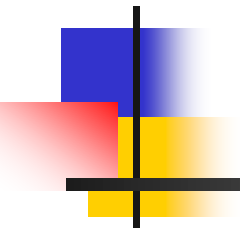


Data Center Network Infrastructure

Data Center Definition

- A data center
  - is a large amount of electronic equipment
    - computers
    - communications equipment
  - is usually maintained by an organization for handling the data operations.
  - enables the consolidation of critical computing resources in controlled environments, under centralized management.
  - permits enterprises to operate around the clock or according to their business needs.

Data Center Architectural Overview

- Data centers provide the following functions:
  - Ensuring network connectivity,
    - including switches and routers.
  - Providing network and server security,
    - including firewalls and intrusion detection systems (IDSs).
  - Enhancing availability and scalability of applications,
    - including load balancers, secure sockets layer (SSL) offloaders and caches.

Critical Requirements

- Designing the data center infrastructure :
  - High Availability—Avoiding a single point of failure and achieving fast and predictable convergence time
  - Scalability—Allowing changes and additions without major changes to the infrastructure, easily adding new services, and providing support for hundreds dual-homed servers
  - Simplicity—Providing predictable traffic paths in steady and failover states, with explicitly defined primary and backup traffic paths
  - Security—Prevent flooding, avoid exchanging protocol information with rogue devices, and prevent unauthorized access to network devices

dual-homed is one of the firewall architectures for implementing preventive security.

Data Center Architecture

- The data center infrastructure must provide:
  - High port density
  - Layer 2 (Data Link layer) connectivity
  - Layer 3 (Network layer) connectivity
- It must support security services provided by
  - Access control lists (ACLs)
  - Firewalls
  - Intrusion detection systems (IDS)
- It must support server farm services such as:
  - Content switching
    - is used to scale application services by front ending servers and load balancing of the incoming requests to those available servers.
  - Caching
  - Secure sockets layer (SSL)
- It must integrate:
  - Multi-tier server farms
  - Mainframes and mainframe services

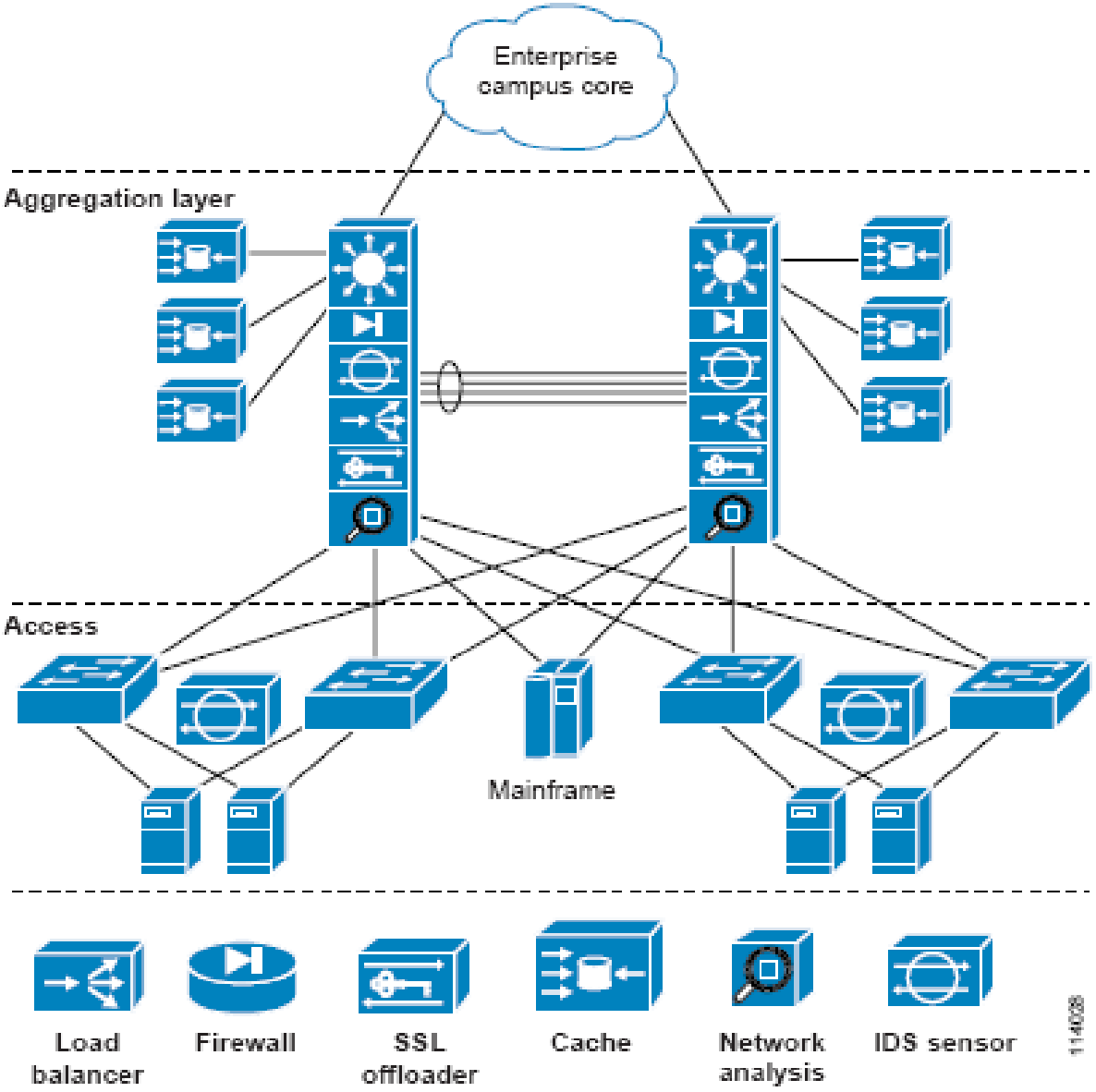
Data Center Architecture

- The data center infrastructure must be scalable and highly available.
- It should still be simple to
  - operate.
  - troubleshoot.
  - easily accommodate new demands.

