CYBERSECURITY SURVIVAL GUIDE

Principles & Best Practices

Third Edition | August 2018

Lawrence C. Miller, CISSP

paloalto NETWORKS
CYBERSECURITY ACADEMY

Advisory Panel:
Brian Adams
Jim Boardman
Steve Bradshaw
Keith Cantillon
James Dalton
Matthew Frohlich
Thomas Trevethan

Palo Alto Networks, Inc.
www.paloaltonetworks.com

Contents

Table of Figures

Foreword

Do you know what I like about my job? It changes all the time. There is always something new to learn; something new to figure out; something new to pull apart. The technology changes so fast. The adversaries are so nimble. There is always something exciting just around the corner that will challenge everything that you thought you had learned to this point. I love that. Do you know what I hate about my job? There is always something new to learn; something new to figure out; something new to pull apart.

That said, having a foundation of basic concepts will allow network defenders to be nimble in their thinking. It will allow them to immediately understand the impact to changes in technology, process and adversary tactics. In fact, without that foundation, network defenders will most likely fail at their threat prevention mission because the operational tempo of change will flow right by them. They will not be able to keep up.

This book provides a view of those basic cybersecurity concepts. I wish I would have had it when I was starting my career. The author, Larry Miller, with Scott Simkin, Sebastian Goodwin, Tim Treat, and other Palo Alto Networks' thought leaders have outdone themselves putting this material together. They provide a framework for how to think about cybersecurity and how the Palo Alto Networks' platform can help you adapt to the constant change more efficiently. In other words, reading this book will keep you on the side of "I love my job" more often than "I hate my job."

Rick Howard

Chief Security Officer – Palo Alto Networks

Introduction

The modern threat landscape continues to evolve and has become more complex and dangerous than ever before. Risks to the enterprise today include new and emerging attack techniques and vectors, accidental data loss and data theft, an ever-expanding network and cloud perimeter, and regulatory non-compliance penalties and other consequences.

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

This guide is divided into the following modules:

Module 1: Cybersecurity Foundation explains the nature and scope of today's cybersecurity challenges. This module explores the cybersecurity landscape, cyberthreats, malware and spamming, and Wi-Fi and advanced threats.

Module 2: Cybersecurity Gateway explains networking concepts, fundamentals, and technologies. This module explores the basic operation of computer networks; common networking devices; routed and routing protocols; network types and topologies; DNS; physical, logical, and virtual addressing; IP addressing and subnetting; the OSI and TCP/IP models; network security models, cloud and data center security, network security technologies, cloud computing, virtualization, and storage technologies; and network operations fundamentals.

Module 3: Cybersecurity Essentials presents detailed information about next-generation cybersecurity solutions available from Palo Alto Networks. This module demonstrates the real-world application of the cybersecurity design best practices and principles presented in Module 2, in order to address the cybersecurity landscape and threat challenges described in Module 1.

Module 1 – Cybersecurity Foundation

Knowledge Objectives

- Describe the cybersecurity landscape including modern computing trends, application frameworks and threat vectors, cloud computing and software as a service (SaaS) application challenges, information security and data protection regulations and standards, and recent cyberattacks.
- Discuss cyberthreats including attacker motivations and the Cyber-Attack Lifecycle.
- Describe endpoint security challenges and solutions.
- Describe cyber-attack techniques and types including malware, vulnerabilities, exploits, spamming, phishing, bots, and botnets.
- Discuss Wi-Fi and advanced threats including Wi-Fi vulnerabilities, Wi-Fi man-in-the-middle attacks, distributed denial-of-service (DDoS) attacks, and advanced persistent threats (APTs).

1.1 Cybersecurity Landscape

The modern cybersecurity landscape is a rapidly evolving, hostile environment fraught with advanced threats and increasingly sophisticated threat actors. This section describes computing trends that are shaping the cybersecurity landscape, application frameworks and threat vectors, cloud computing and SaaS application security challenges, various information security and data protection regulations and standards, and some recent cyberattack examples.

1.1.1 Modern computing trends

Note

The terms "enterprise" and "business" are used throughout this guide to describe organizations, networks, and applications in general. The use of these terms is not intended to exclude other types of organizations, networks, or applications, and should be understood to include not only large businesses and enterprises, but also small and medium-size businesses (SMBs), government, state-owned enterprises (SOEs), public services, military, healthcare, and non-profits, among others.

The nature of enterprise computing has changed dramatically over the past decade. Core business applications are now commonly installed alongside *Web 2.0* "apps" on a variety of *endpoints*, and networks that were originally designed to share files and printers are now used to collect massive volumes of data, exchange real-time information, transact online business, and enable global collaboration.

Key Terms

Web 2.0 is a term popularized by Tim O'Reilly and Dale Dougherty that unofficially refers to a new era of the World Wide Web, which is characterized by dynamic or user-generated content, interaction, and collaboration, and the growth of social media.

An *endpoint* is a computing device such as a desktop or laptop computer, handheld scanner, point-of-sale (POS) terminal, printer, satellite radio, security or videoconferencing camera, self-service kiosk, server, Internet of things (IoT) device or sensor (such as a smart meter, smart appliance, wearable device, or autonomous vehicle), smart TV, smartphone, tablet, or voice over internet protocol (VoIP) phone. Although endpoints can include servers and network equipment, the term is generally used to describe end-user devices.

Similarly, *Web 3.0* will transform the enterprise computing landscape over the next decade and beyond. Web 3.0, as defined on ExpertSystem.com, is characterized by five main features:

- **Semantic web.** "The semantic web improves web technologies in order to generate, share and connect through search and analysis based on the ability to understand the meaning of words, rather than on keywords and numbers."

- **Artificial intelligence.** "…computers can understand information like humans in order to provide faster and more relevant results."
- **3D graphics.** 3D design is "…used extensively in websites and services."
- **Connectivity.** "…information is more connected thanks to semantic metadata. As a result, the user experience evolves to another level of connectivity that leverages all the available information."
- **Ubiquity.** "Content is accessible by multiple applications, every device is connected to the web, [and] the services can be used everywhere."[1]

Many Web 2.0 apps are available as *software as a service* (SaaS), web-based, or mobile apps that can be easily installed by end users, or can be run without installing any local programs or services on the endpoint. The use of Web 2.0 apps in the enterprise is sometimes referred to as *Enterprise 2.0,* although not all Web 2.0 apps are considered to be Enterprise 2.0 applications.

Key Terms

Software as a service (SaaS) is a cloud computing service model, defined by the U.S. National Institute of Standards and Technology (NIST), in which "the capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through either a thin client interface, such as a web browser, or a program interface. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings."

Enterprise 2.0 is a term introduced by Andrew McAfee and defined as "the use of emergent social software platforms within companies, or between companies and their partners or customers."

[1] Expert System. 2017. "5 main features of Web 3.0." Accessed June 3, 2018. http://www.expertsystem.com/web-3-0/.
[2] "Application Usage & Risk Report: Fall 2009." Palo Alto Networks. November 2009. https://researchcenter.paloaltonetworks.com/2009/11/application-usage-risk-report-fall-2009/.

Typical core business applications include:

- **Accounting software** is used to process and record accounting data and transactions such as accounts payable (AP), accounts receivable (AR), payroll, trial balances, and general ledger (GL) entries. Examples of accounting software include Intacct, Microsoft Dynamics AX and GP, NetSuite, Quickbooks, and Sage.
- **Business intelligence (BI) and business analytics software** consists of tools and techniques used to surface large amounts of raw unstructured data from a variety of sources (such as data warehouses and data marts). BI and business analytics software performs a variety of functions, including business performance management, data mining, event processing, and predictive analytics. Examples of BI and analytics software include IBM Cognos, MicroStrategy, Oracle Hyperion, and SAP.
- **Content management systems (CMS) and enterprise content management (ECM) systems** are used to store and organize files from a central management interface, with features such as indexing, publishing, search, workflow management, and versioning. Examples of CMS and ECM software include EMC Documentum, HP Autonomy, Microsoft SharePoint, and OpenText.
- **Customer relationship management (CRM)** software is used to manage an organization's customer (or client) information including lead validation, past sales, communication and interaction logs, and service history. Examples of CRM suites include Microsoft Dynamics CRM, Salesforce.com, SugarCRM, and ZOHO.
- **Database management systems (DBMS)** are used to administer databases including the schemas, tables, queries, reports, views, and other objects that comprise a database. Examples of DBMS software include Microsoft SQL Server, MySQL, NoSQL, and Oracle Database.
- **Enterprise resource planning (ERP)** systems provide an integrated view of core business processes such as product and cost planning, manufacturing or service delivery, inventory management, and shipping and payment. Examples of ERP software include NetSuite, Oracle JD Edwards EnterpriseONE and PeopleSoft, and SAP.
- **Enterprise asset management (EAM)** software is used to manage an organization's physical assets throughout their entire lifecycle including acquisition, upgrade, maintenance, repair, replacement, decommissioning, and disposal. EAM is commonly implemented as an integrated module of ERP systems. Examples of EAM software include IBM Maximo, Infor EAM, and SAP.

- **Supply chain management (SCM)** software is used to manage supply chain transactions, supplier relationships, and various business processes such as purchase order processing, inventory management, and warehouse management. SCM software is commonly integrated with ERP systems. Examples of SCM software include Fishbowl Inventory, Freightview, Infor Supply Chain Management, and Sage X3.
- **Web content management (WCM)** software is used to manage website content including administration, authoring, collaboration, and publishing. Examples of web content management software include Drupal, IBM FileNet, Joomla, and WordPress.

Common Web 2.0 apps and services (many of which are also SaaS apps) include:

- **File sync and sharing services** are used to manage, distribute, and provide access to online content, such as documents, images, music, software, and video. Examples include Apple iCloud, Box, Dropbox, Google Drive, Microsoft OneDrive, Spotify, and YouTube.
- **Instant messaging (IM)** is used to exchange short messages in real-time. Examples include Facebook Messenger, Skype, Snapchat, and WhatsApp.
- **Microblogging** web services allow a subscriber to broadcast short messages to other subscribers. Examples include Tumblr and Twitter.
- **Office productivity suites** consist of cloud-based word processing, spreadsheet, and presentation software. Examples include Google Apps and Microsoft Office 365.
- **Remote access software** is used for remote sharing and control of an endpoint, typically for collaboration or troubleshooting purposes. Examples include Ammyy Admin, LogMeIn, and TeamViewer.
- **Social curation** shares collaborative content about a particular topic(s) or theme(s). Social bookmarking is a type of social curation. Examples include Cogenz, Instagram, Pinterest, and Reddit.
- **Social networks** are used to share content with business or personal contacts. Examples include Facebook, Google+, and LinkedIn.
- **Web-based email** is an Internet email service that is typically accessed via a web browser. Examples include Gmail, Outlook.com, and Yahoo! Mail.
- **Wikis** enable users to contribute, collaborate, and edit site content. Examples include Socialtext and Wikipedia.

According to research from McKinsey & Company and the Association for Information and Image Management (AIIM), many organizations are recognizing significant benefits from the use of Enterprise 2.0 applications and technologies including better collaboration, increased knowledge sharing, and reduced expenses (for example, for travel, operations, communications).[2] Thus, enterprise infrastructures (systems, applications, and networks) are rapidly converging with personal and Web 2.0 technologies and apps, making definition of where the Internet begins and the enterprise infrastructure ends practically impossible. This convergence is being driven by several important trends including:

- **Cloud computing.** The popularity of cloud computing service models in general, and SaaS application services in particular, continues to surge. According to a January 2018 McKinsey and Company article, even though adoption of the public cloud has been limited to date, the outlook is markedly different. Just 40 percent of the companies studied have more than 10 percent of their workloads on public-cloud platforms; in contrast, 80 percent plan to have more than 10 percent of their workloads in public-cloud platforms in three years or plan to double their cloud penetration.[3]

- **Consumerization.** The process of consumerization occurs as end users increasingly find personal technology and apps that are more powerful or capable, more convenient, less expensive, quicker to install, and easier to use than enterprise IT solutions.

- **Bring your own device (BYOD).** Closely related to consumerization is BYOD, a policy trend in which organizations permit end users to use their own personal devices, primarily smartphones and tablets, for work-related purposes. BYOD relieves organizations from the cost of providing equipment to employees, but creates a management challenge because of the vast number and type of devices that must be supported.

- **Bring your own apps (BYOA).** Web 2.0 apps on personal devices are increasingly being used for work-related purposes. As the boundary between work and personal lives

---

[2] "Application Usage & Risk Report: Fall 2009." Palo Alto Networks. November 2009. https://researchcenter.paloaltonetworks.com/2009/11/application-usage-risk-report-fall-2009/.

[3] Elumalai, Arul, James Kaplan, Mike Newborn, and Roger Roberts. "Making a secure transition to the public cloud." McKinsey and Company. January 2018. https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/making-a-secure-transition-to-the-public-cloud.

becomes less distinct, end users are practically demanding that these same apps be available to them in their workplaces.

- **Mobile computing.** The appetite for rapid, on-demand access to apps and data from anywhere, at any time, on any device is insatiable. There are now more than 4.4 billion smartphone subscriptions worldwide, and total mobile monthly data traffic (including audio, file sharing, social networking, software uploads and downloads, video, web browsing, and other sources) in the third quarter of 2017 was about 14 exabytes![4]

Organizations are often unsure of the potential business benefits – and the inherent risks – of these trends, and therefore either:

- Implicitly allow personal technologies and apps 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

Whether personal technologies and apps are implicitly allowed (and ignored) or explicitly prohibited (but not enforced), the adverse results of ineffective policies include:

- **Lost productivity** because users must either find ways to integrate these unsupported technologies and apps (when allowed) with the enterprise infrastructure, or use applications that are unfamiliar to them or less efficient (when personal technologies and apps are prohibited).
- **Potential disruption of critical business operations** because of 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 are fully dependent on their use of personal technologies and apps.
- **Exposure to additional risks** for the enterprise due to unknown – and therefore unpatched – vulnerabilities in personal technologies and apps, and a perpetual cat-and-mouse game between employees that circumvent controls (for example, with external

[4] "Ericsson Mobility Report, November 2017." Ericsson. November 2017. https://www.ericsson.com/assets/local/mobility-report/documents/2017/ericsson-mobility-report-november-2017.pdf.

proxies, encrypted tunnels, and remote desktop applications) and security teams that manage these risks.

- **Penalties for regulatory non-compliance**, for example, the E.U. General Data Protection Regulation (GDPR), the U.S. Health Insurance Portability and Accountability Act (HIPAA), and the Payment Card Industry Data Security Standard (PCI DSS).

As these trends continue to blur the distinction between the Internet and the enterprise network, new security challenges and risks emerge, including:

- New application threat vectors
- Turbulence in the cloud
- SaaS application risks

1.1.2 New application framework and threat vectors

Next-generation firewalls (NGFWs) disrupted the traditional port-based and unified threat management firewalls by capitalizing on the advances in hardware parallel processing to quickly inspect all network traffic and provide better attack prevention capability. It's true that NGFWs have provided much improved protection for on-site physical networks, but now a new cloud-based application consumption model is revolutionizing the way organizations do business. In this new model, applications such as Microsoft Office 365 can now be consumed and updated through cloud networks with no additional on-premise infrastructure. Attackers also are constantly innovating, and organizations must be able to rapidly evaluate and deploy new capabilities that detect and prevent successful cyberattacks in a highly agile, automated way, without deploying new infrastructure that needs to be purchased (capital expenditure) and managed (operating expenditure). To contend with these changes, a new framework to provide cybersecurity protection is needed. This new framework will likely disrupt the traditional point-based cybersecurity protection model. The new framework will have to leverage innovation and entrepreneurship, big data, machine learning, and advances in cloud technology to provide superior security with high consistency.

Beyond managing the risks associated with a relatively limited, known set of core applications that are authorized and supported in the enterprise, security teams must now manage the risks associated with an ever-increasing number of unknown personal technologies and apps that may be used in the organization.

Classification of applications as either "good" (allowed) or "bad" (blocked) in a clear and consistent manner has also become increasingly difficult. 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 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.

Many applications are designed to circumvent traditional port-based firewalls (discussed in Section 2.6.1), so that they can be easily installed and accessed on any device, anywhere and anytime, using techniques such as:

- **Port hopping**, in which ports and protocols are randomly changed during a session
- **Use of nonstandard ports**, such as running Yahoo! Messenger over TCP port 80 (HTTP) instead of the standard TCP port for Yahoo! Messenger (5050)
- **Tunneling within commonly used services**, such as when peer-to-peer (P2P) file sharing or an instant messenger (IM) client such as Meebo is running over HTTP
- **Hiding within SSL encryption**, which masks the application traffic, for example, over TCP port 443 (HTTPS). More than half of all web traffic is now encrypted

Many traditional client-server business applications are also being redesigned for web use, and employ these same techniques for ease of operation while minimizing disruptions. For example, *remote procedure call* (RPC) and Microsoft SharePoint both 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.

Key Terms

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.

An *attack* (or *threat*) *vector* is a path or tool that an attacker uses to target a network.

Applications can also be hijacked and repurposed by malicious actors, such as was done in the 2014 Heartbleed attack. According to an April 2014 Palo Alto Networks article:

> *"*[T]*he story of Heartbleed's impact has been focused on the compromise of HTTPS-enabled websites and web applications, such as Yahoo!, Google, Dropbox, Facebook, online banking, and the thousands of other vulnerable targets on the web. These are of huge impact, but those sites will all be updated quickly….*
>
> *"For security professionals,* [the initial Heartbleed attack] *is only the tip of the iceberg. The vulnerability puts the tools once reserved for truly advanced threats into the hands of the average attacker – notably, the ability to breach organizations, and move laterally within them. Most enterprises of even moderate size do not have a good handle on what services they are running internally using SSL encryption. Without this baseline knowledge, it is extremely difficult for security teams to harden their internal attack surface against the credential and data stealing tools Heartbleed enables. All footholds for the attacker with an enterprise network are suddenly of equal value."*[5]

As new applications are increasingly web-enabled and browser-based, HTTP and HTTPS now account for about two-thirds of all enterprise network traffic. Traditional port-based firewalls and other security infrastructure cannot distinguish whether these applications, riding on HTTP and HTTPS, are being used for legitimate business purposes.

Thus, applications (including malware) have become the predominant attack vector to infiltrate networks and systems.

1.1.3 Turbulence in the cloud

Cloud computing technologies enable organizations to evolve their data centers from a hardware-centric architecture where applications run on dedicated servers to a dynamic and automated environment where pools of computing resources are available on-demand, to support application workloads that can be accessed anywhere, anytime, and from any device.

However, many organizations have been forced into significant compromises with regard to their public and private cloud environments — trading function, visibility, and security, for simplicity, efficiency, and agility. If an application hosted in the cloud isn't available or

[5] Simkin, Scott. "Real-world Impact of Heartbleed (CVE-2014-0160): The Web is Just the Start." Palo Alto Networks. April 2014. https://researchcenter.paloaltonetworks.com/2014/04/real-world-impact-heartbleed-cve-2014-0160-web-just-start.

responsive, network security controls, which all too often introduce delays and outages, are typically “streamlined” out of the cloud design. Cloud security trade-offs often include

- Simplicity *or* function
- Efficiency *or* visibility
- Agility *or* security

Many of the features that make cloud computing attractive to organizations also run contrary to network security best practices. For example:

- **Cloud computing doesn’t mitigate existing network security risks.** The security risks that threaten your network today don’t go away when you move to the cloud. The shared responsibility model defines who (customer and/or provider) is responsible for what (related to security) in the public cloud. In general terms, the cloud provider is responsible for security ”of” the cloud, including the physical security of the cloud data centers, and for foundational networking, storage, compute, and virtualization services. The cloud customer is responsible for security ”in” the cloud, which is further delineated by the cloud service model. For example, in an infrastructure-as -a-service (IaaS) model, the cloud customer is responsible for the security of the operating systems, middleware, runtime, applications, and data. In a platform-as-a-service (PaaS) model, the cloud customer is responsible for the security of the applications and data – the cloud provider is responsible for the security of the operating systems, middleware, and run time. In a SaaS model, the cloud customer is responsible only for the security of the data, and the cloud provider is responsible for the full stack, from the physical security of the cloud data centers to the application.
- **Separation and segmentation are fundamental to security; the cloud relies on shared resources.** Security best practices dictate that mission-critical applications and data be separated in secure segments on the network, based on Zero Trust principles (discussed in Section 2.4.2). On a physical network, Zero Trust is relatively straightforward, using firewalls and policies based on application and user identity. In a cloud environment, direct communication between virtual machines (VMs) within a server host occurs constantly — in some cases, across varied levels of trust, thus making segmentation a real challenge. Mixed levels of trust, combined with a lack of intra-host traffic visibility by virtualized port-based security offerings, may weaken your security posture.
- **Security deployments are process-oriented; cloud computing environments are dynamic.** The creation or modification of your cloud workloads can often be done in

minutes, yet the security configuration for this workload may take hours, days, or weeks. Security delays aren't designed to be burdensome; they're the result of a process that is designed to maintain a strong security posture. Policy changes need to be approved, the appropriate firewalls need to be identified, and the relevant policy updates need to be determined. In contrast, the cloud is a highly dynamic environment, with workloads being added, removed, and changed rapidly and constantly. The result is a disconnect between security policy and cloud workload deployments, which lead to a weakened security posture. Thus, security technologies and processes must be able to auto scale to take advantage of the elasticity of the cloud while maintaining a strong security posture.

1.1.4 SaaS application risks

Data is located everywhere in today's enterprise networks, including many locations that are not under the organization's control. New data security challenges emerge for organizations that permit SaaS usage in their networks.

With SaaS applications, data is often stored where the application resides – in the cloud. Thus, the data is no longer under the organization's control, and visibility is often lost. SaaS vendors do their best to protect the data in their applications, but it is ultimately not their responsibility. Just as in any other part of the network, the IT team is responsible for protecting and controlling the data, regardless of its location.

Because of the nature of SaaS applications, their use is very difficult to control – or have visibility into – once the data leaves the network perimeter. This lack of control presents a significant security challenge: End users are now acting as their own "shadow" IT department, with control over the SaaS applications they use and how they use them. But they have little or no understanding of the inherent data exposure and threat insertion risks of SaaS, including:

- **Malicious outsiders.** The most common source of breaches for networks overall is also a critical concern for SaaS security. The SaaS application becomes a new threat vector and distribution point for malware used by external adversaries. Some malware will even target the SaaS applications themselves, for example, by changing their shares to "public" so the data can be retrieved by anyone.

- **Accidental data exposure.** Well-intentioned end users are often untrained and unaware of the risks their actions pose in SaaS environments. Because SaaS applications are designed to facilitate easy sharing, it's understandable that data often becomes unintentionally exposed. Accidental data exposure by end users is surprisingly common and includes:

- **Accidental share:** A share meant for a particular person is accidently sent to the wrong person or group. Accidental shares are common when a name auto fills, or is mistyped, which may cause an old email address or the wrong name, group, or even an external user, to have access to the share.

- **Promiscuous share:** A legitimate share is created for a user, but that user then shares with other people who shouldn't have access. Promiscuous shares often result in the data being publicly shared because it can go well beyond the control of the original owner.

- **Ghost (or stale) share:** A share remains active for an employee or vendor that is no longer working with the company, or should no longer have access. Without visibility and control of the shares, tracking and fixing shares to ensure they are still valid is very difficult.

- **Malicious insiders.** The least common but real SaaS application risk is the internal user who maliciously shares data for theft or revenge purposes. For example, an employee who is leaving the company might set a folder's share permissions to "public," or share it with an external email address to later steal the data from a remote location.

1.1.5 Compliance and security are not the same

A rapidly and ever-increasing number of international, multinational, federal, regional, state, and local laws and regulations mandate numerous cybersecurity and data protection requirements for businesses and organizations worldwide. Various industry directives, such as the Payment Card Industry's Data Security Standard (PCI DSS), also establish their own cybersecurity standards and best practices for businesses and organizations operating under their purview.

This complex regulatory environment is further complicated by the fact that many laws and regulations are obsolete, ambiguous, not uniformly supported by international communities, and/or inconsistent (with other applicable laws and regulations), thus requiring legal interpretation to determine relevance, intent, and/or precedence. As a result, businesses and organizations in every industry struggle to achieve and maintain compliance.

You should understand that compliance and security are not the same thing. An organization can be fully compliant with the various cybersecurity laws and regulations that are applicable for that organization, yet still not be secure. Conversely, an organization can be secure, yet not be fully compliant. As if to underscore this point, the compliance and security functions in many organizations are separate.

Pertinent examples (neither comprehensive nor exhaustive) of current cybersecurity laws and regulations include:

- **Canada Personal Information Protection and Electronic Documents Act (PIPEDA).** PIPEDA defines individual rights with respect to the privacy of their personal information, and governs how private sector organizations collect, use, and disclose personal information in the course of business.
- **European Union (EU) General Data Protection Regulation (GDPR).** The GDPR applies to any organization that does business with EU citizens. It strengthens data protection for E.U. citizens and addresses the export of personal data outside the EU.
- **EU Network and Information Security (NIS) Directive:** An EU directive that imposes network and information security requirements for banks, energy companies, healthcare providers and digital service providers, among others.
- **North American Electric Reliability Corporation (NERC) Critical Infrastructure Protection (CIP).** NERC CIP defines cybersecurity standards to protect the physical and cyber assets necessary to operate the Bulk Electric System (BES) – the "power grid" – in the United States and Canada. The standards are mandatory for all BES-generating facilities with different criteria based on a tiered classification system (high, medium, or low impact).
- **Payment Card Industry Data Security Standards (PCI DSS).** PCI DSS applies to any organization that transmits, processes, or stores payment card (such as debit and credit cards) information. PCI DSS is mandated and administered by the PCI Security Standards Council (SSC) comprising Visa, MasterCard, American Express, Discover, and JCB.
- **U.S. Cybersecurity Enhancement Act of 2014.** This act provides an ongoing, voluntary public-private partnership to improve cybersecurity and to strengthen cybersecurity research and development, workforce development and education, and public awareness and preparedness.
- **U.S. Cybersecurity Information Sharing Act (CISA).** This act enhances information sharing about cybersecurity threats by allowing Internet traffic information to be shared between the U.S. government and technology and manufacturing companies.
- **U.S. Federal Exchange Data Breach Notification Act of 2015.** This act further strengthens HIPAA by requiring health insurance exchanges to notify individuals whose personal information has been compromised as the result of a data breach as soon as possible, but no later than 60 days after breach discovery.

- **U.S. Federal Information Security Modernization Act (FISMA).** Known as the Federal Information Security Management Act prior to 2014, FISMA implements a comprehensive framework to protect information systems used in federal government agencies.
- **U.S. Gramm-Leach-Bliley Act (GLBA).** Also known as the Financial Services Modernization Act of 1999, relevant provisions of GLBA include the Financial Privacy Rule and the Safeguards Rule, which require financial institutions to implement privacy and information security policies to safeguard the nonpublic personal information of clients and consumers.
- **U.S. Health Insurance Portability and Accountability Act (HIPAA).** The HIPAA Privacy Rule establishes national standards to protect individuals' medical records and other personal health information. It requires appropriate safeguards for *protected health information* (PHI) and applies to *covered entities* and their business associates.
- **U.S. National Cybersecurity Protection Advancement Act of 2015.** This act amends the Homeland Security Act of 2002 to enhance multi-directional sharing of information related to cybersecurity risks and strengthens privacy and civil liberties protections.
- **U.S. Sarbanes-Oxley (SOX) Act.** This act was enacted to restore public confidence following several high-profile corporate accounting scandals, most notably Enron and Worldcom, SOX increases financial governance and accountability in publicly traded companies. Section 404 of SOX specifically addresses internal controls, including requirements to safeguard the confidentiality, integrity, and availability of IT systems.

Key Terms

Protected health information (PHI) is defined by HIPAA as information about an individual's health status, provision of healthcare, or payment for healthcare that includes identifiers such as names, geographic identifiers (smaller than a state), dates, phone and fax numbers, email addresses, Social Security numbers, medical record numbers, or photographs, among others.

A *covered entity* is defined by HIPAA as a healthcare provider that electronically transmits PHI (such as doctors, clinics, psychologists, dentists, chiropractors, nursing homes, and pharmacies), a health plan (such as a health insurance company, health maintenance organization, company health plan, or government program including Medicare, Medicaid, military and veterans' healthcare), or a healthcare clearinghouse.

1.1.6 Recent high-profile cyber-attack examples

Thousands of cyberattacks are perpetrated against enterprise networks every day. Unfortunately, many more of these attacks succeed than are typically reported in mass media. Recent high-profile examples of such attacks include:

- **Target.** In late 2013, Target discovered that credit card data and debit card data from 40 million of its customers, and the personal information of an additional 70 million of its customers, had been stolen over a period of about 19 days, from November 27 to December 15, 2013. The attackers were able to infiltrate Target's Point of Sale (POS) systems by installing malware (believed to be a variant of the ZeuS financial botnet) on an HVAC (heating, ventilation, and air conditioning) contractor's computer systems to harvest credentials for an online portal used by Target's vendors. Target's 2016 annual report disclosed that the total cost of the breach was US$292 million.

- **Home Depot.** In September 2014, Home Depot suffered a data breach that went unnoticed for about five months. As with the Target data breach (see the previous attack example), the attackers used a vendor's credentials and exploited a *zero-day threat*, based on a Windows vulnerability, to gain access to Home Depot's network. Memory scraping malware was then installed on more than 7,500 self-service POS terminals to collect 56 million customer credit card numbers throughout the United States and Canada. Home Depot's 2016 annual report disclosed that the total cost of the breach was US$298 million.

- **Anthem.** In February 2015, Anthem disclosed that its servers had been breached and Personally Identifiable Information (PII) including names, Social Security numbers, birthdates, addresses, and income information, for about 80 million customers had been stolen. The breach occurred on December 10, 2014, when attackers compromised an Anthem database using a database administrator's credentials. The breach wasn't found until January 27, 2015, when the database administrator discovered a questionable query being run with his credentials. The total cost of the breach is expected to reach US$31 billion.

- **U.S. Office of Personnel Management (OPM).** Two separate data breaches discovered in April 2015 and June 2015 resulted in the compromise of personal information including names, Social Security numbers, birthdates, and other sensitive information of about 24 million current and prospective federal employees (along with their spouses and partners). The breaches are believed to have been linked to the Anthem data breach (see the previous attack example) and may have originated in China as early as

March 2014. By some estimates, the total cost of the breach could exceed US$1 billion over the next decade.

- **Yahoo!** While in negotiations to sell itself to Verizon in September 2016, Yahoo! announced it had been the victim of a data breach in 2014, likely by a "state-sponsored actor." The attack compromised the real names, email addresses, birthdates, and phone numbers of about 500 million users and is the largest data breach to date. Yahoo! said the vast majority of the passwords involved had been hashed using the robust bcrypt algorithm. As a direct result of the breach, Yahoo! reduced its sale price to Verizon by US$350 million.
- **Equifax.** In July 2017, Equifax discovered a data breach that had exploited an unpatched security vulnerability (Apache Struts CVE-2017-5638 published March 10, 2017). From mid-May to July 2017, cybercriminals compromised various personal information of nearly 148 million U.S. consumers (as of March 2018), including passport and driver's license data, and Social Security numbers. The total cost of the breach at the end of 2017 was US$439 million and could ultimately exceed US$600 million.

Key Terms

A *zero-day threat* is the window of vulnerability that exists from the time a new (unknown) threat is released until security vendors release a signature file or security patch for the threat.

Personally Identifiable Information (PII) is defined by the U.S. National Institute of Standards and Technology (NIST) as "any information about an individual maintained by an agency, including (1) any information that can be used to distinguish or trace an individual's identity… and (2) any other information that is linked or linkable to an individual…." Examples of PII include:

- **Name** (such as full name, maiden name, mother's maiden name, or alias)
- **Personal identification number** (such as Social Security number, passport number, driver's license number, and financial account number or credit card number)
- **Address information** (such as street address or email address)
- **Asset information** (such as IP or MAC address)
- **Telephone numbers** (such as mobile, business, and personal numbers)
- **Personal characteristics** (such as photographs, x-rays, fingerprints, and biometric data)
- **Information about personally owned property** (such as vehicle registration number or title information)
- **Information that is linked or linkable to any of the above** (such as birthdate, birthplace, race, religion, height, weight, and employment, medical, education, and financial records)

Important lessons to be learned from these attacks include:

- A "low and slow" cyberattack can go undetected for weeks, months, or even years.
- An attacker doesn't necessarily need to run a sophisticated exploit against a hardened system to infiltrate a target organization. Often, an attacker will target an auxiliary system or other vulnerable endpoint, then pivot the attack toward the primary target.
- Unpatched vulnerabilities are a commonly exploited attack vector.

- The direct and indirect financial costs of a breach can be devastating for both the targeted organization and individuals whose personal and financial information is stolen or compromised.

1.1 Knowledge Check

Test your understanding of the fundamentals in the preceding section. Review the correct answers in the Appendix at the end of this guide.

1. **Multiple Choice.** In which cloud computing service model does a provider's applications run on a cloud infrastructure and the consumer does not manage or control the underlying infrastructure? (Choose one.)

    a) platform as a service (PaaS)
    b) infrastructure as a service (IaaS)
    c) software as a service (SaaS)
    d) public cloud

2. **True or False.** Business intelligence (BI) software consists of tools and techniques used to surface large amounts of raw unstructured data to perform a variety of tasks including data mining, event processing, and predictive analytics.

3. **True or False.** The process in which end users find personal technology and apps that are more powerful or capable, more convenient, less expensive, quicker to install, and easier to use than enterprise IT solutions is known as *consumerization*.

4. **True or False.** An organization can be compliant with all applicable security and privacy regulations for its industry, yet still not be secure.

5. **Fill in the Blank.** The U.S. law that establishes national standards to protect individuals' medical records and other health information is known as the _____.

6. **Classroom Discussion.** What are lessons or common themes that can be derived from the Target, Home Depot, Anthem, OPM, Yahoo!, and Equifax cyberattack examples?

7. **Classroom Discussion.** What other organizations have been the victim of a recent cyberattack? How did the attacker gain access? How long did the attack go undetected? What was the apparent objective of the attack? Was the attacker a criminal organization, nation-state, or hacktivist? What, if anything, do these attacks have in common with the Target, Home Depot, Anthem, OPM, Yahoo!, and Equifax attacks?

1.2 Cyberthreats

This section describes cybersecurity adversaries – the various threat actors, their motivations, and the cyber-attack strategy.

1.2.1 Attacker profiles and motivations

In *The Art of War*, Sun Tzu teaches "know thy enemy, know thy self. A thousand battles, a thousand victories" (translated in various forms) to instill the importance of understanding the strengths, weaknesses, strategies, and tactics of your adversary as well as you know your own. Of course, in modern cyber warfare a thousand battles can occur in a matter of seconds and a single victory by your enemy can imperil your entire organization. Thus, knowing your enemies – including their means and motivations – is more important than ever.

In the relatively innocuous "good ol' days" of *hackers* and *script kiddies*, the primary motivation for a cyberattack was notoriety, and the attack objective was typically limited to defacing or "owning" a website to cause inconvenience and/or embarrassment to the victim.

Key Terms

The term *hacker* was originally used to refer to anyone with highly specialized computing skills, without connoting good or bad purposes. However, common misuse of the term has redefined a hacker as someone that circumvents computer security with malicious intent, such as a cybercriminal, cyberterrorist, or hacktivist, cracker, and/or black hat.

A *script kiddie* is someone with limited hacking and/or programming skills that uses malicious programs (malware) written by others to attack a computer or network.

Modern cyberattacks are perpetrated by far more sophisticated and dangerous adversaries, motivated by far more sinister purposes including:

- **Cybercriminals.** Acting independently or as part of a criminal organization, cybercriminals commit acts of data theft, embezzlement, fraud, and/or extortion for financial gain. According to the RAND Corporation, "In certain respects, the black market

[for cybercrime] can be more profitable than the illegal drug trade,"[6] and by many estimates, cybercrime is now a US$1 trillion industry.

- **State-affiliated groups.** Sponsored by or affiliated with nation-states, these organizations have the resources to launch very sophisticated and persistent attacks, have great technical depth and focus, and are well funded. They often have military and/or strategic objectives such as the ability to disable or destroy critical infrastructure including power grids, water supplies, transportation systems, emergency response, and medical and industrial systems. The Center for Strategic and International Studies reports that "At the nation-state level, Russia, Iran, and North Korea are using coercive cyberattacks to increase their sphere of influence, while China, Russia and Iran have conducted reconnaissance of networks critical to the operation of the U.S. power grid and other critical infrastructure without penalty."[7]

- **Hacktivists.** Motivated by political or social causes, hacktivist groups (such as Anonymous) typically execute denial-of-service (DoS) attacks against a target organization by defacing their websites or flooding their networks with traffic.

- **Cyberterrorists.** Terrorist organizations use the Internet to recruit, train, instruct, and communicate, and to spread fear and panic to advance their ideologies. Unlike other threat actors, cyberterrorists are largely indiscriminate in their attacks and their objectives include physical harm, death, and destruction.

External threat actors including organized crime, state-affiliated groups, activists, former employees, and other unaffiliated or otherwise unknown attackers account for the majority of data breaches. Internal threat actors are responsible for about 28 percent of reported data breaches.[8]

---

[6] Lillian Ablon, Martin Libicki, and Andrea Golay. "Markets for Cybercrime Tools and Stolen Data." RAND Corporation, National Security Research Division. 2014. https://www.rand.org/content/dam/rand/pubs/research_reports/RR600/RR610/RAND_RR610.pdf.

[7] Zheng, Denise E. "Global Forecast 2016: Disrupting the Cyber Status Quo." Center for Strategic and International Studies. November 16, 2015. https://www.csis.org/analysis/disrupting-cyber-status-quo.

[8] "2018 Data Breach Investigations Report, 11th Edition." Verizon Enterprise Solutions. 2018. https://www.verizonenterprise.com/resources/reports/rp_DBIR_2018_Report_en_xg.pdf.

1.2.2 Modern cyber-attack strategy

Modern cyber-attack strategy has evolved from a direct attack against a high-value server or asset ("shock and awe") to a patient, multi-step process that blends exploits, malware, stealth, and evasion in a coordinated network attack ("low and slow").

The Cyber-Attack Lifecycle (see Figure 1-1) illustrates the sequence of events that an attacker goes through to infiltrate a network and exfiltrate (or steal) valuable data. Blocking of just one step breaks the chain and can effectively defend an organization's network and data against an attack.

Figure 1-1: The Cyber-Attack Lifecycle

1. **Reconnaissance.** Like common criminals, attackers meticulously plan their cyberattacks. They research, identify, and select targets, often extracting public information from targeted employees' social media profiles or from corporate websites, which can be useful for social engineering and phishing schemes. Attackers will also use various tools to scan for network vulnerabilities, services, and applications that they can exploit, such as:

    - **Network analyzers** (also known as packet analyzers, protocol analyzers, or packet sniffers) are used to monitor and capture raw network traffic (packets). Examples include tcpdump and Wireshark (formerly Ethereal).
    - **Network vulnerability scanners** typically consist of a suite of tools including password crackers, port scanners, and vulnerability scanners and are used to probe a network for vulnerabilities (including configuration errors) that can be exploited. Examples include Nessus and SAINT.
    - **Password crackers** are used to perform brute-force dictionary attacks against password hashes. Examples include John the Ripper and THC Hydra.
    - **Port scanners** are used to probe for open TCP or UDP (including ICMP) ports on an endpoint. Examples include Nmap ("network mapper") and Nessus.

- **Web application vulnerability scanners** are used to scan web applications for vulnerabilities such as cross-site scripting, SQL injection, and directory traversal. Examples include Burp Suite and OWASP Zed Attack Proxy (ZAP).

- **Wi-Fi vulnerability scanners** are used to scan wireless networks for vulnerabilities (including open and misconfigured access points), to capture wireless network traffic, and to crack wireless passwords. Examples include Aircrack-ng and Wifite.

Breaking the Cyber-Attack Lifecycle at this phase of an attack begins with proactive and effective end-user security awareness training that focuses on topics such as social engineering techniques (for example, phishing, piggybacking, and shoulder surfing), social media (for example, safety and privacy issues), and organizational security policies (for example, password requirements, remote access, and physical security). Another important countermeasure is continuous monitoring and inspection of network traffic flows to detect and prevent unauthorized port and vulnerability scans, host sweeps, and other suspicious activity. Effective change and configuration management processes help to ensure that newly deployed applications and endpoints are properly configured (for example, disabling unneeded ports and services) and maintained.

2. **Weaponization.** Next, attackers determine which methods to use to compromise a target endpoint. They may choose to embed intruder code within seemingly innocuous files such as a PDF or Microsoft Word document or email message. Or, for highly targeted attacks, attackers may customize deliverables to match the specific interests of an individual within the target organization.

   Breaking the Cyber-Attack Lifecycle at this phase of an attack is challenging because weaponization typically occurs within the attacker's network. However, analysis of artifacts (both malware and weaponizer) can provide important threat intelligence to enable effective zero-day protection when delivery (the next step) is attempted.

3. **Delivery.** Attackers next attempt to deliver their weaponized payload to a target endpoint, for example, via email, instant messaging (IM), drive-by download (an end user's web browser is redirected to a webpage that automatically downloads malware to the endpoint in the background), or infected file share.

   Breaking the Cyber-Attack Lifecycle at this phase of an attack requires visibility into all network traffic (including remote and mobile devices) to effectively block malicious or risky websites, applications, and IP addresses, and preventing known and unknown malware and exploits.

4. **Exploitation**. After a weaponized payload is delivered to a target endpoint, it must be triggered. An end user may unwittingly trigger an exploit, for example, by clicking a malicious link or opening an infected attachment in an email, or an attacker may remotely trigger an exploit against a known server vulnerability on the target network.

   Breaking the Cyber-Attack Lifecycle at this phase of an attack, as during the Reconnaissance phase, begins with proactive and effective end-user security awareness training that focuses on topics such as malware prevention and email security. Other important security countermeasures include vulnerability and patch management, malware detection and prevention, threat intelligence (including known and unknown threats), blocking risky, unauthorized, or unneeded applications and services, managing file or directory permissions and root or administrator privileges, and logging and monitoring network activity.

5. **Installation**. Next, an attacker will escalate privileges on the compromised endpoint, for example, by establishing remote shell access and installing root kits or other malware. With remote shell access, the attacker has control of the endpoint and can execute commands in privileged mode from a command line interface (CLI), as if physically sitting in front of the endpoint. The attacker will then move laterally across the target's network, executing attack code, identifying other targets of opportunity, and compromising additional endpoints to establish persistence.

   The key to breaking the Cyber-Attack Lifecycle at this phase of an attack is to limit or restrict the attackers' lateral movement within the network. Use network segmentation and a Zero Trust model that monitors and inspects all traffic between zones or segments, and granular control of applications that are allowed on the network.

6. **Command and Control**. Attackers establish encrypted communication channels back to command-and-control (C&C) servers across the Internet so that they can modify their attack objectives and methods as additional targets of opportunity are identified within the victim network, or to evade any new security countermeasures that the organization may attempt to deploy if attack artifacts are discovered. Communication is essential to an attack because it enables the attacker to remotely direct the attack and execute the attack objectives. C&C traffic must therefore be resilient and stealthy for an attack to succeed. Attack communication traffic is usually hidden with various techniques and tools including:

   - **Encryption** with SSL, SSH (Secure Shell), or some other custom or proprietary encryption.

- **Circumvention** via proxies, remote access tools, or tunneling. In some instances, use of cellular networks enables complete circumvention of the target network for attack C&C traffic.
- **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 endpoints or multiple ever-changing C&C servers to reroute traffic, and make determination of the true destination or attack source difficult.

Breaking the Cyber-Attack Lifecycle at this phase of an attack requires inspection of all network traffic (including encrypted communications), blocking of outbound C&C communications with anti-C&C signatures (along with file and data pattern uploads), blocking of all outbound communications to known malicious URLs and IP addresses, blocking of novel attack techniques that employ port evasion methods, prevention of the use of anonymizers and proxies on the network, monitoring of DNS for malicious domains and countering with DNS sinkholing or DNS poisoning, and redirection of malicious outbound communications to honeypots to identify or block compromised endpoints and analyze attack traffic.

7. **Actions on the Objective**. Attackers often have multiple, different attack objectives including data theft; destruction or modification of critical systems, networks, and data; and denial-of-service (DoS). This last stage of the Cyber-Attack Lifecycle can also be used by an attacker to advance the early stages of the Cyber-Attack Lifecycle against another target. The 2018 Verizon *Data Breach Investigations Report* (DBIR) describes this strategy as a secondary motive in which "[web applications] are compromised to aid and abet in the attack of another victim."[9] For example, an attacker may compromise a company's extranet to breach a business partner that is the primary target. According to the DBIR, in 2014 there were 23,244 "incidents where web applications were compromised with a secondary motive."[10] The attacker pivots the attack against the

[9] Ibid.

[10] Ibid.

initial victim network to a different victim network, thus making the initial victim an unwitting accomplice.

1.2 Knowledge Check

Test your understanding of the fundamentals in the preceding section. Review the correct answers in the Appendix at the end of this guide.

1. **True or False.** Most cyberattacks today are perpetrated by internal threat actors such as malicious employees engaging in corporate espionage.

2. **Classroom Discussion.** Describe the different motivations of various adversaries including cybercriminals, cyberterrorists, state-sponsored organizations, and hacktivists.

3. **True or False.** The Cyber-Attack Lifecycle is a five-step process that an attacker goes through to attack a network.

4. **Multiple Answer.** List and briefly describe the steps of the Cyber-Attack Lifecycle.

5. **True or False**. An attacker needs to succeed in executing only one step of the Cyber-Attack Lifecycle to infiltrate a network, whereas a defender must "be right every time" and break every step of the chain to prevent an attack.

6. **Multiple Choice.** Which technique is *not* used to break the command and control (C&C) phase of the Cyber-Attack Lifecycle? (Choose one.)

    a) blocking outbound traffic to known malicious sites and IP addresses
    b) DNS sinkholing and DNS poisoning
    c) vulnerability and patch management
    d) all of the above

7. **True or False.** The key to breaking the Cyber-Attack Lifecycle during the Installation phase is to implement network segmentation, a Zero Trust model, and granular control of applications to limit or restrict an attacker's lateral movement within the network.

1.3 Endpoint security

Most organizations deploy a number of security products to protect their endpoints, including personal firewalls, Host-Based Intrusion Prevention Systems (HIPS), mobile device management (MDM), mobile application management (MAM), data loss prevention (DLP), and antivirus software. Nevertheless, cyber breaches continue to increase in frequency, variety, and sophistication. Faced with the rapidly changing threat landscape, traditional endpoint security solutions and antivirus can no longer prevent security breaches on the endpoint.

Endpoint security is an essential element of cybersecurity because the network firewall cannot completely protect hosts from zero-day exploits. Zero-day exploits target unknown vulnerabilities in operating system and application software on host machines. Network firewalls may not be able to block an attacker's delivery of a zero-day exploit until a new signature identifying the zero-day attack has been developed and delivered to the firewall.

Network firewalls also may be restricted from decrypting all traffic because of regulations and laws. This restriction provides a window of opportunity for attackers to bypass a firewall's protection and exploit a host machine necessitating endpoint security protection. Endpoint security protection is provided by an application that runs on the host machine. Effective endpoint security must be able to stop malware, exploits, and ransomware before they can compromise the host; provide protection while endpoints are online and offline; and detect threats and automate containment to minimize impact.

1.3 Knowledge Check

Test your understanding of the fundamentals in the preceding section. Review the correct answers in the Appendix at the end of this guide.

1. **True or False.** Network firewalls cannot completely protect hosts from zero-day exploits.
2. **Fill in the Blank.** _____ exploits target unknown vulnerabilities in operating system and application software on a host machine.

1.4 Cyberattack Techniques and Types

Attackers use a variety of techniques and attack types to achieve their objectives. *Malware* and *exploits* are integral to the modern cyber-attack strategy. Spamming and phishing are commonly employed techniques to deliver malware and exploits to an endpoint via an email executable or a web link to a malicious website. Once an endpoint is compromised, an attacker

typically installs back doors, remote access Trojans, and other malware to ensure persistence. Compromised endpoints ("bots") under the control of an attacker are often used to perpetrate much larger-scale attacks against other organizations or networks as part of a botnet. This section describes different types of malware, vulnerabilities, and exploits; email spamming and phishing techniques; and how bots and botnets function, along with different types of botnets.

Key Terms

Malware is malicious software or code that typically damages, takes control of, or collects information from an infected endpoint. Malware broadly includes viruses, worms, Trojan horses (including remote access Trojans, or RATs), ransomware, anti-AV, logic bombs, back doors, rootkits, bootkits, spyware, and (to a lesser extent) adware.

An *exploit* is a small piece of software code, part of a malformed data file, or a sequence (string) of commands that leverages a vulnerability in a system or software, causing unintended or unanticipated behavior in the system or software.

A *vulnerability* is a bug or flaw that exists in a system or software, and creates a security risk.

1.4.1 Malware

Malware is malicious software or code that typically damages, takes control of, or collects information from an infected endpoint. Malware broadly includes:

- **Viruses:** Malware that is self-replicating but must first infect a host program and be executed by a user or process
- **Worms:** Malware that typically targets a computer network by replicating itself to spread rapidly. Unlike viruses, worms do not need to infect other programs and do not need to be executed by a user or process.
- **Trojan horses:** Malware that is disguised as a harmless program, but actually gives an attacker full control and elevated privileges of an endpoint when installed. Unlike other types of malware, Trojan horses are typically not self-replicating.
- **Ransomware:** Malware that locks a computer or device (Locker ransomware) or encrypts data (Crypto ransomware) on an infected endpoint with an encryption key that only the attacker knows, thereby making the data unusable until the victim pays a ransom (usually cryptocurrency, such as Bitcoin). Reveton and LockeR are two examples

of Locker ransomware. Locky, TeslaCrypt/EccKrypt, Cryptolocker, and Cryptowall are examples of Crypto ransomware.

- **Anti-AV:** Malware that disables legitimately installed antivirus software on the compromised endpoint, thereby preventing automatic detection and removal of other malware
- **Logic bombs:** Malware that is triggered by a specified condition, such as a given date or a particular user account being disabled
- **Back doors:** Malware that allows an attacker to bypass authentication to gain access to a compromised system
- **Rootkits:** Malware that provides privileged (root-level) access to a computer. Rootkits are installed in the BIOS of a machine, which means operating system-level security tools cannot detect them.
- **Bootkits:** Malware that is a kernel-mode variant of a rootkit, commonly used to attack computers that are protected by full-disk encryption
- **Spyware and adware:** Malware that collects information, such as Internet surfing behavior, login credentials, and financial account information on an infected endpoint. Spyware often changes browser and other software settings, and slows computer and Internet speeds on an infected endpoint. Adware is spyware that displays annoying advertisements on an infected endpoint, often as popup banners.

Early malware typically consisted of viruses that displayed annoying – but relatively benign – errors, messages, or graphics.

The first computer virus was Elk Cloner, written in 1982 by a ninth grade high school student near Pittsburgh, Pennsylvania. Elk Cloner was a relatively benign *boot sector* virus that displayed a poem on the fiftieth time that an infected *floppy disk* was inserted into an Apple II computer.

The first PC virus was a boot sector virus, written in 1986, called Brain. Brain was also relatively benign and displayed a message with the actual contact information for the creators of the virus. Brain was written by two Pakistani brothers who created the virus so that they could track piracy of their medical software.

One of the first computer worms to gain widespread notoriety was the Morris worm, written by a Harvard and Cornell University graduate student, Robert Tappan Morris, in 1988. The worm exploited weak passwords and known vulnerabilities in several Unix programs and spread rapidly across the early Internet (the worm infected up to an estimated 10 percent of all Unix machines connected to the Internet at that time – about 6,000 computers), sometimes infecting a computer numerous times to the point that it was rendered useless – an example of an early DoS attack. The U.S. Government Accountability Office (GAO) estimated the damage caused by the Morris worm between US$100,000 and US$10 million.

Key Terms

A *boot sector virus* targets the boot sector or master boot record (MBR) of an endpoint's storage drive or other removable storage media.

A *boot sector* contains machine code that is loaded into an endpoint's memory by firmware during the startup process, before the operating system is loaded.

A *master boot record* (MBR) contains information about how the logical partitions (or file systems) are organized on the storage media, and an executable boot loader that starts up the installed operating system.

A *floppy disk* is a removable magnetic storage medium commonly used from the mid-1970s until about 2007, when it was largely replaced by compact discs and removable USB storage devices. Floppy disks were typically available in 8-inch, 5¼-inch, and 3½-inch sizes with capacities from 90 kilobytes to 200 megabytes!

Unfortunately, more than 35 years since these early examples of malware, modern malware has evolved and is used for far more sinister purposes. Examples of modern malware include:

- **WannaCry.** In a period of just 24 hours in May 2017, the WannaCry ransomware attack infected more than 230,000 vulnerable Windows computers in more than 150 countries worldwide. Although the attack was quickly halted after the discovery of a "kill switch", the total economic damage is estimated between hundreds of millions to as much as US$4 billion, despite the perpetrators collecting only 327 ransom payments totaling about US$130,000.
- **HenBox.** HenBox typically masquerades as legitimate Android system and VPN apps, and sometimes embeds legitimate apps. the primary goal of the HenBox apps appears to be to spy on those who install them. By using similar traits as legitimate apps, for

example, copycat iconography and app or package names, HenBox lures victims into installing the malicious apps, especially when available on so-called third-party (that is, non-Google Play) app stores that often have fewer security and vetting procedures for the apps they host. As with other Android malware, some apps may also be available on forums or file-sharing sites, or even may be sent to victims as email attachments.

- **TeleRAT.** Telegram Bots are special accounts that do not require an additional phone number to set up and are generally used to enrich Telegram chats with content from external services or to get customized notifications and news. TeleRAT abuses Telegram's Bot API for C&C and data exfiltration.
- **Rarog.** Rarog is a cryptocurrency mining Trojan that has been sold on various underground forums since June 2017 and has been used by countless criminals since then. Rarog has been primarily used to mine the Monero cryptocurrency. However, it can mine others. It comes equipped with a number of features, including providing mining statistics to users, configuring various processor loads for the running miner, the ability to infect USB devices, and the ability to load additional DLLs on the victim. Rarog provides an affordable way for new criminals to gain entry into this particular type of malware.

Key Terms

A *dynamic link library (DLL)* is a type of file used in Microsoft operating systems that enables multiple programs to simultaneously share programming instructions contained in a single file to perform specific functions.

Modern malware is typically stealthy and evasive, and now plays a central role in a coordinated attack against a target (see Section 1.2.2).

Advanced malware leverages networks to gain power and resilience, and can be updated — just like any other software application — so that an attacker can change course and dig deeper into the network or make changes and enact countermeasures.

This is a fundamental shift compared to earlier types of malware, which were more or less a swarm of independent agents that simply infected and replicated themselves. Increasingly, advanced malware has become a centrally coordinated, networked application in a very real sense. In much the same way that the Internet changed what was possible in personal computing, ubiquitous network access is 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 endpoint expanding the attack foothold and increasing the potential damage to the organization.

Some important characteristics and capabilities of advanced malware include:

- **Distributed, fault-tolerant architecture.** Advanced malware takes full advantage of the resiliency built into the Internet itself. Advanced malware can have multiple control servers distributed all over the world with multiple fallback options, and can also leverage other infected endpoints as communication channels, thus providing a near infinite number of communication paths to adapt to changing conditions or update code as needed.

- **Multi-functionality.** Updates from C&C servers can also completely change the functionality of advanced malware. This multifunctional capability enables an attacker to use various endpoints strategically to accomplish specific desired tasks such as stealing credit card numbers, sending spam containing other malware payloads (such as spyware), or installing ransomware for the purpose of extortion.

