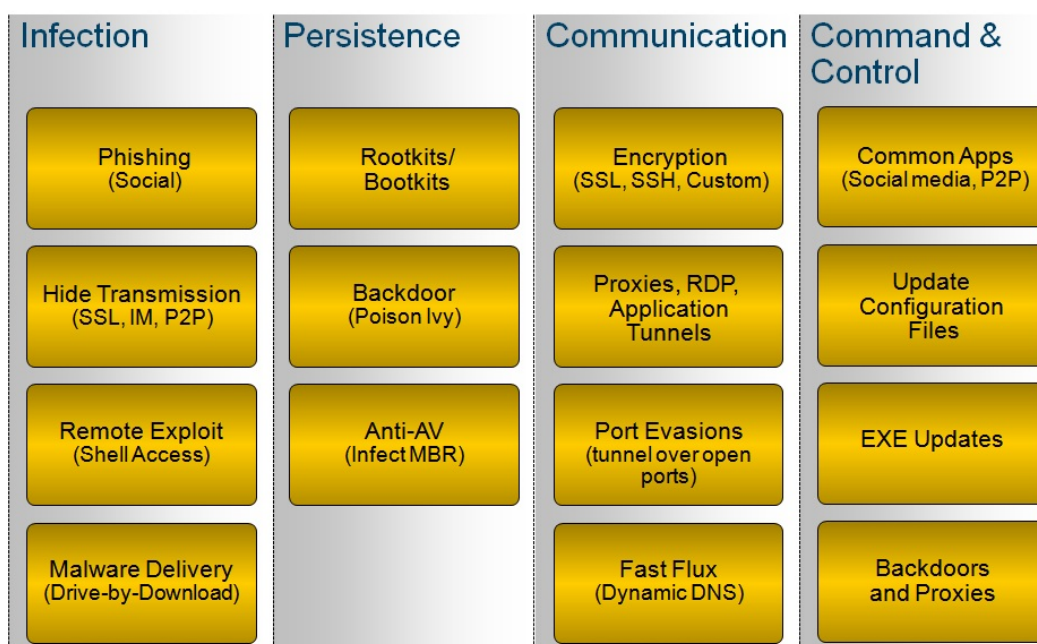


Figure 1-3: The lifecycle of a modern attack.

### The Path of Attack

A typical attack begins with the infection of a target endpoint, for example, an end user within an organization is tricked into clicking on an infected link in a phishing e-mail or in an IM.

- The user's web browser is redirected to a web page that automatically downloads malware to the user's computer in the background—a drive-by download.
- The infected endpoint is now a bot under the control of the attacker and acts as the attacker's entry point into the corporate network.
- From there, the attack is escalated by exploiting other internal assets on the network, and information is slowly and quietly stolen.

## Infection

**The first step in the modern attack lifecycle is *infection*.**

Infection typically has a social engineering aspect to it that requires some human action, such as clicking on a link in an e-mail or running an application on a social networking site. Infections can also occur without humans falling for such tricks, such as a SQL injection which feeds encoded database commands into a web form.

Once a target machine is infected with malware, an exploit can be run to take advantage of a software coding weakness, perhaps causing a buffer overflow or other error state. If exploitation is successful, it may provide shell access to the target machine.

- With shell access, the attacker has control of the system. The attacker is then able to type commands at the C:\ prompt as if the user were typing them.
- Other exploit modules, such as Meterpreter, provide the attacker with full graphical user interface (GUI) access, as if they were sitting at the user's desk logged in, able to click, type, and operate the user's computer.

***The key to infection is stealth and evasion.*** This is often accomplished by hiding traffic from traditional security measures, for example using SSL or some other proprietary encryption used in IM or P2P networks.

## Persistence

**Next, the attacker needs to ensure *persistence* of the bot.**

This is commonly achieved by installing a rootkit, bootkit, backdoor or anti-AV malware on the compromised machine which ensures the attacker can maintain control of the bot and adapt the malware as needed to continue evading detection.

- A *rootkit* is malware that provides privileged (root-level) access to a computer. Rootkits live in the BIOS of a machine, which means operating system level security tools have no visibility to them.
- A *bootkit* is a kernel-mode variant of a rootkit, commonly used to attack computers that are protected by full-disk encryption.
- *Backdoors* allow an attacker to bypass authentication to gain access to a compromised system.
- *Anti-AV malware* disables legitimately installed anti-virus software on the compromised machine, thereby preventing automatic detection and removal of other malware.

Figure 1-4 shows the frequency of various persistence tactics found in malware including creating copies, creating executable in Windows folders, and masquerading as Windows System program files.

## Persistence Tactics Found in Malware

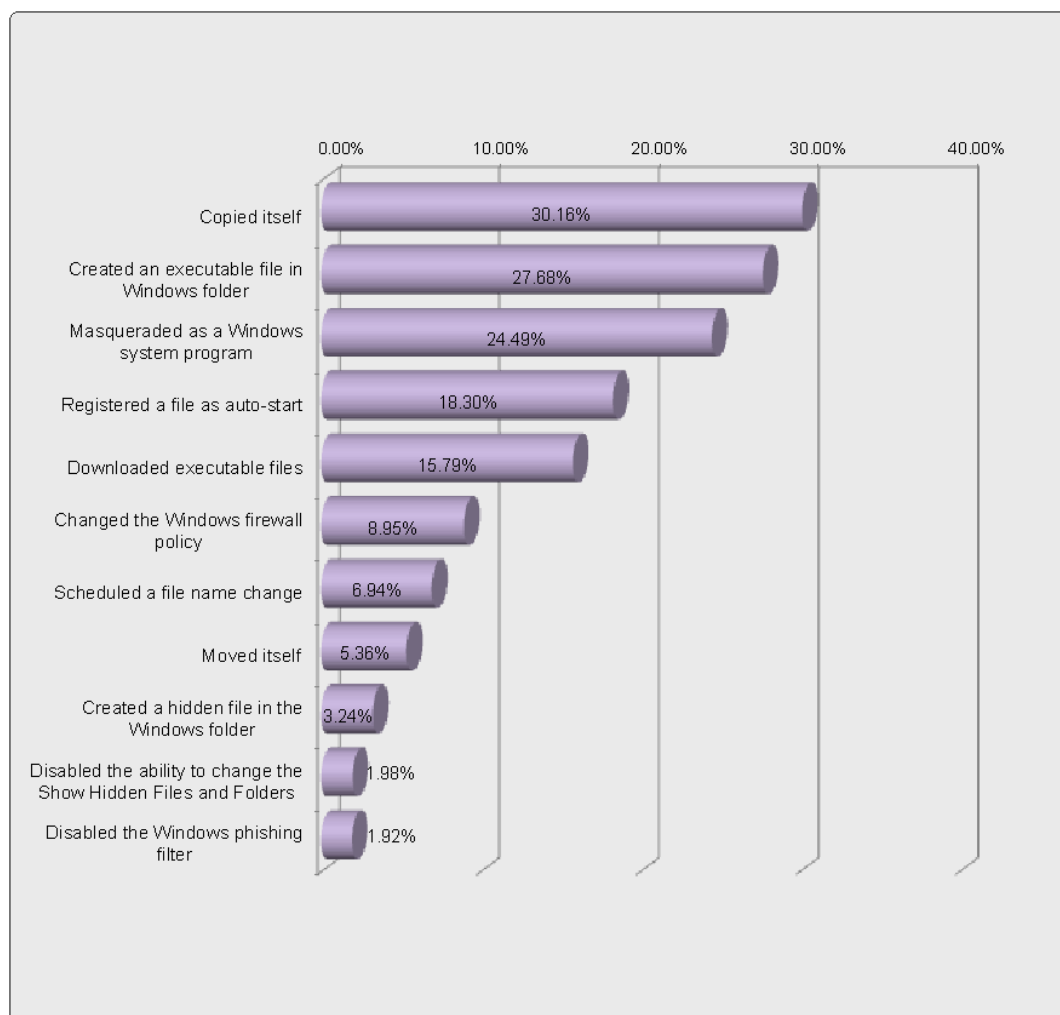


Figure 1-4: Most common persistence tactics found in malware

## Communication

Communication is the lifeblood of an attack. The attacker must be able to communicate with some form of command and control network to be able to steal data from a target system or network.

Attack communications traffic must be stealthy and is usually hidden through:

- **Encryption** with SSL, SSH (Secure Shell), or some other custom or proprietary encryption.
- **Circumvention** via proxies, remote desktop access tools, or tunneling. In some instances, use of cellular networks enables complete circumvention of the target network. PWNplug is a portable penetration testing device that covers the entire OSI stack from the physical to application layers.
- **Port evasion** using network anonymizers or port hopping to traverse over any available open ports.
- **Fast Flux (or Dynamic DNS)** to proxy through multiple infected hosts or multiple ever-changing command and control servers in order to reroute traffic, and make it difficult to determine the true destination or attack source.

## Command and Control

Command and control ensures that the malware or attack is controllable, manageable, and updateable by the attacker. This is the connection to the outside world that modern malware relies on as a means for:

- offloading data
- receiving commands
- initiating attack
- or even for cleanup (self-deletion) operations



Viruses were originally written as self-contained, self-propagating operators which did not require command and control networks. After launch, they performed a sequence of pre-programmed steps to do the virus writer's bidding.

As motives and methods evolved, the need for more flexible, configurable, and communicative tools arose. Modern malware writers found a need to instantly respond to detection, adapt to new exploitation opportunities, and receive plug-and-play updates with additional functionality—all of this *after* the initial infection.

Most importantly, modern malware needed a way to effectively transmit stolen goods out of the victim organizations. Thus, surreptitious communication channels to malicious servers, or networks of servers, became the modus operandi of virtually all theft-oriented malware.

## **The Central Role of Malware**

The rise of botnets and modern malware is reshaping the threat landscape and forcing enterprises to reassess how they protect themselves.

These modern threats have outpaced traditional anti-malware strategies and in the process, have established a foothold within the enterprise that criminals and nation-states can use to steal information and attack sensitive assets.

Attack techniques have also evolved and malware now plays a central role in the hacker's arsenal and in the lifecycle of an attack.

Attackers have developed new methods for delivering malware (such as drive-by-downloads), hiding malware communications (with encryption), and avoiding traditional signature-based detection.

### **Battle with Malware**

Information security professionals have been doing battle with malware for over twenty years. Unfortunately, all of this hard-earned experience does not necessarily mean we are winning the war. This is due in large part to the evolution of malware that either mutates or can be updated to avoid detection by traditional malware signatures.

### **Battle with Bots and Botnets**

Bots (*individual infected machines*) and botnets (*the broader network of bots working together*) play a major role in this evolution of polymorphic malware, and are notoriously difficult for traditional anti-virus/anti-malware solutions to detect. As such it is important to understand what it is that makes botnets so different from previous generations of malware, and how it impacts our efforts to control them.

### **Why a Botnet is called a Botnet?**

A potentially obvious but important distinction is evident in the name “botnet” itself. Literally, a botnet is a network of bots (infected machines).

Unlike earlier types of malware, which were more or less a swarm of independent agents, bots are centrally coordinated and retain a communication channel to the outside world (see Figure 1-5).

## Botnet as a Communication Channel

Botnets are remotely controlled by command and control servers.

### Botnet remotely controlled

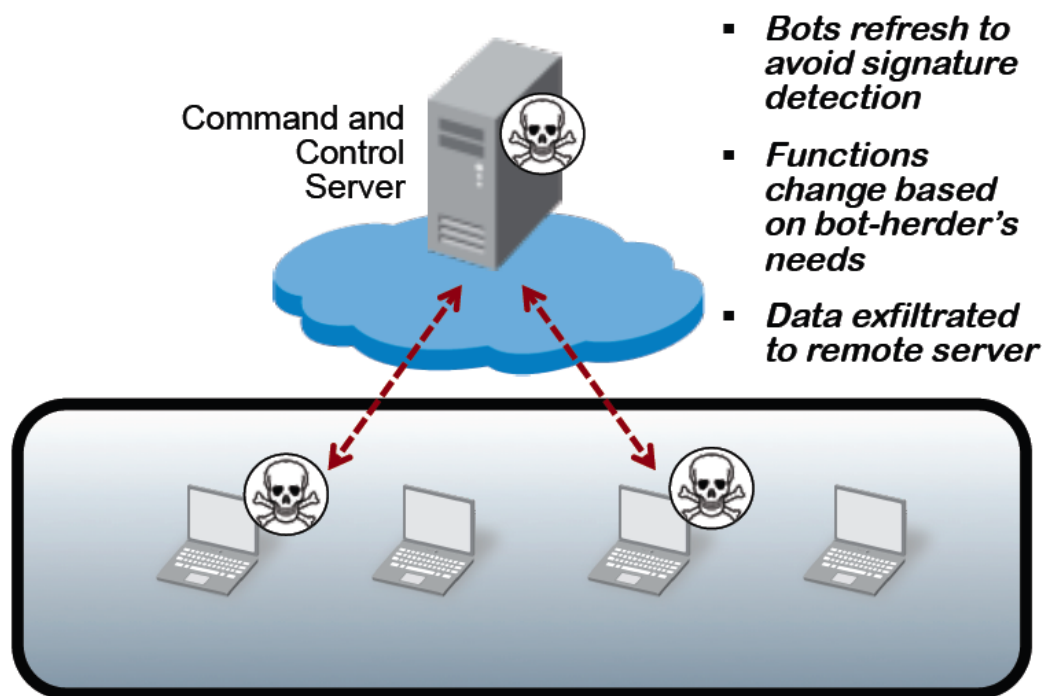


Figure 1-5: A remotely controlled botnet

In much the same way that the Internet changed what was possible in personal computing, botnets are changing what is possible in the world of malware. Now all malware of the same type can work together toward a common goal, with each infected machine growing the power and destructiveness of the overall botnet.

## **Botnets are changing the world of Malware**

Botnets can evolve to pursue new goals or adapt as different security countermeasures are deployed. Some of the most important and unique functional traits of botnets are:

### **Distributed and Fault-Tolerant**

- Botnets are malware that take full advantage of the resiliency built in the Internet itself.
- A botnet can have numerous control servers distributed all over the world, with multiple fallback options.
- Bots can also potentially leverage other infected bots as communication channels, providing them with a near infinite number of communication paths to adapt to changing access options.

### **Multifunctional**

- Updates from the command and control servers can also completely change the bot's functionality. This enables a new economic approach for a botnet owner, who can now use portions of the botnet for a particular task such as collecting credit card numbers, while other segments of the botnet could be sending spam.
- The important point is that the infection is the most important step, because the functionality can always be changed later as needed.

## Persistence and Intelligence

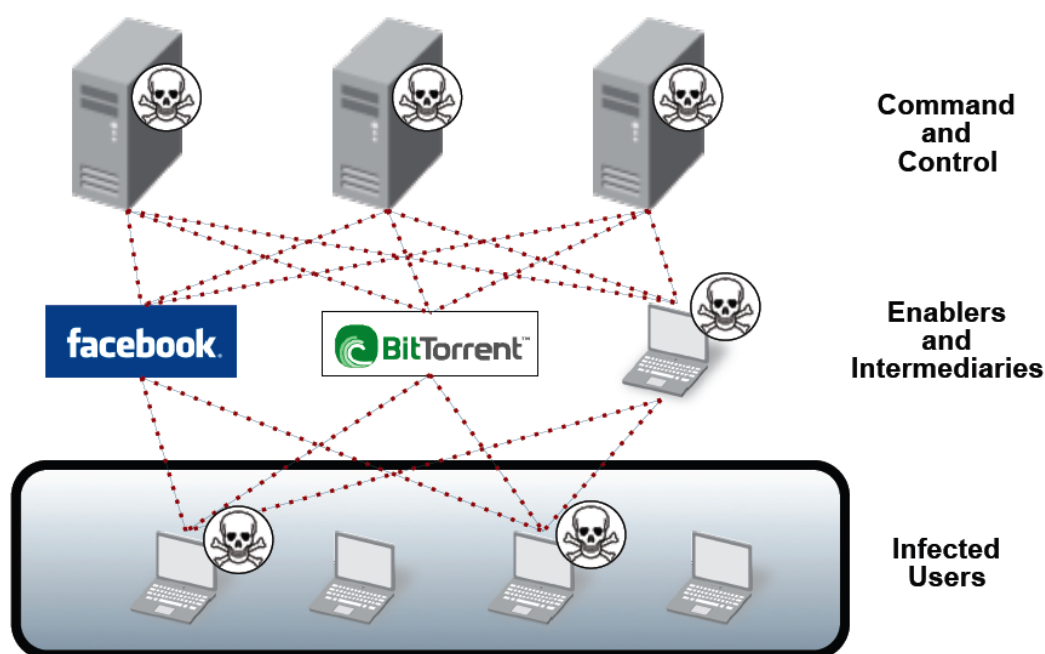
- Given that bots are both hard to detect and can easily change function, they are particularly well-suited for targeted and long-term intrusions into a network.
- Since bots are under the control of a remote human, a botnet is more like having a malicious hacker inside your network as opposed to a malicious executable program, such as a virus.
- For instance, a bot can be used to learn more about the organization of the network, find targets to exploit, and install additional backdoors into the network in case the bot is ever discovered.

## Threats to the Enterprise

Given their flexibility and ability to evade defenses, botnets present an enormous amount of risk to the enterprise. Botnets are virtually unlimited in terms of their functionality, ranging from sending spam to the theft of classified information and trade secrets. The ultimate impact of a botnet is largely left up to the bot-herder—a botnet that was sending spam one day could be stealing credit card information the next.

Intensive investigation is required to map the distributed command and control infrastructure of a botnet.

### Infrastructure of a Botnet



**Figure 1-6: The distributed command and control infrastructure of a botnet**

The following sections go into the classes of botnets: spamming, DDos, financial, and targeted intrusive botnets. This is followed by what the industry is doing about botnets.

#### **Key Term**

*A bot-herder is a person (or persons) that remotely control a botnet through command and control servers.*

## Spamming Botnets

***The largest botnets are often dedicated to sending spam. They are called spamming botnets.***

The premise is fairly straightforward—the bot-herder attempts to infect as many endpoints as possible, which can then be used without the owner’s knowledge to send out spam e-mail messages.

The relative impact of this type of bot on the enterprise may seem low initially.

- An infected user sending spam could consume additional bandwidth and ultimately reduce the productivity of the user and even the network itself.
- The company’s e-mail domain and IP addresses could also easily become listed by various real-time blackhole lists (RBL’s), causing legitimate corporate e-mails to be labeled as spam and blocked by other organizations.

***The Rustock botnet is an example of a spamming botnet.***

### Example of the Rustock Botnet

*The Rustock botnet is an example of a spamming botnet. It was capable of sending up to 25,000 spam e-mail messages per hour from an individual bot and, at its peak, sent an average of 192 spam e-mails per minute per bot. You can just imagine the impact that this infection potentially had on a computer’s or network’s performance! Rustock is estimated to have infected more than 2.4 million computers worldwide.*

## **DDoS Botnets**

***A slight twist on the spamming botnet model is to use bots as part of a distributed denial-of-service attack (DDoS).***

The DDoS botnets attempt to overwhelm a target system with traffic from a large number of endpoints.

In these cases, the enterprise with the infected clients is often not the target of the attack itself.

- Instead the bot-herder is simply using the compromised hosts in the enterprise to flood a remote target with traffic.
- The bot-herder leverages the massive scale of a botnet to generate an amount of traffic that can overwhelm server and network resources at the target.
- These DDoS attacks often target specific companies either for personal/political reasons, or to extort payment from the target in return for halting the DDoS attack.

***DDoS botnets represent a dual risk for the enterprise.***

The enterprise itself can potentially be the target of a DDoS attack resulting in downtime and lost productivity. Even if the enterprise is not the ultimate target, any infected endpoints participating in the attack could consume valuable network resources, while unwittingly facilitating a cybercrime.



### Example of the Skunkx Botnet

*The Skunkx botnet is an example of a DDoS botnet. Skunkx is believed to have originated in the U.S. and was discovered in 2011. It conducted DDoS attacks using UDP, SYN and HTTP floods, and was commonly spread over USB devices, MSN, and YahooMessenger.*

### Financial Botnets

***Financial botnets have had widespread coverage in the press, largely due to the spectacular amounts of damage they have caused in the market.***

Banking botnets, such as ZeuS and SpyEye are responsible for the direct theft of funds from all types of enterprises. These botnets are typically not as large and monolithic as the spamming botnets, which grow as large as possible for a single owner. Instead, banking botnets are often sold as kits allowing large numbers of attackers to license the code and set about building their own botnets and targets.

Even with their smaller size, the impact of these botnets can be enormous:

- ZeuS botnets have repeatedly been able to steal millions of dollars from enterprises in very short periods of time.
- Other financial botnets focus on the theft of credit card information, or faking ACH (Automated Clearing House) bank transfers.

***The impact of a financial breach can be enormous for an enterprise.***

The breach of customer credit card information can lead to serious financial, legal and brand damage for the enterprise. Additionally, if Human Resources, Finance or Accounting groups are compromised, the enterprise can lose money that may potentially never be recovered.

**Example of the Zeus/SpyEye Botnet**

*The Zeus/SpyEye botnet is a financial botnet that has stolen more than \$100 million from small and medium-sized businesses worldwide. It relies on peer-to-peer (P2P) networks and UDP (User Datagram Protocol) communication for command and control, and to gather information from the victim network.*

**Targeted Intrusive Attack Botnet**

Botnets are also a key component of targeted, sophisticated and ongoing attacks.

These types of botnets are very different than their larger counterparts. Instead of attempting to infect large numbers of users to launch actions on a large scale, these smaller botnets aim to compromise specific high-value machines that can be used to further the penetration and surveillance into the target network.

In these cases, an infected machine can be used to gain access to protected systems, and to establish a backdoor into the network in case any part of the intrusion is discovered.

***Targeted and intrusive attack botnets represent one of the most dangerous threats for an enterprise.***

These threats are almost always unknown to anti-virus vendors, and they represent one of the most dangerous threats for an enterprise as they are premeditated and specifically target the highest value information such as research and development, source code, planning data, financial data, and customer lists.

#### **Example of the Aurora Botnet**

*The Aurora botnet is an example of a targeted intrusion that caused extensive damage and resulted in stolen intellectual property. The attacks began in 2009 and targeted various organizations including Adobe Systems, Google, Juniper Networks, and Rackspace.*

### **What is the Industry doing about Botnets?**

The explosion of botnets has certainly not gone unnoticed in the industry, and large companies have begun to team up with law enforcement to take action against some of the largest and most notorious botnets. To understand how this is done, it is important to understand the infrastructure of a botnet. Intensive investigation is required to map the distributed command and control infrastructure of a botnet.

