CHAPTER 3

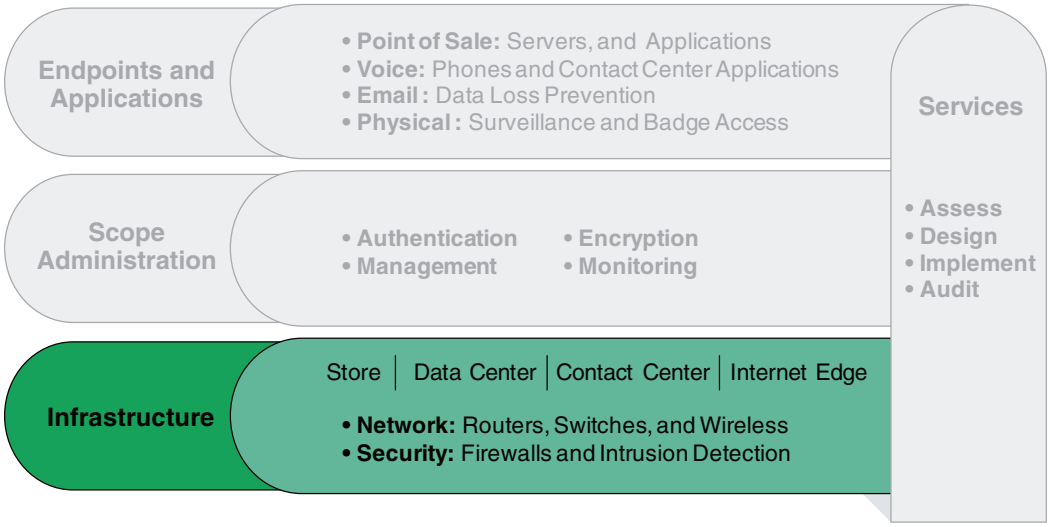
Solution Architecture

The Cisco PCI Solution for Retail 2.0 is a set of architectures, strategic principles, and tactical designs that details a holistic approach to addressing the requirements of PCI DSS 2.0. The Cisco Connected Retail architecture is used as a baseline for demonstrating the range of places that typically exist within an enterprise retailer. This chapter describes the Connected Retail Architecture in detail, so that when the discussion of specific PCI controls is discussed, the controls can be placed in context with that enterprise-wide view. The solution looks at an enterprise from an end-to-end perspective; from the store, where someone swipes the credit card, to the back-end of the data center, where the transaction leaves the retailers network to be processed by the acquiring bank. For more information on the Cisco Connected Retail Architecture, see http://www.cisco.com/go/retail.

For more information on the individual components used to build these architectures, see Chapter 4, "Component Assessment."

Chapter 2, "PCI and the Solution Framework," describes the elements that make up the solution framework. The solution framework organizes the scope of the cardholder data environment for contextual reference. The bottom layer of the model shows the organization of the enterprise into places such as the store, data center, and the Internet edge. (See Figure 3-1.)

Figure 3-1 Solution Framework

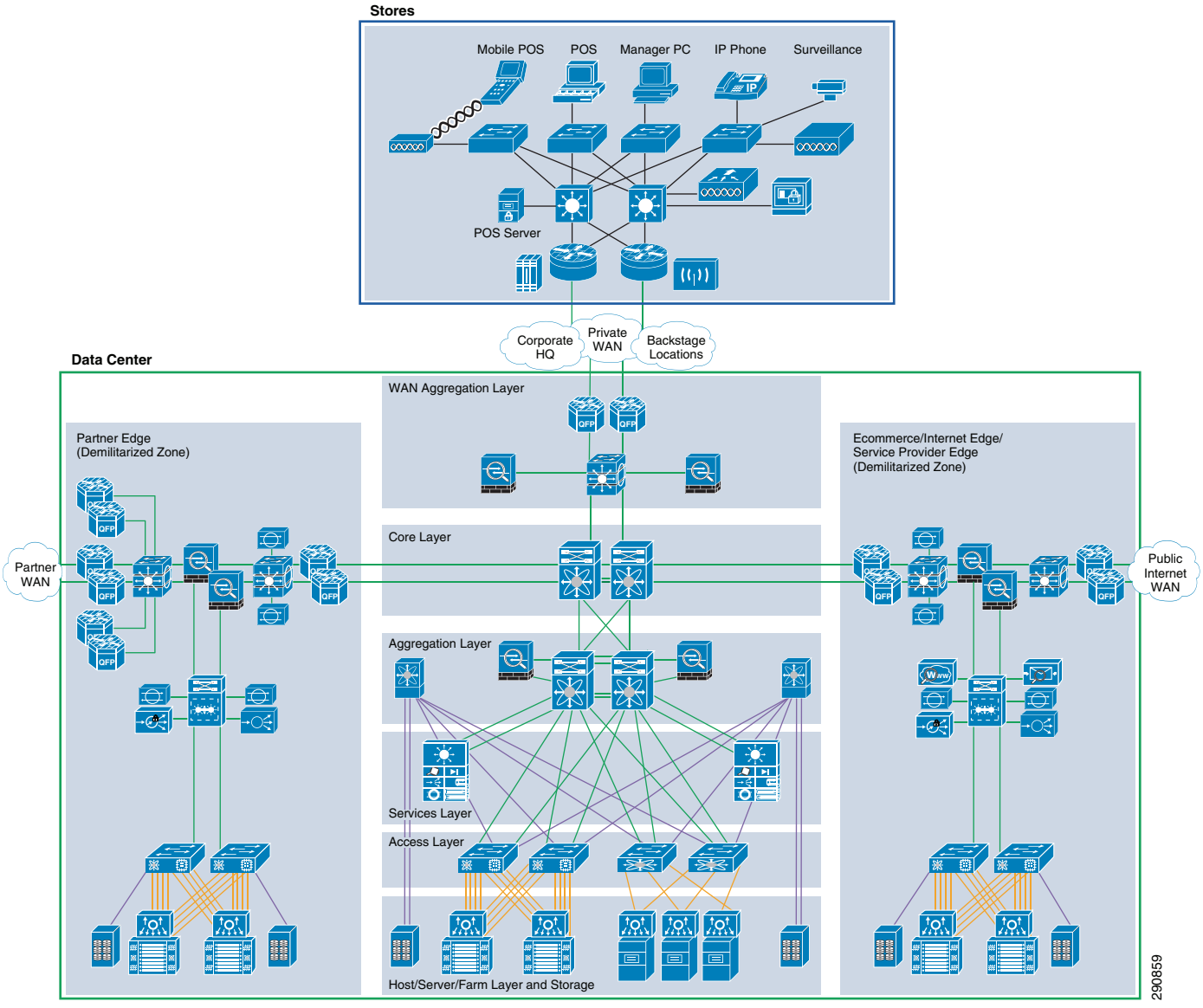


Enterprise Architecture and PCI Design Considerations

PCI compliance affects the overall enterprise architecture, depending on the requirements of the business. For example, a new business requirement for direct customer Internet connectivity at the store level extends the firewall and IDS/IPS perimeter requirements to the branch level, whereas before it might exist only at the headend data center. Without this contextual reference, it is difficult to discuss specific controls.