- **Polymorphism and metamorphism.** Some advanced malware has entire sections of code that serve no purpose other than to change the signature of the malware, thus producing an infinite number of unique signature hashes for even the smallest of malware programs. Techniques such as *polymorphism* and *metamorphism* are used to avoid detection by traditional signature-based anti-malware tools and software. For example, a change of just a single character or bit of the file or source code completely changes the *hash signature* of the malware.

- **Obfuscation.** Advanced malware often uses common *obfuscation* techniques to hide certain binary strings that are characteristically used in malware and therefore are easily detected by anti-malware signatures, or to hide an entire malware program.

Key Terms

Polymorphism alters part of the malware code with every iteration, such as the encryption key or decryption routine, but the malware payload remains unchanged.

Metamorphism uses more advanced techniques than polymorphism to alter malware code with each iteration. Although the malware payload changes with each iteration – for example, by using a different code structure or sequence, or by inserting garbage code to change the file size – the fundamental behavior of the malware payload remains unchanged.

A *hash signature* is a cryptographic representation of an entire file or program's source code.

Obfuscation is a programming technique used to render code unreadable. It can be implemented using a simple substitution cipher, such as an *exclusive or* (XOR) operation, in which the output is true only when the inputs are different (for example, TRUE and TRUE equals FALSE, but TRUE and FALSE equals TRUE), or more sophisticated encryption algorithms such as the *Advanced Encryption Standard* (AES). Alternatively, a *packer* can be used to compress a malware program for delivery, then decompress it in memory at run time.

1.4.2 Vulnerabilities and exploits

An exploit is a type of malware that takes advantage of a vulnerability in installed endpoint or server software such as a web browser, Adobe Flash, Java, or Microsoft Office. An attacker crafts an exploit that targets a software vulnerability, causing the software to perform functions or execute code on behalf of the attacker.

Vulnerabilities are routinely discovered in software at an alarming rate. Vulnerabilities may exist in software when the software is initially developed and released, or vulnerabilities may be inadvertently created, or even reintroduced, when subsequent version updates or security patches are installed. According to research by Palo Alto Networks, 78 percent of exploits take advantage of vulnerabilities that are less than two years old.

Security patches are developed by software vendors as quickly as possible after a vulnerability has been discovered in their software. However, an attacker may learn of a vulnerability and begin exploiting it before the software vendor is aware of the vulnerability or has an opportunity to develop a patch. This delay between the discovery of a vulnerability and development and release of a patch is known as a zero-day threat (or exploit). It may be months or years before a vulnerability is announced publicly. After a security patch becomes available, time inevitably is required for organizations to properly test and deploy the patch on

all affected systems. During this time, a system running the vulnerable software is at risk of being exploited by an attacker (see Figure 1-2).

Figure 1-2: Vulnerabilities can be exploited from the time software is deployed until it is patched.

Exploits can be embedded in seemingly innocuous data files (such as Microsoft Word documents, PDFs, and webpages), or they can target vulnerable network services. Exploits are particularly dangerous because they are often packaged in legitimate files that do not trigger anti-malware (or antivirus) software and are therefore not easily detected.

Creating an exploit data file is a two-step process. The first step is to embed a small piece of malicious code within the data file. However, the attacker still must trick the application into running the malicious code. Thus, the second part of the exploit typically involves memory corruption techniques that allow the attacker's code to be inserted into the execution flow of the vulnerable software. Once that happens, a legitimate application, such as a document viewer or web browser, will perform actions on behalf of the attacker, such as establishing communication and providing the ability to upload additional malware to the target endpoint. Because the application being exploited is a legitimate application, traditional signature-based antivirus and whitelisting software have virtually no effectiveness against these attacks.

Although there are many thousands of exploits, they all rely on a small set of core techniques that change infrequently. For example, a *heap spray* is an attempt to insert the attacker's code into multiple locations within the memory heap, hoping that one of those locations will be called by the process and executed. Some attacks may involve more steps, some may involve fewer, but typically three to five core techniques must be used to exploit an application. Regardless of the attack or its complexity, for the attack to be successful, the attacker must execute a series of these core exploit techniques in sequence, like navigating a maze to reach its objective (see Figure 1-3).

Key Terms

Heap spray is a technique used to facilitate arbitrary code execution by injecting a certain sequence of bytes into the memory of a target process.

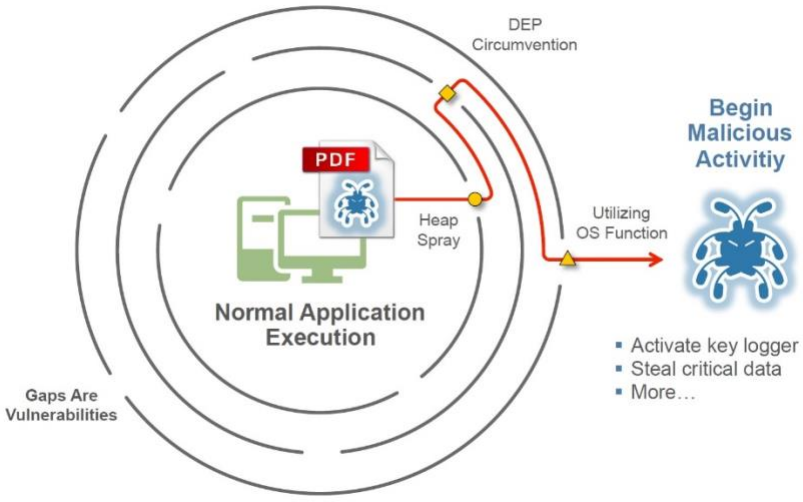

Figure 1-3: Exploits rely on a series of core attack techniques to succeed.

1.4.3 Spamming and phishing

Spam and phishing emails are the most common delivery methods for malware. The volume of spam email as a percentage of total global email traffic fluctuates widely from month to month – typically 45 to 75 percent. Although most end users today are readily able to identify spam emails and are more savvy about not clicking links, opening attachments, or replying to spam emails, spam remains a popular and effective infection vector for the spread of malware.

Phishing attacks, in contrast to spam, are becoming more sophisticated and difficult to identify.

Spear phishing is a targeted phishing campaign that appears more credible to its victims by gathering specific information about the target, and thus has a higher probability of success. A spear phishing email may spoof an organization (such as a financial institution) or individual that the recipient actually knows and does business with, and may contain very specific information (such as the recipient's first name, rather than just an email address). According to Symantec's *2018 Internet Security Threat Report*, "Spear-phishing emails emerged as by far the

most widely used infection vector, employed by 71 percent of [140 known targeted attack] groups."[11]

Whaling is a type of spear phishing attack that is specifically directed at senior executives or other high-profile targets within an organization. A whaling email typically purports to be a legal subpoena, customer complaint, or other serious matter.

Spear phishing, and phishing attacks in general, are not always conducted via email. A link is all that is required, such as a link on Facebook or on a message board, or a shortened URL on Twitter. These methods are particularly effective in spear phishing attacks because they allow the attacker to gather a great deal of information about the targets and then lure them through dangerous links into a place where the users feel comfortable.

Watering hole attacks compromise websites that are likely to be visited by a targeted victim, for example, an insurance company website that may be frequently visited by healthcare providers. The compromised website will typically infect unsuspecting visitors with malware (known as a "drive-by-download"). Watering hole attacks are the second most popular infection vector for targeted attack groups (24 percent), according to Symantec.[12]

A *pharming* attack redirects a legitimate website's traffic to a fake site, typically by modifying an endpoint's local hosts file or by compromising a DNS server ("DNS poisoning").

[11] "Internet Security Threat Report, Volume 23." Symantec. 2018. https://www.symantec.com/security-center/threat-report.

[12] Ibid.

Key Terms

Spear phishing is a highly targeted phishing attack that uses specific information about the target to make the phishing attempt appear legitimate.

Whaling is a type of spear phishing attack that is specifically directed at senior executives or other high-profile targets within an organization.

Watering hole attacks compromise websites that are likely to be visited by a targeted victim to deliver malware via a drive-by-download. A *drive-by-download* is a software download, typically malware, that happens without a user's knowledge or permission.

Pharming is a type of attack that redirects a legitimate website's traffic to a fake site.

1.4.4 Bots and botnets

Bots and *botnets* are notoriously difficult for organizations to detect and defend against using traditional anti-malware solutions.

Key Terms

Bots (or *zombies*) are individual endpoints that are infected with advanced malware that enables an attacker to take control of the compromised endpoint.

A *botnet* is a network of bots (often tens of thousands or more) working together under the control of attackers using numerous C&C servers.

In a botnet, advanced malware works together toward a common objective, with each bot growing the power and destructiveness of the overall botnet. The botnet can evolve to pursue new goals or adapt as different security countermeasures are deployed. Communication between the individual bots and the larger botnet through C&C servers provides resiliency in the botnet (see Figure 1-4).

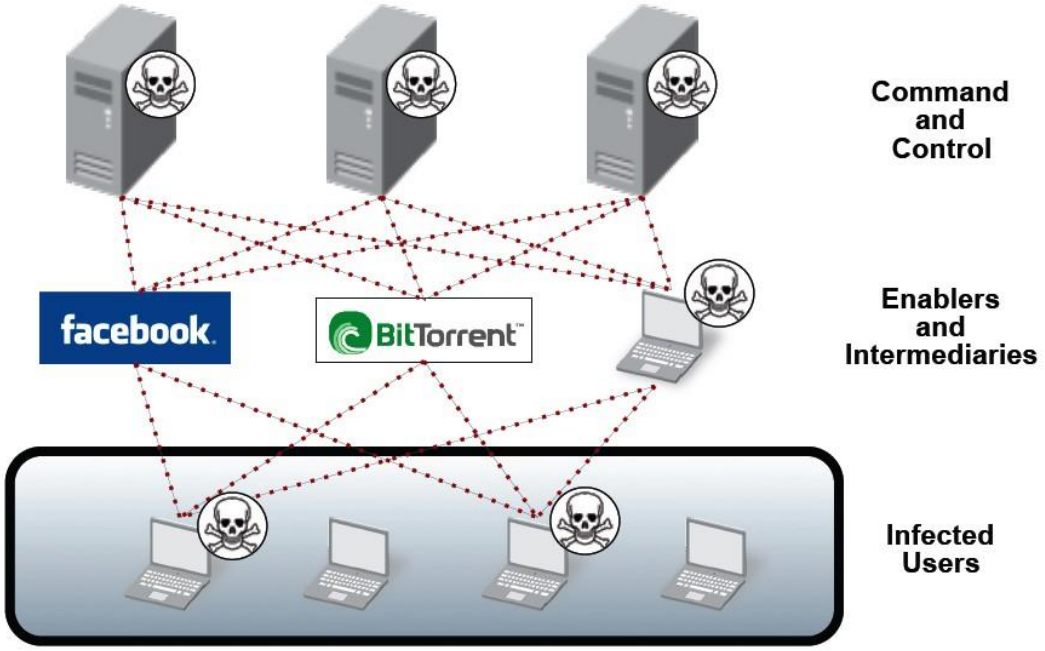

Figure 1-4: The distributed C&C infrastructure of a botnet.

Given their flexibility and ability to evade defenses, botnets present an enormous threat to organizations. The ultimate impact of a botnet is largely left up to the attacker, from sending spam one day to stealing credit card data the next – and far beyond, as many cyberattacks go undetected for months or even years.

Botnets themselves are dubious sources of income for cybercriminals. Botnets are created by cybercriminals to harvest computing resources (bots). Control of botnets (through C&C servers) can then be sold or rented out to other cybercriminals.

The key to "taking down" or "decapitating" a botnet is to separate the bots (infected endpoints) from their brains (C&C servers). If the bots cannot get to their servers, they cannot get new instructions, upload stolen data, or do any of the other things that make botnets so unique and dangerous.

Although this approach may seem straightforward, extensive resources are typically required to map the distributed C&C infrastructure of a botnet, and this approach almost always requires an enormous amount of investigation, expertise, and coordination between numerous industry, security, and law enforcement organizations worldwide.

Disabling of C&C servers often requires both physically seizing the servers and 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 C&C

infrastructure of a botnet. Many botnets have C&C servers all over the world, and will specifically function in countries that have little or no law enforcement for Internet crimes.

Further complicating takedown efforts, a botnet almost never relies on a single C&C server, but rather uses multiple C&C servers for redundancy purposes. Each server also 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. These evasion techniques make simply finding C&C servers a considerable challenge.

Most botnets are also designed to withstand the loss of a C&C server, meaning that the *entire* botnet C&C infrastructure must be disabled almost simultaneously. If any C&C server is accessible or any of the fallback options survive, the bots will be able to get updates, rapidly populate a completely new set of C&C servers, and the botnet will quickly recover. Thus, even a single C&C server remaining functional for even a small amount of time can give an attacker the window needed to update the bots and recover the entire botnet.

According to a 2017 botnet threat report, Spamhaus Malware Labs identified and issued Spamhaus Block List (SBL) listings for more than 9,500 botnet C&C servers on 1,122 different networks.[13] Botnet C&C servers are used to control infected endpoints (bots) and to exfiltrate personal and/or valuable data from bots. Botnets can be easily scaled up to send massive volumes of spam, spread ransomware, launch DDoS attacks, commit click-fraud campaigns, and/or mine cryptocurrency (such as Bitcoin).

1.4.4.1 Spamming botnets

The largest botnets are often dedicated to sending spam. The premise is fairly straightforward – the attacker attempts to infect as many endpoints as possible, which can then be used to send out spam email messages without the end users' knowledge. The relative impact of this type of bot on an organization may seem low initially, but an infected endpoint sending spam could consume additional bandwidth and ultimately reduce the productivity of the user and even the network itself. Perhaps more consequential, the organization's email domain and IP addresses could also easily become listed by various real-time blackhole lists (RBLs), causing legitimate

[13] "Spamhaus Botnet Threat Report 2017." Spamhaus Malware Labs. January 2018. https://www.spamhaus.org/news/article/772/spamhaus-botnet-threat-report-2017.

emails to be labeled as spam and blocked by other organizations, and damaging the reputation of the organization.

The Rustock botnet is an example of a spamming botnet. It could send up to 25,000 spam email messages per hour from an individual bot and, at its peak, sent an average of 192 spam emails per minute per bot. Rustock is estimated to have infected more than 2.4 million computers worldwide. In March 2011, the U.S. Federal Bureau of Investigation (FBI), working with Microsoft and others, was able to take down the Rustock botnet, which had operated for more than 5 years and at the time was responsible for sending up to 60 percent of the world's spam.

1.4.4.2 DDoS botnets

A *distributed denial-of-service (DDoS)* attack is a type of cyberattack in which extremely high volumes of network traffic such as packets, data, or transactions are sent to the target victim's network to make their network and systems (such as an e-commerce website or other web application) unavailable or unusable. A DDoS botnet uses bots as part of a DDoS attack, overwhelming a target server or network with traffic from a large number of bots. In such attacks, the bots themselves are not the target of the attack. Instead, the bots are used to flood some other remote target with traffic. The attacker leverages the massive scale of the botnet to generate traffic that overwhelms the network and server resources of the target.

Unlike other types of cyberattacks, a DDoS attack does not typically employ a prolonged, stealthy approach. Instead, a DDoS attack more often takes the form of a highly visible brute-force attack that is intended to rapidly cause damage to the victim's network and systems infrastructure, and to their business and reputation.

DDoS attacks often target specific organizations for personal or political reasons, or to extort a ransom payment in exchange for stopping the DDoS attack. DDoS attacks are often used by hacktivists (discussed in Section 1.2.1) to promote or protest a particular political agenda or social cause. DDoS attacks may also be used for criminal extortion purposes to extract a hefty ransom payment in exchange for ending the attack.

DDoS botnets represent a dual risk for organizations: The organization itself can be the target of a DDoS attack, and even if the organization isn't the ultimate target, any infected endpoints participating in the attack will consume valuable network resources and facilitate a criminal act, albeit unwittingly.

A DDoS attack can also be used as part of a targeted strategy for a later attack. While the victim organization is busy defending against the DDoS attack and restoring the network and systems, the attacker can deliver an exploit to the victim network (for example, by causing a buffer overflow in a SQL database) that will enable a malware infection and establish a foothold in the

network. The attacker can then return later to expand the (stealthy) attack and extract stolen data.

Some examples of recent DDoS attacks include attacks against domain name registrars Melbourne IT and Dreamhost in April 2017 and August 2017, respectively. The UK National Lottery was targeted in September 2017. Electroneum, a cryptocurrency startup, was the victim of a DDoS attack just prior to the launch of its mobile mining app in November 2017. The Boston Globe was also targeted in November 2017, not only disrupting the bostonglobe.com website, but also the newspaper's telephones, editing system, and other company-owned websites.

1.4.4.3 Financial botnets

Financial botnets, such as ZeuS and SpyEye, are responsible for the direct theft of funds from all types of enterprises. These types of botnets are typically not as large as spamming or DDoS botnets, which grow as large as possible for a single attacker. Instead, financial botnets are often sold as kits that allow attackers to license the code and build their own botnets.

The impact of a financial breach can be enormous, including the breach of sensitive consumer and financial information leading to significant financial, legal, and brand damage. As reported by Tech Republic:

> "*A Mirai botnet variant was used in attacks against at least one financial sector company in January 2018—possibly the first time an IoT botnet has been observed in use in a DDoS attack since the Mirai botnet took down multiple websites in 2017, according to a Thursday report from Recorded Future*."[14]

[14] Rayome, Alison DeNisco. "Mirai variant botnet launches IoT DDoS attacks on financial sector." Tech Republic. April 5, 2018. https://www.techrepublic.com/article/mirai-variant-botnet-launches-iot-ddos-attacks-on-financial-sector/.

1.4 Knowledge Check

Test your understanding of the fundamentals in the preceding section. Review the correct answers in the Appendix at the end of this guide.

1. **Multiple Choice.** Which option describes malicious software or code that typically damages, takes control of, or collects information from an infected endpoint? (Choose one.)

    a) exploit
    b) malware
    c) vulnerability
    d) none of the above

2. **Multiple Choice.** Which of the following is an important characteristic or capability of advanced malware? (Choose one.)

    a) distributed, fault-tolerant architecture
    b) multi-functionality
    c) hiding techniques such as polymorphism, metamorphism, and obfuscation
    d) all of the above

3. **True or False.** A vulnerability is a small piece of software code, part of a malformed data file, or a sequence (string) of commands created by an attacker to cause unintended or unanticipated behavior in a system or software.

4. **True or False.** New exploits can be crafted from any number of more than a thousand core exploit techniques.

1.5 Wi-Fi and Advanced Threats

This section describes Wi-Fi vulnerabilities and attacks, and advanced persistent threats (APTs).

1.5.1 Wi-Fi vulnerabilities

With the explosive growth in the number of mobile devices over the past decade, wireless (Wi-Fi) networks are now everywhere. Whether you're in an office, hotel, airport, school, or coffee shop, you're likely in range of a Wi-Fi network somewhere.

Of course, as a security professional, your first concern when trying to get connected is "how secure is this Wi-Fi network?" But for the average user, the unfortunate reality is that Wi-Fi connectivity is more about convenience than security.

Thus, the challenge is to not only secure your Wi-Fi networks, but also to protect the mobile devices that your organization's employees use to perform work and access potentially sensitive data — no matter where they are or whose network they're on.

Wi-Fi security begins — and ends — with authentication. If you can't control who has access to your wireless network, then you can't protect your network.

1.5.1.1 Wired equivalent privacy

The wired equivalent privacy (WEP) protocol was the wireless industry's first attempt at security. As its name falsely implies, WEP was intended to provide data confidentiality equivalent to the security of a wired network. However, WEP had many well-known and well-publicized weaknesses, and simply wasn't effective for establishing a secure wireless network. Today, WEP is not even an option in most wireless network configurations and, according to statistics from Kaspersky Security Network (KSN), WEP was used in about 3 percent of nearly 32 million Wi-Fi hotspots accessed by KSN users worldwide in 2016.[15]

One critical weakness in WEP is in how it handles the *initialization vector* (IV) for WEP's RC4 (Rivest Cipher 4) stream cipher. In WEP, the IV is a 24-bit key that is transmitted in the clear (or unencrypted). With a 24-bit key, generation of unique values becomes impossible after sending 2^24 (or 16,777,216) packets, and the IVs will repeat. In a secure environment, the key should be replaced before exhausting the IVs, but you have no way to automate the process in WEP. Thus, given enough traffic, IV collisions will occur, which, in conjunction with other techniques, can help an attacker to deduce the WEP key.

A WEP key can be deduced by passively monitoring and examining the network traffic using a wireless network card in *promiscuous mode*. Passive monitoring leaves no indication that the *wireless access point* (AP) is under attack, because the attacker is doing nothing more than making copies of the packets on the network.

---

[15] Legezo, Denis. "Research on unsecured Wi-Fi networks across the world." SecureList. November 24, 2016. https://securelist.com/research-on-unsecured-wi-fi-networks-across-the-world/76733/.

An attacker sending traffic directly to the target AP can reduce the time to break into a WEP-enabled network from days to minutes, thus making WEP not much better than the 22 percent of open Wi-Fi networks worldwide in 2016.[16]

Key Terms

An *initialization vector* (IV) or *nonce* is a random number used only once in a session, in conjunction with an encryption key, to protect data confidentiality.

In computer networking, *promiscuous mode* refers to Ethernet hardware, typically a network interface card (NIC), that receives all traffic on a network segment, even if the traffic is not addressed to the hardware.

A *wireless access point* (AP) is a network device that connects to a router or wired network and transmits a Wi-Fi signal so that wireless devices can connect to a wireless (or Wi-Fi) network.

Attacks on WEP don't depend on having a massive amount of computing power and aren't greatly affected by the size of the encryption key. The attack isn't dependent on how complex the original passphrase is either. It's simply a matter of being able to collect enough traffic.

Once it became apparent that WEP had critical, unfixable security flaws, efforts took place immediately to develop a successor. Because a replacement for WEP was urgently needed, an interim standard, Wi-Fi Protected Access (WPA) was published in 2003. WPA was further refined as WPA2 in 2004, and WEP was then deprecated as a Wi-Fi security standard.

1.5.1.2 Wi-Fi protected access (WPA/WPA2/WPA3)

WPA was published as an interim standard in 2003, quickly followed by WPA2 in 2004. WPA/WPA2 contains improvements to protect against the inherent flaws in WEP. These improvements included changes to the encryption to avoid many of the problems that plagued WEP.

WPA2 can be implemented in different ways. WPA2-Enterprise, also known as WPA2-802.1x mode, uses the *extensible authentication protocol* (EAP) and *remote authentication dial-in user*

[16] Ibid.

service (RADIUS) for authentication. Numerous EAP types are also available for use in WPA2-Enterprise.

However, the use of a *pre-shared key* (PSK) is by far the most common, particularly in homes, small businesses, and guest Wi-Fi networks. WPA2-PSK can be implemented with just the AP and the client, requiring neither a third-party 802.1*x* authentication server nor individual user accounts.

Key Terms

The *extensible authentication protocol* (EAP) is a widely used authentication framework that includes about 40 different authentication methods.

remote authentication dial-in user service (RADIUS) is a client/server protocol and software that enables remote access servers to communicate with a central server to authenticate users and authorize access to a system or service.

A *pre-shared key* (PSK) is a shared secret, used in symmetric key cryptography, which has been exchanged between two parties communicating over an encrypted channel.

WPA2-PSK supports 256-bit keys, which require 64 hexadecimal characters. Because requiring users to enter a 64-hexadecimal character key is impractical, WPA2 includes a function that generates a 256-bit key based on a much shorter passphrase created by the administrator of the Wi-Fi network and the *service set identifier* (SSID) of the AP used as a *salt* for the *one-way hash function*.

In WPA2, the name of the SSID is used for the salt. An easy way to make your Wi-Fi security stronger (and make *rainbow table* attacks impractical) is to simply change your SSID to something that isn't common or easily guessed.

Key Terms

A *service set identifier* (SSID) is a case-sensitive, 32-character alphanumeric identifier that uniquely identifies a Wi-Fi network.

In cryptography, a *salt* is randomly generated data that is used as an additional input to a one-way hash function that "hashes" a password or passphrase. The same original text hashed with different salts results in different hash values.

A *one-way (hash) function* is a mathematical function that creates a unique representation (a hash value) of a larger set of data in a manner that is easy to compute in one direction (input to output), but not in the reverse direction (output to input). The hash function can't recover the original text from the hash value. However, an attacker could attempt to guess what the original text was and see if it produces a matching hash value.

A *rainbow table* is a pre-computed table used to find the original value of a cryptographic hash function.

To execute an attack on a WPA2 passphrase, an attacker needs to be able to test a large number of passphrase candidates. So although WPA2 remains cryptographically secure (namely the key isn't recoverable by simple observation of the traffic as with WEP), methods do exist to test passphrases offline by gathering the handshake packets between the AP and a legitimate user.

To collect the necessary packets to crack a WPA2 passphrase, an attacker could passively gather traffic when a legitimate user joins the network. This method requires time however, because the attacker does not know when someone will join the network.

For an impatient attacker, the solution is to employ an active attack. As long as a legitimate user is already online, the attacker can kick the user's client device off the AP with forged de-authentication packets. After getting knocked off, the client device will automatically attempt to reconnect, thus providing the attacker with the handshake packets needed for offline passphrase analysis. Thus, unlike with WEP, attacks on WPA2 can be done without spending a significant amount of time in the proximity of the target network, once the handshake packets have been captured.

Next, the attacker must recover (or find) the passphrase itself, which requires the following:

- A test to check millions of potential passphrases until it finds the correct passphrase. To avoid detection, an attacker can't use the actual target, because the victim would be

able to see this attack activity. The alternative is to use an offline method of testing that uses the handshake packets.

- A methodology to guess passphrases. The worst-case scenario is to "brute force" the passphrase, trying every possible combination of numbers and characters until a correct value is found. This effort can produce a correct result given enough time and computing power. However, it's much faster to take educated guesses without having to resort to brute force. By using educated guesses on possible passphrase candidates, the attacker can attempt a much shorter list.

This basic process for recovering Wi-Fi passphrases is similar to cracking user passwords. In the early days of password cracking, an attacker might have knowledge of a target system's one-way hash function and a list of the system's user password hash values. However, the attacker had no way to decrypt the password, because the original text isn't recoverable from a hash. But by encrypting a list of words with the same one-way hash function (a dictionary attack), an attacker can then compare the resulting hash values with the hash values stored for the various user accounts on the system. So, although the password itself isn't decrypted, a given input that produces a given result, a password match, can be found. With the addition of more computing power, an attacker could try longer word lists and a greater number of variations of each word. The process for attacking WPA2 passphrases is similar.

WPA3 was published in 2018 and introduces security enhancements such as more robust brute-force attack protection, improved "hotspot" and guest access security, simpler integration with devices that have limited or no user interface (such as IoT devices), and a 192-bit security suite. Newer Wi-Fi routers and client devices will likely support both WPA2 and WPA3 to ensure backward compatibility in mixed environments.

According to the Wi-Fi Alliance, WPA3 features include improved security for Internet of things (IoT) devices such as smart bulbs, wireless appliances, smart speakers, and other screen-free gadgets that make everyday tasks easier. The Wi-Fi Alliance hasn't outlined the specifics yet, but WPA3 is expected to support a one-touch setup system that'll make devices without screens (such as IoT devices and smart speakers like Google Home and Amazon Echo) easier to connect. It will be similar to the existing Wi-Fi Protected Setup protocol, which involves pushing a button on the router to connect a device.

According to a recent VentureBeat article, WPA3 also "supports a much stronger encryption algorithm than WPA2… intended for industrial, defense, and government applications rather than homes and offices. Specifically, it includes a 192-bit security suite that's aligned with the

Commercial National Security Algorithm (CNSA) Suite, a feature requested by the Committee on National Security Systems (CNSS), a part of the U.S. National Security Agency [NSA]."[17]

WPA3 provides protection against brute-force dictionary attacks by implementing "a robust handshake [called the Dragonfly protocol, also referred to as Simultaneous Authentication of Equals] that isn't vulnerable to wireless exploits like KRACK, and it hardens security at the time when the network key is exchanged between a device and the access point."[18] By limiting the number of network password attempts on a per-user basis, WPA3 also reduces the efficacy of common dictionary attacks.

"WPA3 introduces Opportunistic Wireless Encryption (OWE), or individualized data encryption, which encrypts every connection between a device and the router with a unique key. Even if the access point doesn't require a password, your device's data won't be exposed to the wider network."[19]

1.5.2 Wi-Fi man-in-the-middle attacks

Instead of breaking into a wireless network, an attacker can trick victims into connecting to a wireless network that the attacker controls. These techniques are part of a larger set of attacks known as man-in-the-middle attacks. With a man-in-the-middle exploit in place on a Wi-Fi network, an attacker can serve up practically any content, for example:

- If a user attempts to download a legitimate file, the attacker can send mobile malware instead.
- When a user attempts to visit a legitimate webpage, the attacker can alter the content to exploit a vulnerability that exists in the device's browser, allowing the attacker to further escalate an attack.

---

[17] Wiggers, Kyle. "What is WPA3, why does it matter, and when can you expect it?" VentureBeat. May 19, 2018. https://venturebeat.com/2018/05/19/what-is-wpa3-why-does-it-matter-and-when-can-you-expect-it/.

[18] Ibid.

[19] Ibid.

- Email addresses and financial account information can be harvested from the connected endpoint, enabling an attacker to create a very targeted and convincing phishing attack to trick even more users on a network into disclosing sensitive information.

1.5.2.1 Evil Twin

Perhaps the easiest way for an attacker to find a victim to exploit is to set up a wireless access point that serves as a bridge to a real network. An attacker can inevitably bait a few victims with "free Wi-Fi access."

The main problem with this approach is that it requires a potential victim to stumble on the access point and connect. The attacker can't easily target a specific victim because the attack depends on the victim initiating the connection.

A slight variation on this approach is to use a more specific name that mimics a real access point normally found at a particular location — the Evil Twin. For example, if your local airport provides Wi-Fi service and calls it "Airport Wi-Fi," the attacker might create an access point with the same name using an access point that has two radios. Average users cannot easily discern when they are connected to the real access point or a fake one, so this approach would catch a greater number of users than trying to attract victims at random. Still, the user has to select the network so there's a bit of chance involved in trying to reach a particular target.

The main limitation of the Evil Twin attack is that the attacker can't choose the victim. In a crowded location, the attacker will be able to get a large number of people connecting to the wireless network to unknowingly expose their account names and passwords. However, it's not an effective approach if the goal is to target employees in a specific organization.

1.5.2.2 Jasager

To understand a more targeted approach than the Evil Twin attack, think about what happens when you bring your wireless device back to a location that you've previously visited. For example, when you bring your laptop home, you don't have to choose which access point to use because your device remembers the details of wireless networks to which it has previously connected. The same goes for visiting the office or your favorite coffee shop.

Your mobile device detects when it's within the proximity of a previously known wireless network by sending a beacon out to see if a preferred network is within range. Under normal conditions, when a wireless device sends out a beacon, the nonmatching access points ignore it. The beacon goes unanswered, except when it comes within the proximity of the preferred network.

The Jasager attack takes a more active approach toward beacon requests. Jasager, German for "the Yes man," responds to all beacon requests, thus taking a very permissive approach toward who can connect. The user doesn't have to manually choose the attacker's access point. Instead, the attacker pretends to be whatever access point the user normally connects to (see Figure 1-5). Instead of trying to get victims to connect at random, now the attacker simply needs to be within the proximity of the target.

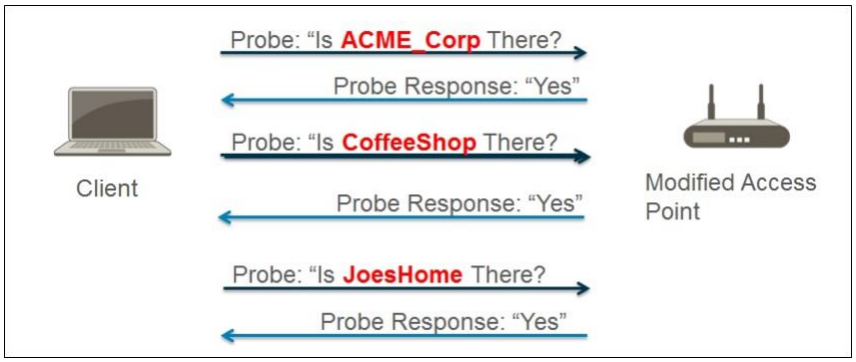

Figure 1-5: Jasager pretends to be whatever access point is requested by the client's beacon.

This process intercepts the communication from laptops, mobile phones, and tablets. Many (if not most) 3G/4G/LTE mobile devices automatically switch to Wi-Fi when they recognize that they are near a network that they know.

An attacker can use the same method to capture WPA2 handshake packets (discussed in Section 1.5.1) to disconnect users from a Wi-Fi network by using forged de-authentication packets. When the users reconnect, they'll unwittingly connect to the modified access point. Unlike the Evil Twin attack, the attacker doesn't have to just wait for a victim to connect to the modified access point; with this approach, everyone that's in the vicinity will automatically connect and become a potential victim.

Jasager runs on any number of devices, but perhaps one of the most effective ways to employ it is with the Pineapple access point. The Pineapple is simply an access point with modified firmware that embeds a number of tools for wireless "penetration" testing. It also has a number of accessories, such as support for cellular USB cards to provide network connectivity when it is otherwise unavailable at the target location, and battery packs to operate as a standalone unit. It's also easily concealed because it can be disguised within any number of housings typically found plugged in at the office.

Once the attacker has the victim connected to a malicious access point, the man-in-the-middle attack can proceed and the attacker can not only observe and capture network traffic, but also modify it.

1.5.2.3 SSLstrip

After a user connects to a Wi-Fi network that's been compromised — or to an attacker's Wi-Fi network masquerading as a legitimate network — the attacker can control the content that the victim sees. The attacker simply intercepts the victim's web traffic, redirects the victim's browser to a web server that it controls, and serves up whatever content the attacker desires.

A man-in-the middle attack can be used to steal a victim's online banking or corporate email account credentials. Normally, this type of traffic would be considered safe because the webpage typically uses secure sockets layer (SSL) encryption. Of course, the average user only knows that a padlock somewhere in the address bar of means that their browser is secure, correct?

But the padlock appears differently, and in different locations, in different browsers. How does the padlock appear in Internet Explorer? What about Mozilla Firefox, Google Chrome, and Apple Safari? And it appears differently on different smartphones and tablets too! It's no wonder that typical end users — even many security professionals — can be easily tricked.

SSLstrip strips SSL encryption from a "secure" session. When a user connected to a compromised Wi-Fi network attempts to initiate an SSL session, the modified access point intercepts the SSL request (see Figure 1-6). The modified access point then completes the SSL session on behalf of the victim's device. Then, the SSL tunnel between the victim's device and the legitimate secure web server is actually terminated — and decrypted — on the modified access point, thus allowing the attacker to see the victim's credentials, and other sensitive information, in cleartext.

With SSLstrip, the modified access point displays a fake padlock in the victim's web browser. Webpages can display a small icon called a *favicon* next to a website address in the browser's address bar. SSLstrip replaces the favicon with a padlock that looks like SSL to an unsuspecting user.

Key Terms

A *favicon* ("favorite icon") is a small file containing one or more small icons associated with a particular website or webpage.

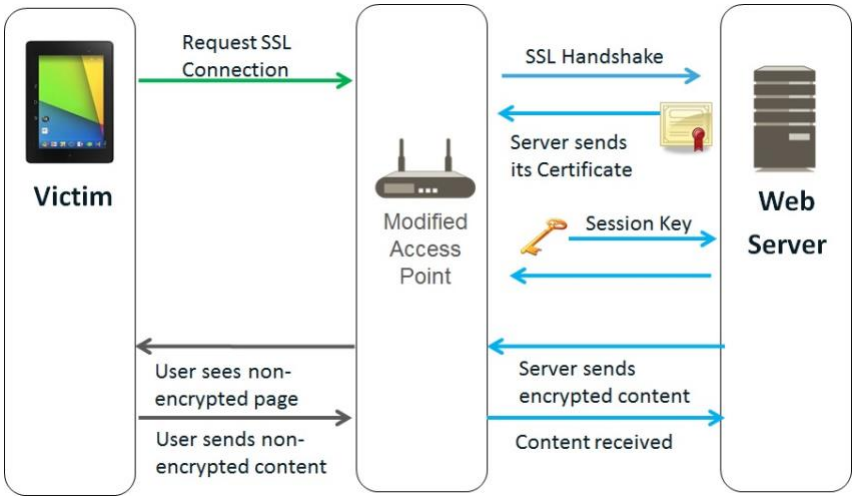

Figure 1-6: Man-in-the-middle with SSLstrip.

1.5.3 Advanced Persistent Threats

Advanced persistent threats (APTs) are a class of threats that are far more deliberate and potentially devastating than other types of cyberattacks. As its name implies, an APT has three defining characteristics. An APT is:

- **Advanced.** Attackers use advanced malware and exploits and typically also have the skills and resources necessary to develop additional cyber-attack tools and techniques, and may have access to sophisticated electronic surveillance equipment, satellite imagery, and even human intelligence assets.
- **Persistent.** An APT may take place over a period of several years. The attackers pursue specific objectives and use a "low-and-slow" approach to avoid detection. The attackers are well organized and typically have access to substantial financial backing to fund their activities, such as a nation-state or organized criminal organization.
- **Threat.** An APT is deliberate and focused, rather than opportunistic. APTs are designed to cause real damage including significant financial loss, destruction of systems and infrastructure, or physical harm and loss of life.

Some recent examples of APT campaigns include:

- **MONSOON.** Monsoon is an APT campaign that appears to have begun in December 2015. According to Forcepoint Security Labs, "The overarching campaign appears to target both Chinese nationals within different industries and government agencies in

Southern Asia."[20] As of July 2016, more than 110 different victim countries and 6,300 victim IP addresses had been identified.[21] "The malware components used in MONSOON are typically distributed through [weaponized] documents sent through e-mail to specifically chosen targets. Themes of these documents are usually political in nature and taken from recent publications on topical current affairs. Several malware components have been used in this operation including Unknown Logger Public, TINYTYPHON, BADNEWS, and an AutoIt [3] back door."[22]

- **1937CN.** In 2017, FortiGuard Labs discovered several malicious documents that exploited the CVE-2012-0158 buffer overflow vulnerability (ListView/TreeView ActiveX controls in the MSCOMCTL.OCX library). "It was believed… that the hacking campaign where these documents were used was led by the Chinese hacking group 1937CN. The link to the group was found through malicious domains used as command and control servers by the attacker…. Similar to other APT attacks, such as MONSOON APT, this APT uses DLL hijacking to evade [Host Intrusion Prevention Systems, or HIPS]."[23]

- **Scarlet Mimic.** The Scarlet Mimic attacks began in 2011. Their targeting pattern suggests that this adversary's primary mission is to gather information about minority rights activists. Although there is no evidence directly linking these attacks to a government source, the information derived from their activities supports an assessment that a group (or groups) with motivations similar to the stated position of the Chinese government in relation to Uyghur and Tibetan activists, and those who are interested in their causes, is involved. "The Scarlet Mimic attacks primarily center around the use of a Windows back door named 'FakeM.' It was first described by Trend Micro in 2013 and was named FakeM because its primary command and control traffic mimicked Windows

---

[20] Settle, Andy, Nicholas Griffin, and Abel Toro. "Monsoon – Analysis of an APT Campaign: Espionage and Data Loss Under the Cover of Current Affairs. Forcepoint™ Security Labs™. 2016. https://www.forcepoint.com/sites/default/files/resources/files/forcepoint-security-labs-monsoon-analysis-report.pdf.

[21] Ibid.

[22] Ibid.

[23] Manuel, Jasper and Artem Semenchenko. "Rehashed RAT Used in APT Campaign Against Vietnamese Organizations." Fortinet®. September 5, 2017. https://www.fortinet.com/blog/threat-research/rehashed-rat-used-in-apt-campaign-against-vietnamese-organizations.html.

Messenger and Yahoo! Messenger network traffic to evade detection."[24] Scarlet Mimic has also "deployed Trojans that target the Mac OS X and Android operating systems."[25]

- **Lazarus.** The Lazarus APT group is a threat actor linked to North Korea and believed to be behind attacks targeting U.S. defense contractors and other worldwide attack campaigns, including the Bangladesh cyber heist (US$81 million was surreptitiously transferred from the New York Federal Reserve Bank account of Bangladesh in February 2016)[26], the Troy Operation (attacks against South Korean infrastructure in 2013)[27], the DarkSeoul Operation (malware-based attacks that wiped tens of thousands of hard drives belonging to South Korean television networks and banks in March 2013)[28], and the Sony Picture hack (employees' emails and personal information including salaries, addresses, and Social Security Numbers revealed, unreleased movies posted on file sharing sites, and internal computer systems shut down in 2014).[29]

---

[24] Falcone, Robert and Jen Miller-Osborn. "Scarlet Mimic: Years-Long Espionage Campaign Targets Minority Activists." Palo Alto Networks. January 24, 2016. https://researchcenter.paloaltonetworks.com/2016/01/scarlet-mimic-years-long-espionage-targets-minority-activists/.

[25] Ibid.

[26] Paganini, Pierluigi. "US blames North Korea for the $81 million Bangladesh cyber heist." Security Affairs. March 24, 2017. http://securityaffairs.co/wordpress/57396/cyber-crime/bangladesh-cyber-heist.html.

[27] Paganini, Pierluigi. "Hackers hit South Korea also spread spyware to steal military secrets." Security Affairs. July 9, 2013. http://securityaffairs.co/wordpress/16014/hacking/hackers-hit-south-korea-spyware-steal-military-secrets.html.

[28] Ibid.

[29] Weisman, Aly. "A Timeline of the Crazy Events in the Sony Hacking Scandal." Business Insider. December 9, 2014. http://www.businessinsider.com/sony-cyber-hack-timeline-2014-12.

1.5 Knowledge Check

Test your understanding of the fundamentals in the preceding section. Review the correct answers in the Appendix at the end of this guide.

1. **True or False.** Wired Equivalent Privacy (WEP) is an effective protocol for securing wireless networks.

2. **Fill in the Blank.** A _____ is a mathematical function that creates a unique representation of a larger set of data in a manner that is easy to compute in one direction, but not in the reverse direction.

Module 2 – Cybersecurity Gateway

Knowledge Objectives

- Describe the basic operation of computer networks and the Internet, including common networking devices, routed and routing protocols, different types of area networks and topologies, and the domain name system (DNS).
- Explain the function of physical, logical, and virtual addressing in networking.
- Discuss IPv4 and IPv6 addressing and subnetting fundamentals.
- Discuss the OSI Reference Model and TCP/IP model including packet analysis, protocol and packet filtering, and TCP/IP encapsulation.
- Explain various network security models, concepts, and principles including perimeter-based security and the Zero Trust model.
- Discuss cloud and data center security design concepts including cloud computing security considerations and requirements, the role of virtualization in cloud computing, existing data security solution weaknesses, east-west traffic protection, and security in virtualized data centers.
- Describe network security technologies including firewalls, Intrusion Detection Systems and Intrusion Prevention Systems (IDS/IPS), web content filters, virtual private networks (VPNs), data loss prevention (DLP), unified threat management (UTM), and security information and event management (SIEM).
- Explain cloud, virtualization, and storage security concepts and challenges.
- Discuss network operations concepts including traffic analysis, troubleshooting, and server and systems administration.
- Discuss directory services, network device optimization, and structured host and network troubleshooting.
- Describe IT Infrastructure Library (ITIL) concepts and help desk and technical support functions.

2.1 The Connected Globe

With more than four billion Internet users worldwide in 2018, which represents well over half the world's population, the Internet connects businesses, governments, and people across the globe. Our reliance on the Internet will continue to grow, with nearly 30 billion devices and "things" – including autonomous vehicles, household appliances, wearable technology, and more – connecting to the Internet of Things (IoT) and more than 7.3 billion worldwide smartphone subscriptions each downloading 17 gigabytes (GB) of monthly data by 2023[30].

2.1.1 The NET: How things connect

In the 1960s, the U.S. Defense Advanced Research Project Agency (DARPA) created ARPANET, the precursor to the modern Internet. ARPANET was the first packet switching network. A packet switching network breaks data into small blocks (packets), transmits each individual packet from node to node toward its destination, then reassembles the individual packets in the correct order at the destination.

Today, hundreds of millions of routers (discussed in Section 2.1.2) deliver transmission control protocol/internet protocol (TCP/IP) packets using various routing protocols (discussed in Section 2.1.3) across local area networks and wide area networks (LANs and WANs, respectively, discussed in Section 2.1.4). The domain name system (DNS, discussed in Section 2.1.5) enables Internet addresses, such as www.paloaltonetworks.com, to be translated into routable IP addresses.

2.1.2 Introduction to networking devices

Routers are physical or virtual devices that send data packets to destination networks along a network path using logical addresses (discussed in Section 2.2). Routers use various routing protocols (discussed in Section 2.1.3) to determine the best path to a destination, based on variables such as bandwidth, cost, delay, and distance. A wireless router combines the functionality of a router and a wireless access point (AP) to provide routing between a wired and wireless network. An AP is a network device that connects to a router or wired network and transmits a Wi-Fi signal so that wireless devices can connect to a wireless (or Wi-Fi)

[30] "Ericsson Mobility Report, November 2017." Ericsson. November 2017. https://www.ericsson.com/assets/local/mobility-report/documents/2017/ericsson-mobility-report-november-2017.pdf.

network. A *wireless repeater* rebroadcasts the wireless signal from a wireless router or AP to extend the range of a Wi-Fi network.

A *hub* (or *concentrator*) is a network device that connects multiple devices – such as desktop computers, laptop docking stations, and printers – together on a local area network (LAN). Network traffic that is sent to a hub is broadcast out of all ports on the hub, which can create network congestion and introduces potential security risks (broadcast data can be intercepted).

A *switch* is essentially an intelligent hub that uses physical addresses (discussed in Section 2.2) to forward data packets to devices on a network. Unlike a hub, a switch is designed to forward data packets only to the port that corresponds to the destination device. This transmission method (referred to as micro-segmentation) creates separate network segments and effectively increases the data transmission rates available on the individual network segments. Additionally, a switch can be used to implement *virtual LANs* (VLANs), which logically segregate a network and limit *broadcast domains* and *collision domains*.

Key Terms

A *router* is a network device that sends data packets to a destination network along a network path.

A *wireless repeater* rebroadcasts the wireless signal from a wireless router or AP to extend the range of a Wi-Fi network.

A *hub* (or *concentrator*) is a device used to connect multiple networked devices together on a local area network (LAN).

A *switch* is an intelligent hub that forwards data packets only to the port associated with the destination device on a network.

A *virtual LAN* (VLAN) is a logical network that is created within a physical local area network.

A *broadcast domain* is the portion of a network that receives broadcast packets sent from a node in the domain.

A *collision domain* is a network segment on which data packets may collide with each other during transmission.

2.1.3 Routed and routing protocols

Routed protocols, such as the *internet protocol* (IP), address packets with routing information that enables those packets to be transported across networks using routing protocols. IP is discussed further in Sections 2.2 and 2.2.1.

Routing protocols are defined at the Network layer of the OSI model (discussed in Section 2.3.1) and specify how routers communicate with one another on a network. Routing protocols can either be *static* or *dynamic*.

> **Key Terms**
> An *internet protocol (IP) address* is a 32- or 128-bit identifier assigned to a networked device for communications at the Network layer of the OSI model (discussed in Section 2.3.1) or the Internet layer of the TCP/IP model (discussed in Section 2.3.2).

A *static* routing protocol requires that routes be created and updated manually on a router or other network device. If a static route is down, traffic can't be automatically rerouted unless an alternate route has been configured. Also, if the route is congested, traffic can't be automatically rerouted over the less congested alternate route. Static routing is practical only in very small networks or for very limited, special-case routing scenarios (for example, a destination that's reachable only via a single router or when used as a backup route). However, static routing has low bandwidth requirements (routing information isn't broadcast across the network) and some built-in security (users can only get to destinations that are specified in statically defined routes).

A dynamic routing protocol can automatically learn new (or alternate) routes and determine the best route to a destination. The routing table is updated periodically with current routing information. Dynamic routing protocols are further classified as:

- **Distance-vector.** A distance-vector protocol makes routing decisions based on two factors: the distance (hop count or other metric) and vector (the egress router interface). It periodically informs its peers and/or neighbors of topology changes. *Convergence*, the time it takes for all routers in a network to update their routing tables with the most current information (such as link status changes), can be a significant problem for distance-vector protocols. Without convergence, some routers in a network may be unaware of topology changes, which causes the router to send traffic to an invalid destination. During convergence, routing information is exchanged between

routers, and the network slows down considerably. Convergence can take several minutes in networks that use distance-vector protocols.

routing information protocol (RIP) is an example of a distance-vector routing protocol that uses *hop count* as its routing metric. To prevent routing loops, in which packets effectively get stuck bouncing between various router nodes, RIP implements a hop limit of 15, which limits the size of networks that RIP can support. After a data packet crosses 15 router nodes (hops) between a source and a destination, the destination is considered unreachable. In addition to hop limits, RIP employs four other mechanisms to prevent routing loops:

- **Split horizon.** Prevents a router from advertising a route back out through the same interface from which the route was learned.
- **Triggered Updates.** When a change is detected the update gets sent immediately instead of waiting 30 seconds to send a RIP update.
- **Route poisoning.** Sets the hop count on a bad route to 16, which effectively advertises the route as unreachable.
- **Holddown timers.** Causes a router to start a timer when the router first receives information that a destination is unreachable. Subsequent updates about that destination will not be accepted until the timer expires. This timer also helps avoid problems associated with flapping. Flapping occurs when a route (or interface) repeatedly changes state (up, down, up, down) over a short period of time.

Key Terms

Convergence is the time it takes for all routers in a network to update their routing tables with the most current routing information about the network.

Hop count generally refers to the number of router nodes that a packet must pass through to reach its destination.

- **Link-state.** A link-state protocol requires every router to calculate and maintain a complete map, or routing table, of the entire network. Routers that use a link-state protocol periodically transmit updates that contain information about adjacent connections, or link states, to all other routers in the network. Link-state protocols are

compute-intensive, but they can calculate the most efficient route to a destination. They consider numerous factors such as link speed, delay, load, reliability, and *cost* (an arbitrarily assigned weight or metric). Convergence occurs very rapidly (within seconds) with link-state protocols.

Open shortest path first (OSPF) is an example of a link-state routing protocol that is often used in large enterprise networks. OSPF routes network traffic within a single *autonomous system* (AS). OSPF networks are divided into areas identified by 32-bit area identifiers. Area identifiers can (but don't have to) correspond to network IP addresses and can duplicate IP addresses without conflicts.

Key Terms

An *autonomous system (AS)* is a group of contiguous IP address ranges under the control of a single Internet entity. Individual autonomous systems are assigned a 16-bit or 32-bit AS Number (ASN) that uniquely identifies the network on the Internet. ASNs are assigned by the Internet Assigned Numbers Authority (IANA).

- **Path-vector.** A path-vector protocol is similar to a distance-vector protocol, but without the scalability issues associated with limited hop counts in distance-vector protocols. Each routing table entry in a path-vector protocol contains path information that gets dynamically updated.
- *border gateway protocol* (BGP) is an example of a path-vector protocol used between separate autonomous systems. BGP is the core protocol used by Internet service providers (ISPs), network service providers (NSPs), and on very large private IP networks.

2.1.4 Area networks and topologies

Most computer networks are broadly classified as either local area networks (LANs) or wide area networks (WANs).

A *local area network* (LAN) is a computer network that connects end-user devices such as laptop and desktop computers, servers, printers, and other devices so that applications, databases, files, file storage, and other networked resources can be shared among authorized users on the LAN. A LAN operates across a relatively small geographic area, such as a floor, a building, or a group of buildings, typically at speeds of up to 10 megabits per second (Mbps - Ethernet), 100 Mbps (Fast Ethernet), 1000 Mbps (or 1 gigabit per second [1 Gbps] – Gigabit

Ethernet) on wired networks and 11 Mbps (802.11b), 54 Mbps (802.11a and g), 450 Mbps (802.11n), 1.3 Gbps (802.11ac), and 14 Gbps (802.11ax – theoretical) on wireless networks. A LAN can be wired, wireless, or a combination of wired and wireless. Examples of networking equipment commonly used in LANs include *bridges*, hubs, *repeaters*, switches, and wireless access points (APs).

Key Terms

A *local area network* (LAN) is a computer network that connects laptop and desktop computers, servers, printers, and other devices so that applications, databases, files and file storage, and other networked resources can be shared across a relatively small geographic area, such as a floor, a building, or a group of buildings.

A *bridge* is a wired or wireless network device that extends a network or joins separate network segments.

A *repeater* is a network device that boosts or re-transmits a signal to physically extend the range of a wired or wireless network.

Two basic network topologies (and many variations) are commonly used in LANs:

- **Star.** Each node on the network is directly connected to a switch, hub, or concentrator, and all data communications must pass through the switch, hub, or concentrator. The switch, hub, or concentrator can thus become a performance bottleneck or single point of failure in the network. A star topology is ideal for practically any size environment and is the most commonly used basic LAN topology. A star topology is also easy to install and maintain, and network faults are easily isolated without affecting the rest of the network.
- **Mesh.** All nodes are interconnected to provide multiple paths to all other resources. A mesh topology may be used throughout the network, or only for the most critical network components, such as routers, switches, and servers to eliminate performance bottlenecks and single points of failure.

Other once-popular network topologies, such as *ring* and *bus*, are rarely found in modern networks.

Key Terms

In a *ring topology*, all nodes are connected in a closed loop that forms a continuous ring. In a ring topology, all communication travels in a single direction around the ring. Ring topologies were common in token ring networks.

In a *bus (or linear bus) topology*, all nodes are connected to a single cable (the backbone) that is terminated on both ends. In the past, bus networks were commonly used for very small networks because they were inexpensive and relatively easy to install, but today bus topologies are rarely used. The cable media has physical limitations (the cable length), the backbone is a single point of failure (a break anywhere on the network affects the entire network), and tracing a fault in a large network can be extremely difficult.

A *wide area network* (WAN) is a computer network that connects multiple LANs or other WANs across a relatively large geographic area, such as a small city, a region or country, a global enterprise network, or the entire planet (for example, the Internet). A WAN connects networks using telecommunications circuits and technologies such as *broadband cable*, *digital subscriber line* (DSL), *fiber optic, optical carrier* (for example, OC-3), and *T-carrier* (for example, T-1), at various speeds typically ranging from 256 Kbps to several hundred Mbps. Examples of networking equipment commonly used in WANs include access servers, Channel Service Units/Data Service Units (CSUs/DSUs), firewalls, modems, routers, virtual private network (VPN) gateways, and WAN switches.

The hierarchical internetworking model is a best practice network design, originally proposed by Cisco, that is comprised of three layers:

- **Access.** User endpoints and servers connect to the network at this layer, typically via network switches. Switches at this layer may perform some Layer 3 (discussed in Section 2.3.1) functions and may also provide electrical power via *power over Ethernet* (PoE) ports to other equipment connected to the network, such as wireless APs or *Voice over IP* (VoIP) phones.
- **Distribution.** This layer performs for any compute-intensive routing and switching functions on the network such as complex routing, filtering, and *quality of service* (QoS). Switches at this layer may be Layer 7 (discussed in Section 2.3.1) switches and connect to lower-end Access layer switches and higher-end Core layer switches.
- **Core.** This layer is responsible for high-speed routing and switching. Routers and switches at this layer are designed for high-speed packet routing and forwarding.

Key Terms

A *wide area network* (WAN) is a computer network that connects multiple LANs or other WANs across a relatively large geographic area, such as a small city, a region or country, a global enterprise network, or the entire planet (for example, the Internet).

Broadband cable is a type of high-speed Internet access that delivers different upload and download data speeds over a shared network medium. The overall speed varies depending upon the network traffic load from all the subscribers on the network segment.

