Network Security in Virtualized Data Centers for Dummies

Learn to:
- Securely enable applications in your virtualized data center
- Ensure hypervisor integrity and control migrations
- Implement a phased approach to data center security

Brought to you by paloalto NETWORKS
the network security company

Lawrence C. Miller, CISSP
About Palo Alto Networks

Palo Alto Networks™ (NYSE: PANW) is the network security company. Its next-generation firewalls enable unprecedented visibility and granular policy control of applications and content — by user, not just by IP address. Based on patent-pending App-ID™ technology, Palo Alto Networks firewalls accurately identify and control threats and applications — regardless of port, protocol, evasive tactic, or SSL encryption — and scan content to stop threats and prevent data leakage. Palo Alto Networks delivers proven threat prevention combining IPS, anti-malware, URL, file and content filtering in a single integrated engine supported by one of the industry’s most prolific vulnerability research teams.

For more information, visit www.paloaltonetworks.com.
# Table of Contents

## Introduction .......................................................... 1

About This Book .......................................................... 1
Foolish Assumptions ...................................................... 2
How This Book Is Organized ........................................... 2
Icons Used in This Book ................................................ 3
Where to Go from Here .................................................. 4

## Chapter 1: Data Center Evolution ............................... 5

What Is Virtualization? .................................................. 5
Why Server Virtualization? ............................................. 7
Moving to the Cloud ..................................................... 9

## Chapter 2: The Application and Threat Landscape in the Data Center ........................................ 13

Security Challenges with Applications in the Data Center ... 14
Applications Aren’t All Good or All Bad ......................... 14
Applications Are Evasive .............................................. 18
Threats Are Taking a Free Ride ..................................... 19
  Management applications .......................................... 20
  Unknown applications .............................................. 21
Recognizing the Challenges of Legacy Security Solutions .... 22

## Chapter 3: The Life Cycle of a Modern Data Center Attack ......................................................... 23

Data Center Attack Vectors .......................................... 23
Rethinking Malware ..................................................... 25
  Modern malware characteristics .................................. 26
  Botnets and other enterprise threats ............................. 28
Hackers — No Longer the Usual Suspects ....................... 30
The Life Cycle of a Modern Attack ............................... 32
  Infection .................................................................. 33
  Persistence .................................................................. 34
  Communication ......................................................... 35
  Command and control ................................................. 35
Virtualization in the data center has become a powerful engine for driving business growth. Virtualization technologies help organizations utilize their existing hardware infrastructure more effectively, leading to significant cost reductions and improvements in operational efficiencies. Many organizations are now moving beyond basic server and workload consolidation and extending their virtualization infrastructure to build their own private cloud environment.

Yet the very benefits of virtualization — for example, the ability to provide self-service resource activation and improve IT responsiveness to business demands — also introduce a myriad of security complexities. These include having visibility into virtual machine (VM) traffic that may not leave the virtual infrastructure, the ability to tie security policies to VM instantiation and movement, segmentation of virtual machines with different trust levels, and a new attack vector — the hypervisor.

About This Book

Tackling the security implications for a virtualized computing environment is essential for the journey to the cloud. This book outlines the challenges of securing the virtualized data center and cloud computing environments and how to address them with next-generation firewalls.

Virtualization topics cover many technologies, including servers, storage, desktops, and applications, among others. The focus of this book is network security in the virtualized data center — specifically, server virtualization.
Foolish Assumptions

It’s been said that most assumptions have outlived their uselessness, but I’ll assume a few things nonetheless! Mainly, I assume that you know a little something about server virtualization, network security, and firewalls. As such, this book is written primarily for technical readers who are evaluating network security solutions to address modern threats and challenges in virtualized data centers.

How This Book Is Organized

This book consists of twelve voluminous tomes that rival the works of Shakespeare, conveniently distilled into six short chapters and a glossary chock-full of just the information you need. Here’s a brief look at what awaits you!

Chapter 1: Data Center Evolution

The book begins with an overview of what virtualization technology is and explains why it’s such a hot trend that is transforming the modern data center from traditional 3-tier architectures to virtualized data centers and cloud environments.

Chapter 2: The Application and Threat Landscape in the Data Center

This chapter explores and maps the current application and threat landscape. I help you identify applications that are good, bad, and perhaps both good and bad — depending on who’s using them and for what purpose!

Chapter 3: The Life Cycle of a Modern Data Center Attack

This chapter delves into viruses, worms, bots and botnets, and other data center threats that make your systems go bump in the night!
Chapter 4: Securing the Virtualized Data Center

Chapter 4 explores some of the unique challenges associated with virtualization technology, including hypervisor security, intra-host communications, and system migrations.

Chapter 5: A Phased Approach to Security: From Virtualization to Cloud

Here, I walk you through the evolution of the virtualized data center and explain what security solutions you need to deploy along each step of your journey to the cloud.


Finally, in that classic For Dummies format, this book ends with a Part of Tens chapter filled with great information to help you evaluate which network security solutions are best for your virtualized data center.

Glossary

And, just in case you get stumped on a technical term or an acronym here or there, I include a glossary to help you sort through it all.

Icons Used in This Book

Throughout this book, you occasionally see special icons that call attention to important information. You won’t find smiley faces winking at you or any other cute little emoticons, but you’ll definitely want to take note! Here’s what you can expect:
This icon points out information that may well be worth committing to your nonvolatile memory, your gray matter, or your noggin’ — along with anniversaries and birthdays!

You won’t discover a map of the human genome or the secret to the blueprints for the next iPhone here (or maybe you will, hmm), but if you seek to attain the seventh level of NERD-vana, perk up! This icon explains the jargon beneath the jargon and is the stuff legends — well, nerds — are made of!

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

Proceed at your own risk . . . well, okay — it’s actually nothing that hazardous. These helpful alerts offer practical advice to help you avoid making potentially costly mistakes.

**Where to Go from Here**

With our apologies to Lewis Carroll, Alice, and the Cheshire cat:

“Would you tell me, please, which way I ought to go from here?”

“That depends a good deal on where you want to get to,” said the Cat — err, the Dummies Man.

“I don’t much care where . . . ,” said Alice.

“Then it doesn’t matter which way you go!”

That’s certainly true of *Network Security in Virtualized Data Centers For Dummies*, which, like *Alice in Wonderland*, is also destined to become a timeless classic!

If you don’t know where you’re going, any chapter will get you there — but Chapter 1 might be a good place to start! However, if you see a particular topic that piques your interest, feel free to jump ahead to that chapter. Each chapter is individually wrapped (but not packaged for individual sale) and written to stand on its own, so feel free to start reading anywhere and skip around! Read this book in any order that suits you (though we don’t recommend upside down or backwards). I promise you won’t get lost falling down the rabbit hole!

These materials are the copyright of John Wiley & Sons, Inc. and any dissemination, distribution, or unauthorized use is strictly prohibited.
Chapter 1
Data Center Evolution

In This Chapter
▶ Defining Type 1 and Type 2 hypervisors
▶ Improving efficiency and application delivery
▶ Clearing the air about what the cloud is — and what it isn’t

“Doing more with less” — it’s become the mantra for businesses and organizations seeking competitive advantage in a challenging global economy. Today’s IT organizations are no exception. Faced with shrinking budgets and constant pressure to drive operational efficiencies and improve responsiveness, many IT organizations are turning to virtualization in the data center to maximize existing resource utilization, increase IT flexibility, and achieve greater agility to enable key business processes.

This chapter talks about server virtualization and its benefits to the modern enterprise, as well as how organizations evolve along the virtualization journey and move their data centers to the cloud.

What Is Virtualization?

Virtualization technology partitions a single physical server into multiple operating systems and applications, thereby emulating multiple servers, known as virtual machines (VMs).

VMs are also commonly referred to as guests, virtual environments (VEs), and workloads.
The hypervisor — a software layer that sits between the hardware and the “virtual” operating system and applications — allocates memory and processing resources to the “virtual” machines, allowing multiple VMs to run concurrently on a single physical server (also known as a host). The hypervisor functions between the computer operating system (OS) and the hardware kernel.

Two types of server virtualization are available: Type 1 (also known as bare metal or native) hypervisors and Type 2 (also known as hosted) hypervisors (see Figure 1-1). A Type 1 hypervisor is the first layer of software running directly on the underlying hardware without a host operating system. A Type 2 hypervisor runs on top of a host operating system and supports the broadest range of hardware operating system including Windows, Linux, or MacOS.

Virtualization is one of the hottest and most disruptive technologies of the past decade and continues to be so today. Gartner, Inc., estimates that almost 50 percent of all x86 server workloads are virtualized today and that this number will grow to 77 percent by 2015.

Bare metal (Type 1) hypervisor architectures run all applications within a virtualized environment, while hosted (Type 2) hypervisors support applications such as web browsers running alongside the hosted virtualized applications. Because bare metal hypervisors are closer to the hardware resources, they are more efficient — and typically more scalable — than hosted hypervisors. From a security perspective, the lack of dependency on the guest operating system provides one less risk to the overall solution. Server virtualization typically utilizes bare metal hypervisors, while desktop virtualization uses hosted hypervisors.

The rest of this chapter focuses on server virtualization.
Chapter 1: Data Center Evolution

Why Server Virtualization?

The most common business reasons for adopting virtualization in the data center today include improving operational efficiencies and optimization of limited resources to drive greater return on investment (ROI) and lower total cost of ownership (TCO) in existing data center infrastructure.

Data center virtualization initiatives often begin with organizations leveraging their existing hardware infrastructure to consolidate multiple applications within the same system. By consolidating underutilized resources on virtualized systems, organizations are able to shrink their server hardware “box count” and data center footprint. This in turn drives additional business benefits that include:

- Reducing capital expenditures (CAPEX) for new servers
- Lowering operating expenses (OPEX) such as power, cooling, and rack space
- Improving IT flexibility and agility through dynamic provisioning of VMs to rapidly deliver new applications as business needs dictate

Server virtualization allows organizations to consolidate multiple, often-unrelated applications from multiple physical servers to a single physical server that has been virtualized. Virtualization prevents potential interoperability issues between multiple applications running in a mixed environment on a single physical server. Server consolidation has become particularly important for controlling costs and server sprawl in the data center, given the trend within the software industry to design applications that run on dedicated, purpose-built servers in order to optimize performance, improve stability, and simplify support.