Figure 3-2 shows the enterprise-wide retail reference architecture and locations that commonly exist in the enterprise retailer domain.

Figure 3-2 Enterprise-wide Retail Reference Architecture



The following sections describe the major places affected by PCI compliance throughout the enterprise. Each section provides design considerations that are affected by PCI controls in more detail.

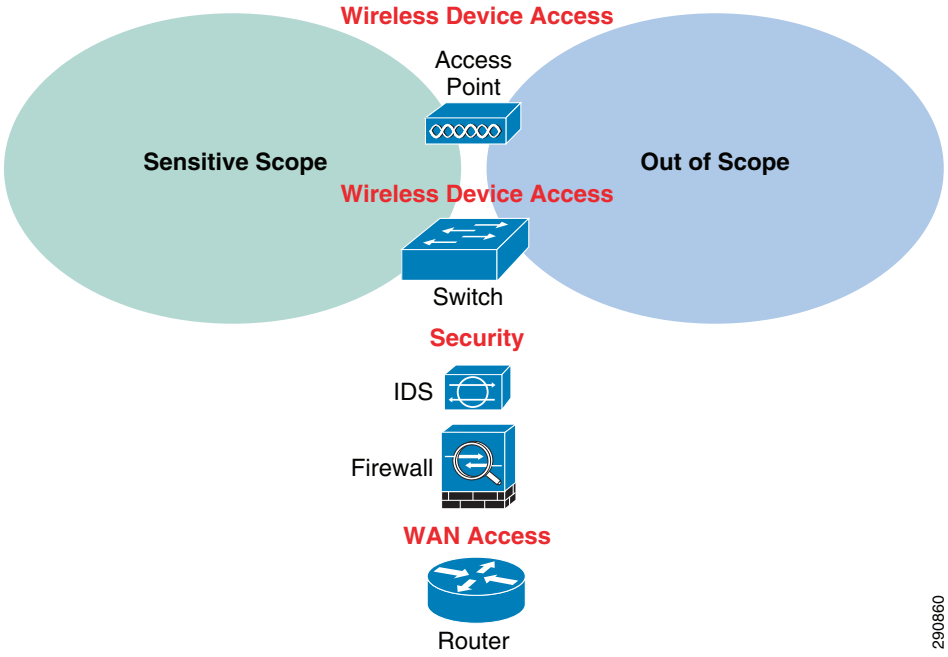
Store Architecture

The store is the location where customers swipe their credit cards to purchase goods. Depending on the type of services that are offered at the store, various levels of security are required. This section discusses those design considerations and relates them to various store formats.

Design Considerations

Figure 3-3 shows the fundamental infrastructure components used within a store location. These components are used in conjunction with each other to segment sensitive data from non-sensitive data. The process of segmenting the network into *scopes* allows a merchant to reduce the amount of branch-level components that need to be audited. Note that devices/endpoints themselves may be cut out of the scope of an audit by putting them onto their own network, but the actual network infrastructure may not necessarily be decreased. For example, a switch can have devices that are both sensitive and non-sensitive attached to it. By putting the non-sensitive devices onto their own VLANs, they can be cut out of the audit by using the VLAN function of the switch. However, the switch itself still remains in scope.

Figure 3-3 Fundamental Store Infrastructure Components



Each store component is used for a different function, as follows:

- The router function can be used for:
  - Accessing the WAN
  - Routing between VLANs
  - Access control lists
- The firewall can be used for:
  - Filtering unnecessary or inappropriate data via a stateful firewall
  - Routing between VLANs

  - Detecting and preventing intrusions; (IPS/IDS devices can also be separate appliances)
- Intrusion Prevention Systems (IPS)/Intrusion Detection Systems (IDS) monitor for anomalous behavior on the network and send alerts.
- The switch can be used for:
  - Segmenting via VLANs
  - Accessing wired devices
- The access point can be used for:
  - Wireless segmentation
  - Accessing wireless devices

The function of each of these devices can be virtualized and consolidated for simplicity, depending on the space and management requirements of the store footprint. For example, some smaller box stores have power, wiring closet, rack, and cabling restraints that would benefit from virtualized devices that reduce the physical footprint of the branch infrastructure.

Conversely, each of these devices can be increased in number depending on the resiliency and redundancy requirements of the business. For example, if store connectivity is a business priority, using redundant routers for redundant WAN access might be a requirement to ensure that store connectivity is maintained.

Regardless of how the store is designed from a redundancy or scale perspective, the same types/locations of controls are consistent across them.

Many retailers use their data center as their centralized location to connect to public networks such as the Internet. This perimeter is typically secured as a demilitarized zone (DMZ) using firewalls and IDS/IPS. Whenever you introduce any type of untrusted network (wireless, Internet, microwave, satellite, cellular, and so on) into the store environment, you have effectively created a new external perimeter that must now be secured with a firewall and intrusion detection/prevention system. Table 3-1 defines the types of factors that affect store controls and requirements.

Table 3-1 Store Services and Corresponding Compliance Controls Located at Store

Store Service Type	Minimum PCI Control Required	Relevant Solution Component
Any store with point-of-sale (POS) systems	Rogue detection	Cisco Identity Services Engine (ISE), wireless IPS, 802.1x switch
POS systems; no direct Internet access, no wireless access, no untrusted networks of any type	Access control lists (ACLs), no state table required	Any router with ACLs
Basic wireless connectivity	Firewall, IDS	Cisco Integrated Services Router (ISR), Cisco Adaptive Security Appliance (ASA), Cisco IPS appliance
Wireless POS	Firewall, IDS, strong client encryption	Cisco ISR, Cisco ASA, Cisco IPS appliance, Cisco Unified Wireless
Public WAN	Firewall, IDS	Cisco Integrated Services Router (ISR), Cisco Adaptive Security Appliance (ASA), Cisco IPS appliance