Digital subscriber line (DSL) is a type of high-speed Internet access that delivers different upload and download data speeds. The overall speed depends upon the distance from the home or business location to the provider's central office (CO).

Fiber optic technology converts electrical data signals to light and delivers constant data speeds in the upload and download directions over a dedicated fiber optic cable medium. Fiber optic technology is much faster and more secure than other types of network technology.

Optical carrier is a standard specification for the transmission bandwidth of digital signals on Synchronous Optical Networking (SONET) fiber optic networks. Optical carrier transmission rates are designated by the integer value of the multiple of the base rate (51.84 Mbps). For example, OC-3 designates a 155.52 Mbps (3 x 51.84) network and OC-192 designates a 9953.28 Mbps (192 x 51.84) network.

T-carrier is a full-duplex digital transmission system that uses multiple pairs of copper wire to transmit electrical signals over a network. For example, a T-1 circuit consists of two pairs of copper wire – one pair transmits, the other pair receives – that are multiplexed to provide a total of 24 channels, each delivering 64 Kbps of data, for a total bandwidth of 1.544 Mbps.

Power over Ethernet (PoE) is a network standard that provides electrical power to certain network devices over Ethernet cables.

Voice over IP (VoIP) or *IP telephony* is technology that provides voice communication over an internet protocol (IP) based network.

Quality of service (QoS) is the overall performance of specific applications or services on a network including error rate, bit rate, throughput, transmission delay, availability, and jitter. QoS policies can be configured on certain network and security devices to prioritize certain traffic, such as voice or video, over other, less performance-intensive traffic.

In addition to LANs and WANs, many other types of area networks are used for different purposes:

- Campus area networks (CANs) and wireless campus area networks (WCANs) connect multiple buildings in a high-speed network (for example, across a corporate or university campus).
- Metropolitan area networks (MANs) and wireless metropolitan area networks (WMANs) extend networks across a relatively large area, such as a city.
- Personal area networks (PANs) and wireless personal area networks (WPANs) connect an individual's electronic devices – such as laptop computers, smartphones, tablets, virtual personal assistants (for example, Amazon Alexa, Apple Siri, Google Assistant, and Microsoft Cortana), and wearable technology – to each other or to a larger network.
- Storage area networks (SANs) connect servers to a separate physical storage device (typically a disk array).
- Value-added networks (VANs) are a type of extranet that allows businesses within an industry to share information or integrate shared business processes.
- Virtual local area networks (VLANs) segment broadcast domains in a LAN, typically into logical groups (such as business departments). VLANs are created on network switches.
- Wireless local area networks (WLANs), also known as Wi-Fi networks, use wireless access points (APs) to connect wireless-enabled devices to a wired LAN.
- Wireless wide area networks (WWANs) extend wireless network coverage over a large area, such as a region or country, typically using mobile cellular technology.

2.1.5 Domain name system (DNS)

The *domain name system* (DNS) is a distributed, hierarchical Internet database that maps *fully qualified domain names* (FQDNs) for computers, services, and other resources – such as a website address (also known as a uniform resource locator, or URL) – to IP addresses (discussed in Sections 2.2 and 2.2.1), similar to how a contact list on a smartphone maps the names of businesses and individuals to phone numbers. To create a new domain name that will be accessible via the Internet, you must register your unique domain name with a *domain name registrar*, such as GoDaddy or Network Solutions, similar to listing a new phone number in a phone directory. DNS is critical to the operation of the Internet.

Key Terms

The *domain name system* (DNS) is a hierarchical distributed database that maps the fully qualified domain name (FQDN) for computers, services, or any resource connected to the Internet or a private network to an IP address.

A *fully qualified domain name* (FQDN) is the complete domain name for a specific computer, service, or resource connected to the Internet or a private network.

A *domain name registrar* is an organization that is accredited by a *top-level domain* (TLD) registry to manage domain name registrations.

A *root name server* is the *authoritative* name server for a DNS root zone. Worldwide, 13 root name servers (actually 13 networks comprised of hundreds of root name servers) are configured, named a.root-servers.net through m.root-servers.net. DNS servers are typically configured with a *root hints file* that contains the names and IP addresses of the root servers.

When a host (such as a web browser on a desktop computer) on a network needs to connect to another host (such as a web server on the Internet), it must first translate the name of the destination host from its URL to an IP address. The connecting host (the DNS client) sends a DNS request to the IP address of the DNS server that is specified in the network configuration of the DNS client. If the DNS server is authoritative for the destination domain, the DNS server resolves the IP address of the destination host and answers the DNS request from the DNS client. For example, you are attempting to connect to an *intranet* server on your internal network from the desktop computer in your office. If the DNS server address that is configured on your computer is an internal DNS server that is authoritative for your intranet domain, the DNS server resolves the IP address of the intranet server. Your computer then encapsulates the resolved destination IP address in the *hypertext transfer protocol* (HTTP) or *hypertext transfer protocol secure* (HTTPS) request packets that are sent to the intranet server.

If a DNS server is not authoritative for the destination domain, for example, an Internet website address, then the DNS server performs a *recursive* query (if it is configured to perform recursive queries) to obtain the IP address of the authoritative DNS server and sends the original DNS request to the authoritative DNS server. This process is a top-down process in which the DNS server first consults its root hints file and queries a root name server to identify the authoritative DNS server for the TLD (for example, .com) associated with the DNS query. The DNS server then queries the TLD server to identify the authoritative server for the specific domain that is being queried (for example, paloaltonetworks.com). This process continues until the authoritative server for the FQDN is identified and queried. The recursive DNS server then

answers the original DNS client's request with the DNS information from the authoritative DNS server.

Key Terms

A *top-level domain* (TLD) is the highest level domain in DNS, represented by the last part of a FQDN (for example, .com or .edu). The most commonly used TLDs are generic top-level domains (gTLD) such as .com, edu, .net, and .org, and country-code top-level domains (ccTLD) such as .ca and .us.

An *authoritative* DNS server is the system of record for a given domain.

An *intranet* is a private network that provides information and resources – such as a company directory, human resources policies and forms, department or team files, and other internal information – to an organization's users. Like the Internet, an intranet uses the HTTP and/or HTTPS protocols, but access to an intranet is typically restricted to an organization's internal users. Microsoft SharePoint is a popular example of intranet software.

Hypertext transfer protocol (HTTP) is an application protocol used to transfer data between web servers and web browsers.

Secure hypertext transfer protocol (HTTPS) is a secure version of HTTP that uses secure sockets layer (SSL) or transport layer security (TLS) encryption.

A *recursive* DNS query is performed (if the DNS server allows recursive queries) when a DNS server is not authoritative for a destination domain. The non-authoritative DNS server obtains the IP address of the authoritative DNS server for the destination domain and sends the original DNS request to that server to be resolved.

The basic DNS record types are:

- **A (IPv4) or AAAA (IPv6)** (Address). Maps a domain or subdomain to an IP address or multiple IP addresses
- **CNAME** (Canonical Name). Maps a domain or subdomain to another hostname
- **MX** (Mail Exchanger). Specifies the hostname or hostnames of email servers for a domain
- **PTR** (Pointer). Points to a CNAME; commonly used for reverse DNS lookups that map an IP address to a host in a domain or subdomain

- **SOA** (Start of Authority). Specifies authoritative information about a DNS zone such as primary name server, email address of the domain administrator, and domain serial number
- **NS** (Name Server)**.** The NS record specifies an authoritative name server for a given host
- **TXT** (Text). Stores text-based information

2.1 Knowledge Check

Test your understanding of the fundamentals in the preceding section. Review the correct answers in the Appendix at the end of this guide.

1. **Fill in the Blank.** A _____ sends data packets to destination networks along a network path using logical addresses.
2. **Multiple Choice.** Which option is an example of a static routing protocol? (Choose one.)
    a) open shortest path first (OSPF)
    b) border gateway protocol (BGP)
    c) routing information protocol (RIP)
    d) split horizon
3. **Multiple Choice.** Which three options are dynamic routing protocols? (Choose three.)
    a) distance-vector
    b) path-vector
    c) link-state
    d) point-to-point
4. **True or False.** The Internet is an example of a wide area network (WAN).
5. **Fill in the Blank.** The _____ is a distributed, hierarchical Internet database that maps FQDNs to IP addresses.

2.2 Physical, Logical, and Virtual Addressing

Physical, logical, and virtual addressing in computer networks requires a basic understanding of decimal (base10), binary (base2), and hexadecimal (base16) numbering (see Table 2-1).

The decimal (base10) numbering system is, of course, what we all learn in school. It is comprised of the numerals 0 through 9. After the number 9, we add a digit ("1") in the "tens" position and begin again at zero in the "ones" position, thereby creating the number 10. We increment the digits in the ones position until we reach the number 19, then increment the tens position to the number "2" to create the number 20. This process of incrementing different numeral positions after the ninth digit continues through the hundreds, thousands, millions, and so on. Humans use the decimal numbering system, quite literally, because we have ten fingers, so a base10 numbering systems is easiest for humans to understand.

Table 2-1 Decimal, Hexadecimal, and Binary Notation

Decimal	Hexadecimal	Binary
0	0	0000
1	1	0001
2	2	0010
3	3	0011
4	4	0100
5	5	0101
6	6	0110
7	7	0111
8	8	1000
9	9	1001
10	A	1010
11	B	1011
12	C	1100
13	D	1101
14	E	1110
15	F	1111

A binary (base2) numbering system is comprised of only two digits – 1 ("on") and 0 ("off"). Binary numbering is used in computers and networking because they use electrical transistors (rather than fingers, like humans) to count. The basic function of a transistor is a gate – when

electrical current is present, the gate is closed (“1” or “on”). When no electrical current is present, the gate is open (“0” or “off”). With only two digits, a binary numbering system increments to the next position more frequently than a decimal numbering system. For example, the decimal number one is represented in binary as “1”, number two is represented as “10”, number three is represented as “11”, and number four is represented as “100”.

A hexadecimal (base16) numbering system is comprised of 16 digits (0 through 9, and A through F). Hexadecimal numbering is used because it is more convenient to represent a byte (which consists of 8 bits) of data as two digits in hexadecimal, rather than eight digits in binary. The decimal numbers 0 through 9 are represented as in hexadecimal “0” through “9”, respectively. However, the decimal number 10 is represented in hexadecimal as “A”, the number 11 is represented as “B”, the number 12 is represented as “C”, the number 13 is represented as “D”, the number 14 is represented as “E”, and the number 15 is represented as “F”. The number 16 then increments to the next numeric position, represented as “10”.

The physical address of a network device, known as a *media access control* (MAC) address (also referred to as a burned-in address [BIA] or hardware address), is used to forward traffic on a local network segment. The MAC address is a unique 48-bit identifier assigned to the network interface controller (NIC) of a device. If a device has multiple NICs, each NIC must have a unique MAC address. The MAC address is usually assigned by the device manufacturer and is stored in the device read-only memory (ROM) or firmware. MAC addresses are typically expressed in hexadecimal format with a colon or hyphen separating each 8-bit section. An example of a 48-bit MAC address is:

00:40:96:9d:68:16

The logical address of a network device, such as an IP address, is used to route traffic from one network to another. An IP address is a unique 32- or 128-bit (IPv4 and IPv6, respectively) address assigned to the NIC of a device. If a device has multiple NICs, each NIC may be assigned a unique IP address or multiple NICs may be assigned a virtual IP address to enable bandwidth aggregation or failover capabilities. IP addresses are statically or dynamically (most commonly using the *dynamic host configuration protocol*, or DHCP) assigned, typically by a network administrator or network service provider. IPv4 addresses are usually expressed in dotted decimal notation with a dot separating each decimal section (known as an *octet*). An example of an IPv4 address is:

192.168.0.1

IPv6 addresses are typically expressed in hexadecimal format (32 hexadecimal numbers grouped into eight blocks) with a colon separating each block of four hexadecimal digits (known as a *hextet*). An example of an IPv6 address is:

2001:0db8:0000:0000:0008:0800:200c:417a

IPv4 and IPv6 addressing is explained further in Section 2.2.1.

The *address resolution protocol* (ARP) translates a logical address, such as an IP address, to a physical MAC address. The *reverse address resolution protocol* (RARP) translates a physical MAC address to a logical address.

The dynamic host configuration protocol (DHCP) is a network management protocol used to dynamically assign IP addresses to devices that do not have a statically assigned (manually configured) IP address on a TCP/IP network. BOOTP is a similar network management protocol that is commonly used on UNIX and Linux TCP/IP networks. When a network-connected device that does not have a statically assigned IP address is powered on, the DHCP client software on the device broadcasts a DHCPDISCOVER message on UDP port 67. When a DHCP server on the same subnet (or a different subnet if a DHCP Helper or DHCP Relay Agent is configured) as the client receives the DHCPDISCOVER message, it reserves an IP address for the client and sends a DHCPOFFER message to the client on UDP port 68. The DHCPOFFER message contains the MAC address of the client, the IP address that is being offered, the subnet mask, the lease duration, and the IP address of the DHCP server that made the offer. When the client receives the DHCPOFFER it broadcasts a DHCPREQUEST message on UDP port 67, requesting the IP address that was offered. A client may receive DHCPOFFER messages from multiple DHCP servers on a subnet, but can only accept one offer. When the DHCPREQUEST message is broadcast, the other DHCP servers that sent an offer which was not requested (in effect, accepted) in the DHCPREQUEST message will withdraw their offers. Finally, when the correct DHCP server receives the DHCPREQUEST message, it sends a DHCPACK (Acknowledgement) message on UDP port 68 and the IP configuration process is completed (see Figure 2-1).

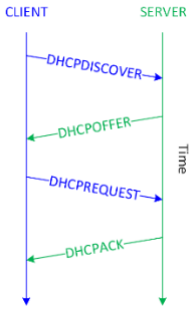

Figure 2-1: DHCP operation.

Network address translation (NAT) virtualizes IP addresses by mapping private, non-routable IP addresses (discussed in Section 2.2.1) that are assigned to internal network devices to public IP addresses when communication across the Internet is required. NAT is commonly implemented on firewalls and routers to conserve public IP addresses.

Key Terms

A *media access control* (MAC) address is a unique 48 or 64-bit identifier assigned to a network interface controller (NIC) for communications at the Data Link layer of the OSI model (discussed in Section 2.3.1).

The *dynamic host configuration protocol* (DHCP) is a network management protocol that dynamically assigns (leases) IP addresses and other network configuration parameters (such as *default gateway* and *domain name system* [DNS] information) to devices on a network.

A *default gateway* is a network device, such as a router or switch, to which an endpoint sends network traffic when a specific destination IP address is not specified by an application or service, or when the endpoint does not know how to reach a specified destination.

The *domain name system* (DNS) is a decentralized, hierarchical directory service that maps IP addresses to domain names for computers, servers, and other resources on a network and the Internet. DNS is analogous to a phone book for the Internet.

An *octet* is a group of 8 bits in a 32-bit IPv4 address.

A *hextet* is a group of four 4-bit hexadecimal digits in a 128-bit IPv6 address.

The *address resolution protocol* (ARP) translates a logical address, such as an IP address, to a physical MAC address. The *reverse address resolution protocol* (RARP) translates a physical MAC address to a logical address.

Network address translation (NAT) virtualizes IP addresses by mapping private, non-routable IP addresses assigned to internal network devices to public IP addresses.

2.2.1 IP addressing basics

Data packets are routed over a transmission control protocol/internet protocol (TCP/IP) network using IP addressing information. IPv4, which is the most widely deployed version of IP, consists of a 32-bit logical IP address. The first four bits in an octet are known as the *high-order* bits (the first bit in the octet is referred to as the *most significant* bit); the last four bits in an

octet are known as the *low-order* bits (the last bit in the octet is referred to as the *least significant* bit).

Each bit position represents its value (see Table 2-2) if the bit is "on" (1); otherwise, the bit's value is zero ("off" or 0).

Key Terms

The first four bits in a 32-bit IPv4 address octet are referred to as the *high-order* bits.

The last four bits in a 32-bit IPv4 address octet are referred to as the *low-order* bits.

The first bit in a 32-bit IPv4 address octet is referred to as the *most significant* bit.

The last bit in a 32-bit IPv4 address octet is referred to as the *least significant* bit.

Table 2-2: Bit Position Values in an IPv4 Address

High-order bits				Low-order bits			
128	64	32	16	8	4	2	1

Each octet contains an 8-bit number with a value of 0 to 255. Table 2-3 shows a partial list of octet values in binary notation.

Table 2-3: Binary Notation of Octet Values

Decimal	Binary	Decimal	Binary	Decimal	Binary	Decimal	Binary
255	1111 1111	200	1100 1000	128	1000 0000	8	0000 1000
254	1111 1110	192	1100 0000	120	0111 1000	7	0000 0111
253	1111 1101	180	1011 0100	110	0110 1110	6	0000 0110
252	1111 1100	172	1010 1100	100	0110 0100	5	0000 0101
251	1111 1011	170	1010 1010	96	0110 0000	4	0000 0100
250	1111 1010	160	1010 0000	90	0101 1010	3	0000 0011
249	1111 1001	150	1001 0110	64	0100 0000	2	0000 0010
248	1111 1000	140	1000 1100	32	0010 0000	1	0000 0001

224	1110 0000	130	1000 0010	16	0001 0000	0	0000 0000

The five different IPv4 address classes (indicated by the high-order bits) are shown in Table 2-4.

Table 2-4: IP Address Classes

Class	Purpose	High-Order Bits	Address Range	Max # of Hosts
A	Large networks	0	1 to 126	16,777,214
B	Medium networks	10	128 to 191	65,534
C	Small networks	110	192 to 223	254
D	Multicast	1110	224 to 239	N/A
E	Experimental	1111	240 to 254	N/A

The address range 127.0.0.1 to 127.255.255.255 is a loopback network used for testing and troubleshooting. Packets sent to a loopback (or localhost) address – such as 127.0.0.1 – are immediately routed back to the source device.

A *subnet mask* is a number that hides the network portion of an IPv4 address, leaving only the host portion of the IP address. The network portion of a subnet mask is represented by contiguous "on" (1) bits beginning with the most significant bit. For example, in the subnet mask 255.255.255.0, the first three octets represent the network portion and the last octet represents the host portion of an IP address. Recall that the decimal number 255 is represented in binary notation as 1111 1111 (refer back to Table 2-2).

Key Terms

A *subnet mask* is a number that hides the network portion of an IPv4 address, leaving only the host portion of the IP address.

The default (or standard) subnet masks for Class A, B, and C networks are:

- **Class A:** 255.0.0.0
- **Class B:** 255.255.0.0
- **Class C:** 255.255.255.0

Several IPv4 address ranges are reserved for use in private networks and are not routable on the Internet, including:

- 10.0.0.0–10.255.255.255 (Class A)
- 172.16.0.0–172.31.255.255 (Class B)
- 192.168.0.0–192.168.255.255 (Class C)

The 32-bit address space of an IPv4 address limits the total number of unique public IP addresses to approximately 4.3 billion. The widespread use of NAT (discussed in Section 2.2) delayed the inevitable depletion of IPv4 addresses but, as of 2018, the pool of available IPv4 addresses that can be assigned to organizations has officially been depleted (a small pool of IPv4 addresses has been reserved by each regional Internet registry to facilitate the transition to IPv6). IPv6 addresses, which use a 128-bit hexadecimal address space providing approximately 3.4×10^{38} (340 hundred undecillion) unique IP addresses, was created to replace IPv4 when the IPv4 address space was exhausted.

IPv6 addresses consist of 32 hexadecimal numbers grouped into eight hextets of four hexadecimal digits, separated by a colon. A hexadecimal digit is represented by 4 bits (refer back to Table 2-1), so each hextet is 16 bits (four 4-bit hexadecimal digits) and eight 16-bit hextets equals 128 bits.

An IPv6 address is further divided into two 64-bit segments: The first (also referred to as the "top" or "upper") 64 bits represent the network part of the address, and the last (also referred to as the "bottom" or "lower") 64 bits represent the node or interface part of the address. The network part is further subdivided into a 48-bit global network address and a 16-bit subnet. The node or interface part of the address is based on the MAC address (discussed in Section 2.2) of the node or interface.

The basic format for an IPv6 address is:

xxxx:xxxx:xxxx:xxxx:xxxx:xxxx:xxxx:xxxx

where x represents a hexadecimal digit (0–f).

This is an example of an IPv6 address:

2001:0db8:0000:0000:0008:0800:200c:417a

There IETF has defined several rules to simplify an IPv6 address:

- Leading zeroes in an individual hextet can be omitted, but each hextet must have at least one hexadecimal digit, except as noted in the next rule. Applying this rule to the previous example yields this result: 2001:db8:0:0:8:800:200c:417a.
- Two colons (::) can be used to represent one or more groups of 16 bits of zeros, as well as leading or trailing zeroes in an address; the :: can appear only once in an IPv6 address. Applying this rule to the previous example yields this result: 2001:db8::8:800:200c:417a.
- In mixed IPv4 and IPv6 environments, the form x:x:x:x:x;x:d.d.d.d can be used, in which x represents the six high-order 16-bit hextets of the address and d represents the four low-order 8-bit octets (in standard IPv4 notation) of the address. For example, 0db8:0:0:0:0:FFFF:129.144.52.38 is a valid IPv6 address. Applying the previous two rules to this example yields this result: db8::ffff:129.144.52.38.

IPv6 security features are specified in Request For Comments (RFC) 7112 and include techniques to prevent fragmentation exploits in IPv6 headers and implementation of internet protocol security (IPsec, discussed in Section 2.6.4) at the Network layer of the OSI model (discussed in Section 2.3.1).

2.2.2 Introduction to subnetting

Subnetting is a technique used to divide a large network into smaller, multiple subnetworks by segmenting an IP address into two parts: the network and the host. Subnetting can be used to limit network traffic or limit the number of devices that are visible to, or able to connect to, each other. Routers examine IP addresses and subnet values (called masks) and determine whether or not to forward packets between networks. With IP addressing the subnet mask is a required element.

Key Terms

Subnetting is a technique used to divide a large network into smaller, multiple subnetworks.

For a Class C IPv4 address, there are 254 possible node (or host) addresses (2^8 or 256 potential addresses, but you lose two addresses for each network: one for the base network address and the other for the broadcast address). A typical Class C network uses a default 24-bit subnet mask (255.255.255.0). This subnet mask value identifies the network portion of an IPv4 address with the first three octets being all ones (11111111 in binary notation, 255 in decimal notation). The mask displays the last octet as zero (00000000 in binary notation). For a Class C IPv4

address with the default subnet mask, the last octet is where the node-specific values of the IPv4 address are assigned.

For example, in a network with an IPv4 address of 192.168.1.0 and a mask value of 255.255.255.0, the network portion of the address is 192.168.1 and there are 254 node addresses (192.168.1.1 through 192.168.1.254) available. Remember, the first address (192.168.1.0) is the base network and the last address (192.168.1.255) is the broadcast address.

Class A and Class B IPv4 addresses use smaller mask values and support larger numbers of nodes than Class C IPv4 addresses for their default address assignments. Class A networks use a default 8-bit (255.0.0.0) subnet mask, which provides a total of more than 16 million (256 x 256 x 256) available IPv4 node addresses. Class B networks use a default 16-bit (255.255.0.0) subnet mask, which provides a total of 65,534 (256x256, minus the network address and the broadcast address) available IPv4 node addresses.

Unlike subnetting, which divides an IPv4 address along an arbitrary (default) classful 8-bit boundary (8 bits for a Class A network, 16 bits for a Class B network, 24 bits for a Class C network), *classless inter-domain routing* (CIDR) allocates address space on any address bit boundary (known as *variable-length subnet masking*, or VLSM). For example, using CIDR, a Class A network could be assigned a 24-bit mask (255.255.255.0, instead of the default 8-bit 255.0.0.0 mask) to limit the subnet to only 254 addresses, or a 23-bit mask (255.255.254.0) to limit the subnet to 512 addresses.

CIDR is used to reduce the size of routing tables on Internet routers by aggregating multiple contiguous network prefixes (known as *supernetting*) and also helped to slow the depletion of public IPv4 addresses (discussed in Section 2.2.1).

Key Terms

classless inter-domain routing (CIDR) is a method for allocating IP addresses and IP routing that replaces classful IP addressing (for example, Class A, B, and C networks) with classless IP addressing.

Variable-length subnet masking (VLSM) is a technique that enables IP address spaces to be divided into different sizes.

Supernetting aggregates multiple contiguous smaller networks into a larger network to enable more efficient Internet routing.

An IP address can be represented with its subnet mask value, using "netbit" or CIDR notation. A netbit value represents the number of ones in the subnet mask and is displayed after an IP address, separated by a forward slash. For example, 192.168.1.0/24 represents a subnet mask consisting of 24 ones:

11111111.11111111.11111111.00000000 (in binary notation)

or

255.255.255.0 (in decimal notation)

2.2 Knowledge Check

Test your understanding of the fundamentals in the preceding section. Review the correct answers in the Appendix at the end of this guide.

1. **Multiple Choice.** Which option is an example of a logical address? (Choose one.)

    a) IP address
    b) hardware address
    c) MAC address
    d) burned-in address

2. **Fill in the Blank.** An IPv4 address consists of four _____-bit octets.

3. **Fill in the Blank.** _____ is a technique used to divide a large network into smaller, multiple subnetworks by segmenting an IPv4 address into a network and host portion.

2.3 Packet Encapsulation and Lifecycle

In a *circuit-switched* network, a dedicated physical circuit path is established, maintained, and terminated between the sender and receiver across a network for each communications session. Prior to the development of the Internet, most communications networks, such as telephone company networks, were circuit-switched. As discussed in Section 2.1.1, the Internet is a *packet-switched* network comprised of hundreds of millions of routers and billions of servers and user endpoints. In a packet-switched network, devices share bandwidth on communications links to transport packets between a sender and receiver across a network. This type of network is more resilient to error and congestion than *circuit-switched* networks.

Key Terms

In a *circuit-switched network*, a dedicated physical circuit path is established, maintained, and terminated between the sender and receiver across a network for each communications session.

In a *packet-switched network*, devices share bandwidth on communications links to transport packets between a sender and receiver across a network.

When an application needs to send data across the network, for example from a server to a client computer, the application first creates a block of data and sends it to the TCP stack on the server. The TCP stack places the block of data into an output buffer on the server and determines the Maximum Segment Size (MSS) of individual TCP blocks (segments) permitted by the server operating system. The TCP stack then divides the data blocks into appropriately sized segments (for example, 1460 bytes), adds a TCP header, and sends the segment to the IP stack on the server. The IP stack adds source (sender) and destination (receiver) IP addresses to the TCP segment (which is now called an IP packet) and notifies the server operating system that it has an outgoing message that is ready to be sent across the network. When the server operating system is ready, the IP packet is sent to the network interface card (NIC), which converts the IP packet to bits and sends the message across the network.

On its way to the destination computer, the packets typically traverse several network and security devices, such as switches, routers, and firewalls before reaching the destination computer, where the encapsulation process described above is reversed.

2.3.1 The OSI and TCP/IP models

The *Open Systems Interconnection* (OSI) and *Transmission Control Protocol/Internet Protocol* (TCP/IP) models define standard protocols for network communication and interoperability. Using a layered approach, the OSI and TCP/IP models:

- Clarify the general functions of communications processes
- Reduce complex networking processes into simpler sublayers and components
- Promote interoperability through standard interfaces
- Enable vendors to change individual features at a single layer rather than rebuilding the entire protocol stack
- Facilitate logical troubleshooting

Defined by the International Organization for Standardization (ISO – not an acronym, but the adopted organizational name from the Greek language, meaning 'Equal'), the *OSI model* consists of seven layers:

- **Application (Layer 7 or L7).** This layer identifies and establishes availability of communication partners, determines resource availability, and synchronizes communication. Examples of protocols that function at the Application layer include:
    - **file transfer protocol (FTP).** Used to copy files from one system to another on TCP ports 20 (the data port) and 21 (the control port)
    - **hypertext transfer protocol (HTTP).** Used for communication between web servers and web browsers on TCP port 80
    - **hypertext transfer protocol secure (HTTPS).** Used for secure sockets layer/transport layer security (SSL/TLS) encrypted communications between web servers and web browsers on TCP port 443 (and other ports, such as 8443)
    - **internet message access protocol (IMAP).** A store-and-forward electronic mail protocol that allows an email client to access, manage, and synchronize email on a remote mail server on TCP and UDP port 143
    - **post office protocol version 3 (POP3).** An email retrieval protocol that allows an email client to access email on a remote mail server on TCP port 110
    - **simple mail transfer protocol (SMTP).** Used to send and receive email across the Internet on TCP/UDP port 25
    - **simple network management protocol (SNMP).** Used to collect network information by polling stations and sending traps (or alerts) to a management station on TCP/UDP ports 161 (agent) and 162 (manager)
    - **telnet.** Provides terminal emulation for remote access to system resources on TCP/UDP port 23
- **Presentation (Layer 6 or L6).** This layer provides coding and conversion functions (such as data representation, character conversion, data compression, and data encryption) to ensure that data sent from the Application layer of one system is compatible with the Application layer of the receiving system. Examples of protocols that function at the Presentation layer include:

- **American Standard Code for Information Interchange (ASCII).** A character-encoding scheme based on the English alphabet, consisting of 128 characters
- **Extended Binary-Coded Decimal Interchange Code (EBCDIC).** An 8-bit character-encoding scheme largely used on mainframe and mid-range computers
- **Graphics Interchange Format (GIF).** A bitmap image format that allows up to 256 colors and is suitable for images or logos (but not photographs)
- **Joint Photographic Experts Group (JPEG).** A photographic compression method used to store and transmit photographs
- **Motion Picture Experts Group (MPEG).** An audio and video compression method used to store and transmit audio and video files

- **Session (Layer 5 or L5).** This layer provides manages communication sessions (service requests and service responses) between networked systems, including connection establishment, data transfer, and connection release. Examples of protocols that function at the Session layer include:
  - **Network File System (NFS).** Facilitates transparent user access to remote resources on a UNIX-based TCP/IP network
  - **remote procedure call (RPC).** A client-server network redirection protocol
  - **secure shell (SSH).** Establishes an encrypted tunnel between a client and server
  - **session initiation protocol (SIP).** An open signaling protocol standard for establishing, managing and terminating real-time communications — such as voice, video, and text — over large IP-based networks
- **Transport (Layer 4 or L4).** This layer provides transparent, reliable data transport and end-to-end transmission control. Specific Transport layer functions include flow control (managing data transmission between devices by ensuring that the transmitting device doesn't send more data than the receiving device can process), multiplexing (enabling data from multiple applications to be simultaneously transmitted over a single physical link), virtual circuit management (establishing, maintaining, and terminating virtual circuits), and error checking and recovery (detecting transmission errors and taking action to resolve any errors that occur, such as requesting that data be retransmitted). TCP and UDP port numbers assigned to applications and services are defined at the Transport layer. Examples of protocols that function at the Transport layer include:

- **transmission control protocol (TCP).** A connection-oriented (a direct connection between network devices is established before data *segments* are transferred) protocol that provides reliable delivery (received segments are acknowledged and retransmission of missing or corrupted segments is requested) of data. TCP connections are established via a *three-way handshake*. The additional overhead associated with connection establishment, acknowledgement, and error correction means that TCP is generally slower than connectionless protocols, such as user datagram protocol (UDP).

- **user datagram protocol (UDP).** A connectionless (a direct connection between network devices is not established before *datagrams* are transferred) protocol that provides best-effort delivery (received datagrams are not acknowledged and missing or corrupted datagrams are not requested) of data. UDP has no overhead associated with connection establishment, acknowledgement, sequencing, or error-checking and recovery. So UDP is ideal for data that requires fast delivery, as long as that data isn't sensitive to packet loss and doesn't need to be fragmented. Examples of applications that use UDP include domain name system (DNS), simple network management protocol (SNMP), and streaming audio or video.

- **stream control transmission protocol (SCTP).** A message-oriented protocol (similar to UDP) that ensures reliable, in-sequence transport with congestion control (similar to TCP).

- **Network (Layer 3 or L3).** This layer provides routing and related functions that enable data to be transported between systems on the same network or on interconnected networks. Routing protocols (discussed in Section 2.1.3) are defined at this layer. Logical addressing of devices on the network is accomplished at this layer using routed protocols, such as the internet protocol (IP). Routers operate at the Network layer of the OSI model.

- **Data Link (Layer 2).** This layer ensures that messages are delivered to the proper device across a physical network link. This layer also defines the networking protocol (for example, Ethernet) used to send and receive data between individual devices and formats messages from layers above into frames for transmission, handles point-to-point synchronization and error control, and can perform link encryption. Switches typically operate at Layer 2 of the OSI model (although multilayer switches that operate at different layers also exist). The Data Link layer is further divided into two sublayers:

- **Logical Link Control (LLC).** The LLC sublayer provides an interface for the MAC sublayer; manages the control, sequencing, and acknowledgement of frames being passed up to the Network layer or down to the Physical layer; and manages timing and *flow control*.
- **Media Access Control (MAC).** The MAC sublayer is responsible for framing and performs error control using a *cyclic redundancy check* (CRC), identifies MAC addresses (discussed in Section 2.2), and controls media access.

Key Terms

A TCP *segment* is a *protocol data unit* (PDU) defined at the Transport layer of the OSI model.

A *protocol data unit* (PDU) is a self-contained unit of data (consisting of user data or control information and network addressing).

In TCP, a *three-way handshake* is used to establish a connection. For example, a PC initiates a connection with a server by sending a TCP SYN (Synchronize) packet. The server replies with a SYN ACK packet (Synchronize Acknowledgement). Finally, the sends an ACK or SYN-ACK-ACK packet, acknowledging the server's acknowledgement, and data communication commences.

A UDP *datagram* is a PDU defined at the Transport layer of the OSI model.

Flow control monitors the flow of data between devices to ensure that a receiving device, which may not necessarily be operating at the same speed as the transmitting device, doesn't drop packets.

A *cyclic redundancy check* (CRC) is a checksum used to create a message profile. The CRC is recalculated by the receiving device. If the recalculated CRC doesn't match the received CRC, the packet is dropped and a request to resend the packet is transmitted back to the device that sent the packet.

- **Physical (Layer 1 or L1).** This layer sends and receives bits across the network medium (cabling or wireless links) from one device to another. It specifies the electrical, mechanical, and functional requirements of the network, including network topology, cabling and connectors, and interface types, as well as the process for converting bits to electrical (or light) signals that can be transmitted across the physical medium.

The *TCP/IP model* was originally developed by the U.S. Department of Defense (DoD) and actually preceded the OSI model. Whereas the OSI model is a theoretical model used to

logically describe networking processes, the TCP/IP model defines actual networking requirements, for example, for frame construction. The TCP/IP model consists of four layers (see Figure 2-2):

- **Application (Layer 4 or L4).** This layer consists of network applications and processes, and it loosely corresponds to Layers 5 through 7 of the OSI model.
- **Transport (Layer 3 or L3).** This layer provides end-to-end delivery and it corresponds to Layer 4 of the OSI model.
- **Internet (Layer 2 or L2).** This layer defines the IP datagram and routing, and it corresponds to Layer 3 of the OSI model.
- **Network Access (Layer 1 or L1).** Also referred to as the Link layer, this layer contains routines for accessing physical networks and it corresponds to Layers 1 and 2 of the OSI model.

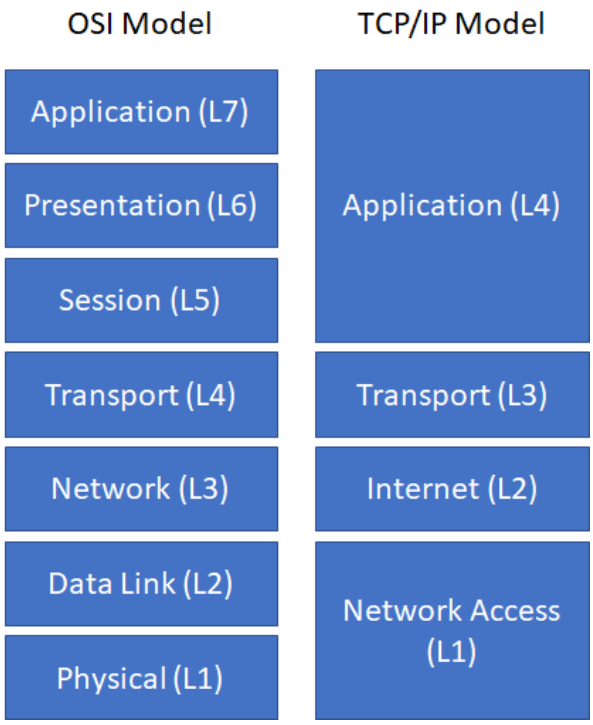

Figure 2-2: The OSI model and the TCP/IP Model.

2.3.2 Data encapsulation

In the OSI and TCP/IP models, data is passed from the highest layer (L7 in the OSI model, L4 in the TCP/IP model) downward through each layer to the lowest layer (L1 in the OSI model and TCP/IP model), and is then transmitted across the network medium to the destination node, where it is passed upward from the lowest layer to the highest layer. Each layer communicates only with the adjacent layer immediately above and below it. This communication is achieved through a process known as *data encapsulation* (or *data hiding*), which wraps protocol information from the layer immediately above in the data section of the layer immediately below.

> **Key Terms**
> *Data encapsulation* (or *data* hiding) wraps protocol information from the (OSI or TCP/IP) layer immediately above in the data section of the layer below.

A protocol data unit (PDU) describes a unit of data at a particular layer of a protocol. For example, in the OSI model, a Layer 1 PDU is known as a bit, a Layer 2 PDU is known as a frame, a Layer 3 PDU is known as a packet, and a Layer 4 PDU is known as a segment or datagram. When a client or server application sends data across a network, a header (and trailer in the case of Layer 2 frames) is added to each data packet from the adjacent layer below it as it passes through the protocol stack. On the receiving end, the headers (and trailers) are removed from each data packet as it passes through the protocol stack to the receiving application.

2.3 Knowledge Check

Test your understanding of the fundamentals in the preceding section. Review the correct answers in the Appendix at the end of this guide.

1. **Multiple Choice.** The OSI model consists of how many layers? (Choose one.)

    a) four
    b) six
    c) seven
    d) nine

2. **Multiple Choice.** Which two protocols function at the Transport layer of the OSI model? (Choose two).

    a) transmission control protocol (TCP)
    b) internet protocol (IP)
    c) user datagram protocol (UDP)
    d) hypertext transfer protocol (HTTP)

3. **Fill in the Blank.** The Data Link layer of the OSI model is further divided into these two sublayers: _____ and_____.

4. **Multiple Choice.** Which four layers comprise the TCP/IP model? (Choose four).

    a) Application
    b) Transport
    c) Physical
    d) Internet
    e) Network Access

5. **Fill in the Blank.** The process that wraps protocol information from the (OSI or TCP/IP) layer immediately above in the data section of the layer immediately below is known as _____.

2.4 Network Security Models

This section describes perimeter-based and Zero Trust network security models.

2.4.1 Perimeter-based network security strategy

Perimeter-based network security models date back to the early mainframe era (circa late 1950s), when large mainframe computers were located in physically secure "machine rooms" that could be accessed by only a relatively limited number of remote job entry (RJE) "dumb" terminals that were directly connected to the mainframe and also located in physically secure areas. Today's data centers are the modern equivalent of machine rooms, but perimeter-based physical security is no longer sufficient for several obvious, but important reasons:

- Mainframe computers predate the Internet. In fact, mainframe computers predate ARPANET, which predates the Internet. Today, an attacker uses the Internet to remotely gain access, rather than physically breaching the data center perimeter.
- Data centers today are remotely accessed by literally millions of remote endpoint devices from anywhere and at any time. Unlike the RJEs of the mainframe era, modern endpoints (including mobile devices) are far more powerful than many of the early mainframe computers and are targets themselves.
- The primary value of the mainframe computer was its processing power. The relatively limited data that was produced was typically stored on near-line media, such as tape. Today, data is the target, it is stored online in data centers and in the cloud, and it is a high value target for any attacker.

The primary issue with a perimeter-based network security strategy in which 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 that may be located in untrusted countries or regions.
- Insiders, whether intentionally malicious or just careless, may present a very real security threat.

Perimeter-based approach strategies fail to account for:

- The potential for sophisticated cyberthreats to penetrate perimeter defenses in which case they would then have free passage on the internal network
- Scenarios where malicious users can 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 that should ideally be separated in any event (for example, research and development and financial systems versus print/file servers)

A broken trust model is not the only issue with perimeter-centric approaches to network security. Another contributing factor is that traditional security devices and technologies (such as port-based firewalls) 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 which devices they're using)
- Filter allowed traffic not only for known application-borne threats but also for unknown ones

The net result is that re-architecting defenses in a way that creates pervasive internal trust boundaries is, by itself, insufficient. You must 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 data.

2.4.2 Zero Trust security

Introduced by Forrester Research, the Zero Trust security model addresses some of the limitations of perimeter-based network security 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.

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 enterprise network. Verification that authorized entities are always doing only what they're allowed to do also is no longer optional in a Zero Trust model; it's now mandatory.

The implications for these two changes are, respectively:

- 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.
- 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 (such as threats).

Benefits of implementing a Zero Trust network include:

- Clearly improved effectiveness in mitigating data loss with visibility and safe enablement of applications, and detection and prevention of cyberthreats
- Greater efficiency for achieving and maintaining compliance with security and privacy mandates, using trust boundaries to segment sensitive applications, systems, and data
- Improved ability to securely enable transformative IT initiatives, such as user mobility, BYOD/BYOA, infrastructure virtualization, and cloud computing
- Lower total cost of ownership (TCO) with a consolidated and fully integrated security operating platform, rather than a disparate array of siloed, purpose-built security point products

2.4.2.1 Core Zero Trust design principles

The core Zero Trust principles that define the operational objectives of a Zero Trust implementation include:

- **Ensure that all resources are accessed securely, regardless of location.** This principle suggests not only the need for multiple trust boundaries but also increased use of secure access for communication to or from resources, even when sessions are confined to the "internal" network. It also means ensuring that the only devices allowed access to

the network have the correct status and settings, have an approved VPN client and proper passcodes, and are not running malware.

Key Terms

The principle of *least privilege* in network security requires that only the permission or access rights necessary to perform an authorized task are granted.

- **Adopt a *least privilege* strategy and strictly enforce access control.** The goal 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.
- **Inspect and log all traffic.** This principle reiterates the need to "always verify" while also reinforcing that adequate protection requires more than just strict enforcement of access control. Close and continuous attention must also be given to exactly what "allowed" applications are actually doing, and the only way to do accomplish these goals is to inspect the content for threats.

2.4.2.2 Zero Trust conceptual architecture

The main components of a Zero Trust conceptual architecture (shown in Figure 2-3) include:

- **Zero Trust Segmentation Platform.** The Zero Trust Segmentation Platform is referred to as a network segmentation gateway by Forrester Research. It is the component used to define internal trust boundaries. That is, it 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 and from resources
    - Continuously monitor allowed sessions for any threat activity

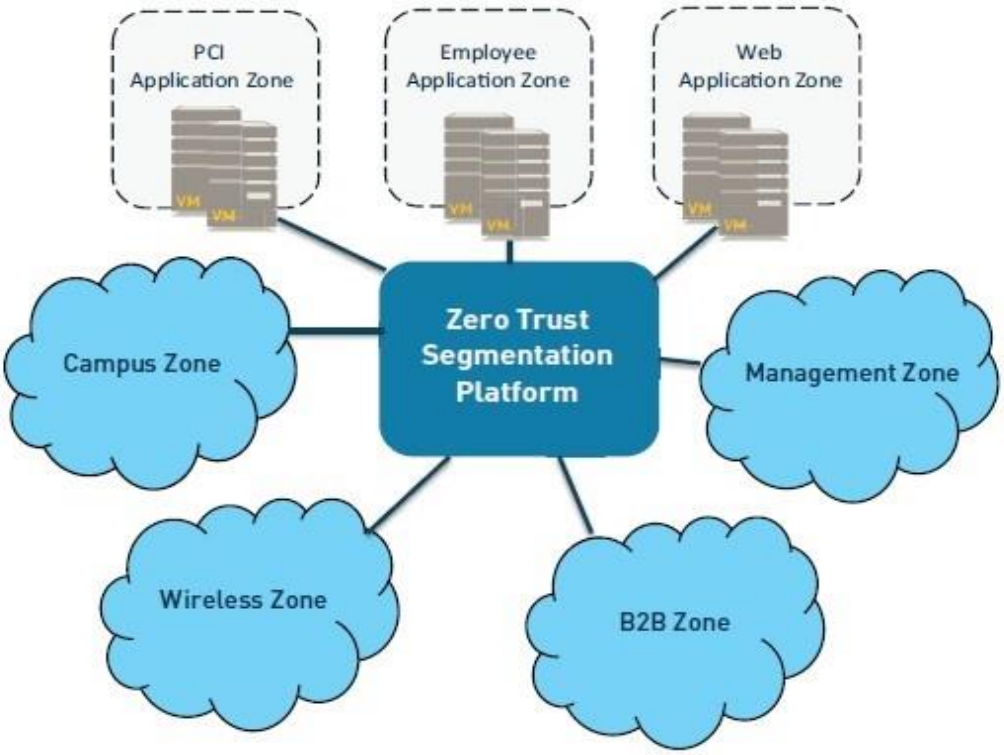

Figure 2-3: Zero Trust conceptual architecture

Although Figure 2-3 depicts the Zero Trust Segmentation Platform as a single component in a single physical location, in practice – because of performance, scalability, and physical limitations – an effective implementation is more likely to entail multiple instances distributed throughout an organization's network. The solution also is designated as a "platform" to reflect that it is an aggregation of multiple distinct (and potentially distributed) security technologies that operate as part of a holistic threat protection framework to reduce the attack surface and correlate information about threats that are found.

- **Trust zones.** Forrester Research refers to a trust zone as a micro core and perimeter (MCAP). 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 of functionality such as protocols and types of transactions is imperative because it 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 2-3 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.

Remember, too, that a trust zone is not intended to be a "pocket of trust" where systems (and therefore threats) within the zone can communicate freely and directly with each other. For a full Zero Trust implementation, the network would be configured to ensure that *all* communications traffic — including traffic 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. A data acquisition network also provides a convenient way to supplement the native monitoring and analysis capabilities for a Zero Trust Segmentation Platform. Session logs that have been forwarded to a data acquisition network 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.

2.4.2.3 Key Zero Trust criteria and capabilities

The core of any Zero Trust network security architecture is the Zero Trust Segmentation Platform, so you must choose the correct solution. Some key criteria and capabilities to consider when selecting a Zero Trust Segmentation Platform, include:

- **Secure access.** Consistent secure IPsec and SSL VPN connectivity is provided for all employees, partners, customers, and guests wherever they're located (for example, at remote or 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, and evasive tactics such as port hopping or encryption. Application identification 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 (such as credit card or Social Security numbers). The result is truly granular access control that safely enables the correct applications for the correct sets of users while automatically preventing unwanted, unauthorized, and potentially harmful traffic from gaining access to the network.

- **Cyberthreat protection.** A combination of anti-malware, intrusion prevention, and cyberthreat prevention technologies provides comprehensive protection against both known and unknown threats, including threats on mobile devices. Support for a closed-loop, highly integrated defense also ensures that inline enforcement devices and other components in the threat protection framework are automatically updated.

- **Coverage for all security domains.** Virtual and hardware appliances establish consistent and cost-effective trust boundaries throughout an organization's entire network, including in remote or branch offices, for mobile users, at the Internet perimeter, in the cloud, at ingress points throughout the data center, and for individual areas wherever they might exist.

2.4.2.4 Implementing a Zero Trust design

Implementation of a Zero Trust network security model doesn't require a major overhaul of an organization's network and security infrastructure. A Zero Trust design architecture can be implemented in a way that requires only incremental modifications to the existing network and is completely transparent to your users. Advantages of such a flexible, non-disruptive deployment approach include minimizing the potential impact on operations and being able to spread the required investment and work effort over time.

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

Once you have a detailed understanding of the network traffic flows in the environment, the next step is to define trust zones and incrementally establish corresponding trust boundaries based on relative risk and/or sensitivity of the data involved:

- Deploy devices in appropriate locations to establish internal trust boundaries for defined trust zones

- Configure the appropriate enforcement and inspection policies to effectively put each trust boundary "online"

Next, you can then progressively establish trust zones and boundaries for other segments of the computing environment based on their relative degree of risk. Examples where secure trust zones can be established are:

- IT management systems and networks (where administrators often hold the proverbial "keys to the kingdom" and a successful breach could lead to compromise of the entire network)
- Partner resources and connections (business-to-business, or B2B)
- High-profile, customer-facing resources and connections (business-to-consumer, or B2C)
- Branch offices in risky countries or regions, followed by all other branch offices
- Guest access networks (both wireless and wired)
- Campus networks

Zero Trust principles and concepts need to be implemented at major access points to the Internet. You will have to replace or augment legacy network security devices with a Zero Trust Segmentation Platform at this deployment stage to gain all of the requisite capabilities and benefits of a Zero Trust security model.

2.4 Knowledge Check

Test your understanding of the fundamentals in the preceding section. Review the correct answers in the Appendix at the end of this guide.

1. **Short Answer.** What is the primary issue with a perimeter-based network security strategy today?
2. **Multiple Choice.** A Zero Trust network security model is based on which security principle? (Choose one.)
    a) due diligence
    b) least privilege
    c) non-repudiation
    d) negative control

2.5 Cloud and Data Center Security

Data centers are rapidly evolving from a traditional, closed environment with static, hardware-based computing resources to one in which traditional and cloud computing technologies are mixed (see Figure 2-4).

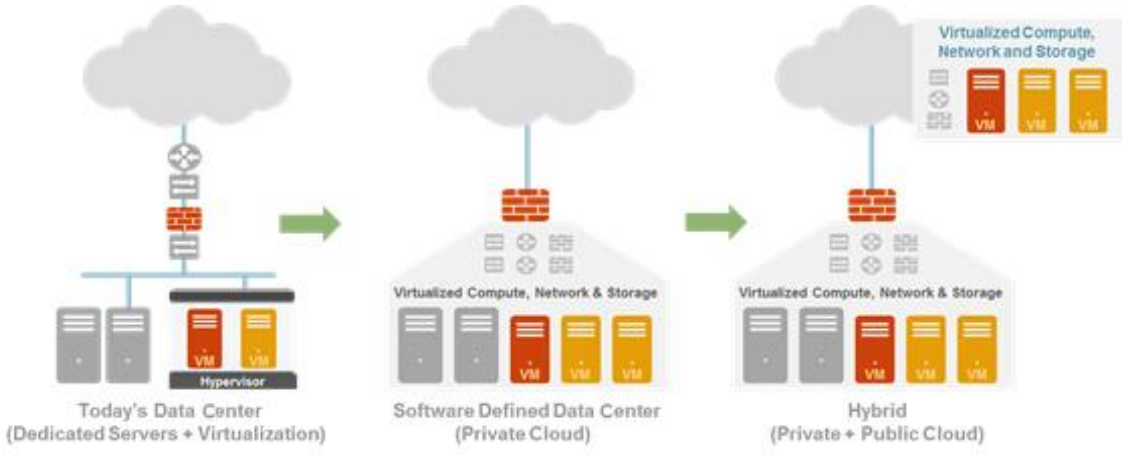

Figure 2-4: Data centers are evolving to include a mix of hardware and cloud computing technologies.

The benefit of moving toward a cloud computing model – private, public, or hybrid – is that it improves operational efficiencies and lowers capital expenditures:

- **Optimizes existing hardware resources.** Instead of using a "one server, one application" model, you can run multiple virtual applications on a single physical server, which means that organizations can leverage their existing hardware infrastructure by running more applications within the same system, provided there are sufficient compute and memory resources on the system.
- **Reduces data center costs.** Reduction of the server hardware "box" count not only reduces the physical infrastructure real estate but also reduces data center costs for power, cooling, and rack space, among others.
- **Increases operational flexibility.** Through the dynamic nature of virtual machine (VM) provisioning, applications can be delivered more quickly than they can through the traditional method of purchasing them, "racking/stacking," cabling, and so on. This operational flexibility 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 use. It also provides a better

way to address server maintenance and backup challenges. For example, IT staff can migrate VMs to other virtualized servers or hypervisors while performing hardware or software upgrades.

2.5.1 Cloud computing depends on virtualization

Cloud computing is not a location, but rather a pool of resources that can be rapidly provisioned in an automated, on-demand manner. 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 value of cloud computing is the ability to pool resources together to achieve economies of scale and agility. This ability to pool resources is true for private or public clouds. Instead of having many independent and often under-used servers deployed for your enterprise applications, pools of resources are aggregated, consolidated, and designed to be elastic enough to scale with the needs of your organization.

The move toward cloud computing not only brings cost and operational benefits but also 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 a cloud computing architecture that, when combined with software orchestration and management tools, allows you to integrate disparate processes so that they can be automated, easily replicated, and offered on an as-needed basis.

2.5.2 Cloud computing security considerations and requirements

With the use of cloud computing technologies, your data center environment can evolve from a fixed environment where applications run on dedicated servers toward an environment that is dynamic and automated, where pools of computing resources are available to support application workloads that can be accessed anywhere, anytime, from any device.

Security remains a significant challenge when you embrace this new dynamic, cloud-computing fabric environment. Many of the principles that make cloud computing attractive are counter to network security best practices:

- **Cloud computing doesn't mitigate existing network security risks.** The security risks that threaten your network today do not change when you move to the cloud. The

shared responsibility model defines who (customer and/or provider) is responsible for what (related to security) in the public cloud. In general terms, the cloud provider is responsible for security ”of” the cloud, including the physical security of the cloud data centers, and foundational networking, storage, compute, and virtualization services. The cloud customer is responsible for security ”in” the cloud, which is further delineated by the cloud service model. For example, in an infrastructure-as-a-service (IaaS) model, the cloud customer is responsible for the security of the operating systems, middleware, runtime, applications, and data. In a platform-as-a-service (PaaS) model, the cloud customer is responsible for the security of the applications and data and the cloud provider is responsible for the security of the operating systems, middleware, and run time. In a SaaS model, the cloud customer is responsible only for the security of the data and the cloud provider is responsible for the full stack from the physical security of the cloud data centers to the application.

- **Security requires isolation and segmentation; the cloud relies on shared resources.** Security best practices dictate that mission-critical applications and data be isolated in secure segments on the network using the Zero Trust (discussed in Section 2.4.2) principle of “never trust, always verify.” On a physical network, Zero Trust is relatively straightforward to accomplish using firewalls and policies based on application and user identity. In a cloud computing environment, direct communication between VMs within a server and in the data center (east-west traffic, discussed in Section 2.5.4) occurs constantly, in some cases across varied levels of trust, making segmentation a difficult task. Mixed levels of trust, when combined with a lack of intra-host traffic visibility by virtualized port-based security offerings, may weaken an organization’s security posture.

- **Security deployments are process-oriented; cloud computing environments are dynamic.** The creation or modification of your cloud workloads can often be done in minutes, yet the security configuration for this workload may take hours, days, or weeks. Security delays are not purposeful; they’re the result of a process that is designed to maintain a strong security posture. Policy changes need to be approved, the appropriate firewalls need to be identified, and the relevant policy updates need to be determined. In contrast, the cloud is a highly dynamic environment, with workloads (and IP addresses) constantly being added, removed, and changed. The result is a disconnect between security policy and cloud workload deployments that leads to a weakened security posture. Security technologies and processes must leverage capabilities such as cloning and scripted deployments to automatically scale and take advantage of the elasticity of the cloud while maintaining a strong security posture.

- **Multi-tenancy is a key characteristic of the public cloud – and a key risk.** Although public cloud providers strive to ensure isolation between their various customers, the infrastructure and resources in the public cloud are shared. Inherent risks in a shared environment include misconfigurations, inadequate or ineffective processes and controls, and the "noisy neighbor" problem (excessive network traffic, disk I/O, or processor utilization can negatively impact other customers sharing the same resource). In hybrid and multi-cloud environments that connect numerous public and/or private clouds, the lines become still more blurred, complexity increases, and security risks become more challenging to address.

