CHAPTER

13 Secure Network Design

Organizations are leveraging the power of the Internet to connect with all types of external entities—such as customers, peer organizations, and suppliers, to name a few. This access is intended to enable a simple and smooth mechanism for dissemination of information, to conduct a variety of business functions, and to provide remote access to systems and data. You would be hard pressed today to find an organization, no matter where or how small, that does not leverage the Internet in some respect as a part of its operation. Fueled by advances in social networking, high performance and miniaturized computing, the mobile technology revolution, and the increasing amount of bandwidth available to portable devices, this trend will continue to accelerate for the foreseeable future.

Unprecedented global access is available, and providing a virtual vehicle to get people connected and operating as efficiently as possible is a standard part of modern IT responsibilities. To put it another way, enterprises must make their information property and platforms available to themselves remotely and to third parties, often including sensitive IP material and data. While in many cases it is important to a business to be able to provide this access, it is also very important to draw a line between where the responsibility for managing security is an organization's role and where it is no longer its jurisdiction. The underlying design of the network plays an integral role in an organization's ability to effectively manage and secure access to its data.

The boundary between an organization's network and the Internet or a peered network, much akin to a parcel property line, is known as an *electronic security perimeter (ESP)*. Thinking of this concept in terms of concentric rings, the network perimeter lies wholly within the ESP and is often confined to a particular physical location or set of locations, while the ESP has other elements like corporate smartphones, tablets, and other mobile devices. These devices may be outside of the network(s) physically, but they are still within the ESP. Within this perimeter you will find all owned computing assets and potential storage locations for organization data, sometimes including third-party systems.

Whether or not an application should be considered "within" the ESP can have a lot to do with how it is used, especially with the prevalence of enterprise-level Software-as-a-Service (SaaS) applications and the rapid adoption of more complex and varied application integrations.

The underlying design of the network will play an integral role both in defining those electronic boundaries and in enabling an organization to effectively protect, manage, and secure access to information assets within that perimeter. Sometimes, an organization's intellectual property can reside outside of this perimeter, which requires additional consideration when planning for the protection of enterprise data.

One of the most significant problem areas for security groups is defining the appropriate boundary for the ESP, and being aware of what actions can inadvertently change that boundary. From the very beginning of a design project, think big in terms of the scale that the project may need to support, which may require you to include placeholders for future technology (in the IP schema, for example). Budgetary constraints and practical matters may impose limits on the approach, and in no way is this advocacy of oversizing systems beyond reason, but from a pure design perspective, spend the time to consider the hardship it will cause *later* if the design cannot accommodate growth or, worse, misses critical gaps and introduces uncomprehended and unforeseen risks.

Introduction to Secure Network Design

All information systems create risks to an organization, and whether or not the level of risk introduced is acceptable is ultimately a business decision. Controls such as firewalls, resource isolation, hardened system configurations, authentication and access control systems, and encryption can be used to help mitigate identified risks to acceptable levels.

Acceptable Risk

What constitutes an acceptable level of risk depends on the individual organization and its ability to tolerate risk. An organization that is risk averse will ultimately accept lower levels of risk and require more security controls in deployed systems. Management's risk tolerance is expressed through the policies, procedures, and guidelines issued to the staff. A complete set of policies outlining management's preferences and its tolerance of information security risks enables employees to make appropriate infrastructure decisions when designing and securing new systems and networks. Thus, the design and configuration of the infrastructure becomes the enforcement of those documents.

Some organizations unintentionally take on more risk than they intend to by being unaware of the legislative instruments that they are subject to within a legal jurisdiction. Computing and information laws have evolved and changed rapidly, and span hundreds of volumes of material and thousands of web pages for the United States alone. That's not even considering the challenges that multinational corporations face when operating on the soil of many nations. During the development of the policies that will guide the design of the systems and networks, management should spend the time and effort necessary to determine if any of these special legal considerations apply.

Many enterprises inadvertently violate certain laws without even knowing that they are doing so (for example, storing credit card numbers without taking into account Payment Card Industry Data Security Standard [PCI DSS], or storing patient data without factoring in Health Insurance Portability and Accountability Act [HIPAA] provisions; compliance with regulations like these was covered in more detail in Chapter 3, and the like). This modifies the level of residual risk actually produced after the controls are applied, since the planned controls may not address risks that are not clearly defined prior to control plan development.

Designing Security into a Network

Security is often an overlooked aspect of network design, and attempts at retrofitting security on top of an existing network can be expensive and difficult to implement properly. Separating assets of differing trust and security requirements should be an integral goal during the design phase of any new project. Aggregating assets that have similar security requirements in dedicated zones allows an organization to use small numbers of network security devices, such as firewalls and intrusion-detection systems, to secure and monitor multiple application systems.

Other influences on network design include budgets, availability requirements, the network's size and scope, future growth expectations, capacity requirements, and management's tolerance of risks. For example, dedicated WAN links to remote offices can be more reliable than virtual private networks (VPNs), but they cost more, especially when covering large distances. Fully redundant networks can easily recover from failures, but having duplicate hardware increases costs, and the more routing paths available, the harder it is to secure and segregate traffic flows.

A significant but often missed or under-considered factor in determining an appropriate security design strategy is to identify how the network will be used and what is expected from the business it supports. This design diligence can help avoid expensive and difficult retrofits after the network is implemented. Let's consider some key network design strategies.

Network Design Models

To paint a clearer picture of how the overall design impacts security, let's examine the designs of a shopping mall and an airport. In a shopping mall, to make ingress and egress as convenient as possible, numerous entrances and exits are provided. However, the large number of entrances and exits makes attempts to control access to the shopping mall expensive and difficult. Screening mechanisms would be required at each door to identify and block unwanted visitors. Furthermore, implementing a screening mechanism isn't the only hurdle; after it is deployed, each mechanism must be kept properly configured and updated to ensure that an unauthorized person doesn't slip through.

In contrast, an airport is designed to funnel all passengers through a small number of well-controlled checkpoints for inspection. Networks built on the shopping mall model are inherently harder to secure than networks designed around the airport model. Networks built with many connections to other networks will be inherently harder to secure due to the number of access control mechanisms (such as firewalls) that must be implemented and maintained.

The design of an airport does much more than just facilitate the passenger screening performed just inside a terminal. Overall, the airport has a highly compartmentalized design that requires an individual to pass through a security check whenever passing between compartments. Not all screening is explicit—some monitoring is passive, involving cameras and undercover police officers stationed throughout the airport. There are explicit checkpoints between the main terminal and the gate areas, as well as between the gate area and the plane. There are security checks for internal airport movements as well, and staff need special access keys to move into the internal areas, such as baggage processing and the tarmac.