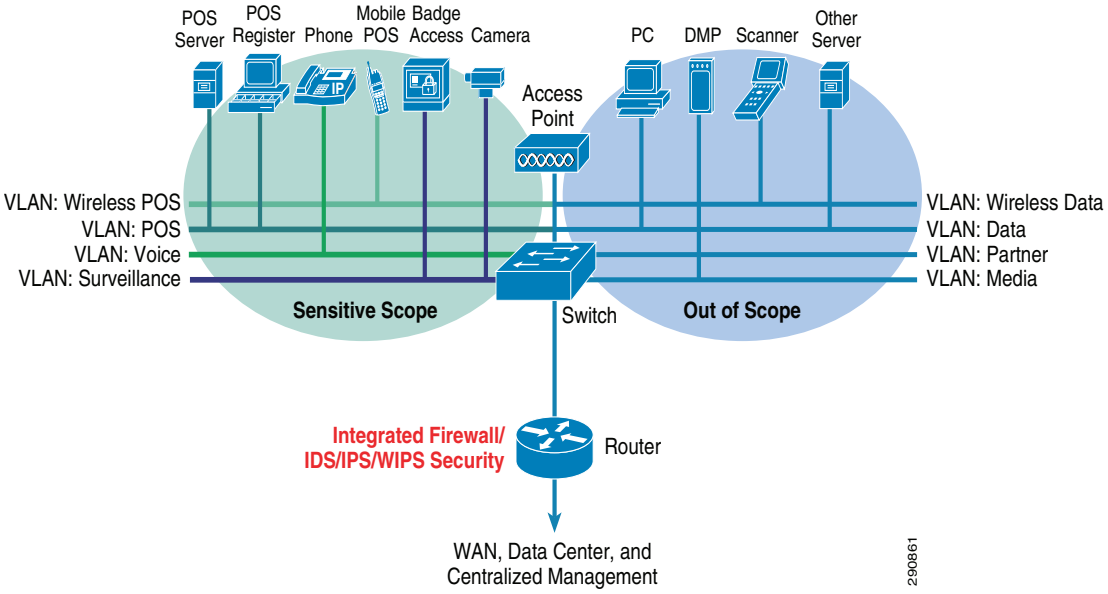
Table 3-1 Store Services and Corresponding Compliance Controls Located at Store (continued)

Internet connectivity	Firewall, IDS	Cisco Integrated Services Router (ISR), Cisco Adaptive Security Appliance (ASA), Cisco IPS appliance
Any untrusted network access	Firewall, IDS	Cisco Integrated Services Router (ISR), Cisco Adaptive Security Appliance (ASA), Cisco IPS appliance

The fundamental reference store architecture assumes that a retailer may eventually need to scale to these levels of services, but not necessarily immediately. From a store perspective, the Cisco Integrated Services Router (ISR) performs each of the functions listed in Table 3-1. This allows merchants to grow with their investment by purchasing a router that can scale by different license keys for different services without having to rip and replace. For example, a merchant can purchase a Cisco ISR for basic WAN connectivity. When the business wants to introduce wireless to the stores, the merchant can unlock the firewall/IPS/IDS feature set with a license.

The fundamental store reference architecture in Figure 3-4 shows the solution framework endpoints/applications within the context of the fundamental store component's infrastructure.

Figure 3-4 Fundamental Reference Store Architecture



In-scope devices can include the following:

- POS devices
- Wireless handheld devices
- Mobile POS
- Voice systems
- Physical badge access
- Video surveillance systems.

In general, an additional VLAN for management of infrastructure should be distinctly defined.

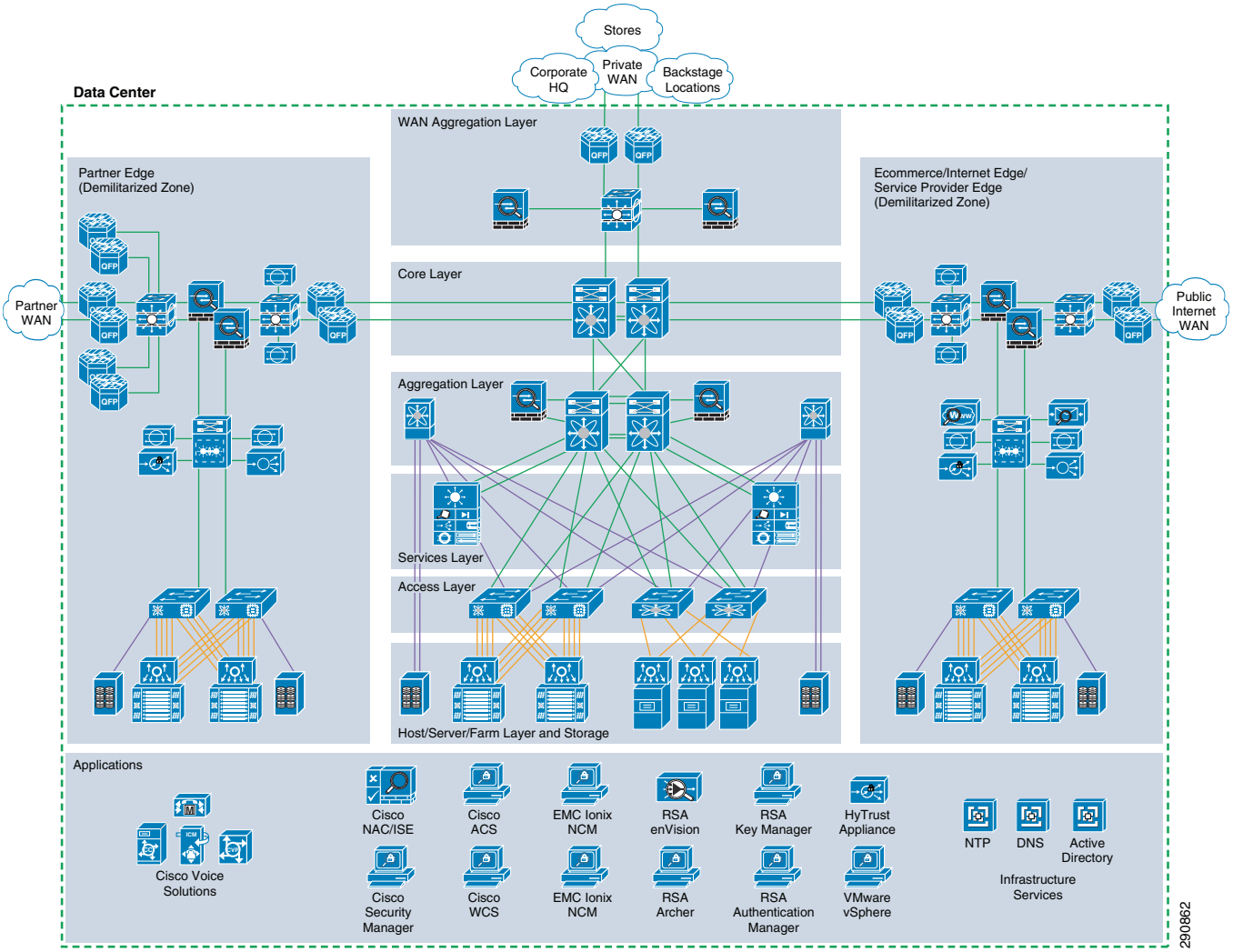
The remaining devices at the store level are considered *out-of-scope* and do not need to be audited, given that they are on their own network and segmented via firewall/IPS/IDS from the sensitive networks.

The PCI store model and its controls were applied to the small, medium, and large Connected Retail Store footprints. This section provides sample addressing plans used by various stores. Many designs can be extracted by understanding and using the PCI solution model shown above, but the overall functions are essentially the same.

Data Center

The data center is where centralized data processing, data storage, and data communications take place (see Figure 3-5). The data center is also the place where management systems are deployed. The data center provides centralized control from an administrative perspective because it is typically where the tools that are used to monitor and enforce compliance are deployed.