### Infrastructure of a botnet

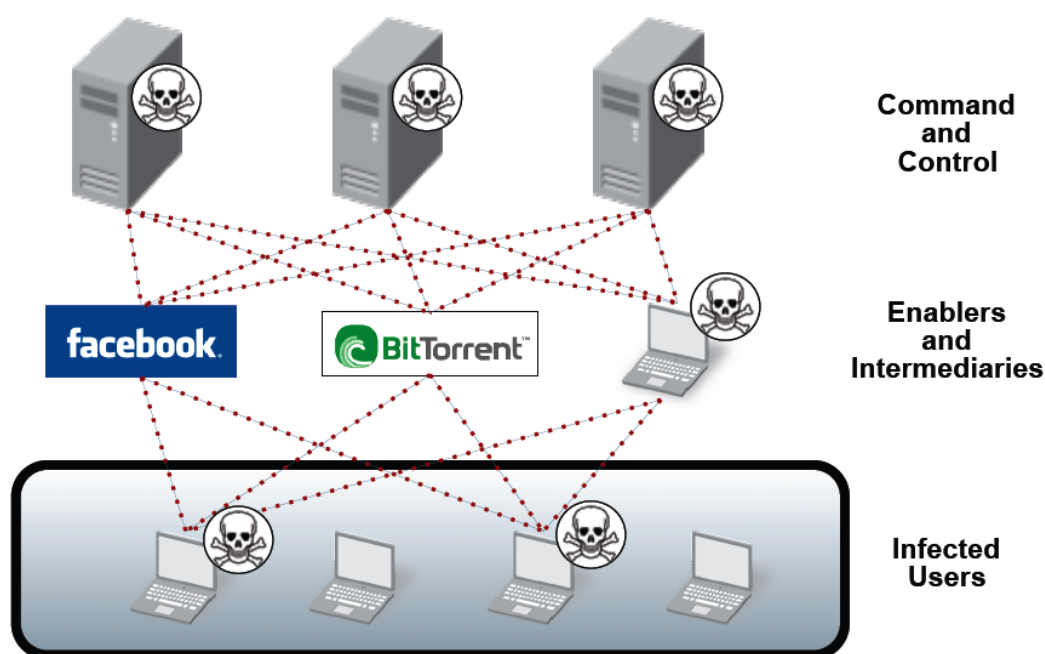


Figure 1-6 (repeated): The distributed command and control infrastructure of a botnet

In short, the goal is to separate the bots (the infected machines) from their brain (the command and control servers).

### Decapitating a Botnet

If the bots cannot get to their servers, they cannot get new instructions, or upload data, or any of the things that make botnets so unique and dangerous. The general approach can be thought of, and is often referred to as “taking down” a botnet, or botnet “decapitation”.

While this strategy may seem straightforward, in reality it requires an enormous amount of investigation, expertise, and coordination between both security researchers and law enforcement.

### **Disabling a Command and Control Server**

Disabling a command and control server often requires both physically seizing the server as well as taking ownership of the domain and/or IP address range associated with the servers.

Very close coordination between technical teams, legal teams, and law enforcement is essential to disabling the botnet (see Figure 1-6).

## **Why are Botnets complex & complicated**

### **Botnets do not rely on a single server**

Making matters worse, a botnet typically does not rely on a single server, but rather has multiple command and control servers for redundancy purposes. Additionally each server is typically insulated by a variety of intermediaries to cloak the true location of the server. These intermediaries include P2P networks, blogs and social networking sites, and even communications that proxy through other infected bots. This means that simply finding the command and control servers is a considerable challenge, requiring time and deep technical investigations.

### **Botnets withstand loss of server**

To make things even more complicated, most botnets are designed to withstand the loss of a command and control server, meaning that researchers must disable ALL of the command and control options at

once. If any of the command and control servers are accessible, or any of the fallback options are successful, the bots will be able to get updates, and can quickly populate a completely new set of servers, and the botnet will quickly recover.

### **Botnets attempt to move to new infrastructure**

It also means that the servers will need to be taken out almost simultaneously. As soon as a botnet owner sees he's under attack, he will immediately attempt to move to a new infrastructure. Thus, even a single server remaining functional for even a small amount of time can give the bot owner the window he needs to update the bots and recover the entire botnet.

### **Recent Win against a Rustock Botnet**

*For all the challenges of this approach, there have nevertheless been some very consistent and high-profile wins in the industry. In March 2011, Microsoft working in concert with industry leaders and the FBI, was able to successfully take down the Rustock botnet, which had operated for more than 5 years and at the time was responsible for sending up to 60% of the world's spam.*

### **Top-down model does not protect Enterprise**

The top-down model described above offers a credible industry response to some of the largest and most notorious botnets. However, it's important as security professionals to understand that these efforts, while good on a global scale, do very little to mitigate the threat that botnets pose to an individual enterprise.

Limitations are largely due to:

- Limited Resources
- Time Intensive Schedules
- Being Impractical to Attack

**LIMITED RESOURCES:** *The most obvious limitation is that the top-down model is incredibly resource intensive, both in terms of time and effort.*

As such, only the largest and most notorious botnets are targeted, typically spamming botnets. While any bot on a network is a threat, spamming botnets do not target a particular enterprise—they are simply looking for more hosts to infect so that they can send more spam, rather than steal sensitive or valuable information.

**TIME INTENSIVE SCHEDULES:** *Secondly, the top-down model is largely reactive, sometimes taking years.*

Enterprise security needs are far more immediate and must ensure that an intrusion or exploit doesn't succeed in the first place. In a very real sense, relying on the industry to disable a botnet is analogous to waiting for government to enact a law when someone is breaking into your business right now.

When an attacker wants to infiltrate an enterprise network and steal information, the botnet is almost always far more customized and difficult to detect. These smaller botnets may be completely unknown and are thus unlikely to warrant enough attention to be targeted for removal by the industry.

**BEING IMPRACTICAL TO ATTACK: *Some botnets will simply be impractical to attack at the command and control level.***

This is based on how key botnet components are distributed around the world. Recall that one of the major challenges for researchers is the requirement to find and take control of all the command and control servers in a short window of time. The more distributed the botnet is, the more difficult it will be to decapitate.

**Conclusions Derived:**

**Rustock**

In the case of Rustock, authorities were somewhat fortunate that almost all of the command and control servers were located within the United States. This allowed federal law enforcement and court rulings to closely coordinate the all-important process of disabling all of the servers at once.

**Problems locating botnets**

Many botnets have servers all over the world, and will specifically function in areas that have very little law enforcement for Internet crimes. This model directly mirrors the Internet itself, which from the beginning was designed to withstand the loss of any one site.

***In short, a distributed network is designed to withstand a decapitation attempt, so it will be very difficult to extend this model to botnets in general.***



### **Responsibility lies in the hands of the Enterprise!**

So while progress has been made fighting botnets at a global level, the simple truth is that the wins are more of an exception to the rule, and will do little to protect enterprises from the threats posed by botnets.

***This puts the responsibility for protecting the enterprise from botnets squarely on the shoulders of the enterprise itself.***

---

## Summary

- It is increasingly complex to distinguish between good and bad applications on a network.
- Malicious and legitimate applications share many of the same techniques and tools.
- Threat actors have become much more organized and their motives have shifted from notoriety to profit and national competitive advantage.
- The modern threat life cycle describes phased attacks which start with initial infection, develop persistence, establish communication, and then work with command and control networks to accomplish their missions.
- Botnets are distributed, fault-tolerant networks of compromised machines, which can be coordinated, and re-purposed, to carry out various malicious tasks.
- Corporations, law enforcement, and government entities have had only limited success in dealing with botnets and malware.

## Discussion

The following are some unique scenarios not covered in this book as yet but presented here for your exploration and research:

1. In your work or school computing environments, can you tell the difference between applications which are hosted onsite, versus cloud based applications? What are some examples of critical applications that you use regularly which are cloud based? Locally hosted?
  2. In your work or school computing environments, have you encountered 'blocked sites' or restricted network resources? What was your reaction? Talking with co-workers or fellow students, are methods of circumventing these restrictions ever discussed?
  3. Have you been part of a data breach or had your identity or financial information stolen? If so describe the experience. Could you personally have done anything better to avoid loss?
  4. Persistence is a defining element of an 'Advanced Persistent Threat (APT)'. Using recent cyber-attacks in the news, contrast between truly persistent attacks and non-persistent smash-and-grab attacks.
  5. This chapter suggests that the current top-down approach to IT security risk management is broken. What would it take to change that paradigm, or restore effectiveness to IT security?
-

## Knowledge Check

1. Which of the following is not an example of a typical 'hole' opened on a traditional firewall?  
**a) Port 194: Internet Relay Chat (IRC) \_\_\_\_\_**  
**b) Port 443: Secure Socket Layer (SSL) \_\_\_\_\_**  
**c) Port 25: Simple Mail Transfer Protocol (SMTP) \_\_\_\_\_**  
**d) Port 53: Domain Name Service (DNS) \_\_\_\_\_**
2. Name two techniques that can be used to evade port-based stateful inspection firewalls?
3. Name two reasons that Enterprise 2.0 applications are a benefit to organizations, and name three examples of enterprise 2.0 applications?  
\_\_\_\_\_
4. True or False: The blurring of personal and corporate devices, as well as personal and corporate applications, is making it easier to secure a network. \_\_\_\_\_
5. True or False: BYOD refers to employees using their own personal devices to attach with corporate network resources. \_\_\_\_\_
6. Name four risks incurred when BYOD policy is ineffective. \_\_\_\_\_
7. True or False: it is more common for new malware to exploit network vulnerabilities than application vulnerabilities. \_\_\_\_\_
8. True or False: A blended threat makes use of known existing malicious code, modified, such that it represents a new 'unknown' threat. \_\_\_\_\_
9. Name three characteristics of a botnet. \_\_\_\_\_
10. What single requirement do all bots and botnets have in common, making them vulnerable to take-down? \_\_\_\_\_

ANSWERS ARE AT THE END OF THIS BOOK.

## 2. Cybersecurity Countermeasures

### Traditional & Next-Generation Countermeasures

This module outlines the various classes of security countermeasures which exist today and contrasts those with next-generation security techniques. While this text is primarily focused on network security concepts, it is important to be aware of its relation to system-level security tools as well.

It addresses the following main topics:

- Traditional Endpoint Security
- Next-Generation Endpoint Security
- Traditional Network Security
- Next-Generation Network Security

## Traditional Endpoint Security

Systems security includes tools that reside on endpoints, such as desktop computers, laptops, servers, virtual servers, industrial control systems, mobile devices, automated teller machines (ATMs), or any other independent computing devices or instances.

This section discusses the following:

- Anti-Virus/Anti-Spyware
- Signature-based Anti-Virus Software
- Zero-Day Exploit Averages
- Web-based Zero Day Attacks

### Anti-Virus/Anti-Spyware

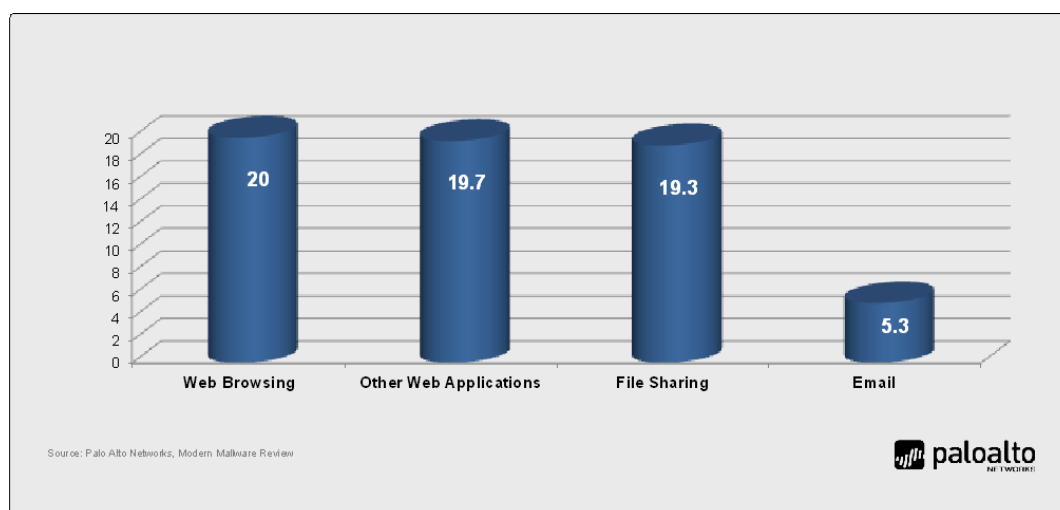
Perhaps one of the longest standing countermeasures, anti-virus software examines files on a system and attempts to match them against a known-bad file, or signature. Installing anti-virus software involves installing an engine, which often has kernel-level access to a system's resources. Signatures must be regularly downloaded and updated. Downloading and processing signature files in this manner can cause noticeable performance reductions on the networks and systems they are running on.

### Signature-based Anti-Virus Software

By its nature, signature-based anti-virus software is a reactive measure that leaves enterprises unprotected against new threats until a new threat is discovered, identified, analyzed, and a signature file created and distributed. The notorious *zero-day attack* is a term that refers to malware,

which has been written and released, but is unknown as of yet to the security industry. However, the term “zero-day” is misleading. Figure 2-1, based on research by Palo Alto Networks, shows that the number of days in a “zero-day” exploit averages from 5 to 20 days depending on the application vector.

### Zero-Day Exploit Averages



**Figure 2-1: Average time to anti-virus detection by application vector**

### Web-based Zero-Day Attacks

Interestingly, web-based zero-day attacks remain unidentified “in the wild” up to four times longer, on average, than e-mail based threats. This is due to a number of factors including user awareness of e-mail borne threats, availability and use of e-mail security solutions, such as anti-spam and anti-virus, and preferred use of the web as a threat vector by malware developers. Viruses traditionally are all-in-one actors that carry all of the needed code to accomplish their purpose upon delivery. Viruses generally



reproduce themselves as a method of spreading. The inflexibility and lack of extensibility of viruses has cause them to be superseded by more sophisticated malware, including Trojans, rootkits, and other exploits.

## **Protection against Viruses and Malware**

Some commonly known ways of protection is to use the following software and firewalls in this section:

- Anit-Spyware Software
- Host Intrusion Prevention Systems
- Personal Firewalls
- Network-Based Firewalls

### **Anti-Spyware Software**

Anti-spyware software is very similar to anti-virus software, in that it uses signatures to look for other forms of malware beyond viruses, such as adware, malicious web application components, and other malicious tools, which share user behaviors without their permission.

### **Host Intrusion Prevention Systems**

Host Intrusion Prevention Systems (HIPS) introduce heuristics, or behavioral anomaly detection, into the process of detecting malware on endpoint devices and machines. Systems are initially base-lined to get a snapshot of what represents 'normal behavior' and then watched for behaviors that vary from that baseline. While it is a cool concept, HIPS has not been widely adopted because of the extreme complexity in deploying it effectively, the need for lengthy 'good behavior' baselines, and its

propensity to lock down machines (and lock-out users) due to false positives.

### **Personal Firewalls**

Personal (or host-based) firewalls are commonly installed and configured on endpoints, such as laptop and desktop computers. Personal firewalls typically operate as application-layer (Layer 7) firewalls that allow or block traffic to an endpoint device based on an individual (or group) security policy. Personal firewalls are particularly helpful on laptops used by remote or traveling user that connect their laptop computers directly to the Internet, for example, over a public WiFi connection. Windows Firewall is an example of a personal firewall that is installed as part of the Windows 7/8 desktop operating system. A personal firewall only protects the endpoint device that it is installed on, but provides an extra layer of protection inside the network.

### **Network-Based Firewalls**

Network-based firewalls typically protect a corporate (trusted) network against threats from an untrusted network, such as the Internet.

However, most traditional network-based firewalls do little to protect endpoints inside the trusted network from threats that originate within the trusted network, such as another device that has been compromised by malware and is propagating throughout the network.

A personal firewall can provide this extra layer of protection. Additionally, a personal firewall can control outbound traffic from the endpoint to help prevent the spread of malware from that endpoint. However, it should

be noted that disabling or otherwise bypassing a personal firewall is a common and basic objective in most malware today.

## **Next-Generation Endpoint Security**

In the system security world, a variety of new methods are being brought to bear against malware. These tools acknowledge many of the weaknesses of traditional anti-virus/anti-spyware and HIPS, and attempt to overcome these weaknesses using new methods of malware discovery, remediation, and removal.

### **Application Whitelisting**

Anti-virus operates by assuming all files are good, unless they match a known-bad. Whitelisting assumes that *'only certain identified files and processes are good, while all others are assumed to be bad.'*

This section details:

- The way Whitelisting works
- When Whitelisting works and when it does not

### **The way Whitelisting Works**

By identifying what exact applications and processes are allowed to run, and cataloging those, a whitelisting approach would then lock out any other files or processes from running at all. In a perfect world, or on a perfect system, this approach is quite effective. Issues arise though when:

- contending with frequent software updates and changes,
- constantly changing application states,

- and the need for users to have the rights to install software for given situations without helpdesk intervention.

### When Whitelisting works and when it does not

Whitelisting works well for single-purpose machines, such as ATMs, which do not get updated very often, and execute the exact same code day in and day out. Single-purpose servers can also use whitelisting effectively.

However, using whitelisting with a user-driven laptop or desktop computer can be very challenging. ***Unfortunately, it is these user-driven platforms that are most often attacked.***

### Exploit Prevention

Signature based anti-malware solutions have the inherent flaw of having to keep up with new signatures for emerging threats, while malware writers outpace them with new versions.

- Using signatures to detect threats is analogous to playing ‘whack-a-mole’, a game where one target is identified and hit down, while 2 new targets pop up. A new paradigm has emerged, which does not attempt to look for the payload and compare it to the ever-growing database of malware signatures.

Exploit Prevention does not come to identify anything, rather *‘it is designed to obstruct and block the attack before it can deliver any malware.’*

### New malicious files released consistently

Each year hundreds of thousands of new malicious files are released on to their targets. All of these malicious payloads rely on only a handful of techniques in order to successfully deliver that payload. To contrast, new techniques are developed at pace of one, or maybe two, per year.

### Stopping the attack before it is begun

Next-generation exploit prevention at the endpoint uses an agent to kill the process that would enable this exploitation technique, before it can open. Thus the attack is stopped before it can begin.

## Mobile Device Management

Mobile device management (MDM) software provides next-generation endpoint security for mobile devices such as smartphones and tablets. MDM provides centralized management capabilities for mobile devices such as:

- **Data loss prevention (DLP)**–Restrict what type of data can be stored on or transmitted from the device.
- **Policy enforcement**–Require passcodes, enable encryption, lockdown security settings, etc.
- **Malware protection**–Scan mobile devices for known malware and prevent “jailbreaking” or “rooting” on mobile devices
- **Software distribution**–remotely install software, including patches and updates over-the-air
- **Remote erase/wipe**–securely and remotely delete the complete contents of a lost or stolen device

### **Key terms**

*Jailbreaking refers to hacking an Apple iOS device to gain root-level access to the device. This is sometimes done by end users to allow them to download and install mobile applications without paying for them, from sources that are not sanctioned and/or controlled by Apple as their AppStore is. Jailbreaking bypasses the security features of the device by replacing the firmware's operating system with a similar, albeit counterfeit version. Doing so makes it vulnerable to malware exploits. Rooting is the equivalent term for a Google Android device.*

## **Traditional Network Security**

Traditional network security comprises of:

- Firewalls
- Intrusion Detection and Prevention Systems
- Web Content Filters, URL Filters, and Web Proxies
- Virtual Private Networks
- Data Loss (or Leakage) Prevention
- Unified Threat Management

### **Firewalls**

Firewalls have been a cornerstone of network security since the early days of the Internet. A firewall is a hardware and/or software platform that controls the flow of traffic between a trusted network (such as a corporate LAN) and a non-trusted network (such as the Internet).

### Packet filtering firewalls

In 1988, Digital Equipment Corporation (DEC) built the first packet filtering firewall (also known as a port-based firewall).

- These first-generation firewalls inspect individual packets to determine whether certain traffic should be allowed between the trusted and untrusted network segments or blocked. This determination is made by matching information contained within each packet—such as source and destination IP address, protocol (TCP, UDP, ICMP) and port number—with a corresponding rule on the firewall that designates whether the packet should be allowed, blocked, or dropped.
- A packet filtering firewall typically operates at the first three layers of the OSI model (Physical, Data Link, and Network).

A variation of the packet-filtering firewall is the dynamic packet-filtering firewall, which uses context-based access control (CBAC) or reflexive access control lists (ACLs) to create and delete firewall rules for individual connections, as needed.

### Stateful inspection firewalls

Second-generation firewalls are known as circuit-level gateways. A stateful inspection firewall is the most common implementation of a circuit-level gateway. Stateful packet inspection was invented in the early 1990's by a research team at Checkpoint Software.

Stateful packet inspection has the following characteristics:

- A stateful inspection firewall maintains state information about the different communication sessions that have been established between hosts on the trusted and untrusted networks.
- Once a permitted connection is established between two hosts, the firewall creates a tunnel for the session that allows traffic to flow between the two hosts without further inspection.
- This type of firewall is very fast, but is highly dependent on the trustworthiness of the two hosts because individual packets aren't inspected after the connection is established.

### **Proxy servers**

Application-layer gateways, most commonly in the form of proxy servers, represent the third-generation of network firewalls.

- Proxy servers do not permit direct communication between hosts. Instead, requests are sent from the originating host to the proxy server, which analyzes the contents of the data packets and, if permitted, sends a copy of the original data packets to the destination host.
- Proxy servers are considered very secure because they perform application-level (Layer 7) inspection of the data packets, can be used to implement strong user authentication, and mask the internal network from untrusted networks. However, proxy servers have a negative impact on overall performance of the network.



## Intrusion Detection and Prevention Systems

While traditional firewalls examine Layer 3 header information, intrusion detection systems (IDS) and intrusion prevention systems (IPS) provide real-time monitoring of network traffic and introduce the concept of deep-packet inspection and analysis of network activity and data.

### Traditional IDS and IPS

Traditional IDS and IPS is similar to anti-virus software in that it attempts to match known-bad, or malicious, patterns found deep within packets. An IDS/IPS is typically deployed to detect and block exploits of software vulnerabilities on target networks.

The primary difference between IDS and IPS is that IDS is considered a passive system, whereas IPS is an active system.

A passive IDS monitors and analyzes network activity and alerts the security administrator to potential attacks and vulnerabilities on the network, but it doesn't typically perform any preventive action to stop an attack.

An IPS, on the other hand, performs all of the same functions as an IDS but also automatically blocks suspected attacks on the network in real-time. However, IPS has some disadvantages that include:

- Must be placed inline along a network boundary and is thus directly susceptible to attack itself
- False alarms must be properly identified and filtered to avoid inadvertently blocking authorized users and applications
- May be used to affect a Denial of Service (DoS) attack by flooding the IPS, causing it to block connections until no connection or bandwidth is available

## Classifications of IDS and IPS

IDS and IPS can also be classified as knowledge-based (or signature-based) or behavior-based (or statistical anomaly-based) systems.