An average big-city airport also maintains multiple terminals to handle the traffic load, which reduces the impact of a security breach in a single terminal. These smaller, higher-security terminals can have more stringent security checks, and it allows passengers with different security requirements, such as politicians and federal prisoners, to be segregated, lowering the risk that one group could affect the other. All of these elements can be translated into network design, such as using firewalls and authentication systems for controlling traffic movement around the network, using the network to segregate traffic of differing sensitivity levels, and using monitoring systems to detect unauthorized activities.

Designing an Appropriate Network

There are invariably numerous requirements and expectations placed upon a network, such as meeting and exceeding the organization's availability and performance requirements, providing a platform that is conducive for securing sensitive network assets, and enabling effective and secure links to other networks. On top of that, the overall network design must provide the ability to grow and support future network requirements. As illustrated earlier with the airport and mall analogies, the overall design of the network will affect an organization's ability to provide levels of security commensurate with any risks associated with the resources or on that network.

To design and maintain a network that supports the needs of its users, network architects and engineers must have a solid understanding of what those needs are. The best way to do this is to involve those architects and engineers in the application development process. By getting involved early in the development cycle, engineers can suggest more secure designs and topologies, and additionally can assure the project team that they have a clear understanding of the security considerations and capabilities. In addition, they can ensure that new projects are more compatible with the existing corporate infrastructure.

Common steps for obtaining such information include meeting with project stakeholders, application and system owners, developers, management, and users. It is important to understand their expectations and needs with regard to performance, security, availability, budget, and the overall importance of the new project. Adequately understanding these elements will ensure that project goals are met, and that appropriate network performance and security controls are included in the design. One of the most common problems encountered in a network implementation is unmet expectations resulting from a difference of assumptions. That's why expectations should be broken down into mutually observable (and measureable) facts as much as possible, so the security designers ensure that there is explicit agreement with any functional proposals clearly understood and agreed.

The Cost of Security

Security control mechanisms have expenses associated with their purchase, deployment, and maintenance, and implementing these systems in a redundant fashion can increase costs significantly. When deciding on appropriate redundancy and security controls for a given system or network, it is helpful to create a number of negative scenarios in which a security breach or an outage occurs, to determine the corporation's costs for each occurrence. This risk-model approach should help management determine the value to the corporation of the various security control mechanisms.

For example, what costs are incurred to recover from a security breach or when responding to a system outage outside of normal business hours? Be sure to include cost estimates for direct items, such as lost sales, reduced productivity, and replacement costs, as well as for indirect items, such as damage to the organization's reputation and brand name, and the resultant loss of customer confidence. Armed with an approximation of expected loss, corporations can determine appropriate expenditure levels. For example, spending $200,000 to upgrade a trading system to achieve 99.999 percent availability may seem overly expensive on the surface, but it is a trivial expense if system downtime can cost the corporation $250,000 per hour of outage.

Performance

The network will play a huge rule in meeting the performance requirements of an organization. Networks are getting faster and faster, evolving from 10 megabit to 100 megabit to gigabit speeds, with 10GE commonly deployed and 40GE, 100GE, and InfiniBand technologies available today. When determining the appropriate network technology, be sure that it can meet the bandwidth requirements projected for three to five years in the future. Otherwise, expensive replacements or upgrades may be required.

Applications and networks that have low tolerance for latency, such as those supporting video and voice streaming, will obviously require higher performance network connections and hardware. What about applications that move data in large chunks (for example, storage snapshots or disk-to-disk offsite replication)? In lieu of an expensive, dedicated, high-bandwidth connection, it may be more economical to implement links that are *burstable*, meaning that the provider will allow short bursts of traffic above the normal subscribed rate. If applications will share common network infrastructure components, the design team may also consider implementing Quality of Service (QoS) technologies to prevent one application from consuming too much bandwidth, or to ensure that higher priority applications always have sufficient bandwidth available.

The legacy Cisco Hierarchical Internetworking model, which most network engineers are intimately familiar with, is a common design implemented in large-scale networks today, although many new types of purposed designs have been developed that support emerging technologies like class fabrics, lossless Ethernet, layer two bridging with trill or IEEE 802.1aq, and other data center–centric technologies.

The three-tier hierarchy still applies to campus networks, but no longer to data centers. This is a "legacy" model socialized by Cisco, but even Cisco has newer thinking for data centers. Networks are becoming much more specialized, and the security thinking for different types of networks is significantly different. The Cisco three-tier model is derived from the Public Switched Telephone Network (PSTN) model, which is in use for much of the world's telephone infrastructure. The Cisco Hierarchical Internetworking model, depicted in Figure 13-1, uses three main layers commonly referred to as the core, distribution, and access layers:

- **Core layer** Forms the network backbone and is focused on moving data as fast as possible between distribution layers. Because performance is the core layer's primary focus, it should not be used to perform CPU-intensive operations such as filtering, compressing, encrypting, or translating network addresses for traffic.

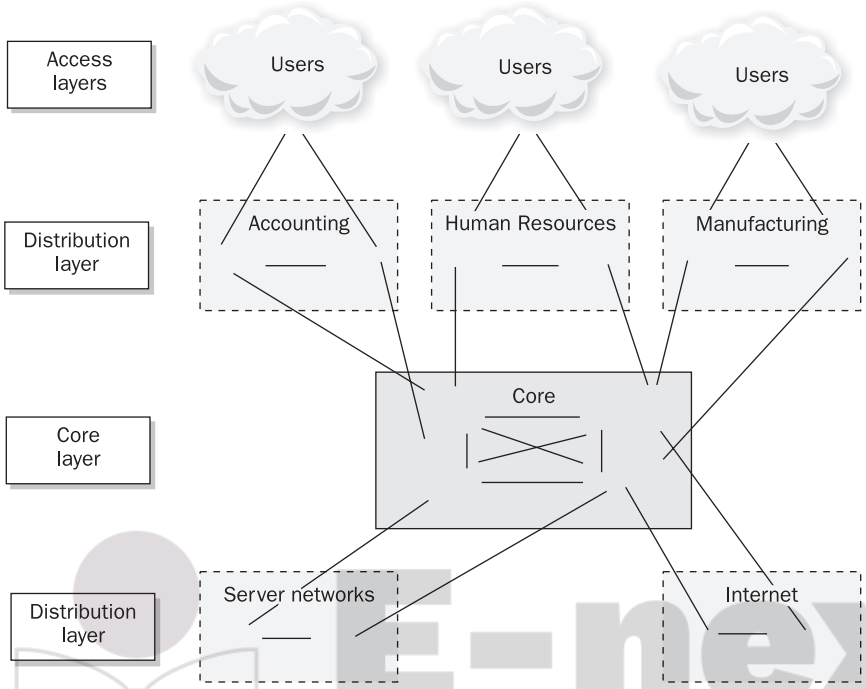

Figure 13-1 The Cisco Hierarchical Internetworking model

- **Distribution layer** Sits between the core and the access layer. This layer is used to aggregate access-layer traffic for transmission into and out of the core.
- **Access layer** Composed of the user networking connections.