Figure 3-5 Data Center Architecture



Design Considerations

Design considerations are as follows:

- Centralized solution management that supports all aspects of network, security, and systems management; and supports remote access from anywhere on the network.
- Standardized equipment and software images, deployed in a modular, layered approach, that simplify configuration management and increase the availability of the system.
- A highly available data center design that permits highly resilient access from stores to core data and storage services.
- WAN aggregation alternatives that allow flexible selection of service provider network offerings.
- A service aggregation design that allows for a modular approach to adding new access layers and managing shared network services (for example, firewall, IDS, application networking, wireless management).
- Firewall, IDS, and application networking services that are available at the service and aggregation layers of the data center.
- Scalability that can accommodate shifting requirements in data center compute and storage requirements.
- Note that WAN access speeds are typically the limiting factor between the store network systems and the WAN aggregation layer. It is typical for retailers to over-subscribe the WAN circuits between the stores and the WAN edge aggregation router. Over-subscription can cause inconsistent results and packet loss of payment card information in the event that more traffic enters the WAN circuit simultaneously.
- Backup network connections from store networks to the data center are recommended when payment card information is transported via the WAN.

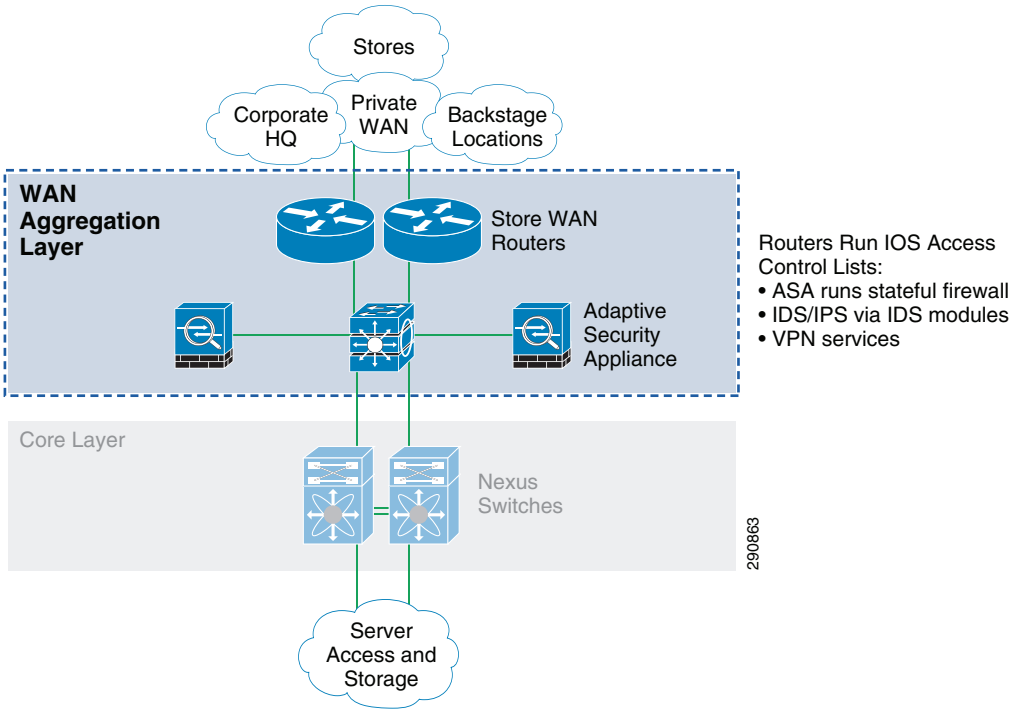
Data centers can house many types of functions, and the term itself can encompass narrow and broad aspects. For the purposes of this guide, data centers include the following functions:

- WAN aggregation layer—Aggregates the store and backstage WAN connections to the core
- Core layer—Highly available, high-speed area that is the central point of connectivity to all data center areas
- Aggregation layer—Aggregates the services of one area and connects that area to the core
- Services layer—Data treatment and manipulation occurs between the access layer and aggregation layer
- Access layer—Server-level access and connectivity between hosts/servers to the services and aggregation layers, depending on the nature of the application
- Host/server farm—Physical servers, virtualized servers, and appliances' host applications
- Storage—Storage area networks (SANs)
- E-commerce—Internet-based transactions
- Internet/service provider edge demilitarized zone (DMZ)—Secure connectivity to the Internet
- Partner edge DMZ—Secure segmented connectivity to partners

WAN Aggregation

The WAN aggregation layer is a transit network that aggregates the connections from the stores, backstage locations, and corporate offices, as shown in Figure 3-6.

Figure 3-6 WAN Aggregation Layer



Design Considerations

The WAN edge routers should not also be used as the Internet gateways for the data center network. By clearly defining zones of separation of responsibility within the infrastructure, it is easier to maintain.

Two options are possible at this layer for Layer 3 filters at the WAN aggregation layer:

- Firewall appliance—Interior to the WAN edge routers, a dedicated firewall appliance is used to secure incoming WAN traffic and to terminate store VPN connections. This design provides the highest scalability.
- Cisco IOS Software firewall routers—Many Cisco routers also support the Cisco IOS Security Software option that includes a firewall feature. Cisco recommends the use of the Cisco IOS Security feature set in stores, branches, and teleworker deployments, because of a much lower number of users and connection rates than at the store WAN aggregation headend location.

There are two typical WAN speeds categories for a WAN aggregation network: less than and up to OC3 (155 Mbps), and OC12 (622 Mbps) and above. The choice of these two network speeds determines the platform set to select from Cisco. In addition, this design creates two profiles for each WAN speed. These profiles are designed to provide guidance when designing a WAN edge network, regardless of which enterprise WAN architecture is selected. The profiles for each WAN speed investigate integrated versus dedicated chassis for each functionality component, as highlighted in the previous section. Some