As organizations transition from a traditional data center architecture to a public, private, or hybrid cloud environment, enterprise security strategies must be adapted to support changing requirements in the cloud. Key requirements for securing the cloud include:

- **Consistent security in physical and virtualized form factors.** The same levels of application control and threat prevention should be used to protect both your cloud computing environment and your physical network. First, you need to be able to confirm the identity of your applications, validating their identity and forcing them to use only their standard ports. You also need to be able to block the use of rogue applications while simultaneously looking for and blocking misconfigured applications. Finally, application-specific threat prevention policies should be applied to block both known and unknown malware from moving into and across your network and cloud environment.

- **Segment your business applications using Zero Trust principles.** To fully maximize the use of computing resources, a relatively common current practice is to mix application workload trust levels on the same compute resource. Although they are efficient in practice, mixed levels of trust introduce new security risks in the event of a compromise. Your cloud security solution needs to be able to implement security policies based on the concept of Zero Trust (discussed in Section 2.4.2) as a means of controlling traffic between workloads while preventing lateral movement of threats.

- **Centrally manage security deployments; streamline policy updates.** Physical network security is still deployed in almost every organization, so it's critical to have the ability to manage both hardware and virtual form factor deployments from a centralized location using the same management infrastructure and interface. To ensure that security keeps pace with the speed of change that your workflows may exhibit, your security solution should include features that will allow you to lessen, and in some cases eliminate the manual processes that security policy updates often require.

2.5.3 Traditional data security solution weaknesses

Traditional data center security solutions exhibit the same weaknesses found when they are deployed at a perimeter gateway on the physical network – they make their initial positive control network access decisions based on port, using stateful inspection, then they make a series of sequential, negative control decisions using bolted-on feature sets. There are several problems with this approach:

- **"Ports first" limits visibility and control.** The "ports first" focus of traditional data security solutions limits their ability to see all traffic on all ports, which means that evasive or encrypted applications, and any corresponding threats that may or may not use standard ports can evade detection. For example, many data center applications such as Microsoft Lync, Active Directory, and SharePoint use a wide range of contiguous ports to function properly. You must therefore open all those ports first, exposing those same ports to other applications or cyberthreats.

- **They lack any concept of unknown traffic.** Unknown traffic epitomizes the 80/20 rule – it is high risk, but represents only a relatively small amount of traffic on every network. Unknown traffic can be a custom application, an unidentified commercial off-the-shelf application, or a threat. The common practice of blocking it all may cripple your business. Allowing it all is highly risky. You need to be able to systematically manage unknown traffic using native policy management tools to reduce your organizational security risks.

- **Multiple policies, no policy reconciliation tools.** Sequential traffic analysis (stateful inspection, application control, IPS, anti-malware, etc.) in traditional data center security solutions requires a corresponding security policy or profile, often using multiple management tools. The result is that your security policies become convoluted as you build and manage a firewall policy with source, destination, user, port, and action; an application control policy with similar rules; and any other threat prevention rules required. Multiple security policies that mix positive (firewall) and negative (application control, IPS, anti-malware) control models can cause security holes by missing traffic and/or not identifying the traffic. This situation is made worse when there are no policy reconciliation tools.

- **Cumbersome security policy update process.** Existing security solutions in the data center do not address the dynamic nature of your cloud environment because your policies have difficulty contending with the numerous dynamic changes that are common in virtual data centers. In a virtual data center, VM application servers often

move from one physical host to another, so your security policies must adapt to changing network conditions.

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

2.5.4 East-west traffic protection

In a virtualized data center (private cloud), there are two different types of traffic, each of which is secured in a different manner (see Figure 2-5):

- **North-south** refers to data packets that move in and out of the virtualized environment from the host network or a corresponding traditional data center. North-south traffic is secured by one or more physical form factor perimeter edge firewalls. The edge firewall is usually a high-throughput appliance working in high availability active/passive (or active/active) mode to increase resiliency. It controls all the traffic reaching into the data center and authorizes only allowed and "clean" packets to flow into the virtualized environment.
- **East-west** refers to data packets moving between virtual workloads entirely within the private cloud. East-west traffic is protected by a local, virtualized firewall instantiated on each hypervisor. East-west firewalls are inserted transparently into the application infrastructure and do not necessitate a redesign of the logical topology.

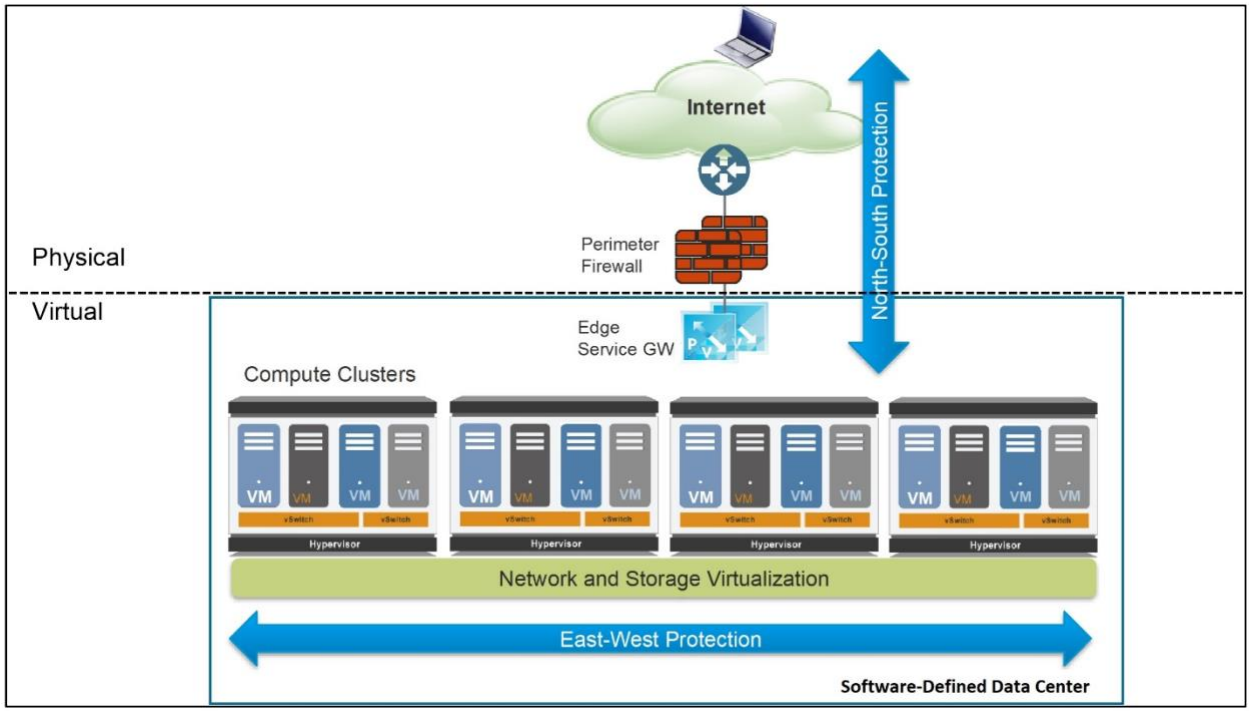

Figure 2-5: Typical virtual data center design architecture

The compute cluster is the building block for hosting the application infrastructure and provides the necessary resources in terms of compute, storage, networking, and security. Compute clusters can be interconnected using OSI Model (discussed in Section 2.3.1) Layer 2 (Data Link) or Layer 3 (Network) technologies such as virtual LAN (VLAN), virtual extensible LAN (VXLAN), or IP, thus providing a domain extension for workload capacity. Innovations in the virtualization space allow VMs to move freely in this private cloud while preserving compute, storage, networking, and security characteristics and postures.

Organizations usually implement security to protect traffic flowing north-south, but this approach is insufficient for protecting east-west traffic within a private cloud. To improve their security posture, enterprises must protect against threats across the entire network, both north-south and east-west.

One common practice in a private cloud is to isolate VMs into different tiers. Isolation provides clear delineation of application functions and allows a security team to easily implement security policies. Isolation is achieved using logical network attributes (such as a VLAN or VXLAN) or logical software constructs (such as security groups). Figure 2-6 displays a simple three-tier application that is composed of a WEB-VM as the front-end, an APP-VM as the application, and a DB-VM providing database services.

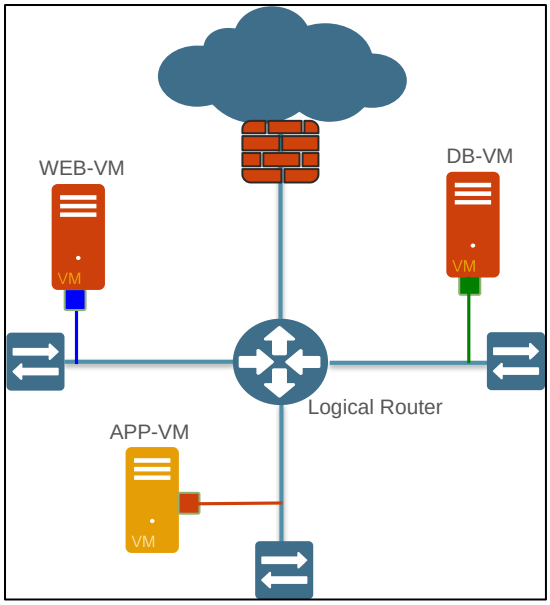

Figure 2-6: Three-tier application hosted in a virtual data center

An attacker has multiple options to steal data from the DB-VM. The first option is to initiate an SQL injection attack by sending HTTP requests containing normalized SQL commands that target an application vulnerability. The second option is to compromise the WEB-VM (using vulnerability exploits), and then move laterally to the APP-VM, initiating a brute-force attack to retrieve the SQL admin password.

Once the DB-VM is compromised, the attacker can hide sensitive data extraction using techniques such as DNS tunneling, or by moving data across the network with NetBIOS, and then off the network via FTP. In fact, attackers using applications commonly found on nearly every network have virtually unlimited options for stealing critical data in this environment. Infiltration into the environment and exfiltration of critical data can be completely transparent and undetected because the data is carried over legitimate protocols (such as HTTP and DNS) that are used for normal business activities.

Virtual data center security best practices dictate a combination of north-south and east-west protection. East-west protection provides the following benefits:

- Authorizes only allowed applications to flow inside the data center, between VMs
- Reduces lateral threat movement when a front-end workload has been compromised (attacker breaches the front-end server using a misconfigured application or unpatched exploit)
- Stops known and unknown threats that are sourced internally within the data center.

- Protects against data theft by leveraging data and file filtering capability and blocking anti-spyware communications to the external world

An added benefit of using virtual firewalls for east-west protection is the unprecedented traffic and threat visibility that the virtualized security device can now provide. Once traffic logs and threat logs are turned on, VM-to-VM communications and malicious attacks become visible. This virtual data center awareness allows security teams to optimize policies and enforce cyberthreat protection (for example, IPS, anti-malware, file blocking, data filtering, and DoS protection) where needed.

2.5.5 Implementing security in virtualized data centers

The following approach to security in the evolving data center — from traditional three-tier architectures to virtualized data centers and to the cloud — aligns with practical realities, such as the need to leverage existing best practices and technology investments, and the likelihood that most organizations will transform their data centers incrementally.

This approach consists of four phases:

- **Consolidating servers within trust levels.** Organizations often consolidate servers within the same trust level into a single virtual computing environment — either one physical host or a cluster of physical hosts. Intra-host communications are generally minimal and inconsequential. As a matter of routine, most traffic is directed "off-box" to users and systems residing at different trust levels. When intra-host communications do happen, the absence of protective safeguards between these virtualized systems is also consistent with the organization's security posture for non-virtualized systems. Live migration features are typically used to enable transfer of VMs only to hosts supporting workloads within the same subnet. Security solutions should incorporate a robust virtual systems capability in which a single instance of the associated countermeasures can be partitioned into multiple logical instances, each with its own policy, management, and event domains. This virtual systems capability enables a single physical device to be used to simultaneously meet the unique requirements of multiple VMs or groups of VMs. Controlling and protecting inter-host traffic with physical network security appliances that are properly positioned and configured is the primary security focus.

- **Consolidating servers across trust levels.** Workloads with different trust levels often coexist on the same physical host or cluster of physical hosts. Intra-host communications are limited, and live migration features are used to enable transfer of VMs only to hosts that are on the same subnet and that are configured identically with regard to routing of VM-to-VM traffic. Intra-host communication paths are intentionally

not configured between VMs with different trust levels. Instead, all traffic is forced "off-box" through a default gateway — such as a physical network security appliance — before it is allowed to proceed to the destination VM. Typically, this off-box routing can be accomplished by configuring separate virtual switches with separate physical network interface cards (NICs) for the VMs at each distinct trust level. As a best practice for virtualization, you should minimize the combination of workloads with different trust levels on the same server. Live migrations of VMs also should be restricted to servers supporting workloads within the same trust levels and within the same subnet. Over time, and in particular, as workloads move to the cloud, maintenance of segmentation based on trust levels becomes more challenging.

- **Selective network security virtualization.** Intra-host communications and live migrations are architected at this phase. All intra-host communication paths are strictly controlled to ensure that traffic between VMs at different trust levels is intermediated either by an on-box, virtual security appliance or by an off-box, physical security appliance. Long-distance live migrations (for example, between data centers) are enabled by combination of native live migration features with external solutions that address associated networking and performance challenges. The intense processing requirements of solutions such as NGFW virtual appliances will ensure that purpose-built physical appliances continue to play a key role in the virtualized data center. However, virtual instances are ideally suited for scenarios where countermeasures need to "migrate" along with the workloads they control and protect.

- **Dynamic computing fabric.** Conventional, static computing environments are transformed into dynamic fabrics (private or hybrid clouds) where underlying resources such as network devices, storage, and servers can be fluidly engaged in whatever combination best meets the needs of the organization at any given point in time. Intra-host communication and live migrations are unrestricted. This phase requires networking and security solutions that are not only capable of being virtualized but are also virtualization-aware and can dynamically adjust as necessary to address communication and protection requirements, respectively. Classification, inspection, and control mechanisms in virtualization-aware security solutions must not be dependent on physical and fixed network-layer attributes. In general, higher-layer attributes such as application, user, and content identification are the basis not only for how countermeasures deliver protection but also for how they dynamically adjust to account for whatever combination of workloads and computing resources exist in their sphere of influence. Associated security management applications also need to be capable of orchestrating the activities of physical and virtual instances of

countermeasures first with each other and subsequently with other infrastructure components. This capability is necessary to ensure that adequate protection is optimally delivered in situations where workloads are frequently migrating across data center hosts.

2.5 Knowledge Check

Test your understanding of the fundamentals in the preceding section. Review the correct answers in the Appendix at the end of this guide.

1. **Short Answer.** List some of the principles of cloud computing that are contrary to network security best practices.

2. **Multiple Choice.** Intra-VM traffic is also known as which type of traffic? (Choose one.)

    a) north-south traffic
    b) unknown traffic
    c) east-west traffic
    d) untrusted traffic

3. **Multiple Choice.** What does the first phase of implementing security in virtualized data centers consist of? (Choose one.)

    a) consolidating servers across trust levels
    b) consolidating servers within trust levels
    c) selectively virtualizing network security functions
    d) implementing a dynamic computing fabric

2.6 Network Security Technologies

This section describes traditional network security technologies including firewalls, intrusion detection systems and Intrusion Prevention Systems (IDS/IPS), web content filters, virtual private networks (VPNs), data loss prevention (DLP), unified threat management (UTM), and security information and event management (SIEM).

2.6.1 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 an untrusted network (such as the Internet)

2.6.1.1 Packet filtering firewalls

First-generation *packet filtering* (also known as *port-based*) firewalls have the following characteristics:

- Operate up to Layer 4 (transport layer) of the OSI model (discussed in Section 2.3.1) and inspects individual packet headers to determine source and destination IP address, protocol (TCP, UDP, ICMP), and port number.
- Match source and destination IP address, protocol, and port number information contained within each packet header to a corresponding rule on the firewall that designates whether the packet should be allowed, blocked, or dropped.
- Inspect and handle each packet individually with no information about context or session.

2.6.1.2 Stateful packet inspection (SPI) firewalls

Second-generation *stateful packet inspection* (also known as *dynamic packet filtering*) firewalls have the following characteristics:

- Operate up to Layer 4 (transport layer) of the OSI model and maintain state information about the different communication sessions that have been established between hosts on the trusted and untrusted networks.
- Inspect individual packet headers to determine source and destination IP address, protocol (TCP, UDP, ICMP), and port number, during session establishment only, to determine if the session should be allowed, blocked, or dropped based on pre-established firewall rules.
- Once a permitted connection is established between two hosts, the firewall creates and deletes firewall rules for individual connections, as needed, thus effectively creating a tunnel that allows traffic to flow between the two hosts without further inspection of individual packets during the session.
- This type of firewall is very fast, but it is port-based and is highly dependent on the trustworthiness of the two hosts because individual packets aren't inspected after the connection is established.

Key Terms

The *Open Systems Interconnection* (OSI) reference model (or "OSI model") defines standard protocols for communication and interoperability using a layered approach in which data is passed from the highest layer (application) downward through each layer to the lowest layer (physical), then transmitted across the network to its destination, then passed upward from the lowest layer to the highest layer. Each layer communicates only with the layer immediately above it and below it using a process called *data encapsulation*, in which protocol information from the layer immediately above is wrapped in the data section of the layer immediately below. The seven layers of the OSI model are:

- **Application** (Layer 7). Identifies and establishes availability of communication partners, determining resource availability and synchronizing communication
- **Presentation** (Layer 6). Provides coding and conversion functions
- **Session** (Layer 5). Establishes, coordinates, and terminates communication sessions
- **Transport** (Layer 4). Provides transparent data transport and end-to-end transmission control.
- **Network** (Layer 3). Provides routing and related functions.
- **Data link** (Layer 2). Ensures that messages are delivered to the proper device across a physical network link.
- **Physical** (Layer 1). Sends and receives bits across the network from one device to another.

2.6.1.3 Application firewalls

Third-generation *application* (also known as *application-layer gateways*, *proxy-based*, and *reverse-proxy*) firewalls have the following characteristics:

- Operate up to Layer 7 (application layer) of the OSI model and control access to specific applications and services on the network.
- Proxy network traffic rather than permitting direct communication between hosts. Requests are sent from the originating host to a 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.

- Inspect application layer traffic and thus can identify and block specified content, malware, exploits, websites, and applications or services using hiding techniques such as encryption and non-standard ports. Proxy servers can also be used to implement strong user authentication and web application filtering, and to mask the internal network from untrusted networks. However, proxy servers have a significant negative impact on the overall performance of the network.

2.6.2 Intrusion detection and prevention systems

Intrusion detection systems (IDS) and Intrusion Prevention Systems (IPS) provide real-time monitoring of network traffic and perform deep-packet inspection and analysis of network activity and data. Unlike traditional packet filtering and stateful packet inspection firewalls that examine only packet header information, IDS/IPS examines both the packet header and payload of network traffic. IDS/IPS attempts to match known-bad, or malicious, patterns (or signatures) found within inspected 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 to be a passive system, whereas IPS is an active system. IDS monitor and analyze network activity and provides alerts to potential attacks and vulnerabilities on the network, but it doesn't perform any preventive action to stop an attack. An IPS, however, performs all of the same functions as an IDS but also automatically blocks or drops suspicious, pattern-matching activity on the network in real-time. However, IPS has some disadvantages, including:

- 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. A false positive occurs when legitimate traffic is improperly identified as malicious traffic. A false negative occurs when malicious traffic is improperly identified as legitimate traffic.

- May be used to deploy a denial-of-service (DoS) attack by flooding the IPS, thus causing it to block connections until no connection or bandwidth is available.

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 signatures 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.

2.6.3 Web content filters

Web content filters are used to restrict the Internet activity of users on a network. Web content filters match a web address (*uniform resource locator*, or URL) against a database of websites, which is typically maintained by the individual security vendors that sell the web content filters and is provided as a subscription-based service. Web content filters attempt to classify websites based on broad categories that are either allowed or blocked for various groups of users on the network. For example, the marketing and human resources 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. Examples of typical website categories include:

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

These sites lower individual productivity but also may be prime targets for malware that users may unwittingly become victims of, via drive-by-downloads. Certain sites may also create liabilities in the form of sexual harassment or racial discrimination suits for organizations that fail to protect other employees from being exposed to pornographic or hate-based websites.

Organizations may elect to implement these solutions in a variety of modes to either block content, warn users before accessing restricted sites, or log 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.

2.6.4 Virtual private networks

A virtual private network (VPN) creates a secure, encrypted connection (or tunnel) across the Internet back to an organization's network. VPN client software is typically installed on mobile endpoints, such as laptop computers and smartphones, to extend a network beyond the physical boundaries of the organization. The VPN client connects to a VPN server, such as a firewall, router, or VPN appliance (or concentrator). Once a VPN tunnel is established, 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.

2.6.4.1 Point-to-point tunneling protocol)

Point-to-point tunneling protocol (PPTP) is a basic VPN protocol that uses transmission control protocol (TCP) port 1723 to establish communication with the VPN peer, and then creates a *generic routing encapsulation* (GRE) tunnel that transports encapsulated *point-to-point protocol* (PPP) packets between the VPN peers. Although PPTP is easy to set up and is considered to be very fast, it is perhaps the least secure of the various VPN protocols. It is commonly used with either the *password authentication protocol* (PAP), *challenge-handshake authentication protocol* (CHAP), or *Microsoft challenge-handshake authentication protocol versions 1 and 2* (MS-CHAP v1/v2), all of which have well-known security vulnerabilities, to authenticate tunneled PPP traffic. The *extensible authentication protocol transport layer security* (EAP-TLS) provides a more secure authentication protocol for PPTP, but requires a *public key infrastructure* (PKI) and is therefore more difficult to set up.

2.6.4.2 Layer 2 tunneling protocol

The Layer 2 tunneling protocol (L2TP) is a tunneling protocol that is supported by most operating systems (including mobile devices). Although it provides no encryption by itself, it is considered secure when used together with IPsec (discussed in Section 2.6.4).

2.6.4.3 Secure socket tunneling protocol

The secure socket tunneling protocol (SSTP) is a VPN tunnel created by Microsoft to transport PPP or L2TP traffic through an SSL 3.0 channel. SSTP is primarily used for secure remote client VPN access, rather than for site-to-site VPN tunnels.

2.6.4.4 Microsoft Point-to-Point Encryption

Microsoft Point-to-Point Encryption (MPPE) encrypts data in PPP-based dial-up connections or PPTP VPN connections. MPPE uses the RSA RC4 encryption algorithm to provide data confidentiality and supports 40-bit and 128-bit session keys.

Key Terms

Generic routing encapsulation (GRE) is a tunneling protocol developed by Cisco Systems that can encapsulate various network layer protocols inside virtual point-to-point links.

Point-to-point protocol (PPP) is a Layer 2 (Data Link) protocol used to establish a direct connection between two nodes.

Password authentication protocol (PAP) is an authentication protocol used by PPP to validate users with an unencrypted password.

Microsoft challenge-handshake authentication protocol (MS-CHAP) is used to authenticate Microsoft Windows-based workstations, using a challenge-response mechanism to authenticate PPTP connections without sending passwords.

Extensible authentication protocol transport layer security (EAP-TLS) is an Internet Engineering Task Force (IETF) open standard that uses the transport layer security (TLS) protocol in Wi-Fi networks and PPP connections.

Public key infrastructure (PKI) is a set of roles, policies, and procedures needed to create, manage, distribute, use, store, and revoke digital certificates and manage public key encryption.

2.6.4.5 OpenVPN

OpenVPN is a highly secure, open source VPN implementation that uses SSL/TLS encryption for key exchange. OpenVPN uses up to 256-bit encryption and can run over TCP or UDP. Although it is not natively supported by most major operating systems, it has been ported to most major operating systems, including mobile device operating systems.

2.6.4.6 Internet protocol security

IPsec is a secure communications protocol that authenticates and encrypts IP packets in a communication session. 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 username and password.

A security association (SA) in IPsec defines how two or more entities will securely communicate over the network using IPsec. A single Internet Key Exchange (IKE) SA is established between communicating entities to initiate the IPsec VPN tunnel. Separate IPsec SAs are then established for each communication direction in a VPN session.

An IPsec VPN can be configured to force all of the user's Internet traffic back through an organization's firewall, thus providing optimal protection with enterprise-grade security, 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 traffic route through the IPsec tunnel, for acceptable protection with much less performance degradation.

If split tunneling is used, a personal firewall should be configured and active on the organization's endpoints, because a split tunneling configuration can create a "side door" into the organization's network. Attackers can essentially bridge themselves over the Internet, through the client endpoint, and into the network over the IPsec tunnel.

2.6.4.7 Secure sockets layer (SSL)

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.

Key Terms

Secure sockets layer (SSL) is a cryptographic protocol for managing authentication and encrypted communication between a client and server to protect the confidentiality and integrity of data exchanged in the session.

Transport layer security (TLS) is the successor to SSL (although it is still commonly referred to as SSL).

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 webpage

using the HTTPS protocol, and log in to the network with their user credentials. An agent-based SSL 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 an endpoint device they do not own or control, such as a hotel kiosk, where full client VPN software cannot be installed.

SSL VPN technology has become the de facto standard and preferred method of connecting remote endpoint devices back to the enterprise network, and IPsec is most commonly used in site-to-site or device-to-device VPN connections such as connecting a branch office network to a headquarters location network or data center.

2.6.5 Data loss prevention

Network *data loss prevention* (DLP) solutions inspect data that is leaving or egressing a network (for example, via email, file transfer, Internet uploads, or by 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:

- Personally Identifiable Information (PII) such as names, addresses, birthdates, Social Security numbers, health records (including *electronic medical records,* or EMRs, and *electronic health records*, or EHRs), and financial data (such as bank account numbers and credit card numbers)
- Classified materials (such as military or national security information)
- Intellectual property, trade secrets, and other confidential or proprietary company information

Key Terms

As defined by HealthIT.gov, an *electronic medical record* (EMR) "contains the standard medical and clinical data gathered in one provider's office."

As defined by HealthIT.gov, an *electronic health record* (EHR) "go[es] beyond the data collected in the provider's office and include[s] a more comprehensive patient history. EHR data can be created, managed, and consulted by authorized providers and staff from across more than one healthcare organization."

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 because 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. DLP solutions often require many moving parts to effectively route traffic to and from inspection engines, which can add to the complexity of troubleshooting network issues.

2.6.6 Unified threat management

Unified threat management (UTM) devices combine numerous security functions into a single appliance, including:

- Anti-malware
- Anti-spam
- Content filtering
- DLP
- Firewall (stateful inspection)
- IDS/IPS
- VPN

UTM devices don't necessarily perform any of these security functions better than their standalone counterparts, but nonetheless serve a purpose in small to medium-size enterprise networks as a convenient and inexpensive solution that gives an organization an all-in-one security device. Typical disadvantages of UTM include:

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

2.6.7 Security information and event management

Security information and event management (SIEM) software tools and managed services provide real-time monitoring, event correlation, analysis, and notification of security alerts generated by various network devices and applications.

2.6 Knowledge Check

Test your understanding of the fundamentals in the preceding section. Review the correct answers in the Appendix at the end of this guide.

1. **True or False.** A dynamic packet filtering firewall inspects each individual packet during a session to determine if the traffic should be allowed, blocked, or dropped by the firewall.

2. **Multiple Choice.** Which are characteristics of application firewalls? (Choose three.)

    a) proxies traffic rather than permitting direct communication between hosts
    b) can be used to implement strong user authentication
    c) masks the internal network from untrusted networks
    d) is extremely fast and has no impact on network performance

3. **Short Answer.** Briefly describe the primary difference between intrusion detection systems and Intrusion Prevention Systems (IDS and IPS, respectively).

4. **Multiple Choice.** Which VPN technology is currently considered the preferred method for securely connecting a remote endpoint device back to an enterprise network? (Choose one.)

    a) Point-to-point tunneling protocol (PPTP)
    b) Secure socket tunneling protocol (SSTP)
    c) secure sockets layer (SSL)
    d) internet protocol security (IPsec)

5. **Multiple Choice.** Which is NOT a characteristic of unified threat management (UTM)? (Choose one.)

    a) combines security functions such as firewalls, intrusion detection systems (IDS), anti-malware, and data loss prevention (IDS) in a single appliance
    b) enabling all of the security functions in a UTM device can have a significant performance impact
    c) UTM fully integrates all of the security functions installed on the device
    d) UTM can be a convenient solution for small networks

2.7 Endpoint security

Traditional endpoint security encompasses numerous security tools, such as anti-malware software, anti-spyware software, personal firewalls, host-based intrusion prevention systems (HIPS), and mobile device management (MDM) software. There are also effective endpoint security best practices, including patch management and configuration management.

2.7.1 Anti-malware

Malware prevention – more specifically, antivirus software – has been one of the first and most basic tenets of information security since the early 1980s. Unfortunately, all of this hard-earned experience doesn't necessarily mean we're winning the war. For example, Trustwave's 2015 *Global Security Report* found that it takes an average of 188 days from infection to detection of malware "in the wild."[31] Interestingly, web-based zero-day attacks, on average, remain "in the wild" up to four times longer than email based threats. This is due to a number of factors, including user awareness of email borne threats, availability and use of email security solutions (such as anti-spam and antivirus), and preferred use of the web as a threat vector by malware developers.

> **Note**
> With the proliferation of advanced malware, such as remote access Trojans (RATs), anti-AV, and rootkits/bootkits (discussed in Section 1.4.1), security vendors have largely rebranded their antivirus solutions as "anti-malware" and expanded their malware protections to encompass the broader malware classifications.

This poor "catch rate" is due to several factors. Some malware has the ability to mutate or can be updated to avoid detection by traditional anti-malware signatures. Additionally, advanced malware is increasingly specialized to the point where an attacker can develop customized malware that is targeted against a specific individual or organization.

Traditional anti-malware software uses various approaches to detect and respond to malware threats, including signature-based, container-based, application whitelisting, and anomaly-based techniques.

---

[31] "2015 Trustwave Global Security Report," Trustwave, 2015, https://www2.trustwave.com/GSR2015.html.

2.7.1.1 Signature-based

Signature-based antivirus (or anti-malware) software is the oldest and most commonly used approach for detecting and identifying malware on endpoints. This approach requires security vendors to continuously collect malware samples, create matching signature files for those samples, and distribute those signature files as updates for their endpoint security products to all of their customers.

Deploying signature-based antivirus software requires installing an engine that typically has kernel-level access to an endpoint's system resources. Signature-based antivirus software scans an endpoint's hard drive and memory, based on a predefined schedule and in real-time when a file is accessed. If a known malware signature is detected, the software performs a predefined action, such as:

- **Quarantine.** Isolates the infected file so that it cannot infect the endpoint or other files.
- **Delete.** Removes the infected file.
- **Alert.** Notifies the user (and/or system administrator) that malware has been detected.

Updated signatures must be regularly and frequently downloaded from the security vendor and installed on the organization's endpoints. Downloading and processing signature files in this manner can cause noticeable performance degradations on the networks and endpoints on which they are running.

Although the signature-based approach is very popular, its effectiveness is limited. By design, it is a reactive countermeasure because a signature file for new malware can't be created and delivered until the malware is already "in the wild," during which time networks and endpoints are blind to the threat – the notorious zero-day threat (or attack). The "zero-day" label is misleading, however, as the number of days from release to detection averages 5 to 20 days (see Figure 2-7).

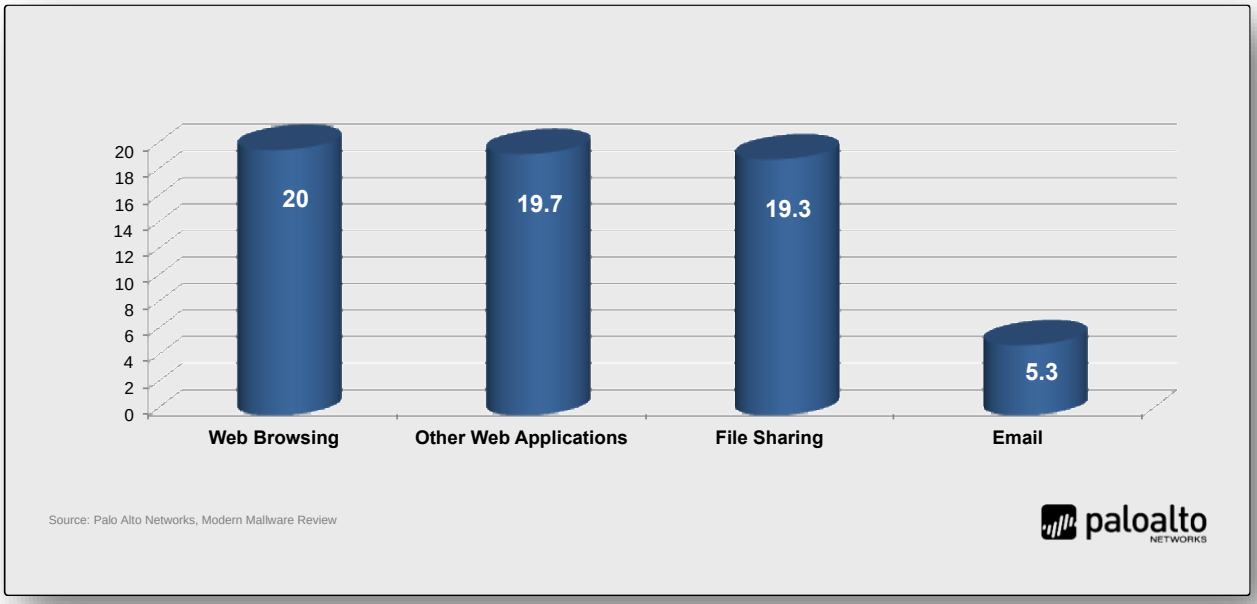

Figure 2-7: Average time to detection by application vector.

A sample of new or unknown suspicious traffic must first be captured and identified before a detection signature can be created by security vendors. The new signature must then be downloaded and installed on an organization's endpoints in order to provide protection.

This means that some users and networks will be successfully breached by new malware until a new detection signature is created, downloaded, and installed. This reactive model creates a window of opportunity for attackers, leaving endpoints vulnerable — sometimes for weeks or even months — until new malware is suspected, collected, analyzed, and identified. During this time, attackers have free rein to infect networks and endpoints.

Another challenge for the signature-based approach is that millions of new malware variations are created each year – on average about 20,000 new forms daily – for which unique signatures must be written, tested, and deployed – after the new malware variation is discovered and sampled. Despite the fact that 70 percent of these millions of malware variations are based on a relatively limited number of malware "families" – numbering just seven in 2005 and

increasing to only 20 over the past decade[32] – this reactive approach is simply not effective for protecting endpoints against modern malware threats.

Additionally, advanced malware uses techniques such as metamorphism and polymorphism to take advantage of the inherent weaknesses of signature-based detection in order to avoid being discovered in the wild and to circumvent signatures that have already been created. These techniques are so commonly used that "70 to 90 percent of malware samples [collected] today are unique to a single organization."[33]

2.7.1.2 Container-based

Container-based endpoint protection wraps a protective virtual barrier around vulnerable processes while they're running. If a process is malicious, the container detects it and shuts it down, preventing it from damaging other legitimate processes or files on the endpoint.

However, the container-based approach typically requires a significant amount of computing resource overhead and attacks have been demonstrated that circumvent or disable container-based protection. This approach also requires knowledge of the applications that need to be protected and how they interact with other software components. So a containerization tool will be developed to support certain common applications, but will not be capable of protecting most proprietary or industry-specific software. Even web browser plug-ins and the like can have problems operating correctly within a container-based environment.

2.7.1.3 Application whitelisting

Application whitelisting is another endpoint protection technique that is commonly used to prevent end users from running unauthorized applications – including malware – on their endpoints.

Application whitelisting requires a positive control model in which no applications are permitted to run on the endpoint unless they're explicitly permitted by the whitelist policy. In practice, this requires a large administrative effort to establish and maintain a list of approved applications. This approach is based on the premise that if you create a list of applications that are specifically allowed and then prevent any other file from executing, you can protect the endpoint. While this basic functionality can be useful to reduce the attack surface, it is by no means a comprehensive approach to endpoint security.

---

[32] Ibid.
[33] Ibid.

Modern trends like cloud and mobile computing, consumerization, and BYOD/BYOA (all discussed in Section 1.1.1) make application whitelisting extremely difficult to enforce in the enterprise. Additionally, once an application is whitelisted it is permitted to run – even if the application has a vulnerability that can be exploited. This means the attacker can simply exploit a whitelisted application and have complete control of the target endpoint regardless of the whitelisting. Once the application has been successfully exploited, the attacker can run malicious code while keeping all of the activity in memory. This means that no new files are created and no new executables attempt to run, rendering the whitelisting software ineffective against this type of attack.

2.7.1.4 Anomaly detection

Endpoint security approaches that use mathematical algorithms to detect unusual activity on an endpoint are known as heuristics-based, behavior-based, or anomaly-detection solutions. This approach relies on first establishing an accurate baseline of what is considered "normal" activity. Although this approach has been around for many years, it requires a very large dataset to reduce the number of false positives.

> **Key Terms**
>
> In anti-malware, a *false positive* incorrectly identifies a legitimate file or application as malware. A *false negative* incorrectly identifies malware as a legitimate file or application. In intrusion detection, a false positive incorrectly identifies legitimate traffic as a threat, while a false negative incorrectly identifies a threat as legitimate traffic.

2.7.2 Anti-spyware

Anti-spyware software is very similar to traditional antivirus 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.

2.7.3 Personal firewalls

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

Personal (or host-based) firewalls are commonly installed and configured on laptop and desktop PCs. Personal firewalls typically operate as Layer 7 (application layer) firewalls that allow or block traffic based on an individual (or group) security policy. Personal firewalls are particularly helpful on laptops used by remote or traveling users that connect their laptop computers directly to the Internet, for example, over a public Wi-Fi connection. 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 advanced malware today.

Windows Firewall is an example of a personal firewall that is installed as part of the Windows desktop or mobile 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.

2.7.4 Host-based intrusion prevention systems (HIPS)

HIPS is another approach to endpoint protection that relies on an agent installed on the endpoint to detect malware. HIPS can be either signature-based or anomaly-based, and is therefore susceptible to the same issues as other signature and anomaly-based endpoint protection approaches.

Additionally, HIPS software often causes significant performance degradation on endpoints. A recent Palo Alto Networks survey found that 25 percent of respondents indicated HIPS solutions “caused significant end user performance impact.”

2.7.5 Mobile device management

Mobile device management (MDM) software provides 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, and prevent *jailbreaking* or *rooting*, for example.
- **Malware protection:** Detect and prevent mobile malware.
- **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.
- **Geo-fencing and location services:** Restrict specific functionality in the device based on its physical location.

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 apps without paying for them, from sources, other than the App Store®, that are not sanctioned and/or controlled by Apple®. Jailbreaking bypasses the security features of the device by replacing the firmware's operating system with a similar, albeit counterfeit version, which makes it vulnerable to malware and exploits. Jailbreaking is known as *rooting* on Google Android™ devices.

2.7 Knowledge Check

Test your understanding of the fundamentals in the preceding section. Review the correct answers in the Appendix at the end of this guide.

1. **True or False.** Signature-based anti-malware software is considered a proactive security countermeasure.
2. **Fill in the Blank.** _____ endpoint protection wraps a protective virtual barrier around vulnerable processes while they're running.
3. **Short Answer.** What is the main disadvantage of application whitelisting related to exploit prevention?
4. **Multiple Choice.** Which capabilities are typical mobile device management software capabilities? (Choose three.)
    a) data loss prevention (DLP)
    b) policy enforcement
    c) intrusion detection
    d) malware prevention

2.8 Cloud, Virtualization, and Storage Security

Cloud computing and virtualization (discussed in Sections 1.1.1, 1.1.3, and 1.1.4) are two important modern computing trends. As data continues to grow exponentially, local and remote storage capacity has become an ever-present challenge and organizations are increasingly relying upon cloud-based storage service offerings. Cloud computing, virtualization, and local and remote storage technologies are discussed in these sections.

2.8.1 Cloud computing

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."

NIST defines three distinct cloud computing service models as follows:

- **Software as a Service (SaaS).** Customers are provided access to an application running on a cloud infrastructure. The application is accessible from various client devices and interfaces, but the customer has no knowledge of, and does not manage or control, the underlying cloud infrastructure. The customer may have access to limited user-specific application settings, and security of the customer data is still the responsibility of the customer.
- **Platform as a Service (PaaS).** Customers can deploy supported applications onto the provider's cloud infrastructure, but the customer has no knowledge of, and does not manage or control, the underlying cloud infrastructure. The customer has control over the deployed applications and limited configuration settings for the application-hosting environment. The company owns the deployed applications and data, and it is therefore responsible for the security of those applications and data.
- **Infrastructure as a Service (IaaS).** Customers can provision processing, storage, networks, and other computing resources and deploy and run operating systems and applications. However, the customer has no knowledge of, and does not manage or control, the underlying cloud infrastructure. The customer has control over operating systems, storage, and deployed applications, as well as some networking components (for example, host firewalls). The company owns the deployed applications and data, and it is therefore responsible for the security of those applications and data.

NIST also defines these four cloud computing deployment models:

- **Public.** A cloud infrastructure that is open to use by the general public. It's owned, managed, and operated by a third party (or parties) and it exists on the cloud provider's premises.
- **Community.** A cloud infrastructure that is used exclusively by a specific group of organizations.
- **Private.** A cloud infrastructure that is used exclusively by a single organization. It may be owned, managed, and operated by the organization or a third party (or a combination of both), and it may exist on or off premises.
- **Hybrid.** A cloud infrastructure that is composed of two or more of the aforementioned deployment models, bound together by standardized or proprietary technology that enables data and application portability (for example, failover to a secondary data center for disaster recovery or content delivery networks across multiple clouds).

The security risks that threaten your network today do not change when you move to the cloud. The *shared responsibility model* defines who (customer and/or provider) is responsible for what (related to security) in the public cloud.

In general terms, the cloud provider is responsible for security "of" the cloud, including the physical security of the cloud data centers, and foundational networking, storage, compute, and virtualization services. The cloud customer is responsible for security "in" the cloud, which is further delineated by the cloud service model (see Figure 2-8).

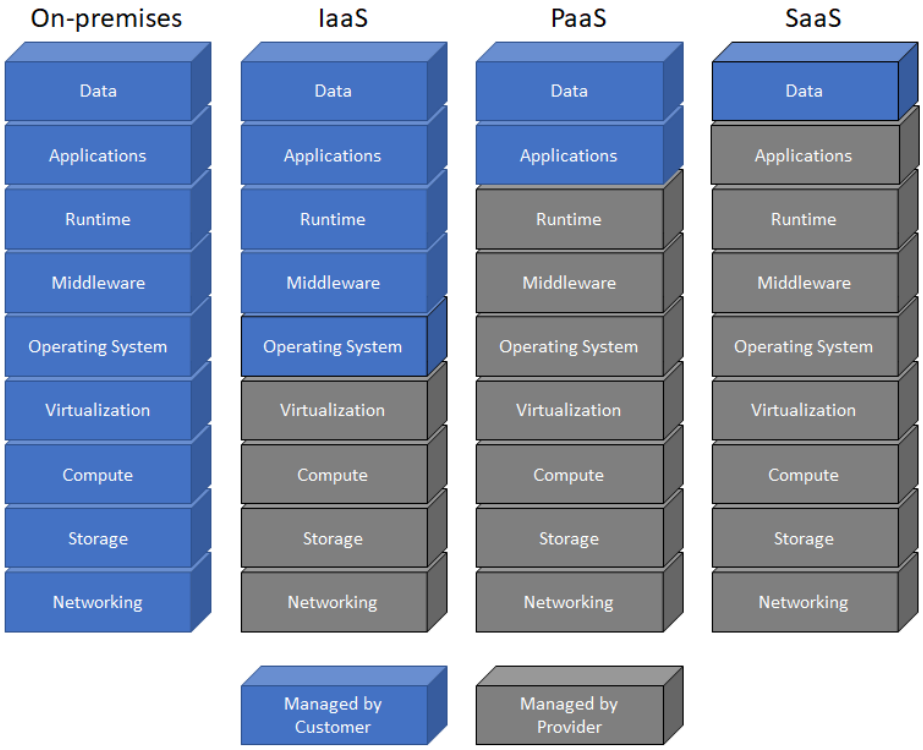

Figure 2-8: The shared responsibility model.

For example, in an infrastructure-as-a-service (IaaS) model, the cloud customer is responsible for the security of the operating systems, middleware, runtime, applications, and data. In a platform-as-a-service (PaaS) model, the cloud customer is responsible for the security of the applications and data and the cloud provider is responsible for the security of the operating systems, middleware, and run time. In a SaaS model, the cloud customer is responsible only for the security of the data and the cloud provider is responsible for the full stack from the physical security of the cloud data centers to the application. Due to multitenancy in cloud environments, particularly in SaaS models, customer controls and resources are necessarily limited by the cloud provider.

2.8.2 Virtualization

Virtualization technology emulates real — or physical — computing resources, such as servers (compute), storage, networking, and applications. Virtualization allows multiple applications or server workloads to run independently on one or more physical resources.

A *hypervisor* allows multiple, virtual ("guest") operating systems to run concurrently on a single physical host computer. The hypervisor functions between the computer operating system (OS) and the hardware kernel. There are two types of hypervisors:

- **Type 1** (*native* or *bare metal*). Runs directly on the host computer's hardware.
- **Type 2** (*hosted*). Runs within an operating system environment.

> **Key Terms**
>
> A *hypervisor* allows multiple, virtual (or guest) operating systems to run concurrently on a single physical host computer.
>
> A *native* (also known as a *Type 1* or *bare metal*) hypervisor runs directly on the host computer hardware.
>
> A *hosted* (also known as a *Type 2*) hypervisor runs within an operating system environment.

Virtualization is a key technology used in data centers and cloud computing to optimize resources. Some important security considerations associated with virtualization include:

- **Dormant virtual machines (VMs).** In many data center and cloud environments, inactive VMs are routinely (often automatically) shutdown when they are not in use. VMs that are shutdown for extended periods of time (weeks or months) may be inadvertently missed when anti-malware updates and security patches are applied.
- **Hypervisor vulnerabilities.** In addition to vulnerabilities within the hosted applications, VMs, and other resources in a virtual environment, the hypervisor itself may be vulnerable, which potentially can expose hosted resources to attack.
- **Intra-VM communications.** Network traffic between virtual hosts, particularly on a single physical server, may not traverse a physical switch. This lack of visibility increases troubleshooting complexity and can increase security risks due to inadequate monitoring and logging capabilities.
- **VM sprawl.** Virtual environments can grow quickly, leading to a breakdown in change management processes and exacerbating security issues such as dormant VMs, hypervisor vulnerabilities, and intra-VM communications.

2.8.3 Local and remote storage

Three basic storage technologies commonly are used in local and remote storage:

- **Block.** In data storage, a block is a sequence of bits or bytes of a fixed length or size, for example, 512 bytes. Devices such as computers, servers, and storage area networks

(SANs), access block-based storage through various interfaces including advanced technology attachment (ATA), fibre channel protocol (FCP), internet SCSI (iSCSI), serial attached SCSI (SAS), serial ATA (SATA), and small computer system interface (SCSI). To use block-based storage you must create a volume, install an operating system, and mount (or attach to) the storage device.

- **File.** File-based storage systems, such as network-attached storage (NAS), typically run their own operating system and manage access control and permissions using *inodes*. File-based storage systems are accessed using protocols such as Common Internet File System (CIFS), Network File System (NFS), and Server Message Block (SMB). File-based storage typically requires mapping a drive letter or network path to the storage device.
- **Object.** Object-based storage use variable-sized data containers, known as objects, which are organized into a flat address space rather than a hierarchical file system, such as a directory-based structure. Object-based storage systems are used to manage large content repositories containing several petabytes of data and billions of objects. Users typically access object-based storage using a web browser or an application that uses an HTTP interface to interact with the storage device.

Key Terms

Inodes store information about files and directories in a file-based storage system, but not the filenames or data content itself.

2.8 Knowledge Check

Test your understanding of the fundamentals in the preceding section. Review the correct answers in the Appendix at the end of this guide.

1. **Multiple Choice.** Which three cloud computing service models are defined by NIST? (Choose three.)

    a) Software as a Service (SaaS)
    b) Platform as a Service (PaaS)
    c) Desktop as a Service (DaaS)
    d) Infrastructure as a Service (IaaS)

2. **Fill in the Blank.** A _____cloud infrastructure is composed of two or more cloud deployment models, bound together by standardized or proprietary technology that enables data and application portability.

3. **Fill in the Blank.** The _____ defines who (customer and/or provider) is responsible for what, related to security, in the public cloud.

4. **Fill in the Blank.** A _____ allows multiple, virtual operating systems to run concurrently on a single physical host computer.

5. **Multiple Choice.** Which three important security considerations are associated with virtualization? (Choose three.)

    a) dormant VMs
    b) hypervisor vulnerabilities
    c) hypervisor sprawl
    d) intra-VM communications

6. **Fill in the Blank.** A Storage Area Network (SAN) uses _____-based storage.

2.9 Networking Concepts

Important networking and operations concepts covered in this section include server and system administration, directory services, troubleshooting, IT best practices, and technical support.

2.9.1 Server and system administration

Server and system administrators perform a variety of important tasks in a network environment. Typical server and system administration tasks include:

- Account provisioning and de-provisioning
- Managing account permissions
- Installing and maintaining server software
- Maintaining and optimizing servers, applications, databases (may be assigned to a database administrator), and network (may be assigned to a network administrator) and security (may be assigned to a security administrator) devices
- Installing security patches
- Managing system and data backup and recovery
- Monitoring network communication and server logs
- Troubleshooting and resolving server and system issues

2.9.1.1 Patch management

New software vulnerabilities and exploits are discovered each year, requiring diligent software patch management by system and security administrators in every organization.

However, patch management only protects an organization's endpoints after a vulnerability has been discovered and the patch installed. Delays of days, weeks, or longer are inevitable as security patches for newly discovered vulnerabilities must be developed, distributed, tested, and deployed. Although patch management is an important aspect of any information security program, like signature-based anti-malware detection, it is an endless race against time that offers no protection against zero-day exploits.

2.9.1.2 Configuration management

Configuration management is the formal process used by organizations to define and maintain standard configurations for applications, devices, and systems throughout their lifecycle. For example, a particular desktop PC model may be configured by an organization with specific security settings such as enabling whole disk encryption and disabling USB ports. Within the desktop operating system, security settings such as disabling unneeded and risky services (for example, FTP and Telnet) may be configured. Maintaining standard configurations on

applications, devices, and systems used by an organization helps to reduce risk exposure and improve security posture.

2.9.2 Directory services

A directory service is a database that contains information about users, resources, and services in a network. The directory service associates users and network permissions to control who has access to which resources and services on the network. Examples of directory services include:

- **Active Directory.** A centralized directory service developed by Microsoft for Windows networks to provide authentication and authorization of users and network resources. Active Directory uses the lightweight directory access protocol (LDAP), *Kerberos*, and the domain name system (DNS, discussed in Section 2.1.5)
- **lightweight directory access protocol (LDAP).** An IP-based client-server protocol that provides access and manages directory information in TCP/IP networks

> **Key Terms**
> *Kerberos* is a ticket-based authentication protocol in which "tickets" are used to identify network users.

2.9.3 Structured host and network troubleshooting

If a network or segment of a network goes down, it could have a negative impact on your organization or business. Network administrators should use a systematic process to troubleshoot network problems when they occur to restore the network to full production as quickly as possible without causing new issues or introducing new security vulnerabilities. The troubleshooting process performed by a network administrator to resolve network problems quickly and efficiently is a highly sought skill in IT.

Two of the most important troubleshooting tasks a network administrator performs occur long before a network problem occurs: baselining and documenting the network.

A baseline provides quantifiable metrics that are periodically measured with various network performance monitoring tools, protocol analyzers, and packet sniffers. Important metrics might include application response times, server memory and processor utilization, average and peak network throughput, and storage input/output operations per second. These baseline metrics provide an important snapshot of "normal" network operations to help network administrators

identify impending problems, troubleshoot current problems, and know when a problem has been fully resolved.

Network documentation should include logical and physical diagrams, application data flows, change management logs, user and administration manuals, and warranty and support information. Network baselines and documentation should be updated any time a significant change to the network occurs, and as part of the change management process of an organization.

Many formal multi-step troubleshooting methodologies have been published and various organizations or individual network administrators may have their own preferred method. Generally speaking, troubleshooting consists of these steps:

1. **Discover the problem.**
2. **Evaluate system configuration against the baseline.**
3. **Track possible solutions.**
4. **Execute a plan.**
5. **Check results.**
6. **Verify solution.** (If unsuccessful, go back to step 2; if successful, proceed to step 7.)
7. **Deploy positive solution.**

Troubleshooting host and network connectivity problems typically starts with analyzing the scope of the problem and identifying the devices and services that are affected. Problems with local hosts are typically much easier to assess and remedy than problems that affect a network segment or service. For an individual device that loses network connectivity, the problem sometimes be easily resolved by simply restarting the device. However, problems with integrated or shared services, for example web or file services, can be complex, and restarting a service or rebooting a device may actually compound the problem. Connectivity problems may be intermittent or difficult to trace, so it's important that your troubleshooting processes follow an approved or standardized methodology.

The OSI model (discussed in Section 2.3.1) provides a logical model for troubleshooting complex host and network issues. Depending on the situation, you might use a bottom-up (discussed in the following paragraphs), top-down, or "divide and conquer" approach when using the OSI model to guide your troubleshooting efforts. In other situations, you might make an "educated guess" about the source of the issue and begin investigating at the corresponding layer of the

OSI model, or use the substitution method (replacing a bad component with a known good component) to quickly identify and isolate the cause of the issue.

Using a bottom-up approach to diagnose connectivity problems, you begin at the Physical layer of the OSI model by verifying network connections and device availability. For example, a wireless device may have power to the antenna or transceiver temporarily turned off. Or, a wireless access point may have lost power because a circuit breaker may have been tripped offline or a fuse may have been blown. Similarly, a network cable connection may be loose or the cable may be damaged. Thus, before you begin inspecting service architectures, it is best to start with the basics: confirm physical connectivity.

Moving up to the Data Link layer, you verify data link architectures, such as compatibility with a particular standard or frame type. While Ethernet is a predominant LAN network standard, devices that roam (such as wireless devices) sometimes automatically switch between Wi-Fi, Bluetooth, and Ethernet networks. Wireless networks usually have specified encryption standards and keys. Connectivity may be lost because a network device or service has been restored to a previous setting and the device is not responding to endpoints requests that are using different settings. Firewalls and other security policies may also be interfering with connection requests. It's never a good idea to disable firewalls, but in a controlled network environment with proper procedures established, you may find that temporarily disabling or bypassing a security appliance resolves a connectivity issue. The remedy then is to properly configure security services to allow the required connections.

Various connectivity problems may also occur at the Network layer. Important troubleshooting steps include confirming proper network names and addresses. Devices may have improperly assigned IP addresses that are causing routing issues or IP address conflicts on the network. A device may have an improperly configured IP address because it is not able to communicate with a DHCP server on the network. Similarly, networks have different identities, such as wireless SSIDs, domain names, and workgroup names. Another common problem exists when a particular network has conflicting names or addresses. Issues with DNS name resolvers may be caused by DNS caching services or connecting to the wrong DNS servers. The *internet control message protocol* (ICMP) is used for network control and diagnostics at the Network layer of the OSI model. Commonly used ICMP commands include *ping* and *traceroute*. These two simple, but powerful commands (as well as other ICMP commands and options) are some of the most commonly used tools for troubleshooting network connectivity issues. You can run ICMP commands using the command-line interface on computers, servers, routers, switches, and many other networked devices.