Filtering, compressing, encrypting, and address-translating operations should be performed at the access and distribution layers.

The Cisco model is highly scalable. As the network grows, additional distribution and access layers can be added seamlessly. As the need for faster connections and more bandwidth arises, the core and distribution equipment can be upgraded as required. This model also assists corporations in achieving higher levels of availability by allowing for the implementation of redundant hardware at the distribution and core layers. And because the network is highly segmented, a single network failure at the access or distribution layers does not affect the entire network.

Although the Cisco three-tier model is perhaps the most commonly known and referenced model for designing LAN environments, it has its limitations and is rapidly being supplanted by newer models aimed at addressing the specific needs of highly virtualized data centers, the specific needs of different industry verticals, and the specific needs of cloud computing and multitenancy environments. Despite the success of the three-tier model over the past two decades, as networks continue to develop and become more specialized for purpose, additional reference models have been developed in order to address the limitations and relatively high costs associated with this approach.

Many modern data center architectures and "cloud" designs do not employ a three-tier model, instead favoring a clustered switching, class fabric, or collapsed two-tier approach that offers higher performance and lower cost but also brings special security considerations into play. Data center performance networking has become both more complex and simpler at the same time, with advanced operations being handled in silicon, enabling in-device functionality not previously possible. Looking at one from farther away, the two-tier model appears simplified (and is simpler to operate and manage), but upon zooming in, it becomes obvious that there is a lot more going on within those network devices to enable the magic. Between class fabrics offering clustered switching approaches and silicon capable of microsecond latency, unprecedented performance and availability have been unlocked. A few of the more well-known and published models are Cisco's FlexPod model (data center in a box), Arista's two-tier CloudVision model, Brocade's Brocade One model, and Juniper's Stratus model.

NOTE These are merely examples from some of the larger manufacturers; there are many more approaches to data center network design.

The following are two-tier network fundamentals (three-tier terminology is used for comparative purposes):

- **Core** The core of the two-tier network is a highly available, horizontally scalable element used for transit and moving data between different areas or zones in the network, much like the core in the three-tier model. The only major difference is that, in general, the core in a two-tier network doesn't see 100 percent of the traffic, as much of the host-to-host traffic transits across the fabric without needing to be handled by the core.
- **Distribution** The distribution layer in some collapsed networks either is eliminated completely or is combined with the access layer as part of the fabric. Although a "distribution" layer may literally exist, it does not logically exist, as it is part of the same switch fabric or switching cluster as the access switching.
- **Access** The access layer is collapsed into the distribution layer, so while physically separate devices may provide the aggregation and access function, both can be part of the same layer-two domain employing trill or 802.1aq for bridging. These combined layers offer active/active connectivity across multiple switches via clustering for high availability and performance. This "fabric" introduces a new dimension for security, as server-to-server, server-to-storage, and virtual host communication can now be fused together in ways not previously possible.

Since the data center network is becoming flatter, faster, and much larger, designing the security components to support the goals of the network is more important than ever. Since virtualization is commonplace and shared server/storage platforms almost always exist underneath, ensure that adequate time is spent designing the networks and topologies to allow security components (firewalls, filtering devices, etc.) to "plug in" to the fabric in a fashion that maintains the integrity of data communications between intended hosts but does not compromise the performance of the data center platform. Techniques like VM fencing, virtual appliance firewalls, hypervisor protection, and segregation of security zones by service type are common approaches to ensuring adequate controls are in place to enforce the security plan.

Availability

Network availability requires that systems are appropriately resilient and available to users on a timely basis (meaning, when users require them). The opposite of availability is denial of service, which is when users cannot access the resources they need on a timely basis. Denial of service can be intentional (for example, the act of malicious individuals) or accidental (such as when hardware or software fails). Unavailable systems cost corporations real dollars in lost revenue and employee productivity, and they can hurt organizations in intangible ways through lost consumer confidence and negative publicity. Business availability needs have driven some organizations to construct duplicate data centers that perform real-time mirroring of systems and data to provide failover and reduce the risk of a natural disaster or terrorist attack destroying their only data center.

Depending on the specific business and risk factors, redundancy often increases both cost and complexity. Determining the right level of availability and redundancy is an important design element, which is best influenced by a balance between business requirements and resource availability.

The best practice for ensuring availability is to avoid single points of failure within the architecture. This can require redundant and/or failover capabilities at the hardware, network, and application functions. A fully redundant solution can be extremely expensive to deploy and maintain, because as the number of failover mechanisms increases, system complexity increases, which alone can raise support costs and complicate troubleshooting.

As mentioned previously, the application's availability requirements should be assessed to determine the financial and business impacts of systems being unavailable. Performing this assessment will help management arrive at the optimal balance between failover mechanisms, cost, and complexity for the particular network or application. Numerous security appliance vendors have failover mechanisms that enable a secondary firewall to assume processing responsibilities in the event that the primary firewall fails. Beyond firewalls, routers can also be deployed in a high-availability configuration.

TIP To understand the kind of redundancy that will be required, try to determine how long the business could function normally, should an outage occur.

Implementing a redundant firewall or router solution is only one step in achieving a full high-availability network architecture. For example, a high-availability firewall solution provides no value when both firewalls are plugged into the same switch. The switch becomes a single point of failure, and any interruption in its normal operation would take both firewalls off the network, negating any benefit of the firewall failover mechanism. The same holds true of a router—if there is only a single router between the firewalls and the rest of the network, the failure of that router would also cause an outage. Figure 13-2 shows a full high-availability network segment without a single hardware point of failure (which in this example uses Cisco's Hot Standby Router Protocol [HSRP], which is a built-in protocol for switching routes if a router or interface goes down).