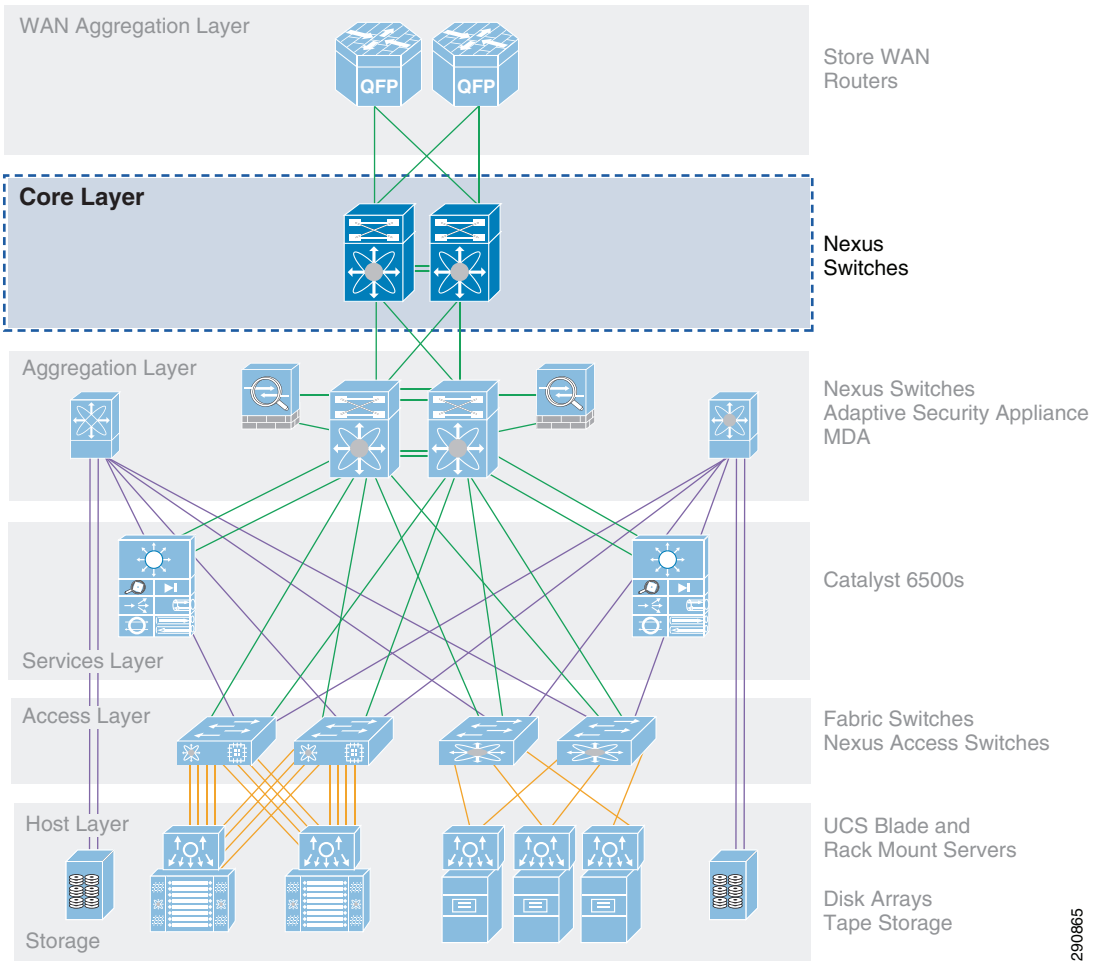
customers prefer a highly integrated solution where most, if not all, of the WAN edge functions described in this document reside on a single or very few network devices. Other customers prefer the granularity and scalability of these same functions separated across multiple network devices.

The WAN aggregation architecture is based on the *Infrastructure Protection and Security Service Integration Design for the Next Generation WAN Edge v 2.0*, which can be found at the following URL: http://www.cisco.com/en/US/docs/solutions/Enterprise/WAN_and_MAN/IPSNGWAN.html

Core Layer

The core layer provides the high-speed packet switching backplane for all flows going throughout of the data center, as shown in Figure 3-7.

Figure 3-7 Core Layer



Design Considerations

The core layer provides connectivity to multiple aggregation layers and provides a resilient Layer 3 routed fabric with no single point of failure. The core layer runs an interior routing protocol, such as Open Shortest Path First (OSPF) or Enhanced Interior Gateway Routing Protocol (EIGRP), and load balances traffic between the core and aggregation layers using the Cisco Express Forwarding-based hashing algorithms.

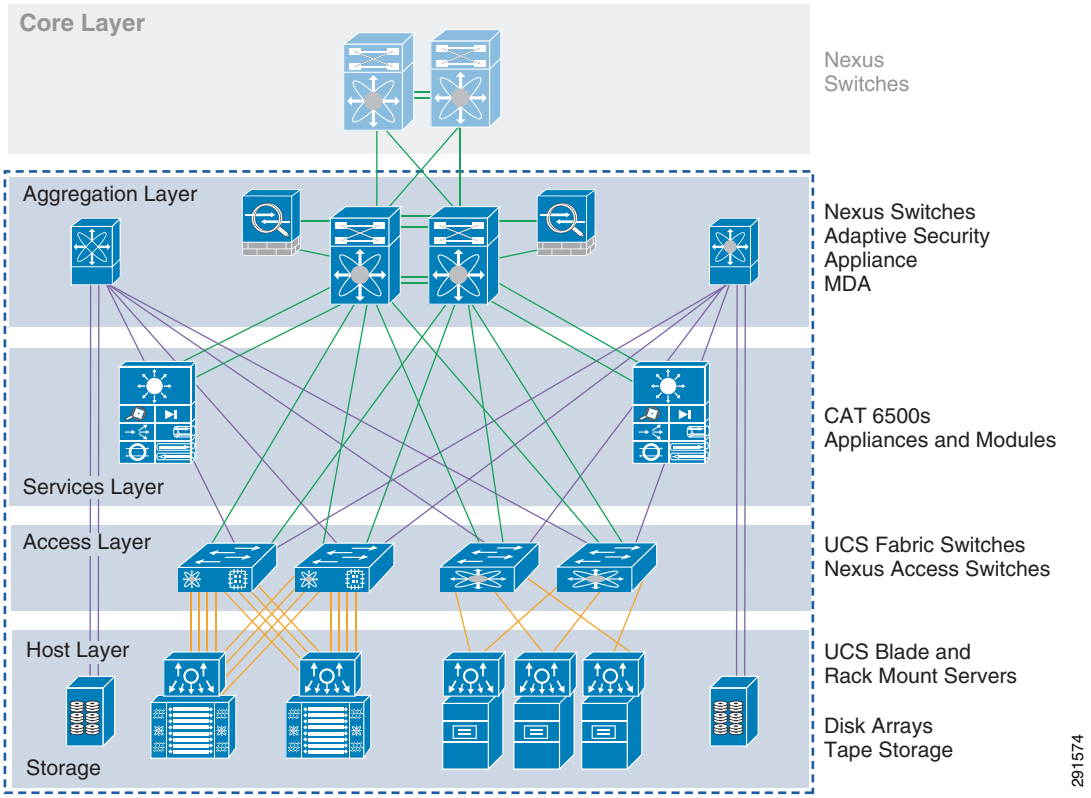
The core is not a perimeter; no security filtration should be performed at this layer.

The core, services aggregation, and server access tiers of the multi-tier data center architecture were based on the design documented in the *Cisco Data Center Infrastructure Design Guide 3.0*, which can be found at the following URL:
http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/DC_3_0/DC-3_0_IPInfra.html

Aggregation Block

An aggregation block is a combination of the aggregation, services, and access layer systems. It represents a repeatable, implementable template for scaling applications and services within the data center. (See Figure 3-8.)

Figure 3-8 Aggregation Block



Design Considerations

Zones are a best practice to isolate applications and services based on their individual policy requirements. You can securely mix in-scope and out-of-scope applications and services within a single aggregation block.