At the Transport layer, communications are more complex. Latency and network congestion can interfere with communications that depend on timely acknowledgements and handshakes. Time to Live (TTL) values sometimes have to be extended in the network service architecture to allow for slower response times during peak network traffic hours. Similar congestion problems can occur when new services are added to an existing network, or when a local device triggers a prioritized service, such as a backup or antivirus scan.

Session layer settings can also be responsible for dropped network connections. For example, devices that automatically go into a power standby mode ("sleep") may have expired session tokens that fail when the device attempts to resume connectivity. At the server, failover communications or 'handshake' negotiations with one server may not translate to other clustered servers. Sessions may have to restarted.

Presentation layer conflicts are often related to changes in encryption keys or updates to service architectures that are not supported by different client devices. For example, an older browser may not interoperate with a script or a new encoding standard.

Application layer network connectivity problems are extremely common. Many applications may conflict with other apps. Apps also may have caching or corrupted files that can be remedied only by uninstalling and reinstalling, or updating to a newer version. Some apps also require persistent connections to updater services or third parties, and network security settings may prevent those connections from being made.

Other troubleshooting steps may include searching log files for anomalies and significant events, verifying that certificates or proper authentication protocols are installed and available, verifying encryption settings, clearing application caches, updating applications, and, for endpoints, possibly removing and reinstalling an application. Search vendor-supported support sites and forums, as well as frequently asked questions (FAQ) pages before you make changes to installed services. It's also important that you are aware of any Service-Level Agreements (SLAs) that your organization must meet.

Always follow proper troubleshooting steps, keep accurate records of any changes that you attempt, document your changes, and publish any remedies so that others can learn from your troubleshooting activities.

2.9.4 ITIL fundamentals

ITIL (formerly known as the Information Technology Infrastructure Library) is a five-volume set of IT service management best practices:

- **Service Strategy.** Addresses IT services strategy management, service portfolio management, IT services financial management, demand management, and business relationship management.
- **Service Design.** Addresses design coordination, service catalog management, service level management, availability management, capacity management, IT service continuity management, Information security management system, and supplier management.
- **Service Transition.** Addresses transition planning and support, change management, service asset and configuration management, release and deployment management, service validation and testing, change evaluation, and knowledge management.
- **Service Operation.** Addresses event management, incident management, service request fulfillment, problem management, and access management.
- **Continual Service Improvement.** Defines a seven-step process for improvement initiatives, including identifying the strategy, defining what will be measured, gathering the data, processing the data, analyzing the information and data, presenting and using the information, and implementing the improvement.

2.9.5 Help desk and technical support

An important function in any IT department is the help desk (or technical support). The help desk is the IT department's liaison to an organization's users (or customers).

Help desks are commonly organized in multiple tiers, for example, User Support (tier 1), Desktop Support (tier 2), Network Support (tier 3). User issues that cannot be resolved at tier 1 are escalated to the appropriate tier.

Typical performance measures for a help desk include service-level agreements (SLAs), wait times, first-call resolution, and mean-time-to-repair (MTTR).

2.9 Knowledge Check

Test your understanding of the fundamentals in the preceding section. Review the correct answers in the Appendix at the end of this guide.

1. **Fill in the Blank.** _____ is a network directory service developed by Microsoft for Windows networks.
2. **Fill in the Blank.** _____ is a set of IT service management best practices.

Module 3 – Cybersecurity Essentials

Knowledge Objectives

- Identify the security solutions that comprise the Palo Alto Networks Security Operating Platform.
- Discuss the fundamental components of a next-generation firewall and explain the basic operation of a next-generation firewall. Compare and contrast traditional port-based firewalls and next-generation firewalls.
- Describe the need for centralized network security management and explain its benefits to an organization.
- Identify the major components of the Palo Alto Networks Traps Advanced Endpoint Protection deployment architecture and explain how Traps protects endpoints from malware and exploits.
- List the requirements to safely enable mobile devices in the enterprise, identify the primary components of GlobalProtect, and describe the basic functionality of GlobalProtect.
- Explain the importance of continuous, real-time monitoring in the public cloud and how Evident enables organizations to protect and segment their public cloud workloads, ensure continuous regulatory and policy compliance, and discover and classify data within containers and buckets.
- Demonstrate an understanding of unique SaaS-based security risks and how Aperture protects SaaS-based applications and data.
- Describe Palo Alto Networks cloud-delivered security services within the Application Framework and Logging Service including behavioral analytics, log management, threat intelligence, threat indicator sharing, and malware analysis.

3.1 Security Operating Platform

The Palo Alto Networks Security Operating Platform (see Figure 3-1) is a purpose-built, fully integrated cybersecurity approach that helps organizations get control of their networks and protect critical assets.

Figure 3-1: Palo Alto Networks Security Operating Platform.

The Security Operating Platform makes prevention, action, and control integral and central to enterprise security strategy. The Security Operating Platform provides visualization, protection, and control capabilities for all network traffic, applications, users, and endpoints. Tight integrations across the platform and with ecosystem partners (third-party vendors) deliver consistent security across clouds, networks, and mobile devices.

The sections that follow describe the different Palo Alto Networks solutions that comprise the Security Operating Platform.

3.1 Knowledge Check

Test your understanding of the fundamentals in the preceding section. Review the correct answers in the Appendix at the end of this guide.

1. **Fill in the Blank.** _____ is a purpose-built, fully integrated cybersecurity approach that helps organizations get control of their networks and protect critical assets.

2. **Multiple Choice.** Which three options are key components of the Security Operating Platform? (Choose three.)

    a) network security
    b) advanced endpoint protection
    c) cloud security
    d) application development security

3.2 Network Security

Network security components in the Security Operating Platform include Palo Alto Networks next-generation firewalls (NGFWs), the Expedition migration tool, and Panorama for centralized network security management.

3.2.1 Next-generation firewalls

Fundamental shifts in application usage, user behavior, and complex network infrastructure create a threat landscape that exposes weaknesses in traditional port-based network firewalls. End users want access to an ever-increasing number of applications, operating across a wide range of device types, often with little regard for the business or security risks. Meanwhile data center expansion, network segmentation, virtualization, and mobility initiatives are forcing organizations to rethink how to enable access to applications and data, while protecting their networks from a new, more sophisticated class of advanced threats that evade traditional security mechanisms.

Palo Alto Networks NGFWs are the core of the Security Operating Platform. The NGFW inspects all traffic – including applications, threats, and content – and ties it to the user, regardless of location or device type. The application, content, and user become integral components of the enterprise security policy.

NGFWs classify network traffic based on the application's identity in order to enable visibility and control of all types of applications running on enterprise networks. The essential functional requirements for an effective NGFW include:

- **Application identification.** Accurately identify applications regardless of port, protocol, evasive techniques, or encryption. Provide visibility of applications and granular policy-based control over applications, including individual application functions.
- **User identification.** Accurately identify users and subsequently use identity information as an attribute for policy control.
- **Content identification.** Content identification controls traffic based on complete analysis of all allowed traffic, using multiple threat prevention and data loss prevention techniques in a single pass architecture that fully integrates all security functions.

Palo Alto Networks NGFWs are built on a single-pass architecture (see Figure 3-2), which is a unique integration of software and hardware that simplifies management, streamlines

processing, and maximizes performance. The single-pass architecture integrates multiple threat prevention disciplines (IPS, anti-malware, URL filtering, etc.) into a single stream-based engine with a uniform signature format. This architecture allows traffic to be fully analyzed in a single pass without the performance degradation seen in multifunction 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.

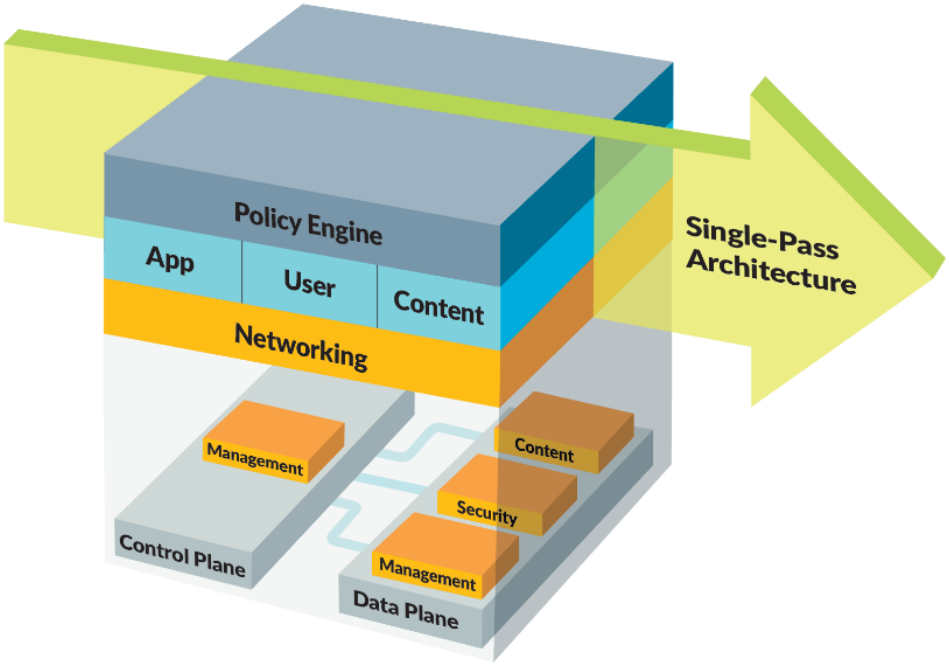

Figure 3-2: Palo Alto Networks NGFWs use a single-pass architecture.

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 in 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.

3.2.1.1 Application identification

Stateful packet inspection technology – the basis for most of today's legacy firewalls – was created more than 25 years ago, 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. As a result, 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.

Establishing port and protocol information is a first step in application identification, but it is insufficient by itself. Robust application identification and inspection in an NGFW enables granular control of the flow of sessions through the firewall. Identification is based on the specific applications (such as Skype, Gmail, and WebEx) that are being used, instead of just relying on the underlying set of often indistinguishable network communication services (see Figure 3-3).

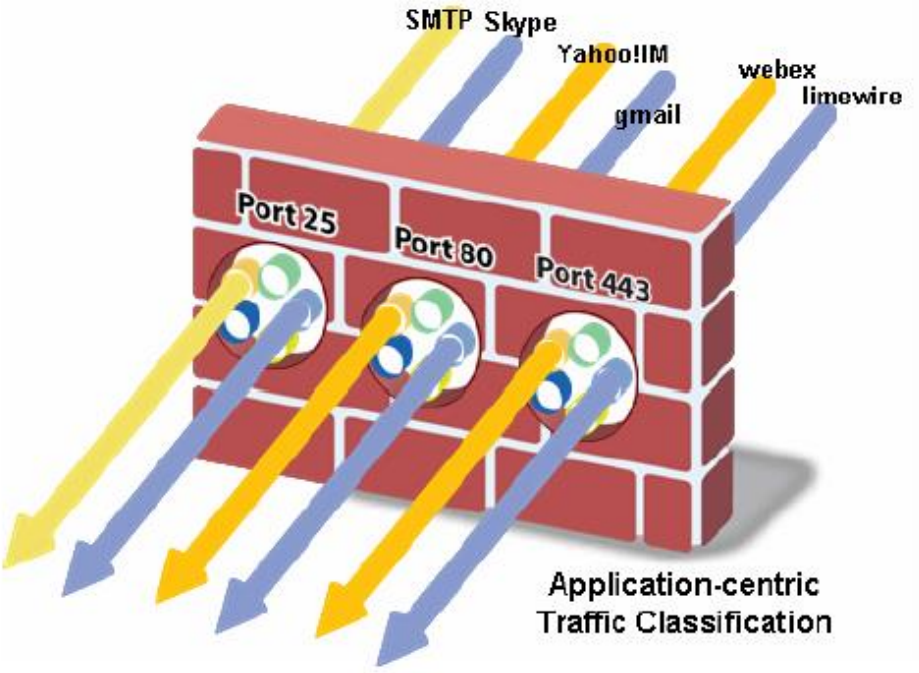

Figure 3-3: Application-centric traffic classification identifies specific applications on the network, irrespective of the port and protocol in use.

Application identification provides visibility and control over work-related and non-work-related applications that can evade detection by legacy port-based firewalls, for example, by masquerading as legitimate traffic, hopping ports, or slipping past the firewall using encryption.

Application identification technology in a Palo Alto Networks NGFW does not rely on a single element, such as 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 that is applied to the session. 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 NGFW executes is the identification of the applications traversing the network using App-ID. Using a multifaceted approach, App-ID determines what the application is, irrespective of port, protocol, encryption (SSL and SSH) or other evasive tactics employed. The number and order of identification mechanisms used to identify the application vary depending on the application. The application identification techniques (see Figure 3-4) used include:

- Application protocol detection and decryption. Determines the application protocol (for example, HTTP) and, if SSL is in use, decrypts the traffic so that it can be analyzed further. Traffic is re-encrypted after all the NGFW technologies have had an opportunity to operate.
- Application protocol decoding. Determines whether the initially detected application protocol is the "real one," or if it is being used as a tunnel to hide the actual application (for example, Tor might be inside of HTTPS).
- Application signatures. Context-based signatures look for unique properties and transaction characteristics to correctly identify the application regardless of the port and protocol being used. These signatures include the ability to detect specific functions within applications (such as file transfers within SaaS applications).
- Heuristics. For traffic that eludes identification by signature analysis, heuristic (or behavioral) analyses are applied, which enable identification of any troublesome applications, such as P2P or VoIP tools that use proprietary encryption.

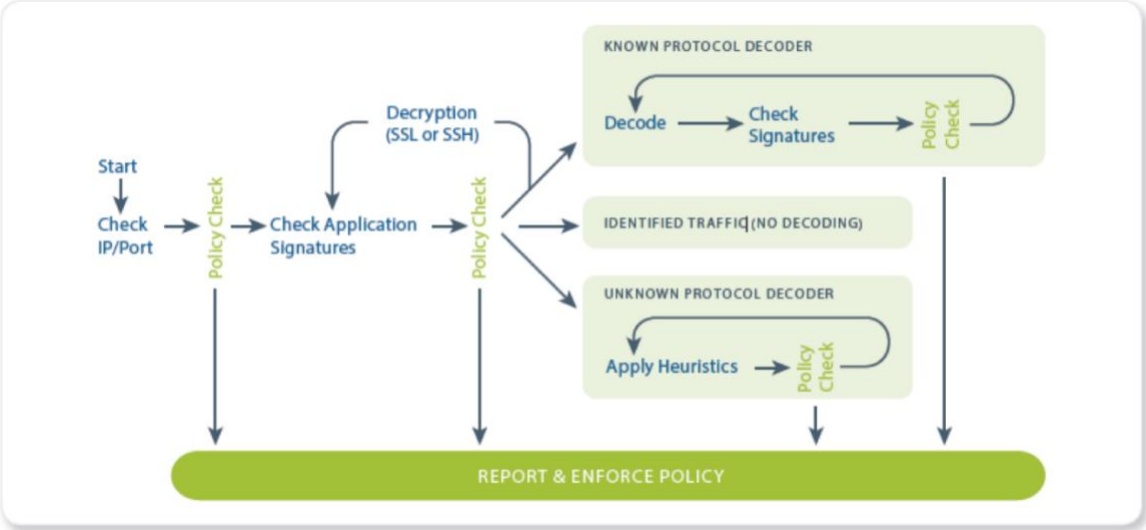

Figure 3-4: How Palo Alto Networks App-ID classifies applications.

Key Terms

Tor ("The Onion Router") is software that enables anonymous communication over the Internet.

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

App-ID: Dealing with custom or unknown applications

Palo Alto Networks adds an average of five new applications to App-ID each week, yet unknown application traffic is still detected on the network, such as:

- **Unknown Commercial Applications.** Using Application Command Center (ACC) and the log viewer, administrators can quickly determine whether an unknown application is a commercial application. Using the packet capture (pcap) feature on the Palo Alto Networks NGFW, administrators can record the traffic and submit it for App-ID development. The new App-ID is developed, tested with the organization, then added to the global database for all users.
- **Internal or Custom Applications.** Using ACC and the log viewer, administrators can quickly determine whether an unknown application is an internal or custom application. You can develop a custom App-ID for the application, using the exposed protocol decoders. The protocol decoders that have been exposed include:

- FTP (file transfer protocol)
- HTTP (hypertext transfer protocol) and HTTPS (HTTP secure, or HTTP over SSL)
- IMAP (internet message access protocol) and SMTP (simple mail transfer protocol)
- RTSP (real time streaming protocol)
- Telnet
- unknown-TCP, unknown-UDP, and file body (for html/pdf/flv/swf/riff/mov)

After the custom App-ID is developed, traffic identified by it is treated in the same manner as the previously classified traffic; it can be enabled via policy, inspected for threats, shaped using quality of service (QoS), etc. Alternatively, an application override can be created and applied, which effectively renames the application. Custom App-ID entries are managed in a separate database on the NGFW to ensure they are not impacted by weekly App-ID updates.

An important point to highlight is that Palo Alto Networks NGFWs 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.

App-ID in action: 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. If there is a matching decryption policy rule, then the decryption engine and protocol decoders are initiated to decrypt the SSL and detect that it is HTTP traffic. After 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”: 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 detects the WebEx Desktop Sharing feature, which is then displayed in ACC. At this stage, an administrator has learned more about the application use and can exert policy control over the use of the WebEx Desktop Sharing feature separately from general WebEx use.

Application identification and policy control

Application identification enables administrators to see the applications on the network, learn how they work, and analyze their behavioral characteristics and relative risk. When application identification is used in conjunction with user identification, administrators can see exactly who is using the application based on their identity, not just an IP address. With this information, administrators can use granular rules – based on a positive security model – to block unknown applications, while enabling, inspecting, and shaping those applications that are allowed.

After an application has been identified and a complete picture of its usage is gained, organizations can apply policies with a range of responses that are far more granular than the "allow" or "deny" actions available in legacy firewalls. 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
- 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 applications, 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 3-5 shows SharePoint as the container application, and the individual functions within.

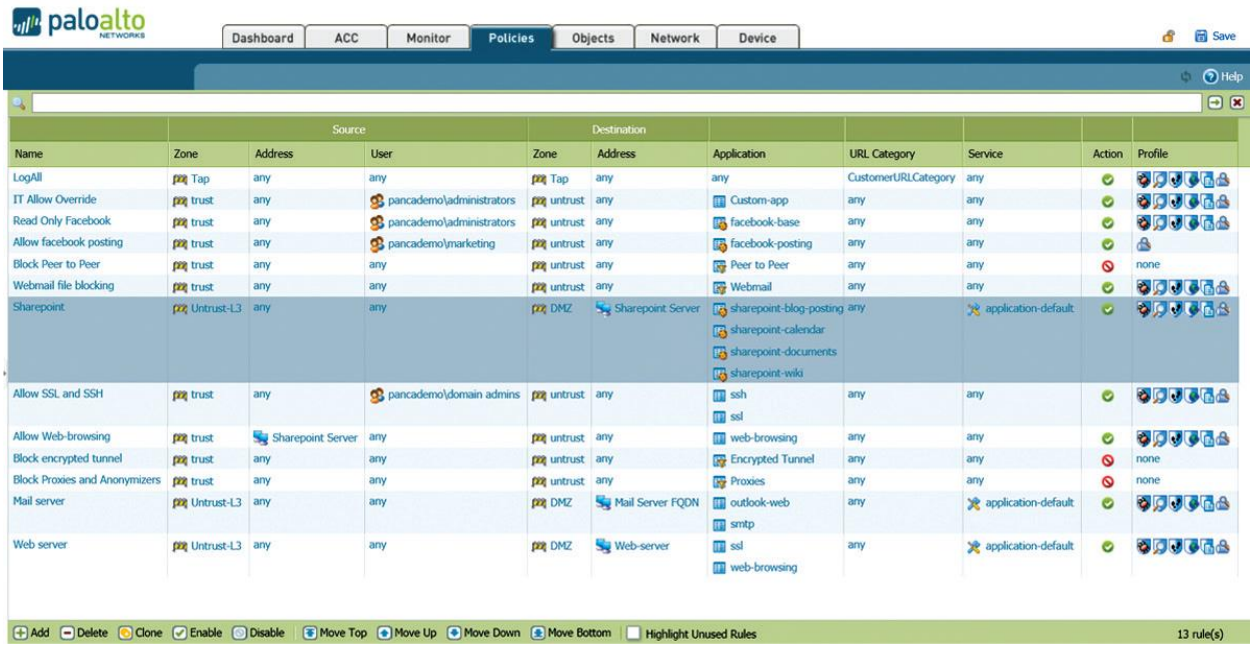

Figure 3-5: Application Function Control maximizes productivity by safely enabling the application itself (Microsoft SharePoint) or individual functions.

Controlling multiple applications: Dynamic filters and groups

In some cases, organizations may want to control applications in bulk, as opposed to controlling them individually. The two mechanisms in the Palo Alto Networks NGFW 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 allow their use for certain users, while blocking their use for others. For example, remote management applications such as remote desktop protocol (RDP), Telnet, and Secure Shell (SSH) are commonly used by IT support personnel, yet employees that fall outside of these groups are also known to use these tools as a means of accessing their home networks. A group of applications can be created and assigned to IT 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, or risk factor. After the desired filter is created, a policy that

blocks or enables and scans the traffic can be applied. As new App-ID files are added that fulfill the filter criteria, the filter is automatically updated as soon as the device is updated, thereby minimizing the administrative effort associated with policy management.

3.2.1.2 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, by itself, is no longer an adequate mechanism for monitoring and controlling user activity.

User-ID: Integrating user information and security policies

Creating and managing security policies on a NGFW, 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 information in legacy, port-based firewalls. Palo Alto Networks User-ID enables organizations to leverage user information stored in a wide range of repositories for the following purposes:

- **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, which provides a more complete picture of the incident.

User-ID in action

User-ID seamlessly integrates Palo Alto Networks NGFWs with a wide range of user repositories and terminal services environments. Depending on the network environment, multiple techniques can be configured to accurately 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 (see Figure 3-6).

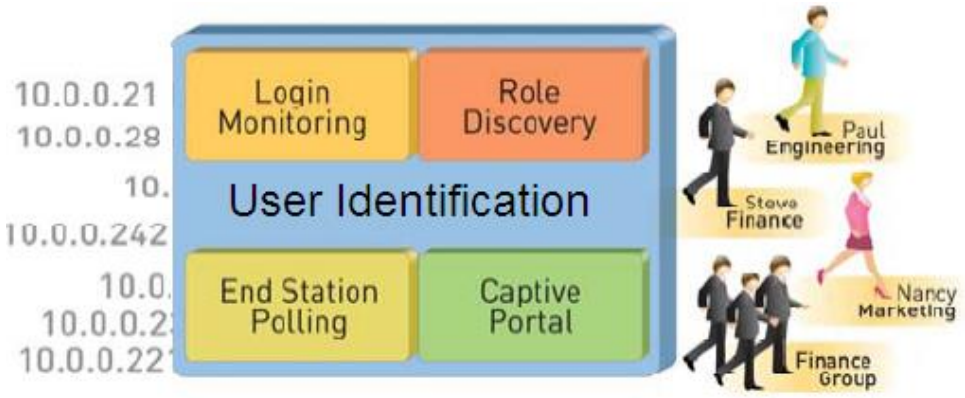

Figure 3-6: User identification integrates enterprise directories for user-based policies, reporting, and forensics.

After the applications and users are identified, full visibility and control within ACC, policy editing, and logging and reporting are available. User-ID tools and techniques include:

- **Authentication events.** Monitoring 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. User-ID can be configured to monitor authentication events for:
    - **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 authentication events to identify users on the network for creation and enforcement of policy.
    - **Microsoft Exchange Server:** User-ID can be configured to constantly monitor Microsoft Exchange logon events produced by clients accessing their email. Using this technique, even Mac OS X, Apple iOS, and 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 lightweight directory access protocol (LDAP) queries on Novell eDirectory servers.
- **User authentication.** This technique allows organizations to configure a challenge-response authentication sequence to collect user and IP address information, using the following tools:

- **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. A 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 NT LAN Manager (NTLM) authentication request to the web browser to make the authentication process transparent to the user.

- **GlobalProtect:** Users logging in to the network with GlobalProtect (discussed in Section 3.3.2) 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 are being accessed by users.

    - **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 username provided to establish the session.

    - **Terminal services:** Users sharing IP addresses while 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 include:

- **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:** 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 organizations 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.

- **Syslog listener.** The agent runs a syslog listener on a designated port that is able to parse the syslog messages and convert the information into appropriate User-ID mappings.

To allow organizations to specify security rules based on user groups and resolve the group members automatically, User-ID integrates with directory servers using a standards-based protocol and a flexible configuration. After 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.

Visibility into a user's 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 business unit requirements and, if appropriate, can choose to inform the user that they are in violation of 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 created based on the application, category and subcategory, underlying technology, or application characteristics. 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 applications for all users.

3.2.1.3 Content identification

Content identification infuses NGFWs with capabilities not possible in legacy, port-based firewalls. Application identification eliminates threat vectors through the tight control of all types of applications. This capability immediately reduces the attack surface of the network, after which all allowed traffic is analyzed for exploits, malware, dangerous URLs, and 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.

Threat prevention

Enterprise networks are facing a rapidly evolving threat landscape full of modern applications, exploits, malware, and attack strategies that can avoid 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 antimalware solutions.

Palo Alto Networks Content-ID addresses these challenges with unique threat prevention capabilities not found in traditional security solutions. First, the NGFW removes the methods that threats use to hide from security through the complete analysis of all traffic, on all ports regardless of any evasion, tunneling, or circumvention techniques that are used. 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, using the following tools and techniques:

- **Application decoders.** Content-ID leverages the more than 100 application and protocol decoders in App-ID to look for threats hidden within application data streams. This tool enables the firewall to detect and prevent threats tunneled within approved applications that would pass by traditional IPS or proxy solutions.
- **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 C&C activity and vulnerability exploits in a single pass.
- **Vulnerability attack protection (IPS).** Robust routines for traffic normalization and defragmentation are joined by protocol-anomaly, behavior-anomaly, and heuristic detection mechanisms to provide protection from the widest range of both known and unknown threats.
- **Cloud-based intelligence.** For unknown content, WildFire (discussed in Section 3.5.5) provides rapid analysis and a verdict that the firewall can leverage.
- **SSL decryption.** More and more web traffic connections are 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.

Stream-based malware scanning

Prevention of known 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. Stream-based scanning minimizes performance and latency issues by receiving, scanning, and sending traffic to its intended destination immediately without having to first buffer and then scan the file (see Figure 3-7).

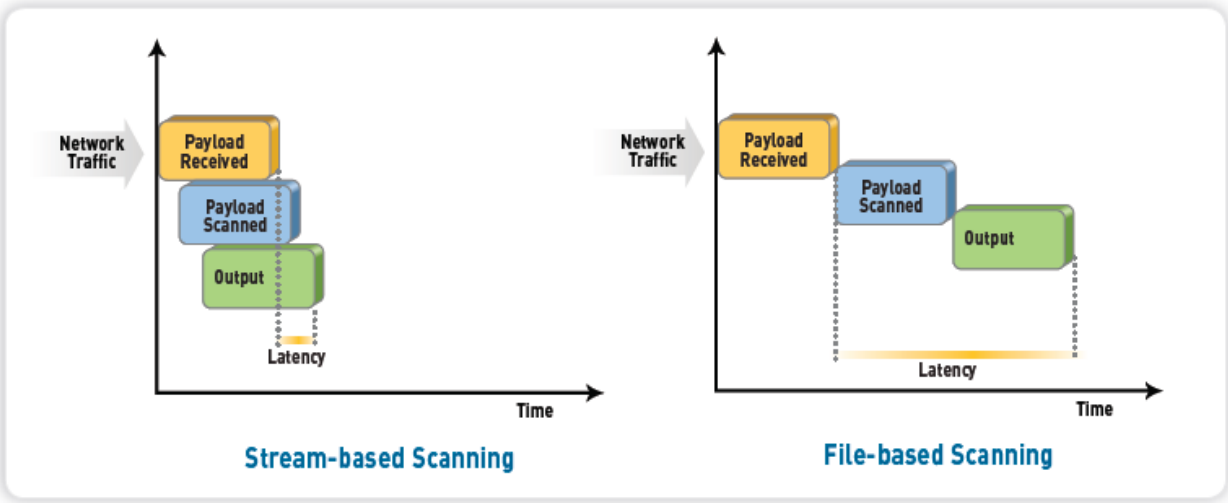

Figure 3-7: Stream-based scanning helps minimize latency and maximize throughput performance.

Intrusion prevention

Content-ID protects networks from all types of vulnerability exploits, buffer overflows, DoS attacks, and port scans that lead to the compromise of confidential and sensitive enterprise information. IPS mechanisms in Content-ID 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 defragmentation is performed to ensure the utmost accuracy and protection despite any packet-level evasion techniques.

URL filtering

To complement the threat prevention and application control capabilities, a fully integrated, on-box URL filtering database 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 URL database. URLs that are not categorized by the local URL database can be pulled into cache from a hosted 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 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 that threats are exposed, and QoS controls can be applied to streaming media sites. URL filtering visibility and policy controls can be tied to specific users through transparent integration with enterprise directory services (such as Active Directory, LDAP, and eDirectory), with additional insight provided through customizable reporting and logging.

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. To place some of the web activity ownership back in the user's hands, administrators can allow users to continue to the website or webpage, after being presented with a warning page, or can use passwords to override the URL filtering policy.

File and data filtering

Taking advantage of in-depth application inspection, file and data filtering enables enforcement of policies that reduce the risk of unauthorized information transfer or malware propagation. File and data filtering capabilities in Content-ID include:

- **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, which allow application use while preventing undesired inbound or outbound file transfer.

3.2.1.4 Log correlation and reporting

Powerful log filtering enables administrators to quickly investigate security incidents by correlating threats with applications and user identity. The Application Command Center (ACC) provides a comprehensive view of current and historical data – including network activity, application usage, users, and threats – in a highly visual, fully customizable, and easy-to-use

interactive format. This visibility enables administrators to make informed policy decisions and respond quickly to potential security threats.

ACC provides a tabbed view of network activity, threat activity, and blocked activity, and each tab includes pertinent widgets for better visualization of traffic patterns on the network (see Figure 3-8).

Figure 3-8: ACC provides a highly visual, interactive, and customizable security management dashboard.

Figure 3-9 shows a core widget of the ACC, the "Application Usage" widget. In this case, the widget shows application traffic in bytes. Applications (colored boxes) are grouped in application categories (gray bars). The size of each box indicates how much traffic a given application consumed during the selected time frame. The color of the box indicates the risk level of an application, with red being critical, orange being medium, and blue being the lowest risk. The tabular listing below the graph shows additional information, such as the number of sessions, threats detected, and content or files included, as well as URLs accessed by these applications.

Figure 3-9: The ACC "Application Usage" widget displays application traffic by type, amount, risk, and category.

Figure 3-10 is another ACC widget example that shows source and destination by region, with a visual display of where traffic is originating and going. The world maps are interactive and provide the ability to get more detail to learn more about traffic to or from individual countries.

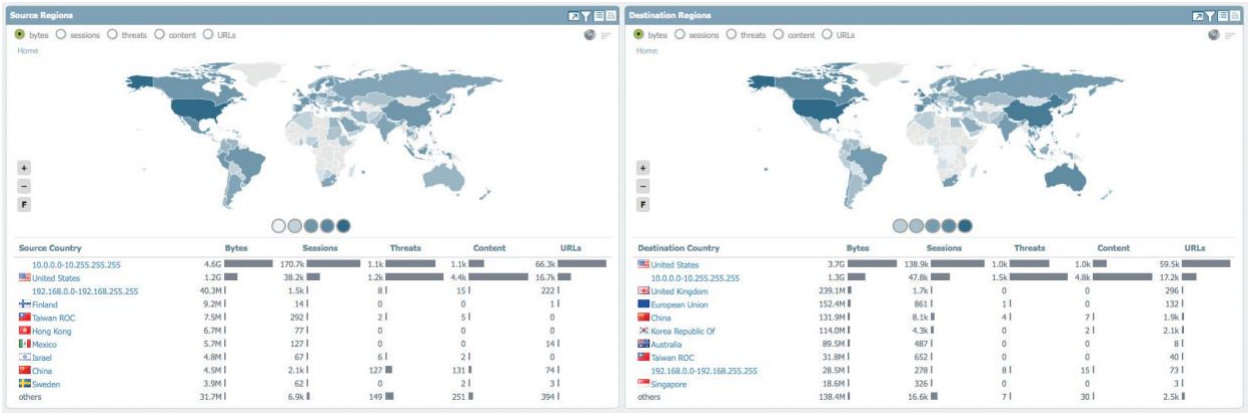

Figure 3-10: Geolocation awareness in ACC provides valuable information about source and destination of all application traffic.

Figure 3-11 is another ACC widget example that shows the power of application control in a NGFW versus a traditional port-based firewall. This widget shows applications with port hopping capabilities using nonstandard ports.

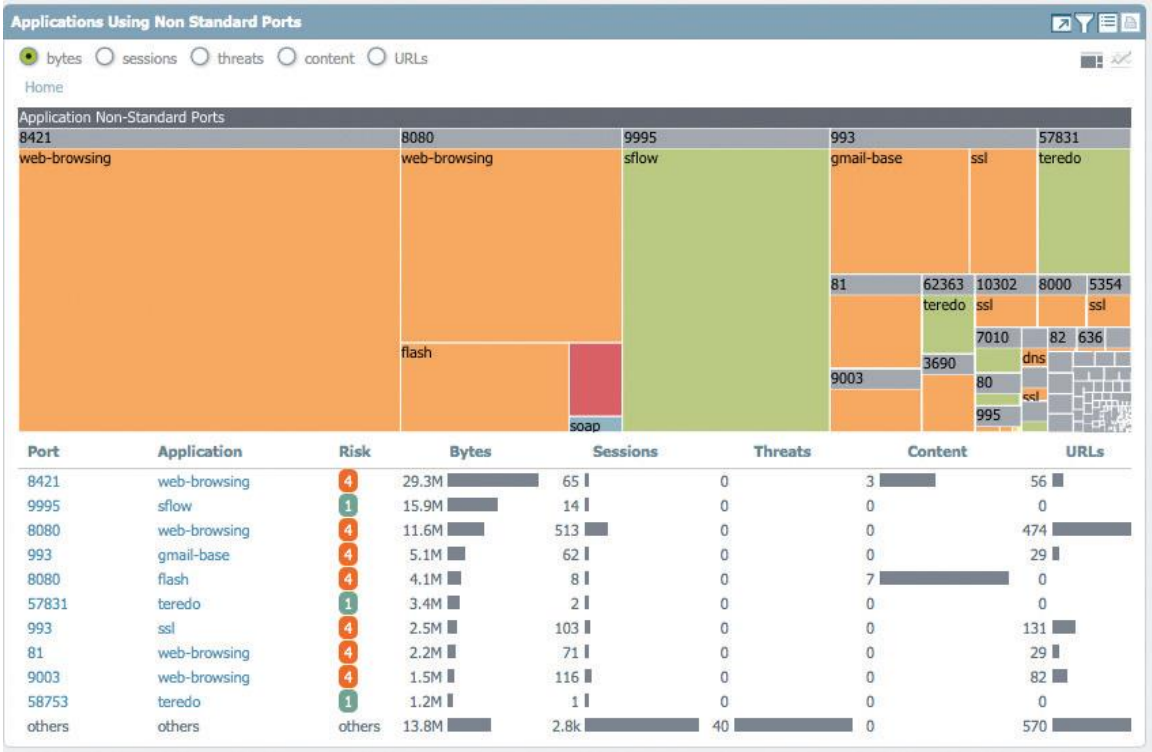

Figure 3-11: The "Applications Using Non Standard Ports" ACC widget highlights port hopping and showcases the importance of application versus port control.

Custom tabs can also be created, which include widgets that enable administrators to view more specific information. With the ACC, every administrator can customize their own views by selecting predesigned widgets from a drop-down list and building their own user interface (see Figure 3-12).

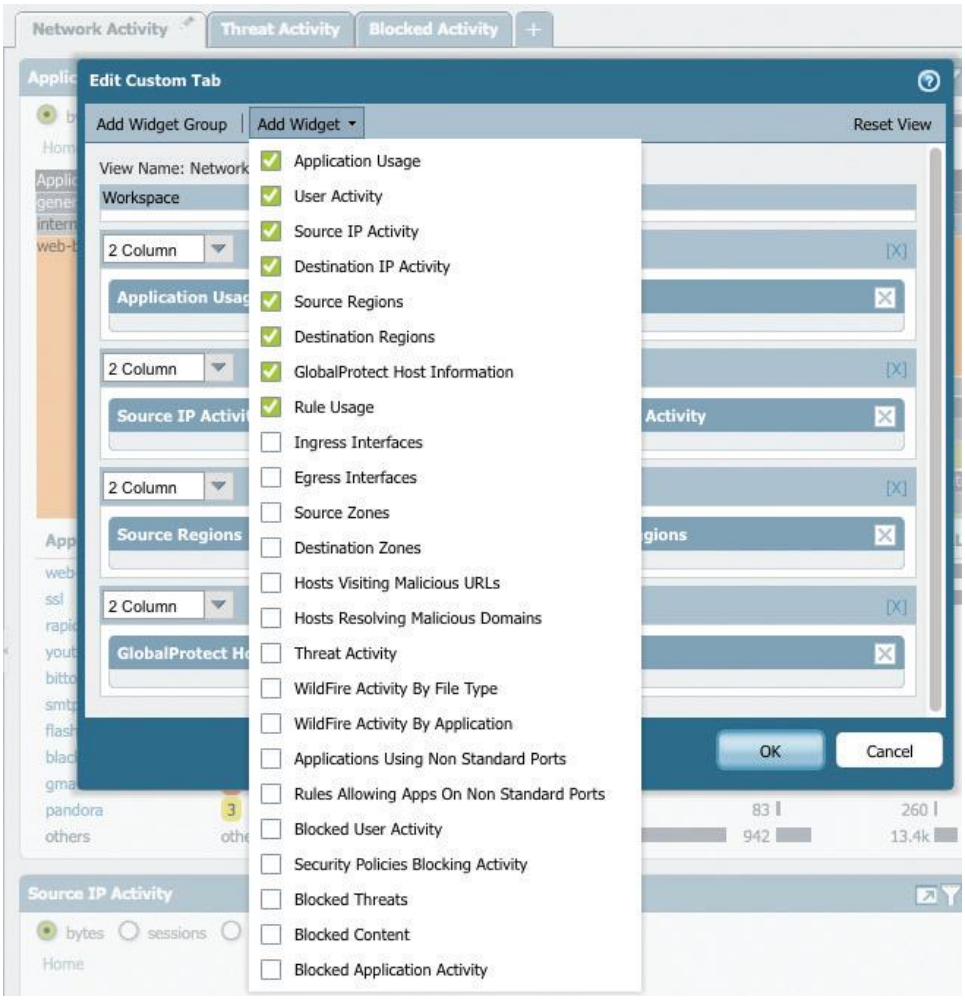

Figure 3-12: A large variety of widgets can be chosen to customize tabs in the ACC.

In addition to customizing existing tabs (network, threat, and blocked activity), new custom tabs can also be created to monitor certain employees, situations, or applications.

With the interactive capabilities of the ACC, you can learn more about applications, URL categories, risk levels, or threats to get a complete picture of network and threat activity (see Figure 3-13).

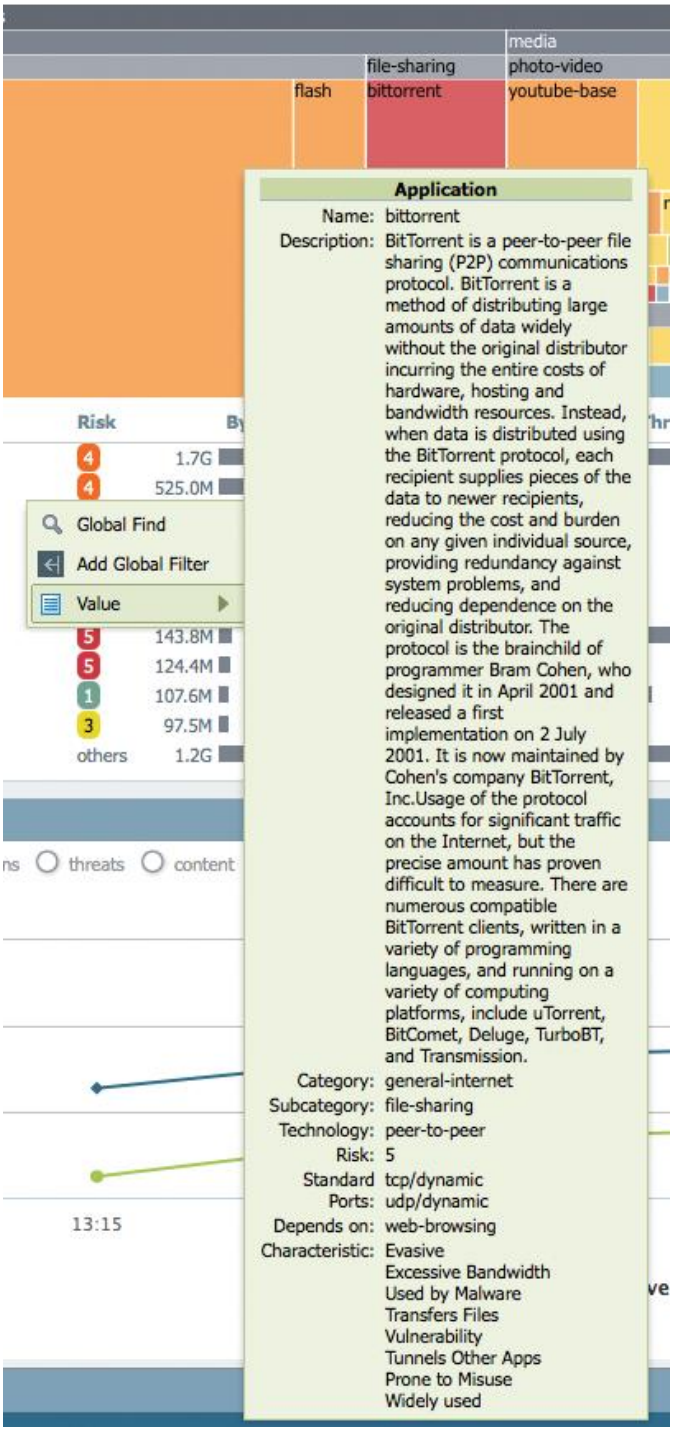

Figure 3-13: One-click, drill-down interactive capabilities provide additional information and the ability to apply any item as a global filter.

The Automated Correlation Engine in the ACC is an analytics tool that surfaces critical threats that may be hidden in the network, which reduces manual data mining and enables faster response times. It scrutinizes isolated events automatically across multiple logs, queries the

data for specific patterns, and correlates network events to identify compromised hosts. It includes correlation objects that are defined by the Palo Alto Networks Malware Research team. These objects identify suspicious traffic patterns, compromised hosts, and other events that indicate a malicious outcome. Some correlation objects can identify dynamic patterns that have been observed from malware samples in WildFire.

Correlation objects trigger correlation events when they match on traffic patterns and network artifacts that indicate a compromised host on your network. In the ACC, correlation triggers are clearly identified and highlighted to enable a fast response (see Figure 3-14).

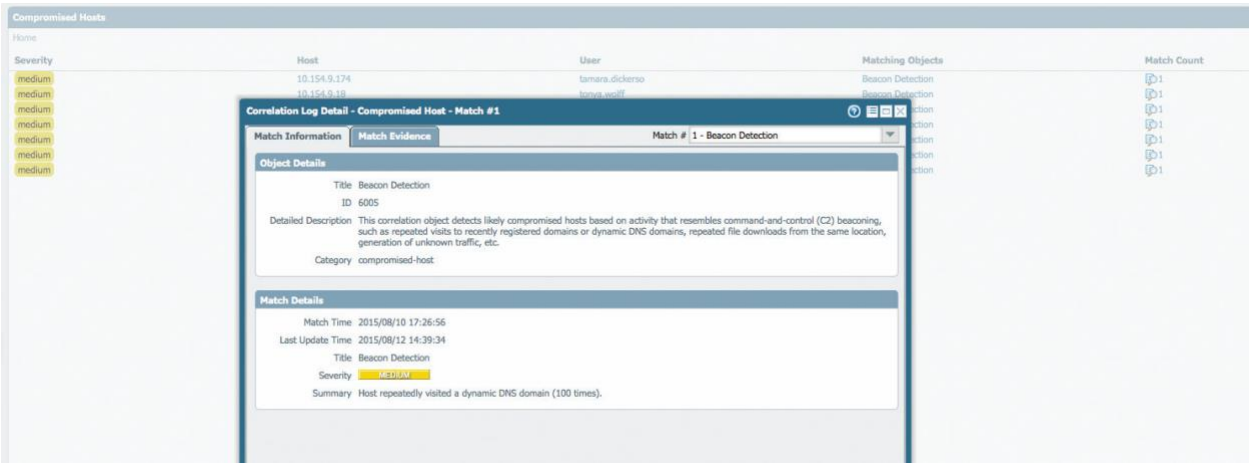

Figure 3-14: The Automated Correlation Engine automatically highlights compromised hosts in the ACC by correlating indicators of compromise (IOCs).

A log is an automatically generated, timestamped file that provides an audit trail for system events on the firewall or network traffic events that the firewall monitors. Log entries contain *artifacts*, which are properties, activities, or behaviors associated with the logged event, such as the application type or the IP address of an attacker. Each log type records information for a separate event type. For example, the firewall generates a Threat log to record traffic that matches a spyware, vulnerability, or virus signature or a DoS attack that matches the thresholds configured for a port scan or host sweep activity on the firewall.

The following logs can be viewed from the Monitor tab on Palo Alto Networks NGFWs:

- **Traffic logs.** These logs display an entry for the start and end of each session. Each entry includes the following information: date and time; source and destination zones, addresses and ports; application name; security rule applied to the traffic flow; rule action (allow, deny, or drop); ingress and egress interface; number of bytes; and session end reason.

- **Threat logs.** These logs display entries when traffic matches one of the Security Profiles attached to a security rule on the firewall. Each entry includes the following information: date and time; type of threat (such as virus or spyware); threat description or URL (Name column); source and destination zones, addresses, and ports; application name; alarm action (such as allow or block); and severity level.
- **URL Filtering logs.** These logs display entries for traffic that matches URL Filtering Profiles attached to security rules. For example, the firewall generates a log if a rule blocks access to specific websites and website categories or if you configured a rule to generate an alert when a user accesses a website.
- **WildFire Submissions logs.** The firewall forwards samples (files and emails links) to the WildFire cloud for analysis based on WildFire Analysis profiles settings. The firewall generates WildFire Submissions log entries for each sample it forwards after WildFire completes static and dynamic analysis of the sample. WildFire Submissions log entries include the WildFire verdict for the submitted sample.
- **Data Filtering logs.** These logs display entries for the security rules that help prevent sensitive information such as credit card numbers from leaving the area that the firewall protects.
- **Correlation logs.** The firewall logs a correlated event when the patterns and thresholds defined in a Correlation Object match the traffic patterns on your network.
- **Config logs.** These logs display entries for changes to the firewall configuration. Each entry includes the date and time, the administrator username, the IP address from where the administrator made the change, the type of client (web, CLI, or Panorama), the type of command executed, the command status (succeeded or failed), the configuration path, and the values before and after the change.
- **System logs.** These logs display entries for each system event on the firewall. Each entry includes the date and time, event severity, and event description.
- **HIP Match logs.** The GlobalProtect host information profile (HIP) feature enables you to collect information about the security status of the end devices accessing your network (such as whether they have disk encryption enabled). The firewall can allow or deny access to a specific host based on adherence to the HIP-based security rules you define. HIP Match logs display traffic flows that match a HIP Object or HIP Profile that you configured for the rules.

- **Alarms logs.** An alarm is a firewall-generated message that indicate that the number of events of a particular type (for example, encryption and decryption failures) has exceeded the threshold configured for that event type.

- **Unified logs.** Unified logs are entries from the Traffic, Threat, URL Filtering, WildFire Submissions, and Data Filtering logs displayed in a single view. Unified log view enables you to investigate and filter the latest entries from different log types in one place, instead of searching through each log type separately.

The reporting capabilities on the Palo Alto Networks NGFW enable you to keep a pulse on your network, validate your policies, and focus your efforts on maintaining network security. The following report types are available:

- **Predefined reports** allow you to view a quick summary of the traffic on your network. Predefined reports are available in four categories: Applications, Traffic, Threat, and URL Filtering.

- **User or Group Activity reports** allow you to schedule or create an on-demand report on the application use and URL activity for a specific user or for a user group. The report includes the URL categories and an estimated browse time calculation for individual users.

- **Custom reports** can be created and scheduled to show exactly the information you want to see by filtering on conditions and columns to include. You can also include query builders for more specific drill downs on report data.

- **PDF Summary reports** aggregate up to 18 predefined or custom reports/graphs from Threat, Application, Trend, Traffic, and URL Filtering categories into one PDF document.

- **Botnet reports** allow you to use behavior-based mechanisms to identify potential botnet-infected hosts in the network.

- **Report Groups** combine custom and predefined reports into report groups and compile a single PDF that is emailed to one or more recipients.

Reports can be generated on demand or on a recurring schedule, and they can be scheduled for email delivery.

3.2.2 Palo Alto Networks Expedition (Migration Tool)

The migration to a Palo Alto Networks NGFW is a critical step toward the prevention and detection of cyberattacks. Today's advanced threats require moving away from port-based

firewall policies, which are no longer adequate to protect against a modern threat landscape, into an architecture that reduces your attack surface by safely enabling only those applications that are critical to your organization, and eliminating applications that introduce risk.

Expedition enables organizations to analyze their existing environment, convert existing security policies to Palo Alto Networks NGFWs, and assist with the transition from proof-of-concept to production.

Primary functions of Expedition include:

- **Third-party migration** transfers the various firewall rules, addresses, and service objects to a PAN-OS XML config file that can be imported into a Palo Alto Networks NGFW. Third-party migration from the following firewall vendors is available:
    - Cisco ASA/PIX/FWSM
    - Check Point
    - Fortinet
    - McAfee Sidewinder
    - Juniper SRX/NETSCREEN
- **Adoption of App-ID** enables organizations to get the most value from their NGFW, while reducing the attack surface and regaining visibility and control over the organization through App-ID.
- **Optimization** keeps NGFWs operating at peak performance with services that include:
    - Architecture review
    - System health check
    - Configuration audit
    - Optional product tuning and configuration change implementation
- **Consolidation** of legacy firewalls to Palo Alto Networks virtual systems enables organizations to customize administration, networking, and security policies for the network traffic that is associated with specific departments or customers. In a standard virtual system interface configuration, each virtual system uses a dedicated interface to the Internet, requiring the use of multiple IP addresses. A shared gateway allows

organizations to create a common virtual interface for the virtual systems that correspond to a single physical interface. This shared gateway is helpful in environments where the ISP provides only a single IP address. All of the virtual systems communicate with the outside world through the physical interface using a single IP address.

- **Centralized management with Panorama** enables organizations to centrally manage the process of configuring devices, deploying security policies, performing forensic analysis, and generating reports across the organization's entire network of Palo Alto Networks NGFWs. Available as either a virtual appliance or a dedicated management platform, Panorama and the individual device management interfaces share the same Web-based look and feel, which ensures workflow consistency while minimizing any learning curve or delay in executing the task at hand.

- **Auto-zoning** automatically adapts security policies from vendors that currently do not use zones and zones-based rules. The mapping of zones depends on the routes and the zone interface IP address. The mappings adjust when you set or change the Interfaces and Zones settings.

- **Customized response pages** can be loaded by administrators to notify end users of policy violations.

With a combination of tools, expertise, and best practices, Palo Alto Networks helps analyze an organization's existing environment, migrate policies and firewall settings to the NGFW, and assist in all phases of the transition.

3.2.3 Network security management (Panorama)

Often, a data security breach occurs not due to a lack of information about a cyberattack, but rather a lack of appropriately prioritized, actionable information. Having actionable, well-organized information about network traffic and threats is more crucial today than ever before. IT and security teams are inundated with unmanageable and uncorrelated amounts of information from multiple, independent security solutions that don't fully integrate with other solutions and lack automation. This complexity makes it almost impossible to find critical threats buried deep in mountains of information. Both teams are simply too overwhelmed to find the proverbial "needle in the haystack" and are therefore unable to prioritize their responses appropriately. As a result, several operational gaps exist between where most organizations are and where they need to be with their network security. These operational gaps exist between:

- **Alert and action**. Network and security teams are often overwhelmed by the volume of data in security logs and are unable to easily determine which alerts are minor and which alerts are critical. Several cyberattacks in recent years (discussed in Section 1.1.6) demonstrate the impact of this first operational gap.

- **Known and unknown**. As the threat landscape grows increasingly complex, we are facing a growing number of unknown threats, and many security teams are struggling to keep pace. Discovering these threats quickly is crucial, but after they are discovered, security professionals must be able to quickly differentiate between the critical and the non-critical.

- **Idea and implementation.** Networks are growing fast and complexity is increasing. Many companies have huge numbers of policies, many of them outdated, because the complexity of provisioning and managing a secure network has become too overwhelming.

Closing these operational gaps requires reducing security management complexity and improving incident response, to enable rapid discovery of threats and quickly surface actionable intelligence.

Palo Alto Networks Panorama network security management reduces security management complexity with consolidated policy creation and centralized management features. The Application Command Center (ACC) in Panorama provides a customizable dashboard for setup and control of Palo Alto Networks NGFWs, with an efficient rule base and actionable insight into network-wide traffic and threats.

Panorama simplifies network security management with a single security rule base for firewall, threat prevention, URL filtering, application awareness, user identification, sandboxing, file blocking, and data filtering, to safely enable applications in the enterprise. Security rules easily can be imported, duplicated, or modified across the network. Centralized management of policies and objects provides consistent global security for the organization, while local administrative control provides flexibility at the local level.

Panorama centrally manages common device and network configurations through templates that can be used to push configuration changes to all managed firewalls. Templates eliminate manual, repetitive, risky, and error-prone configuration changes to multiple, individual firewalls deployed throughout the enterprise network. Templates can also be stacked and used as building blocks for streamlined device and network configuration (see Figure 3-15).

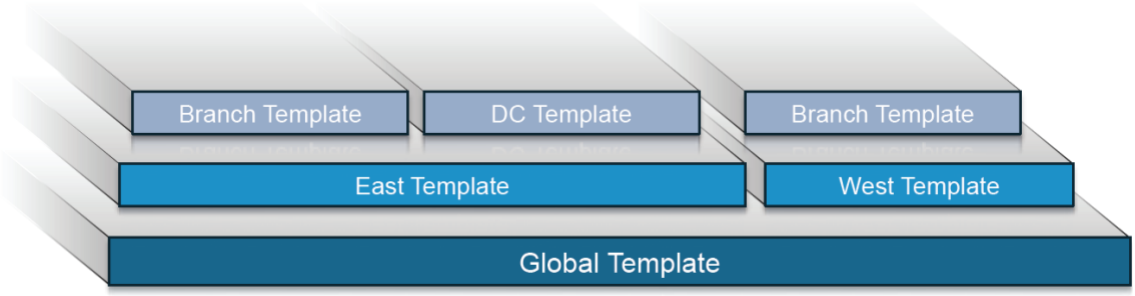

Figure 3-15: Template stacking in Panorama.

Panorama manages common policies and objects through hierarchical device groups (see Figure 3-16). Multilevel device groups are used to centrally manage the policies across all deployment locations with common requirements. Deploying hierarchical device groups ensures that lower-level groups inherit the settings of higher-level groups. This streamlines central management and enables you to organize devices based on function and location without redundant configuration.