A true high-availability design will incorporate redundant hardware components at the switch, network, firewall, and application levels. When eliminating failure points, be sure to consider all possible components. You may want to guarantee reliable power via a battery back-up, commonly called an uninterruptible power supply (UPS), or even an emergency

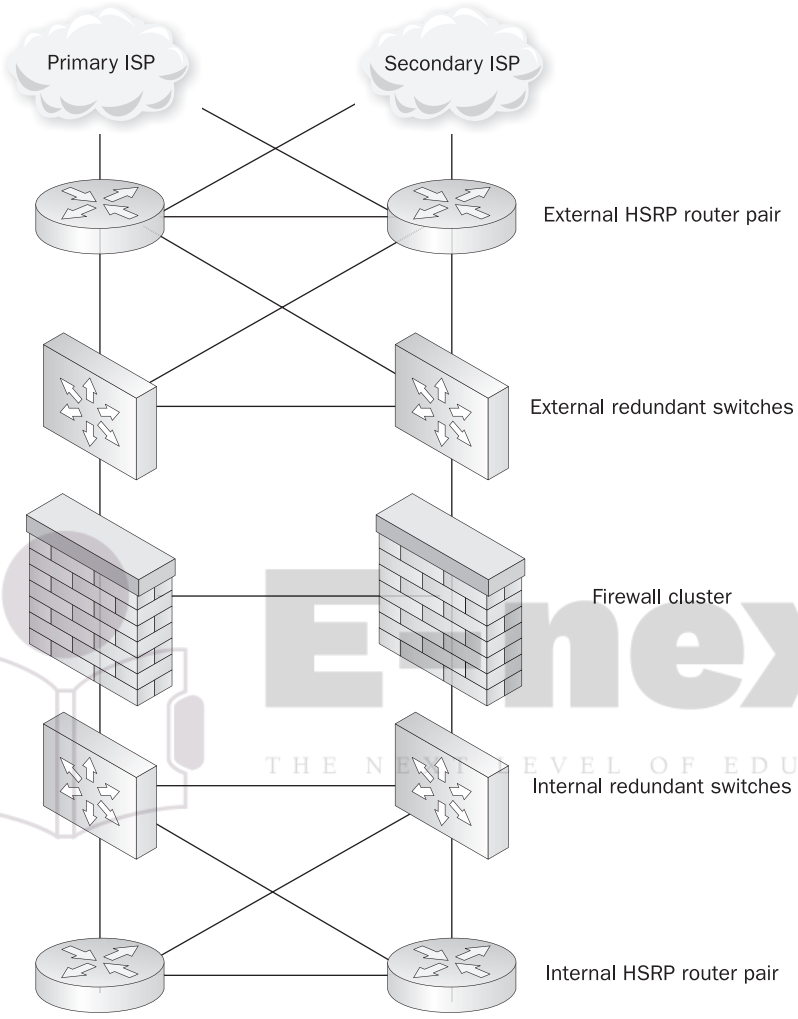

Figure 13-2 A full high-availability network design

generator for potential long-term interruptions. Designers can and should consider maintaining multiple Internet links to different Internet service providers to insulate an organization from problems at any one provider.

Today's high-availability designs have reached a high level of sophistication in modern data centers and network and computing architectures, from the facility itself down to the application running in front of the end user. Load balancers also play an important role in maintaining the availability and performance of network-based services. Today's application delivery technologies are being used for both security and availability. In some cases, organizations have gotten rid of their web tier completely, and host it directly on application delivery controllers (ADCs), which provide optimized application and network performance.

Security

Each element on a network performs different functions and contains data of differing security requirements. Some devices contain highly sensitive information that could damage an organization if disseminated to unauthorized individuals, such as payroll records, internal memorandums, customer lists, and even internal job-costing documents. Other devices have more exposure due to their location on the network. For example, internal file servers will be protected differently than publicly available web servers.

When designing and implementing security in network and system architectures, it is helpful to identify critical security controls and understand the consequences of a failure in those controls. For example, firewalls protect hosts by limiting what services users can connect to on a given system. Firewalls can allow different sets of users selective access to different services, such as allowing system administrators to access administrative services while preventing non-administrative users from accessing those same services. This provides an additional level of control over that provided by the administrative mechanisms themselves. By denying a non-administrative user the ability to connect to the administrative service, that user is prevented from mounting an attack directly on that service without first circumventing the firewall.

However, simply restricting users to specific services may be insufficient to achieve the desired level of security. For example, it is necessary to allow traffic through the firewall to connect to various authorized services. In order for an organization to send and receive e-mail, firewalls must be configured to permit e-mail traffic. As Chapter 15 will discuss, firewalls have limited capability in preventing attacks directed at authorized applications, so overall network security is dependent on the proper and secure operation of those applications.

Flaws, such as a buffer overflows, can allow an attacker to turn a vulnerable server into a conduit through the firewall. Once through the firewall, the attacker can mount attacks against infrastructure behind the protection of the firewall. If the server is on the internal network, the entire network could be attacked without the protection provided by the firewall, but if the server is on a separate firewalled segment instead of the internal network, only the hosts on the same subnet could be directly attacked. Because all traffic exiting that subnet still must pass back through the firewall, it can still be relied upon to protect any additional communications from this compromised subnet to any other internal subnets.

In addition to the best practice of segmenting the traffic, using the advanced inspection capabilities and application-layer gateways of current-generation firewalls can help protect segmented networks by ensuring that traffic being sent as a particular service over a particular port is in fact well-formed traffic for that service. For example, if a server in a segregated network zone is compromised via an http exploit and the attacker attempts to create a connection to another host within a different firewall zone using ssh but over port 80, the firewall should be able to detect that ssh is not http traffic, and warn or block accordingly (based on how it is configured to behave).

Thus, the network design can increase security by segregating servers from each other with firewalls. However, this is not the only control mechanism that can and should be used. While it may not be initially obvious, the proper operation of the service itself is a security

control, and limiting the privileges and capabilities of that service provides an additional layer of control. For example, it is good practice to run services without administrative privileges wherever possible.

In addition to securing individual elements on the network, it is important to secure the network as a whole. The *network perimeter* consists of all the external-most points of the internal network and is a definable inner boundary within the electronic security perimeter. Each connection to another network, whether to the Internet or to any external third party (be it business partner, data provider, and so on), creates an entry point in the perimeter that must be secured.

Perimeter security is only as strong as its weakest link. Without adequate security on each external connection, the security of the internal network becomes dependent on the security of these other connected networks. Because the organization cannot control the security of these external networks and the connections that they maintain, a security breach in those networks can pose significant risks that must be mitigated through appropriate firewalling and application-layer controls.

NOTE All too often, there are connections to a network that the security team is not aware of, due to shadow IT or some "exception" that was not well documented. Having the capability to learn about these unauthorized gateways is very important, as it is impossible to secure the network perimeter without fully understanding where it reaches.