I refer to a “single physical server” for virtualization throughout this book to illustrate the benefits of reduced server hardware compared to a nonvirtualized data center with numerous physical servers. However, a virtualized data center typically includes multiple (but fewer) physical servers that are virtualized to provide load balancing and fault tolerance.

Beyond consolidation, server virtualization maximizes the efficient use of underutilized resources within the data center.

These materials are the copyright of John Wiley & Sons, Inc. and any dissemination, distribution, or unauthorized use is strictly prohibited.
Many physical application servers today experience asynchronous or bursty demand loads. For example, an organization’s e-mail system typically sees heavy use during normal business hours but significantly diminished demand after hours. However, that same organization’s backup system may be idle during the normal business day but then peak during the organization’s backup window after hours. Similarly, many applications are characterized by short bursts of computationally intensive processing, followed by extended periods of little or no activity. Virtualization technologies, such as resource schedulers and VM migrations, provide intelligent, automated management of asynchronous and bursty applications to prevent resource contention issues and maximize server utilization.

You need to carefully assess all of your physical server workloads before deciding which applications to virtualize and consolidate.

Server virtualization also provides greater scalability than physical servers. Rather than purchasing newer, bigger, more expensive servers, reinstalling software, and then restoring its associated data when an application outgrows its server infrastructure, virtualization gives organizations several (better) options.

These options include the migration of:

- Other VMs off the physical host server to provide more resources for the application
- The application to a more powerful physical host server to provide the necessary resources

*Migration* is the process of moving a VM from one physical server to another. Migration types include *cold* (for example, when a VM is halted, moved, and then rebooted, and any open sessions or transactions are reset), *warm* (migration of a suspended VM to another host), and *live* (migration of a running VM to another host; the VM continues to operate during the migration with zero downtime).

Server virtualization has also become a key component of organizational disaster recovery strategies. Virtualization enables organizations to quickly migrate their entire server infrastructure to a secondary data center in the event of a disaster or outage.
Finally, server virtualization helps organizations improve operational efficiencies, such as:

- **Reduced downtime due to hardware and software upgrades.** Rather than scheduling upgrades during maintenance windows, IT staff can simply migrate VMs to a different physical server while performing hardware upgrades or create a clone of an existing VM for software upgrades.

- **Standardization of server configurations.** Master server builds can be created and cloned for different deployment scenarios.

- **Flexible, rapid provisioning.** Rather than spending days (or longer) to purchase, “rack and stack,” cable, power, install, and configure new systems, provisioning a VM takes minutes and simply involves configuring storage, selecting and fine-tuning a master OS image, and installing the application.

**Moving to the Cloud**

As enterprise IT needs continue to evolve toward on-demand services, many organizations move beyond data center virtualization to cloud-based services and infrastructure.

The “cloud” is many things to many people. Unfortunately, defining the cloud has become more difficult as the term has become more commercialized and different “cloud service providers” attempt to capitalize on this popular trend.

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 NIST cloud model is composed of five essential characteristics, three service models, and four deployment models.
The five essential characteristics of cloud computing are

- **On-demand self-service.** Computing capabilities (such as server resources) can be unilaterally and automatically provisioned without service provider human interaction.

- **Broad network access.** Services are available over the network through various platforms, such as PCs, laptops, smartphones, and tablets.

- **Resource pooling.** Computing resources (such as processing, memory, storage, and network bandwidth) are dynamically assigned and reassigned according to demand and pooled to serve various customers (multitenancy).

- **Rapid elasticity.** Capabilities can be provisioned and released, in some cases automatically, to scale with demand.

- **Measured service.** Resource usage can be monitored, controlled, optimized, and reported.

Virtualization is the fundamental cloud-enabling technology that delivers the five essential characteristics of the cloud computing model in the virtualized data center.

The three service models defined for cloud computing include

- **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.

- **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.
Infrastructure as a Service (IaaS). Customers can provision processing, storage, networks, and other computing resources and deploy and run operating systems and applications, but 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).

Finally, NIST defines four cloud computing deployment models, as follows:

- **Public.** A cloud infrastructure that is open to use by the general public. It is owned, managed, and operated by a third party (or parties) and exists on the cloud provider’s premises.

- **Community.** A cloud infrastructure that is used exclusively by a specific group of organizations with a shared concern (for example, a mission or business objective). It may be owned, managed, and operated by one or more of the organizations or a third party (or a combination of both), and may exist on or off premises.

- **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 may exist on or off premises.

- **Hybrid.** A cloud infrastructure that is comprised 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).

Cloud service providers such as Amazon (EC2), Microsoft (Azure), Rackspace (Cloud), and VMware (vCloud) provide customers with flexible deployment options for public, community, private, and hybrid cloud infrastructures.
Chapter 2

The Application and Threat Landscape in the Data Center

In This Chapter
▶ Identifying applications as good, bad, or good and bad
▶ Understanding accessibility and evasion tactics
▶ Recognizing how threats use applications as an attack vector

The purpose of a data center is to serve up applications. In the data center, business applications constitute good traffic that should be allowed on the network; other non-business applications that constitute bad traffic should be blocked from the network.

However, the lines between business applications and non-business applications are blurring. Many personal applications such as social media are now being used by enterprises as marketing enablement tools. The ability to classify types of applications as good or bad is no longer a relatively straightforward exercise. Understanding the changes in the application landscape and the threat vector that these applications bring is essential to understanding what types of enablement policies are appropriate for the data center.

This chapter explores the security challenges with application enablement in the data center and dives into the new application landscape.
Security Challenges with Applications in the Data Center

Application developers in the data center are challenged with delivering and supporting hundreds, if not thousands, of applications to employees. As business needs evolve, applications continue to be developed and improved to meet specific user-community needs. These applications can range from enterprise off-the-shelf applications to custom and home-grown applications.

The challenge for security organizations is keeping up with these application developers. In many cases, to expedite delivery of these applications, developers have been known to implement applications on any port that is convenient, or bypass security controls altogether. When they create security backdoors to manage these applications from home or on the road, these can become avenues for attackers to infiltrate.

The ease of application creation and delivery due to virtualization technologies exacerbates the problem. This creates a paradox for security teams that are forced to either be a barrier to business growth by enforcing strict security controls for application delivery or to become helpless to manage security risks in the face of application proliferation.

Rather than attempt to control application developers, the answer lies in controlling the applications. In order to adopt secure application enablement firewall policies, having full and comprehensive visibility into all applications on your network is essential. Understanding how the application landscape has changed is also critical to determine which applications carry threats and whether they should be authorized.

Applications Aren’t All Good or All Bad

Applications in the data center can largely be divided into:
Corporate-supported applications — enterprise off-the-shelf, custom, and home-grown
Management applications using RDP, Telnet, and SSH to control the enterprise applications
Rogue or misconfigured applications such as peer-to-peer applications for personal use within the data center

The first set of applications described in the preceding list should be allowed for authorized employees, the second set should be enabled only for a select group of IT users, and the third set should be remediated or dropped.

This seems simple enough. However, over the past decade, the application landscape has changed dramatically for organizations. Corporate productivity applications have been joined by a plethora of personal and consumer-oriented applications. This convergence of corporate infrastructures and personal technologies is being driven by two significant trends: Bring Your Own Device (or BYOD) and consumerization.

The BYOD trend has taken hold in corporate networks as businesses and organizations are increasingly allowing their employees to use their personal mobile devices — such as smartphones and tablets — in the workplace, for both personal and work-related use.

BYOD isn’t just an endpoint challenge. It becomes a data center issue when these personal devices are used to access corporate applications.

The process of consumerization occurs as users increasingly find personal technology and applications that are more powerful or capable, more convenient, less expensive, quicker to install, and easier to use than corporate IT solutions. These user-centric “lifestyle” applications and technologies enable individuals to improve their personal efficiency, handle their nonwork affairs, maintain online personas, and more. Common examples include Google Docs, instant messaging applications, and web-based e-mail.

Enterprise 2.0 applications highlight the dissolution of the traditional distinctions between business and personal use. More often than not, the same applications used for social interaction are being used for work-related purposes. And, as
the boundary between work and their personal lives becomes less distinct, users are practically demanding that these same tools be available to them in their workplace.

The adoption of Enterprise 2.0 applications is being driven by users, not by IT. The ease with which they can be accessed, combined with the fact that today’s knowledge workers are accustomed to using them, points toward a continuation of the consumerization trend. Defined by Appopedia (www.theappgap.com) as “a system of web-based technologies that provide rapid and agile collaboration, information sharing, emergence and integration capabilities in the extended enterprise,” Enterprise 2.0 applications have taken the world by storm. What started as a few applications that were mostly focused on searching, linking, and tagging, rapidly shifted to a horde of applications that enable authoring, networking, and sharing, among other things.

In the data center, examples of Enterprise 2.0 applications can vary widely, from content management tools (such as SharePoint) to collaborate internally or with external business partners, to complex social networks and posting tools (such as Facebook and Twitter) for marketing outreach programs.

Unsure of how to leverage the BYOD and consumerization trends in their business processes, many organizations take one of two approaches. They either implicitly allow these personal technologies and Enterprise 2.0 applications by simply ignoring their use in the workplace, or they explicitly prohibit their use, which then makes them unable to effectively enforce such policies with traditional firewalls and security technologies. Neither of these two approaches is ideal, and both incur inherent risks for the organization. In addition to lost productivity, adverse issues include

- Creating a subculture of back-channel or underground workflow processes that are critical to the business’s operations but are known only to a few users and fully dependent on personal technologies and applications.

- Introducing new risks to the entire networking and computing infrastructure. This is due to the presence of unknown, and therefore unaddressed and unpatched, vulnerabilities, as well as threats that target normal application and user behavior — whether a vulnerability exists in the application or not.
Being exposed to noncompliance penalties for organizations that are subject to regulatory requirements such as Financial Industry Regulatory Authority (FINRA), Health Insurance Portability and Accountability Act (HIPAA), North American Electric Reliability Corporation Critical Infrastructure Protection (NERC CIP), and Payment Card Industry Data Security Standard (PCI DSS).

Having employees circumvent controls with external proxies, encrypted tunnels, and remote desktop applications, making it difficult, if not impossible, to manage the risks.