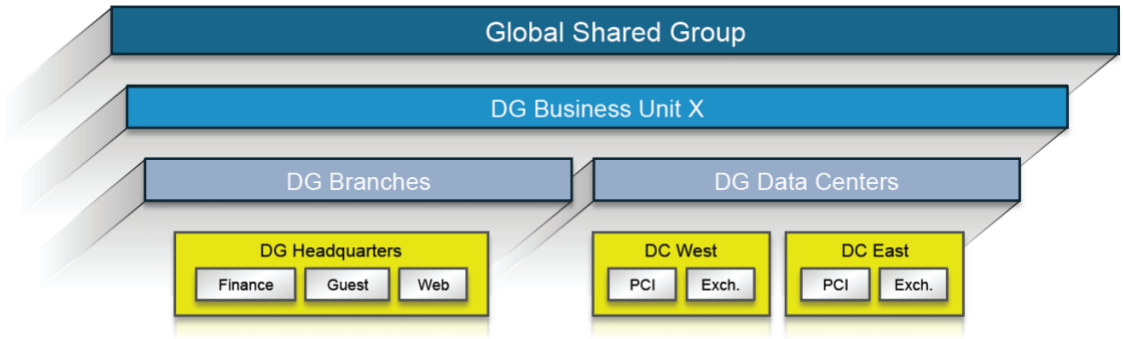

Figure 3-16: Hierarchical device groups in Panorama.

You can use shared policies for central control while still providing your local firewall administrator with the autonomy to make specific adjustments for local requirements. At the device group level, you can create shared policies that are defined as the first set of rules (pre-rules) and the last set of rules (post-rules) to be evaluated against match criteria. Pre- and post-rules can be viewed on a managed firewall, but they can be edited from Panorama only within the context of the administrative roles that have been defined. Local device rules (those between pre- and post-rules) can be edited by either a local firewall administrator or by a Panorama administrator who has switched to a local firewall context. In addition, an organization can use shared objects defined by a Panorama administrator, which can be referenced by locally managed device rules.

Panorama uses the same set of powerful monitoring and reporting tools available at the local device management level. As you perform log queries and generate reports, Panorama dynamically pulls the most current data directly from NGFWs under management or from logs forwarded to Panorama. Logging and reporting capabilities in Panorama include:

- **Log viewer.** For either an individual device or all devices, you can quickly view log activities using dynamic log filtering by clicking a cell value and/or using the expression builder to define the sort criteria. Results can be saved for future queries or exported for further analysis.

- **Custom reporting.** Predefined reports can be used as is, customized, or grouped together as one report to suit specific requirements.

- **User activity reports.** A user activity report shows the applications used, URL categories visited, websites visited, and all URLs visited over a specified period of time for individual users. Panorama builds the reports using an aggregate view of users' activity, no matter which firewall they are protected by, or which IP or device they may be using.

- **Log forwarding.** Panorama aggregates logs collected from all of your Palo Alto Networks firewalls, both physical and virtual form factor, and forwards them to a remote destination for purposes such as long-term storage, forensics, or compliance reporting. Panorama can forward all or selected logs, simple network management protocol (SNMP) traps, and email notifications to a remote logging destination, such as a Syslog server (over UDP, TCP, or SSL).

Panorama can be deployed in a centralized architecture with all Panorama management and logging functions consolidated into a single device, or in a distributed architecture with separate management units and Log Collectors in a hierarchical deployment architecture:

- **Panorama manager.** The Panorama manager is responsible for handling the tasks associated with policy and device configuration across all managed devices. The manager does not store log data locally, but rather uses separate Log Collectors for handling log data. The manager analyzes the data stored in the Log Collectors for centralized reporting.

- **Panorama Log Collector.** Organizations with high logging volume and retention requirements can deploy dedicated Panorama Log Collector devices that will aggregate log information from multiple managed firewalls.

Finally, Palo Alto Networks and Splunk have partnered to extend the powerful visibility into network traffic from Panorama to other network components. The combined solution delivers highly effective, coordinated detection, incident investigation, and response for cyberthreats. With the Splunk App for Palo Alto Networks (see Figure 3-17), enterprise security teams have a powerful platform for security visualization, monitoring, and analysis that enables them to fully leverage the extensive application, user, content, and threat data generated by Palo Alto Networks devices.

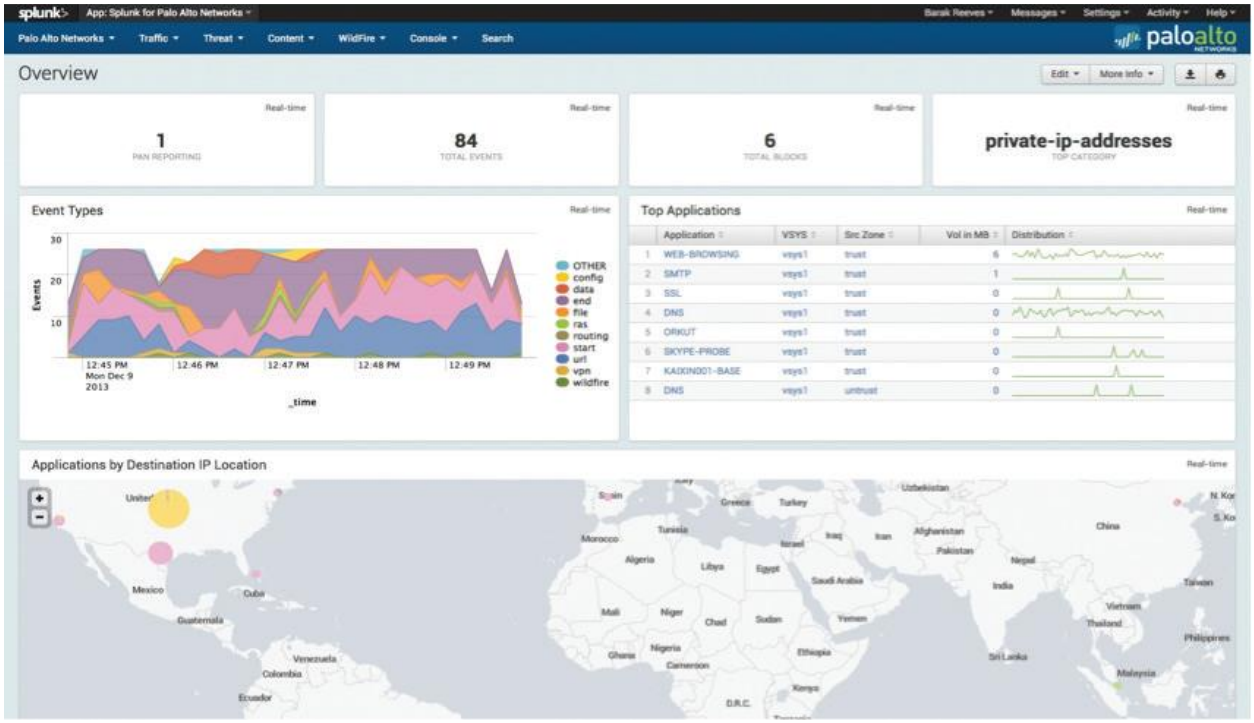

Figure 3-17: Integration with Splunk extends visibility and prevention capabilities to your entire network infrastructure.

The integrated solution not only combines several approaches for identifying cyberthreats — including dynamic sandbox analysis, statistical anomaly detection, and infrastructure-wide event correlation — but also enables security administrators to expedite incident response by automating the steps needed to block malicious sources and quarantine compromised devices.

3.2 Knowledge Check

Test your understanding of the fundamentals in the preceding section. Review the correct answers in the Appendix at the end of this guide.

1. **Multiple Choice.** Which option is NOT a defining characteristic of an NGFW? (Choose one.)

    a) low latency packet processing with minimal throughput loss
    b) adherence to strict port and protocol enforcement for allow/block decisions
    c) integrated security tools
    d) bidirectional full-stack analysis of packets

2. **Short Answer.** List the three core capabilities of a NGFW.

3. **Multiple Choice.** Which option is not a core technique for identifying applications in Palo Alto Networks NGFWs? (Choose one.)

    a) packet headers
    b) application signatures
    c) protocol decoding
    d) behavioral analysis

4. **Short Answer.** List three methods of mapping user identification to an IP address within a NGFW.

5. **Short Answer.** Describe stream-based malware scanning and explain its benefits.

6. **Short Answer.** What is the advantage of using templates in Panorama?

7. **Multiple Choice.** Panorama does not integrate with which option? (Choose one.)

    a) WildFire
    b) Splunk
    c) Palo Alto Networks NGFWs
    d) traditional port-based firewalls

3.3 Endpoint Protection

Endpoint protection components in the Security Operating Platform include Traps advanced endpoint protection and GlobalProtect mobile security.

3.3.1 Advanced endpoint protection (Traps)

Advanced endpoint protection is a new security product innovation that requires a different mindset from traditional security methodologies. Rather than a reactive "detect and respond" approach as with traditional anti-malware software, advanced endpoint protection employs a proactive prevention strategy. Advanced endpoint protection must do the following:

- Prevent all exploits, including those using unknown zero-day vulnerabilities
- Block all malware, without requiring any prior knowledge of specific malware signatures
- Provide detailed forensics against prevented attacks to strengthen all areas of the organization by pinpointing the targets and techniques used
- Be highly scalable and lightweight to seamlessly integrate into existing operations with minimal to no disruption
- Integrate closely with network and cloud security for quick data exchange and cross-organization protection

Palo Alto Networks Traps provides advanced endpoint protection that prevents sophisticated vulnerability exploits and malware-driven attacks, both known and unknown. The key to Traps is blocking core exploit and malware techniques, not the individual attacks. Traps automatically detects and blocks a core set of techniques that an attacker must link together to execute any type of attack, regardless of its complexity. Preventing just one technique in the Cyber Attack Lifecycle (see Section 1.2.2) is all that is needed to thwart the entire attack before it can do any damage.

The Traps agent injects itself into each process as it's started and automatically blocks advanced attacks that would otherwise evade detection. If an exploit attempt is made using one of the attack techniques, Traps immediately blocks that technique, terminates the process, and notifies the user and the security team that an attack was prevented (see Figure 3-18).

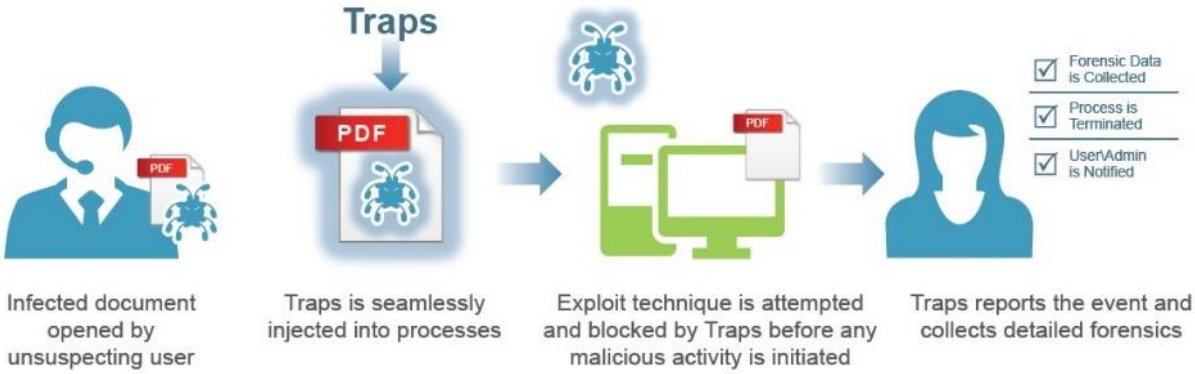

Figure 3-18: Traps blocks a core set of techniques to stop advanced attacks.

Throughout each event, Traps collects detailed forensics and reports this information to the Endpoint Security Manager (ESM), resulting in better visibility and an understanding of attacks that were prevented. With Traps, endpoints are always protected, regardless of patch, signature, or software-update levels; plus, it requires no prior knowledge of an attack to prevent it.

To prevent the execution of malicious executables on the endpoint, Traps focuses on three key areas to ensure comprehensive protection (see Figure 3-19). When combined, these methods offer unparalleled malware prevention and include the following:

- **Policy-based restrictions:** Organizations can easily set up policies restricting specific execution scenarios. For example, you may want to prevent the execution of files from the Outlook TMP directory, prevent execution of unsigned files, or prevent the execution of a particular file type directly from a USB drive.
- **WildFire inspection and analysis:** Traps queries Palo Alto Networks WildFire threat prevention cloud (discussed in Section 3.5.5) with a hash and submits any unknown .EXE files to assess their risk within the global threat community.
- **Malware techniques mitigation:** Traps implements technique-based mitigations that prevent attacks by blocking techniques such as thread injection.

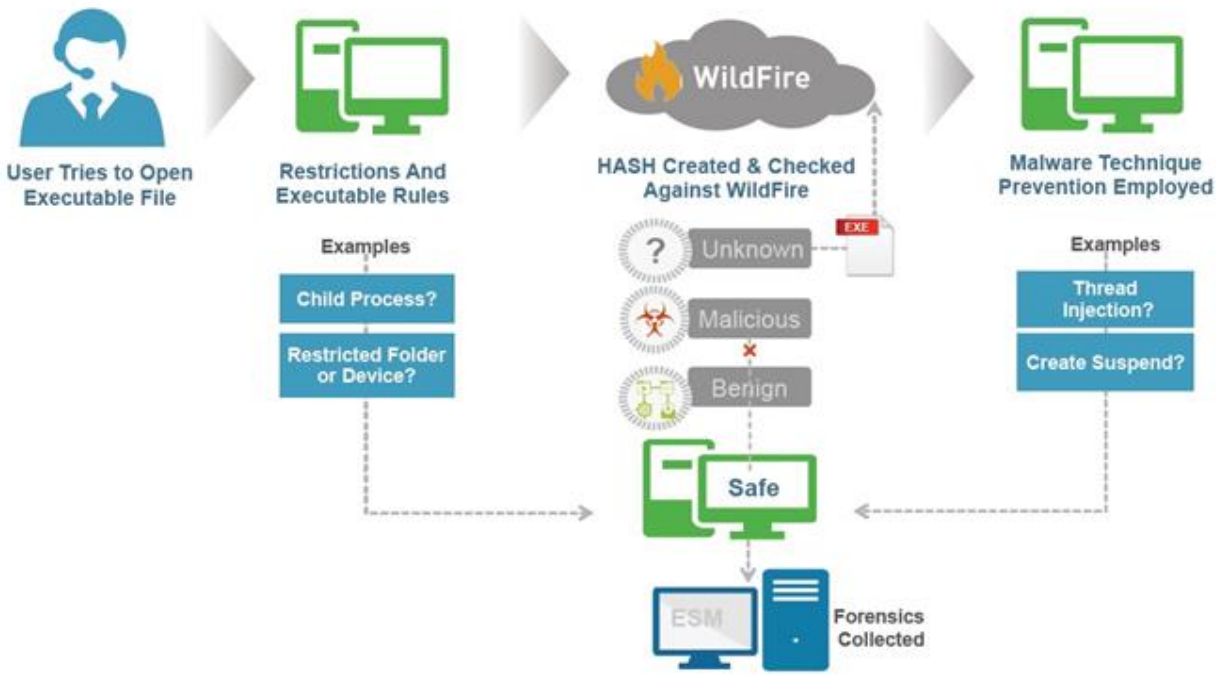

Figure 3-19: Prevention of malicious executables, a multi-tier approach.

3.3.1.1 Malware prevention

The Traps malware prevention engine uses advanced execution control, WildFire integration, and malware prevention modules (MPMs) to prevent the execution of malware.

When a user or endpoint attempts to open an executable, Traps first verifies that the executable doesn't violate any policy-based restrictions. Policy-based restrictions drastically reduce the attack surface by preventing file execution in high-risk scenarios. For example, you may want to prevent execution of the following:

- Particular (or any) file types directly from a USB drive
- Files from certain paths (such as the Outlook TMP folder) or network locations where applications don't reside
- Child processes created by specific applications (such as Microsoft PowerPoint)
- Unsigned executables or executables with an invalid certificate

Alternatively, highly granular restrictions are available to define trusted processes or file types, locations, and registry paths that these processes can read from and write to. If any of the

restriction rules apply to the executable, Traps blocks the file from executing and reports the security event to the ESM.

Traps provides static and dynamic execution control. Basic whitelisting and blacklisting of applications can be managed easily from the ESM console. Every executable that has ever been run in the organization is listed in the ESM console along with the WildFire verdict (discussed in Section 3.5.5). The administrator can easily override the verdict with the click of a button. For relatively static environments or specialized systems, like point-of-sale (POS) or supervisory control and data acquisition (SCADA), endpoints can be hardened with a strict execution-control policy. For more dynamic environments like end-user workstations, dynamic execution analysis and control is accomplished through integration with WildFire.

Traps Advanced Endpoint Protection is natively integrated with WildFire (discussed in Section 3.5.5) to provide zero-day protection against new and unknown exploits and malware threats. WildFire integration provides the capability to have the security of granular execution control and the manageability of a dynamic security policy driven by automated analysis of unknown executables (see Figure 3-20).

Figure 3-20: WildFire integration with Traps enables real-time evaluation of hash verdicts.

If an executable file has never been seen before on the endpoint, Traps can submit the file hash for immediate identification by WildFire. If WildFire identifies the file as malicious, Traps will prevent execution before any damage is done. With over one million samples analyzed each day, there is a good chance WildFire has seen the file and can alert Traps if it is malicious. If the

file hasn't been seen by WildFire, it can be automatically uploaded for rapid analysis to determine whether it's malicious. Traps and Palo Alto Networks NGFWs can submit files to WildFire, so this integration allows for seamless sharing of threat intelligence between NGFWs and the endpoints.

If a malicious file is not blocked by advanced execution control or WildFire evaluation and is allowed to execute, malicious activity can still be blocked by Traps malware prevention modules (MPM). MPMs focus on core techniques leveraged by many types of malware. For example, they will prevent malicious code from being injected into trusted applications.

A malware protection rule prevents the execution of malware, often disguised as or embedded in non-malicious files, by using malware modules to target common process behavior triggered by malware. You can enable injection of MPMs into all processes or enable protection into one or more protected processes in your organization. To allow legitimate processes to run, you can whitelist parent processes that inject into other processes. MPM rules include:

- **Suspend Guard:** Protects against a common malware technique where the attacker creates processes in a suspended state to inject and run code before the process starts. You can enable suspend guard on a source process mode, configure user notification, and optionally whitelist function modules that can call child processes.

- **Thread Injection:** Another common entry point for malicious code is through the creation of remote threads and processes. You can enable thread injection to stop remote thread and process creation and specify the limitation on either the source or destination process or thread. You can whitelist specific folders to make exceptions to the general execution restriction rule.

Traps prevents the execution of malicious files with a tailored approach to combating traditional and modern attacks (see Figure 3-21). Additionally, administrators can utilize periodic scanning to identify dormant threats, comply with regulatory requirements, and accelerate incident response with endpoint context.

- **WildFire threat intelligence:** In addition to third-party feeds, Traps leverages the intelligence obtained from tens of thousands of subscribers to the WildFire cloud-based threat analysis service to continuously aggregate threat data and maintain the collective immunity of all users across endpoints, networks, and cloud applications.

    - Traps queries WildFire with the hash of any Windows or macOS executable file, DLL, or Office file before the file runs to assess its standing within the global threat community. WildFire returns a near-instantaneous verdict on whether the

file is malicious or benign. If the file is unknown, Traps proceeds with additional prevention techniques to determine whether it is a threat that should be terminated.

    - If the file is deemed malicious, Traps automatically terminates the process and optionally quarantines it.

- **Local analysis via machine learning:** If a file remains unknown after the initial hash lookup and has not been identified by administrators, Traps uses local analysis via machine learning on the endpoint – trained by the rich threat intelligence of WildFire – to determine whether the file can run, even before receiving a verdict from the deeper WildFire inspection. By examining hundreds of file characteristics in real time, local analysis can determine whether a file is likely malicious or benign without relying on signatures, scanning, or behavioral analysis.

- **WildFire inspection and analysis:** In addition to local analysis, Traps sends unknown files to WildFire for discovery and deeper analysis to rapidly detect potentially unknown malware. WildFire brings together the benefits of independent techniques for high-fidelity and evasion-resistant discovery that go beyond legacy approaches. These techniques include:

    - **Static analysis via machine learning** – a more powerful version of local analysis, based in the cloud, that detects known threats by analyzing the characteristics of samples prior to execution.

    - **Dynamic analysis** – a custom-built, evasion-resistant virtual environment in which previously unknown submissions are detonated to determine real-world effects and behavior.

    - **Bare metal analysis** – a hardware-based analysis environment specifically designed for advanced threats that exhibit highly evasive characteristics and can detect virtual analysis.

If WildFire determines a file to be a threat, it automatically creates and shares a new prevention control with Traps and other components of Palo Alto Networks Next-Generation Security Platform in as few as five minutes. This control ensures that the threat is immediately classified as malicious and prevented if it is encountered again.

Additional prevention capabilities include:

- **Granular child process protection:** Traps prevents script-based and fileless attacks, by default, with out-of-the-box, fine-grained controls over the launching of legitimate applications, such as script engines and command shells. The number of available controls continue to grow with regular content updates from the Palo Alto Networks threat research team, Unit 42. Administrators have the ability to whitelist or blacklist child processes, and command-line comparisons help to increase detection without negatively impacting process performance or shutting processes down.
- **Behavior-based ransomware protection:** In addition to existing multi-method prevention measures, including exploit prevention, local analysis and WildFire, Traps monitors the system for ransomware behavior. Upon detection, it immediately blocks attacks and prevents encryption of customer data.
- **Scanning:** Administrators can scan endpoints and attached removable drives for dormant malware, with an option to automatically quarantine it for remediation when found. Periodic or on-demand scanning can be configured as part of a security profile on one or more endpoints.
- **Admin override policies:** Traps enables organizations to define policies based on the hash of an executable file to control what is or isn't allowed to run in their environments. This capability reduces the attack surface and eliminates the negative impact on homegrown or heavily customized applications.
- **Malware quarantine:** The quarantine is particularly useful in preventing the inadvertent dissemination of malware in organizations where network- or cloud-based data storage and SaaS applications automatically sync files across multiple users and systems. Traps immediately quarantines malicious executable files, DLLs, and Office files to prevent propagation or execution attempts of infected files.
- **Grayware classification:** Traps enables organizations to identify non-malicious but otherwise undesirable software, such as adware, and prevent it from running in their environments.
- **Execution restrictions:** Traps enables organizations to easily define policies to restrict specific execution scenarios to reduce the attack surface of any environment. For example, Traps can prevent the execution of files from the Outlook temp directory or a particular file type from a USB drive.

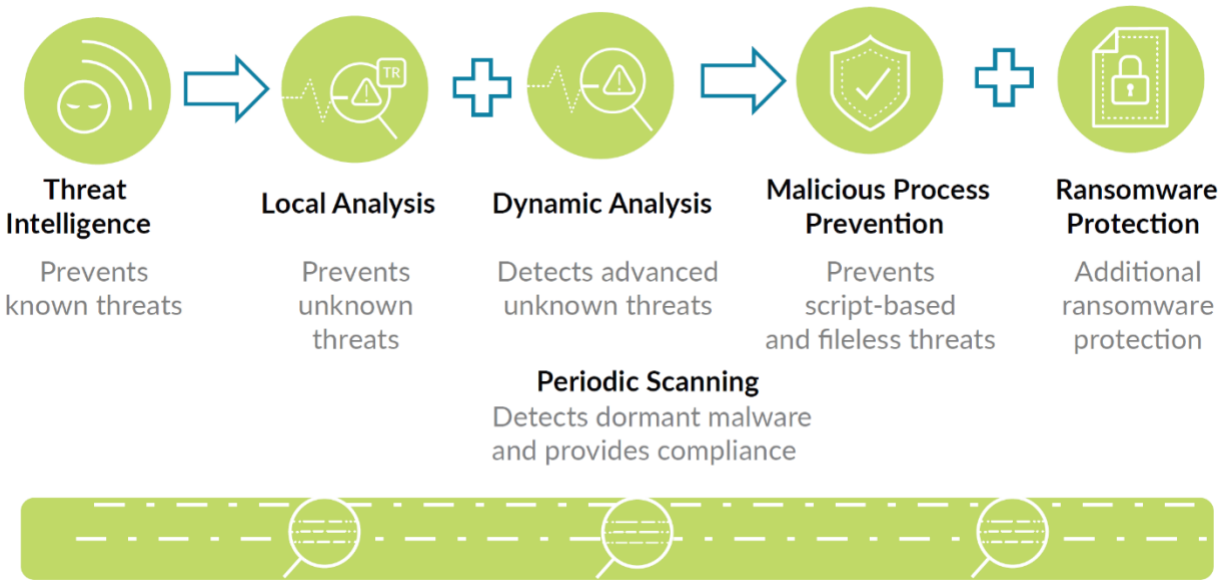

Figure 3-21: Traps multi-method malware prevention.

3.3.1.2 Exploit prevention

Traps focuses on the core techniques used by all exploits to render those techniques ineffective, which means the application is no longer vulnerable.

The Traps agent injects itself into each process as it is started. If the process attempts to execute any of the core attack techniques, the corresponding exploit prevention module (EPM) prevents that exploit by killing the process, and it reports all of the details to the ESM as depicted in Figure 3-22.

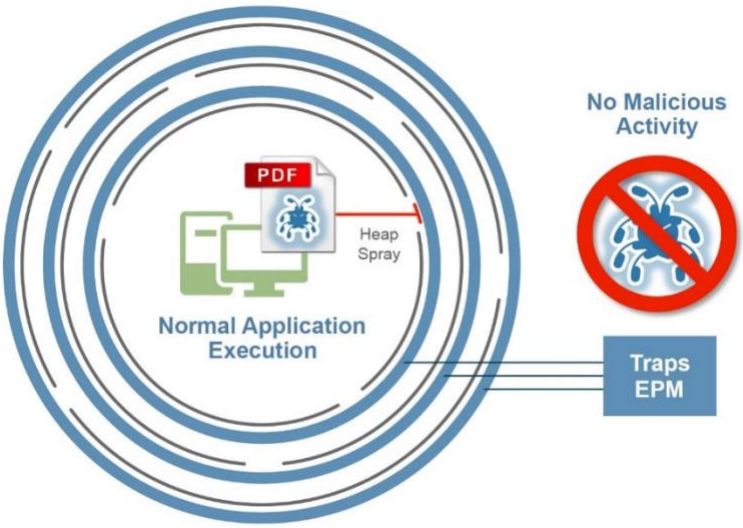

Figure 3-22: Traps EPMs protect application processes against vulnerabilities.

By default, Traps policy is configured to protect over 100 processes — each one with dozens of proprietary EPMs. Traps isn't limited to protecting only those processes or applications. Organizations use Traps to protect all manner of processes and applications by simply adding them to the policy configuration. Processes that have been run on the endpoint automatically show up in the ESM console, which makes it easy to protect those processes with the click of a button. This capability is especially useful for organizations running industry-specific applications, such as point-of-sale (POS) systems, ATM terminals, and supervisory control and data acquisition (SCADA).

If for some reason an application conflicts with one of the EPMs, security administrators can simply disable that EPM for the specific application and endpoint. The application is still protected by dozens of other EPMs (see Figure 3-23). Exploits rely on a series of techniques to successfully run, so the other EPMs continue to protect that application and block at least one of the techniques, which breaks the sequence.

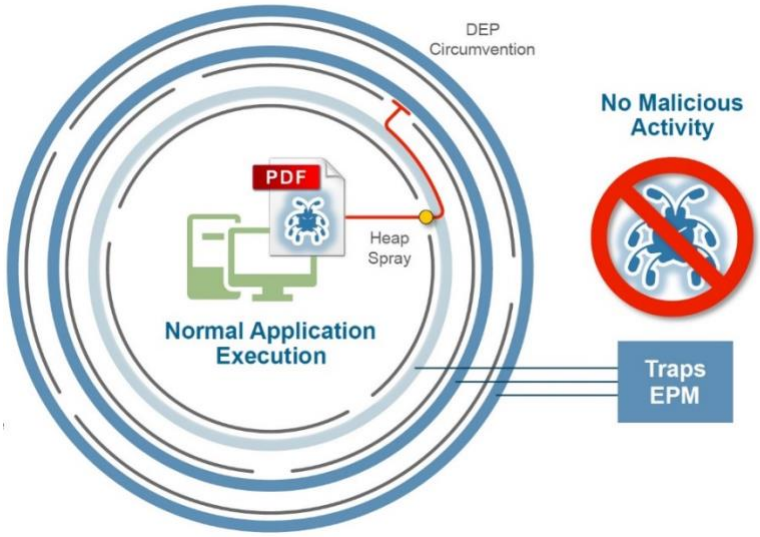

Figure 3-23: Only one technique needs to be blocked for an exploit to fail.

Examples of attacks that the EPMs can prevent include:

- Dynamic link library (DLL) hijacking — replacing a legitimate DLL with a malicious one of the same name
- Hijacking program control flow
- Inserting malicious code as an exception handler

Rather than relying on signatures or behavior-based detection to identify exploit-based attacks, Traps takes the unique approach of targeting the limited number of techniques – the tools, if you will – any exploit-based attack must use to manipulate a software vulnerability. By preventing the techniques instead of identifying each individual attack, Traps protects unpatched systems, unsupported legacy systems, applications IT is unaware of, and never-before-seen exploits – also called zero-day exploits. Traps delivers exploit prevention using multiple methods:

- **Pre-exploit protection:** Traps prevents the vulnerability-profiling techniques exploit kits use prior to launching attacks. By blocking these techniques, Traps prevents attackers from targeting vulnerable endpoints and applications, effectively preventing the attacks before they begin.

- **Technique-based exploit prevention:** Traps prevents known, zero-day and unpatched vulnerabilities by blocking the exploitation techniques that attackers use to manipulate applications. Although there are thousands of exploits, they typically rely on a small set of exploitation techniques that change infrequently. Traps blocks these techniques, which prevents exploitation attempts before they can compromise endpoints.

- **Kernel exploit prevention:** Traps prevents exploits that leverage vulnerabilities in the operating system kernel to create processes with escalated (that is, system-level) privileges. Traps also protects against new exploit techniques used to execute malicious payloads, such as those seen in 2017's WannaCry and NotPetya attacks. By blocking processes from accessing the injected malicious code from the kernel, Traps can prevent the attack early in the attack lifecycle without affecting legitimate processes. This capability enables Traps to block advanced attacks that target or stem from the operating system itself.

By blocking the techniques common to all exploit-based attacks, Traps provides three important benefits:

- **Protects applications that cannot be patched and shadow IT applications:** Providing a positive work experience is critical to the productivity of any organization, but running unsupported legacy applications or granting users the flexibility to download and run programs as they please introduces risk. Traps enables organizations to run any applications, including those developed in-house, no longer receiving updates or security support, or running in their environment without IT's awareness, without opening the network to the threat of exploit-based attacks.

- **Eliminates the urgency to patch applications as soon as possible:** Organizations using Traps can apply security patches when it is appropriate for the business and after sufficient testing. Traps prevents the exploitation of application vulnerabilities regardless of when an organization applies security patches issued by application vendors.

- **Prevents zero-day exploits from succeeding:** Traps blocks the limited set of exploitation techniques zero-day exploits typically use, so Traps protects organizations against attacks that utilize zero-day exploits.

3.3.1.3 Traps deployment architecture

Traps is a highly scalable advanced endpoint protection solution that consists of an Endpoint Security Manager (ESM) Console, Endpoint Security Manager Server(s), lightweight Traps agents (installed on individual endpoints), and optional external logging.

The Traps infrastructure supports various architectural options to allow for scalability to a large distributed environment. Installation of the ESM creates a database on Microsoft SQL Server and installs the administrative console within Internet Information Server (IIS).

ESM servers essentially act as proxies between Traps agents and the ESM database. Communications from Traps agents to ESM servers occur over HTTPS. ESM servers don't store data and therefore can be easily added and removed from the environment as needed to ensure adequate geographic coverage and redundancy.

To ensure global connectivity, organizations that don't use a mobility solution like Palo Alto Networks GlobalProtect (discussed in Section 3.3.2) may opt to put an ESM server in the DMZ or in a cloud-based environment with external connectivity.

The Traps agent installer can be deployed using your software deployment tool of choice. Subsequent updates to the agent can be deployed via the ESM. The agent consumes less than 25 MB on disk and less than 40 MB while running in memory. Observed CPU use is less than 0.1 percent. The agent also employs various tamper-proofing methods that prevent users and malicious code from disabling protection or tampering with agent configuration.

The lightweight structure allows for the Traps environment to scale horizontally and support large deployments of up to 50,000 agents per ESM, while still maintaining a centralized configuration and database for policies. Traps can coexist with most major endpoint security solutions, and the CPU use and I/O remains incredibly low. With minimal disruption, Traps is optimal for critical infrastructures, specialized systems, and virtual desktop infrastructure (VDI) environments.

The ESM can write logs to an external logging platform, such as an SIEM solution or any system that supports syslog, in addition to storing its logs internally.

3.3.1.4 Traps in action

To understand how Traps prevents an attack from succeeding, take a look at an actual cyberattack example in which a PDF file with an embedded exploit is sent to an unsuspecting user (see Figure 3-24). The user opens the PDF file, which does the following:

- Exploits Adobe Reader
- Causes Adobe Reader to create an Internet Explorer (IE) child process
- Causes IE to download an executable (EXE) file from a malicious website
- Executes the new EXE file, which then performs malicious activities on the endpoint, including thread injection into IE

This chain of events is common in many attacks. The specific file type, exploit, and malicious executable payload may vary, but the steps are largely the same from one attack to another.

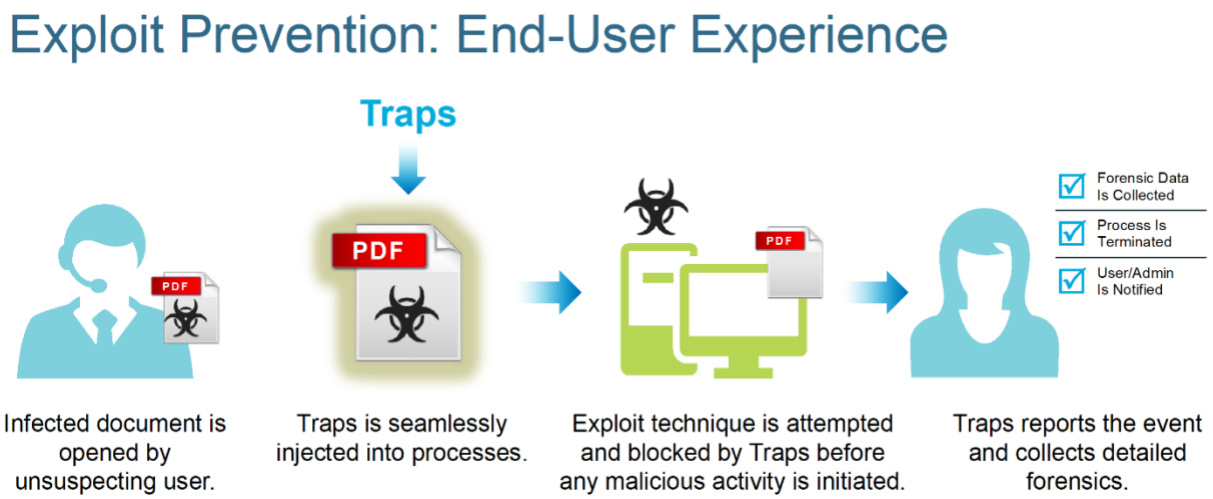

Figure 3-24: When an exploit is attempted, Traps blocks it before any malicious activity is initiated.

The key to stopping an attack is to break this chain of events at the earliest possible stage of the attack.

To prevent an attack from succeeding, Traps provides prevention capabilities and multiple layers of protection to block the attacker's ability to access the network and move laterally within the enterprise. In this particular attack example, Traps can prevent the attack from succeeding by using any of the following techniques (see Figure 3-25):

1. **Exploit Technique 1.** The exploit uses a series of techniques in order to take advantage of the vulnerability in the targeted application, Adobe Reader in this case. Although the exploit could be a new zero-day threat, the techniques it has to use are common and new techniques are very rare (typically two to four per year). In this example, the exploit uses several operating system (OS) functions.

2. **Exploit Technique 2.** In this example, just-in-time (JIT) spraying is used to exploit a just-in-time compiler, which sets up the third exploit technique (heap spraying) to be used in the attack. This type of exploit is commonly used against PDF file formats and Adobe Flash Player. Again, Traps prevents the exploit from executing so that even if the first exploit technique for some reason succeeds, the second exploit fails and the attack is thwarted.

3. **Exploit Technique 3.** In this example, heap spraying is used next to facilitate arbitrary code execution. This technique allows the attacker to place a byte sequence in the memory of a target process. Heap spraying is another commonly used exploit technique that Traps prevents from executing.

4. **Execution Restriction 1.** In this example, Adobe Reader creates a child process (a technique commonly used to avoid anti-malware detection). Traps restricts child processes from executing arbitrarily and thus prevents the attack from succeeding.

5. **Execution Restriction 2.** In this example, the attacker attempts to run an unsigned executable. Here again, Traps prevents the executable from running, based on rules that can be customized by an administrator.

6. **Execution Restriction 3.** In this example, an executable program attempts to run from a restricted location, the IE TMP folder. These restricted locations can be customized by an administrator if needed.

7. **Local Verdict Check.** A local verdict check compares the file against an administrator-configured blacklist to determine whether the file is explicitly blocked, or against a whitelist to determine if the file has been explicitly allowed regardless of its WildFire verdict.

8. **WildFire Known Verdict.** Traps EPM checks the file against WildFire by sending the file hash. In this example, WildFire responds that the file is known to be malicious and therefore is not allowed to execute.

9. **WildFire On-Demand Inspection.** If WildFire has never seen the file, it can be uploaded for analysis and not allowed to run until WildFire provides a verdict.

10. **Malware Prevention Module.** If the malicious executable is allowed to run, it attempts a thread injection into IE. This malware technique is blocked by the Thread Injection malware prevention module in Traps.

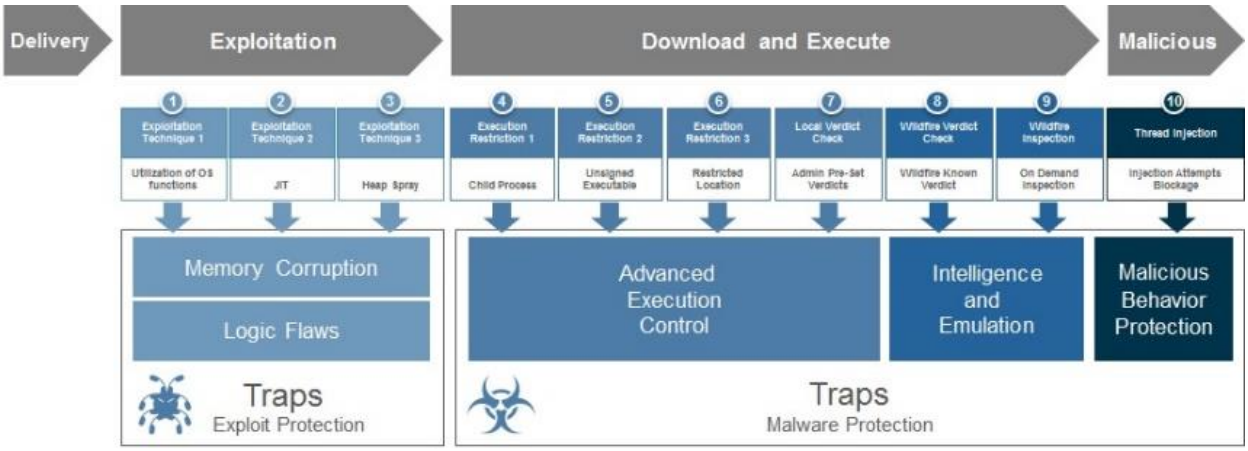

Figure 3-25: Traps prevents this attack example at any one of ten "kill points."

Key Terms

In multitasking operating systems, a *child process* is a subprocess created by a parent process that is currently running on the system.

While this is just one example, most modern attacks use some combination of these steps and various exploit and malware techniques. Whereas most endpoint protection approaches focus on one blocking method (whitelisting, for example), Traps takes advantage of every opportunity to prevent compromise. Any one of these "kill points" is enough to prevent the attack.

3.3.2 Mobile security and VPN management (GlobalProtect)

Mobile computing is revolutionizing how and where employees work, and the tools that they use to perform their jobs. As enterprise mobile strategies mature, mobile capabilities become more advanced with new applications and greater access to data, opening the door to new opportunities — and new risks. For organizations to adopt more sophisticated uses of mobile devices, enterprise security teams must ensure that they address concerns about the inherent risks to sensitive information and network assets that mobility brings.

Unfortunately, many traditional mobile security tools tend to focus on very basic use cases and may be as limited in their security capabilities as the use cases themselves. The path to unlocking the full value of the mobile device depends on security, which provides the means to

extend applications safely. Security should be seen as a way to enable mobile initiatives rather than a limitation to mobile strategies. To fully realize all of mobility's benefits and safely enable mobile devices, enterprises must:

- **Manage the device.** Ensure that mobile devices are safely enabled by configuring the device with proper security settings. Simplify deployment and setup by provisioning common configurations like account setting for email and credentials such as certificates.
- **Protect the device.** Protect the mobile device from exploits and malware. Protecting the device also plays an important role for protecting the data as well, because data is not safe on a compromised device.
- **Control the data.** Control access to data and control the movement of data between applications. Establish policies that define who can access sensitive applications, and the particular devices that can be used.

Palo Alto Networks GlobalProtect safely enables mobile devices for business use by providing a unique solution to manage the device, protect the device, and control the data. It blends together the necessary technology and intelligence to provide a comprehensive solution for mobile security. This solution enables the organization to stop mobile threats, enforce security policies, and protect networks from compromised and non-compliant mobile devices.

GlobalProtect has three primary components:

- **GlobalProtect Gateway:** Delivers mobile threat prevention and policy enforcement based on apps, users, content, device, and device state. GlobalProtect gateways provide security enforcement for traffic from GlobalProtect agents and apps. Additionally, if the host information profile (HIP) feature is enabled, the gateway generates a HIP report from the raw host data the clients submit and can use this information in policy enforcement. GlobalProtect gateways are configured on an interface on any Palo Alto Networks NGFW. You can run a gateway and a portal on the same firewall, or you can have multiple, distributed gateways throughout the enterprise. There are two types of GlobalProtect gateways:
    - **External gateways:** Provide security enforcement and/or virtual private network (VPN) access for remote users.
    - **Internal gateways:** An interface on the internal network configured as a GlobalProtect gateway for applying security policy for access to internal

resources. When used in conjunction with User-ID (discussed in Section 3.2.1.2) and/or HIP checks, an internal gateway can be used to provide a secure, accurate method of identifying and controlling traffic by user and/or device state. Internal gateways are useful in sensitive environments where authenticated access to critical resources is required. You can configure an internal gateway in either tunnel mode or non-tunnel mode.

- **GlobalProtect Client:** Enables device management, provides device state information, and establishes secure connectivity. Extends a VPN tunnel to Apple iOS, Android, and Windows 10 (and Universal Windows Platform) mobile devices with GlobalProtect App, and Windows, Mac, and Google Chrome operating systems with GlobalProtect Agent. Connects to the GlobalProtect Gateway to access applications and data in accordance with policy. GlobalProtect client software runs on endpoints and enables access to network resources via the GlobalProtect portals and gateways that have been deployed. There are two types of GlobalProtect clients:
    - **GlobalProtect Agent** runs on Windows, Mac, and Chrome operating systems and is deployed from the GlobalProtect portal. You configure the behavior of the agent – for example, which tabs the users can see and whether or not users can uninstall the agent – in the client configuration(s) you define on the portal.
    - **GlobalProtect App** runs on Apple iOS, Android, and Windows 10 mobile devices and establishes a device-level VPN connection to the GlobalProtect Gateway to protect traffic and enforce security policies. GlobalProtect App can automatically select the optimal gateway for a given location to provide a transparent user experience for security. On Apple iOS devices, GlobalProtect App can be configured for app-level VPN.
- **GlobalProtect Portal:** Directs all client traffic to the appropriate gateway and is accessed first by the client device. The GlobalProtect portal provides the management functions for the GlobalProtect infrastructure. Every client system that participates in the GlobalProtect network receives configuration information from the portal, including information about available gateways as well as any client certificates that may be required to connect to the GlobalProtect gateway(s). In addition, the portal controls the behavior and distribution of the GlobalProtect client. If you are using the HIP feature, the portal also defines what information to collect from the host, including any custom information you require. The GlobalProtect Portal can be configured on an interface on any Palo Alto Networks NGFW.

Figure 3-26 illustrates how the GlobalProtect portals, gateways, and agents/apps work together to enable secure access for all your users, regardless of what devices they are using or where they are located.

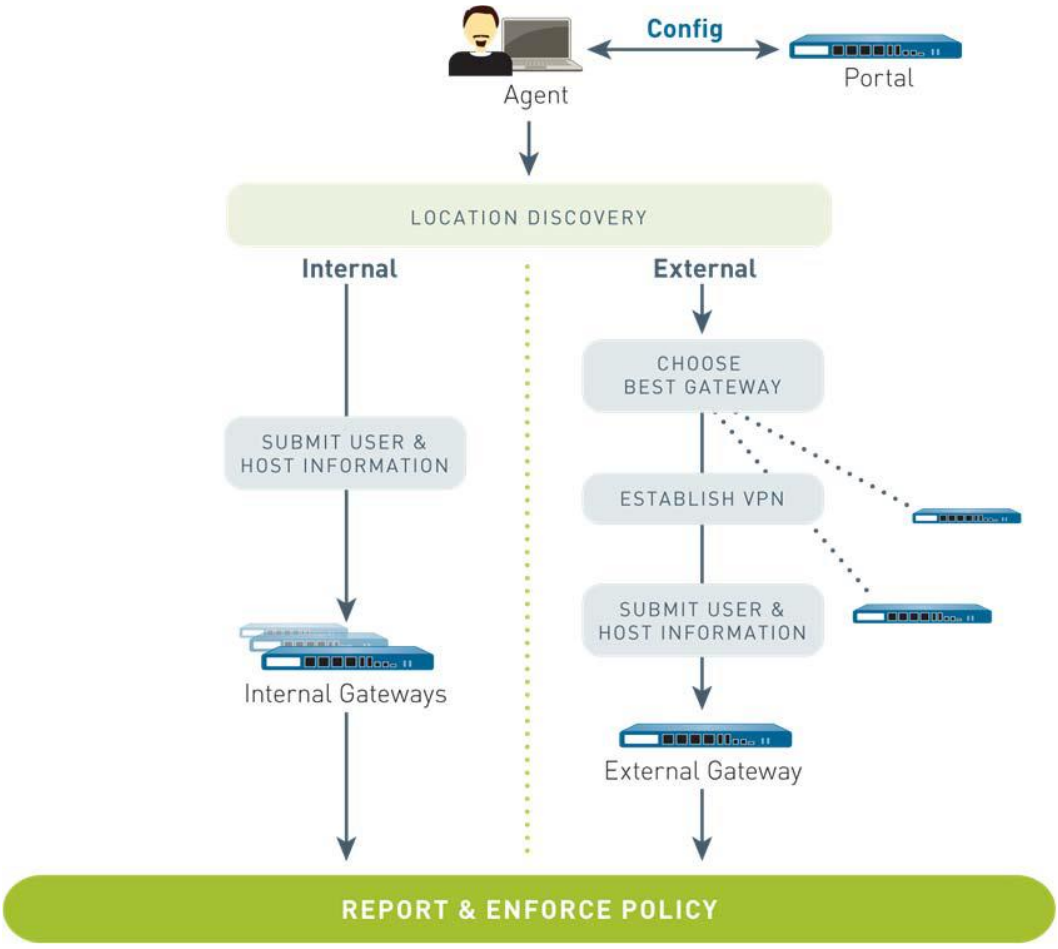

Figure 3-26: GlobalProtect components work together to secure access for all users in the enterprise, regardless of location or device.

GlobalProtect also provides a complete infrastructure for managing secure access to enterprise resources from remote sites. The GlobalProtect Large Scale VPN (LSVPN) feature on Palo Alto Networks NGFWs simplifies the deployment of traditional hub and spoke VPNs. The LSVPN feature enables security teams to quickly extend enterprise networks to multiple branch offices with a minimum amount of configuration required on the remote satellite devices. LSVPN uses certificates for device authentication and IPsec to secure data. The LSVPN infrastructure consists of the following components (see Figure 3-27):

- **GlobalProtect Portal**: Provides the management functions for the GlobalProtect LSVPN infrastructure. Every satellite that participates in the GlobalProtect LSVPN receives

configuration information from the portal, including configuration information to enable the satellites (the spokes) to connect to the gateways (the hubs). The portal can be configured on an interface on any Palo Alto Networks NGFW.

- **GlobalProtect Gateways:** A Palo Alto Networks NGFW that provides the tunnel end point for satellite connections. The resources that the satellites access are protected by security policy on the gateway. A separate portal and gateway is not required; a single firewall can function as portal and gateway.
- **GlobalProtect Satellite:** A Palo Alto Networks NGFW at a remote site that establishes IPsec tunnels with the gateway(s) at the corporate office(s) for secure access to centralized resources. Configuration on the satellite firewall is minimal, enabling security teams to quickly and easily scale the VPN as new sites are added.

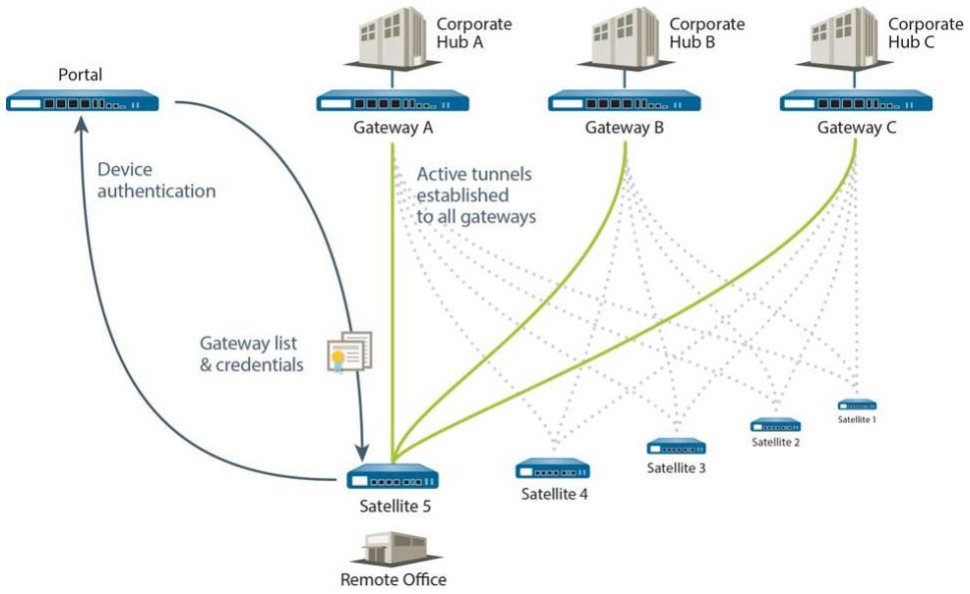

Figure 3-27: The GlobalProtect LSVPN components work together to securely extend an enterprise network to remote offices.

- **GlobalProtect cloud service** is a cloud-based security infrastructure service that simplifies the process of scaling the Palo Alto Networks next-generation security platform so that organizations can extend the same best-in-breed security to remote network locations and mobile users without having to build out their own global security infrastructure and expand their operational capacity. With the GlobalProtect cloud service, Palo Alto Networks automatically deploys NGFWs and GlobalProtect portals and gateways in the locations where the organization needs them (see Figure 3-28).

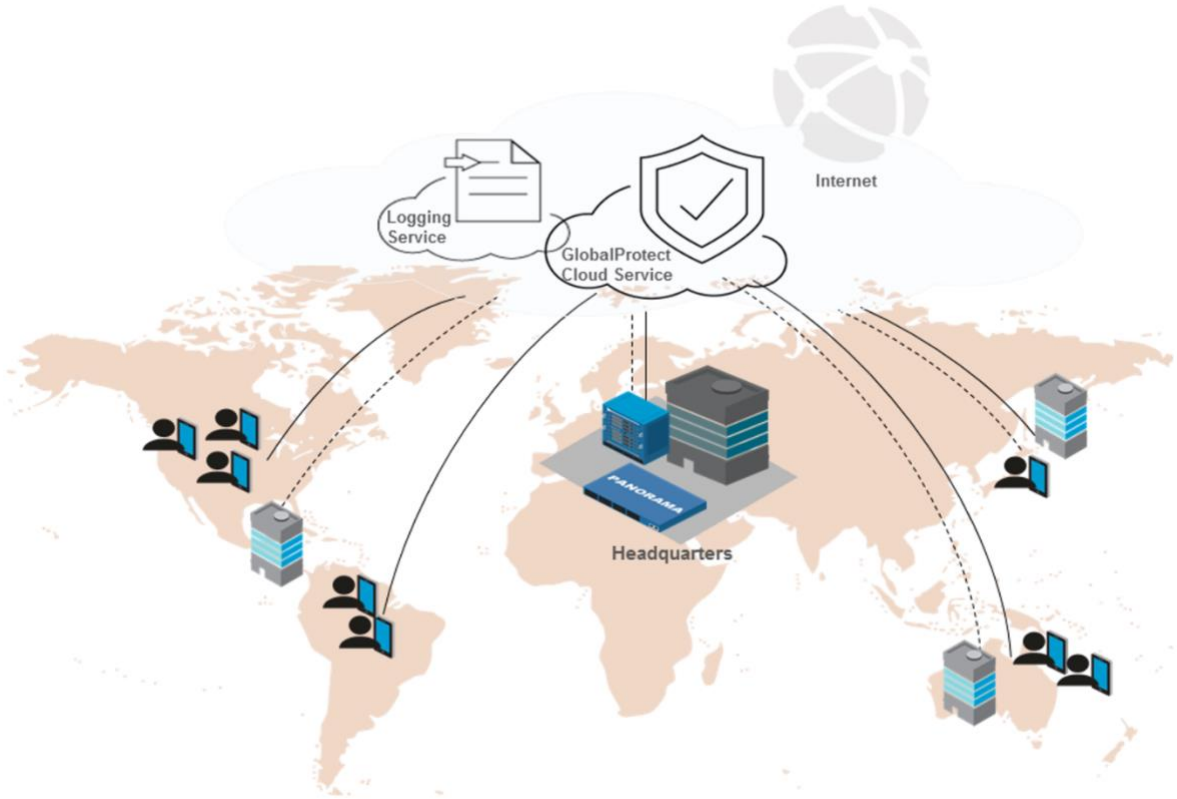

Figure 3-28: GlobalProtect cloud service.

With the GlobalProtect cloud service, Palo Alto Networks deploys and manages the security infrastructure globally to secure your remote networks and mobile users. The GlobalProtect cloud service is comprised of five components:

- **Cloud Services Plugin**—Panorama (discussed in Section 3.2.3) plugin that enables the GlobalProtect cloud service and the Logging Service. This plugin provides a simple and familiar interface for viewing the status of the service, and it configures the settings to begin directing traffic from your remote network locations and mobile users to the cloud service. To enable you to quickly enforce consistent security policy across all locations, you can leverage the Panorama templates and device groups you may have already created to push configurations to the firewalls, portals, and gateways in the GlobalProtect cloud service.
- **Service Infrastructure**—For the GlobalProtect cloud service to create an infrastructure in the cloud for your remote network locations and mobile users, you must supply a subnet that does not overlap with other IP addresses you use internally. The

GlobalProtect cloud service uses the IP addresses within this subnet to establish a network infrastructure between your remote network locations and mobile users and service connections to your headquarters and/or data center (if applicable). Internal communication within the cloud is established using dynamic routing.