Strong security will ensure that these connections cannot be used as a back door into the internal network. While the risk associated with an Internet user breaking into that network and using it as a conduit into yours may seem remote, unintentional risks, such as virus propagation, still exist. Good practices for reducing risks include periodic auditing of the external networks to ascertain their overall security posture, as well as implementing firewalls to permit only those communications required to conduct business.

Wireless Impact on the Perimeter

Network perimeter security is only useful if there are adequate physical security controls to prevent an unauthorized user from simply walking up to and plugging into the internal network. Thus, without physical access to the network, a malicious user is required to exploit a weakness in the corporate perimeter security controls to gain access. Organizations that deploy wireless solutions must recognize and mitigate risks associated with an unauthorized individual gaining connectivity to the corporate LAN via wireless signal leakage outside of the corporate-controlled premises. By simply getting physically close enough, a malicious user with a laptop and a wireless LAN card may be able to get an IP address on the network.

While the signals from wireless access points degrade quickly when passing through walls and over distance, more powerful and specialized directional antennas can pick up signals at significant distances. These antennas, called Yagi antennas, can pick up wireless signals at distances approaching one mile. While commercial Yagi antennas can be costly, inexpensive ones can be built at home out of an empty potato chip can and some wire. And Yagi antennas are not the only type that can be used for long-range Wi-Fi; there are also backfire and other types of relatively small but powerful antennas that can be used in this fashion.

In addition to signal-leakage problems, flaws have been discovered in the encryption mechanisms used to protect wireless traffic. Thus, wireless networks are at significant risk for having network communications intercepted and monitored by unauthorized parties. To mitigate the risks created by poor encryption and signal monitoring, it has become commonplace to segregate wireless connectivity from the rest of the corporate LAN. As shown in Figure 13-3, administrators have augmented wireless control mechanisms with VPN solutions to provide strong authentication and encryption of wireless traffic to achieve appropriate levels of security for wireless data and for accessing internal resources.

Finally, network design must also factor in the impact of the explosion of mobile devices into the wireless network, the ways in which the wireless design needs to support and accommodate many more varieties of devices, and how that in turn is forcing the advancement of technologies like mobile device fingerprinting and identity management. The sheer volume of mobile devices has created significant security challenges and unanticipated risks for the wireless network, creating a new dynamic that has expanded the network beyond traditional boundaries.

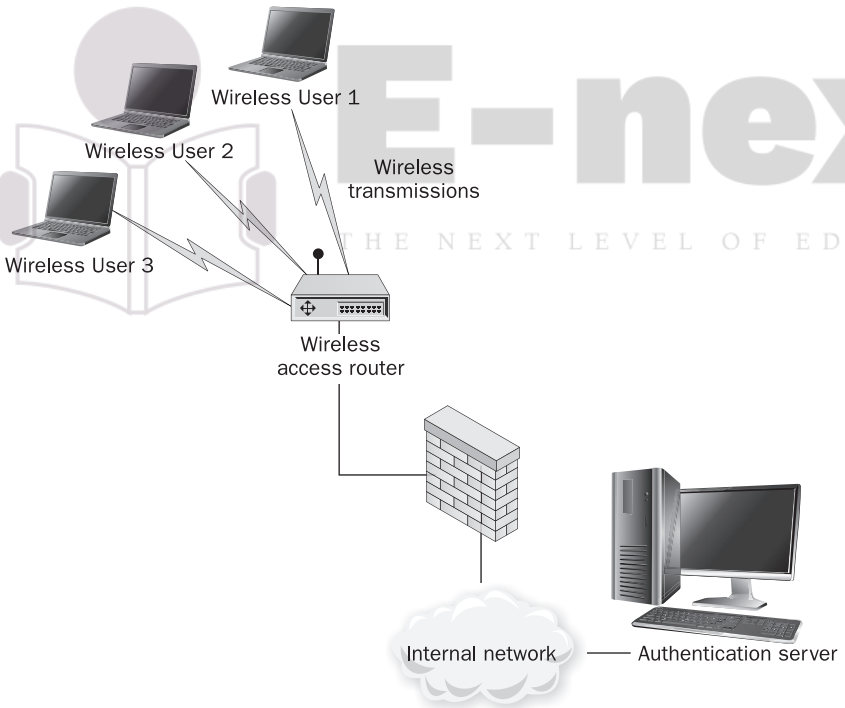

Figure 13-3 Wireless deployment through a VPN server

Remote Access Considerations

Most corporate networks permit user access to internal resources from remote locations. Historically this was done via a dial-up connection to an internally maintained modem bank. While some corporations still maintain dial-up access as a backup or secondary solution, remote access is now generally provided via a VPN solution. This type of VPN, which connects remotely located people to the organization's network, is a remote access VPN (as distinguished from a site-to-site or LAN-to-LAN VPN, which connects two networks together). VPNs provide a means to protect data while it travels over an untrusted network, they provide authentication services before permitting VPN traffic, and they function at network speeds. Chapter 16 provides more detailed information on VPNs.

Despite their usefulness, VPNs have a significant impact on the corporate network perimeter. Depending on how they are configured, VPNs can enable remote workstations to connect as if they were physically connected to the local network, though they remain outside the protection of the corporate security infrastructure. When VPN peers consist of remote users accessing the corporate network over the Internet, the overall security of the corporate network becomes dependent on the security of that employee's remote PC. Should a hacker gain access to an unprotected PC, the VPN may be used to tunnel traffic past the corporate firewalls and the protection they provide.

To protect the corporate network when VPNs are used for remote user access, security administrators should ensure that adequate protection is implemented over the endpoints. Most major firewall and VPN vendors include firewalling functionality in their clients.

While a hijacked VPN tunnel may seem like a remote possibility, it has happened. In October of 2000, sensitive Microsoft internal systems were accessed. The intrusion was traced to a VPN user PC that had been compromised by an e-mail worm known as Qaz. This event also points out another highly dangerous element of VPNs, the ability to propagate viruses. Home users are not protected by the up-to-date corporate antivirus infrastructure when they use their Internet and external e-mail accounts. These risks should be considered and mitigated when deploying VPNs. Posture validation, which is a feature of many remote access VPN products, is a technique for checking the security software and configuration of remote systems before they are allowed to connect to the network. It's a good way to reduce the risk of unsecure, infected, or compromised systems spreading risks onto the organization's network.

Internal Security Practices

Organizations that deploy firewalls strictly around the perimeter of their network leave themselves vulnerable to internally initiated attacks, which are statistically the most common threats today. Internal controls, such as firewalls and early detection systems (IDS, IPS, and SIEM, as described in Chapter 18), should be located at strategic points within the internal network to provide additional security for particularly sensitive resources such as research networks, repositories containing intellectual property, and human resource and payroll databases.