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

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

Translated into real-world examples in the data center, secure application enablement policies might include allowing

- IT staff to use a fixed set of remote management applications (such as SSH, RDP, and Telnet) across their standard ports but blocking their use for all other users.
- Streaming media applications by category, but applying QoS policies to limit their impact on business VoIP applications.
- The marketing team to use a social networking application such as Facebook to share product documentation with customers, while allowing read access by other users in the organization but blocking posting access.
Today’s network security solution in the data center, therefore, must be able not only to distinguish one type of application from the next, but also to account for other contextual variables surrounding its use and to vary the resulting action that will be taken accordingly.

**Applications Are Evasive**

Although “distinguishing one type of application from the next” sounds simple, it really isn’t — for a number of reasons. In order to maximize their accessibility and use, many applications are designed from the outset to circumvent traditional firewalls by dynamically adjusting how they communicate. For the end-user, this means an application can be used from anywhere, at any time.

Common evasion tactics include

- **Hiding within SSL encryption**, which masks the traffic, for example, over TCP port 443 (HTTPS)
- **Port hopping**, where ports/protocols are randomly shifted over the course of a session
- **Tunneling within commonly used services**, such as when P2P file sharing or an IM client runs over HTTP
- **Use of nonstandard ports**, such as running SSH on ports other than port 22

SSL is commonly viewed as a means of encrypting traffic to keep it secure. Financial transactions, healthcare records, retail purchases, and strategic collaboration are some common examples of SSL usage in organizations. In these cases, SSL is used to protect sensitive content from data theft. But in other cases, SSL is used merely as a means to evade detection.

Dynamic applications that can hop ports eliminate barriers to access and are relatively easy to use — wherever the user is. The slippery nature of applications that can hop ports means that organizations will continually struggle to identify and control them. From a security perspective, many of these applications are known to have vulnerabilities and can act as a malware vector (discussed in Chapter 3). The business risks
include whether they are “approved for use,” and many (P2P file-sharing applications, in particular) introduce the potential risk of loss of confidential data.

Applications that can tunnel other applications, for good or bad, expand far beyond the traditional view of SSH-, SSL-, and VPN-related applications. One ironic example of this type of application is web browsing. Many years ago, anti-malware vendors began using TCP port 80 to update their pattern engines quickly and easily. To most security infrastructure components, this traffic appears as if it is web browsing.

Finally, applications can no longer be characterized by the ports they use. In the past, applications were identified by standard TCP and UDP ports. E-mail would typically flow through port 25, FTP was assigned to port 20, and “web surfing” was on port 80. Everybody played by the rule that “ports + protocols = applications.” Nice and simple. Many applications now exist that insist on making their own rules by using nonstandard ports as a basic evasion tactic.

Many standard client-server applications are being redesigned to take advantage of web technologies. Enterprises are increasingly embracing cloud-based web services such as Salesforce.com and WebEx — which often initiate in a browser but then quickly switch to more client-server behavior (rich client, proprietary transactions, and others).

Many new business applications also use these same techniques to facilitate ease of operation while minimizing disruptions for customers, partners, and the organization’s own security and operations departments. For example, RPC and Sharepoint use port hopping because it is critical to how the protocol or application (respectively) functions, rather than as a means to evade detection or enhance accessibility.

**Threats Are Taking a Free Ride**

The increasing prevalence of application-layer attacks is yet another disturbing trend. Threats that directly target applications can pass right through the majority of enterprise defenses, which have historically been built to provide network-layer or port-based protection. Threat developers exploit the same methods (described in the previous section)
to infiltrate networks that application developers utilize to promote ease of use and widespread adoption, such as tunneling within applications.

The evasion techniques built into these and many other modern applications are being leveraged to provide threats with “free passage” into enterprise networks. It is no surprise, therefore, that greater than 80 percent of all new malware and intrusion attempts (discussed in Chapter 3) are exploiting weaknesses in applications, as opposed to weaknesses in networking components and services. Together with the implicit trust that users place in their applications, all of these factors combine to create a “perfect storm.”

Management applications

The weakest security link in data centers is often what’s used to manage them. Management application such as SSH, RDP, Telnet, and VNC enable applications to be easily accessed and managed from anywhere, at any time, but can also serve as a threat vector for the data center. IT administrators and application developers require access to the applications, and occasionally these backdoors are either mistakenly left open or proper access control rules are not enforced.

What’s interesting is that these types of applications are not only being used by IT, but also by sophisticated employees who want to access their home machine — or someone else’s — while they are at work. These intrepid users are accessing their machines and, in the process, exposing themselves and their company to numerous business and security risks.

According to analysis in the December 2011 Palo Alto Networks Application Usage and Risk Report, an average of eight remote-access applications were found in 96 percent of organizations that participated in the study. When viewed across the past two years of data collected and analyzed, the top five remote access tools have remained consistent in terms of the frequency of usage and include RDP, Teamviewer, Logmein, Telnet, and Citrix.

A recent Verizon Data Breach report analyzed 900 incidents worldwide and found that 320 of the initial penetrations could
be tracked back to remote access errors. Attackers actively scan for these open backdoors, and when any backdoors are found in a vendor’s application, attacks are quickly extended to the vendor’s customers and business partners.

The impact from a breach can be catastrophic. Exposed applications could be vulnerable to brute-force password attacks, or the initial access level could enable control of the entire application and become a stepping stone to other intrusions in the enterprise. Remote desktop and management applications must be properly controlled in the data center, and access control must be strictly enforced.

**Unknown applications**

Being able to monitor, manage, and control a known application is one thing, but not every application on a network is known and instantly recognized. Most companies accept a variant of what is known as the 80:20 rule; most of the traffic is known, but the rest, which is a small amount, is unknown.

Most unknown applications fall into one of three categories:

- An internal home-grown application
- A commercial application that hasn’t yet been identified
- A potential threat

A network security solution in the data center should be able to identify unknown traffic and drill down into specific communications and logs to understand the threat impact. Once home-grown applications and commercial applications not previously identified have been characterized and appropriate security policies implemented, it is a reasonable assumption that any other unknown is likely a potential threat.

While a data center with zero unknown applications may be a challenging goal, recognition and active management of the unknowns will go a long way toward reducing the risks of application-enabled threats in the data center.
Recognizing the Challenges of Legacy Security Solutions

As discussed earlier, the application landscape has changed. Corporate applications hosted in the data center include a variety of applications that can all exhibit a variety of characteristics, from port hopping to tunneling.

Organizations need firewall policies that understand business-relevant elements such as application identity, user identity (who is using the application), and the types of content or threats embedded within the application.

Using business-relevant elements, you can transform your traditional “allow or deny” firewall policy into a secure application enablement policy. This means more than allowing only what you expressly define and blocking everything else. It means you can build firewall policies that are based on application/application feature, users and groups, and content, as opposed to port, protocol, and IP address.

This is specifically why traditional network security solutions are not effective in the data center. Port-based rules may allow other applications that should not be allowed in the data center. The strict adherence to relying on port as the initial classification mechanism means that applications directed over nonstandard ports are missed completely, introducing unnecessary business and security risks. Finding tech-savvy employees using remote access tools on nonstandard ports is not uncommon.

To implement application control, legacy security vendor solutions require that you first build a firewall policy with source, destination, user, port, and action (for example, allow, deny, drop, log). Then, to control applications, you move to a different configuration tab or a separate management application and duplicate information from the firewall policy, adding application and action. Maintaining and reconciling even a small set of firewall and application control policies is challenging. Most medium to large organizations have hundreds — even thousands — of firewall rules, and the multiple policy rulebase approach not only increases the administrative overhead, but it also increases both business and security risks.
The attack vector for the data center has expanded significantly. As the enterprise becomes more distributed, empowered users and an ecosystem of partners, contractors, and customers now require access to the data center — and can introduce potential compromise to data center security.

While physical intrusions and direct attacks such as distributed denial-of-service attacks on data centers continue to be a problem, the modern attack strategy is now a patient, multi-step process that takes advantage of a variety of different threat techniques to penetrate the network.

This chapter explains different types of attacks against the data center and dives into the rise of modern malware, the role and motives of today’s hackers, and the threat/attack life cycle.

Data Center Attack Vectors

The most common Hollywood premise of a data center attack is the physical intrusion in which attackers accomplish the impossible and gain interior access to a data center to disable...
specific servers or retrieve proprietary information. Today, data centers are located in remote regions, armed with the best physical security systems with their locations hand-selected based on the propensity to be able to handle natural and man-made disasters. Physical perimeter security incorporates not only chain-link fences, armed guards, and advanced video surveillance systems, but also sophisticated physical access control including biometrics systems. Actual physical access is also limited to only key personnel. Therefore, the reality of an actual physical attack to the data center, while not impossible, is highly unlikely.

The more common scenario is thus the cyber attack. In Internet data center environments (see Chapter 4 for more details on Internet versus enterprise data centers), the most common type of attack is the denial-of-service attack. Denial-of-service (DoS) attacks generate large volumes of traffic that consume server or network resources such as CPU and memory and, in the process, take these limited resources away from legitimate traffic. A distributed denial-of-service (DDoS) attack floods a target server or network with traffic from a large number of infected endpoints rendering the server or network unresponsive or otherwise unusable.

DDoS campaigns are commonly used by hacktivists to embarrass or otherwise disrupt a target company or government agency. Botnets controlled by criminal groups can recruit thousands and even millions of infected machines to join in a truly global DDoS attack, enabling the gang to essentially extort a ransom from the target network in exchange for stopping the attack.

Internet-facing data center attacks, such as script kiddie attacks, tend to be more automated. Script kiddies are attackers who are not directly targeting an organization but are looking for an easy way to leverage known vulnerabilities by scanning the Internet. Common exploits and vulnerabilities that haven’t been patched in an organization become the initial entry point for attack.

Although script kiddies don’t directly target an organization, they are still dangerous because they are able to leverage the vast amounts of exploit knowledge already accumulated by others to launch an attack.
In an enterprise data center environment (see Chapter 4), sophisticated modern malware attacks are more common. Modern malware has outpaced traditional anti-malware strategies and, in the process, has established a foothold within the enterprise that criminals and nation-states can use to steal information and attack sensitive assets. In particular, initial compromise of a user or asset ultimately leads to a data center breach as information within the data center is what holds the most promise of financial gain for these attackers.

The rest of this chapter talks about the new threat landscape shaped by modern malware.

**Rethinking Malware**

Attack techniques have become more sophisticated over the past several years, and malware is now a major weapon in the hacker’s arsenal. According to Verizon’s 2012 Data Breach Investigations Report, 69 percent of all breaches last year incorporated malware in the attack. New methods for delivering malware include drive-by-downloads and encrypting malware communications to avoid detection by traditional signature-based anti-virus software.