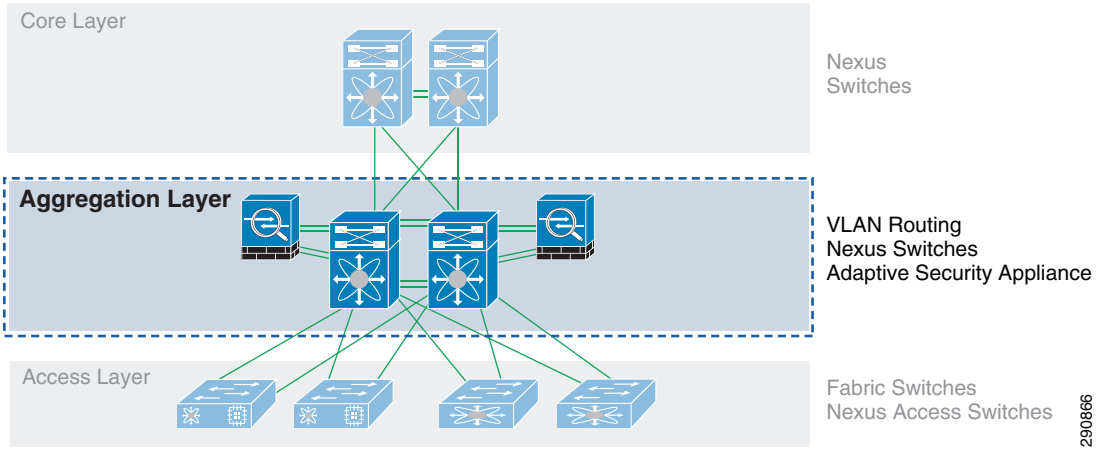
The layers that comprise the aggregation block are described in more detail below.

For more information, see the following URL:
http://www.cisco.com/en/US/solutions/ns340/ns414/ns742/ns743/ns994/landing_aggregationlayer.html

Aggregation Layer

The aggregation layer aggregates the connections from the services layer and the access layer to the centralized core, as shown in Figure 3-9.

Figure 3-9 Aggregation Layer



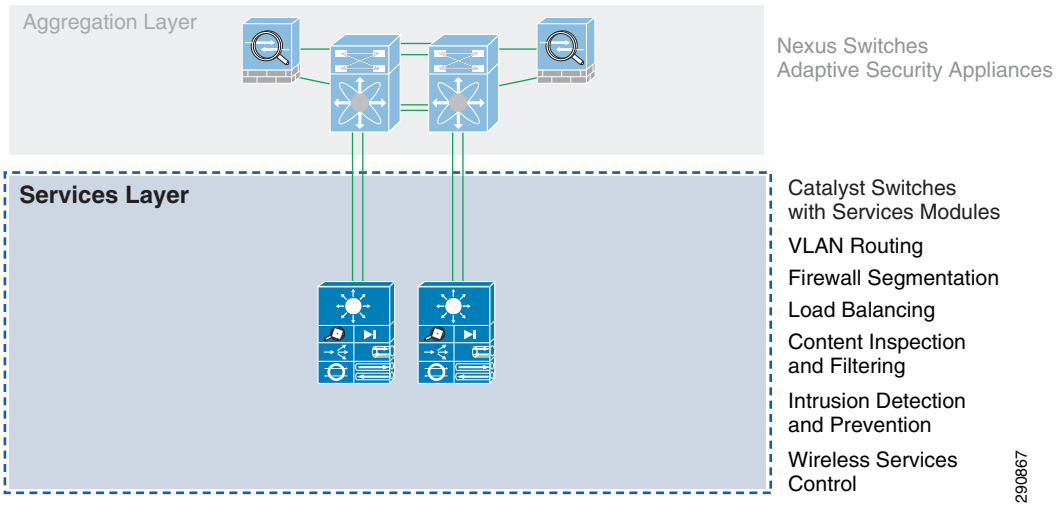
Design Considerations

The aggregation layer uses Layer 3 filters to segregate and protect the edge of the scope of compliance.

Services Layer

The services layer provides important functions, such as service module integration, Layer 2 domain definitions, spanning tree processing, and default gateway redundancy. (See Figure 3-10.)

Figure 3-10 Services Layer



Design Considerations

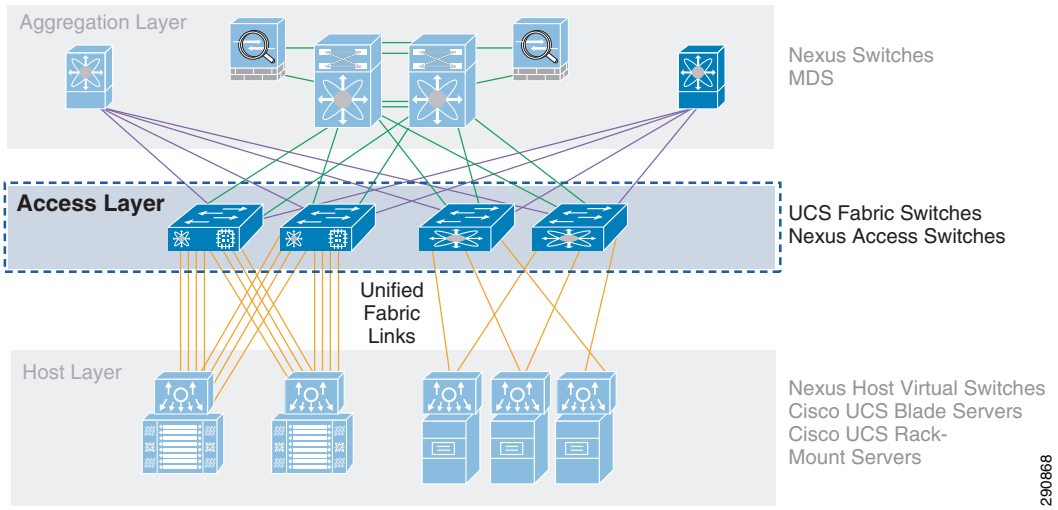
Services such as server load balancing and wide-area application services (WAAS) are used at this layer to optimize applications. Optimizing devices used within the scope of PCI are also brought into scope and are susceptible to the same controls as traditional network devices. For more information on understanding these controls, consult the capability assessment logic in Chapter 4, "Component Assessment."

Services such as content switching, SSL offload, intrusion detection, and network analysis are provided by hardware-based service modules or standalone appliances.

Access Layer

The access layer is where the servers physically attach to the network, as shown in Figure 3-11.

Figure 3-11 Access Layer



In typical data centers, the server components consist of appliances, 1RU servers, blade servers with integral switches, blade servers with pass-through cabling, clustered servers, and mainframes with OSA adapters. The access layer network infrastructure consists of modular switches, fixed configuration 1RU or 2RU switches, and integral blade server switches.