Dedicated internal firewalls, as well as the ability to place access control lists on internal network devices, can slow the spread of a virus. Figure 13-4 depicts a network utilizing internal firewalls.

When designing internal network zones, if there is no reason for two particular networks to communicate, explicitly configure the network to block traffic between those networks, and log any attempts that hosts make to communicate between them. With modern VoIP

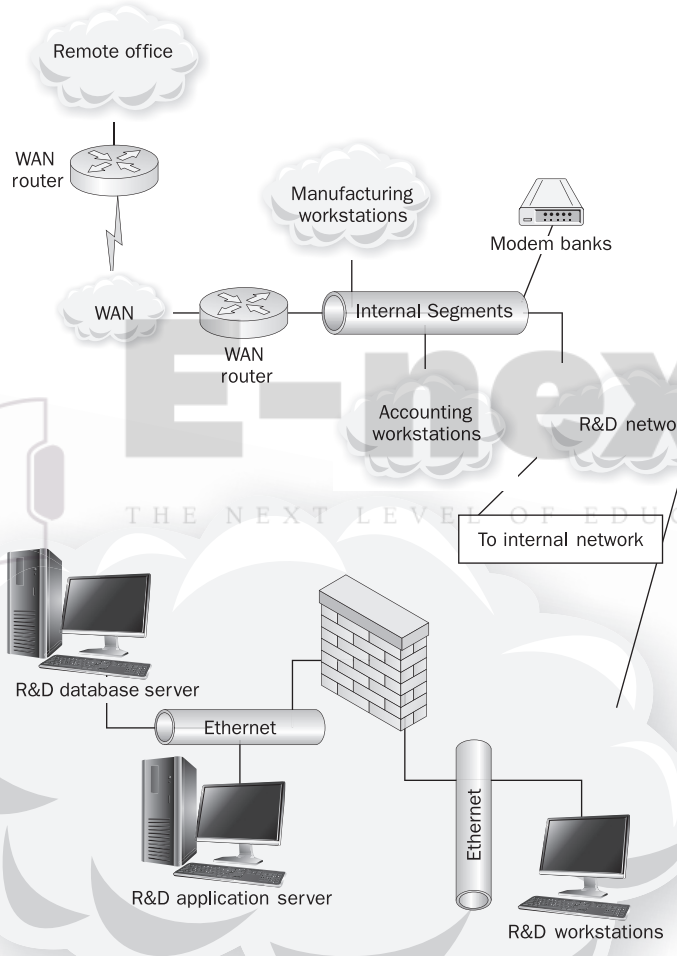

Figure 13-4 Internal firewalls can be used to increase internal security.

networks, this can be a challenge as VoIP streams are typically endpoint to endpoint, but consider only allowing the traffic you know to be legitimate between any two networks. A common technique used by hackers is to target an area of the network that is less secure, and then work their way in slowly via "jumping" from one part of the network to another. If all of the internal networks are wide open, there is little hope of detecting, much less preventing, this type of threat vector.

Intranets, Extranets, and DMZs

Organizations need to provide information to internal and external users and to connect their infrastructure to external networks, so they have developed network topologies and application architectures that support that connectivity while maintaining adequate levels of security. The most prevalent terms for describing these architectures are *intranet*, *extranet*, and *demilitarized zone (DMZ)*. Organizations often segregate the applications deployed in their intranets and extranets from other internal systems through the use of firewalls. An organization can exert higher levels of control through firewalling to ensure the integrity and security of these systems.

Intranets

The main purpose of an *intranet* is to provide internal users with access to applications and information. Intranets are used to house internal applications that are not generally available to external entities, such as time and expense systems, knowledge bases, and organization bulletin boards. The main purpose of an intranet is to share organization information and computing resources among employees. To achieve a higher level of security, intranet systems are aggregated into one or more dedicated subnets and are firewalled.

From a logical connectivity standpoint, the term *intranet* does not necessarily mean an internal network. Intranet applications can be engineered to be universally accessible. Thus, employees can enter their time and expense systems while at their desks or on the road. When intranet applications are made publicly accessible, it is a good practice to segregate these systems from internal systems and to secure access with a firewall. Additionally, because internal information will be transferred as part of the normal application function, it is commonplace to encrypt such traffic. It is not uncommon to deploy intranet applications in a DMZ configuration to mitigate risks associated with providing universal access.

Extranets

Extranets are application networks that are controlled by an organization and made available to trusted external parties, such as suppliers, vendors, partners, and customers. Possible uses for extranets are varied and can include providing application access to business partners, peers, suppliers, vendors, partners, customers, and so on. However, because these users are external to the corporation, and the security of their networks is beyond the control of the corporation, extranets require additional security processes and procedures beyond those of intranets. As Figure 13-5 shows, access methods to an extranet can vary greatly—VPNs, direct connections, and even remote users can connect.