- **Service Connections**—A GlobalProtect cloud service license includes the option to establish IPsec tunnels (discussed in Section 2.6.4) to up to three of your headquarters or data center sites. This service is optional and enables the GlobalProtect cloud service to connect to your authentication servers and give your mobile users and remote network users access to corporate resources. To set up a service connection, you must set up an IPSec tunnel from each HQ/data center location to the GlobalProtect cloud service. You then set up routing to enable traffic to and from the tunnel to the subnetworks that contain the resources to which your remote network and mobile users need access. All GlobalProtect gateways can then connect to the service connection firewall in a hub-and-spoke architecture to provide access to the internal networks in your GlobalProtect cloud service infrastructure.

- **Remote Networks**—The GlobalProtect cloud service for remote networks automatically deploys NGFWs in the regions you specify in the Cloud Services plugin during the onboarding steps. You will need an IPsec-compliant firewall, router, or software-defined WAN (SD-WAN) device that can establish a tunnel to the GlobalProtect cloud service for remote networks, and you must route traffic from users at the remote network location through the IPsec tunnel so that the policy you have pushed to the service can be enforced by the cloud service. You can enable access to the subnetworks at each remote network location using either static routes, dynamic routing using the border gateway protocol (BGP, discussed in Section 2.1.3), or a combination of static and dynamic routes. All remote network locations that you onboard are fully meshed.

- **Mobile Users**—The GlobalProtect cloud service for mobile users automatically deploys GlobalProtect portals and gateways in the cloud. Mobile users then connect to the GlobalProtect cloud service for mobile users to receive their VPN configuration, which routes them to the closest GlobalProtect cloud service gateway for policy enforcement. To configure this service, you must designate an IP address pool for the service to use to assign IP addresses for the client VPN tunnels. The addresses in this pool must not overlap with other address pools you use internally or pools you assign for the Service Connections.

Additionally, the cloud firewalls, gateways, and portals that are deployed as part of the GlobalProtect cloud service infrastructure must forward all logs to the Logging Service. You can

then view the logs, Application Command Center (ACC), and reports from Panorama for an aggregated view into your remote network and mobile user traffic.

All of the GlobalProtect cloud service firewalls deployed for your organization are fault tolerant. All of the cloud firewalls deployed to secure your remote network locations and enable service connections are in a high availability configuration, with state synchronization across multiple availability zones. In addition, if you configure a backup WAN link, tunnel failover time is less than 10 seconds from the time of detection (depending on your Internet provider). To ensure availability for mobile users, the GlobalProtect cloud service deploys multiple firewalls in all regions to enable reliable, global coverage. Failover between a primary gateway and a backup gateway is less than 20 seconds. The service is secure, resilient, up to date, and available to you when you need it, so you can focus on defining polices that meet your corporate usage guidelines for consistent policy enforcement.

3.3 Knowledge Check

Test your understanding of the fundamentals in the preceding section. Review the correct answers in the Appendix at the end of this guide.

1. **True or False.** The key to Traps is blocking core exploit and malware techniques, not the individual attacks.
2. **Short Answer.** Describe the basic function of Traps exploit prevention modules (EPMs).
3. **Multiple Answer.** What are the three keys to safely enabling mobile devices in the enterprise?

3.4 Cloud Security

Cloud security components in the Security Operating Platform include Evident cloud monitoring and compliance and Aperture SaaS security.

3.4.1 Cloud monitoring and compliance (Evident)

Driven by the flexibility and ease of public clouds, Agile development and the DevOps movement have accelerated the speed of application development cycles. Security teams can no longer depend on pre-deployment scanning, penetration tests, or presence-based discovery methods. To get the visibility they need they require automated, API-based tools that can handle the volumes of data produced in the cloud. The key for today's enterprise is to remove

the human element from repeatable processes and tasks so teams can better protect their cloud environments. Human errors can become amplified quickly and create major risks. Cloud environments are continuously changing and connecting with more services, so errors are bound to occur: private keys inadvertently are made public, ports are left open, and stored data gets exposed.

Evident was developed specifically to help modern IT, DevOps, and risk and compliance teams implement and maintain security within the cloud shared responsibility model (discussed in Sections 1.1.3 and 2.8.1). Evident provides public cloud infrastructure services security that enables organizations to automate the management of cloud security and compliance risks, so they can minimize the attack surface and protect their public cloud deployments.

Evident provides continuous monitoring of public clouds, which enables organizations to deploy applications confidently, knowing the cloud environment is securely configured. With continuous monitoring enabled, Evident also helps organizations achieve a continuous state of compliance by analyzing the configurations of all the services and account settings against strict security and compliance controls. Data stored within cloud storage services is classified and checked for data exposure and malware.

The capabilities and benefits of Evident include:

- **Continuous visibility and monitoring.** Evident provides security and compliance teams with a view into the risks across all their cloud accounts, services, and regions by automating monitoring, inspection, and assessment of the organization's cloud infrastructure services. With real-time visibility into the security posture of their environment, security and compliance teams can be alerted on issues that do not comply with the organization's required controls and settings.

- **Compliance validation.** Taking a security-first approach to compliance helps organizations go beyond compliance requirements and adopt best practices that keep their environments and data secure. Evident simplifies measurement and reporting of compliance with prebuilt, one-click compliance reports for Center for Internet Security (CIS) Foundations, General Data Protection Regulation (GDPR), Health Insurance Portability and Accountability Act (HIPAA), U.S. National Institute of Standards and Technology (NIST), Payment Card Industry (PCI) and Service Organization Control (SOC) 2, and allows users to create custom reports to measure specific organizational goals.

- **Securing cloud storage services.** Evident helps identify and classify data stored in Amazon S3 buckets, Microsoft Azure Blob Storage, and Google Cloud Storage. Powered by machine learning, Evident provides awareness of the type of data in these services so

organizations can identify data exposure risks and automatically remediate policy violations as soon as they occur.

- **Automated remediation.** Evident enables automated remediation to enforce policies as defined by the organization quickly. Risks can be addressed quickly, and necessary changes can be made to configurations and settings without manual intervention, which gets the environment back to a compliant state faster.

The Evident automated approach to securing public cloud workloads incorporates three critical security components – continuous monitoring, compliance validation, and secure cloud storage. It is fully customizable and can be adapted to identify and alert enterprises about risks and vulnerabilities that are specific to their data and usage policies. The Evident API-based approach allows all three security components to be embedded directly into the application development process without compromising on agility.

3.4.2 SaaS security (Aperture)

To safely enable SaaS usage in your organization, start by clearly defining the SaaS applications that should be used and which behaviors within those applications are allowed. This step requires a clear definition of which applications are:

- **Sanctioned** (allowed and provided by IT),
- **Tolerated** (allowed because of a legitimate business need, with restrictions, but not provided by IT), and
- **Unsanctioned** (not allowed), then controlling their usage with granular policies.

Sanctioned SaaS applications provide business benefits and are fast to deploy, require minimal cost, and are infinitely scalable. Tolerated SaaS applications fulfill a legitimate business need, but certain usage restrictions may be necessary to reduce risk. Unsanctioned SaaS applications either clearly provide no business benefits, or the security risks of the application outweigh the business benefits. For example, an unsanctioned SaaS application may violate regulatory compliance mandates, create an unacceptable risk of loss of corporate intellectual property or other sensitive data, or enable malware distribution (see Figure 3-29).

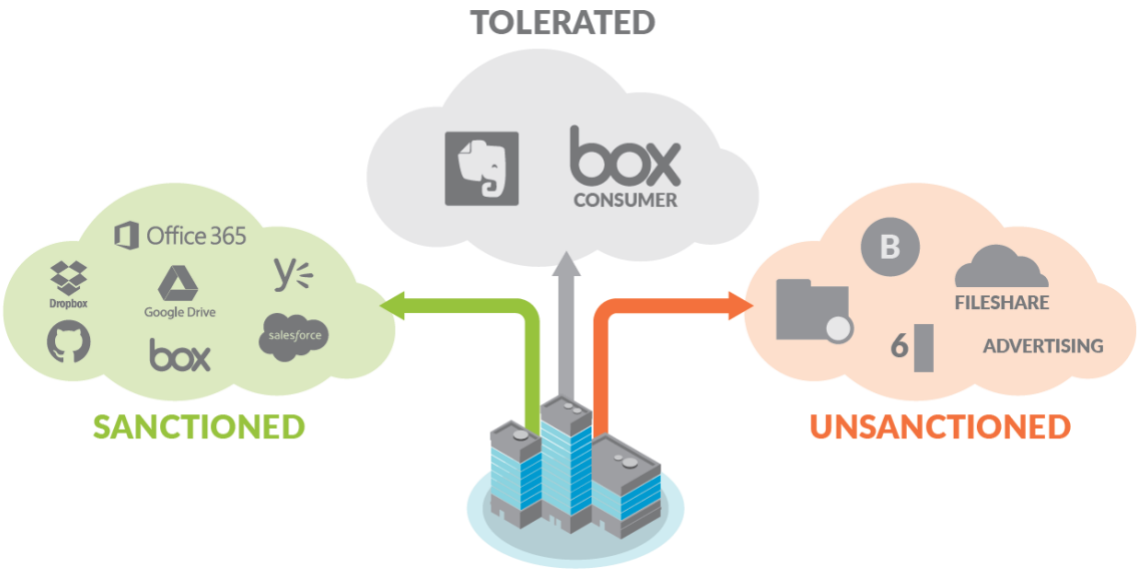

Figure 3-29: Impacts of sanctioned and unsanctioned SaaS applications.

To control sanctioned SaaS usage, an enterprise security solution must provide the following:

- **Threat prevention.** SaaS applications introduce new threat risks that need to be understood and controlled. Many SaaS applications sync files with users automatically, and users often share data in SaaS applications with third parties that are out of an organization's control. These two aspects of SaaS environments create a new insertion point for malware that can not only get in from external shares, but can also automatically sync those infected files across the organization without any user intervention. To address SaaS-based malware threats, a security solution must be able to prevent known and unknown malware from residing in sanctioned SaaS applications, regardless of the source.

- **Visibility and data exposure control.** After sanctioned SaaS usage is defined and controlled with a granular policy, data residing in those SaaS applications is no longer visible to the organization's perimeter firewalls. This loss of visibility creates a blind spot for IT. Additional data exposure controls are needed to specifically address the unique risks associated with SaaS environments, with a focus on data protection. Visibility of data stored and used in SaaS applications is critical to ensuring a deep understanding of users, the data they have shared, and how they have shared it.

- **Prevent risk, don't just respond.** It's very common for an organization's users to already be using certain SaaS applications long before the organization officially sanctions those applications. Even after a SaaS application is sanctioned, data is often shared with third

parties that don't necessarily have next-generation security solutions to effectively safeguard SaaS data from malware threats and data exposure risks. Threat prevention and data exposure control in a SaaS-based environment requires visibility and control not just from the time that a SaaS application is sanctioned going forward. You need visibility and control of *all* your data, including data that was being stored – and shared – before the SaaS application was sanctioned.

Data residing within enterprise-enabled SaaS applications is not visible to an organization's network perimeter. Palo Alto Networks Aperture connects directly to sanctioned SaaS applications to provide data classification, sharing/permission visibility, and threat detection within the application. This capability yields unparalleled visibility, which allows organizations to inspect content for data exposure violations and control access to shared data via a contextual policy.

Aperture builds upon the existing SaaS visibility and granular control capabilities of the Palo Alto Networks Security Operating Platform provided through App-ID with detailed SaaS-based reporting and granular control of SaaS usage. See Figure 3-30 for an example of the granular controls for SaaS applications supported with App-ID.

Application	Control	Feature
Box	Box – Personal	App-ID
	Box – Corporate	App-ID
	Upload control	File Blocking
	Download control	File Blocking
	Malware detection	WildFire & protection profile
	User-based control	User-ID

Figure 3-30: Example of granular controls supported with App-ID.

Aperture is a completely cloud-based, end-to-end security solution that provides visibility and control within SaaS applications, without the need for any proxies, agents, software, additional hardware, or network changes (see Figure 3-31). Aperture isn't an inline service, so it doesn't impact latency, bandwidth, or end-user experience. Aperture communicates directly with the SaaS applications themselves and looks at data from any source, regardless of the device or location from which the data was sent.

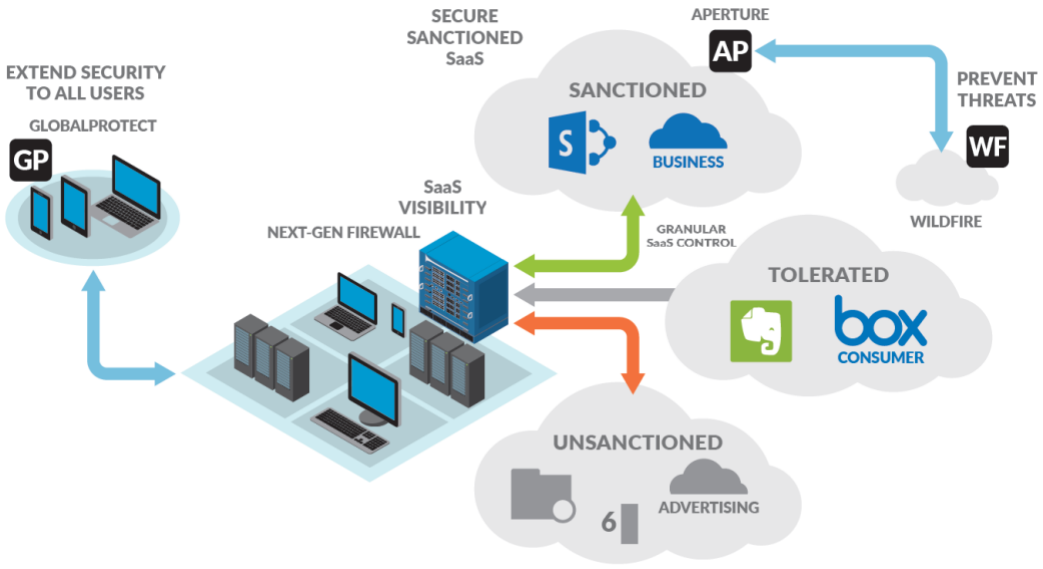

Figure 3-31: Complete SaaS visibility and control with Palo Alto Networks Security Operating Platform.

3.4.2.1 SaaS threat prevention

WildFire (discussed in Section 3.5.5) threat cloud integration with Aperture provides cyberthreat prevention to block known malware and identify and block unknown malware. This integration extends the existing integration of WildFire to prevent threats from spreading through the sanctioned SaaS applications, which prevents a new insertion point for malware. When new malware is discovered by Aperture, the threat information is shared with the rest of the Security Operating Platform, even if it is not deployed in-line with the SaaS applications.

3.4.2.2 Data exposure visibility

Aperture provides complete visibility across all user, folder, and file activity, which provides detailed analysis that helps you transition from a position of speculation to one of knowing exactly what's happening at any given point in time. With the ability to view deep analytics into day-to-day usage, you can quickly determine if there are any data risk or compliance-related policy violations. This detailed analysis of user and data activity allows for granular data governance and forensics.

Aperture connects directly to the applications themselves, so it provides continuous silent monitoring of the risks within the sanctioned SaaS applications, with detailed visibility that is not possible with traditional security solutions.

3.4.2.3 Contextual data exposure control

Aperture enables you to define granular, context-aware policy control that provides you with the ability to drive enforcement and the quarantine of users and data as soon as a violation

occurs. This control enables you to quickly and easily satisfy data risk compliance requirements such as PCI and PII while still maintaining the benefits of cloud-based applications.

Aperture prevents data exposure in unstructured (hosted files) and structured (application entries such as Salesforce.com) data. Both data types are common sources of improper data shares.

3.4.2.4 Advanced document classification

Aperture inspects documents for common sensitive data strings such as credit card numbers, SSH keys, and Social Security numbers, and flags them as risks if they are improperly shared. Unique to Aperture is the ability to identify documents by type, through advanced document classification regardless of the data that is contained in the document itself. Aperture has been predesigned to automatically identify sensitive documents, such as medical, tax, and legal.

3.4.2.5 Retroactive policy

A traditional network security solution can see only inline data and apply security policies to data that is accessed inline, after the policy is created. This approach doesn't effectively prevent SaaS data exposure, however, because SaaS data may have been shared long before the policy was created. This data may not be accessed inline for many months or years, potentially leaving sensitive data exposed to malware infection and unauthorized access indefinitely.

Aperture retroactively applies security policies to all users and data from the beginning of the SaaS account's creation, rather than the policy creation, to identify any potential vulnerabilities or policy violations. Aperture does not wait for someone to access the data in-line to apply policies and resolve any vulnerabilities or violations; SaaS data and shares are proactively discovered, protected, and resolved, no matter when they were created.

Policies are context-driven to allow for granular definitions of data exposure risks. This granularity is necessary to enable SaaS use by users while still preventing accidental data exposure. Policies take a number of factors in context to create an overall data exposure risk profile. One or two factors may not provide enough insight into the potential risk of the share. The overall risk of exposure is determined only after the full context of the share is understood.

Risks are calculated by user type, document type, sensitive data contained, how they are shared, and whether there is malware present. This capability provides the ability to control the exposure at a granular level based on a number of important factors.

For example, a financial team may be able to share financial data with other people on their team, but not beyond that. Even though the original share is allowed, they cannot share data that is infected with malware. Finance may, however, be allowed to share non-sensitive data companywide or, in some cases, with external vendors. The key to enabling this level of granularity is the ability to look at the share in the context of all the factors.

3.4 Knowledge Check

Test your understanding of the fundamentals in the preceding section. Review the correct answers in the Appendix at the end of this guide.

1. **Fill in the Blank.** _____ provides continuous monitoring of public clouds and helps organizations achieve a continuous state of compliance in their public cloud workloads.
2. **Short Answer.** What are some of the organizational security risks associated with unsanctioned SaaS application usage?
3. **Short Answer.** Explain why traditional perimeter-based firewalls are not effective for protecting data in SaaS environments.
4. **True or False.** Aperture is deployed as a standalone inline service between the organization's traditional perimeter-based firewalls and requires a software agent to be installed on mobile devices.
5. **True or False.** Aperture protects data in hosted files and application entries.

3.5 Application Framework and Logging Service

The Application Framework and Logging Service in the Security Operating Platform provides cloud-delivered security services including behavioral analytics (Magnifier), log management (Logging Service), threat intelligence (AutoFocus), threat indicator sharing (MineMeld), and malware analysis and threat prevention (WildFire).

3.5.1 Behavioral analytics (Magnifier)

Many organizations can't find intrusions quickly because security analysts are inundated with log messages generated by their infrastructure. They try to find high-priority threats by correlating logs, but they rarely have the right data or tools to accurately detect attacks. So,

they're left with endless alerts to review, many false positives and an unwieldy list of correlation rules to maintain.

As a result, security analysts operate in firefighting mode, attempting to review as many alerts as possible each day. These alerts often lack the context needed to confirm threats, so analysts waste valuable time chasing down additional information rather than stopping attacks.

Palo Alto Networks Magnifier behavioral analytics helps security analysts quickly find and stop the stealthiest network threats. By analyzing rich network, endpoint, and cloud data with machine learning, Magnifier accurately identifies targeted attacks, malicious insiders, and compromised endpoints. Security analysts can rapidly confirm threats by reviewing actionable alerts with investigative detail, and then leverage the NGFW to block threats before the damage is done.

By thwarting every step of an attack, organizations can limit any opportunity for an attack to succeed. Magnifier detects and stops command and control, lateral movement, and data exfiltration by detecting behavioral anomalies indicative of attack. Magnifier delivers powerful behavior-based protection, augmenting the Security Operating Platform to stop attacks across the attack lifecycle (see Figure 3-32).

Figure 3-32: The Security Operating Platform prevents threats across the attack lifecycle.

Magnifier automatically pinpoints active attacks, which allows security analysts to focus on the threats that matter. Magnifier starts by analyzing rich data stored in Logging Service by Palo Alto Networks NGFWs, including information on users, devices, and applications. Magnifier examines multiple logs, including Enhanced Application Logs, which provide data specifically designed for analytics. Analyzing multiple logs allows Magnifier to track attributes that are nearly impossible to ascertain from traditional threat logs or high-level network flow data. Magnifier uses the following machine learning techniques to analyze logs:

- **Unsupervised machine learning:** Magnifier uses unsupervised machine learning to model user and device behavior, perform peer group analysis, and cluster devices into relevant groups of behavior. Based on these profiles, Magnifier detects anomalies compared to past behavior and peer behavior.
- **Supervised machine learning:** Magnifier monitors multiple characteristics of network traffic to classify each device by type, such as a desktop computer, mobile device, or mail server. Magnifier also learns which users are IT administrators or normal users. With supervised machine learning, Magnifier recognizes deviations from expected behavior based on the type of user or device, reducing false positives.

Magnifier leverages a pre-compute detection framework to maximize speed, efficiency and accuracy. This framework processes log data stored in Logging Service by NGFWs and calculates the values it needs to track user and device behavior. Each Magnifier detection algorithm can analyze large amounts of data over long periods of time because the inputs have been pre-calculated. Instead of relying on correlation rules to parse large volumes of raw data and find one or two signs of malicious behavior, the Magnifier detection algorithms can evaluate past behavior, peer behavior, the type of entity, and many other attributes simultaneously to avoid false positives and produce higher-fidelity results.

By integrating attack detection algorithms with data collected from the Security Operating Platform and applying a pre-compute detection framework, Magnifier identifies active attacks with unparalleled precision.

To reduce investigation time, Magnifier produces a small number of accurate, actionable alerts, as well as information about the user, application, and device obtained through User-ID and App-ID technology. Magnifier also eliminates lengthy forensics investigations by interrogating endpoints to determine which process or executable initiated an attack. Then, Magnifier ascertains whether the endpoint process is malicious by integrating with WildFire cloud-based threat analysis service to analyze the process. Magnifier makes it easy for security analysts to verify attacks by presenting all the necessary information in an intuitive web interface (see Figure 3-33).

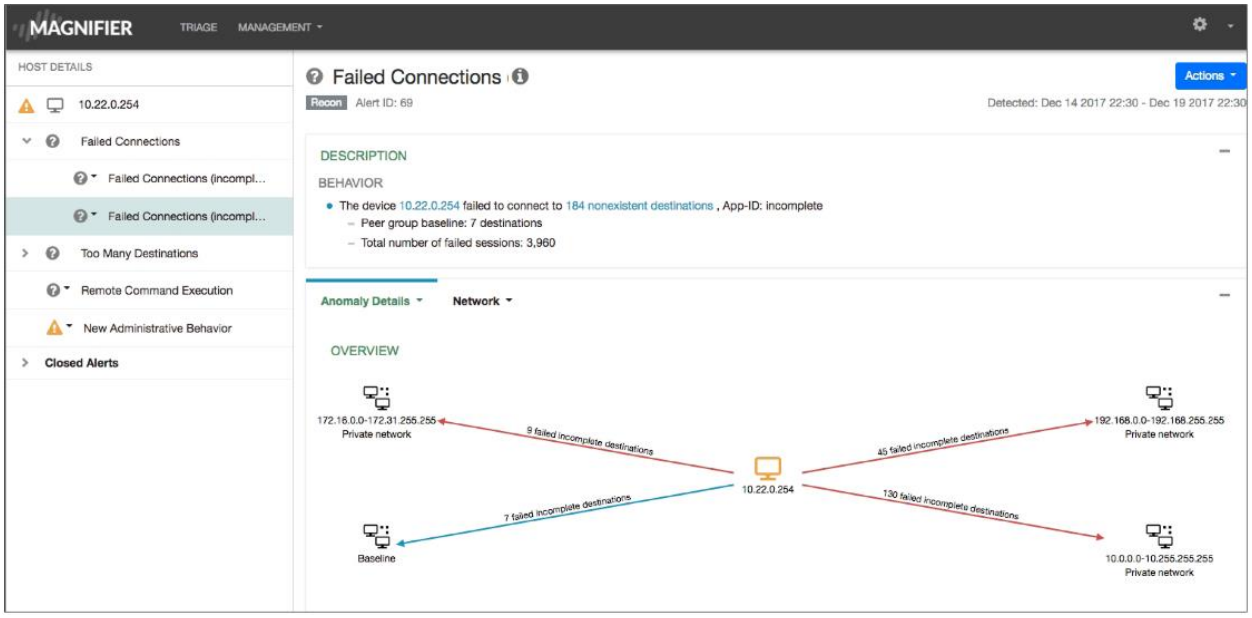

Figure 3-33: Magnifier web interface.

Magnifier behavioral analytics identifies behavioral anomalies to expose hard-to-detect threats, such as:

- **Targeted attacks.** Attackers attempt to blend in with legitimate users as they explore and exploit targeted networks. Magnifier detects the anomalous behavior that attackers cannot avoid as they traverse the network and look for valuable data.
- **Malicious insiders.** With their trusted credentials and access, malicious insiders can cause massive damage. Magnifier spots changes in user behavior to detect attacks like internal reconnaissance and lateral movement.
- **Risky behavior.** Well-meaning but reckless employees can expose organizations to undue risk. Magnifier allows organizations to follow security best practices by monitoring user activity and identifying risky behavior.
- **Compromised endpoints.** Attackers often use malware to infiltrate targeted networks. Magnifier identifies anomalous traffic generated by malware and confirms infections using Pathfinder endpoint analysis and WildFire threat analysis services.

Palo Alto Networks NGFWs monitor network traffic and extract metadata expressly designed for analytics. Magnifier uses this data, along with Pathfinder endpoint analysis, to profile user and device behavior without requiring organizations to provision new network sensors or agents (see Figure 3-34). Palo Alto Networks Logging Service delivers efficient log storage that scales to handle the large volumes of data needed for behavioral analytics. Organizations can

quickly deploy Magnifier and the Logging Service and avoid the time-consuming process of setting up new equipment.

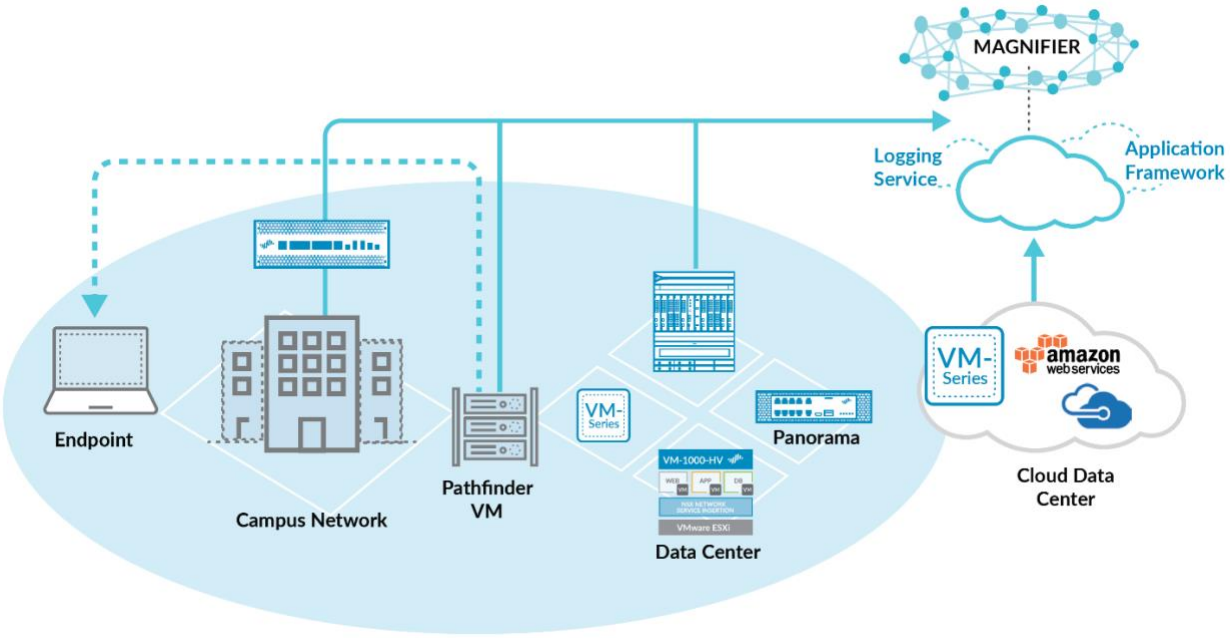

Figure 3-34: Magnifier uncovers attacks by analyzing data from NGFWs and Pathfinder endpoint analysis

As a cloud-based application for Palo Alto Networks Application Framework, Magnifier overcomes the scaling challenges of on-premises analytics and allows Palo Alto Networks researchers to roll out security innovations more quickly.

The Logging Service scales on demand with elastic cloud computing, providing an intelligent, operationally efficient, cost-effective way to store the large volumes of data needed for behavioral analytics.

As a cloud-delivered application, Magnifier increases the speed of technical innovation while streamlining IT operations. Magnifier researchers can rapidly roll out new behavioral analytics detection algorithms to all subscribers, review anonymized metrics to gauge their efficacy, and adjust as needed. Organizations no longer need to maintain or upgrade on-premises software because Magnifier is always up-to-date.

3.5.2 Log management (Logging Service)

Network security log analysis is an important cybersecurity practice that organizations perform to correlate potential threats and prevent breaches, but managing logs from various security tools and services takes effort and resources. To convert these logs into actionable information,

organizations need an affordable way to store, process, and analyze as much log data as possible. Unfortunately, traditional hardware-based log collection comes with administrative overhead and scale limitations that make otherwise useful data unwieldy or unavailable.

To protect their networks, organizations must be able to perform advanced analytics on all available data. Security applications that perform such analytics need access to scalable storage capacity and processing power. In the case of hardware-based log management products, such infrastructure and processing power may not be readily available, which makes these offerings less responsive to changing business needs.

Palo Alto Networks Logging Service is a cloud-based offering for context-rich enhanced network logs generated by Palo Alto Networks security products, including NGFWs, GlobalProtect cloud service, and Traps advanced endpoint protection. The cloud-based Logging Service lets organizations collect ever-expanding volumes of data without needing to plan for local compute and storage, and it is always ready to scale. Palo Alto Networks handles all the infrastructure needs, including storage and compute, to provide insights customers can use. If an organization already has on-premises log collectors, the Logging Service complements them by providing a logical extension of log storage to the cloud.

The Logging Service is the cornerstone of the Palo Alto Networks Application Framework (see Figure 3-35): a scalable ecosystem of security applications that can apply advanced analytics in concert with Palo Alto Networks enforcement points to prevent the most advanced attacks. Organizations are no longer limited by how much hardware is available or how quickly sensors can be deployed pervasively throughout the network.

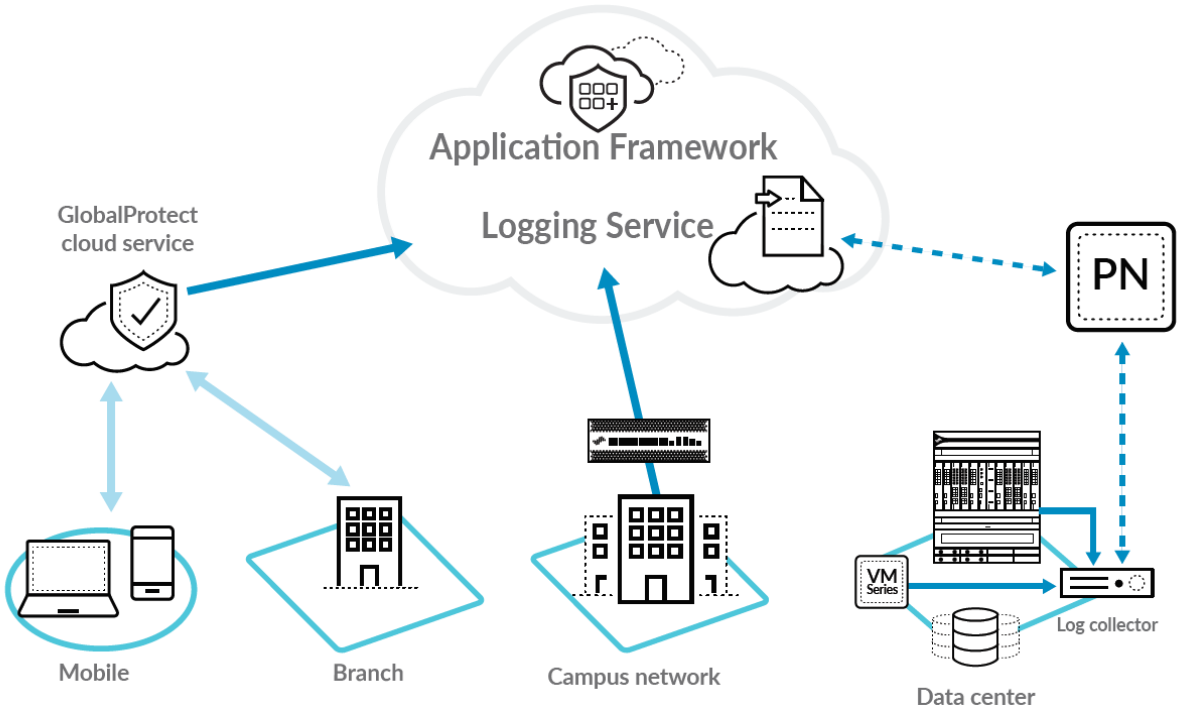

Figure 3-35: The Palo Alto Networks Logging Service.

Logging Service capabilities include:

- **Central repository for NGFW and cloud services logs.** The Logging Service can collect logs from NGFWs of all form factors as well as Palo Alto Networks cloud-based services. Logs are available in one location, which makes it easy to apply analytics and correlation capabilities to identify threats.
- **Logging infrastructure that scales with changing business needs.** The Logging Service was designed to scale quickly, and changes can easily be made.
- **Insight into network, application, and user behavior.** The Application Command Center – part of Panorama network security management (discussed in Section 3.2.3) – and its reporting capabilities gives security analysts critical insights into network, application, and user behavior. With this level of context, analysts can make informed decisions about how to eliminate open attack vectors and improve the organization's security posture.
- **Integration with other security infrastructure.** You can make the data and information hosted by the Logging Service available to your choice of third-party or custom security applications. You can also automate security workflows using Palo Alto Networks security infrastructure through the Application Framework.

3.5.3 Threat intelligence (AutoFocus)

Highly automated and increasingly sophisticated cyberattacks are occurring in greater volumes than ever before. Overburdened security teams, futilely attempting to investigate every threat in the enterprise network, have little time to analyze and understand truly advanced attacks – they're too busy fighting, rather than preventing, fires.

Palo Alto Networks AutoFocus enables a proactive, prevention-based approach to network security that puts automation to work for security professionals. Threat intelligence from the service is made directly accessible in the Palo Alto Networks platform, including PAN-OS software and Panorama. AutoFocus speeds the security team's existing workflows, which allows for in-depth investigation into suspicious activity, without additional specialized resources.

AutoFocus is built on a large-scale, distributed computing environment hosted in the Palo Alto Networks threat intelligence cloud. Unlike other solutions, the service makes threat data accessible and actionable at the IOC level and goes beyond simply showing summarized logs from multiple sources in a dashboard. AutoFocus has unprecedented visibility into the threat landscape, with the collective insight of thousands of global enterprises, service providers, and governments feeding the service (see Figure 3-36).

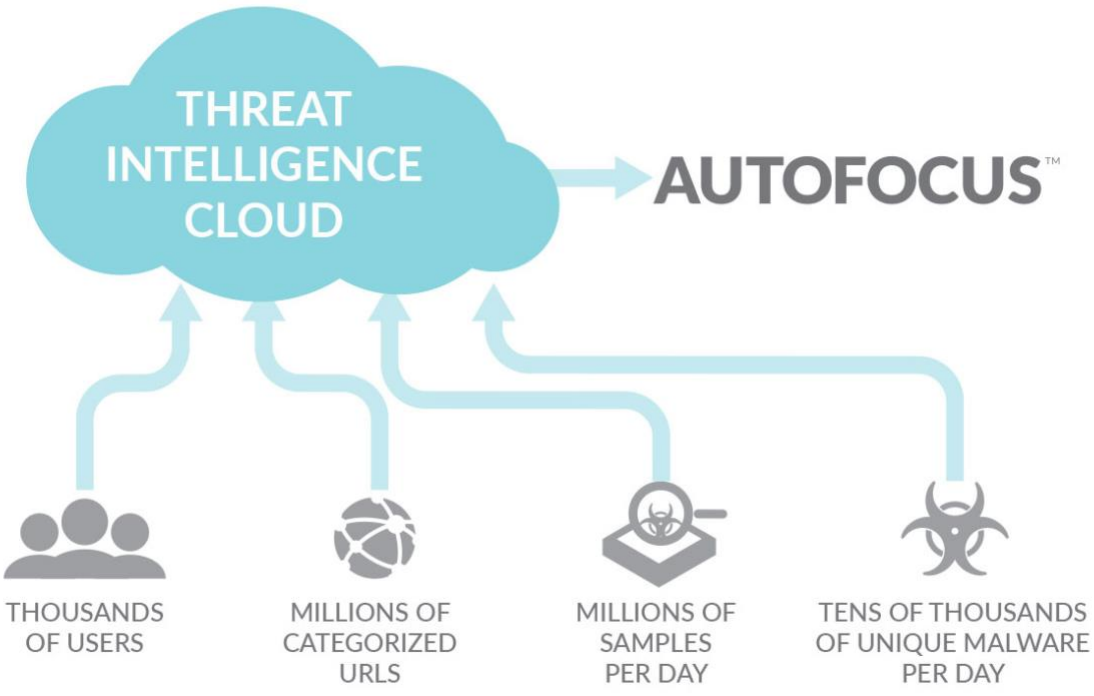

Figure 3-36: Palo Alto Networks AutoFocus threat intelligence cloud.

The service correlates and gains intelligence from:

- WildFire (discussed in Section 3.5.5)
- URL filtering with PAN-DB service
- Palo Alto Networks global passive DNS network
- Palo Alto Networks Unit 42 threat intelligence and research team
- Third-party feeds, including closed and open-source intelligence

AutoFocus makes over a billion samples and sessions, including billions of artifacts, immediately actionable for security analysis and response efforts. AutoFocus extends the Security Operating Platform with the global threat intelligence and attack context needed to accelerate analysis, forensics, and hunting workflows. Together, the platform and AutoFocus move security teams away from legacy manual approaches that rely on aggregating a growing number of detection-based alerts and post-event mitigation, to preventing sophisticated attacks and enabling proactive hunting activities.

3.5.3.1 Priority alerts and tags

AutoFocus enables you to distinguish the most important threats from everyday commodity attacks, contextualizing events on your network with tags. Unique to AutoFocus, tags enrich your visibility into the most critical threats, with contextual intelligence that let you know which malware families, campaigns, threat actors, malicious behaviors, and exploits, are being used against you.

When a tag matches an event on your network, a priority alert is sent via email, within the AutoFocus dashboard, or via HTTP post, with the full tag context included. Alerts are highly customizable, which enhances your existing security workflow with prioritization and context for the most critical threats.

Tags can be created for any host or network-based indicator in AutoFocus to alert you when a specific threat has been observed in your organization or industry. In addition to priority alerts, all tags are searchable so you can instantly pivot to associated malicious samples or indicators.

As new threats are identified, Palo Alto Networks Unit 42, your own organization, and the global community of AutoFocus experts add new tags to the service. AutoFocus is the primary analysis tool used by Unit 42 to identify new threats, correlate global data, identify connections between malicious samples, and build adversary or campaign profiles.

With AutoFocus and the Security Operating Platform, security teams can:

- Determine how targeted or unique a threat seen on their network is
- Investigate related malicious samples
- Identify suspicious DNS queries with domain resolution history

3.5.3.2 Threat correlation

When conducting threat analysis, security teams must quickly identify which IOCs represent the best path to remediation. For an active or ongoing compromise, the speed of investigation and the ability to meaningfully correlate data is critical. Each file has hundreds, potentially thousands, of artifacts, with only a small number of unique IOCs able to tie back to the larger profile of an adversary or related attacks.

AutoFocus uses an innovative statistical analysis engine to correlate billions of artifacts across a global data set and bring forward unique IOCs likely associated with targeted attacks. The service automatically applies a unique visual weighting system to identify unique and critical IOCs, which guides analysis and incident response efforts down the most relevant path.

AutoFocus allows you to build sophisticated multi-layer searches at the host and network-based artifact levels, and target your search within industry, time period, and other filters. These searches allow you to make previously unknown connections between attacks and pivot your incident response actions accordingly.

When further analysis is required, security teams can switch between AutoFocus and PAN-OS software or Panorama, with pre-populated searches for both systems. AutoFocus provides the entirety of Palo Alto Networks threat intelligence, which dramatically reduces the time it takes to conduct analysis, forensics, and hunting tasks.

3.5.3.3 Actionable intelligence

Security teams require more than a way to prioritize, analyze, and correlate threat intelligence – they need a way to convert it into actionable controls to prevent future attacks. AutoFocus enables you to create new protections for the Security Operating Platform by exporting high-value IOCs from the service into PAN-OS software external dynamic lists to block malicious URLs, domains, or IP addresses instantly. AutoFocus can also export IOCs to third-party security devices via a standard CSV format. Security teams can use AutoFocus to identify unique, targeted attacks against their organization and take direct action to mitigate and prevent them.

Threat analysis, forensics, and incident response teams often rely on a broad range of scripts, open-source tools, security devices, and services to investigate potential security incidents. AutoFocus can dramatically cut the time required to investigate by enriching third-party services through:

- **Open API support.** The AutoFocus API is built on an easy-to-use, *representational state transfer* (RESTful) framework, and allows for integrations into hundreds of use cases, such as feeding intelligence into existing SIEM tools. This framework makes data available for additional threat analysis or custom threat blocking automations.
- **Remote sweeping capability.** Security teams can sweep from indicators in the service to internal and third-party external systems directly from AutoFocus. Teams can define up to 10 external systems, which lets them continue their analysis seamlessly across their entire infrastructure, such as correlating logs from NGFWs or triggering searches in SIEM tools.
- **Support for STIX data format.** AutoFocus provides out-of-the-box integration with *structured threat information expression* (STIX) infrastructure and makes data available for export in the STIX data format.

Key Terms

Representational state transfer (REST) is an architectural programming style that typically runs over HTTP, and it is commonly used for mobile apps, social networking websites, and mashup tools.

Structured threat information expression (STIX) is an *extensible markup language* (XML) format for conveying data about cybersecurity threats in a standardized format.

Extensible markup language (XML) is a programming language specification that defines a set of rules for encoding documents in a human- and machine-readable format.

3.5.4 Threat indicator sharing (MineMeld)

To prevent successful cyberattacks, many organizations collect indicators of compromise (IOCs) from various threat intelligence providers with the intent of creating new controls for their security devices. Unfortunately, legacy approaches to aggregation and enforcement are highly manual in nature, often creating complex workflows and extending the time needed to identify and validate which IOCs should be blocked.

MineMeld is an open-source application that streamlines the aggregation, enforcement, and sharing of threat intelligence. MineMeld is available directly on GitHub, as well as on pre-built virtual machines (VMs) for easy deployment. With an extensible modular architecture, anyone can add to the MineMeld functionality by contributing code to the open-source repository.

MineMeld (see Figure 3-37) supports a variety of use cases, with more being added each day by the community, including:

- Aggregating and correlating threat intelligence feeds
- Enforcing new prevention controls, including IP blacklists
- Evaluating the value of a specific threat intelligence feed for your environment
- Extracting indicators from Palo Alto Networks device logs and sharing them with other security tools
- Sharing indicators with trusted peers
- Identifying incoming sessions from Tor exit nodes for blocking or strict inspection
- Tracking Office365 URLs and IPs

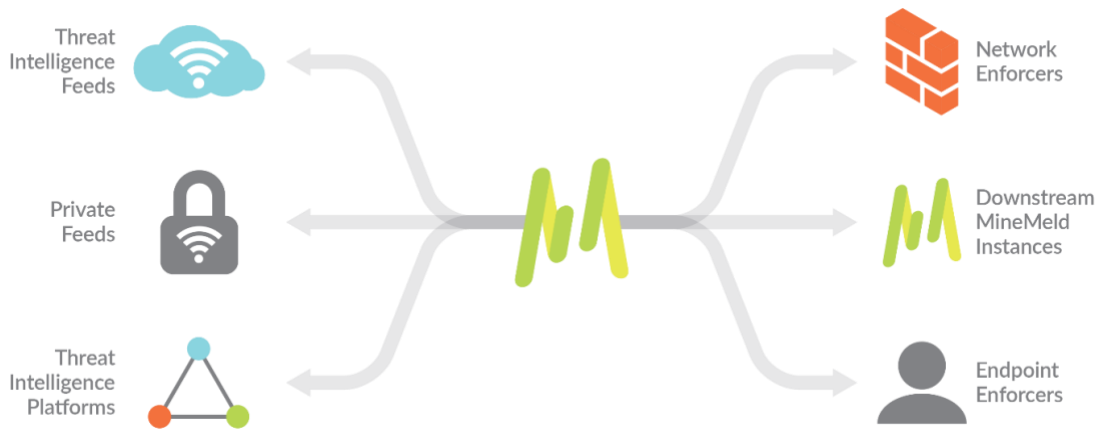

Figure 3-37: MineMeld aggregates and correlates threat intelligence feeds.

MineMeld allows you to aggregate threat intelligence across public, private, and commercial intelligence sources, including between government and commercial organizations.

MineMeld simplifies the collection and correlation of intelligence across:

- Commercial threat intelligence feeds
- Open-source intelligence (OSINT) providers
- Threat intelligence platforms

- Information sharing and analysis centers (ISACs)
- Computer emergency response teams (CERTs)
- Other MineMeld users

After indicators are collected, MineMeld can filter, deduplicate, and consolidate metadata across all sources, which allows security teams to analyze a more actionable set of data, enriched from multiple sources, for easier enforcement.

MineMeld natively integrates with Palo Alto Networks security platforms to automatically create new prevention-based controls for URLs, IPs, and domain intelligence derived from all sources feeding into the tool. Organizations can simplify their workflows for blocking IOCs with external dynamic lists and dynamic address groups, without spending additional resources to manage block lists, including the automated timeout of expired indicators. MineMeld also integrates with the Palo Alto Networks AutoFocus contextual threat intelligence service to allow organizations to identify high-value, targeted indicators – in AutoFocus – and block them on their NGFWs with export lists and MineMeld.

3.5.5 Malware analysis (WildFire)

Advanced cyberattacks employ stealthy and persistent methods to evade traditional security measures. Skilled adversaries require modern security teams to re-evaluate their prevention tactics to better address the volume and sophistication of today's attacks.

The Palo Alto Networks WildFire cloud-based malware analysis environment is a cyberthreat prevention service that identifies unknown malware, zero-day exploits, and advanced persistent threats (APTs) through static and dynamic analysis in a scalable, virtual environment. WildFire automatically disseminates updated protections in near real-time to immediately prevent threats from spreading – without manual intervention.

WildFire significantly improves security posture and protection against unknown malware. WildFire processes approximately 5 million unique files daily and approximately 30,000 to 50,000 unique malware files that are sent to WildFire by customer deployed Palo Alto Networks NGFWs. Typically, 60 percent of these malware files are not detected by any of the major antivirus vendors when first submitted to WildFire, and 30 days later 25 to 50 percent are still not detected by the major antivirus vendors.

To support dynamic malware analysis across the network at scale, WildFire is built on a cloud-based architecture (see Figure 3-38). Where regulatory or privacy requirements prevent the use of public cloud infrastructure, a private cloud solution can be built on premises.

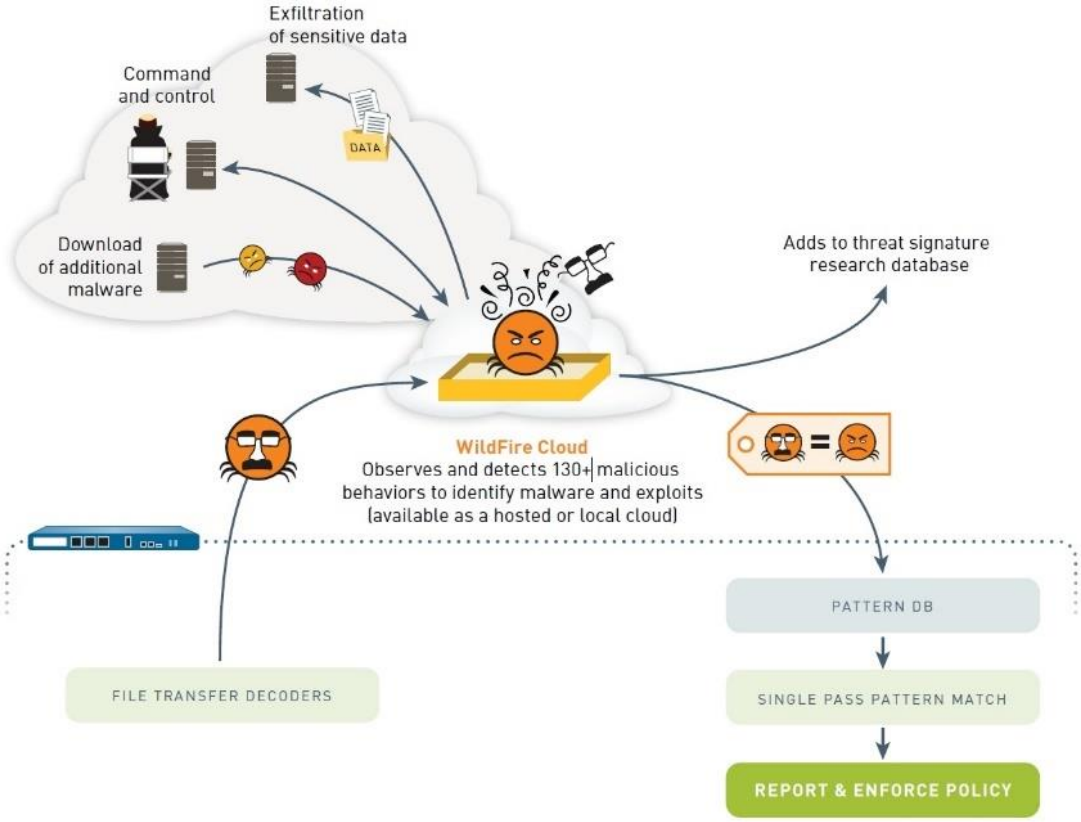

Figure 3-38: WildFire provides cloud-based malware analysis and threat prevention.

In addition to either public or private cloud deployments, leveraging both within the same environment is also an option. The hybrid cloud capabilities of WildFire allows security teams more file analysis flexibility because they can define which file types are sent to the WildFire public cloud versus the on-premises appliance, or private cloud. The WildFire hybrid cloud capability enables organizations to alleviate privacy or regulatory concerns by using the WildFire appliance for file types containing sensitive data. Organizations also benefit from the comprehensive analysis and global threat intelligence services of the WildFire public cloud for all others.

The Security Operating Platform proactively blocks known threats, which provides baseline defenses against known exploits, malware, malicious URLs and C&C activity. When new threats emerge, the Security Operating Platform automatically routes suspicious files and URLs to WildFire for deep analysis.

WildFire inspects millions of samples per week from its global network of customers and threat intelligence partners looking for new forms of previously unknown malware, exploits, malicious domains, and outbound C&C activity. The cloud-based service automatically creates new

protections that are capable of blocking targeted and unknown malware, exploits, and outbound C&C activity by observing their actual "behavior," rather than relying on pre-existing signatures. The protections are delivered globally in minutes. The result is a closed-loop, automated approach to preventing cyber threats that includes:

- Positive security controls to reduce the attack surface
- Inspection of all traffic, ports, and protocols to block all known threats
- Rapid detection of unknown threats by observing the actions of malware in a cloud-based execution environment
- Automatic deployment of new protections back to the frontline to ensure threats are known to all and blocked across the attack lifecycle

3.5.5.1 Behavior-based cyber threat discovery

To find unknown malware and exploits, WildFire executes suspicious content in the Windows, Android, and Mac OS X operating systems, with full visibility into common file types, including:

- Executables (EXEs), dynamic link libraries (DLLs), compressed files (ZIP), and portable document format (PDF)
- Microsoft Office documents, spreadsheets, and presentations
- Java files
- Android application packages (APKs)
- Adobe Flash applets and webpages (including high-risk embedded content, such as Java and Adobe Flash files/images)

WildFire identifies hundreds of potentially malicious behaviors to uncover the true nature of malicious files based on their actions, including:

- **Changes made to host:** WildFire observes all processes for modifications to the host, including file and registry activity, code injection, memory heap spray (exploit) detection, addition of auto-run programs, *mutexes*, Windows services, and other suspicious activities.
- **Suspicious network traffic:** WildFire performs analysis of all network activity produced by the suspicious file, including backdoor creation, downloading of next-stage malware, visiting low-reputation domains, network reconnaissance, and more.

- **Anti-analysis detection:** WildFire monitors techniques used by advanced malware that is designed to avoid virtual machine (VM)-based analysis, such as debugger detection, hypervisor detection, code injection into trusted processes, disabling of host-based security features, and more.

Key Terms

A *mutex* is a program object that allows multiple program threads to share the same resource, such as file access, but not simultaneously.

WildFire is natively integrated with the Security Operating Platform, which is capable of classifying all traffic across hundreds of applications. WildFire uniquely applies this behavioral analysis to web traffic, email protocols (SMTP, IMAP, POP), and FTP, regardless of ports or encryption.

3.5.5.2 Threat prevention with global intelligence sharing

When an unknown threat is discovered, WildFire automatically generates protections to block it across the cyberattack lifecycle, and it shares these updates with all global subscribers in as little as 5 minutes. These quick updates are able to stop rapidly spreading malware; and these updates are payload-based, so they can block proliferation of future variants without any additional action or analysis.

In addition to protecting organizations from malicious and exploitive files and links, WildFire looks deep into malicious outbound communication and disrupts C&C activity with anti-C&C signatures and DNS-based callback signatures. The information is also fed into URL Filtering with PAN-DB, where newly discovered malicious URLs are automatically blocked. This correlation of threat data and automated protections is key to identifying and blocking ongoing intrusion attempts and future attacks against your organization.

3.5.5.3 Integrated logging, reporting, and forensics

WildFire provides access to integrated logs, analysis, and visibility into WildFire events, through the management interface, the WildFire portal, AutoFocus (discussed in Section 3.5.3), and Panorama (discussed in Section 3.2.3). This access enables security teams to quickly investigate and correlate events observed in their networks to rapidly locate the data needed for timely investigations and incident response.

Host-based and network-based *indicators of compromise* (IOCs) become actionable through log analysis and custom signatures. To aid security and incident response teams in discovering infected hosts, WildFire also provides:

- Detailed analysis of every malicious file sent to WildFire across multiple operating system environments, including host-based and network-based activity.
- Session data associated with the delivery of the malicious file, including source, destination, application, User-ID, URL, and more.
- Access to the original malware sample for reverse-engineering and full *packet captures* (pcaps) of dynamic analysis sessions.
- An open *application programming interface* (API) for integration with best-in-class SIEM tools, such as the Palo Alto Networks application for Splunk, and leading endpoint agents. This analysis provides a wealth of IOCs that can be applied across the attack lifecycle.
- Native integration with Traps advanced endpoint protection (discussed in Section 3.3.1) and Aperture advanced SaaS protection (discussed in Section 3.4.2).
- Access to the actionable intelligence and global context provided by AutoFocus threat intelligence (discussed in Section 3.5.3).
- Natively integrated with the correlation engine in Palo Alto Networks NGFWs (discussed in Section 3.2.1).

Key Terms

An *indicator of compromise* (IOC) is a network or operating system (OS) artifact that provides a high level of confidence that a computer security incident has occurred.

An *application programming interface* (API) is a set of routines, protocols, and tools for building software applications and integrations.

A *packet capture* (pcap) is a traffic intercept of data packets that can be used for analysis.