*Malware* is malicious software or code that typically damages or disables, takes control of, or steals information from a computer system. Malware broadly includes botnets, viruses, worms, Trojan horses, logic bombs, rootkits, bootkits, backdoors, spyware, and adware. See the glossary for definitions of these various types of malware.

Modern malware is somewhat like the pea in a shell game. A street con running a shell game on the sidewalk lures the mark (or victim) into trying to follow the pea, when actually it’s an exercise in sleight of hand (see Figure 3-1).

Similarly, the modern threat life cycle relies on sleight of hand — how to infect, persist, and communicate without being detected. Unfortunately, our traditional view of malware and old security habits make us think of malware as the pea — an executable payload, perhaps attached to an e-mail. To understand, control, and successfully counter modern threats, we need to focus on not just the pea (malware) but on all the moving parts.
Modern malware characteristics

Organizations and computer users have been dealing with various types of malware for many years. Unfortunately, industry solutions to combat malware are not necessarily keeping pace with new threats. An alarming percentage of active malware “in the wild” goes undetected, and a recent NSS Labs study (www.nsslabs.com) found that malware protection in general ranges between 54 and 90 percent, giving cybercriminals a 10 to 45 percent chance of getting past your defenses with malware and a 25 to 97 percent chance of compromising your systems using exploits.

This poor effectiveness can be attributed to several factors. For example, some malware can mutate or be updated to avoid detection. Additionally, malware can increasingly be customized to target a specific individual or network.

Botnets illustrate many of the unique characteristics of modern malware. Bots (individual infected machines) and botnets (the broader network of bots working together) are notoriously difficult for traditional antivirus/anti-malware solutions to detect. Bots leverage networks to gain power and resilience. A bot under the remote control of a human attacker (or bot-herder) can be updated — just like any other application — so that the attacker can change course and dig deeper into the network.

Early types of malware operated more or less as swarms of independent agents that simply infected and replicated themselves. Botnets, in comparison, and a great deal of
modern malware, essentially function as centrally coordinated, networked applications. In much the same way that the Internet has changed what is possible in personal computing, ubiquitous network access is changing what is possible in the world of malware. Now, similar types of malware can work together against a common target, with each infected machine expanding the power and destructiveness of the overall botnet. The botnet can evolve to pursue new objectives or adapt to changes in security countermeasures.

Some of the most important and unique functional traits of botnets (see Figure 3-2) are discussed in the following sections.

![Figure 3-2: Key characteristics of botnets.](image)

### Fault-tolerant and distributed
Modern malware takes full advantage of the Internet’s resilient design. A botnet can have multiple control servers distributed anywhere in the world, with numerous fallback options. Bots can also potentially leverage other infected bots as communication channels, providing them with a nearly infinite number of communication paths to adapt to changing access options or to update their code as needed.

### Multifunctional
Updates from the botnet’s command-and-control servers can also completely change the bots functionality. This multifunctional capability enables a bot-herder (botnet operator)
to use portions of the botnet for a particular task such as collecting credit card numbers, while other segments of the botnet might be sending spam. The important point is that the infection is the most important step because the functionality can always be changed later as needed.

**Persistent and intelligent**

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

An APT is a sustained Internet-borne attack usually perpetrated by a group of individuals with significant resources, such as organized crime or a rogue nation-state.

**Botnets and other enterprise threats**

Botnets are a major threat to organizations due to their ability to evade traditional security measures and their practically limitless functionality — from sending spam to stealing trade secrets. A botnet that is sending spam one day could be stealing credit card data the next.

**Spamming botnets**

Some of the largest botnets primarily send spam. A bot-herder infects as many computers as possible, which can then be used without the user’s knowledge to send out thousands of spam messages. Some of the worst spamming botnets are capable of sending thousands of spam messages every hour from each infected computer. This type of botnet affects not only the performance of the infected computer but the network it is attached to as well.
**DDoS and botnets**

A slight twist on spamming botnets uses bots as part of a DDoS attack. In such cases, the infected machine(s) is often not the target of the attack itself. Instead, the infected machine(s) is used to attack some other remote target, such as another organization or network, with traffic. The bot-herder leverages the massive scale of the botnet to generate traffic that overwhelms the network and server resources of the target. DDoS attacks often target specific companies for personal or political reasons or to extort payment from the target in return for stopping the DDoS attack.

DDoS botnets represent a dual risk for the enterprise. The enterprise itself can potentially be the target of a DDoS attack, resulting in downtime and lost productivity. Even if the enterprise is not the ultimate target, any infected machines participating in the attack will consume valuable computer and network resources and facilitate a criminal act, albeit unwittingly.

**Financial botnets**

Financial botnets can cause significant monetary damage to individuals and organizations. These botnets are typically not as large and monolithic as spamming botnets, which grow as large as possible for a single bot-herder. Instead, financial botnets are often sold as kits that allow large numbers of attackers to license the code and set about building their own botnets and targets.

The smaller size of these botnets helps them evade detection for as long as possible in order to steal as much as possible. Even with their smaller size, the impact of these botnets can be enormous. The breach of customer credit card information, for example, can lead to serious financial, legal, and brand damage, and the enterprise could lose money that potentially may never be recovered.

**Targeted intrusions**

Botnets are also used extensively in targeted, sophisticated, and ongoing attacks against specific organizations. Instead of attempting to infect large numbers of machines to launch malicious large-scale attacks, these smaller botnets compromise specific high-value systems that can be used to further penetrate and intrude into the target network. In these cases,
an infected machine can be used to gain access to protected systems and to establish a backdoor into the network in case any part of the intrusion is discovered (I talk more about the life cycle of an attack later in this chapter).

These types of threats almost always evade detection by antivirus software. They represent one of the most dangerous threats to the enterprise because they specifically target an organization’s most valuable information, such as research and development, intellectual property, strategic planning, financial data, and customer information.

Modern malware depends on the enterprise network in order to survive. In the truest sense, modern malware consists of networked applications that are uniquely designed to evade traditional security solutions. To detect and stop these threats, security teams need to regain full visibility into network and data center traffic, reduce the exposure of the network and user, and establish new techniques to detect and prevent malware.

**Hackers — No Longer the Usual Suspects**

Hackers today have evolved into bona fide cybercriminals, often motivated by significant financial gain and sponsored by criminal organizations, nation-states, or radical political groups. Today’s hacker fits the following profile:

- Has far more resources available to facilitate an attack
- Has greater technical depth and focus
- Is well funded and better organized

Why is it important to understand who hackers are and what motivates them? Because a hacker sitting in his parents’ basement may be able to break into a corporate network and snoop around, but he doesn’t necessarily know what to do with, say, intellectual property or sensitive personnel data. On the other hand, a rogue nation-state or criminal organization knows all about extortion and exactly what to do or who to sell stolen intellectual property to on the gray or black market.
According to Verizon’s 2012 Data Breach Investigations Report, 97 percent of all external breaches last year were motivated by financial or personal gain. Additionally, criminal organizations and nation-states have far greater financial resources than do independent hackers. Many criminal hacking operations have been discovered, complete with all the standard appearance of a legitimate business with offices, receptionists, and cubicles full of dutiful hackers. These are criminal enterprises in the truest sense, and their reach extends far beyond that of an individual hacker. Hackers today are focused on stealing valuable information. Consequently, it isn’t in a hacker’s best interests to devise threats that are “noisy” or that are relatively benign. To be successful, a hacker must be fast, or stealthy — or both.

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

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

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

Targeted attacks against specific organizations or individuals are another major concern. In this case, hackers often develop customized attack mechanisms to take advantage of the specific equipment, systems, applications, configurations, and even personnel employed in a specific organization or at a given location. According to Verizon’s 2012 Data Breach Investigations Report, 98 percent of data breaches resulted from external agents — a sharp increase from the 70 percent attributed to external agents just two years earlier!

**The Life Cycle of a Modern Attack**

As with hackers and their motives, the modern attack strategy has also evolved. Instead of directly attacking a high-value server or asset, today’s strategy employs a patient, multi-step process that blends exploits, malware, and evasion into a coordinated network attack (see Figure 3-3).

As an example, an attack often begins by simply luring an individual into clicking on an infected website link on a web page or in an e-mail. The resulting page remotely exploits the...
individual's computer and downloads malware to the user’s computer in the background. The malware then acts as a control point inside the network, allowing the attacker to further expand the attack by finding other assets in the internal network, escalating privileges on the infected machine, and/or creating unauthorized administrative accounts — just to name a few tactics.

Instead of malware and network exploits being separate disciplines as they were in the past, they are now integrated into an ongoing attack. Malware, which is increasingly customized to avoid detection, provides a remote attacker with a mechanism of persistence, and the network enables the malware to adapt and react to the environment it has infected. Key components of the modern attack strategy include infection, persistence, communication, and command and control (see Figure 3-4).

**Figure 3-4**: Key components and tools in the modern attack strategy.

### Infection

Infection almost always has a social aspect, such as getting users to click on a bad link in a phishing e-mail, luring them to a social networking site, or sending them to a web page with an infected image, for example. Understanding how malware
and exploits have become closely interrelated in the modern attack life cycle is important. Exploits used to be directed at vulnerabilities on target servers. Most exploits today are used to crack a target system to infect it with malware: an exploit is run, causing a buffer overflow, which allows the attacker to gain shell access.

With shell access, the attacker can deliver pretty much any payload desired. The first step is to exploit the target and then deliver the malware in the background through the application or connection that is already open. This is known as a *drive-by-download* and is far and away the most common delivery mechanism for modern malware today.

Infection relies heavily on hiding from and evading traditional security solutions. Targeted attacks will often develop new and unique malware that is customized specifically for the target network. This technique allows the attacker to send in malware knowing that it is unlikely to be detected by traditional antivirus tools. Another common way to avoid security measures is to infect the user over a connection that security tools can’t see into, such as an encrypted channel. Attack transmissions are often obscured in SSL-encrypted (Secure Sockets Layer) traffic or other proprietary encryption used in P2P (peer-to-peer) networking applications and IM (instant messaging), for example.

Threats today do not necessarily come as an executable attachment in an e-mail. A link is all that is required. This is why social media, webmail, message boards, and microblogging platforms such as Twitter are rapidly becoming favorite infection vectors for attackers.