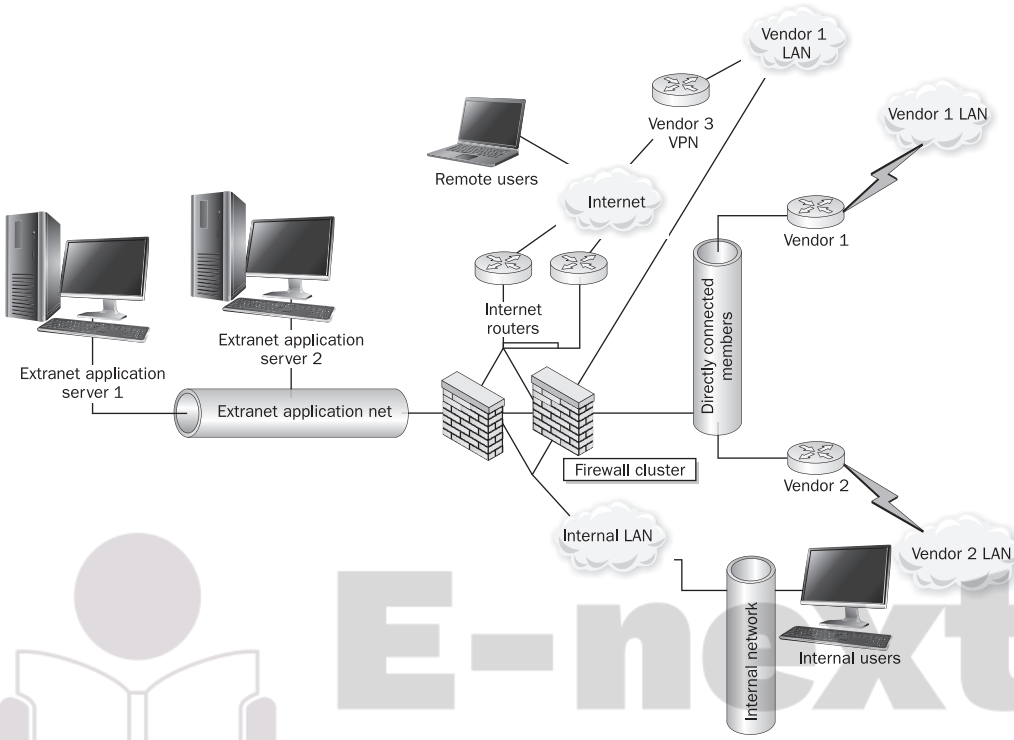

Figure 13-5 A possible extranet design

DMZ Networks and Screened Subnets

An organization may want to provide public Internet access to certain systems. For example, for an organization to receive Internet e-mail, the e-mail server must be made available to the Internet. As shown in Figure 13-6, it is good practice to deploy these systems on a dedicated subnet, commonly referred to as a *demilitarized zone (DMZ)* or *screened subnet*, separate from internal systems. Because these systems are publicly accessible, they can and will come under attack from malicious users. By housing them on a segregated network, a successful attack against these systems still leaves a firewall between the successful attacker and more sensitive internal resources.

NOTE While the terms DMZ and screened subnet have been used interchangeably, there is a small difference between the two terms. A DMZ is technically the small subnet between your Internet router and the external interface of your firewall. A screened subnet is really an isolated network available only through a firewall interface and is not directly connected to the internal network. The term DMZ was originally a military term used to describe a buffer area between a trusted zone and an untrusted zone, in which no military hardware was permitted.

As the number of publicly accessible systems grows, it is commonplace to create multiple DMZs to limit the breadth of a single security breach. For example, a corporation that puts its

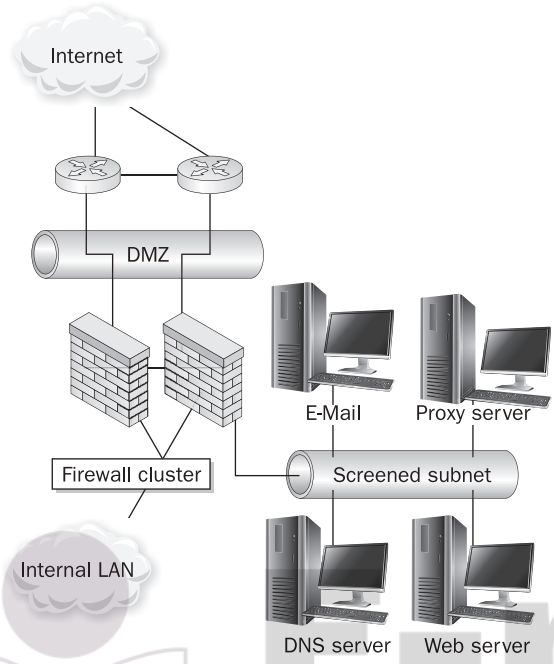

Figure 13-6 A sample DMZ configuration

web servers and its e-mail system in different DMZs protects each system from a vulnerability in the other. If a hacker is able to exploit a flaw in the web server, a firewall still stands between the hacker and the e-mail system.

Multiple DMZs can also be used to separate components of a single application system. As shown in Figure 13-7, application systems can consist of three separate tiers, referred to as the presentation, application, and database tiers. The *presentation* layer consists of a web server that interacts with end users, accepting input, sending that input to the application layer for processing, and returning the output back to the end user. The *application* layer contains the logic necessary for processing those queries and extracting the data that is stored in a database housed on a separate database server. Other services that aren't directly supporting the application but provide other functions can be further segregated into a fourth DMZ subnet.

Outbound Filtering

Up to this point, we have focused almost entirely on securing inbound access to a corporate network. While it may not initially be obvious, outbound filtering of network traffic can be nearly as important. Failure to restrict outbound access creates a number of significant risks to the corporation and its infrastructure, such as users accessing services that do not comply with corporate security policies or that do not have legitimate business purposes.

Additionally, failure to filter traffic leaving the corporate network may allow an attacker to use the network to launch attacks on other networks. There are precedents for organizations

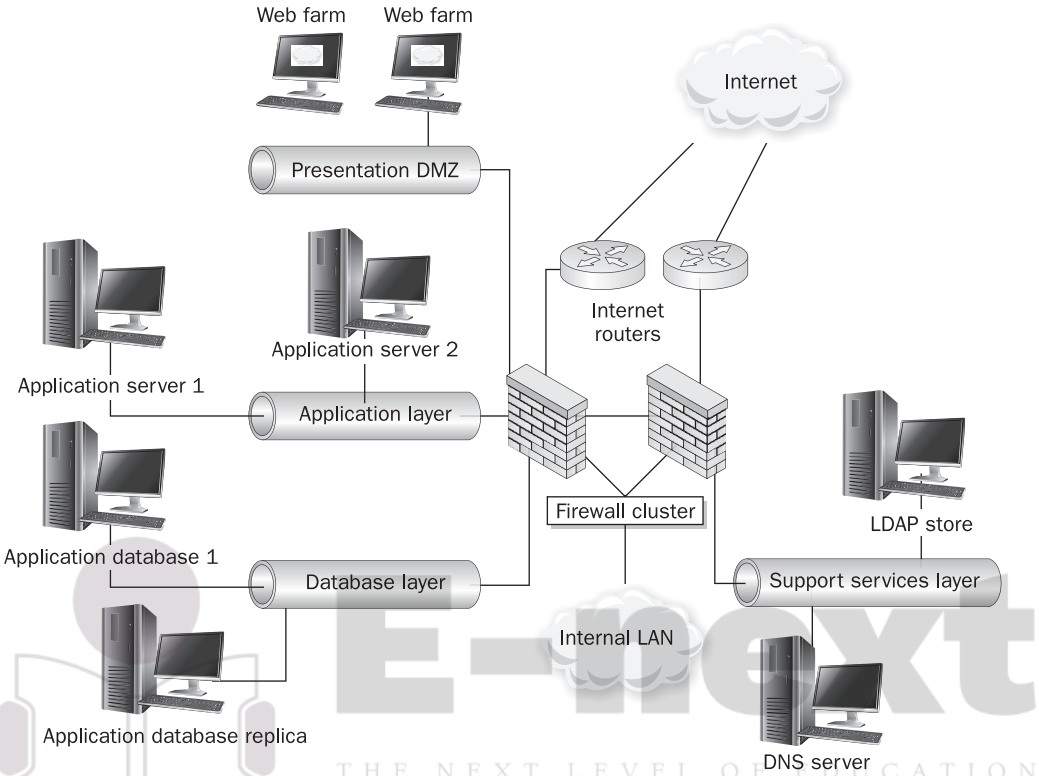

Figure 13-7 Example of a multitier application infrastructure

being held legally liable for the behaviors of their employees (and networks), in some cases due to a lack of effort in securing outbound traffic. This means that there is potentially significant liability for organizations that don't properly control their outbound network traffic.