Design Considerations

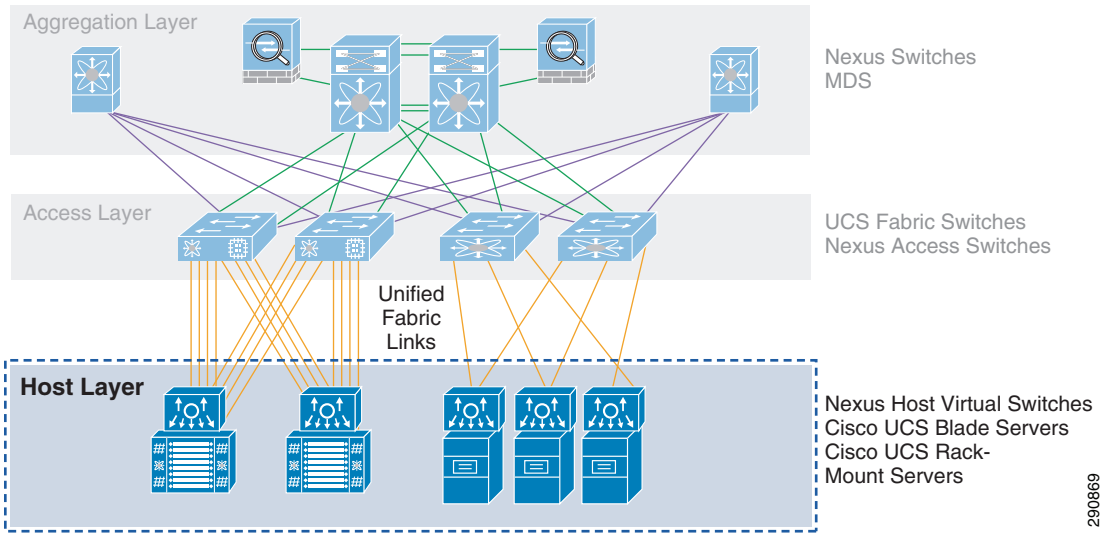
Switches provide both Layer 2 and Layer 3 topologies, fulfilling the various server broadcast domain or administrative requirements.

The solution management servers connect to the network in this layer. They are centralized, segmented from other business application servers, and protected by firewall services from the service aggregation layer above. Business servers, consisting of POS transaction log servers, database, and data warehouse servers also exist at this layer but are segmented via separate VLANs and firewall policy.

Host/Server Farm Layer

The host/server farm layer is where the centralized administrative applications reside, as shown in Figure 3-12.

Figure 3-12 Host/Server Farm Layer



Design Considerations

Network addressing should be used per business function. This allows the discrete manipulation of data traffic as requirements arise. For example, both POS applications and network management are used within the scope of PCI compliance but should be segregated onto their own subnets.

Virtualization technology can be used within a data center server farm. Individual blades within a blade server chassis can be used to segment sensitive and non-sensitive applications because they run independent hypervisors. Because hypervisors are considered insecure, when mixing sensitive applications with non-sensitive applications (mixed-mode) across the same hypervisor, the non-sensitive applications are now in scope.

For more information, see the PCI Virtualization Guidelines whitepaper at the following URL: https://www.pcisecuritystandards.org/documents/Virtualization_InfoSupp_v2.pdf.

Multiple internal Network Time Protocol (NTP) servers should be deployed for consistent log synchronization in the event of failure. Those internal NTP servers should use more than one external source in the event of an external failure.

Although virtualization can be used for a variety of services, NTP requires a high resolution system clock and accurate response times to clock interrupts that virtual machines cannot provide. For these reasons, it is recommended not to run NTP on virtual machines. Instead, NTP should be run on the base OS of the hypervisor, and the virtual machine should use VMware Tools Clock synchronization to sync with the base host. NTP servers should also not run on virtual machines but on physical devices (for example, on the Cisco Catalyst 6509 Services switches in the services layer of the data center aggregation block). For more details, see the following URL: http://www.vmware.com/files/pdf/Timekeeping-In-VirtualMachines.pdf.

Table 3-2 lists descriptions of applications for administrators.

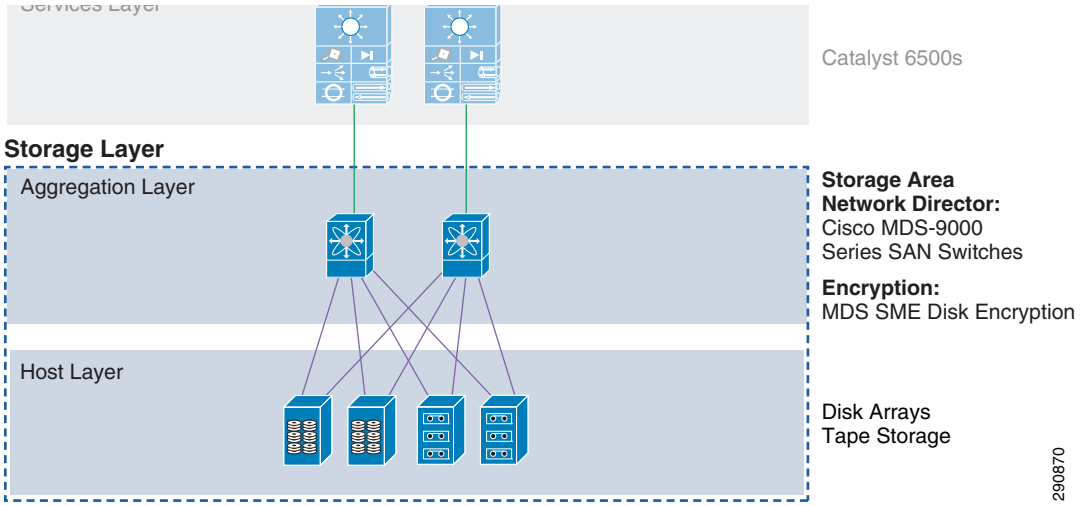
Table 3-2 Central Toolkit Description of Applications for Administrators

Function	Solution Component Options
Authentication	
Device AAA	Cisco Secure ACS, Cisco ISE
Two-factor remote	RSA Authentication Manager
Directory services	Active Directory
Network Management	
Device configuration	Cisco LMS, EMC Ionix NCM
Security configuration	Cisco Security Manager
Wireless configuration	Cisco WCS
Monitoring	
Event correlation	RSA enVision
Policy enforcement	EMC Ionix NCM
Corporate policy	RSA Archer
Virtualization	EMC Unified Infrastructure Manager, VMware vSphere
Physical Security	
Video surveillance	Cisco Video Surveillance Manager
Building access	Cisco Physical Access Manager
Encryption	
Storage	Cisco Key Manager, RSA Data Protection Manager
Remote access/VPN	Cisco Security Manager

Storage Layer

The storage layer is where sensitive data is stored, as shown in Figure 3-13.

Figure 3-13 Storage Layer



Design Considerations

A combination of disk encryption provided by Cisco MDS, Fibre-Channel zoning, and masking were used in the storage implementation of this solution. By deploying zoning within a Fibre Channel fabric, device access is limited to devices within the zone. This allows the user to segregate devices based on access to a particular storage device (disk array). This is a requirement in a data center environment in which multiple file servers in the data center server farm are connected to the same SAN fabric, and access to cardholder data must be restricted to a subset of servers. LUN masking takes zoning beyond the Fibre Channel switchport level, by restricting access to specific LUNs on a given disk array. Only specific devices belonging to the LUN zone are able to access those sections of the disk.