3.5 Knowledge Check

Test your understanding of the fundamentals in the preceding section. Review the correct answers in the Appendix at the end of this guide.

1. **Fill in the Blank.** Magnifier leverages _____ to analyze network, endpoint, and cloud data, which helps security analysts rapidly confirm threats by reviewing actionable alerts.

2. **Multiple Choice.** Which three options are threat intelligence sources for AutoFocus? (Choose three.)

    a) WildFire
    b) URL filtering with PAN-DB service
    c) Unit 42 threat intelligence and research team
    d) third-party intrusion prevention systems

3. **True or False.** AutoFocus is an optional module that can be installed on NGFWs.

4. **Fill in the Blank.** _____ is an open-source application, available directly on GitHub, that streamlines the aggregation, enforcement, and sharing of threat intelligence.

5. **Multiple Choice.** WildFire operates on which concept? (Choose one.)

    a) file-based scanning against a signature database
    b) IPS and SIEM tool correlation
    c) cloud-based reputation service
    d) virtualized sandbox

6. **True or False.** WildFire prevents known and unknown malware threats.

7. **True or False.** WildFire performs deep packet inspection of malicious outbound communications to disrupt C&C activity.

Appendix A – Knowledge Check Answers

Section 1.1 Knowledge Check

1. [c] software as a Service (SaaS)
2. True
3. True
4. True
5. Health Insurance Portability and Accountability Act (HIPAA)
6. Discussion
7. Discussion

Section 1.2 Knowledge Check

1. False. External threat answers have accounted for the majority of data breaches over the past five years. According to Verizon's *2018 Data Breach Investigations Report*, internal threat actors are responsible for approximately 28 percent of reported data breaches.
2. Discussion
3. False. The Cyber-Attack Lifecycle is a seven-step process.
4. Reconnaissance, Weaponization, Delivery, Exploitation, Installation, Command and Control, Actions on the Objective
5. False. A defender only needs to break a single step in the Cyber Attack Lifecycle framework to prevent an attack from succeeding.
6. [c] vulnerability and patch management
7. True

Section 1.3 Knowledge Check

1. True
2. Zero-day

Section 1.4 Knowledge Check

1. [b] malware
2. [d] all of the above
3. False. A vulnerability is a bug or flaw that exists in a system or software, and creates a security risk.
4. False. The number of core exploit techniques is relatively small, typically as few as two to four new techniques are developed each year.

Section 1.5 Knowledge Check

1. False. WEP has many well-known and well-publicized weaknesses and is not considered effective for establishing a secure wireless network.
2. one-way hash

Section 2.1 Knowledge Check

1. router
2. [b] routing information protocol (RIP)
3. [a] Distance-vector, [b] Path-vector, and [c] Link-state
4. True
5. domain name system (DNS)

Section 2.2 Knowledge Check

1. [a] IP address
2. 8
3. Subnetting

Section 2.3 Knowledge Check

1. [c] seven
2. [a] transmission control protocol (TCP), [c] user datagram protocol (UDP)

3. Media Access Control (MAC), Logical Link Control (LLC)
4. [a] Application, [b] Transport, [d] Internet, [e] Network Access
5. data encapsulation

Section 2.4 Knowledge Check

1. The primary issue with a perimeter-centric network security strategy is that it relies on the assumption that everything on the internal network can be trusted.
2. [b] least privilege

Section 2.5 Knowledge Check

1. Cloud computing doesn't mitigate existing network security risks; security requires isolation and segmentation, whereas the cloud relies on shared resources; security deployments are process-oriented, whereas cloud computing environments are dynamic.
2. [c] East-west traffic
3. [b] consolidating servers within trust levels

Section 2.6 Knowledge Check

1. False. A dynamic packet filtering (also known as stateful packet inspection) firewall only inspects individual packet headers during session establishment to determine if the traffic should be allowed, blocked, or dropped by the firewall. Once a session is established, individual packets that are part of the session are not inspected.
2. [a] proxies traffic rather than permitting direct communication between hosts, [b] can be used to implement strong user authentication, [c] masks the internal network from untrusted networks
3. IDS is considered to be a *passive* system because it only monitors, analyzes, and alerts. IPS is an *active* system that performs all of the functions of IDS, but can also block or drop suspicious, pattern-matching traffic on the network.
4. [c] secure sockets layer (SSL)
5. [c] UTM fully integrates all of the security functions installed on the device

Section 2.7 Knowledge Check

1. False. Signature-based anti-malware software is considered a reactive countermeasure because a signature file for new malware can't be created and delivered until the malware is already "in the wild."
2. Container-based
3. The main disadvantage of application whitelisting related to exploit prevention is that an application that has been whitelisted is permitted to run – even if the application has a vulnerability that can be exploited.
4. [a] data loss prevention (DLP), [b] policy enforcement, [d] malware prevention

Section 2.8 Knowledge Check

1. [a] Software as a Service (SaaS), [b] Platform as a Service (PaaS), [d] Infrastructure as a Service (IaaS)
2. hybrid
3. Shared Responsibility Model
4. hypervisor
5. [a] dormant VMs, [b] hypervisor vulnerabilities, [d] intra-VM communications
6. block

Section 2.9 Knowledge Check

1. Active Directory
2. ITIL

Section 3.1 Knowledge Check

1. Security Operating Platform
2. [a] network security, [b] advanced endpoint protection, [c] cloud security

Section 3.2 Knowledge Check

1. [b] Adherence to strict port and protocol enforcement for allow/block decisions

2. Application identification, user identification, and content identification
3. [a] packet headers
4. Any three of the following: Security event log monitoring (Active Directory, Novell eDirectory, Microsoft Exchange), user provided credentials, client probing, receiving user information through XML API from an external LDAP directory
5. Unlike file-based malware scanning that waits until an entire file is loaded into memory to begin scanning, stream-based malware scanning begins scanning as soon as the first packets of the file are received. Stream-based malware scanning reduces latency and improves performance by receiving, scanning, and sending traffic to its intended destination immediately, without having to first buffer and then scan the file.
6. Templates eliminate manual, repetitive, risky, and error-prone configuration changes to multiple, individual firewalls deployed throughout the enterprise network.
7. [d] traditional port-based firewalls

Section 3.3 Knowledge Check

1. True
2. The Traps agent injects itself into each process as it is started. If the process attempts to execute any of the core attack techniques, the corresponding EPM kills the process and prevents the exploit.
3. Manage the device, protect the device, control the data

Section 3.4 Knowledge Check

1. Evident
2. The organizational security risks associated with unsanctioned SaaS application usage include regulatory non-compliance or compliance violations, loss of corporate intellectual property (IP) or other sensitive data, and malware distribution.
3. Traditional perimeter-based firewalls only have visibility of traffic that passes through the firewall. SaaS applications and data can be accessed from mobile devices that don't necessarily traverse a perimeter-based firewall, and many SaaS-based applications are designed to circumvent firewalls for performance and ease of use.

4. False. Aperture is used to protect sanctioned SaaS usage, as part of an integrated security solution that includes NGFWs to prevent unsanctioned SaaS use. Aperture communicates directly with the SaaS applications themselves and therefore does not need to be deployed in-line and does not require any software agents, proxies, additional hardware, or network configuration changes.

5. True

Section 3.5 Knowledge Check

1. machine learning

2. [a] WildFire, [b] URL filtering with PAN-DB service, [c] Unit 42 threat intelligence and research team

3. False. AutoFocus is a subscription-based threat intelligence cloud that fully integrates with the Security Operating Platform, but does not require any configuration changes to NGFWs or Traps Advanced Endpoint Protection.

4. MineMeld

5. [d] Virtualized sandbox

6. False. WildFire prevents unknown malware threats. Known malware threats are prevented by the other components of the Security Operating Platform, including NGFWs, Traps Advanced Endpoint Protection, and Aperture SaaS-based security.

7. True

Appendix B – Glossary

access point (AP): See *wireless access point (AP)*.

address resolution protocol (ARP): A protocol that translates a logical address, such as an IP address, to a physical MAC address. The *reverse address resolution protocol* (RARP) translates a physical MAC address to a logical address. See also *IP address*, *media access control (MAC) address*, and *reverse address resolution protocol (RARP)*.

Advanced Encryption Standard (AES): A symmetric block cipher based on the Rijndael cipher.

AES: See Advanced Encryption Standard (AES).

AP: See *wireless access point (AP)*.

API: See application programming interface (API).

application programming interface (API): A set of routines, protocols, and tools for building software applications and integrations.

application whitelisting: A technique used to prevent unauthorized applications from running on an endpoint. Authorized applications are manually added to a list that is maintained on the endpoint. If an application is not on the whitelist, it cannot run on the endpoint. However, if it is on the whitelist the application can run, regardless of whether or not vulnerabilities or exploits are present within the application.

ARP: See *address resolution protocol (ARP).*

AS: See *autonomous system (AS).*

attack vector: A path or tool that an attacker uses to target a network.

authoritative DNS server: The system of record for a given domain. See also *domain name system (DNS)*.

autonomous system (AS): A group of contiguous IP address ranges under the control of a single Internet entity. Individual autonomous systems are assigned a 16-bit or 32-bit AS number (ASN) that uniquely identifies the network on the Internet. ASNs are assigned by the Internet Assigned Numbers Authority (IANA). See also *internet protocol (IP) address* and *Internet Assigned Numbers Authority (IANA)*.

bare metal hypervisor: See *native hypervisor*.

BES: See *bulk electric system (BES)*.

boot sector: Contains machine code that is loaded into an endpoint's memory by firmware during the startup process, before the operating system is loaded.

boot sector virus: Targets the boot sector or master boot record (MBR) of an endpoint's storage drive or other removable storage media. See also *boot sector* and *master boot record (MBR)*.

bot: Individual endpoints that are infected with advanced malware that enables an attacker to take control of the compromised endpoint. Also known as a zombie. See also *botnet* and *malware*.

botnet: A network of bots (often tens of thousands or more) working together under the control of attackers using numerous command and control (C&C) servers. See also *bot*.

bridge: A wired or wireless network device that extends a network or joins separate network segments.

bring your own apps (BYOA): Closely related to BYOD, BYOA is a policy trend in which organizations permit end users to download, install, and use their own personal apps on mobile devices, primarily smartphones and tablets, for work-related purposes. See also *bring your own device (BYOD)*.

bring your own device (BYOD): A policy trend in which organizations permit end users to use their own personal devices, primarily smartphones and tablets, for work-related purposes. 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 that must be supported. See also *bring your own apps (BYOA)*.

broadband cable: A type of high-speed Internet access that delivers different upload and download data speeds over a shared network medium. The overall speed varies depending upon the network traffic load from all the subscribers on the network segment.

broadcast domain: The portion of a network that receives broadcast packets sent from a node in the domain.

bulk electric system (BES): The large interconnected electrical system, consisting of generation and transmission facilities (among others), that comprises the "power grid."

***bus (or linear bus) topology*:** A LAN topology in which all nodes are connected to a single cable (the backbone) that is terminated on both ends. In the past, bus networks were commonly used

for very small networks because they were inexpensive and relatively easy to install, but today bus topologies are rarely used. The cable media has physical limitations (the cable length), the backbone is a single point of failure (a break anywhere on the network affects the entire network), and tracing a fault in a large network can be extremely difficult. See also *local area network (LAN)*.

BYOA: See *bring your own apps (BYOA).*

BYOD: See *bring your own device (BYOD).*

child process: In multitasking operating systems, a sub-process created by a parent process that is currently running on the system.

CIDR: See *classless inter-domain routing (CIDR)*.

CIP: See *Critical Infrastructure Protection (CIP)*.

circuit-switched network: A network in which a dedicated physical circuit path is established, maintained, and terminated between the sender and receiver across a network for each communications session.

classless inter-domain routing (CIDR): A method for allocating IP addresses and IP routing that replaces classful IP addressing (for example, Class A, B, and C networks) with classless IP addressing. See also *internet protocol (IP) address*.

collision domain: A network segment on which data packets may collide with each other during transmission.

consumerization: A computing trend that describes the process that occurs as end users increasingly find personal technology and apps that are more powerful or capable, more convenient, less expensive, quicker to install, and easier to use, than enterprise IT solutions.

convergence: The time it takes for all routers in a network to update their routing tables with the most current routing information about the network.

covered entity: Defined by HIPAA as a healthcare provider that electronically transmits PHI (such as doctors, clinics, psychologists, dentists, chiropractors, nursing homes, and pharmacies), a health plan (such as a health insurance company, health maintenance organization, company health plan, or government program including Medicare, Medicaid, military and veterans' healthcare), or a healthcare clearinghouse. See also *Health Insurance Portability and Accountability Act (HIPAA)* and *protected health information (PHI)*.

CRC: See *cyclic redundancy check (CRC).*

Critical Infrastructure Protection (CIP): Cybersecurity standards defined by NERC to protect the physical and cyber assets necessary to operate the bulk electric system (BES). See also *bulk electric system (BES)* and *North American Electric Reliability Corporation (NERC)*.

Cybersecurity Enhancement Act of 2014: A U.S. regulation which provides an ongoing, voluntary public-private partnership to improve cybersecurity and to strengthen cybersecurity research and development, workforce development and education, and public awareness and preparedness.

Cybersecurity Information Sharing Act (CISA): A U.S. regulation which enhances information sharing about cybersecurity threats by allowing Internet traffic information to be shared between the U.S. government and technology and manufacturing companies.

cyclic redundancy check (CRC): A checksum used to create a message profile. The CRC is recalculated by the receiving device. If the recalculated CRC doesn't match the received CRC, the packet is dropped and a request to resend the packet is transmitted back to the device that sent the packet.

data encapsulation: A process in which protocol information from the OSI or TCP/IP layer immediately above is wrapped in the data section of the OSI or TCP/IP layer immediately below. Also referred to as data hiding. See also *Open Systems Interconnection (OSI) reference model* and *Transmission Control Protocol/Internet Protocol (TCP/IP) model*.

data hiding: See *data encapsulation.*

DDOS: See *distributed denial-of-service (DDOS).*

default gateway: A network device, such as a router or switch, to which an endpoint sends network traffic when a specific destination IP address is not specified by an application or service, or when the endpoint does not know how to reach a specified destination. See also *router* and *switch*.

DHCP: See *dynamic host configuration protocol (DHCP).*

digital subscriber line (DSL): A type of high-speed Internet access that delivers different upload and download data speeds. The overall speed depends upon the distance from the home or business location to the provider's central office (CO).

distributed denial-of-service (DDOS): A type of cyberattack in which extremely high volumes of network traffic such as packets, data, or transactions are sent to the target victim's network to

make their network and systems (such as an e-commerce website or other web application) unavailable or unusable.

DLL: See *dynamic link library (DLL).*

DNS: See *domain name system (DNS).*

domain name registrar: An organization that is accredited by a *top-level domain* (TLD) registry to manage domain name registrations. See also *top-level domain (TLD)*.

domain name system (DNS): A hierarchical distributed database that maps the fully qualified domain name (FQDN) for computers, services, or any resource connected to the Internet or a private network to an IP address. See also *fully qualified domain name (FQDN)*.

drive-by-download: A software download, typically malware, that happens without a user's knowledge or permission.

DSL: See *digital subscriber line (DSL).*

dynamic host configuration protocol (DHCP): A network management protocol that dynamically assigns (leases) IP addresses and other network configuration parameters (such as *default gateway* and *domain name system* [DNS] information) to devices on a network. See also *default gateway* and *domain name system (DNS)*.

dynamic link library (DLL): A type of file used in Microsoft operating systems that enables multiple programs to simultaneously share programming instructions contained in a single file to perform specific functions.

EAP: See *extensible authentication protocol (EAP)*.

EAP-TLS: See *extensible authentication protocol transport layer security (EAP-TLS)*.

EHR: See *electronic health record (EHR)*.

electronic health record (EHR): As defined by HealthIT.gov, an EHR "goes beyond the data collected in the provider's office and include[s] a more comprehensive patient history. EHR data can be created, managed, and consulted by authorized providers and staff from across more than one healthcare organization."

electronic medical record (EMR): As defined by HealthIT.gov, an EMR "contains the standard medical and clinical data gathered in one provider's office."

EMR: See *electronic medical record (EMR)*.

endpoint: A computing device such as a desktop or laptop computer, handheld scanner, point-of-sale (POS) terminal, printer, satellite radio, security or videoconferencing camera, self-service kiosk, server, smart meter, smart TV, smartphone, tablet, or Voice over internet protocol (VoIP) phone. Although endpoints can include servers and network equipment, the term is generally used to describe end user devices.

Enterprise 2.0: A term introduced by Andrew McAfee and defined as "the use of emergent social software platforms within companies, or between companies and their partners or customers." See also *Web 2.0*.

exclusive or (XOR): A Boolean operator in which the output is true only when the inputs are different (for example, TRUE and TRUE equals FALSE, but TRUE and FALSE equals TRUE).

exploit: A small piece of software code, part of a malformed data file, or a sequence (string) of commands, that leverages a vulnerability in a system or software, causing unintended or unanticipated behavior in the system or software.

extensible authentication protocol (EAP): A widely used authentication framework that includes approximately 40 different authentication methods.

extensible authentication protocol transport layer security (EAP-TLS): An Internet Engineering Task Force (IETF) open standard that uses the transport layer security (TLS) protocol in Wi-Fi networks and PPP connections. See also *Internet Engineering Task Force (IETF)*, *point-to-point protocol (PPP)* and *transport layer security (TLS)*.

extensible markup language (XML): A programming language specification that defines a set of rules for encoding documents in a human- and machine-readable format.

false negative: In anti-malware, malware that is incorrectly identified as a legitimate file or application. In intrusion detection, a threat that is incorrectly identified as legitimate traffic. See also *false positive*.

false positive: In anti-malware, a legitimate file or application that is incorrectly identified as malware. In intrusion detection, legitimate traffic that is incorrectly identified as a threat. See also *false negative*.

favicon ("favorite icon"): A small file containing one or more small icons associated with a particular website or webpage.

Federal Exchange Data Breach Notification Act of 2015: A U.S. regulation which further strengthens HIPAA by requiring health insurance exchanges to notify individuals whose

personal information has been compromised as the result of a data breach as soon as possible, but no later than 60 days after breach discovery. See also *Health Insurance Portability and Accountability Act (HIPAA)*.

Federal Information Security Management Act (FISMA): See *Federal Information Security Modernization Act (FISMA)*.

Federal Information Security Modernization Act (FISMA): A U.S. law that implements a comprehensive framework to protect information systems used in U.S. federal government agencies. Known as the Federal Information Security Management Act prior to 2014.

fiber optic: Technology that converts electrical data signals to light and delivers constant data speeds in the upload and download directions over a dedicated fiber optic cable medium. Fiber optic technology is much faster and more secure than other types of network technology.

Financial Services Modernization Act of 1999: See *Gramm-Leach-Bliley Act (GLBA)*.

FISMA: See *Federal Information Security Modernization Act (FISMA)*.

floppy disk: A removable magnetic storage medium commonly used from the mid-1970s until approximately 2007, when they were largely replaced by removable USB storage devices.

flow control: A technique used to monitor the flow of data between devices to ensure that a receiving device, which may not necessarily be operating at the same speed as the transmitting device, doesn't drop packets.

fully qualified domain name (FQDN): The complete domain name for a specific computer, service, or resource connected to the Internet or a private network.

GDPR: See *General Data Protection Regulation (GDPR)*.

General Data Protection Regulation (GDPR): A European Union (EU) regulation that applies to any organization that does business with EU citizens. It strengthens data protection for EU citizens and addresses the export of personal data outside the EU.

generic routing encapsulation (GRE): A tunneling protocol developed by Cisco Systems that can encapsulate various network layer protocols inside virtual point-to-point links.

GLBA: See *Gramm-Leach-Bliley Act (GLBA)*.

Gramm-Leach-Bliley Act (GLBA): A U.S. law that requires financial institutions to implement privacy and information security policies to safeguard the non-public personal information of clients and consumers. Also known as the Financial Services Modernization Act of 1999.

GRE: See *generic routing encapsulation (GRE)*.

hacker: Originally used to refer to anyone with highly specialized computing skills, without connoting good or bad purposes. However, common misuse of the term has redefined a hacker as someone that circumvents computer security with malicious intent, such as a cybercriminal, cyberterrorist, or hacktivist.

hash signature: A cryptographic representation of an entire file or program's source code.

Health Insurance Portability and Accountability Act (HIPAA): A U.S. law that defines data privacy and security requirements to protect individuals' medical records and other personal health information. See also *covered entity* and *protected health information (PHI).*

heap spray: A technique used to facilitate arbitrary code execution by injecting a certain sequence of bytes into the memory of a target process.

hextet: A group of four 4-bit hexadecimal digits in a 128-bit IPv6 address. See also *internet protocol (IP) address*.

high-order bits: The first four bits in a 32-bit IPv4 address octet. See also *internet protocol (IP) address*, *octet*, and *low-order bits*.

HIPAA: See *Health Insurance Portability and Accountability Act (HIPAA).*

hop count: The number of router nodes that a packet must pass through to reach its destination.

hosted hypervisor: A hypervisor that runs within an operating system environment. Also known as a Type 2 hypervisor. See also *hypervisor* and *native hypervisor*.

HTTP: See *hypertext transfer protocol (HTTP).*

HTTPS: See *hypertext transfer protocol secure (HTTPS).*

hub (or concentrator): A device used to connect multiple networked devices together on a local area network (LAN).

hypertext transfer protocol (HTTP): An application protocol used to transfer data between web servers and web browsers.

hypertext transfer protocol secure (HTTPS): A secure version of HTTP that uses secure sockets layer (SSL) or transport layer security (TLS) encryption. See also *secure sockets layer (SSL)* and *transport layer security (TLS)*.

hypervisor: Technology that allows multiple, virtual (or guest) operating systems to run concurrently on a single physical host computer.

IaaS: See *Infrastructure as a Service (IaaS)*.

IANA: See *Internet Assigned Numbers Authority (IANA).*

IETF: See *Internet Engineering Task Force (IETF)*.

indicator of compromise (IOC): A network or operating system (OS) artifact that provides a high level of confidence that a computer security incident has occurred.

Infrastructure as a Service (IaaS). A cloud computing service model in which customers can provision processing, storage, networks, and other computing resources and deploy and run operating systems and applications. However, the customer has no knowledge of, and does not manage or control, the underlying cloud infrastructure. The customer has control over operating systems, storage, and deployed applications, as well as some networking components (for example, host firewalls). The company owns the deployed applications and data, and it is therefore responsible for the security of those applications and data.

initialization vector (IV): A random number used only once in a session, in conjunction with an encryption key, to protect data confidentiality. Also known as a nonce.

inodes: Used to store information about files and directories in a file-based storage system, but not the filenames or data content itself.

Internet Assigned Numbers Authority (IANA): A private, nonprofit U.S. corporation that oversees global IP address allocation, autonomous system (AS) number allocation, root zone management in the domain name system (DNS), media types, and other internet protocol-related symbols and Internet numbers. See also *autonomous system (AS)* and *domain name system (DNS)*.

Internet Engineering Task Force (IETF): An open international community of network designers, operators, vendors, and researchers concerned with the evolution of the Internet architecture and the smooth operation of the Internet.

internet protocol (IP) address: A 32- or 128-bit identifier assigned to a networked device for communications at the Network layer of the OSI model or the Internet layer of the TCP/IP model. See also *Open Systems Interconnection (OSI) reference model* and *Transmission Control Protocol/Internet Protocol (TCP/IP) model*.

intranet: A private network that provides information and resources – such as a company directory, human resources policies and forms, department or team files, and other internal information – to an organization's users. Like the Internet, an intranet uses the HTTP and/or HTTPS protocols, but access to an intranet is typically restricted to an organization's internal users. Microsoft SharePoint is a popular example of intranet software. See also *hypertext transfer protocol (HTTP)* and *hypertext transfer protocol secure (HTTPS)*.

IOC: See *indicator of compromise (IOC)*.

IP address: See *internet protocol (IP) address*.

IP telephony: See *voice over internet protocol (VoIP).*

IV: See *initialization vector (IV)*.

jailbreaking: 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 apps without paying for them, from sources, other than the App Store, that are not sanctioned and/or controlled by Apple. Jailbreaking bypasses the security features of the device by replacing the firmware's operating system with a similar, albeit counterfeit version, which makes it vulnerable to malware and exploits. See also *rooting*.

Kerberos: A ticket-based authentication protocol in which "tickets" are used to identify network users.

LAN: See *local area network (LAN).*

least privilege: A network security principle in which only the permission or access rights necessary to perform an authorized task are granted.

least significant bit: The last bit in a 32-bit IPv4 address octet. See also *internet protocol (IP) address*, *octet*, and *most significant bit*.

linear bus topology: See *bus topology.*

local area network (LAN): A computer network that connects laptop and desktop computers, servers, printers, and other devices so that applications, databases, files and file storage, and other networked resources can be shared across a relatively small geographic area, such as a floor, a building, or a group of buildings.

low-order bits: The last four bits in a 32-bit IPv4 address octet. See also *internet protocol (IP) address*, *octet*, and *high-order bits*.

MAC address: See *media access control (MAC) address.*

malware: Malicious software or code that typically damages, takes control of, or collects information from an infected endpoint. Malware broadly includes viruses, worms, Trojan horses (including Remote Access Trojans, or RATs), anti-AV, logic bombs, backdoors, rootkits, bootkits, spyware, and (to a lesser extent) adware.

master boot record (MBR): Contains information on how the logical partitions (or file systems) are organized on the storage media, and an executable boot loader that starts up the installed operating system.

MBR: See *master boot record (MBR).*

media access control (MAC) address: A unique 48 or 64-bit identifier assigned to a network interface controller (NIC) for communications at the Data Link layer of the OSI model. See also *Open Systems Interconnection (OSI) reference model*.

metamorphism: A programming technique used to alter malware code with every iteration, to avoid detection by signature-based anti-malware software. Although the malware payload changes with each iteration – for example, by using a different code structure or sequence, or inserting garbage code to change the file size – the fundamental behavior of the malware payload remains unchanged. Metamorphism uses more advanced techniques than polymorphism. See also *polymorphism*.

Microsoft challenge-handshake authentication protocol (MS-CHAP): A protocol used to authenticate Microsoft Windows-based workstation, using a challenge-response mechanism to authenticate PPTP connections without sending passwords. See also *point-to-point tunneling protocol (PPTP)*.

most significant bit: The first bit in a 32-bit IPv4 address octet. See also *internet protocol (IP) address*, *octet*, and *least significant bit*.

MS-CHAP: See *Microsoft challenge-handshake authentication protocol (MS-CHAP)*.

mutex: A program object that allows multiple program threads to share the same resource, such as file access, but not simultaneously.

NAT: See *network address translation (NAT).*

National Cybersecurity Protection Advancement Act of 2015: A U.S. regulation which amends the Homeland Security Act of 2002 to enhance multi-directional sharing of information related to cybersecurity risks and strengthens privacy and civil liberties protections.

native hypervisor: A hypervisor that runs directly on the host computer hardware. Also known as a Type 1 or bare metal hypervisor. See also *hypervisor* and *hosted hypervisor*.

NERC: See *North American Electric Reliability Corporation (NERC)*.

network address translation (NAT): A technique used to virtualize IP addresses by mapping private, non-routable IP addresses assigned to internal network devices to public IP addresses.

Network and Information Security (NIS) Directive: A European Union (EU) directive that imposes network and information security requirements for banks, energy companies, healthcare providers and digital service providers, among others.

NIS Directive: See *Network and Information Security (NIS) Directive*.

nonce: See *initialization vector (IV).*

North American Electric Reliability Corporation (NERC): A not-for-profit international regulatory authority responsible for assuring the reliability of the bulk electric system (BES) in the continental U.S., Canada, and the northern portion of Baja California, Mexico. See also *bulk electric system (BES)* and *Critical Infrastructure Protection (CIP)*.

obfuscation: A programming technique used to render code unreadable. It can be implemented using a simple substitution cipher, such as an *exclusive or* (XOR) operation, or more sophisticated encryption algorithms, such as the *Advanced Encryption Standard* (AES). See also *Advanced Encryption Standard (AES)*, *exclusive or (XOR)*, and *packer*.

octet: A group of 8 bits in a 32-bit IPv4 address. See *internet protocol (IP) address*.

one-way (hash) function: A mathematical function that creates a unique representation (a hash value) of a larger set of data in a manner that is easy to compute in one direction (input to output), but not in the reverse direction (output to input). The hash function can't recover the original text from the hash value. However, an attacker could attempt to guess what the original text was and see if it produces a matching hash value.

Open Systems Interconnection (OSI) reference model: A seven-layer networking model consisting of the Application (Layer 7 or L7), Presentation (Layer 6 or L6), Session (Layer 5 or L5), Transport (Layer 4 or L4), Network (Layer 3 or L3), Data Link (Layer 2 or L2), and Physical (Layer 1 or L1) layers. Defines standard protocols for communication and interoperability using a layered approach in which data is passed from the highest layer (application) downward through each layer to the lowest layer (physical), then transmitted across the network to its

destination, then passed upward from the lowest layer to the highest layer. See also *data encapsulation*.

optical carrier: A standard specification for the transmission bandwidth of digital signals on Synchronous Optical Networking (SONET) fiber optic networks. Optical carrier transmission rates are designated by the integer value of the multiple of the base rate (51.84 Mbps). For example, OC-3 designates a 155.52 Mbps (3 x 51.84) network and OC-192 designates a 9953.28 Mbps (192 x 51.84) network.

OSI model: See *Open Systems Interconnection (OSI) reference model.*

PaaS: See *Platform as a Service (PaaS)*.

packer: A software tool that can be used to obfuscate code by compressing a malware program for delivery, then decompressing it in memory at runtime. See also *obfuscation*.

packet capture (PCAP): A traffic intercept of data packets that can be used for analysis.

packet-switched network: A network in which devices share bandwidth on communications links to transport packets between a sender and receiver across a network.

PAP: See *password authentication protocol (PAP)*.

password authentication protocol (PAP): An authentication protocol used by PPP to validate users with an unencrypted password. See also *point-to-point protocol (PPP)*.

Payment Card Industry Data Security Standards (PCI DSS): A proprietary information security standard mandated and administered by the PCI Security Standards Council (SSC), and applicable to any organization that transmits, processes, or stores payment card (such as debit and credit cards) information. See also *PCI Security Standards Council (SSC)*.

PCAP: See *packet capture (PCAP).*

PCI: See *Payment Card Industry Data Security Standards (PCI DSS).*

PCI DSS: See *Payment Card Industry Data Security Standards (PCI DSS)*.

PCI Security Standards Council (SSC): Comprised of Visa, MasterCard, American Express, Discover, and JCB, the SSC maintains, evolves, and promotes PCI DSS. See also *Payment Card Industry Data Security Standards (PCI DSS)*.

PDU: See *protocol data unit (PDU).*

Personal Information Protection and Electronic Documents Act (PIPEDA): A Canadian privacy law that defines individual rights with respect to the privacy of their personal information, and governs how private sector organizations collect, use, and disclose personal information in the course of business.

personally identifiable information (PII): Defined by the U.S. National Institute of Standards and Technology (NIST) as "any information about an individual maintained by an agency, including (1) any information that can be used to distinguish or trace an individual's identity… and (2) any other information that is linked or linkable to an individual…."

pharming: A type of attack that redirects a legitimate website's traffic to a fake site.

PHI: See *protected health information (PHI).*

PII: See *Personally Identifiable Information (PII).*

PIPEDA: See *Personal Information Protection and Electronic Documents Act (PIPEDA).*

PKI: See *public key infrastructure (PKI)*.

Platform as a Service (PaaS): A cloud computing service model in which customers can deploy supported applications onto the provider's cloud infrastructure, but the customer has no knowledge of, and does not manage or control, the underlying cloud infrastructure. The customer has control over the deployed applications and limited configuration settings for the application-hosting environment. The company owns the deployed applications and data, and it is therefore responsible for the security of those applications and data.

PoE: See *power over Ethernet (PoE).*

point-to-point protocol (PPP): A Layer 2 (Data Link) protocol layer used to establish a direct connection between two nodes.

point-to-point tunneling protocol (PPTP): An obsolete method for implementing virtual private networks, with many known security issues, which uses a TCP control channel and a GRE tunnel to encapsulate PPP packets. See also *transmission control protocol (TCP)*, *generic routing encapsulation (GRE)*, and *point-to-point protocol (PPP)*.

polymorphism: A programming technique used to alter a part of malware code with every iteration, to avoid detection by signature-based anti-malware software. For example, an encryption key or decryption routine may change with every iteration, but the malware payload remains unchanged. See also *metamorphism*.

power over Ethernet (PoE): A network standard that provides electrical power to certain network devices over Ethernet cables.

PPP: See *point-to-point protocol (PPP)*.

PPTP: See *point-to-point tunneling protocol (PPTP).*

pre-shared key (PSK): A shared secret, used in symmetric key cryptography which has been exchanged between two parties communicating over an encrypted channel.

promiscuous mode: Refers to Ethernet hardware used in computer networking, typically a network interface card (NIC), that receives all traffic on a network segment, even if the traffic is not addressed to the hardware.

protected health information (PHI): Defined by HIPAA as information about an individual's health status, provision of healthcare, or payment for healthcare that includes identifiers such as names, geographic identifiers (smaller than a state), dates, phone and fax numbers, email addresses, Social Security numbers, medical record numbers, or photographs, among others. See also *Health Insurance Portability and Accountability Act (HIPAA)*.

protocol data unit (PDU): a self-contained unit of data (consisting of user data or control information and network addressing).

PSK: See *pre-shared key (PSK)*.

public key infrastructure (PKI): A set of roles, policies, and procedures needed to create, manage, distribute, use, store, and revoke digital certificates and manage public key encryption.

QoS: See *quality of service (QoS)*.

quality of service (QoS): The overall performance of specific applications or services on a network including error rate, bit rate, throughput, transmission delay, availability, jitter, etc. QoS policies can be configured on certain network and security devices to prioritize certain traffic, such as voice or video, over other, less performance-intensive traffic, such as file transfers.

RADIUS: See *remote authentication dial-in user service (RADIUS)*.

rainbow table: A pre-computed table used to find the original value of a cryptographic hash function.

RARP: See reverse address resolution protocol (RARP).

recursive DNS query: A DNS query that is performed (if the DNS server allows recursive queries) when a DNS server is not authoritative for a destination domain. The non-authoritative DNS server obtains the IP address of the authoritative DNS server for the destination domain and sends the original DNS request to that server to be resolved. See also *domain name system (DNS)* and *authoritative DNS server*.

remote authentication dial-in user service (RADIUS): A client/server protocol and software that enables remote access servers to communicate with a central server to authenticate users and authorize access to a system or service.

remote procedure call (RPC): 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.

repeater: A network device that boosts or re-transmits a signal to physically extend the range of a wired or wireless network.

representational state transfer (REST): An architectural programming style that typically runs over HTTP, and is commonly used for mobile apps, social networking websites, and mashup tools. See also *hypertext transfer protocol (HTTP)*.

REST: See *representational state transfer (REST)*.

reverse address resolution protocol (RARP): Translates a physical MAC address to a logical address. See also *media access control (MAC) address*.

ring topology: A LAN topology in which all nodes are connected in a closed loop that forms a continuous ring. In a ring topology, all communication travels in a single direction around the ring. Ring topologies were common in token ring networks. See also *local area network (LAN)*.

rooting: The Google Android equivalent of jailbreaking. See *jailbreaking*.

router: A network device that sends data packets to a destination network along a network path.

RPC: See *remote procedure call (RPC)***.**

SaaS: See *Software as a Service (SaaS).*

salt: Randomly generated data that is used as an additional input to a one-way has function that hashes a password or passphrase. The same original text hashed with different salts results in different hash values.

Sarbanes-Oxley (SOX) Act: A U.S. law that increases financial governance and accountability in publicly traded companies.

script kiddie: Someone with limited hacking and/or programming skills that uses malicious programs (malware) written by others to attack a computer or network.

secure sockets layer (SSL): A cryptographic protocol for managing authentication and encrypted communication between a client and server to protect the confidentiality and integrity of data exchanged in the session.

service set identifier (SSID): A case sensitive, 32-character alphanumeric identifier that uniquely identifies a Wi-Fi network.

Software as a Service (SaaS): A cloud computing service model, defined by the U.S. National Institute of Standards and Technology (NIST), in which "the capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through either a thin client interface, such as a web browser, or a program interface. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings."

SONET: See *Synchronous Optical Networking (SONET).*

SOX: See *Sarbanes-Oxley (SOX) Act*.

spear phishing: A highly targeted phishing attack that uses specific information about the target to make the phishing attempt appear legitimate.

SSID: See *service set identifier (SSID)*.

SSL: See *secure sockets layer (SSL).*

STIX: See *structured threat information expression (STIX).*

structured threat information expression (STIX): An XML format for conveying data about cybersecurity threats in a standardized format. See also *extensible markup language (XML)*.

subnet mask: A number that hides the network portion of an IPv4 address, leaving only the host portion of the IP address. See also *internet protocol (IP) address*.

subnetting: A technique used to divide a large network into smaller, multiple subnetworks.

supernetting: A technique used to aggregate multiple contiguous smaller networks into a larger network to enable more efficient Internet routing.

switch: An intelligent hub that forwards data packets only to the port associated with the destination device on a network.

Synchronous Optical Networking (SONET): A protocol that transfer multiple digital bit streams synchronously over optical fiber.

T-carrier: A full-duplex digital transmission system that uses multiple pairs of copper wire to transmit electrical signals over a network. For example, a T-1 circuit consists of two pairs of copper wire – one pair transmits, the other pair receives – that are multiplexed to provide a total of 24 channels, each delivering 64 Kbps of data, for a total bandwidth of 1.544 Mbps.

TCP: See *transmission control protocol (TCP).*

TCP segment: A protocol data unit (PDU) defined at the Transport layer of the OSI model. See also *protocol data unit (PDU)* and *Open Systems Interconnection (OSI) reference model*.

three-way handshake: A sequence used to establish a TCP connection. For example, a PC initiates a connection with a server by sending a TCP SYN (Synchronize) packet. The server replies with a SYN ACK packet (Synchronize Acknowledgement). Finally, the sends an ACK or SYN-ACK-ACK packet, acknowledging the server's acknowledgement, and data communication commences. See also *transmission control protocol (TCP)*.

TCP/IP model: See *Transmission Control Protocol/Internet Protocol (TCP/IP) model*.

threat vector: See *attack vector.*

TLD: See *top-level domain (TLD).*

TLS: See *transport layer security (TLS)*.

top-level domain (TLD): The highest level domain in DNS, represented by the last part of a FQDN (for example, .com or .edu). The most commonly used TLDs are generic top-level domains (gTLD) such as .com, edu, .net, and .org, and country-code top-level domains (ccTLD) such as .ca and .us.

Tor ("The Onion Router"): Software that enables anonymous communication over the Internet.

transmission control protocol (TCP): A connection-oriented (a direct connection between network devices is established before data segments are transferred) protocol that provides

reliable delivery (received segments are acknowledged and retransmission of missing or corrupted segments is requested) of data.

Transmission Control Protocol/Internet Protocol (TCP/IP) model: A four-layer networking model consisting of the Application (Layer 4 or L4), Transport (Layer 3 or L3), Internet (Layer 2 or L2), and Network Access (Layer 1 or L1) layers.

transport layer security (TLS): The successor to SSL (although it is still commonly referred to as SSL). See also *secure sockets layer (SSL)*.

Type 1 hypervisor: See *native hypervisor*.

Type 2 hypervisor: See *hosted hypervisor.*

UDP: See *user datagram protocol (UDP).*

UDP datagram: A protocol data unit (PDU) defined at the Transport layer of the OSI model. See also *user datagram protocol (UDP)* and *Open Systems Interconnection (OSI) reference model*.

uniform resource locator (URL): A unique reference (or address) to an Internet resource, such as a webpage.

URL: See *uniform resource locator (URL)*.

user datagram protocol (UDP): A connectionless (a direct connection between network devices is not established before datagrams are transferred) protocol that provides best-effort delivery (received datagrams are not acknowledged and missing or corrupted datagrams are not requested) of data.

variable-length subnet masking (VLSM): A technique that enables IP address spaces to be divided into different sizes. See also *internet protocol (IP) address*.

virtual LAN (VLAN): A logical network that is created within a physical local area network.

VLAN: See *virtual LAN (VLAN)*.

VLSM: See *variable-length subnet masking (VLSM)*.

voice over internet protocol (VoIP): Technology that provides voice communication over an internet protocol (IP) based network. Also known as IP telephony.

VoIP: See *voice over internet protocol (VoIP)*.

vulnerability: A bug or flaw that exists in a system or software, and creates a security risk.

WAN: See *wide area network (WAN).*

watering hole: An attack which compromises websites that are likely to be visited by a targeted victim to deliver malware via a drive-by-download. See also *drive-by-download*.

Web 2.0: A term popularized by Tim O'Reilly and Dale Dougherty, unofficially referring to a new era of the World Wide Web, which is characterized by dynamic or user-generated content, interaction, and collaboration, and the growth of social media. See also *Enterprise 2.0*.

whaling: A type of spear phishing attack that is specifically directed at senior executives or other high-profile targets within an organization. See also *spear phishing*.

wide area network (WAN): A computer network that connects multiple LANs or other WANs across a relatively large geographic area, such as a small city, a region or country, a global enterprise network, or the entire planet (for example, the Internet). See also local area network (LAN).

wireless access point (AP): A network device that connects to a router or wired network and transmits a Wi-Fi signal so that wireless devices can connect to a wireless (or Wi-Fi) network.

wireless repeater: A device that rebroadcasts the wireless signal from a wireless router or AP to extend the range of a Wi-Fi network.

XML: See *extensible markup language (XML)*.

XOR: See *exclusive or (XOR)*.

zero-day threat: The window of vulnerability that exists from the time a new (unknown) threat is released until security vendors release a signature file or security patch for the threat.

zombie: See *bot*.

Appendix C – Palo Alto Networks Technical Training and Certification Programs

Palo Alto Networks offers technical training and certification programs that provide you with the advanced knowledge you need to secure enterprise networks and safely enable applications. Training from Palo Alto Networks and Palo Alto Networks Authorized Training Centers delivers knowledge and expertise that prepare you to protect our digital way of life. Palo Alto Networks trusted security certifications validate your knowledge of the Palo Alto Networks next-generation security platform and your ability to help prevent successful cyberattacks and safely enable applications. You can learn more about these programs at www.paloaltonetworks.com/services/education.

Firewall 8.1 Essentials: Configuration and Management (EDU-210)

The Palo Alto Networks Firewall 8.1 Essentials: Configuration and Management (EDU-210) course is five days of instructor-led training that should enable you to:

- Configure and manage the essential features of Palo Alto Networks next-generation firewalls (NGFWs)
- Configure and manage GlobalProtect to protect systems that are located outside of the data center perimeter
- Configure and manage firewall high availability
- Monitor network traffic using the interactive web interface and firewall reports

Course Objectives

Successful completion of this five-day, instructor-led course should enhance the student's understanding of how to configure and manage Palo Alto Networks NGFWs. The student will get hands-on experience configuring, managing, and monitoring a firewall in a lab environment.

Scope

- **Course level:** Introductory
- **Course duration:** 5 days
- **Course format:** Combines lecture and hands-on labs

- **Platform support:** Palo Alto Networks next-generation enterprise firewalls running the PAN-OS operating system

Target Audience

Security Engineers, Security Administrators, Security Operations Specialists, Security Analysts, Network Engineers, and Support Staff

Prerequisites

Students must have a basic familiarity with networking concepts including routing, switching, and IP addressing. Students also should be familiar with basic security concepts. Experience with other security technologies (IPS, proxy, and content filtering) is a plus.

Sessions

- **Module 1:** Next Generation Security Platform and Architecture
- **Module 2:** Virtual and Cloud Deployment
- **Module 3:** Initial Configuration
- **Module 4:** Interface Configuration
- **Module 5:** Security and NAT Policies
- **Module 6:** App-ID
- **Module 7:** Content-ID
- **Module 8:** URL Filtering
- **Module 9:** Decryption
- **Module 10:** WildFire
- **Module 11:** User-ID
- **Module 12:** GlobalProtect
- **Module 13:** Site-to-Site VPNs
- **Module 14:** Monitoring and Reporting
- **Module 15:** Active/Passive High Availability

- **Module 16:** Next-Generation Security Practices

Firewall 8.1: Optimizing Firewall Threat Prevention (EDU-214)

The Palo Alto Networks Firewall 8.1: Optimizing Firewall Threat Prevention (EDU-214) course is four days of instructor-led training that emphasizes the PAN-OS threat prevention capabilities. After completing this course, you should be able to:

- Describe the cyber-attack lifecycle and recognize common forms of attack
- Describe PAN-OS threat prevention capabilities
- Use firewall logs and reports to make better configuration decisions
- Configure the firewall to detect, block, and record threats

Course Objectives

Successful completion of this four-day, instructor-led course will enhance the student's understanding of how to better configure, manage, and monitor PAN-OS threat prevention features. The student will get hands-on experience configuring, managing, and monitoring threat prevention features in a lab environment.

Scope

- **Course level:** Intermediate
- **Course duration:** 4 days
- **Course format:** Combines lecture and hands-on labs
- **Platform support:** Palo Alto Networks next-generation enterprise firewalls running the PAN-OS operating system

Target Audience

Security Engineers, Security Administrators, Security Operations Specialists, Security Analysts, Network Engineers, and Support Staff

Sessions

- **Module 1:** The Cyber-Attack Lifecycle
- **Module 2:** Blocking Packet- and Protocol-Based Attacks

- **Module 3:** Blocking Threats from Known-Bad Sources
- **Module 4:** Blocking Threats Using AppID
- **Module 5:** Blocking Threats Using Custom Signatures
- **Module 6:** Creating Custom Threat Signatures
- **Module 7:** Blocking Threats in Encrypted Traffic
- **Module 8:** Blocking Threats in Allowed Traffic
- **Module 9:** Authenticating Firewall User Accounts
- **Module 10:** Blocking Threats from Phishing and Stolen Credentials
- **Module 11:** Viewing Threat and Traffic Information

Panorama 8.1: Manage Firewalls at Scale (EDU-220)

The Palo Alto Networks Panorama 8.1: Managing Firewalls at Scale (EDU-220) course is two days of instructor-led training that will help you:

- Learn how to configure and manage the next-generation Panorama management server
- Gain experience configuring templates (including template variables) and device groups
- Gain experience with administration, log collection, and logging and reporting
- Gain experience with Panorama High Availability and Panorama troubleshooting
- Become familiar with new Panorama features such as Panorama in the public cloud, the Logging Service, and GlobalProtect cloud service

Course Objectives

This course will help students to gain in-depth knowledge about how to configure and manage their Palo Alto Networks Panorama management server. Upon completion of this course, administrators should be familiar with the Panorama management server's role in managing and securing their overall network. Network professionals will be shown how to use Panorama aggregated reporting to provide them with a holistic view of a network of Palo Alto Networks NGFWs.

Scope

- **Course level:** Intermediate
- **Course duration:** 2 days
- **Course format:** Combines lecture with hands-on labs

Target Audience

Security Administrators, Security Operations Specialists, Security Analysts, Security Engineers, and Security Architects

Prerequisites

Students must complete the Firewall 8.1 Essentials: Configuration and Management (EDU-210) class, and be familiar with Palo Alto Networks NGFW management and basic networking concepts, including routing and IP addressing.

Sessions

- **Module 1:** Panorama Overview
- **Module 2:** Initial Configuration
- **Module 3:** Adding Firewalls to Panorama
- **Module 4:** Panorama High Availability
- **Module 5:** Templates
- **Module 6:** Device Groups
- **Module 7:** Administrative Accounts
- **Module 8:** Log Forwarding and Collection
- **Module 9:** Aggregated Monitoring and Reporting
- **Module 10:** Troubleshooting

Firewall 8.1: Troubleshooting (EDU-330)

The Palo Alto Networks Firewall 8.1: Troubleshooting course is three days of instructor-led training that will help you:

- Investigate networking issues using firewall tools including the CLI
- Follow proven troubleshooting methodologies specific to individual features
- Analyze advanced logs to resolve various real-life scenarios
- Solve advanced, scenario-based challenges

Course Objectives

Successful completion of this three-day, instructor-led course will enhance the participant's understanding of how to troubleshoot the full line of Palo Alto Networks NGFWs.

Participants will have opportunities to perform hands-on troubleshooting of common problems related to the configuration and operation of the features of the Palo Alto Networks PAN-OS operating system.

Completion of this class will help participants develop an in-depth knowledge of how to troubleshoot visibility and control over applications, users, and content.

Scope

- **Course level:** Advanced
- **Course duration:** 3 days
- **Course format:** Lecture and hands-on labs
- **Platform support:** Palo Alto Networks next-generation enterprise firewalls running the PAN-OS operating system

Target Audience

Security Engineers, Security Administrators, Security Operations Specialists, Security Analysts, Network Engineers, and Support Staff

Prerequisites

Participants must complete the Firewall 8.1 Essentials: Configuration and Management (EDU-210) course. Participants must have strong practical knowledge of routing and switching, IP addressing, and network security concepts, and at least six months of on-the-job experience with Palo Alto Networks firewalls.

Sessions

- **Module 1:** Tools and Resources
- **Module 2:** CLI Primer
- **Module 3:** Flow Logic
- **Module 4:** Packet Captures
- **Module 5:** Packet-Diagnostics Logs
- **Module 6:** Host-Inbound Traffic
- **Module 7:** Transit Traffic
- **Module 8:** System Services
- **Module 9:** SSL Decryption
- **Module 10:** User-ID
- **Module 11:** GlobalProtect
- **Module 12:** Escalation and RMAs

Traps 4.2: Install, Configure, and Manage (EDU-281)

Palo Alto Networks Traps Advanced Endpoint Protection prevents sophisticated vulnerability exploits and unknown malware-driven attacks. Successful completion of this two-day, instructor-led course helps prepare students to install on-premises Traps in basic configurations.

Course Objectives

Students should learn how Traps protects against exploits and malware-driven attacks. In hands-on lab exercises, students will install and configure the Endpoint Security Manager (ESM) and Traps endpoint components, build rules, enable and disable process protections, and integrate Traps with Palo Alto Networks WildFire, which provides protection from known and unknown malware.

Scope

- **Course level:** Introductory

- **Course duration:** 2 days
- **Course format:** Combines lecture and hands-on labs
- **Software version:** Palo Alto Networks Traps Advanced Endpoint Protection ESM 4.2

Target Audience

Security Engineers, System Administrators, and Technical Support Engineers

Prerequisites

Students must have Windows system administration skills and familiarity with enterprise security concepts.

Sessions

- **Module 1:** Traps Overview
- **Module 2:** Installing Traps
- **Module 3:** Malicious Software Overview
- **Module 4:** Consoles Overview
- **Module 5:** Traps Protection Against Exploits
- **Module 6:** Traps Protection Against Malware
- **Module 7:** Prevention Event Exceptions
- **Module 8:** Managing Traps
- **Module 9:** Traps Forensics Capabilities
- **Module 10:** Basic Traps Troubleshooting

Traps 4.2: Deploy and Optimize (EDU-285)

Palo Alto Networks Traps Advanced Endpoint Protection prevents sophisticated vulnerability exploits and unknown malware-driven attacks. Successful completion of this two-day, instructor-led course should prepare the student to deploy on-premises Traps in large-scale or complex configurations and optimize its configuration.

Course Objectives

Students should learn how to design, build, implement, and optimize large-scale Traps deployments: those with multiple servers and/or thousands of endpoints. In hands-on lab exercises, students will distribute Traps endpoint software in an automated way, prepare master images for VDI deployment, create a Traps Linux installation package and install the agent onto a Linux endpoint, build multi-ESM deployments, design and implement customized policies, test Traps with exploits created using Metasploit, and examine prevention dumps with windbg.

Scope

- **Course level:** Intermediate
- **Course duration:** 2 days
- **Course format:** Combines lecture and hands-on labs
- **Software version:** Palo Alto Networks Traps Advanced Endpoint Protection ESM 4.2

Target Audience

Security Engineers, System Administrators, and Technical Support Engineers

Prerequisites

Students should have completed “Traps 4.2: Install, Configure, and Manage” or (for Palo Alto Networks employee and partner SEs) “PSE: Endpoint Associate” training. Windows system administration skills and familiarity with enterprise security concepts also are required. An elementary level of Linux shell experience is needed only for the Linux lab activity.

Sessions

- **Module 1:** Scaling Server Infrastructure
- **Module 2:** Scaling Agent Deployment
- **Module 3:** ESM Tuning
- **Module 4:** Windows Migrations for Traps
- **Module 5:** Advanced Traps Forensics
- **Module 6:** Advanced Traps Troubleshooting

Traps: Cloud Service Operations (EDU-290)

Palo Alto Networks Traps Advanced Endpoint Protection prevents sophisticated vulnerability exploits and unknown malware-driven attacks. Successful completion of this two-day, instructor-led course helps prepare the student to configure the Traps Management Service and to install Traps onto endpoints.

Course Objectives

Students should learn how Traps protects against exploits and malware-driven attacks. In hands-on lab exercises, students will explore and configure new cloud-based Traps Management Service and install Traps endpoint components; build policy rules and profiles; enable and disable process protections; and integrate Traps with Palo Alto Networks WildFire cloud service, which provides prevention and detection of zero-day malware.

Scope

- **Course level:** Introductory
- **Course duration:** 2 days
- **Course format:** Combines instructor-facilitated lecture with hands-on labs
- **Software version:** Palo Alto Networks Traps Advanced Endpoint Protection

Target Audience

Endpoint Security Engineers, System Administrators, and Technical Support Engineers

Prerequisites

Students must have familiarity with enterprise security concepts.

Sessions

- **Module 1:** Traps Overview
- **Module 2:** Cloud Services
- **Module 3:** Cloud- Based Management
- **Module 4:** Policy Rules and Profiles
- **Module 5:** Malware Protection Flow

- **Module 6:** Exploits and Exploitation Techniques
- **Module 7:** Exploit Protection Modules
- **Module 8:** Event Management
- **Module 9:** Basic Traps Troubleshooting
- **Module 10:** Traps Architecture
- **Module 11:** Directory Sync Service

Certifications

Accredited Configuration Engineer (ACE)

The Accredited Configuration Engineer (ACE) exam tests your knowledge of the core features and functions of Palo Alto Networks NGFWs, and serves as an objective indication of your ability to configure Palo Alto Networks firewalls using PAN-OS.

Passing the ACE exam indicates that you possess the basic knowledge to successfully configure Palo Alto Networks firewalls using PAN-OS. The exam also serves as a study aid to prepare for PCNSE certification.

The ACE exam is web-based and consists of 40-50 multiple-choice questions. The exam is not timed, and you can retake it as many times as necessary to earn a passing score.

To pass the ACE exam, you need foundational knowledge of, and hands-on familiarity with, PAN-OS configuration. The best way to gain this knowledge is to take the Firewall 8.1 Essentials: Configuration and Management course. This courseware is available in both instructor-led training and self-paced e-Learning formats.

Palo Alto Networks Certified Network Security Engineer (PCNSE)

A Palo Alto Networks Certified Network Security Engineer (PCNSE) is capable of designing, deploying, configuring, maintaining and troubleshooting the vast majority of Palo Alto Networks-based network security implementations. Passing the PCNSE and exhibiting solid professional behavior, are the requirements for becoming a PCNSE. The formal certification exam is hosted and proctored by the third-party testing company Pearson VUE.

The PCNSE is a formal, third-party proctored certification that indicates those who have passed it possess the in-depth knowledge to design, install, configure, maintain and troubleshoot the vast majority of implementations based on the Palo Alto Networks platform. The PCNSE exam

should be taken by anyone who wishes to demonstrate a deep understanding of Palo Alto Networks technologies, including customers who use Palo Alto Networks products, value-added resellers, pre-sales system engineers, system integrators, and support staff.