Web Access Considerations

As Chapter 15 discusses, it is possible to prevent direct connections between internal and external users via proxy services or web filtering. Proxy servers can be configured to block connections to URLs that are considered likely to be malicious or unnecessary for normal operation, such as those containing certain scripts or other executable files. Proxy services are hardened processes that can run internally on a firewall or be provided separately by a dedicated server. Web filtering today can be handled via a variety of specialized products and appliances, including some cloud-based offerings.

The use of a proxy service gives a corporation several additional options when controlling user traffic. For example, the corporation may wish to scan downloaded files for viruses before transmission to the final user. A proxy server can also log, record, and report on user Internet usage, which can deter employees from wasting their days browsing web sites or visiting web sites not appropriate or relevant to their job function.

Beyond protecting users' browsers, corporations may wish to filter employee web access for a number of additional reasons. The Internet is filled with many interesting things that may not have a legitimate business use. Access to such distractions reduces employee productivity and consumes costly resources. For example, high-bandwidth music and video downloads can quickly saturate an organization's Internet link, slowing other critical business systems that share the connection. In addition, it is also common for organizations to implement acceptable Internet usage policies for their employees. To reduce the temptation to access non-business sites and to enforce such policies, corporations may wish to restrict the web sites that employees can access.

Outbound Port Filtering

Outbound filtering goes way beyond simple web site filtering. Another reason to filter outbound traffic is to ensure that only authorized traffic traverses controlled links. While this may seem like a terribly obvious statement, users and application developers left to their own devices will build and deploy applications without understanding the security risks they are bringing down on the organization (and other organizations to which the enterprise is connected).

To restrict outbound access, it is necessary to implement outbound filters on perimeter firewalls. As with inbound access, restrictive filters will limit which services can be used by default. This will also require security administrators to relax filters as new applications are deployed and business requirements demand access to new services.

By limiting outbound traffic to authorized applications, outbound filtering will prevent users from using applications that are dangerous or are not business related in the corporate environment. It can also reduce the chance that the organization network can be used to launch an attack against another network—such an attack could damage or cause losses for its victim, and the organization could end up being sued. Regardless of the outcome of that proceeding, it is expensive and time consuming to mount a defense, and it can focus negative publicity on the organization's security practices. To simply avoid the risk of a lawsuit, it is prudent to block unneeded access at the corporate perimeter.

Compliance with Standards

If you are following a specific security framework, here's how NIST, ISO 27002, and COBIT tie in to this chapter. NIST is mainly focused on wireless, COBIT has some general high-level guidance without going into details, and ISO 27002 provides the most specific guidance for network design considerations.

NIST

The following NIST Special Publications offer specific guidance for securing wireless networks:

- SP 800-153: Guidelines for Securing Wireless Local Area Networks (WLANs)
- SP 800-120: Recommendation for EAP Methods Used in Wireless Network Access Authentication
- SP 800-97: Establishing Wireless Robust Security Networks: A Guide to IEEE 802.11i
- SP 800-48: Guide to Securing Legacy IEEE 802.11 Wireless Networks

ISO 27002

ISO 27002 contains the following provisions, to which this chapter's contents are relevant:

- **10.1.4** Development and testing facilities are separated from operational facilities. Where necessary, development and production networks should be separated from one another.
- **10.4.1** All traffic originating from untrusted networks is checked for malware.
- **11.4.3** Access to the network is limited to specifically identified devices or locations.
- **11.4.5** Groups of computers, users, and services are segregated into logical network domains protected by security perimeters.
- **11.4.6** Network traffic is filtered by connection type, such as messaging, e-mail, file transfer, interactive access, and applications access.
- **10.6.1** Network controls including management and remote access should have effective operational controls such as separate network and system administration facilities; responsibilities and procedures for management of equipment are established; and special controls safeguard confidentiality and integrity of data processing over the public network.
- **10.6.2** Security features, service levels, and management requirements of all network services are identified and included in any network services agreement.
- **11.4.1** Protection of network services are defined, including parts of the network to be accessed, authorization services to determine who is allowed to do what, and procedures to protect the access to network connections and network services.
- **11.4.2** User authentication for external connections is performed with an authentication mechanism for challenging external connections.
- **11.4.3** Equipment is identified before it is allowed on remote connections.
- **11.4.4** Access (physical and logical) to diagnostic ports is protected by a security mechanism.
- **11.4.5** Networks are segregated using perimeter security mechanisms such as firewalls.
- **11.4.6** Network connection controls are used for services that extend beyond the organizational boundaries.
- **11.4.7** Network routing control, based on the positive source and destination identification, is used to ensure that computer connections and information flows do not violate the access control policy of the business applications.
- **11.6.1 and 11.1.1** Access is restricted based on a defined access control policy.
- **11.6.2** Systems on the network are segmented and isolated based on their risk or sensitivity.

COBIT

COBIT contains the following provisions, to which this chapter's contents are relevant:

- **DS5.9** Use preventive, detective, and corrective measures, especially regular security patching and virus control, across the organization to protect against malware such as viruses, worms, spyware, and spam.
- **DS5.10** Use firewalls, security appliances, network segmentation, and intrusion detection to manage and monitor access and information among networks.

Summary

The ultimate goal of network security is to enable authorized communications while mitigating information risk to acceptable levels. Design elements such as segregating and isolating high risk or other sensitive assets as well as defining and maintaining a strong network perimeter go a long way toward achieving those goals. As networks become ever more interconnected, a thorough and strongly typed network architecture/design will be required to achieve and maintain a well-secured network.

References

Convery, Sean. *Network Security Architectures*. Cisco Press, 2004.

Ghosh, Sumit, and H. Lawson. *Principles of Secure Network Systems Design*. Springer, 2001.

Northcutt, Stephen, et al. *Inside Network Perimeter Security*. New Riders Publishing, 2005.

Strassberg, Keith, Richard Gondek, and Gary Rollie. *Firewalls: The Complete Reference*. McGraw-Hill/Osborne, 2002.

Zwicky, Elizabeth, Simon Cooper, and D. Brent Chapman. *Building Internet Firewalls*. 2nd ed. O'Reilly & Associates, Inc., 2000.