- A knowledge-based system uses a database of known vulnerabilities and attack profiles to identify intrusion attempts. These types of systems have lower false-alarm rates than behavior-based systems, but must be continuously updated with new attack signature to be effective.
- A behavior-based system uses a baseline of normal network activity to identify unusual patterns or levels of network activity that may be indicative of an intrusion attempt. These types of systems are more adaptive than knowledge-based systems and may therefore be more effective in detecting previously unknown vulnerabilities and attacks, but they have a much higher false positive rate than knowledge-based systems.

## Web Content Filters, URL Filters, and Web Proxies

Web content filters, URL filters, and web proxies attempt to restrict the Internet activity of users on a corporate network. These security solutions attempt to classify web sites based on broad categories that are either allowed or blocked for various groups of users on the network.

For example, the marketing and HR departments may have access to social media sites such as Facebook and LinkedIn for legitimate online marketing and recruiting activities, while other users are blocked.

## Website Categories

Examples of website categories include:

- Gambling and online gaming
- Hacking
- Hate crimes and violence
- Pornographic
- Social media
- Web-based e-mail

### Impact of Such Sites

In addition to lowering productivity, these sites may be prime targets for malware that users may unwittingly fall victim to, via drive-by-downloads. Certain sites may also create liabilities in the form of sexual harassment or racial discrimination suits for companies that fail to protect other employees from being exposed to pornographic or hate-based websites.

### How Organizations Employ Such Sites

Organizations may elect to implement these solutions in a variety of modes to either block content, warn users before accessing restricted sites, or logging all activity. The disadvantage of blocking content is that false positives require the user to contact a security administrator to allow access to websites that have been improperly classified and blocked, or need to be accessed for a legitimate business purpose.

URL filtering works on a similar concept to that of anti-virus signatures except that the uniform resource locator (URL) or web address is matched against a database of websites. These databases and categories

are typically maintained by the individual security vendors that provide these types of security solutions.

### **Virtual Private Networks**

Virtual private network (VPN) software is often installed on mobile endpoints, such as laptop computers, tablets, and smartphones, to extend a corporate network beyond the physical boundaries of the organization.

A VPN creates a secure, encrypted connection (or tunnel) across the Internet back to the corporate network. Once connected via a VPN, a remote user can access network resources, such as file servers, printers, and voice-over-IP (VoIP) phones, just the same as if they were physically located in the office.

VPNs secure against intelligible data being intercepted while in transit, but provide no security for what might go into, and then come out of, a secure tunnel. While there are still limited use cases for older VPN protocols such as PPTP (point-to-point tunneling protocol) and L2TP (Layer 2 tunneling protocol), the two most common types of VPNs used in enterprise networking are IPsec and SSL.

#### **Key terms**

*Internet Protocol Security (IPsec) is a secure communications protocol that authenticates and encrypts IP packets in a communication session. Secure Sockets Layer (SSL) is an asymmetric encryption protocol used to secure communication sessions. SSL has been superseded by Transport Layer Security (TLS), although SSL is still the more commonly used terminology.*

## IPsec VPN

An IPsec VPN requires compatible VPN client software to be installed on the endpoint device. A group password or key is required for configuration.

Client-to-server IPsec VPNs typically require user action to initiate the connection, such as launching the client software and logging in using a user name and password.

An IPsec VPN can be configured to force all of the user's Internet traffic back through the corporate firewall, providing optimal protection through enterprise-grade tools residing at the enterprise datacenter, but with some performance loss.

Alternatively, split tunneling can be configured to allow Internet traffic from the device to go directly to the Internet, while other specific types of enterprise traffic route through the IPsec tunnel, for acceptable protection with much less performance degradation. In split-tunneling cases, a personal firewall (discussed in the previous section) should be configured and active, as this configuration can create a 'side-door' into an enterprise network.

*NOTE: A savvy hacker can essentially bridge themselves over the internet, through the client machine, and into the enterprise network through the secure tunnel.*

## SSL VPN

An SSL VPN can be deployed as an agent-based or 'agentless' browser-based connection.

An agentless SSL VPN only requires users to launch a web browser, open a VPN portal or web page using the HTTPS protocol, and log into the corporate network with their user credentials.

A dissolvable client is used within the browser session, which persists only as long as the connection is active, and removes itself when the connection is closed.

This type of VPN can be particularly useful for remote users that are connecting from a device they do not own or control, such as a hotel kiosk, where full client VPN software cannot be installed.

**NOTE:** *SSL VPN technology has become the de facto standard and preferred method of connecting client/user machines back to the enterprise remotely, while IPsec VPN maintains its hold on the majority of site-to-site, or machine-to-machine VPN connections, such as connecting an entire branch office network to a headquarters location network.*

## Data Loss Prevention

Network Data Loss Prevention (DLP) or Leakage solutions inspect data that is leaving or egressing a network, for example via e-mail, file transfer, Internet uploads, or copying to a USB thumb drive, and prevent certain sensitive data—based on defined policies—from leaving the network.

Examples of sensitive data may include:

- Personal information such as names, addresses, birthdates, and social security numbers
- Personal health records

- Financial data such as bank account numbers and credit card numbers
- Intellectual property and other confidential or proprietary company information

A DLP security solution prevents sensitive data from being transmitted outside the network by a user, either inadvertently or maliciously.

A robust DLP solution can detect the presence of certain data patterns even if the data is encrypted. However, these solutions introduce a potential new vulnerability in the network as they have visibility into—and the ability to decrypt—all data on the network.

Other methods rely on decryption happening elsewhere, such as on a Web Security appliance or other man-in-the-middle decryption engine. Often times DLP solutions require many moving parts to effectively route traffic to and from inspection engines, which can add to the complexity of troubleshooting network issues.

### **Unified Threat Management**

Unified Threat Management (UTM) devices are all-inclusive security solutions. They combine the security functionality of:

- a (stateful inspection) firewall
- IDS, anti-virus
- anti-spam
- VPN, content filtering
- DLP into a single appliance

UTM devices do not necessarily perform any of these security functions better than their standalone counterparts:

- In some cases lack the rich feature sets to make them more affordable.
- Since all functions in a UTM pull from the same set of processor and memory resources, enabling all the functions of a UTM can result in up to a 97 percent drop in throughput/performance as compared to top-end throughput without security features enabled.
- Despite these functions all residing on the same platform, the individual engines operate in silos with little or no integration or cooperation between them.

For all of their shortcomings, UTM devices serve a purpose in the small-to-medium enterprise as a convenient and inexpensive solution giving an organization an all-in-one security device.

### **Next-Generation Network Security**

The next-generation firewall is well defined by Gartner as something new and enterprise-focused *“incorporating full-stack inspection to support intrusion prevention, application-level inspection and granular policy control.”*

Most network security vendors are now offering some level of application awareness and control by either adding application signatures to their IPS engine, or offering you an add-on license for an application control module. In either case, these options are “bolt-on” features to a port-based firewall, and do little to help you focus on the fundamental tasks a firewall is designed to execute.



How effectively a business operates is heavily dependent upon the applications its employees use and the content that the applications themselves carry. Merely allowing some, then blocking others, may inhibit the business.

### Evaluating Next-Generation Firewalls

When evaluating next-generation firewalls, consider the following:

- Will the next-generation firewall increase visibility and understanding of the application traffic on your network?
- Will administrator be required to define exactly which applications the next-generation firewall should look for, or will the next-generation firewall provide the administrator with visibility over all applications that are flowing by default?
- Will the traffic control policy response options be broader than just allow or deny?
- Will your network be protected from threats and cyberattacks—both known and unknown?
- Can you systematically identify and manage unknown traffic?
- Can you implement the desired security policies without compromising on performance?
- Will the administrative efforts your team devotes to firewall management be reduced?
- Will your job of managing risk be easier and more effective?
- Can the policies you enable help contribute to the business bottom line?

## Characteristics of Next-Generation Firewalls

Next-generation network security is defined by several characteristics including:

- A departure from reliance on mere port and protocol for allow/block decisions, replaced by allow/block based on applications, users, and content.
- A truly integrated set of security tools, on a single platform, where individual security functions support their peer functions and share threat intelligence automatically, and in immediately usable ways.
- A unique hardware architecture which supports low latency, multi-function processing through an ASIC-based (application specific integrated circuit), single-pass parallel processing architecture.
- Full-stack analysis of packets performed upfront, and in both directions, allowing efficient subsequent processing and maximum visibility.

## 3 Firewall Selection Criteria

Firewall selection criteria will typically fall into three areas:

- **Security Functions:** The security functional elements correspond to the efficacy of the security controls, and the ability to manage the risk associated with the applications traversing the network.
- **Operations:** From an operations perspective, the big question is, “where does application policy live, and how hard or complex is it to manage?”
- **Performance:** The performance difference is simple: Can the firewall do what it is supposed to do at the required throughput the organization requires?

## 10 Things a Next-Generation Firewall Must Do

While each organization will have varied requirements and priorities within the three selection criteria, the ten things a next-generation firewall (or any firewall for that matter) must do are:

### Identify & Control Applications on Ports

Application developers no longer adhere to standard port/protocol/application development methodologies. More and more applications are capable of operating on non-standard ports or can hop ports (e.g., instant messaging applications, peer-to-peer file sharing, or VoIP). Additionally, users are increasingly savvy enough to force applications to run over non-standard ports (e.g., RDP, SSH).

In order to enforce application specific firewall policies where ports are increasingly irrelevant, a firewall must assume that any application can run on any port. The concept of any application on any port is one of the fundamental changes in the application landscape that is driving the migration from port-based firewalls to next-generation firewalls.

Any application on any port also underscores why a negative control model cannot solve the problem. If an application can move to any port, a product based on negative control would require beforehand knowledge or would have to run all signatures on all ports, all the time.

You must assume that any application can run on any port and your firewall must classify traffic by application on all ports all the time, by default. Port-based controls will continue to be outwitted by the same techniques that have plagued them for years.

## Identify and Control Circumventors

***A small number of the applications on the network may be used to purposely evade the very security policies that are put in place to protect an organization's digital assets.***

Two classes of applications fall into the security evasion tools—those that are expressly designed to evade security (e.g., external proxies, non-VPN related encrypted tunnels) and those that can be adapted to easily achieve the same goal (e.g., remote server/desktop management tools):

- External proxies and non-VPN related encrypted tunnel applications are specifically used to circumvent in-place security controls using a range of evasion techniques. These applications have no business value as they are designed to evade security, introducing unseen business and security risks.
- Remote server/desktop management tools, such as RDP and TeamViewer, are typically used by support and IT professionals to work more efficiently. They are also frequently used by employees to bypass the firewall, establishing connections to their home or other computer outside of the network.

To be clear, not all of these applications carry the same risks—remote access applications have legitimate uses, as do many encrypted tunnel applications. However, these same tools are increasingly being adopted by attackers as part of ongoing persistent attacks. Without the ability to control these security evasion tools, organizations cannot enforce their security policies, exposing themselves to the very risks they thought their controls mitigated.

## Understanding Circumvention Applications

***There are different types of circumvention applications—each using slightly different techniques.***

There are both public and private external proxies (see [proxy.org](http://proxy.org) for a large database of public proxies) that can use both HTTP and HTTPS.

- Private proxies are often set up on unclassified IP addresses (e.g., home computers) with applications like PHPProxy or CGIProxy.
- Remote access applications like RDP, TeamViewer or GoToMyPC have legitimate uses, but due to the associated risk, should be managed more closely.
- Most other circumventors (e.g., Ultrasurf, Tor, and Hamachi) have no legitimate business use case. Regardless of your organization's security policy stance, your firewall needs to have specific techniques to identify and control all of these applications, regardless of port, protocol, encryption, or other evasive tactic.

One more consideration: applications that enable circumvention are regularly updated to make them harder to detect and control, so it is important to understand not only that your firewall can identify these circumvention applications, but, it is also important to know how often that firewall's application intelligence is updated and maintained.

## Decrypt Outbound SSL and Control SSH

***Many applications use SSL in some way, shape, or form on today's corporate networks.***

Given the increasing adoption of HTTPS for many high-risk, high-reward

applications that end-users employ (e.g., Gmail, Facebook), and users' ability to force SSL on many websites, there is large and growing blind spot in firewalls that don't have the ability to decrypt, classify, control, and scan SSL-encrypted traffic.

A next-generation firewall must be flexible enough that certain types of SSL-encrypted traffic can be left alone (e.g., web traffic from financial services or health care organizations) while other types (e.g., SSL on non-standard ports, HTTPS from unclassified websites in Eastern Europe) can be decrypted via policy.

SSH is used nearly universally and can be easily configured by end-users for non-work purposes in the same manner that a remote desktop tool is used. The fact that SSH is encrypted also makes it a useful tool to hide non-work related activity.

The ability to decrypt SSL is a foundational element—not just because it's an increasingly significant percentage of enterprise traffic, but also because it enables a few other key features that would end up incomplete or ineffective without the ability to decrypt SSL. Key elements to look for include recognition and decryption of SSL on any port, inbound and outbound; policy control over decryption, and the necessary hardware and software elements to perform SSL decryption across tens of thousands of simultaneous SSL connections with predictable performance.

Additional requirements to consider are the ability to identify and control the use of SSH. Specifically, SSH control should include the ability to determine if it is being used for port forwarding (local, remote, X11) or

native use (SCP, SFTP and shell access). Knowledge of how SSH is being used can then be translated into appropriate security policies.

### **Provide Application Function Control**

Application platform developers such as Google, Facebook, Salesforce.com or Microsoft provide users with a rich set of features and functions that help to ensure user loyalty but may represent very different risk profiles.

For example, allowing Webex is a valuable business tool, but using Webex Desktop Sharing to take over an employees' desktop from an external source may be an internal or regulatory compliance violation. Another example may be Google Mail (Gmail) and Google Talk (Gtalk). Once a user is signed into Gmail, which may be allowed by policy, they can easily switch context to Gtalk, which may not be allowed. A firewall must be able to recognize and delineate individual features and functions so that an appropriate policy response can be implemented.

A next-generation firewall must continually classify each application, monitoring for changes that may indicate a different function is now being used. The concept of "one and done" traffic classification is not an option as it ignores the fact that these commonly used applications share sessions and support multiple functions.

If a different function or feature is introduced in the session, the firewall must note it within the state tables and perform a policy check. Continual state tracking to understand the different functions that each application may support, and the different associated risks, is a critical requirement.

## **Systematically Manage Unknown Traffic**

Unknown traffic exists on every network and represents a significant risk. There are several important elements to consider with unknown traffic—is it categorized, can you minimize it through policy control, can your firewall easily characterize custom applications so they are “known” within your security policy, and does your firewall help you determine if the unknown traffic is a threat?

## **Unknown traffic is also strongly tied to threats in the network**

Attackers are often forced to modify a protocol in order to exploit a target application. For example, to attack a webserver, an attacker may need to modify the HTTP header so much that the resulting traffic is no longer identified as web traffic. Such an anomaly can be an early indication of an attack. Similarly, malware will often use customized protocols as part of their command and control model, enabling security teams to root out any unknown malware infections.

By default, a firewall must classify all traffic on all ports—this is one area where the earlier architecture and security control model discussion becomes very important. Positive (default deny) models classify everything, negative (default allow) models classify only what they’re told to classify.

Classifying everything is only a small part of the challenge that unknown traffic introduces. A firewall must give you the ability to see all unknown traffic, on all ports, in one management location and quickly analyze the traffic to determine if it is one of the following:



- an internal or custom application
- a commercial application without a signature
- a threat.

Additionally, a firewall must provide the necessary tools to not only see the unknown traffic, but to systematically manage it by controlling it via policy, creating a custom signature, submitting a commercial application packet capture (PCAP) for further analysis, or performing forensic investigation to determine if it a threat.

### **Scan for Viruses/Malware in Apps on Ports**

Enterprises continue to adopt a wide range of applications to enable the business—they may be hosted internally, or outside of a physical location. Whether it's hosted SharePoint, Box.net, Google Docs, Microsoft Office365, or even an extranet application hosted by a partner, many organizations have a requirement to use an application that may use non-standard ports, SSL or can share files. In other words, these applications may enable the business, but they can also act as a threat vector.

Furthermore, some of these applications (e.g., SharePoint) rely on supporting technologies that are regular targets for exploits (e.g., IIS, SQL Server). Blocking the application isn't appropriate, but neither is blindly allowing the applications along with the (potential) associated business and cybersecurity risks.

This tendency to use non-standard ports is highly accentuated in the world of malware. Since malware resides in the network, and most communication involves a malicious client (the malware) communicating

to a malicious server (command and control), then the attacker has full freedom to use any port and protocol combination he chooses.

Part of safe enablement is allowing an application and scanning it for threats. These applications can communicate over a combination of protocols (e.g., SharePoint uses CIFS, HTTP and HTTPS, and requires a more sophisticated firewall policy than “block the application.”)

The first step is to identify the application (regardless of port or encryption), determine the functions you may want to allow or deny, and then scan the allowed components for any of the appropriate threats—exploits, viruses/malware, or spyware—or even confidential, regulated, or sensitive information.

### **Enable the same application visibility and control**

Users are increasingly located outside the four walls of the enterprise, often times accessing the corporate network on smartphones or tablets. Once the domain of road warriors, this is now a significant portion of the workforce is capable of working remotely. Whether working from a coffee shop, home, or a customer site, users expect to connect to their applications via Wi-Fi, wireless broadband, or by any means necessary.

Regardless of where the user is, or even where the application they’re employing might be, the same standard of firewall control should apply. If your firewall enables application visibility and control over traffic inside the four walls of the enterprise, but not outside, it misses the mark on some of the riskiest traffic.

Your firewall must have consistent visibility and control over traffic

regardless of where the user is. This is not to say that the organization will have the exact same policy for both; for example, some organizations might want employees to use Skype when on the road, but not inside headquarters, whereas others might have a policy that says if outside the office, users may not download salesforce.com attachments unless they have hard disk encryption turned on.

This should be achievable on your firewall without introducing significant latency for the end user or undue operational hassles for the administrator.

### **Make network security simpler**

Many enterprises struggle with incorporating more information feeds, policies, and management into overloaded security processes and people. In other words, if a team cannot manage what they've already got, adding more devices, managing interfaces along with associated policies and information doesn't help reduce the administrative effort, nor does it help reduce incident response time.

The more distributed the policy is (e.g., port-based firewall allows port 80 traffic, IPS looks for/blocks threats and applications, secure web gateway enforces URL filtering), the harder it is to manage that policy.

- Which policy does the security team use to enable WebEx?
- How do they determine and resolve policy conflicts across these different devices?

Given that typical port-based firewall installations have rule bases that include thousands of rules, adding thousands of application signatures

across tens of thousands of ports increases complexity by several orders of magnitude.

### **Deliver the same throughput and performance**

Many organizations struggle with the forced compromise between performance and security. All too often, turning up security features on a firewall means accepting significantly lower throughput and performance. If a next-generation firewall is built the right way, this compromise is unnecessary.

The importance of architecture is obvious here too—in a different way. Cobbling together a port-based firewall and other security functions from different technology origins usually means there are redundant networking layers, scanning engines and policies—which translate to poor performance. From a software perspective, the firewall must be designed to do this from the beginning.

Furthermore, given the requirement for computationally intensive tasks (e.g., application identification, threat prevention on all ports, etc.) performed on high traffic volumes and with the low tolerance for latency associated with critical infrastructure, a firewall must have hardware designed for the task as well—meaning dedicated, specific processing for networking, security and content scanning.

### **Support the same functionality**

The explosive growth of virtualization and cloud computing introduces new security challenges that are difficult or impossible for legacy firewalls to effectively manage due to inconsistent functionality, disparate

management, and a lack of integration points with the virtualization environment. In order to protect traffic flowing in and out of the data center as well as within virtualized environments, a firewall must support the exact same functionality in both a hardware and virtualized form factor.

The dynamic setup and tear down of applications within a virtualized datacenter exacerbates the challenges of identifying and controlling applications using a port- and IP address-centric approach. In addition to delivering the features described above in both hardware and virtualized form factors, it is imperative that a firewall provide in-depth integration with the virtualization environment to streamline the creation of application-centric policies as new virtual machines and applications are established and taken down.

This is the only way to ensure you can support evolving data center architectures with operational flexibility while addressing risk and compliance requirements.

## Summary

- Endpoint security refers to protection mechanisms which reside on systems or devices. Examples: anti-virus software, anti-spyware software, desktop firewall, or host intrusion prevention.
- Signature based detection and prevention compares files with known-bad databases. Next-generation Exploit prevention obstructs the processes which allow an attack to occur.
- Network security refers to protection mechanisms which reside 'on the wire'. Examples: firewall, web security gateway, anti-spam gateway, or network intrusion prevention systems.
- Virtual Private Network or VPN is a method of providing encrypted tunnels between systems which secures data in transit from being intelligible if intercepted.
- Unified Threat Management or UTM devices combine multiple network security functions on a single platform.
- Next-generation network security implies new methods of processing network traffic and preventing threats which are truly integrated, providing broad visibility to network activity, and thus greater control.

## Discussion

1. The Wall Street Journal published a quote May 4, 2014 by Brian Dye, Sr. VP of Information Security at Symantec: "Antivirus is dead". Why would Mr. Dye make such a statement, when Symantec is well known for its Norton Antivirus product?
2. What is the difference between a virus and spyware, and how might the methods of preventing each differ?
3. Describe some methods and examples of how endpoint security and network security work together to provide protection to enterprise computing resources.
4. In the not-so-distant past, there was a debate in the industry as to whether or not to even provide employees with access to the internet. How has that debate changed today, and what security functions have shaped that debate?
5. Next-generation firewalls replace port-and-protocol based allow/block decision making, with allow/block based on applications, users, and content. Why that was change necessary?

## Knowledge Check

Try the following knowledge checks:

1. **True or False:** A Zero-day attack is one that lasts less than a day.
2. Name one advantage and one disadvantage of running HIPS on endpoint machines.
3. Two examples of next-generation endpoint security are Application Whitelisting and Exploit Prevention. How do their methods differ?
4. Name three traditional network security countermeasures.
5. Which of the following is not a defining characteristic of next-generation network security?
  - A) Integrated security tools.
  - B) Low latency packet processing with minimal throughput loss.
  - C) Adherence to strict port-and-protocol enforcement for allow/block decisions.
  - D) Bi-directional full-stack analysis of packets.
6. Name three commonly used applications which have the capability of circumventing security policy.
7. **True or False:** Secure Socket Layer (SSL) protects networks from attack.
8. **True or False:** Adherence to usage policies always requires blocking applications from operating at all.
9. What is the difference between a positive enforcement model and a negative enforcement model?
10. Name one of the greatest single challenges to securing virtualized data center environments.



## 3. Cybersecurity Best Practices & Principles

### Enterprise Security Design Elements

After being exposed to the various threats and security breaches that could impact your daily life in the previous chapters, it would be nice to read about some recommended principles and security solutions. Using the best practices referred to in this section, you can design and implement a secure network for an organization using the best practices and principles accounted for in this chapter. Before undertaking any complex projects, refer to the best practices in documented enterprise security designs.