**Persistence**

After a target machine is infected, the attacker needs to ensure *persistence* (the resilience or survivability of the bot). Rootkits and bootkits are commonly installed on compromised machines for this purpose.

Backdoors enable an attacker to bypass authentication mechanisms in order to access a compromised system. Backdoors are often installed as a failover in case other malware is detected and removed from the system. Finally, anti-AV malware may be installed to disable any legitimately
installed antivirus software on the compromised machine, thereby preventing automatic detection and removal of malware that is subsequently installed by the attacker. Many anti-AV programs work by infecting the Master Boot Record (MBR) of a target machine.

**Communication**

Communication is fundamental to a successful attack. Malware must be able to communicate with other infected systems or controllers to enable command and control and to extract stolen data from a target system or network. Attack communications must be stealthy and cannot raise any suspicion on the network. Such traffic is usually obfuscated or hidden through techniques that include

- **Encryption** with SSL, SSH (Secure Shell), or some other custom application. Proprietary encryption is also commonly used. For example, BitTorrent is known for its use of proprietary encryption and is a favorite hacker tool — both for infection and ongoing command and control.

- **Circumvention** via proxies, remote desktop access tools (such as LogMeIn!, RDP, and GoToMyPC), or by tunneling applications within other (allowed) applications or protocols.

- **Port evasion** using network anonymizers or port hopping to tunnel over open ports. For example, botnets are notorious for sending command-and-control instructions over IRC (Internet Relay Chat) on nonstandard ports.

- **Fast Flux (or Dynamic DNS)** to proxy through multiple infected hosts, reroute traffic, and make it extremely difficult for forensic teams to figure out where the traffic is really going.

**Command and control**

Command and control rides on top of the communication platform that is established. Its purpose is to ensure that the malware or attack is controllable, manageable, and updatable. Command and control is often accomplished through common applications including webmail, social media, P2P networks, blogs, and message boards. Command-and-control traffic
doesn’t stand out or raise suspicion, is often encrypted, and frequently makes use of backdoors and proxies.

TDL-4: The indestructible botnet

In mid-2011, security researchers began tracking a new version of the TDL botnet, which is alternatively known as TDSS or Alureon. This new variant — TDL-4 — has built-in mechanisms that protect the botnet from a traditional decapitation takedown, such as Microsoft’s takedown against the Rustock botnet in early 2011. These “features” have led some in the security industry to label TDL-4 as “indestructible.” With TDL-4, as with most modern malware, the threat is more about the framework than the actual payload or application.

TDL-4 is primarily spread through affiliates — often pornographic, piracy (software, movie, and music), and video/file-sharing websites — that are paid as much as $200 USD for every 1,000 computers that they infect.

Persistence is achieved through installation of a bootkit that infects the Master Boot Record (MBR) of the victim machine and more than 20 additional malware programs, including fake antivirus programs, adware, and a spamming bot. Very cleverly, TDL-4 actually removes approximately 20 common malware programs — such as Gbot and ZeuS — to avoid drawing unwanted attention to a victim computer when legitimately installed antivirus software detects these common malware programs on the computer!

Communications are concealed using proprietary encryption that is tunneled within SSL. TDL-4 can also install a proxy server on an infected machine, which can then be rented out as an anonymous browsing service that proxies traffic through numerous infected machines. That’s right! You’re familiar with Software as a Service (SaaS), Infrastructure as a Service (IaaS), and Platform as a Service (PaaS) — get ready for Malware as a Service (MaaS)!

For command and control, TDL-4 uses the Kad P2P network, a publicly accessible P2P file exchange network. TDL-4 updates and distributes information about infected machines over the Kad network, so that even if a command-and-control server is taken down, other infected bots can be found to maintain the botnet — without command-and-control servers.
Chapter 4

Securing the Virtualized Data Center

In This Chapter
▶ Recognizing security challenges in the data center
▶ Distinguishing between enterprise and Internet-facing data centers
▶ Getting real about network security in virtual environments

In principle, data center network security is relatively straightforward — prevent threats and comply with regulations and enterprise policies, all without hindering business. In practice, however, the ever-increasing demands for application availability and performance, the constantly evolving threat landscape, and the need to understand what is happening with applications from a security perspective combine to make data center network security requirements much more difficult to meet.

Compounding the issue, the advent of the virtualized data center introduces new security challenges such as hypervisor integrity, intra-host communications, and VM migration.

In this chapter, you learn about network security challenges in the virtualized data center and how to address them with next-generation firewalls.

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

Data center network security traditionally lags perimeter network security as application availability and performance typically trump security in most organizations. If an application hosted in a data center isn’t available or responsive, network security controls, which all too often introduce delays and outages, are typically “streamlined” out of the data center design. Many organizations have been forced into significant compromises — trading security, function, and visibility for performance, simplicity, and efficiency. Data center network security trade-offs include

- Performance or security
- Simplicity or function
- Efficiency or visibility

These compromises are often “hardwired.” For example, an organization with an Internet-facing data center may have to choose between performance and security in its equipment choice: a service provider-class firewall with ample performance capacity but limited security functionality, or an enterprise-class firewall with plenty of security functionality but lower performance capacity and fewer reliability features. The problem is, once an organization chooses, it is often stuck — new designs and new products have to be implemented to shift the balance between performance and security.

Major demands in data center network security include

- Prevent threats
- Comply and compartmentalize
- Maintain application performance and availability

Preventing threats has become more difficult in the last several years (refer to Chapters 2 and 3). Basic attacks on the infrastructure have given way to multivector, application-borne, sophisticated attacks that are stealthy, profit driven, unwittingly aided by enterprise users, and in many cases, polymorphic. The level of organization associated with the development of these threats is also unprecedented.

These materials are the copyright of John Wiley & Sons, Inc. and any dissemination, distribution, or unauthorized use is strictly prohibited.
Regulatory and compliance requirements — such as PCI’s Data Security Standards (DSS), U.S. healthcare mandates, and European privacy regulations — are pushing network segmentation deeper into organizations generally, and into data centers specifically.

Finally, maintaining performance and availability usually translates to simplicity. Needless complexity can introduce additional integration issues, outages, and latency. Keeping the data center design and architecture simple is essential.

**Not All Data Centers Are Created Equal**

Internal enterprise data centers have very different missions and security requirements from Internet-facing data centers. Application and user characteristics, regulatory requirements, and additional, unique security concerns all vary between these two types of data center.

**Enterprise data centers**

Enterprise data centers typically host more applications but have fewer users than Internet-facing data centers (discussed in the next section). Applications might be packaged, home-grown, or customized, browser-based or client/server, and terminal-based or virtualized. Users are typically known to the organization and include employees, contractors, and partners.

Unique network security issues for enterprise data centers include

- Keeping pace with application developers
- Segmenting the network appropriately (typically driven by compliance)
- Ensuring that network security “fits” into a variety of data center designs
- Controlling the surge of “rogue” applications
Network security in the enterprise data center is often based on network segmentation. Although any firewall can be used to segment a network, port- and IP-address-based segmentation is as meaningless in the data center as it is at the perimeter — practically worthless in the face of an application and threat mix that can, for the most part, use any open port. Furthermore, controlling access by IP address or IP address pool is an equally poor approximation for users. What’s needed is network segmentation by user and application.

For example, an organization might segregate servers containing sensitive credit card data and only permit access to that segment by employees who require access to the payment application — thus containing and restricting access while maintaining individual accountability.

Another key attribute of enterprise data centers is the diversity of architectures. For many organizations, an “internal data center” may be comprised of several data centers in separate geographic locations throughout the world, or incorporate a variety of different networking designs.

These diverse data center designs typically make extensive use of VLANs and distributed application components, which means that the usual stack of routers, switches, and other networking equipment can look quite different in the modern data center. The ability to integrate a firewall at OSI layer 1, 2, or 3, as well as the ability to trunk VLANs, aggregate ports, and perform role-based administration across security zones and virtual firewalls, enables flexibility to support any data center architecture or operational model.

Another issue in enterprise data centers is a surge in “rogue” applications. Whether it’s a rogue SharePoint installation or an administrator using SSH on nonstandard ports, these applications need to be identified and controlled on data center networks. The ability to accurately identify and control applications enforces access control policies and helps reduce the attack surface in the data center.

**Internet-facing data centers**

Compared to enterprise data centers, there are relatively fewer applications in an Internet-facing data center, and those applications are usually web (browser-based) applications.
Often, these applications will use one of the common web infrastructure “stacks” (such as IBM, LAMP, Microsoft, or Oracle). There are many users, and they are often unknown or otherwise “untrusted.”

Network security issues for Internet-facing data centers are typically related to security and performance. They include

- Maintaining performance while delivering threat protection
- Design integration into data center architectures
- Visibility into applications and what ports are being used

For example, determining where to place the IPS — a common performance choke point — in order to not impact throughput can be a challenge. And, should the IPS be deployed inline or in alert-only promiscuous mode?

Like enterprise data centers, network security integration must also be considered. Integration of the network security solution in a flexible manner allows additional threat prevention capabilities to be enabled without re-architecting the network as the security posture changes. It also enables security administrators to consider the simplest design possible without negatively affecting data center performance.

Finally, hand-in-hand with threat prevention capabilities are application visibility and control. Internet-facing data centers host fewer applications than enterprise data centers, but they are at greater risk from Internet-based attacks. Visibility into application and threat traffic and the ability to review usage logs will help prevent threats and ensure appropriate use.

**Virtualized Data Center Network Security Challenges**

The security issues in a traditional data center — performance, modern threats, access control, networking integration, and high availability — are also applicable in a virtualized data center. But a “virtual infrastructure” also introduces new elements in the data center architecture that are vulnerable and therefore require additional security considerations. A virtualized computing environment consists of many different
components, including hypervisors (refer to Chapter 1), guest operating systems, and applications that must be secured to ensure protection for the entire virtualized data center.

The major network security challenges in the virtualized data center include

- **Hypervisor integrity.** A successful attack against a host’s hypervisor can compromise all of the workloads being delivered by the host.

- **Intra-host communications.** Communications traffic between different VMs on the same physical host is often not visible and therefore cannot be controlled by traditional physical firewalls and IPS.

- **VM migration.** When VMs migrate from one physical host to another or from one physical site to another, they tend to break network security tools that rely on physical and/or network-layer attributes.