Encryption keys for storage are managed by Cisco Key Manager and RSA Data Protection Manager.

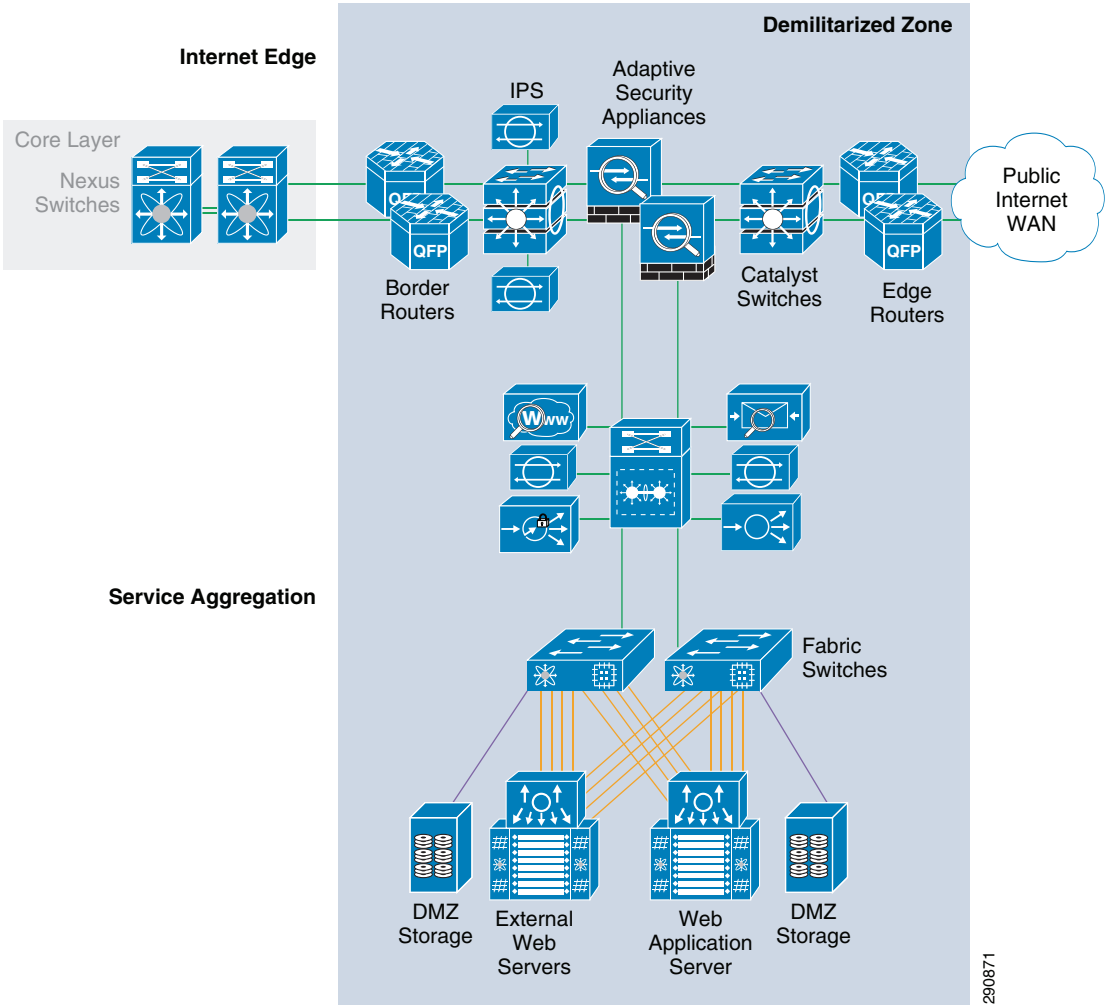
A subtle, yet potentially significant change to key management has been introduced with the PCI 2.0 standard. With past versions of the DSS, annual key rotations were required for encryption keys. DSS 2.0 now requires that keys are rotated at the end of their *cryptoperiod*, and references the NIST 800-57 Special Publication to determine what an appropriate cryptoperiod is. The NIST 800-57 Special Publication is a 324-page, three-part document. Merchants, and even QSAs, may not have the expertise to fully understand such a document that includes countless encryption scenarios, with cryptoperiods ranging from as short as a day to as long as three years.

In an ideal world, with all parties being expert cryptographers, this risk-based change to the standard would be very appropriate and most welcome. However, given the number of scenarios and criteria for determining an appropriate cryptoperiod, it could suggest that this change is too subjective and may become a point of contention between a merchant and QSA assessor, as to what is an appropriate cryptoperiod; whereas the former, more prescriptive control, did not allow for flexibility in this area.

E-commerce/Internet Edge/Service Provider Edge/Partner Edge

The solution uses a collapsed Internet edge and extranet network to support Internet connectivity and business partner connectivity, as shown in Figure 3-14.

Figure 3-14 E-commerce/Internet Edge/Service Provider Edge



Design Considerations

The design does the following:

- Provides an enterprise connection to the Internet
- Secures the Internet edge design using Cisco firewall and intrusion detection systems
- Provides a dual-threaded design for network resiliency
- Provides a collapsed Internet edge and extranet network for a highly centralized and integrated edge network
- Provides remote VPN access to enterprise users/telecommuters

This design takes into account best practices from the *Data Center Networking: Internet Edge Design Architecture Design Guide* (http://www.cisco.com/go/designzone) and customizes these recommendations for the Internet edge and extranet networks of retail businesses. The edges connect Internet services to the complete enterprise environment (that is, from headquarters to Internet service providers), and branch office connections that use a Cisco secure VPN to connect to headquarters. The collapsed design provides highly centralized and integrated edge networks, and transports the aggregated traffic through various service modules (Cisco ACE, Cisco FWSM, and Cisco IDSM2) within a pair of Cisco Catalyst 6500 Switch chassis. The Internet edge provides the following security functions:

- Secure configurations and management.
- IP anti-spoofing.
- Access control lists (ACLs) that provide explicitly permitted and/or denied IP traffic that may traverse between inside, outside, and DMZ.
- Stateful inspection provides the ability to establish and monitor session states of traffic permitted to flow across the Internet edge, and to deny traffic that fails to match the expected state of existing or allowed sessions.
- Intrusion detection using Cisco IDSM2 provides the ability to promiscuously monitor traffic across discrete points within the Internet edge, and to alarm and/or take action after detecting suspect behavior that may threaten the enterprise network.
- Applications servers that need to be directly accessed from the Internet are placed in a quasi-trusted secure area (DMZ) between the Internet and the internal enterprise network, which allows internal hosts and Internet hosts to communicate with servers in the DMZ.
- All public-facing web applications should be developed using the security best practices to prevent known attacks, and must be reviewed annually or after changes.