One of the most impactful strategies is zero-trust networking. This is explained in this section where we counter it with perimeter-based security approaches dealt with in earlier times.

## Zero Trust-Based Approach—The Old Way

***The Zero trust security model is the preferred model for security in today's threat-riddled environment.***

According to the 2014 Cyberthreat Defense Report, more than 60 percent of organizations fell victim to one or more successful cyberattacks in 2013. Given the extent to which today's organizations continue to rely on perimeter-centric strategies, this is no surprise. The failure of resulting architectures is due not only to an outdated assumption that everything on the inside of an organization's network can be trusted, but also the inability of legacy countermeasures to provide adequate visibility, control, and protection of application traffic on the network.

The perimeter-based approaches to security that were allowable yesterday are no longer effective.

## Perimeter-Based Approach—The Old Way

***Perimeter-based trust relies on the assumption that everything on the internal network can be trusted.***

The primary issue with a perimeter-centric security strategy—where countermeasures are deployed at a handful of well-defined ingress/egress points to the network—is that it relies on the assumption that everything on the internal network can be trusted. However, this assumption is no longer a safe one to make given modern business conditions and computing environments where:

- Remote employees, mobile users, and cloud computing solutions blur the distinction between “internal” and “external”
- wireless technologies, the proliferation of partner connections, and the need to support guest users introduce countless additional pathways into the network branch offices may be located in untrusted countries or regions
- insiders, whether intentionally malicious or just careless, may present a very real security threat

### **Failures in Perimeter-Based Approach**

Perimeter-based approach strategies fail to account for:

- The potential for sophisticated cyber threats to penetrate perimeter defenses—in which case they would then have free passage on the internal network
- scenarios where malicious users are able to gain access to the internal network and sensitive resources by using the stolen credentials of trusted users
- the reality that internal networks are rarely homogeneous but instead include pockets of users and resources with inherently different levels of trust/sensitivity which should ideally be separated in any event (e.g., research and development and financial systems versus print/file servers)

### **Inadequate Capabilities of Perimeter-Based Approach**

It’s important to realize that a broken trust model is not the only issue responsible for the diminishing effectiveness of perimeter-centric

approaches to network security. Another contributing factor is that legacy devices and technologies commonly used to build network perimeters let too much unwanted traffic through. Typical shortcomings in this regard include the inability to:

- Definitively distinguish good applications from bad ones (which leads to overly permissive access control settings)
- adequately account for encrypted application traffic
- accurately identify and control users (regardless of where they're located or what devices they're using)
- filter allowed traffic not only for known application-borne threats, but also unknown ones

The net result is that merely re-architecting one's defenses in a way that delivers pervasive internal trust boundaries will not be sufficient. Care must be taken to also ensure that the devices and technologies used to implement these boundaries actually provide the visibility, control, and threat inspection capabilities needed to securely enable essential business applications while still thwarting modern malware, targeted attacks, and the unauthorized exfiltration of sensitive business data.

### **Zero Trust Security Model-The New Way**

First introduced by Forrester Research, Zero Trust is an alternative security model that addresses the shortcomings of failing perimeter-centric strategies by removing the assumption of trust from the equation. With Zero Trust, essential security capabilities are deployed in a way that provides policy enforcement and protection for all users, devices,

applications, data resources, and the communications traffic between them, regardless of location. Benefits of implementing a Zero Trust network include:

- Clearly improved effectiveness in mitigating data loss via visibility and safe enablement of applications, and detection and prevention of advanced threats;
- greater efficiency for achieving compliance with security and privacy mandates;
- increased ability to securely enable transformative IT initiatives—such as user mobility and infrastructure virtualization

### **Effective Security for Modern Networks—Zero Trust**

A promising alternative model for IT security, Zero Trust is intended to remedy the deficiencies with perimeter-centric strategies and the legacy devices and technologies used to implement them. It does this by promoting “never trust, always verify” as its guiding principle. This differs substantially from conventional security models which operate on the basis of “trust but verify.”

In particular, with Zero Trust there is no default trust for any entity—including users, devices, applications, and packets—regardless of what it is and its location on or relative to the corporate network. In addition, verifying that authorized entities are always doing only what they’re allowed to do is no longer optional; it’s now mandatory.

The implications for these two changes are, respectively:

1. The need to establish trust boundaries that effectively compartmentalize different segments of the internal computing environment. The general idea is to move security functionality closer to the different pockets of resources that require protection. This way it can always be enforced regardless of the point of origin of associated communications traffic.
2. The need for trust boundaries to do more than just initial authorization and access control enforcement. To “always verify” also requires ongoing monitoring and inspection of associated communications traffic for subversive activities (i.e., threats).

### **The Premise of Zero Trust Security**

The core Zero Trust principles listed in the following sections define the operational objectives of a Zero Trust implementation.

#### **Concept #1: Ensure that all resources are accessed securely regardless of location.**

This suggests not only the need for multiple trust boundaries but also increased use of secure access for communication to/from resources, even when sessions are confined to the “internal” network. It also means ensuring that only devices with the right status and settings (e.g, ones that are managed by corporate IT, have an approved VPN client and proper passcodes, and are not running malware) are allowed access to the network.

#### **Concept #2: Adopt a least privilege strategy and strictly enforce access control.**

The goal in this case is to absolutely minimize allowed access to resources as a means to reduce the pathways available for malware and attackers to gain unauthorized access—and subsequently to spread laterally and/or infiltrate sensitive data.

### Concept #3: Inspect and log all traffic.

This reiterates the need to “always verify” while also making it clear that adequate protection requires more than just strict enforcement of access control. Close and continuous attention must also be paid to exactly what is happening in “allowed” applications, and the only way to do this is to inspect the content for threats.

*Least privilege... or just least effective? Traditional security gateways and other devices that rely on stateful inspection technology are actually incapable of enforcing a least privileges policy (i.e., where only what's needed to support this business is allowed to pass). The issue with these devices is that their classification engines only understand IP addresses, ports, and protocols—and, therefore, can't distinguish the specific applications that reside behind/within these low-level “wrappers.” With a stateful inspection device, for example, a rule permitting traffic using the HTTP protocol on TCP port 80 would allow the passage of not only a legitimate e-commerce application, but potentially numerous other web applications and utilities as well, such as those used for web mail, social networking, and countless other purposes. The net result is that such devices are, in fact, poor candidates for implementing a Zero Trust security model.*

## Zero Trust Conceptual Architecture

To help understand what Zero Trust looks like in practice, a conceptual architecture is shown in Figure 3-1.

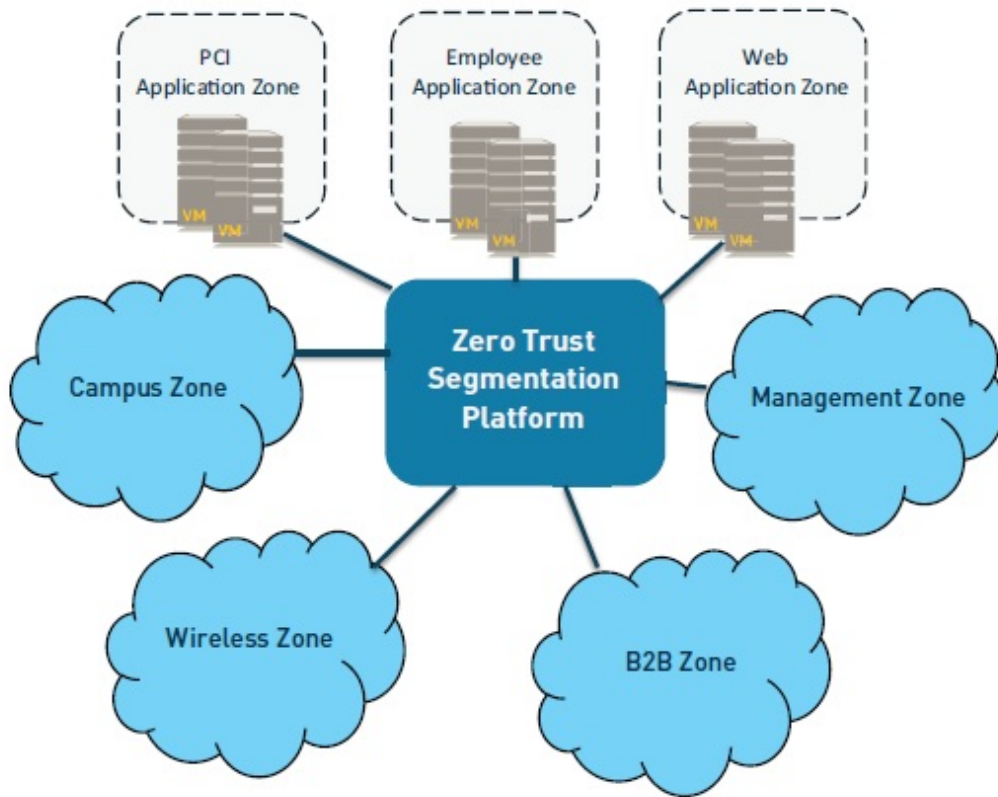


Figure 3-1: “Zero Trust” conceptual architecture.

The main components include the Zero Trust Segmentation Platform, trust zones, and associated management infrastructure.

### **Zero Trust Segmentation Platform**

Referred to as a network segmentation gateway by Forrester Research, the Zero Trust Segmentation Platform is the component used to define internal trust boundaries. In other words, it is what provides the majority of the security functionality needed to deliver on the Zero Trust operational objectives—including the ability to:



- enable secure network access
- granularly control traffic flow to/from resources
- continuously monitor allowed sessions for any of threat activity

Although Figure 3-1 depicts the Zero Trust Segmentation Platform as a single component in a single physical location, in practice—due to performance, scalability, and physical limitations—an effective implementation is more likely to entail multiple instances distributed throughout an organization’s network. In addition, the solution is designated as a “platform” not only to reflect that it is an aggregation of multiple distinct (and potentially distributed) security technologies, but also that they operate as part of a holistic threat protection framework to reduce the attack surface and correlate information about threats that are found.

### **Trust Zones**

Referred to as a micro core and perimeter (MCAP) by Forrester Research, a trust zone is a distinct pocket of infrastructure where the member resources not only operate at the same trust level but also share similar functionality. Sharing functionality such as protocols and types of transactions is imperative because this is what is needed to actually minimize the number of allowed pathways into and out of a given zone and, in turn, minimize the potential for malicious insiders and other types of threats to gain unauthorized access to sensitive resources.

Examples of trust zones shown in Figure 3-1 include the user (or campus) zone, a wireless zone for guest access, a cardholder data zone, database

and application zones for multi-tier services, and a zone for public-facing web applications.

It is important to note, too, that a trust zone is not intended to be a “pocket of trust” where systems (and therefore threats) within the zone are able to communicate freely/directly with each other. For a full zero trust implementation, the network would be configured to ensure that ALL communications traffic—including that between devices in the same zone—is intermediated by the corresponding Zero Trust Segmentation Platform.

### **Management Infrastructure**

Centralized management capabilities are crucial to enabling efficient administration and ongoing monitoring, particularly for implementations involving multiple distributed Zero Trust Segmentation Platforms. In addition, a data acquisition network provides a convenient way to supplement the native monitoring and analysis capabilities for a Zero Trust Segmentation Platform. By forwarding all session logs to a data acquisition network, this data can then be processed by any number of out-of-band analysis tools and technologies intended, for example, to further enhance network visibility, detect unknown threats, or support compliance reporting.

### **Zero Trust with Next-Generation Firewalls**

Because the heart of any Zero Trust network security architecture is the Zero Trust Segmentation Platform, it is imperative that organizations choose the right solution. Accordingly, this section identifies a set of key criteria and capabilities for IT security managers and architects to

consider when making a selection. In each case, a brief synopsis is also provided of how a next-generation security platform meets the corresponding requirements. The following are the comprehensive security functionality.

### **Secure Access**

Consistent secure IPsec and SSL VPN connectivity is provided for all employees, partners, customers, and guests wherever they're located (e.g., at remote/branch offices, on the local network, or over the Internet). Policies to determine which users and devices can access sensitive applications and data can be defined based on application, user, content, device, and device state.

### **Inspection of all Traffic**

Application identification accurately identifies and classifies all traffic, regardless of ports and protocols, evasive tactics such as port hopping, or encryption. This eliminates methods that malware may use to hide from detection and provides complete context into applications, associated content, and threats.

### **Least Privileges Access Control**

The combination of application-, user-, and content identification delivers a positive control model that allows organizations to control interactions with resources based on an extensive range of business-relevant attributes, including the specific application and individual functions being used, user and group identity, and the specific types or pieces of data being accessed (e.g., credit card or social security numbers). Compared to traditional firewalls which let too much traffic through

because they're limited to port and protocol level classification, the result is truly granular access control that safely enables the right applications for the right sets of users while automatically eliminating unwanted, unauthorized, and potentially harmful traffic from gaining access to the network.

### **Advanced Threat Protection**

A combination of anti-virus/malware, intrusion prevention, and advanced threat prevention technologies provide comprehensive protection against both known and unknown threats, including threats on mobile devices. In addition, support for a closed-loop, highly integrated defense ensures that inline enforcement devices and other components in the threat protection framework are automatically updated.

### **Coverage for all IT domains**

Virtual and hardware appliances enable trust boundaries to consistently and cost-effectively be established throughout an organization's entire network, including in remote/branch offices, for mobile users, at the Internet perimeter, in the cloud, at the ingress to the datacenter, and for individual enclaves wherever they might exist.

### **High-performance design**

By definition, a Zero Trust Segmentation Platform aggregates numerous security and networking capabilities. However, it must also be capable of delivering all of these features without becoming a performance bottleneck. Several design elements of a high-performance Zero Trust Segmentation platform should include:

- **Single-pass software architecture.** This minimizes latency and processing requirements as there is no need for traffic streams to be processed multiple times (e.g., once for each security function).
- **Separate control and data planes and function-specific, parallel processing hardware engines (i.e., custom chips).** This provides core packet processing, acceleration of standard security functions, and dedicated content scanning.

### Flexible, non-disruptive deployment

Ideally, it should be possible to implement a Zero Trust approach in a way that requires no modification to the existing network and is completely transparent to your users. Opportunities to take advantage of major network overhauls are rare, and disrupting operations is not a good career choice. Thus, IT security managers will need to make due as best they can, typically by converting to Zero Trust on the fly. A next-generation security platform can support this requirement in numerous ways. For example:

- A single hardware appliance can support multiple different connection modes (Layer 1, Layer 2, or Layer 3), thereby maximizing its ability to accommodate trust zones with different needs.
- Support for a broad range of networking technologies (e.g., L2/L3 switching, dynamic routing, 802.1Q VLANs, trunked ports, and traffic shaping) guarantees the ability to integrate into practically any environment.
- Multiple management domains (see Figure 1) can be accommodated by taking advantage of a virtual systems capability that enables separate,

isolated Zero Trust virtual instances on a physical appliance. Virtual systems allow you to segment the administration of all policies (security, NAT, QoS, etc.) as well as all reporting and visibility functions.

### **A Progressive Approach for Implementing Zero Trust**

In terms of moving forward with a Zero Trust design, it is important for IT security managers and architects to realize that it's not necessary to instigate or wait for the next comprehensive overhaul of their organization's network and security infrastructure. Indeed, one of the great advantages of a Zero Trust architecture is that it is conducive to progressive implementation.

To get started, IT security teams can configure a Zero Trust Segmentation Platform in listen-only mode to obtain a detailed picture of transaction flows throughout the network, including where, when and to what extent specific users are using specific applications and data resources.

Armed with these details, the security team would then be in an excellent position to incrementally:

- deploy devices in appropriate locations to establish internal trust boundaries for identified trust zones
- configure the appropriate enforcement and inspection policies to effectively put each trust boundary “online”

## Advantages

Advantages of a progressive approach such as this include minimizing the potential impact on IT operations and being able to spread the required investment and work effort over time.

## An Alternative Approach

For those security teams that already have a good understanding of the transaction flows in their environment, an alternate approach is to map out trust zones and begin to establish corresponding trust boundaries based on relative risk and/or sensitivity of the data involved. A logical starting point in this case is to begin by identifying well-defined pockets of users and systems involving high concentrations of sensitive data—such as the 4Ps:

- Payment card industry (PCI) or other financial data
- Personal healthcare information (PHI)
- Personally identifiable information (PII)
- Intellectual property (IP)

From there, it then makes sense to consider progressively establishing trust zones/boundaries for other segments of the computing environment based on their relative degree of risk—for example:

- IT management systems/networks (where administrators often hold the proverbial “keys to the kingdom” and a successful attack could lead to compromise of the entire network)
- Partner resources and connections (B2B)
- High-profile, customer-facing resources and connections (B2C)

- Branch offices in risky countries or regions, followed by all other branch offices
- Guest access networks (both wireless and wired)
- Campus networks

Adopting Zero Trust principles and concepts at major access points to the Internet also makes sense. However, this will probably require replacing or augmenting entrenched, legacy security devices with a Zero Trust Segmentation Platform to obtain all of the requisite capabilities.

### **Benefits of Adopting Zero Trust Principles and Practices**

There are several technical and business advantages associated with using a Zero Trust security architecture. These include being able to:

- Incrementally and non-disruptively make the transition to a Zero Trust model
- Obtain unparalleled situational awareness of enterprise computing activity, legitimate and otherwise
- Fully implement all Zero Trust principles and concepts, including strict enforcement of a least privileges access control policy (which is essential to reducing attack surface)
- Dramatically enhance the organization's security posture and ability to prevent the exfiltration of sensitive data
- Simplify achieving and maintaining compliance with applicable standards and regulations (by using highly effective trust boundaries to segment off sensitive resources)



- Securely enable and easily adapt to accommodate business-driven IT initiatives—such as user mobility, social networking, infrastructure virtualization, and cloud computing
- Reduce total cost of ownership (by using a single consolidated security platform across the entire computing environment, instead of a disparate collection of disconnected point products)

Perimeter-centric security strategies continue to be sorely challenged. The issue is not only increasingly sophisticated cyber threats, but also major changes to the technology and business landscape—such as user mobility, hyper inter-connectivity, and globalization—that invalidate the assumption that everything “on the inside” can be trusted. The bottom line is that such strategies—along with the legacy technologies used to implement them—are, for the most part, no longer effective.

Organizations looking to substantially improve their defensive posture against modern cyber threats and more reliably prevent exfiltration of sensitive data should consider migrating to a Zero Trust security architecture. An alternative model for IT security, Zero Trust eliminates the faulty assumption of trust and rectifies the shortcomings of traditional perimeter-centric architectures by promoting the use of a Zero Trust Segmentation Platform to establish secure “trust boundaries” throughout a computing environment and, in general, in closer proximity to sensitive resources.

## **Next-Generation Network Security Principles**

In this section we introduce methods and best practices to control botnets and related malware. This methodology is designed to supplement existing security strategies of an enterprise, as part of a modern coordinated approach to defense-in-depth. The methodology addresses techniques to limit the exposure to botnets, as well as the detection and remediation of endpoints that may already be infected.

### **Incorporating Next-Generation Firewalls**

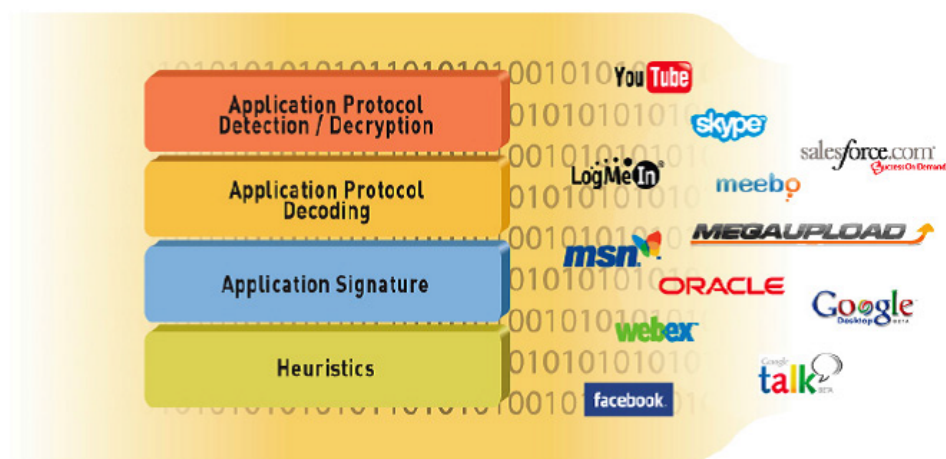
As malware evolves from individual endpoints to coordinated bot networks, enterprises similarly need to expand their analytical perspective to incorporate network level intelligence and controls. Network security allows us to focus on the very trait that distinguishes botnets from earlier forms of malware—its reliance on communication with a larger bot network. To twist John Gage’s famous phrase, “the network is the computer”, in a very real sense the threat has become a network. If our security measures don’t operate at this same level, we run a very real risk of missing the forest for the trees.

Additionally, network security mechanisms provide an independent layer of monitoring and control, unlike the end-points themselves, which can be compromised by malware. Botnets and modern malware can include rootkits, whose purpose is to gain root access on the target machine in an effort to subvert anti-virus protections or other security mechanisms on the machine. This creates a “Catch 22” for the security team since any security software running on a compromised host cannot truly be trusted. This certainly shouldn’t imply that host-based security is obsolete, but rather simply to point out that the host layer certainly needs additional layers of defense in depth.

The points above could apply to network security in general. However, the next-generation firewall in particular provides arguably the most important addition to the fight against botnets—the reliable visibility and control of all traffic on the network regardless of the evasive tactics that are employed. By understanding the full stack behavior of all traffic on the network, we can finely control the behaviors that are allowed in the corporate environment, while simultaneously eliminating the shadows that botnets rely on to remain hidden. Bots quite simply must talk in order to function, and finding these telltale communications is becoming a critical component of controlling botnets and the threats they pose to the enterprise.

### **Classification of Traffic**

A next-generation firewall performs a true classification of traffic based not simply on signatures, but an ongoing process of application analysis, decryption, decoding and heuristics to progressively peel back the layers of a traffic stream to determine its true identity (see Figure 3-2). This ability to pinpoint and analyze even unknown traffic without regard to port or encryption is a defining characteristic of a true next-generation firewall, and as we shall see this ability is invaluable in the fight against botnets.



**Figure 3-2: Next-generation firewalls classify the true nature of traffic using a variety of identification methods**

## Integration

Additionally, a true next-generation firewall provides a fully integrated approach to threat prevention. The distinction in this case is the true coordination of multiple security disciplines as opposed to simply co-locating them on the same box. For example the application identity, malware detection, intrusion prevention, URL filtering, file type controls and inspection of the content should all be integrated into a unified context. This integration provides a far more intelligent and definitive understanding of botnets than any individual technology can provide by itself. This collective intelligence is needed in order to see and understand the telltale signs of unknown threats.