## Securing the Data Center with Next-Generation Firewalls

Today’s application and threat landscape, described in Chapter 2, renders traditional port-based firewalls and other security solutions largely ineffective at protecting the two types of data centers described earlier in this chapter.

Next-generation firewalls provide key differentiating features to uniquely address the traditional trade-offs between security and other critical requirements, such as performance, flexible integration, and visibility of traffic.

## Visibility and control of traffic

A next-generation firewall performs a true classification of data center traffic, based not simply on port and protocol but on an ongoing process of application analysis, decryption, decoding, and heuristics as well. These capabilities progressively peel back the layers of a traffic stream to determine its true application identity.
By understanding the types of applications running in the data center, you can granularly control the applications that are allowed. And within those applications, you can control the behaviors that are allowed based on a user’s role. For example, common applications such as SSH that are used to manage data center applications need to be tightly controlled, restricted to approved users, and monitored and logged.

This helps in two ways. Enforcing the usage of applications by the appropriate users or groups ensures that business policies are being followed and compliance needs are met. By using application visibility and control to prevent or limit high-risk applications, you reduce the attack surface and limit threats.

This is particularly useful in an enterprise data center environment, where meaningful segmentation by user and application with full content scanning can address compliance requirements and auditibility in the data center. For example, an organization can segment servers containing credit card data and only permit access to that segment for finance users employing a payment application, thus containing and limiting access while maintaining individual accountability. Additionally, threat scanning can be enabled for the segment and data filtering can be enabled to prevent credit card data from flowing out of that segment.

**Prevent data center threats**

Next-generation firewalls also provide a fully integrated approach to threat prevention in a unified context. This means true coordination of multiple security disciplines — such as application-enabled threat vectors, malware detection, intrusion prevention, file-type controls, and content inspection — for a more intelligent and definitive understanding of threats in the data center.

In an Internet-facing data center that may be faced with DDoS attacks, ISP monitoring and traffic throttling can be augmented with next-generation DDoS protection. In addition to being able to identify and block DDoS tools or exploits used by attackers, DDoS protection profiles allow you to control various types of traffic floods, such as SYN-, UDP-, and ICMP floods. You also can set rules for the maximum number of
concurrent sessions to ensure that sessions don’t overwhelm available resources. DDoS protection profiles allow you to set independent limits on aggregate, as well as same-source, sessions.

The ability to pinpoint and analyze unknown traffic is essential in a data center, where unknown traffic should ideally constitute a very small percentage of total traffic. A next-generation firewall provides the ability to categorize and analyze unknown traffic in the network to determine whether the traffic is being generated by a legitimate application that isn’t recognized or by a potential malware infection.

Unusual traffic patterns such as traffic going to known malware sites, recently registered domains, IP addresses instead of domain names, and the presence of IRC traffic may indicate the presence of botnets in the data center. Manual investigation usually means reviewing voluminous log files, much like looking for a needle in a haystack. But next-generation firewalls provide threat logs, automated reports, and easy correlation in a single user interface that eliminates the need to manually track and correlate events. The next-generation firewall threat prevention solution also includes the ability to automatically analyze unknown files in a sandbox environment to identify the behaviors of malware.

Performance and security — a policy choice

First and foremost, data center network security infrastructure must perform. A security device that doesn’t perform doesn’t get deployed in the data center. Next-generation firewalls offer a software architecture that optimizes performance to deliver computationally intensive functions at wire speed.

A next-generation firewall that utilizes a single-pass software architecture performs operations once per packet. As a packet is processed, networking functions, policy lookup, application identification, and decoding and signature matching for any and all threats and content are all performed just once. This significantly reduces the amount of processing overhead required to perform multiple functions in one security device.
Content-scanning within the single-pass architecture utilizes stream-based technology that streamlines processing. Instead of using separate engines and signature sets requiring multipass scanning and file proxies requiring file download prior to scanning, the single-pass software scans content once and in a stream-based fashion to avoid introducing latency. This also facilitates a single, fully integrated policy that simplifies operation and management.

The final element of the architecture revolves around the built-in resiliency delivered by the physical separation of data and control planes. This separation means the heavy utilization of one won’t negatively impact the other.

Deployments with physical next-generation firewalls (as opposed to virtual next-generation firewalls) will benefit from the dedicated hardware architecture of the next-generation firewall — a multicore parallel processing architecture that dedicates processing engines and cores to high-processing functions.

**Flexibility of feature deployment**

By leveraging the high-performance architecture described in the previous section, next-generation firewalls can deliver computationally intensive security functions without impacting performance in the data center. But, this is only useful if the integration of the network security is not overly complex or doesn’t require a complete re-architecting of the data center network. Next-generation firewalls simplify this by offering the flexibility of integration into any architecture and operational model by integrating at OSI Layer 1, Layer 2, and Layer 3, or all three layers in a port-dense appliance.

This design flexibility allows enterprises to implement a network security architecture that doesn’t force an either/or choice between performance and security and doesn’t require retrofitting data center designs to incorporate security. Data center network architects no longer have to figure out how to incorporate various security appliances, such as IPS, which are notorious network performance choke points. Changing the security posture becomes a policy setting that can be accommodated at any point in time, for example, from application- and user-specific firewall policies to full IPS.
Secure remote users

Network perimeters between the inside and outside of a network are blurring due to empowered users accessing the network from practically any device from anywhere in the world. This is a security challenge because traditional security solutions can only enforce security policies when the user is on premise. A user’s behavior also tends to be riskier when outside of the office. This behavior increases the likelihood of clicking on a dangerous link or visiting a site that serves up a malicious drive-by-download.

Next-generation firewalls protect remote users by enforcing full enterprise firewalling and threat prevention, regardless of user location. Endpoint protection enables administrators to allow access only after validation of the health or status of the device — such as the latest operating system patches or other settings — required by the corporate security policies. Finally, access to the data center network is secured using IPsec or SSL VPN connections.

Addressing Virtualization-Specific Security Challenges

The introduction of a new technology — even one, such as virtualization, that has the potential to completely transform the physical composition and arrangement of an organization’s computing infrastructure — does nothing to alter the fundamental principles that define good network security. Many of the core security principles that define good network security policies still apply in virtualized data centers. These include

国防深度。拥有多个互补的防护层可以抵御更广泛的威胁，并减少任何单一防护措施被绕过的风险。例如，在虚拟化数据中心，安全的第一步是从未经授权的用户和外部攻击中抵御未经授权的用户，然后部署防火墙在数据中心的网络边界，以及在数据中心内部进行额外的网络分割或虚拟防火墙。
Segmentation based on trust levels. A critical best practice is to evaluate the trust levels of the server resources in the data center and group them based on similar risk factors and security classifications. Groups with different trust levels should then be segmented and communication allowed only via the firewall.

Separation of duties. To avoid conflicts of interest that may lead to security challenges, operations and security responsibilities should be assigned to different personnel in the data center. For example, the networking team responsible for routing and switching should be different from the server and application team; both of these teams should be different from the security team that’s responsible for implementing and enforcing a cohesive, overall security strategy in the data center.

Separation of countermeasures. As part of the defense-in-depth approach, security solutions shouldn’t just be resident on the resources being protected. Countermeasures should also be available external to the resources to better protect those resources and to avoid any exploits. For example, in virtualized server data centers, having just a virtual firewall protect a server isn’t adequate. Attacks that exploit the virtual server or hypervisor can only be adequately addressed with an external firewall.

Overall, the goal of these core principles and commonsense strategies is to achieve an appropriate degree of separation and protection, regardless of whether the data center consists of traditional 3-tier infrastructure or is virtualized. It is also important to remember that “just because a thing can be done, doesn’t mean it should be done.” For example, although server virtualization can be used to consolidate trust levels and thereby reduce infrastructure costs, the trade-off is that this approach increases configuration complexity and, as a result, risk. The point is that all relevant factors and implications need to be carefully weighed when making such choices.

Finally, it is important to recognize that a successful network security solution is one that is adaptable. Rather than remaining fixed over time, it makes far greater sense to use tactics and tools that can steadily evolve to account for the inevitable changes to the extent and nature of how virtualization technology is used in the data center.
Ensuring hypervisor integrity

The hypervisor — the virtualization layer between the operating system and the hardware (refer to Chapter 1) — creates a new attack surface in the virtualized data center. If it is compromised, the hypervisor can put all VMs hosted on it at risk — even individual VMs that are properly secured with other controls and countermeasures.

Hypervisor attacks range from vulnerabilities that crash hypervisors to complex “breakout” exploits and hyperjacking.

In a “breakout” or “VM escape,” the guest VM escapes and infiltrates its own host operating system in order to gain control of the parent hypervisor and other VMs. In a hyperjacking (or “hypervisor stack jacking”) attack, a rogue hypervisor is installed as malware (for example, a VM rootkit) that exploits the processor and takes control of the server.

While publicized hypervisor security breaches are rare, the risk of hypervisor attacks is significant because the hypervisor provides the opportunity to infiltrate the entire machine regardless of the security controls available within each server.

The foundation of hypervisor security is pretty basic and starts with hardened software. In general, treating hypervisors the same as other critical server operating systems makes sense. The organization’s security policy for software operating systems on physical servers should be extended to virtualized servers. Meticulous patch and change management, continuous monitoring and SIEM (Security Information and Event Management), and strict access control based on the principle of least privilege is essential. Other security best practices for the hypervisor include

- **Isolation between components within the server virtualization platform.** The guest OS should never be allowed access to the hypervisor.

- **Controlled access to the hypervisor.** Access to the hypervisor management system must be protected and restricted to ensure that management functions are only
available to authorized IT administrators. This means locking down the plethora of management options for virtualized platforms, including browser and client-based console access to scripts, command-line interfaces, and centralized management tools.

- **Native security features that are part of the server virtualization platform.** These include memory or process “firewalls” that are functionally similar to traditional host intrusion prevention technology.

- **External next-generation firewalls.** Using a firewall to granularly control access to a system effectively reduces its attack surface, while using IPS can stop threats specifically targeted against the hypervisor. In addition to addressing exploits, external next-generation firewalls can enforce appropriate policies and access to critical virtualized systems.

It is also important to know that not all server virtualization platforms are the same. For example:

- Native or bare-metal (Type 1) hypervisors do not have a guest operating system and are therefore better than hosted (Type 2) hypervisors from a security perspective.

- “Thin virtualization” solutions that avoid the need for a parent partition and minimize hypervisor code provide a reduced attack surface.