## Reduction and Elimination

One of the most important steps that an enterprise can take to control malware is to reduce the vectors of infection and eliminate the ability for

the bots to hide. Today, the majority of the attack vectors used by botnets are virtually unchecked, and botnet traffic is typically small enough that it can easily blend into the background of “normal” network traffic. By regaining full visibility and control of exactly what traffic is allowed into the network and why, security teams can go a long way to countering modern malware threats.

### **Reduce the Attack Surface**

An important first step for the enterprise is to return to a positive control model. Positive control simply means specifically allowing the traffic you want as opposed to blocking everything that you don't. The notion of positive control has long been one of the defining characteristics of network firewalls that separate them from other types of network security.

For example, if you want to use Telnet, then you open TCP port 23 to allow Telnet without necessarily allowing all other types of traffic. Unfortunately, traditional firewalls have progressively lost the ability to enforce positive control in any reliable way, as applications have learned to use non-standard ports or commonly open ports (TCP ports 80, 443, 53), or simply hop between open ports.

### **Enforce Positive Control**

Enforcing positive control is essential in the fight against malware as it provides an easy way to greatly reduce the attack surface of the enterprise and reduce overall risk. The simple truth is that the number and diversity of applications has exploded, and almost all of them can introduce some level of risk. The rise of Web and Enterprise 2.0, widgets

and a variety of scripting options have empowered individuals to develop powerful applications or services, most of which are designed to connect or be used with other applications and sites.

Making matters worse, very few of the applications that are written on a daily basis have any real value to an enterprise. By incorporating a positive control model, security teams can focus on enabling the approved applications, as opposed to constantly trying to stay up to speed with all of the applications that they want to block. This approach can immediately preclude large numbers of applications from ever touching the network, while dramatically reducing the number of vectors that botnets can use to get in or out of the network.

### **Consult and Conclude**

However, enforcing positive control is not as easy as simply flipping a switch. Some applications may be used by staff and have business value that is not readily apparent. As such, IT and security teams should plan to consult with a variety of groups within the organization in order to establish an appropriate set of approved applications and use roles.

### **Plan on Refinements**

Additionally, some applications, such as Facebook, will have both business and personal uses. In these cases, security policies should be further refined to allow only certain users to access an application or limit the use of an application to certain approved features (see Figure 3-3). Reduce the attack surface by restricting certain applications to specific users and groups based on business need.

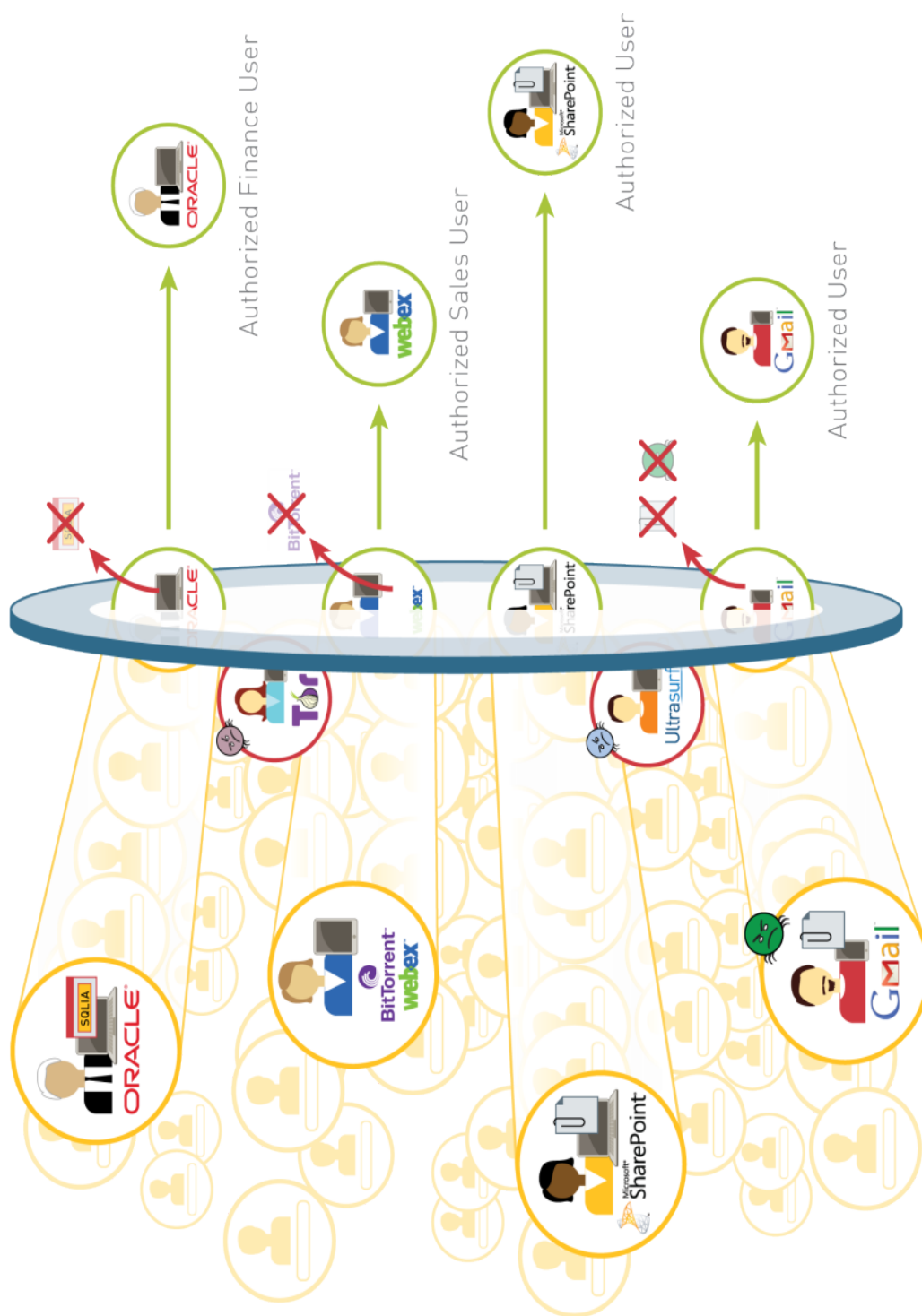


Figure 3-3: Defending Against Attach Surfaces in Business Environments

## Establish Policies and Controls

*Establish policies of approved applications and uses based on company needs and culture:*

- *Establish a baseline of what is on the network—applications, protocols, services, etc.*
- *What applications are in use?*
- *What applications are required for the business and who needs to use them?*
- *What dual-use or personal applications does the enterprise want to allow?*

*Enforce positive control of all traffic*

- *Prevent unnecessary or high risk traffic*
- *Regardless of port evasion or encryption techniques employed*

## Investigate Unknowns

Once the enterprise has regained the ability to accurately classify the approved traffic on the network, we then have a base from which to investigate any remaining unknown traffic in the network.

## Track and Correlate

The presence and behavior of unknown traffic is a critical clue in the identification of botnets, which will often present as “unknown” traffic due to their use of proprietary encryption and unique behavior (see Figure 3-4).



(receive\_time in last-hour) AND (app eq unknown-udp)









	Receive Time	Application	Name
	04/29 10:32:08	unknown-udp	Bot: Mariposa Command and Control
	04/29 10:17:24	unknown-udp	Bot: Mariposa Command and Control
	04/29 10:17:24	unknown-udp	Bot: Mariposa Command and Control
	04/29 10:16:50	unknown-udp	Bot: Mariposa Command and Control
	04/29 10:02:09	unknown-udp	Bot: Mariposa Command and Control
	04/29 09:47:23	unknown-udp	Bot: Mariposa Command and Control
	04/29 09:47:23	unknown-udp	Bot: Mariposa Command and Control
	04/29 09:46:52	unknown-udp	Bot: Mariposa Command and Control

Figure 3-4: Investigate any unknown traffic

### Define and Collect

Custom UDP and TCP traffic, including modified P2P, IM and file transfer, is commonly used for command and control in a botnet. The use of non-standard ports is a favorite tactic of malware for detection avoidance using signature-based anti-malware and is also used to download new payloads. To put the significance of unknown traffic into context, according to research by Palo Alto Networks, unknown UDP traffic typically accounts for only 2 percent of total network bandwidth, yet constitutes 51 percent of malware logs.

## Investigate

Unknown traffic regularly sent by the same client machine should be investigated to determine if the individual is using a legitimate application that is not recognized versus a potential botnet infection. Teams can also investigate where the traffic is going:

- Does it reach out to websites known to serve malware or to social networking sites?
- Does it transmit on a regular schedule or follow a recognizable pattern?
- Does someone attempt to download or upload files to an unknown URL?

## Detection

All of these behaviors can identify the presence of a client machine that is infected by a bot. By positively identifying the approved traffic on the network, any “unknown” traffic should become increasingly rare, thus enabling potentially malicious botnet traffic to be found and analyzed quickly.

- *Track source and destination, volumes of traffic*
- *Correlate against URL, IPS (intrusion prevention system), malware, and file transfer records*
- *Define custom application IDs for any internal or custom applications, as needed*
- *Collect PCAPs (packet captures) for any unrecognized, publicly available applications and deliver to trusted security vendors for analysis*
- *Investigate “unknown” traffic for potential unauthorized user behavior or potential botnet behavior*

## **Control the Enabling Applications**

Applications are an indispensable part of the botnet lifecycle. They are crucial to the initial infection stage and the ongoing command and control of the botnet. This association of malware and applications is nothing new.

### **Target Software Applications**

In the past, the de facto enabling application for malware was corporate e-mail. From a security perspective, viruses and e-mail simply went hand-in-hand. However, even though e-mail is still used by attackers, it has gradually lost its luster for attackers as e-mail applications are now typically heavily secured by the enterprise, and e-mail messages can be analyzed at considerable depth while the message is at rest on the e-mail server.

### **Target Software Users**

Bot-herders have now shifted much of their attention to softer target applications that interact with users in real-time and offer far more flexibility than corporate e-mail.

### **Social Networking**

All applications are not created equal, and botnet owners have gravitated to applications that support the bot's ultimate aim to facilitate social engineering while remaining hidden. Social networking and personal use applications meet both of these criteria, and have become some of the most common sources for malware infection and the subsequent command and control of botnets. This includes applications such as social networking apps themselves, web-based e-mail, instant messaging (IM) applications, P2P networks and a variety of file transfer and file sharing applications.

These applications are fundamentally designed to easily share information in a variety of ways, and users often bring a more cavalier attitude to these applications because they may be accustomed to using them outside of the office. This provides an intelligent attacker with a multitude of infection options to pursue and develop.

### **Social Engineering**

Social applications also present an ideal environment for social engineering, enabling an attacker to impersonate a friend or colleague in order to lure an unsuspecting victim into clicking on a dangerous link. For all of its sophistication, malware infection continues to rely on enticing an unsuspecting user into an ill-advised click. Instead of opening an e-mail attachment, the click may be a link in a tweet or a link on a Facebook page that appears to be from a friend. Cross-site scripting can populate dangerous links among friends and sniffing technologies, such as FireSheep, allow hackers to take over social networking accounts.

This connection between social networking and malware has been observed in the real world. Research from the Information Warfare Monitor and Shadowserver Foundation has provided compelling evidence for the role of social networking applications in the botnet lifecycle.

*In the 2010 paper *Shadows in the Cloud*, the group tracked a very targeted and persistent intrusion into a network using a customized botnet. In their analysis, the group found that the bot-infected machines rarely, if ever, communicated directly with the command and control servers. Instead, the initial malware traffic from the infected host would go to popular blogs, Google Groups, Twitter accounts, and Yahoo! Mail accounts, which allowed the malware communications to blend in with “normal” traffic. This illustrates the key lesson that botnets will often attempt to blend in with what is considered normal, but low value traffic, in the network. How often would a security administrator investigate what appears to be a user simply posting something to an innocuous blog?*

## **Securely Enable**

The ability to securely enable applications is a critical requirement for enterprises. Simply blocking all access to blogs, webmail, IM and social networking applications would be both impractical and unduly constrain the enterprise’s ability to communicate and stay connected with the outside world.

A first step to securely enabling an application is to limit access to the application by users or user groups that have an approved need for the application. For example, access to Facebook and its underlying applications can be limited to sales and marketing teams who are responsible for maintaining the company’s online identity, while other employees are not allowed access, or have tighter restrictions. This again can significantly reduce the attack surface of the enterprise and reduce the risk of an infection.

Security teams also need the capability to allow certain applications, while preventing the use of specific features that introduce unnecessary risk.

For example, the company could allow access to a social networking application but disable the posting functionality or prevent the application from downloading files or other risky behaviors, such as tunneling other applications or sharing the user's desktop. This step can significantly limit the ability for a malware payload to be transferred to a specific target.

An infected webpage can easily cause a target end-user to automatically and unwittingly download a malicious file from a website in the background. This is commonly referred to as a drive-by download, and can occur even on perfectly valid web pages that have been compromised. Drive-by downloads are critical for social networking attacks where links may redirect to sites that are serving up exploits which are then used to download malware or droppers in the background. Enabling drive-by-download protection features in next-generation firewalls protects against this type of infection by prompting users to verify that they really intended to download a file and that files are not pulled down without the user's knowledge.

### **Adoption on SSL**

The reliance on the Internet and cloud computing is leading to a widespread adoption of SSL for a variety of technologies and industries. Social networking sites are also inadvertently making it easier for malware to remain hidden by moving to the default use of SSL to protect user communications. This is a needed improvement given that hackers can eavesdrop and hijack unprotected HTTP sessions.

Tools such as FireSheep have made this process simple for anyone and threaten the notion of privacy on the Internet. On the other hand, most enterprises lack the ability to dynamically look within SSL encrypted communications, thus making the social networking traffic more or less invisible to the enterprise. This represents a net loss for the enterprise security team-the user gets improved privacy for their social traffic, but in the process it establishes an invisible communication infrastructure for the same sites and applications favored by malware.

This move to SSL by default can realistically make social applications as valuable to attackers as P2P applications, such as BitTorrent, have been for the past several years. The enterprise must establish a strategy for dealing with SSL encrypted traffic. A next-generation firewall with on-box SSL decryption can be leveraged based on application or application type. This allows staff to specifically target social networking applications for SSL decryption and content inspection. Security teams can also use URL filtering categories to avoid decrypting sensitive personal information such as financial and personal health information.

- *Prevent use of known “bad” applications*
  - *P2P*
  - *Limit application usage to users/groups who have a need*
- *Prevent the use of dangerous features*
  - *File transfer*
  - *Desktop sharing*
  - *Tunneling of other applications*
- *Prevent drive-by-downloads by educating users and enabling drive-by-download protection features in a next-generation firewall*
- *Selectively decrypt SSL based on application and URL category*
  - *Decrypt social networking, webmail, Instant Message*
  - *Do not decrypt traffic to/from health care or financial sites*
- *Inspect all allowed risky application traffic using:*
  - *Intrusion and threat prevention*
  - *Malware protection*
  - *URL filtering*

## **Prevent the Use of Circumventors**

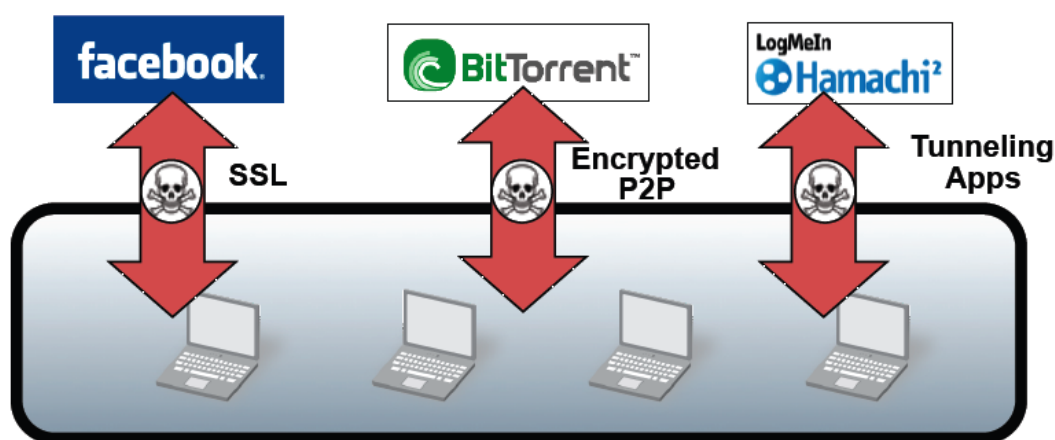
The previous section focused on the common end-user and Web 2.0 applications that can be co-opted by malware for use against the enterprise. However, there is a second class of applications that are proactively designed to pass through traditional network security. This



includes a variety of remote desktop technologies, proxies, and purpose-built circumventing applications, which will also require tight control (see Figure 3-5).

### Prevent Opening of Unmanaged Threat Vectors

Some of these applications have valid enterprise uses, while others are a sure sign of unapproved and dangerous behavior. In all cases they will require tight control by IT to prevent the opening of unmanaged threat vectors into the enterprise.



**Figure 3-5: Circumventors use SSL, encrypted P2P, and tunneling applications to avoid detection**

Remote desktop technologies have become wildly popular, both with end-users and IT support teams. Even most web-conferencing applications have added the ability to give a remote user control over a user's machine. Such technologies introduce two risks. First, when a user connects to his remote PC, he is free to surf to any destination and use any application without that traffic being inspected by the firewall. In addition to

circumventing policy, the remote desktop opens an unmanaged threat vector by allowing a user to remotely undertake all kinds of risky behavior and then have the results tunneled back to his machine inside the enterprise.

### **Prevent an Intrusion**

Secondly, remote desktop technologies provide the risk of an outside user gaining full access to a machine inside the trusted enterprise network. This type of remote control is one of the ultimate goals of malware in the first place, and as such it obviously creates a dangerous opening to launch an intrusion.

There are also common enterprise applications which have valid uses within the enterprise but can easily create unintentional exposures if misused or used by unauthorized or untrained users. For example, many enterprises use SSH (Secure Shell) to manage systems and applications in a corporate datacenter. However, if SSH is used to open tunnels into the datacenter, it can open direct, unmanaged access into the enterprise's most critical assets. These applications will need tight control in the enterprise with only select individuals approved and any tunneling features closely regulated.

Additionally, a variety of web proxies and encrypted tunneling applications have evolved whose primary goal is to provide secure and anonymous communication across firewalls and other security infrastructure. Proxy technologies, such as CGIProxy or PHPProxy, provide a relatively easy way for users to surf anonymously without enterprise control and have been found in more than 75% of enterprise networks.

Applications such as UltraSurf, Hamachi, and Tor are purpose-built to traverse security infrastructure and are regularly updated in order to remain undetected. These applications have very few, if any, valid reasons for use within the enterprise, and their presence generally indicates an intentional attempt to avoid enterprise security. These tools not only pass traffic without being inspected, but they also tend to be used for high-risk behaviors such as file sharing or expressly blocked content and sites, which in turn carry a significantly higher risk of malware infection. As a result, these applications should be blocked in almost all cases.

- *Limit remote desktop usage to IT support only*
- *Securely enable SSH by allowing, but preventing SSH tunneling*
- *Block use of unapproved proxies*
- *Block encrypted tunnels such as UltraSurf and Hamachi*

## **Protect Remote Users**

Thus far we have assumed a fairly traditional network topology with a clear separation between the inside and outside of the network. However, enterprise computing has evolved to reach well beyond the traditional physical boundaries of the enterprise. Users take their laptops home with them and expect to be able to connect and work literally from anywhere and at any time, on any device. This creates an imbalance in security posture in that users expect to be able to do the same work from any location, while the lion's share of security infrastructure and countermeasures (firewalls, IPS, etc.) are applied only when the user is

inside the traditional physical perimeter of the enterprise. To make matters worse, a user's browsing and application behaviors often tend to be riskier when outside the office than when inside, as users inevitably revert back to their personal behaviors at home. This behavior greatly increases the likelihood of clicking on a dangerous link or visiting a website that serves up a drive-by download.

- *Enforce full enterprise firewalling and threat prevention regardless of user location*
- *Enforce drive-by-download protections*
- *Enforce customized policies based on user location—do not allow files to be downloaded from secure systems when a user is remote*

### **Finding Infected Hosts**

In spite of the security team's best efforts at prevention, enterprise machines will inevitably be infected with malware. This could be via an unknown type of malware, an unknown vector, or by physical connections such as a USB drive. Malware has proven time and again that it is possible to infect even the most heavily secured systems in the world. As a result, it is important for teams to assume user devices are infected and develop the skills needed to find the infected hosts in the network. This can be a challenging task given that the malware may have already avoided traditional malware signatures and may already have root access on the infected machine.

To pinpoint these infected hosts, we must shift our attention from malware signatures to instead analyze behaviors that are observed in

the network. For all of their secrecy and ingenuity, botnets need to communicate in order to function, and they also need to make themselves difficult to find and trace. These basic requirements create patterns that can be used to identify bot traffic or behaviors that stand out from the normal end-user traffic, even if the bot is completely unknown in the industry.

### **Detection of Command and Control Traffic**

One of the major advantages of a next-generation firewall is the ability to classify potentially complex streams of traffic at the application level. This includes the ability to progressively scan within traffic to peel back protocols running within protocols until the true underlying application is identified. This expertise in identifying complex traffic is very valuable when identifying the unique command and control traffic of particular botnets.

### **Leveraging IPS to Detect Botnets**

Custom IPS signatures can also be created to identify potentially polymorphic malware based on specific components within the malware. For example, SpyEye a very popular and growing banking botnet reserves space so that the malware can constantly change its size, and therefore change its signature. However, SpyEye periodically downloads an encrypted configuration file to update the bot. Researchers at Palo Alto Networks were able to break into this configuration file and were able to find a unique pattern across all configuration files, that enables an IPS to identify the presence of the bot, even if the malware itself is not recognized. This is just one example where IPS and malware detection can intersect to find a modern threat.

## Automation of Bot Characteristics

Unfortunately, most enterprises don't have the time or resources to conduct manual investigations and to research new and emerging threats. Security industry vendors can provide enterprises with the necessary tools and information to protect corporate systems and networks. For example, the Palo Alto Networks Behavioral Botnet Report automates the process of tracking and correlating the behaviors that indicate the presence of a bot. This report looks for several bot characteristics including:

- **Unknown TCP/UDP**—Botnet traffic is regularly encrypted and unknown. Since a next-generation firewall identifies all traffic, tracking unknown TCP and UDP traffic can be a perfect starting point for finding bot-infected machines. The report allows staff to track unknown traffic by sessions, destinations and bytes.
- **Presence of Dynamic DNS**—Malware will often use dynamic DNS in order to make botnet communications more difficult to track. By bouncing traffic between multiple infected hosts with an ever-changing list of IP addresses, it can become very difficult to track the path of the bot and its true source and destination.
- **Activity on Known Malware Sites**—Palo Alto Networks constantly tracks sites that have hosted malware whether intentionally or unintentionally. A next-generation firewall with a URL filtering capability can track if a user is repeatedly (and unwittingly) visiting one of these sites and attempting to download files.
- **Visiting Recently Registered Domains**—Botnets are constantly moving around in order to avoid detection and to recover as servers are

discovered or disabled. As a result, botnets will often have to use new domains to support the command and control infrastructure. A user repeatedly visiting a newly registered domain will certainly not be conclusive, but may help to provide corroborating evidence of an infection.