- Virtual switching and networking features such as support for Layer 2 access control, promiscuous-mode off by default, and the ability to prevent MAC address changes provide additional security controls.

- Robust resource reservation capabilities are crucial to preventing scenarios where excessive utilization by one VM starves out all other VMs on the same system.

In addition to being a single point of access, the hypervisor is potentially a single point of failure in the virtualized data center. It is important to include redundancy and high availability in the design of your virtualized data center to provide resiliency and protect against denial-of-service (DoS) attacks (including those that are self-inflicted!).
**Controlling intra-host communications**

Virtualized computing environments enable direct communication between VMs within a physical host server. Intra-host communications may not be visible to network-based security appliances deployed outside a virtual server.

As new and potentially vulnerable VMs are created, the same set of security best practices applicable in a physical data center needs to be extended to the virtual environment. This means visibility and control of all application traffic, which ultimately provides more comprehensive threat protection for all application threat-vectors.

Additionally, the inspection of intra-host communications is essential for compliance purposes, particularly if VMs of different trust levels coexist on a single physical host. Just like in a traditional data center environment, a *port-based* virtual firewall has a lack of visibility into all applications, for example, virtualized applications that use nonstandard ports, and port hop, encrypt their communications, and utilize evasive techniques. Malware that uses various techniques to hide its true nature or existence on the network will circumvent traditional port-based virtual firewalls.

A virtualized next-generation firewall brings to the virtualized data center the benefits of application visibility and control, user-based policies, and comprehensive threat protection. The virtualized next-generation firewall also protects against virtualization-specific attacks, such as VM vulnerabilities and exploits.

Meaningful segmentation policies can now extend to virtual workloads as well. For example, controlling and monitoring access to specific virtual servers that store confidential financial data.

**Securing VM migrations**

When VMs migrate from one physical host to another (refer to Chapter 1), they tend to break traditional network security tools that rely on physical and/or network-layer attributes to identify and protect servers and applications. In a virtualized
data center, network security solutions need to be capable of ensuring that applicable access control and traffic inspection rules continue to be enforced, regardless of the potential for VMs to move from one physical host to another and from one physical data center to another.

In a data center environment, IP addresses used in policies may need to be updated or created dynamically. One example is in a virtualized data center where users are creating VMs automatically. Every time a VM is provisioned, or moved, a configuration change on the firewall is needed to ensure the security policy appropriately reflects the updated network topology or VM change.

Next-generation firewalls that bind security policies to VMs automate the process of keeping security policies in sync with the creation of VMs. The dynamic nature of these security objects also enable security policies to be maintained with VM workload movement.

This can also help address challenges with “VM sprawl” where the number of VMs running in a virtualized environment increases because of the ease of creating a new VM rather than making modifications to an existing one.

Additionally, the network link that is used for VM migrations must be secured. A successful attack against the network can compromise the integrity of VMs as they are migrated or can prevent the VM from being migrated in a timely manner, effectively resulting in a denial of service.

Cloud-readiness

As virtualized data center environments move toward the cloud, the management of a virtualized computing environment must be automated in order for applications and resources to be provisioned on an as-needed basis.

The management and orchestration of a virtualized computing environment is complicated and can involve multiple phases. The first phase includes the provisioning of the virtual server, followed by the provisioning of the networking elements in the virtual environment, such as virtual switches. Finally, the last phase is the security provisioning. In order to properly scale, orchestration software is needed to automate many of
these processes. In particular, the speed of the security provisioning cannot slow down the other processes.

Next-generation firewalls offer APIs that enable external orchestration software to connect over an encrypted SSL link for management and configuration. By allowing configuration parameters to be seen, set, and modified as needed, turnkey service templates — where existing security service definitions are defined — can be enabled. This management API enables the proper integration with data center orchestration software so that the security features within the next-generation firewall become part of the data center workflow.

Choosing Physical or Virtual Next-Generation Firewalls

Choosing a physical or virtual form factor to protect the virtualized data center depends on a number of factors.

One consideration is the use case. Specific use cases, for example, visibility of intra-host VM communications, can only be addressed with a virtual firewall. Similarly, segmentation and access control of virtual applications within the same server require a virtual firewall.

Extremely mobile organizations that require network security to be quickly enabled, such as military deployments, may also prefer virtual firewalls due to the ease in provisioning these systems.

Physical firewalls can be used if all servers host applications at the same trust level or for well-defined, high-volume network choke points where performance is absolutely essential, such as the data center core. While a software architecture designed to optimize throughput and performance is available on both physical and virtual next-generation firewalls, physical firewalls directly benefit from hardware-accelerated processing of specific security functions.

In many cases, both physical and virtual firewalls can be used in a hybrid data center environment with physical and virtualized servers.
Chapter 5

A Phased Approach to Security: From Virtualization to Cloud

In This Chapter

▶ Recognizing how security needs change with the data center
▶ Deploying physical and virtual security solutions

As organizations move from traditional 3-tier data centers to virtualized data centers, and eventually to private and public clouds, security architectures must evolve to support changing requirements. This chapter presents a phased plan for moving from virtualized data centers to the cloud.

Journey to the Cloud — One Step at a Time

As organizations implement server virtualization in the data center, the normal progression is toward cloud deployments. As shown in Figure 5-1, as you move from traditional 3-tier architectures to virtualized data centers and the cloud, the same sets of requirements still apply, but additional new requirements need to be addressed. (See Chapter 4 for a discussion about virtualization-specific challenges, including “cloud-readiness.”) The cloud brings additional security requirements.

These materials are the copyright of John Wiley & Sons, Inc. and any dissemination, distribution, or unauthorized use is strictly prohibited.
The “cloud” is not a location but rather a framework of elements, such as automation, orchestration, and service monitoring, that allows disparate processes to be stitched together in a seamless manner. By doing so, the individual elements can be easily replicated and offered on an as-needed basis. Security solutions must provide the appropriate hooks into the orchestration and automation planes in order to not slow down workload creation processes. Because cloud services are extended to multiple organizations with different risk levels, multitenant segmentation is critical.

**A Phased Approach to Security in Virtualized Data Centers**

The following approach to security in the evolving data center — from traditional 3-tier architectures to virtualized data centers and to the cloud — aligns with practical realities, such as the inertia behind existing best practices and technology investments, the need for IT security staff to gain experience and build necessary skill sets, and the likelihood that organizations will transform their data centers incrementally rather than at one fell swoop. This approach consists of four phases:

- **Phase 1**: Consolidating servers within trust levels
- **Phase 2**: Consolidating servers across trust levels
- **Phase 3**: Selective Network security virtualization
- **Phase 4**: Dynamic computing fabric

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

In the first phase of data center virtualization, 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.

Hypervisor protection at this early stage is usually limited to patch and configuration management consistent with other nonvirtualized systems in the data center. 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 take place, the absence of protective safeguards between these virtualized systems is also consistent with the organization’s security posture for nonvirtualized systems. Finally, live migration features (see Chapter 1) are typically used to enable transfer of VMs only to hosts supporting workloads with the same trust level within the same subnet. Because VMs retain their IP and MAC addresses when they move, there should be no loss of context/control for intermediate network security devices.

Security solutions used in this phase 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 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 during this phase.

Consolidating servers across trust levels

In the next phase of data center virtualization, workloads with different trust levels often coexist on the same physical host or cluster of physical hosts. In addition to patch and configuration management, hypervisor protection in this phase typically includes configuring native hypervisor
security features and implementing networking intrusion prevention systems (IPS). 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.

Security solutions used in this phase are similar to those used in the first phase, and additional capabilities are generally not required. 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 (see Figure 5-2) — before being allowed to proceed to the destination VM. Typically, this can be accomplished by configuring separate virtual switches with separate physical network interface cards (NICs) for the VMs at each distinct trust level.

![Figure 5-2: “Off-box” intra-host traffic between different trust levels.](image)

As a best practice for virtualization, combining workloads with different trust levels on the same server should be minimized. Additionally, live migrations of VMs 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, maintaining segmentation based on trust levels becomes more challenging.

**Selective network security virtualization**

The third phase of virtualized data center evolution involves selective virtualization of network security capabilities. Hypervisor protection is expanded to include system-wide anti-malware agents capable of scanning for malware at the hypervisor level and on all VMs (including dormant VMs), without negatively impacting individual VM performance. 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 an off-box, physical security appliance. Long-distance live migrations (for example, between data centers) are enabled by combining native live migration features with external solutions that address associated networking and performance challenges.

A combination of both physical and virtualized network security solutions is appropriate for this phase. The intense processing requirements of solutions such as next-generation firewall 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.

Gartner, Inc., estimates that by 2015, 40 percent of all security controls used in enterprise data centers will be virtualized.

**Dynamic computing fabric**

In the final phase, conventional, static computing environments are transformed into dynamic fabrics (or “private 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. Hypervisor protection at this phase may include hypervisor-level firewalls, and 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 is necessary to ensure not only that adequate protection is provided regardless of the fact that workloads may be frequently migrating across data center locations, but also that it is delivered in an optimal manner.
Chapter 6


In This Chapter

▶ Understanding what to look for in network security solutions for the virtualized data center!

This chapter helps you assess network security solutions for your virtualized data center by presenting several important features and criteria for you to consider.

Safe Application Enablement of Data Center Applications

Accurate traffic classification — regardless of ports, protocols, evasive tactics, and SSL encryption — is important in any data center. This is even more critical in a virtualized environment where VMs can be quickly instantiated to support application needs — often without appropriate policies or risk analysis. Applications must be identified regardless of whether the traffic only traverses within the virtual infrastructure or crosses physical server boundaries. Unknown traffic must also be characterized.
Once a complete picture of applications is gained, safe application enablement of applications is essential to deliver the right security policies in the data center. This includes more fine-grained and appropriate application functions than simply “allow” or “deny,” such as allow but enforce traffic shaping through QoS or allow based on schedule, users, or groups.

Application visibility and control allows organizations to reduce the attack surface by blocking rogue and misconfigured applications, such as unauthorized management tools and P2P file-sharing software. It also enables the protection of high-value targets, such as domain controllers, finance servers, and e-mail and database servers with meaningful network segmentation.

**Identification Based on Users, Not IP Addresses**

User identification tied to the actual user instead of their IP address is essential in enterprises where users are not only dynamic but also mobile — increasingly accessing data center applications from a variety of mobile devices.