- **Browsing to IP domains Instead of URLs**—In a similar vein, bots will often use hard-coded IP addresses or known IP ranges in order to communicate as opposed to users, who typically use URLs to browse Internet web sites. As with tracking newly registered domains, tracking connections using IP domains can sometimes indicate the presence of a bot at work as opposed to a human.
- **IRC (Internet Relay Chat) traffic**—IRC traffic is one of the most well-known communication methods for botnets, and provides an additional strong piece of correlating data for finding a bot.

*The Behavioral Botnet Report takes all of the factors above and automatically correlates them to find hosts that are likely infected with a bot. When run on a Palo Alto Networks next-generation firewall, the report provides specific directory user names of the users or machines that are likely infected along with what behaviors contributed to the analysis. Each user is also provided a score based on how many of the factors listed above were correlated, allowing staff to focus on the devices that are the most likely to be infected.*

## Using DNS Sinkholing to identify infected hosts

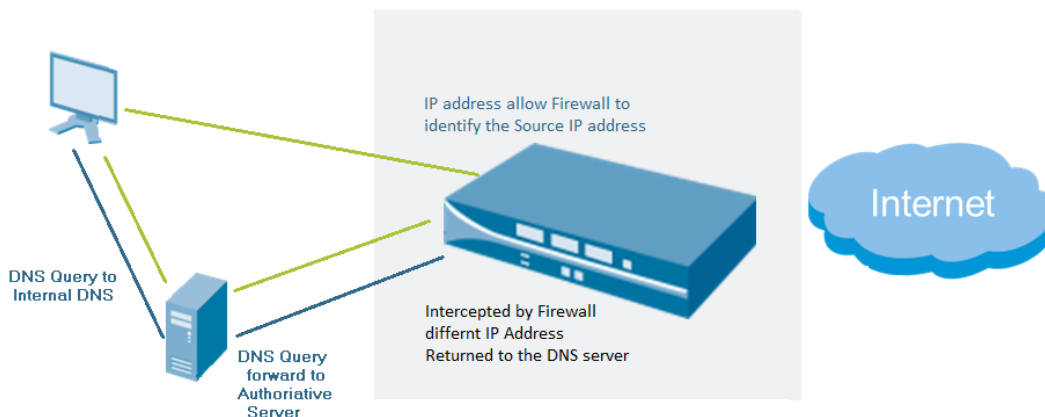
Domain Name Service (DNS) records can't always accurately identify infected hosts because the local DNS server is tasked with resolving the DNS query. To the firewall, the local DNS server is the originator of the query. DNS sinkholing controls the session, by intercepting the DNS

request and responding with a different IP address that directs the session to a specific server. Information is gathered from the session and used for threat research and to better protect the network from malware.

There are two DNS sinkhole approaches: *public sinkhole* and *local sinkhole*.

- Public sinkholing is used to locate malicious domains that are being used by malware, such as command and control servers. This practice requires publishing the authoritative records of the malicious domain to point to sinkhole servers controlled by the security vendor. Information collected from the user sessions improves the current DNS-based signatures. This is sometimes referred to as a “honeysink”.
- Local sinkholing is responding to DNS queries for malicious domains with a local sinkhole IP address. A next-generation firewall intercepts DNS queries and responds to DNS requests with fake DNS responses (see Figure 3-6). The IP address returned points to an internal sinkhole server or a non-routable IP address. Subsequently, all following sessions will connect to the sinkhole server or non-routable IP address, which makes it easy to identify the infected host in the traffic log files.





**Figure 3-6: Local sinkholing can be used to identify infected hosts.**

## **Securing Virtualized Data Centers with Next-Generation Firewalls**

Virtualization is helping organizations utilize their data center hardware infrastructure more effectively, leading to reduction in costs, and improvements in operational efficiencies. Gartner estimates that almost 50% of all x86 server workloads are virtualized today with this number expected to grow to 77% in 2015. Many organizations are also evolving their virtualization infrastructure to build their own automated, self-service, private cloud environments.

As organizations evolve from traditional data centers to virtualized and cloud computing environments, security architectures must support the changing set of requirements. This includes not only addressing fundamental table stakes functionality such as safe application enablement, threat protection and flexible networking integration, but also new challenges brought on by the virtualized infrastructure, and the dynamic and automated nature of the virtualized environment. These

include having visibility into virtual machine traffic that may not leave the virtual infrastructure, the ability to tie security policies to virtual machine instantiation and movement, and orchestration of security policies in lock step with application workflows.

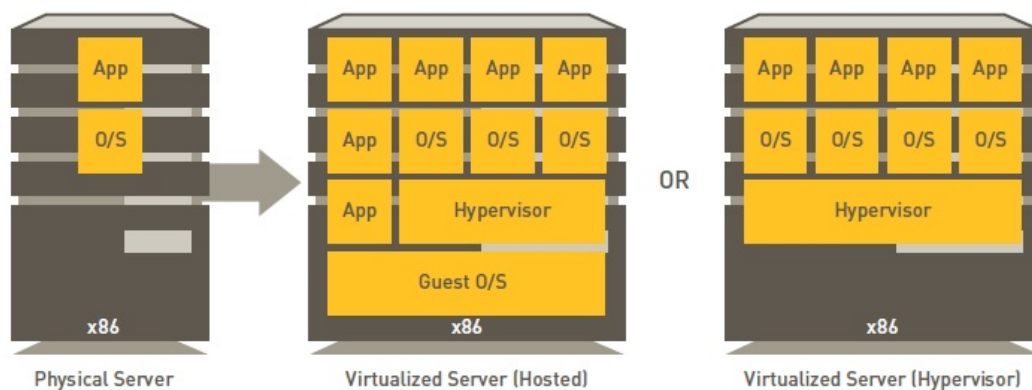
### **The Virtualization and Cloud Computing Evolution**

Today's IT organizations are increasingly tasked with doing more with less. In these challenging economic conditions, IT organizations are faced not only with shrinking budgets but are being asked to improve operational efficiencies and drive responsiveness for business processes. For many IT organizations, the adoption of technologies like virtualization and cloud computing provide many benefits from operational efficiencies to speed in application delivery.

### **Virtualization**

Virtualization technology partitions a single physical server into virtual machines running multiple operating systems and applications. The hypervisor, a software layer that sits between the hardware and the "virtual" operating system and applications, is what allocates memory and processing resources to the "virtual" machines.

Two types of virtualization are available—hypervisor virtualization and hosted virtualization. In hypervisor architectures, also known as bare metal or native virtualization, the hypervisor is the first layer of software running on the underlying hardware without a host operating system. In hosted virtualization, the hypervisor runs on top of the host operating system. This configuration supports the broadest range of hardware operating system including Windows, Linux or MacOS.



**Figure 3-7: Virtualization architectures**

Figure 3-7 shows both architectures. Server virtualization typically utilizes hypervisor architectures while desktop virtualization uses hosted virtualization architectures.

### Why Server Virtualization?

Most data center virtualization initiatives begin with the consolidation of data centers running applications on dedicated, purpose-built servers into an optimized number of data centers with applications on standardized virtualized servers. Server virtualization improves operational efficiencies and lowers capital expenditure for organizations:

- **Optimizes existing hardware resources.** Instead of a “one server, one application” model, multiple virtual applications can be run on a single physical server. This means that organizations can leverage their existing hardware infrastructure by running more applications within the same system.
- **Reduces data center costs.** Reducing the server hardware “box” count

not only reduces the physical infrastructure real-estate but also reduces data center related costs such as power, cooling and rack space.

- **Gain operational flexibility.** Through the dynamic nature of virtual machine provisioning, applications can be delivered quicker rather than the process of purchase, “rack/stack”, cabling, O/S configuration. This helps improve the agility of the IT organization.
- **Maximizes efficiency of data center resources.** Because applications can experience asynchronous, or bursty demand loads, virtualization provides a more efficient way to address resource contention issues and maximize server utilization. It also provides a better way to deal with server maintenance and backup challenges. For example, IT staff can migrate virtual machines to other virtualized servers while performing hardware or software upgrades.

### Why Cloud Computing?

Virtualization is often the first step in an organization’s strategy to move towards automated, on-demand services. Cloud, unlike common misconceptions, is not a location but rather a pool of resources that can be rapidly provisioned. The U.S. National Institute of Standards and Technology (NIST) defines cloud computing in Special Publication (SP) 800-145 as “a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (such as networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.”

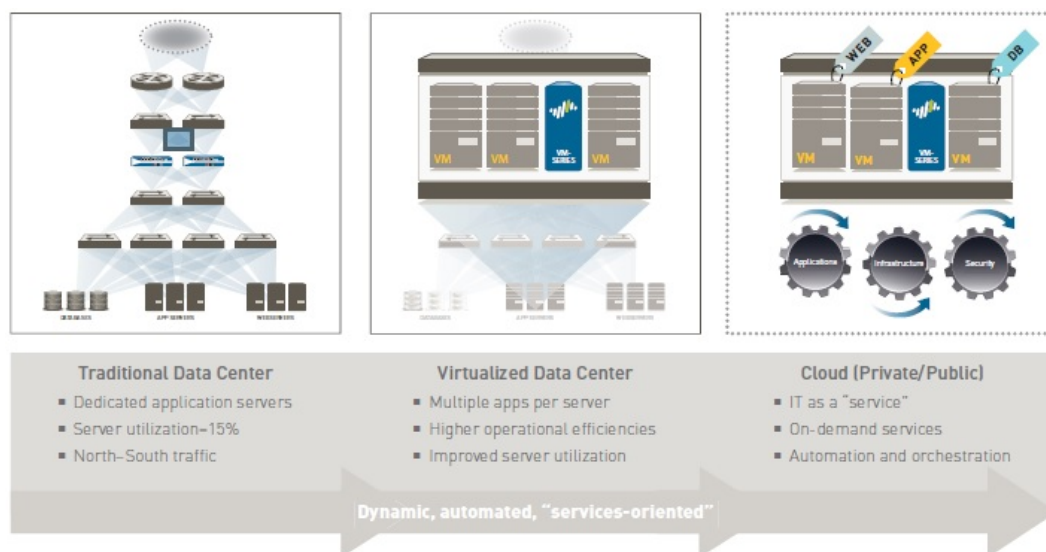
The business value of cloud computing is the ability to pool resources

together to achieve economies of scale. This is true for private or public clouds. Instead of multiple organizations or groups within an organization independently building a data center infrastructure, pools of resources are aggregated and consolidated, and designed to be elastic enough to scale with organizational demand. This not only brings cost and operational benefits but technology benefits. Data and applications are easily accessed by users no matter where they reside, projects can scale easily, and consumption can be tracked effectively.

Virtualization is a critical part of this architecture, enabling applications to be delivered efficiently, and in a more dynamic manner. However, another critical aspect of cloud computing is software orchestration that enables disparate processes to be stitched together in a seamless manner, so that they can be automated, easily replicated and offered on an as-needed basis. The IT organizational model also needs to evolve towards a “services-centric”, multi-tenant model, where consumption needs to be measured, and segmentation between multiple tenants needs to be provisioned.

### **Security Considerations in Securing the Journey to the Cloud**

With virtualization and cloud technologies, the data center environment has evolved from rigid, fixed environments where applications run on dedicated servers towards dynamic, automated, orchestrated environments where pools of computing resources are available to support any application to be accessed anywhere, anytime, from any device (see Figure 3-8).



**Figure 3-8: Evolution of data center architecture**

Security is the biggest hurdle to embrace this new dynamic, automated, services-oriented architecture. The process to configure network security appliances today is excruciatingly painful and slow. Policy changes need to be approved, the appropriate firewalls need to be identified, and the relevant ports and protocols determined. While the creation of a virtual workload may take minutes, the security configuration for this workload may take weeks.

### **Dynamic Nature of Virtualization and Cloud**

Security also cannot keep up with the dynamic nature of virtualization and cloud. Virtual machines can be highly dynamic, with frequent add, move and change operations. This complicates the ability to track security policies to virtual machine creation and movement so that requirements and regulatory compliance continue to be met. Virtualized computing

environments also enable direct communication between virtual machines within a server. Intra-host communications may not be visible to network-based security appliances residing outside a virtual server. The routing of intra-host virtual machine traffic to external security appliances for inspection may not be ideal because of performance and latency requirements.

At the same time, the existing trends that have impacted the security landscape in the virtualized data center —changing application landscape, distributed enterprise, and modern threats—do not go away. The changing application landscape means that the identification, control and safe enablement of applications can no longer be accomplished via ports and protocols. The distributed enterprise of mobile users and extended enterprise, and the evolution of threats towards sophisticated, multi-vector, targeted attacks require user-based policies and a complete threat framework. In summary, next-generation firewalling capabilities to safely enable applications, protect against all known and unknown threats without performance impact, and integrate flexibly into the data center continue to be critical, fundamental security requirements.

Therefore, security for the virtualized data center must exhibit the following characteristics:

- Deliver all the features that are table stakes: These include safe application enablement, threat protection without impacting the performance of the data center, and flexible integration into the data center design. These features must be available within a virtualized firewall to secure intra-host communications or East-West traffic.

- Must become more dynamic:
- Security policies must be applied as soon as a virtual machine is created.
- Security policies must follow virtual machine movement.
- Security workflows must be automated and orchestrated so it doesn't slow down virtual workload provisioning.
- Centralized, consistent management: Centralized management is critical, and must be consistent for all environments—physical, hybrid or mixed environments. The management configuration must provide one unified policy rule base for ease of configuration and complete visibility into the policies being enabled in the data center. In fact, Gartner advocates that organizations “favor security vendors that span physical and virtual environments with a consistent policy management and enforcement framework.”

### **Existing Security Solutions in the Data Center Do Not Deliver**

Existing security solutions in the data center make their access control decisions based on ports and protocol. Many security solutions also bolt on application control and threat prevention features to their stateful inspection firewalls.

There are several problems with this approach. The lack of visibility into all traffic means that evasive applications, applications that use non-standard ports and threats that leverage the same behavior as applications may be missed. Security policies also become convoluted as you build and manage a firewall policy with source, destination, user, port and action, an application control policy with similar rules, in addition to



other threat prevention rules. Policy gaps appear and grow because of the difficulty in managing and monitoring multiple appliances. A multiple policy rule base approach not only increases administrative overhead, but may increase business and security risks with policy gaps that may be hard to see. This multi-platform or multi-module approach also degrades data center performance as more and more features are enabled.

Finally, existing security solutions in the data center do not address the dynamic nature of the virtualized environment, and cannot track policies to virtual machine creation or movement.

Many virtualized security offerings are virtualized versions of port- and protocol-based security appliances, delivering the same inadequacies as their physical counterparts.

### **Mobile Device Considerations**

In recent years, the landscape for mobile devices has changed in a dramatic fashion. No longer limited in capability and functionality, the modern smartphone and tablet are becoming a constant companion (and in some cases, a replacement) for the traditional laptop computer. In fact, with more of these devices connecting directly to the network, security teams are facing growing concerns about the limitations of what they can and cannot protect.

Without a complete picture of what's happening on the network, many organizations assume a fear-based policy towards mobility. Instead of embracing the positive benefits of mobility and the Bring Your Own Device (BYOD) trend—such as increased productivity and higher morale—security teams attempt to block devices from their network

hoping to recreate the relatively controlled environment that once existed. Such efforts are often limited in scope, and fail to deliver an appropriate balance between what the business needs and what users want. Worse still, such efforts can create a perpetual cat-and-mouse game between security teams trying to police the network and users trying to evade detection, rather than allowing security teams to focus on the real threats outside the enterprise. Such a myopic approach to mobility can also create a dangerous subculture in which critical business processes are performed on personal technologies and only known to a few individuals.

Using a next-generation firewall with VPN capabilities, organizations can extend network security and data protection to mobile devices, including smartphones and tablets. Mobile devices running VPN client software can automatically connect to the best available VPN gateway on a next-generation firewall. As a result, organizations can consistently enforce security policies based on application, user, content and device, regardless of where the user is located. Through the safe enablement of applications, users can access business and productivity tools while enjoying protection from mobile threats to the device and data. By placing the security in the network and leveraging an always-on connection to the next-generation firewall, enterprises can now let employees take full advantage of the mobile device of their choice without security compromises.

### **Benefits and Limitations of Existing Approaches**

When looking at the landscape for mobile device security, there are a number of approaches that are available. There are container and virtual

desktop infrastructure (VDI) technologies that isolate data. There are many mobile device management (MDM) products designed to manage the settings on a device. There are legacy VPN products which were originally designed for the remote access use case scenarios. Each one of these pieces provides an element for a mobility solution, but it's important to understand the role, scope and limitations of what each can do.

### **Containers and Virtual Desktops Infrastructure (VDI)**

Containers and virtual desktop infrastructure (VDI) provide a method to isolate data. Containers can partition data on a device in a sandbox. VDI allows users to access a desktop remotely, thus separating the entire desktop from the device itself. In both cases, there is an implicit assumption that the only sensitive data on the device is in the container/virtual desktop. However, users may use other productivity applications on the device, which saves the sensitive data in other, less secure storage areas on the device.

### **Mobile Device Management (MDM)**

Mobile device management (MDM) provides the means to configure certain mobile device settings and provision the device for use. MDM may be considered a baseline requirement for managed devices, but there are additional considerations that one must also add to address security.

### **Traditional SSL VPN**

An SSL VPN provides temporary, secure access to a corporate network. However, it does not provide any network security controls, such as enforcing appropriate application usage, blocking undesirable traffic and preventing inappropriate file sharing. As a result, an SSL VPN provides the

means for mobile devices to access corporate applications without the necessary security controls to protect the device.

In addition to the technologies listed above, there are a number of security products designed for network access, such as network access control and wireless network security, which are often used as part of comprehensive enterprise mobility strategy. There are also stacks of traditional network security products such as the stateful inspection firewall, IPS, and proxies, which are managed in separate contexts and disjointed policies. With a bewildering list of technologies to consider, developing a comprehensive mobile strategy can pose a formidable challenge.

### **Safely Enabling Mobile Devices**

To safely enable mobile devices, you need to:

- Protect traffic
- Protect data
- Ensure the device is OK

Through a combination of technologies, an organization can ensure that only authorized devices have access to sensitive information, maintain consistent security throughout the organization, and protect mobile devices from vulnerabilities and malware.

### **Provide Next-Generation Network Security with an Always-on Connection**

The network is the link between enterprise applications, data and users.

It provides the obvious location for providing policy enforcement and safe enablement of devices. However, the traditional firewall does not differentiate between users, nor does it identify applications. It simply permits any traffic allowed on a particular network segment to pass as long as it follows basic port-based policy guidelines. Mobility exposes this pre-existing condition, as users are able to access anything they want once they are able to get their device on the network. Traditional firewalls and other security solutions do not provide visibility and control.

### **Remote Users Pose Challenges**

Remote users pose a second set of challenges because network security is ephemeral with the traditional VPN. Once the user disconnects from the corporate wireless network or the VPN, the user has a direct path to the Internet without any security in the network traffic path and outside of the jurisdiction of safe practices.

### **Pair Up Next-Generation Firewalls with VPN Gateway**

The next-generation firewall paired with a VPN gateway solves both conditions. Instead of treating all network traffic in the same generic manner, the next-generation firewall provides the means to classify traffic by application, user and content. Thus the protection lies within the network, ensuring that policy enforcement is always in place, rather than being dependent upon blocking technologies to keep mobility off. The VPN provides a secure connection in a manner that's not predicated on temporary connections, but rather an always-on connection to the corporate network, regardless of location. Whether in a hotel room or in the office, the user stays on the network with the same policy enforcement. This approach provides consistency for the protection and

safe application enablement capabilities of the next-generation firewall, regardless of whether the user is in the office or on the road.

## Mobile Protection

All of these technologies work together to provide the foundation for mobile protection that extends to all users. With a strong foundation for network security, an organization can provide a choice in devices with an approach that welcomes change rather than resisting it.

## Next-Generation Firewalls provide Protection

The next-generation firewall provides a number of protections for traffic in order to provide safe enablement. Some of the key capabilities include:

- **Application policy.** Ensure that users have access to the proper applications while removing dangerous or risky elements.
- **URL filtering.** Restrict access—by specific website or by category—to content that may be inappropriate or unauthorized. For instance, an organization may want to make sure that apps can only be downloaded from authorized app stores, while blocking all others.
- **Malware protection.** The next-generation firewall analyzes traffic for malicious content, providing the means to stop dangerous files before they reach devices on the network. Many enterprises do not have antivirus clients running on their mobile devices, which only reinforces the importance of stopping malware in the network. The next-generation firewall scans content that endangers mobile platforms, thus providing the device with protection that's always in place.
- **Vulnerability protection.** Mobile devices pose a special challenge for

organizations in terms of maintaining protection against newly discovered vulnerabilities in the operating system. Due to the vast number of devices in use, and the inconsistent application of operating system updates, it's not easy for an organization to ascertain just how much risk they face against a particular threat. By placing vulnerability protection in the network, the next-generation firewall intercepts an exploit before it reaches the user's device, thus providing protection even in advance of patch installation.

## Data Protection

One of the principal concerns about mobility is the risk of corporate data being placed outside of controlled boundaries, namely a mobile device which may or may not be owned by the organization or in applications outside of its control. One approach to solve this challenge is to leverage an always-on connection to extend the boundary of protection to all locations. This concept, the logical perimeter, differs from a physical perimeter in that it doesn't require the user to be on-premise to benefit from its protection.

The next-generation firewall includes file and data filtering technology to protect data and is applied to all devices on the network, including mobile devices. Data filtering features enable administrators to implement policies that will reduce the risks associated with the transfer of unauthorized files and 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).

- **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 while preventing undesired inbound or outbound file transfer.

For organizations with highly sensitive data, and with regulatory and compliance requirements, the implementation of containers or Virtual Desktop Infrastructure (VDI) can isolate such data from the rest of the device. A container takes the approach of using a sandbox on the device to separate certain data from the rest of the device. VDI keeps the application and the data within the data center, and provides access through a client on various platforms.

The next-generation firewall can add further protections such as restricting access to the data center to allow only virtual desktops, thus providing access to the application without the risk of placing data on the mobile device itself.

All of the measures above work in conjunction with device management, should it be necessary to remotely wipe a device that's been lost or de-provisioned.

## Device Management

Device management plays a role in a mobility solution as it establishes the fundamental profiles that govern device settings and device state. In a nutshell, it provides the means to ensure that the device is appropriate for



use in a manner consistent with the organization's policies and to bring it under management. Some of the things that device management controls include settings for passcode requirements, remote wiping and device wiping after a number of failed unlock attempts.

“Jailbreaking” (for Apple iOS devices) or “rooting” (for Google Android devices) a mobile device removes code signing requirements and bypasses numerous other native security protections in a mobile iOS. With a device management solution, organizations can check to see if a device has been jailbroken or rooted and use the device state as part of its security policy decisions.