User identification is important to not only get full and accurate visibility of user activity on the network, but (together with application identification) it also can provide appropriate control of applications in the data center. User-based policy control along with log viewing and reporting are key requirements for security in the virtualized data center.

**Comprehensive Threat Protection**

The modern threat landscape has evolved into intelligent, targeted, persistent, multiphase intrusions. Threats are delivered via applications that dynamically hop ports, use nonstandard ports, tunnel within other applications, and hide within proxies, SSL, or other types of encryption.
Additionally, enterprises are exposed to targeted and customized malware, which can easily pass undetected through traditional port-based firewalls and antivirus software. A fully integrated threat solution that addresses a variety of advanced threats is needed to properly secure the virtualized data center.

In addition, file and data filtering options — for example, the ability to block files by their actual type and the ability to control the transfer of sensitive data patterns, such as credit card numbers — address important compliance use cases.

One of the limitations of traditional anti-malware security signatures is the ability to only protect against malware that has been previously detected and analyzed. This reactive approach creates a window of opportunity for malware. To supplement this, the data center network security solution should provide the ability to directly analyze unknown executables for malicious behavior.

Flexible, Adaptive Integration

One of the key integration challenges in the data center is security design. Network architectures must often be redesigned when security requirements evolve due to changing applications and threats, new compliance mandates, and shifting risk postures. A new paradigm that enables network security to be flexible and adaptive is needed.

Networking flexibility helps ensure compatibility with practically any organization’s data center environment. Enabling integration without the need for redesign or reconfiguration depends not only on supporting a wide range of networking features and options, such as 802.1Q and port-based VLANs, but also on the ability to integrate at OSI layer 1 (Physical), layer 2 (Data Link), or layer 3 (Network). In addition, the network security solution should be able to turn on additional security features as the security posture changes.
**High-Throughput, Low-Latency Performance**

One key trade-off in traditional data center security designs has been between security and performance. When organizations enable additional threat-prevention capabilities, this additional security protection typically impacts overall performance in the data center.

This paradox ultimately creates friction between network and security teams, with each team pointing fingers because the other team is making their job harder. Many security vendors that claim to offer next-generation security solutions offer them in the form of additional security appliances or add-on “features” that negatively impact performance.

The ability to deliver highly available, highly responsive applications is clearly a foundational requirement for the data center. A high-throughput, low-latency network security design that aligns with the requirements of the data center is therefore a must for any data center security solution.

**Secure Access for Mobile and Remote Users**

The modern enterprise continues to become far more distributed than in the past. Users simply expect to be able to connect and work from any location, whether at an airport, a coffee shop, a hotel room, or at home. Employees, partners, contractors, and supply chains are all accessing data center resources from beyond the traditional perimeter of the enterprise.

The requirement to protect these mobile and remote users is a way to enable the same application, user, and content protections they receive while on premise. Network security solutions for the data center must deliver secure access for these users to the data center, in addition to addressing the use of endpoint devices other than standard corporate-issued equipment.
One Comprehensive Policy, One Management Platform

Most data centers today are hybrid environments, comprised of both physical and virtual infrastructures. Network security solutions in the data center will also likely include both physical and virtual solutions.

The network security policy management platform for both physical and virtual data centers must be the same; otherwise, security policies can become convoluted, leading to needless complexity, misconfigurations, and security blind spots. In addition, a single, comprehensive security policy that fully integrates application control, threat management, and user identification is a must.

Cloud-Readiness

Data center tasks and processes that help IT teams execute change with greater speed, quality, and consistency are typically automated using workflows. However, deployment of security capabilities typically lags orchestration software provisioning in virtual environments, leading to security risks and considerable integration challenges. Automated provisioning of network security capabilities, in line with other orchestration elements of the virtualized data center environment, is essential.

Choice of Form Factor

The choice of whether a physical or virtual network security appliance should be deployed in the data center depends on the specific issues to be addressed.

Physical network security appliances are often adequate if the same trust levels are maintained within a single cluster of virtual hosts. In this scenario, visibility of east–west traffic (internal communications between servers) is less critical and can be forced off-box through a default security appliance, if necessary.
However, when applications of different trust levels exist within the same virtual cluster, full visibility of intra-host communications can only be achieved with virtual firewalls.

Additionally, specific server-level or hypervisor attacks can only be addressed with firewalls protecting these servers — either virtual or physical.

Finally, firewalls deployed to the public cloud may need to be virtualized if the responsible service providers do not allow customers to deploy their physical hardware in the cloud. Virtualized firewalls may also be appropriate in private clouds and virtualized data centers, where rack space is at a premium or where data center mobility (for example, a temporary data center in a remote region) is needed.
802.1q: The IEEE standard for VLAN tagging on Ethernet networks.

adware: Pop-up advertising programs that are commonly installed with freeware or shareware.


APT: Advanced Persistent Threat. A sustained Internet-borne attack usually perpetrated by a group with significant resources, such as organized crime or a rogue nation-state.

backdoor: Malware that enables an attacker to bypass normal authentication to gain access to a compromised system.

bare metal (Type 1) hypervisor: See native hypervisor.

bootkit: A kernel-mode variant of a rootkit, commonly used to attack computers that are protected by full-disk encryption.

bot: A target machine that is infected by malware and is part of a botnet (also known as a zombie).

bot-herder: The owner or individual that controls a botnet.

botnet: A broad network of bots working together.

BYOD: Bring Your Own Device is a current trend in which organizations are increasingly allowing their users to bring personal mobile devices — such as smartphones and tablets — into the workplace and connect these devices to the network for both personal and work-related purposes.

cold migration: A migration process requiring the halting of the original physical server or VM, transfer of associated data, and rebooting of the VM on the physical host. See also live migration and warm migration.
consumerization: A current trend in which users increasingly find personal technology and applications that are more powerful or capable, more convenient, less expensive, quicker to install, and easier to use than corporate IT solutions.

DDNS: Dynamic DNS is a technique used to update domain name system (DNS) records in real-time.

DDoS: Distributed denial-of-service is a large-scale attack that typically uses bots to crash a targeted network or server.

drive-by-download: Software, often malware, downloaded onto a computer from the Internet without a user’s knowledge.


hosted (Type 2) hypervisor: A hypervisor that runs within an OSE.

HTTP: Hypertext Transfer Protocol.

HTTPS: Hypertext Transfer Protocol over SSL/TLS.

hypervisor: Also known as a virtual machine manager (VMM). Allows multiple “guest” operating systems to run concurrently on a single physical host computer.

IDS/IPS: Intrusion Detection System/Intrusion Prevention System.

IEEE: Institute of Electrical and Electronics Engineers.

IM: Instant Messenger.

inline mode: An IDS/IPS mode in which the IDS/IPS is positioned directly in the packet flow and the IDS/IPS can perform actions (such as block, drop, log, or alert) directly on network traffic. See also promiscuous mode.

IPSec: Internet Protocol Security is a protocol suite for protecting communications over IP networks using authentication and encryption.

IRC: Internet Relay Chat.

**live migration**: A migration process in which a new instance of a VM is created before migrating the existing VM. A live migration doesn’t require halting of the VM, transfer of associated data, or a reboot of the VM, and all session information is maintained (stateful). See also **cold migration** and **warm migration**.

**logic bomb**: A program that performs a malicious function when a predetermined circumstance occurs.

**malware**: Malicious code that typically damages or disables, takes control of, or steals information from a computer. Malware includes viruses, worms, Trojan horses, logic bombs, rootkits, bootkits, backdoors, spyware, and adware.

**MBR**: Master Boot Record.

**MS-RPC**: Microsoft Remote Procedure Call is a communications protocol used on MS Windows networks.

**native (Type 1) hypervisor**: A hypervisor that runs directly on the host computer’s hardware. Also known as a **bare metal hypervisor**.

**OSE**: Operating system environment.

**OSI**: Open Systems Interconnection.

**P2P**: Peer-to-Peer.

**promiscuous mode**: An IDS/IPS mode in which the IDS/IPS captures a copy of network traffic but cannot block or drop any packets. In this mode, an IDS/IPS can only detect and log or alert. See also **inline mode**.

**QoS**: Quality of Service.

**rootkit**: Malware that provides privileged (root-level) access to a computer.

**RPC**: Remote Procedure Call.

**RSA**: Rivest-Shamir-Adelman.

**SIEM**: Security incident and event management.

**SMB**: Server Message Block is an application-layer protocol, also known as Common Internet File System (CIFS).
**SMTP**: Simple Mail Transfer Protocol.

**spyware**: A form of malware that’s installed on a user’s computer, often for the purpose of collecting information about Internet usage or for taking control of a computer.

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

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

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

**TCP**: Transmission Control Protocol.

**Trojan horse**: A program that purports to perform a given function, but which actually performs some other function.

**UDP**: User Datagram Protocol.

**virus**: A set of malicious computer instructions that embeds itself within another program in order to replicate itself.

**VoIP**: Voice over Internet Protocol (IP).

**VLAN**: Virtual Local Area Network.

**VM**: Virtual machine.

**VNC**: Virtual Network Computing.

**VPN**: Virtual Private Network.

**warm migration**: A migration process that does not require halting of the VM, transfer of associated data, or a reboot of the VM, and all session information is maintained (stateful). See also **cold migration** and **live migration**.

**workload**: The amount of processing to be done by a specified computer. Increasingly synonymous with VM.

**worm**: Malware that usually has the ability to replicate itself across networks without the need for human intervention.
Restore control and protect your applications and data in virtualized data centers and in the cloud!

Virtualization is one of the hottest trends of the past decade and a key enabling technology in virtualized data centers and cloud computing strategies. But the modern application and threat landscape, coupled with sophisticated malware and attackers, has also evolved. Virtualization also introduces new security challenges at the hypervisor and virtual machine level. In order for organizations to realize the benefits of virtualization and cloud computing, they must adapt their security architectures to address these new realities.

- **How virtualization delivers operational benefits — and introduces new security challenges**

- **How modern threats hide in applications — and expose your data center to new risks**

- **Why traditional security controls alone aren’t enough in virtual data centers — and how to protect hypervisors and control intra-host communications and migrations**

- **Where to deploy next-generation firewalls — and how to use their capabilities to protect your virtualized data center**

**Lawrence Miller** has worked in information security for more than 20 years. He is the coauthor of *CISSP For Dummies* and more than 30 other titles. He is also a Palo Alto Networks customer and liked it so much he bought the company — well, he’s not that rich (yet) — but he did write this book!

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