## Summary

- *Forrester's Zero Trust Model calls for segmentation and least-privilege network access, regardless of a user location or role, such that users are never fully trusted and always verified.*
- *The next-generation network security model calls for reduction of attack surface, investigation of unknown traffic and prevention of circumvention for users inside and outside the physical network.*
- *Cloud services and virtualized data center environments offer new challenges to securing data mainly due to the highly dynamic nature of those environments.*
- *Mobile devices, and their connections back to the enterprise, should be provided the same levels of protection as internal devices using methods such as mobile device management and virtual private networks.*

## Discussion

1. What stages are present in a Security Systems Development Lifecycle (SecSDLC), and who are some of the publishers of such lifecycle models?
2. This text presents two security models. At which phase of the SecSDLC would these be applied?
3. Cloud computing is in an explosive growth mode, for its scalability, accessibility, and relative ease of management. What are some of the challenges presented with securing such environments? How are those different from securing traditional LAN/WAN environments?
4. Consider your own mobile device and its applications. What risks would be introduced if your device was admitted to an enterprise network? Which applications might pose the greatest risks?
5. IT Security must always balance usability and access with security and control. What would be some examples of imbalance, where IT security might induce users to attempt to circumvent enterprise controls?

## Knowledge Check

1. Related to the Zero-Trust security model, choose all that apply:
  - A)** If an attacker can remotely control a machine behind an organization's security perimeter, they are then considering 'insiders' to the network.
  - B)** Trust refers to, and applies to, the inside of a network. Untrust refers to, and applies to, the outside internet facing side of a network
  - C)** A customer service representative residing physically and logically inside the network is accessing an internal database, housed in the same network and building. This is a trusted connection
  - D)** The least privilege strategy must adopted, enforced, and verified
2. Name three benefits to adopting the Zero Trust security model.
3. What trait distinguishes botnets and modern malware from earlier forms of malware?
4. True or False: Reduction of attack surface implies a return to positive control model, where only specific applications, for specific users are allowed to traverse a network.
5. Unknown traffic, whether TCP, UDP, or P2P, represents which 3 of the following scenarios?
  - A)** Custom built in-house application traffic.
  - B)** Incomplete packets or prematurely terminated sessions.
  - C)** Newly available commercial applications which have yet to have application identifiers built.
  - D)** Malicious traffic.
6. Name three indicators on the network that may point to a host machine being compromised.

7. The most efficient and effective way to secure a virtualized data center is to use which of the following methods?
  - A)** Server whitelisting ensuring only known good apps can run.
  - B)** Dynamic, automated, and centrally manageable security that delivers full-featured threat prevention to and between hosts.
  - C)** Host Intrusion Prevention in combination with access control lists.
8. True or False: Virtualized data centers and cloud computing environments must be secured in very different ways to accommodate their respective architectures
9. Name three approaches to securing mobile devices with access to enterprise resources.
10. True or False: A principal focus around securing mobility is protecting the data on the mobile device.

## 4. Cybersecurity Solutions from Palo Alto Networks

### Next-Generation Firewall Technologies

The following module explains various core capabilities and technologies found in next-generation firewalls. Although this discussion is specific to technologies and capabilities used in Palo Alto Networks next-generation firewalls, it provides a foundation for technologies and capabilities that define next-generation firewalls in general.

#### Application Identification

As the foundational element of a next-generation firewall, application identification provides visibility and control over work-related and non-work-related applications that can evade detection by masquerading as legitimate traffic, hopping ports or sneaking through the firewall using encryption (SSL and SSH).

In the past, unapproved or non-work-related applications on the

corporate network were summarily removed or blocked. However, in today's business environment, the response options are not nearly as clear because many of the same applications are helping employees get their jobs done.

Application identification enables administrators to see the applications on the network, learn how they work, their behavioral characteristics, and their relative risk. When used in conjunction with user identification (discussed later in this module, administrators can see exactly who is using the application based on their identity, not just an IP address.

Armed with this information, administrators can use positive security model rules to block unknown applications, while enabling, inspecting and shaping those that are allowed.

### **Firewall Traffic Classification: Applications, not Ports**

Stateful inspection, the basis for most of today's firewalls, was created at a time when applications could be controlled using ports and source/destination IPs. The strict adherence to port-based classification and control methodology is the primary policy element; It is hard-coded into the foundation and cannot be turned off.

This means that many of today's applications cannot be identified, much less controlled by the firewall and no amount of "after the fact" traffic classification by firewall helpers can correct the firewall port-based classification.

Application identification technology in next-generation firewalls does not rely on any one single element like port or protocol. Instead, application identification uses multiple mechanisms to determine what the application is, first and foremost, and the application identity then becomes the basis for the firewall policy.

Application identification is highly extensible and as applications continue to evolve, application detection mechanisms can be added or updated as a means of keeping pace with the ever-changing application landscape.

### **App-ID™ Traffic Classification Technology**

The first task that a Palo Alto Networks next-generation firewall executes is the identification of the applications traversing the network using App-ID™. Using as many as four different techniques, App-ID™ determines what the application is, irrespective of port, protocol, encryption (SSL and SSH) or other evasive tactics employed (see Figure 4-1). The number and order of identification mechanisms used to identify the application will vary depending on the application. The general flow is as follows:

#### **Application Signatures**

Signatures are used first to look for unique application properties and related transaction characteristics to correctly identify the application regardless of the protocol and port being used. The signature also determines if the application is being used on its default port or it is using a non-standard port (for example, RDP across port 80 instead of port 3389, its standard port). If the identified application is allowed by security policy, further analysis of the traffic is done to identify more granular applications as well as scan for threats.

## SSL and SSH Decryption

If App-ID™ determines that SSL encryption is in use and a decryption policy is in place, the traffic is decrypted and then passed to other identification mechanisms as needed. If no policy is in place, then SSL decryption is not employed. Once the application is identified, and deemed acceptable by policy, threat prevention profiles are applied and the traffic is then delivered to its destination. A similar approach is used with SSH to determine if port forwarding is in use as a means to tunnel traffic over SSH. Such tunneled traffic is identified as ssh-tunnel and can be controlled via security policy.

## Application Protocol Decoding

Decoders for known protocols are used to apply additional context-based signatures to detect other applications that may be tunneling inside of the protocol (e.g., Yahoo! Instant Messenger used across HTTP). Decoders validate the traffic conforms to the protocol specification and provide support for NAT traversal and opening dynamic pinholes for applications such as VoIP or FTP. Decoders for popular applications are used to identify the individual functions within the application as well (e.g., webex-filesharing). In addition to identifying applications, decoders also identify files and other content that should be scanned for threats or sensitive data.

## Heuristics

In certain cases, evasive applications still cannot be detected even through advanced signature and protocol analysis. In those situations, it is necessary to apply additional heuristics, or behavioral analysis to identify certain applications such as peer-to-peer file-sharing or VoIP applications



that use proprietary encryption. Heuristic analysis is used as needed, with the other App-ID™ techniques discussed here, to provide visibility into applications that might otherwise elude positive identification. The actual heuristics used are specific to an application and include checks based on such things as the packet length, session rate, packet source, etc.

With App-ID™ as the foundational element for every Palo Alto Networks next-generation firewall, administrators can regain visibility into, and control over, the applications traversing the network.

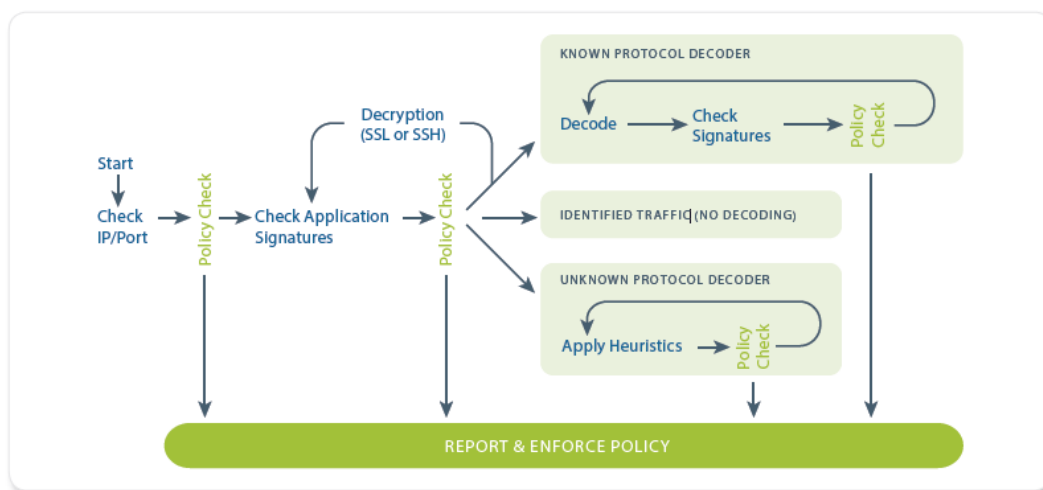


Figure 4-1: How Palo Alto Network App-ID™ classifies traffic

### App-ID™: Dealing with Custom or Unknown Applications

Palo Alto Networks adds an average of five new applications to App-ID™ each week, yet there are cases where unknown application traffic will be detected. There are typically two scenarios where unknown traffic will appear: a commercially available application that does not have an App-ID™ or an internal, custom application is in use.

### Unknown Commercial Applications

Using ACC™ (Application Command Center) and the log viewer, users can quickly determine that the application is used commercially or not. Using the packet capture feature on the Palo Alto Networks firewall, customers can record the traffic and submit it for App-ID™ development. The new App-ID™ is developed, tested with the customer, then added to the database for all users in the form of a weekly update.

### Internal or Custom Applications

Once it has been determined with ACC™ and the log viewer, that the application in question is internal or custom, then you have several options. First off, an application override can be applied, effectively renaming the application. Alternatively, you can develop a custom App-ID™ for the application using the exposed protocol decoders.

The protocol decoders that have been exposed include: FTP, HTTP, HTTPs (SSL), IMAP, SMTP, RTSP, Telnet, unknown-TCP, unknown-UDP, and file body (for html/pdf/flv/swf/riff/mov). Once developed, traffic identified by the custom App-ID™ is treated in the same manner as the previously classified traffic; it can be enabled via policy, inspected for threats, shaped using QoS and so on. Custom App-ID™s are managed in a separate database on the device, ensuring they are not impacted by the weekly App-ID™ updates.

An important point to highlight is that Palo Alto Networks next-generation firewalls use a positive enforcement model, which means that all traffic can be denied except those applications that are expressly allowed via policy. This means that in some cases, the unknown traffic can be easily blocked or tightly controlled. Alternative offerings that are

based on IPS will allow unknown traffic to pass through without providing any semblance of visibility or control.

### **An Example of How App-ID™ Works: Identifying WebEx**

When a user initiates a WebEx session, the initial connection is an SSL-based communication. With App-ID™, the device sees the traffic and the signatures determine that it is using SSL. The decryption engine and protocol decoders are then initiated to decrypt the SSL and detect that it is HTTP traffic. Once the decoder has the HTTP stream, App-ID™ can apply contextual signatures and detect that the application in use is WebEx.

WebEx is then displayed within ACC™ and can be controlled via a security policy. If the end user were to initiate the WebEx Desktop Sharing feature, WebEx undergoes a “mode-shift” to where the session has been altered from a conferencing application to a remote access application.

In this scenario, the characteristics of WebEx have changed and App-ID™ will detect the WebEx Desktop Sharing feature which is then displayed in ACC™. At this stage, an administrator has learned more about the application usage and can exert policy control over the use of the WebEx Desktop Sharing feature separately from general WebEx use.

### **Application Identity: The Heart of Policy Control**

Identifying the application is the first step in learning more about the traffic traversing the network. Learning what the application does, the ports it uses, its underlying technology, and its behavioral characteristics is the next step towards making a more informed decision about how

to treat the application. Once a complete picture of usage is gained, organizations can apply policies with a range of responses that are more fine-grained than allow or deny. Examples 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 (Quality of Service)
- Apply policy-based forwarding
- Allow certain application functions
- Any combination of the above

### **Application Function Control**

For many organizations, secure application enablement means striking an appropriate security policy balance by enabling individual application functionality while blocking other functions within the same application. Examples may include:

- Allowing SharePoint Documents, but blocking the use of SharePoint Administration.
- Block Facebook-mail, -chat, -posting and -apps, but allow Facebook itself, effectively only allowing users to browse Facebook.
- Enable the use of MSN, but disable the use of MSN-file transfer and only allow certain file types to be transferred using the file blocking feature.

Using an application hierarchy that follows a container and supporting function model, App-ID™ makes it easy for administrators to choose which applications to allow, while blocking or controlling functions within the application. Figure 4-2 shows SharePoint as the container application, and the individual functions within.

### **Application Function Control Table**

**NOTE:** The figure that is to follow may appear on the next page. Please print the page for clarity in viewing the image and its details.

**paloalto NETWORKS**

Dashboard ACC Monitor Policies Objects Network Device

Save Help

Name	Source		Destination		Application	URL Category	Service	Action	Profile
	Zone	Address	User	Zone					
Logall	DMZ Tap	any	any	DMZ Tap	any	Customer/URL Category	any	✓	Logall
IT Allow Override	DMZ trust	any	pancademo/administrators	DMZ untrust	Custom-app	any	any	✓	Logall
Read Only Facebook	DMZ trust	any	pancademo/administrators	DMZ untrust	facebook-base	any	any	✓	Logall
Allow facebook posting	DMZ trust	any	pancademo/marketing	DMZ untrust	facebook-posting	any	any	✓	Logall
Block Peer to Peer	DMZ trust	any	any	DMZ untrust	Peer to Peer	any	any	✗	none
Webmail file blocking	DMZ trust	any	any	DMZ untrust	Webmail	any	any	✓	Logall
Sharepoint	DMZ Untrust-L3	any	any	DMZ DMZ	Sharepoint Server	sharepoint-blog-posting	any	✓	Logall
					sharepoint-calendar	sharepoint-documents			
					sharepoint-wiki				
Allow SSL and SSH	DMZ trust	any	pancademo/domain admins	DMZ untrust	ssh	any	any	✓	Logall
Allow Web-browsing	DMZ trust	Sharepoint Server	any	DMZ untrust	ssl	any	any	✓	Logall
Block encrypted tunnel	DMZ trust	any	any	DMZ untrust	web-browsing	any	any	✗	none
Block Proxies and Anonymizers	DMZ trust	any	any	DMZ untrust	Encrypted Tunnel	any	any	✗	none
Mail server	DMZ Untrust-L3	any	any	DMZ DMZ	Proxies	any	any	✗	none
					Mail Server FQDN	outlook-web	application-default	✓	Logall
Web server	DMZ Untrust-L3	any	any	DMZ DMZ	snmp	any	application-default	✓	Logall
					Web-server	any	application-default	✓	Logall
					ssl				
					web-browsing				

+ Add - Delete Clone Enable Disable Move Top Move Up Move Down Move Bottom Highlight Unused Rules

13 rule(s)

Figure 4-2: Application Function Control maximizes productivity by safely enabling the application itself (Microsoft SharePoint) or individual functions.

## **Controlling Multiple Applications: Dynamic Filters and Groups**

There are many cases where customers may want to control applications “in bulk”, as opposed to controlling them individually. The two mechanisms that address this need are application groups and dynamic filters.

### **Application groups**

A group of applications is a static list of applications that can be used to enable use for certain users while blocking their use for others. An example may be the use of remote management applications such as RDP, Telnet, and SSH. Each of these applications are known to be used by support and IT personnel, yet employees that fall outside of these groups are also known to use them as a means of accessing their home networks. A group of applications can be created and assigned to IT and support through User-ID™ (discussed later in this module), tying the groups to the policy. As new employees are added, they only need to be added to the directory group. No updates are needed to the policy itself.

### **Dynamic filters**

A dynamic filter is a set of applications that is created based on any combination of the filter criteria: category, subcategory, behavioral characteristic, underlying technology and risk factor. Once the desired results for the filter are achieved, a policy that blocks or enables and scans the traffic can be applied. As new App-ID™ that fulfills the filter criteria is added in the weekly content updates, the filter is automatically updated as soon as the device is updated, thereby minimizing the administrative effort associated with policy management.

### Filter Category and Subcategory options:

- Business: Authentication services, database, ERP, general management, office programs, software updates, storage/backup
- General Internet: File sharing, Internet utilities (web-browsing, toolbars, etc.)
- Collaboration: Email, instant messaging, Internet conferencing, social networking, social business, VoIP/video, web posting
- Media: Audio streaming, gaming, photo/video
- Networking: Encrypted tunnel, infrastructure, IP protocol, proxy, remote access, routing
- Able to transfer files from one network to another
- Used to propagate malware
- Consumes 1Mbps or more regularly through normal use
- Evades detection using a port or protocol for something other than its intended purpose or intent
- Has been widely deployed
- Application has had known vulnerabilities
- Prone to misuse or is easily configured to expose more than intended
- Tunnels other applications
- Client-server based
- Browser-based
- Peer-to-peer based



- Network protocol

## **Application Behavioral Characteristics**

### **Underlying Application Technology**

- Client-server based
- Browser-based
- Peer-to-peer based
- Network protocol

## **Content Identification**

Enterprises of all sizes are at risk from a variety of increasingly sophisticated network-borne threats that have evolved to avoid many of the industry's traditional security measures. Content identification is a new approach to traffic control based on the complete analysis of all allowed traffic using multiple threat prevention and data-loss prevention techniques in a single unified engine.

Unlike traditional solutions, content identification actually controls the threat vectors themselves through the tight control of all types of applications. This immediately reduces the attack surface of the network, after which all allowed traffic is analyzed for exploits, malware, dangerous URLs, dangerous or restricted files or content. Content identification then goes beyond stopping known threats to proactively identify and control unknown malware, which is often used as the leading edge of sophisticated network attacks.

Palo Alto Networks' Content-ID™ is built on a single-pass architecture, which is a unique integration of software and hardware that simplifies management, streamlines processing and maximizes performance. The single-pass architecture (SP3) integrates multiple threat prevention disciplines (IPS, anti-malware, URL filtering, etc.) into a single stream-based engine with a uniform signature format. This allows traffic to be fully analyzed in a single pass without the incremental performance degradation seen in other multi-function gateways. The software is tied directly to a parallel processing hardware platform that uses function specific processors for threat prevention to maximize throughput and minimize latency.

### **Threat Prevention**

Enterprise networks are facing a rapidly evolving threat landscape full of modern applications, exploits, malware and attack strategies that are capable of avoiding traditional methods of detection. Threats are delivered via applications that dynamically hop ports, use non-standard ports, tunnel within other applications or hide within proxies, SSL or other types of encryption. These techniques can prevent traditional security solutions such as IPS and firewalls from ever inspecting the traffic, thus enabling threats to easily and repeatedly flow across the network. Additionally, enterprises are exposed to targeted and customized malware, which may pass undetected through traditional antivirus solutions.

Palo Alto Networks' Content-ID™ addresses these challenges with unique threat prevention capabilities not found in traditional security solutions. First, the next-generation firewall removes the methods that

threats use to hide from security through the complete analysis of all traffic, on all ports regardless of evasion, tunneling or circumvention techniques. Simply put, no threat prevention solution will be effective if it does not have visibility into the traffic. Palo Alto Networks ensures that visibility through the identification and control of all traffic.

- **Application decoders.** Content-ID™ leverages the more than 100 application and protocol decoders in App-ID™ to look for threats hidden within streams of application data. This enables the firewall to detect and prevent threats tunneled within approved applications that would pass by traditional IPS or proxy solutions.
- **SSL decryption.** More and more web traffic is encrypted with SSL by default. This can provide some protection to end-users, but it also can provide attackers with an encrypted channel to deliver exploits and malware. Palo Alto Networks ensures visibility by giving security organizations the flexibility to, by policy, granularly look inside of SSL traffic based on application or URL category.
- **Control of circumventing technologies.** Attackers and malware have increasingly turned to proxies, anonymizers and a variety of encrypted proxies to hide from traditional network security products. Palo Alto Networks provides the ability to tightly control these technologies and limit them to approved users, while blocking unapproved communications that could be used by attackers.
- **Uniform threat signature format.** Rather than use a separate set of scanning engines and signatures for each type of threat, Content-ID™ leverages a uniform threat engine and signature format to detect and

block a wide range of malware including viruses, spyware, and vulnerability exploits in a single pass.

### **Multiple Security Disciplines**

Secondly, Palo Alto Networks brings multiple security disciplines into a single context and a single threat prevention engine. This context enables security teams to easily see beyond individual security events and recognize the full extent of a threat. Security managers can now see the interconnection of applications, exploits, malware, URLs, anomalous network behaviors and management and reporting and ensures predictable performance by analyzing traffic once instead of progressive scanning in multiple engines.

### **Integrated by Design**

Palo Alto Networks next-generation firewalls are purpose-built platforms that utilize a single pass parallel processing architecture to maximize throughput and minimize latency. Traditional blade or UTM architectures notoriously introduce performance penalties for each feature that is enabled due to repeatedly processing traffic for each blade or feature. Palo Alto Networks designed a unique approach that performs Content-ID™ in a single unified engine and leverages a common signature format. This means that content is processed only once, and performance remains steady even as additional Content-ID™ features are enabled.

### **Single Pass Software**

The single pass software uses a stream-based, uniform signature-matching engine for content inspection. Instead of using separate engines and signature sets (requiring multi-pass scanning) and instead of using

file proxies (requiring file download prior to scanning), the single pass architecture scans traffic for all signatures once and in a stream-based fashion to avoid latency introduction.

### **Stream-Based Engine**

The use of a stream-based engine replaces several components commonly used in other solutions—a file proxy for data, virus, and spyware, a signature engine for vulnerability exploits, and an http decoder for URL filtering. By using one common engine, two key benefits are realized. First, unlike file proxies that need to download the entire file before they can scan the traffic, a stream-based engine scans traffic real time, only reassembling packets as needed and only in very small amounts. Second, unlike traditional approaches, all traffic can be scanned with a single engine, instead of multiple scanning engines.

True integration of security functions is achieved through a combination of behavior-only (WildFire) and signature-based (AV, AS, DNS, Malware URLs, and exploit protection) which constantly improve upon themselves through an automated threat intelligence feedback loop. As forensic data is gathered by observation in the WildFire sandbox, or supplied through external threat intelligence feeds, it is used to create new signatures feeding the other security functions. This actionable protection is shared across the entire global install base rapidly, shortening the ‘zero-day window’ to less than an hour.

### **Intrusion Prevention**

Content-ID™ protects networks from all types of vulnerability exploits, buffer overflows, DoS attacks and port scans that lead to the compromise

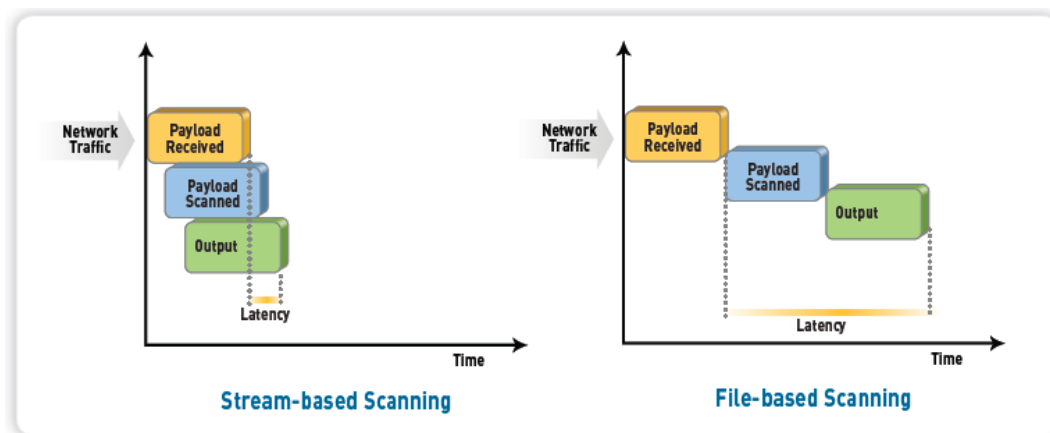
and damage of enterprise information resources. IPS mechanisms include:

- Protocol decoders and anomaly detection
- Stateful pattern matching
- Statistical anomaly detection
- Heuristic-based analysis
- Invalid or malformed packet detection
- IP defragmentation and TCP reassembly
- Custom vulnerability and spyware phone-home signatures

Traffic is normalized to eliminate invalid and malformed packets, while TCP reassembly and IP de-fragmentation is performed to ensure the utmost accuracy and protection despite any packet-level evasion techniques.

### **Stream-based Malware Scanning**

Prevention of known viruses and malware is performed through the use of stream-based scanning, a technique that begins scanning as soon as the first packets of the file are received as opposed to waiting until the entire file is loaded into memory to begin scanning. This means that performance and latency issues are minimized by receiving, scanning, and sending traffic to its intended destination immediately without having to first buffer and then scan the file (see Figure 4-3).



**Figure 4-3: Stream-based scanning helps minimize latency and maximize throughput performance**

Palo Alto Networks maintains an independent database of millions of malware samples, with more than 50,000 samples analyzed daily. Virus and spyware signatures control a wide range of malware including PDF, HTML and Javascript viruses, spyware downloads, spyware phone home, trojans, key-loggers and botnets. Palo Alto Networks provides coverage for signatures for all types of malware are generated directly from millions of live virus samples collected by Palo Alto Networks from several sources including a worldwide network of honeypots deployed around the world, from the WildFire™ malware analysis service and from other leading third-party research organizations around the world. The Palo Alto Networks threat team analyzes the samples and quickly eliminates duplicates and redundancies. New signatures for new malware variants are then generated (using a uniform signature format) and delivered through scheduled daily or emergency updates.

## **WildFire™: Threat Intelligence Cloud to Prevent Unknown Threats**

Criminals have increasingly turned to customized and targeted malware to avoid traditional antivirus controls. Palo Alto Networks has addressed this challenge with WildFire™, which identifies malware by observing the actual behavior of a suspect file in a virtualized environment instead of relying solely on pre-existing signatures.

- **Integration of Firewall and the Cloud**—WildFire™ makes use of a customer's on-premises firewalls in conjunction with Palo Alto Networks cloud-based analysis engine to deliver an ideal blend of protection and performance. The in-line firewall captures unknown files and performs in-line enforcement while maintaining high network throughput and low latency. The analysis of unknown files is offloaded to a secure cloud-based engine to identify unknown malware and subsequently deliver protections to all locations.
- **WildFire™ Virtualized Sandbox**—When the Palo Alto Networks firewall encounters an unknown file (initially portable executable files, and expanding to other file types in the future), the file can be submitted to the hosted WildFire™ virtualized sandbox. Submissions can be made manually or automatically based on policy. The sandbox provides virtual targets for the suspected malware where Palo Alto Networks can directly observe more than 100 malicious behaviors that can reveal the presence of malware.
- **Automated Signature Generator**—When a sample is identified as malware, the sample is then passed on to the signature generator, which automatically writes a signature for the sample and tests it for accuracy.
- **Deep Visibility and Analysis**—In addition to providing protection from



modern malware, users can see a wealth of information about the detected malware in reports available on the WildFire™ Portal. This includes the ability to see all behaviors of the malware: the user that was targeted, the application that delivered the malware, registry changes, file adds/deletes, DNS requests made, and all URLs involved in delivery or phone-home of the malware.

- **Integration with Endpoint Security**—WildFire produces system level forensics which are rich sources of threat intelligence for use by endpoint security products. Bit9 is an example of a next-generation endpoint security company which has integrated with the WildFire cloud to corroborate findings, enforce policy, and remediate infections. Palo Alto Networks acquired next-generation exploit protection company Cyvera, to build out its integrated platform across cloud, endpoint, and network.

## URL Filtering

Complementing the threat prevention and application control capabilities is a fully integrated, on-box URL filtering database that enables security teams to not only control end-user web surfing activities, but also combine URL context with application and user rules.

The on-box URL database can be augmented to suit the traffic patterns of the local user community with a custom, one million URL database. URLs that are not categorized by the local URL database can be pulled into cache from a hosted, 180 million URL database. In addition to database customization, administrators can create custom URL categories to further tailor the URL controls to suit their specific needs.

## URL Categorization

URL categorization can be combined with application and user classification to further target and define policies. For example, SSL decryption can be invoked for select high-risk URL categories to ensure threats are exposed, QoS controls can be applied to streaming media sites, URL filtering visibility and policy controls can be tied to specific users through the transparent integration with enterprise directory services (Active Directory, LDAP, eDirectory) with additional insight provided through customizable reporting and logging.

## Customizing a Custom Block Page

Administrators can configure a custom block page to notify end users of any policy violations. The page can include references to the username, IP address, the URL they are attempting to access and the URL category. In order to place some of the web activity ownership back in the user's hands, administrators can allow users to continue after being presented with a warning page, or can use passwords to override the URL filtering policy.

## File and Data Filtering

Data filtering features enable administrators to implement policies that will reduce the risks associated with the transfer of unauthorized files and 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).
- **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.

### Log Correlation and Reporting

Powerful log filtering enables administrators to quickly investigate security incidents by correlating threats with applications and user identity. The log viewer enables an administrator to click on a cell value to immediately create a filter that can be narrowed down further by combining multiple criteria using an expression builder and additional log fields, even if they are not visible in the log viewer.

To tie the user identity to the threat, the log viewer leverages the integration with enterprise directory services. Log results can be exported to a CSV file for offline archival or further analysis. The trace session tool accelerates forensics and incident investigation with a centralized, correlated view across all of the logs for traffic, threats, URLs, and applications related to an individual session.

### Reporting

Reporting is enabled through a set of predefined reports that can be customized, pulling data from any of the log databases and then saving them for future use. Once the desired report is created, it can be configured to run on a regular basis, emailing a set of PDF reports or exporting them to CSV or PDF.

## User Identification

Compounding the visibility problem in an increasingly mobile enterprise, where employees access the network from virtually anywhere around the world, internal wireless networks re-assign 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.

## User-ID™: Integrating User Information into Security Policies

Creating and managing security policies based on the application and the identity of the user, regardless of device or location, is a more effective means of protecting the network than relying solely on port and IP address. Palo Alto Networks' User-ID™ enables organizations to leverage user information stored in a wide range of repositories for the following uses:

- **Visibility:** Improved visibility into application usage based on user and group information can help organizations maintain a more accurate picture of network activity.
- **Policy control:** Tying user information to the security policy to safely enable applications or specific application functions while reducing the administrative effort associated with employee moves, adds and changes.
- **Logging and reporting:** In the event that a security incident occurs, forensics analysis and reporting can include user information, again, providing a more complete picture of the incident.

## How User-ID™ Works

User-ID™ seamlessly integrates Palo Alto Networks next-generation

firewalls with a wide range of user repositories and terminal services environments. Depending on the network environment, multiple techniques can be configured to map the user identity to an IP address. Events include authentication events, user authentication, terminal services monitoring, client probing, directory services integration and a powerful XML API. Once the applications and users are identified, full visibility and control within ACC™, policy editing, logging and reporting is available.

### Authentication Events

User-ID™ can be configured to monitor authentication events for Microsoft Active Directory, Microsoft Exchange and Novell eDirectory environments. Monitoring of the authentication events on a network allows User-ID™ to associate a user with the IP address of the device the user logs in from to enforce policy on the firewall.

- **Microsoft Exchange Server:** User-ID™ can be configured to constantly monitor the Microsoft Exchange logon events produced by clients accessing their email. Using this technique, even MAC OS X, Apple iOS, Linux/UNIX client systems that don't directly authenticate to Microsoft Active Directory can be discovered and identified.
- **Novell eDirectory:** User-ID™ can query and monitor logon information to identify users and group memberships via standard LDAP queries on the Novell eDirectory servers.
- **Microsoft Active Directory:** User-ID™ constantly monitors domain controller event logs to identify users when they log onto the domain. When a user logs onto the Windows domain, a new authentication event

is recorded on the corresponding Windows Domain Controller. By remotely monitoring the authentication events on Windows Domain Controllers, User-ID can recognize those authentication events to identify users on the network for creation and enforcement of policy.

### User Authentication

This technique allows organizations to configure a challenge-response authentication sequence to collect user and IP address information.

- **Captive portal:** In cases where administrators need to establish rules under which users are required to authenticate to the firewall prior to accessing the internet, a captive portal can be deployed. Captive portal is used in cases where the user cannot be identified using other mechanisms. In addition to an explicit username and password prompt, a captive portal can also be configured to send an NTLM authentication request to the web browser in order to make the authentication process transparent to the user.
- **GlobalProtect™:** Remote users logging into the network with GlobalProtect™ will provide user and host information to the firewall that in turn, can be used for policy control.

### Client probing and terminal services

This technique allows organizations to configure User-ID to monitor Windows clients or hosts to collect the identity and map it to the IP address. In environments where the user identity is obfuscated by Citrix XenApp or Microsoft terminal Services, the User-ID™ Terminal Services Agent can be deployed to determine which applications users are accessing.

- **Client probing:** If a user cannot be identified via monitoring authentication events, User-ID™ actively probes Microsoft Windows clients on the network for information on the currently logged on user. Using this mechanism, laptop users who often switch from wired to wireless networks can be reliably identified.
- **Host probing:** User-ID™ can also be configured to probe Microsoft Windows servers for active network sessions of a user. As soon as a user accesses a network share on the server, User-ID™ identifies the origin IP address and maps it to the user name provided to establish the session.
- **Terminal services:** Users sharing IP addresses working on Microsoft Windows Terminal Services or Citrix can be identified. Completely transparent to the user, every user session is assigned a certain port range on the server, which allows the firewall to associate network connections with users and groups sharing one host on the network.

## XML API

In some cases, organizations may already have a user repository or an application that is used to store information on users and their current IP address. In these scenarios, the XML API within User-ID™ enables rapid integration of user information with security policies. Examples of how the XML API can be used to collect user and IP address information are described below.

- **Wireless environments:** Organizations using 802.1x to secure corporate wireless networks can leverage a syslog based integration with the Palo Alto Networks User-ID™ XML API, to identify users as they authenticate to the wireless infrastructure.

- **Proxies:** Similarly, authentication prompted by a proxy server can be provided to Palo Alto Networks User-ID™ via its XML API by parsing the authentication log file for user and IP address information.
- **Network Access Control (NAC):** The XML API allows customers to harvest user information from NAC environments. As an example, Bradford Networks, a NAC solution provider uses the User-ID™ XML API to populate user logons and logoffs of its 802.1x solution. This integration allows organizations to identify users as soon as they connect to the network and set user-based enablement policies.

### Directory Integration

To allow organizations to specify security rules based on user groups and resolve the group members automatically, User-ID™ integrates with nearly every directory server using a standards-based protocol and a flexible configuration. Once configured, the firewall automatically retrieves user and user group information and keeps the information updated to automatically adjust to changes in the user base or organization.

- **Microsoft Active Directory:** In any other LDAP based directory service, the firewall can retrieve user and group information via standard LDAP from most LDAP-based directory servers. The association of users to computers can be achieved through other means, for example Captive Portal or XML API.

### Visibility into a User's Application Activity

The power of User-ID™ becomes evident when a strange or unfamiliar application is found on the network by App-ID™. Using either ACC™ or



the log viewer, an administrator can discern what the application is, and who is using the application, the bandwidth and session consumption, the sources and destinations of the application traffic as well as any associated threats.

Visibility into the application activity at a user level, not just an IP address level, allows organizations to more effectively enable the applications traversing the network. Administrators can align application usage with the business unit requirements and if appropriate, can choose to inform the user that they are in violation of corporate policy, or take a more direct approach of blocking the user's application usage outright.

### User-based Policy Control

User-based policy controls can be assembled based on the application, which category and subcategory it belongs in, its underlying technology or what the application characteristics are. Policies can be used to safely enable applications based on users or groups, in either an outbound or an inbound direction.

Examples of user-based policies might include:

- Enable only the IT department to use tools such as SSH, telnet, and FTP on their standard ports.
- Allow the Help Desk Services group to use Yahoo Messenger.
- Allow Facebook for all users, yet allow only marketing to use Facebook-posting and block the use of Facebook-apps for all users.

## **Returning the Firewall to the Cornerstone of Security**

The intent of this book is to provide a foundation for understanding where next-generation network security fits into the overall scheme of IT Security. Palo Alto Networks has built a security platform triad: endpoint, cloud, and network security. The platform satisfies a wide range of security requirements for enterprise computing, and reaches even greater effectiveness when integrated with other compatible technologies such as Security Information and Event Management, Vulnerability Management, Wireless Security, Intelligent Switching and Routing, and Data Loss Prevention systems.

The next-generation firewall restores the role of the firewall as the cornerstone of network security. When integrated into an organization's overall security strategy, the next-generation firewall makes significant contributions in terms of flexibility, in-line threat prevention of the latest attack methods, intelligence sharing, and overall effectiveness.

Students of information security should be aware of the direction that network security tools are evolving and seek to maintain the most current skill set possible. It is the hope that the materials presented here will support that learning and foster a desire to remain cutting-edge.

## Summary

- The next generation of firewalls are defined by their paradigm shift away from port and protocol for allow/block decisions, toward applications, users and content as the allow/block decision maker.
- Platforms, such as Microsoft SharePoint or Facebook, present multiple functions, where each function is distinguishable by its traffic signature. Next-generation firewalls provide visibility and granular control over these sub-functions as a result.
- Palo Alto Networks adopted a single-pass parallel architecture which uses stream-based malware detection to ensure maximum throughput with all security functions turned on.
- Integration of content identification functions allows automated threat intelligence sharing, and cross-referencing between functions in policy.
- User authentication identification provides personal context to network traffic and can be obtained through methods such as domain controller log scraping, captive portal, and client probing, or fed to the platform from outside systems via XML API.

## Discussion

1. If a malicious actor is able to obtain legitimate credentials on a given network, their traffic would appear to be originating from that legitimate user. What additional steps might be taken to detect malicious use of otherwise legitimate credentials?
2. Encrypted SSL traffic is a perfect hiding spot for malicious actors to operate undetected, through a generally allowed and open port and protocol. To seamlessly perform man-in-the-middle style decryption and inspection of SSL traffic requires adherence to the basic tenets of public key infrastructure (PKI). What are the components of PKI and how does it work?
3. Your next-generation firewall reports detection of command-and-control traffic originating from several of your client machines. How might you investigate, verify and remediate this situation?
4. Sophisticated attackers can piggy back their way into networks using seemingly legitimate applications. Search security news websites for instances where common applications were hijacked or repurposed, and discuss your findings.
5. How can an integrated security platform be leveraged to validate application traffic, and counter the scenario presented in discussion question #4 above?

## Knowledge Check

1. Name the three core capabilities of a next-generation firewall.
2. Which is not a technique for determining application **identification**?  
**A)** Packet Headers.                      **C)** Protocol Decoding  
**B)** Application Signatures              **D)** Behavioral Analysis
3. True or False: To securely enable use of applications, it may be necessary to restrict certain functionality within that app.
4. True or False: Intrusion prevention generally refers to blocking viruses and spyware at the network layer.
5. WildFire operates on which of the following concepts?  
**A)** File based scanning against a signature database.      **C)** Cloud based reputation service  
**B)** IPS and SIEM tool correlation      **D)** Virtualized sandbox
6. Forensic data uncovered by sandbox examination feeds threat intelligence to which of the following functions?  
**A)** Anti-virus signatures.      **C)** Malware category URL filtering  
**B)** Malicious DNS security      **D)** Command and control spyware signatures
7. True or False: URL filtering examines the requested web address against a categorized database of URLs to determine whether to allow or block.
8. Name three possible methods of mapping user identification information to IP address with a next-generation firewall.
9. True or False: User-ID fails when one user logs in to multiple devices simultaneously.
10. Which of the following are valid policy/mechanism combinations in a next-generation firewall?  
**A)** Selectively decrypting SSL traffic by excluding certain URL categories from decryption.  
**B)** Allowing social media applications while blocking the use of imbedded mail or chat functions to transfer files of specific types  
**C)** Allowing point to point protocols to operate only when the cross reference with certain URLs  
**D)** Blocking unknown (non-authenticated) users from accessing certain segments of the network while allowing them to have URL filtered internet access

# **Knowledge Check Answers**

## ANSWERS-CHAPTER 1 KNOWLEDGE CHECK

1. IRC is not a common application used on corporate networks, therefore most network administrators keep port 143 closed.
2. Port hopping, SSL encryption, use of non-standard ports, or tunneling within other allowed services–any two.
3. Better collaboration, expense reduction or increased knowledge sharing–any two. Microsoft Office 365, DropBox, Instant Messenger, Skype, Social Networks (LinkedIn, Facebook), Wiki, NewsGator, Google Docs, or other valid examples of cloud based applications.
4. False: The blurring of personal and corporate presents new challenges to network security.
5. True.
6. Lost productivity, potential disruption of operations, exposure to threats, regulatory penalties.
7. False: Layer 7 attacks are increasingly the norm, making use of existing entry points on the network layer that is allowed by traditional firewalls.
8. True. Blended threats are also known as ‘variants.’
9. Distributed, fault-tolerant, multifunctional, persistent, or intelligent–any 3.
10. The need for command and control, or network access to command and control.

## ANSWERS-CHAPTER 2 KNOWLEDGE CHECK

1. False: Zero-day commonly refers to threats that are unknown to the security industry.
2. Advantage: behavioral or anomaly based detection is better than pure signature based detection. Disadvantage: difficult deployments or lock-out from false positives.
3. Application Whitelisting only allows known good applications to run, while Exploit Prevention kills specific processes before they can be used to deliver an attack.
4. Firewall, proxy servers, intrusion detection/prevention, web content/URL filters, virtual private networks, data loss prevention, or unified threat management—any three.
5. C. Next-generation network security replaces legacy port/protocol allow/block decision making with applications, users and content.
6. Http or https proxies, remote desktop protocol, team viewer, goto meeting, webex, or other valid-use applications which provide remote desktop control.
7. False: SSL protects data in motion from interception, but is commonly used to mask attacks, command and control communications, and/or stolen data streams.
8. False: next-generation firewalls can enable certain functions of an application, while only blocking certain higher risk features of that same application.
9. A positive model denies all by default and classifies everything, a negative model allows all by default and requires specific manual classification.
10. Dynamic setup and tear down of applications.



## ANSWERS-CHAPTER 3 KNOWLEDGE CHECK

1. A. and D. Zero trust challenges perimeter-centric security models, that consider resources inside of the security perimeter to be trusted, while those outside to be untrusted. It also does not make assumptions an internal user is to be trusted, so B. and C. would be incorrect.
2. Can transition to Zero Trust incrementally and non-disruptively, obtain true situational awareness, strict enforcement of least privilege, enhanced security posture, compliance with regulations, adaptable to business driven IT initiatives, reduced TCO.
3. Reliance on network communication of some kind with a larger botnet or command and control network.
4. True.
5. A, B, and C are true. Unknown traffic falls into **one** of those **three scenarios**. Incomplete sessions are indicated as 'incomplete' in **next-generation** firewalls.
6. Presence of command and control traffic, unknown TCP/UDP, presence of dynamic DNS or malicious DNS requests, activity on known malware sites, visiting recently registered domains, use of IP instead of URL for browsing, IRC traffic presence—any **three**.
7. B. While whitelisting and HIPS can offer system level protection per host, they are not dynamic enough to be easily implemented at the same rate of change that occurs within virtualized server environments.
8. False. Most cloud environments are simply massive, publically accessible virtualized data centers **sharing** many **similar** security architectural **factors**.
9. Containers (or VDI), Mobile Device Management, Virtual Private Networks.
10. True. The device itself is often considered expendable, while sensitive data it may contain is not.

## ANSWERS-CHAPTER 4 KNOWLEDGE CHECK

1. Application identification, content identification, and user identification.
2. A. Simply checking a packet header is not a sound method of determining application.
3. True. Some functions of applications carry a higher risk of data loss, misuse, and/or present threat vectors, more than other functions.
4. False. Intrusion prevention generally refers to exploit protection, which is different from payload protection.
5. D. A, B, and C are alternative security methods, but are not the basis for WildFire.
6. A, B, C, and D are true. WildFire reveals unknown threats, purely by behavior, and makes them known threats by cycling their behaviors back to signature based methods.
7. True. When cross referenced against application, URL filtering can be implemented with greater contextual accuracy, but the underlying mechanism is to match URL against a known, categorized list.
8. Security event log monitoring (Active Directory, Novell, MS Exchange), user provided credentials, client probing, receive user information through XML API from external system—any 3.
9. False. Provided that those authentication mechanisms used are integrated with the next-generation firewall, a given user may have many IPs, or their IPs may change, but still logs will reflect the user associated with a given IP at the given time.
10. A, B, C, and D are all valid examples of how next-generation firewall features can be used in combination to provide flexible enforcement of security policy.

# Acknowledgements

*Cybersecurity Survival Guide-Principles & Best Practices* would not have been possible without the hard work, dedication, and expertise of Matthew Ancelin—the key contributor to this textbook. Thank you Matthew!

I would also like to thank Shoba Trivadi whose patience, understanding, and vision throughout this process has been absolutely critical to the success of the entire project. Finally, I'd like to thank Roger Connolly, Judith Backel, Jay Mackey, and all the great people at Palo Alto Networks for their contributions and expertise.

I look forward to the opportunity to work together again!

Author

**Lawrence Miller**

## A FINAL WORD

The idea in bringing this book to you is to give you a perspective of the various levels of threats and what you can do to prevent, manage, and eliminate it. We invite you to send us your comments and suggestions that we promise to incorporate as applicable in the next version of the book.

***Shoba Trivadi***