Data Center Architecture

- The architecture of enterprise data centers is determined by
  - the business requirements
  - the application requirements
  - the traffic load
- The extent of the data center services offered translates into the actual design of the architecture.

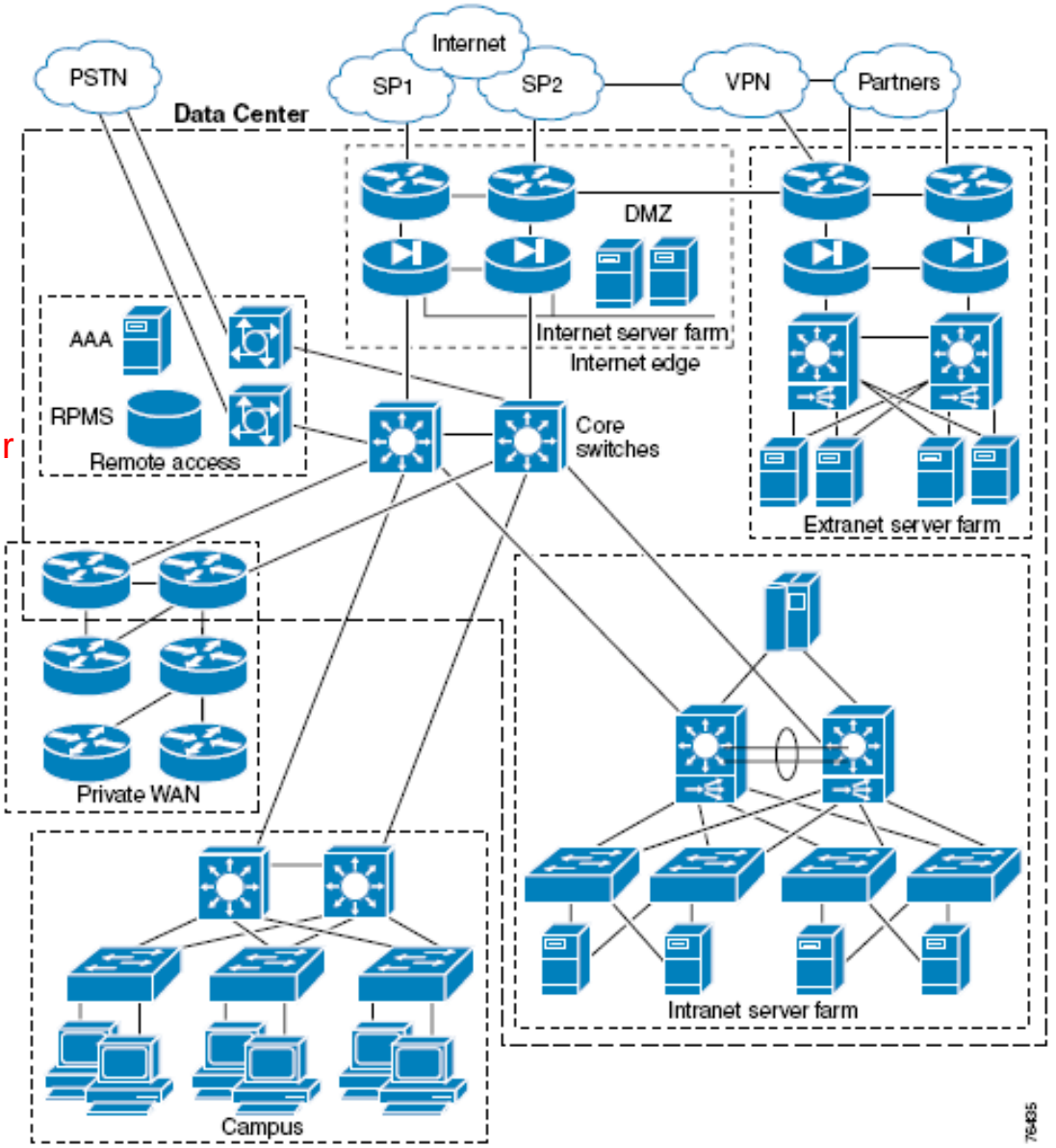
Data Center Architecture



Enterprise Network Infrastructure

- Typical Enterprise network include:
  - Campus
  - Private WAN
  - Remote Access
  - Internet server farm
  - Intranet server farm
  - Extranet server farm

Enterprise Network Infrastructure Example



Demilitarized Zone

Remote Power Manager

Enterprise Network Infrastructure

- Data centers house many network infrastructure components
  - the core switches of the campus network or the edge routers of the private WAN.
- Data Center designs include at least one type of server farm.
  - These server farms may or may not be built as separate physical entities, depending on the business requirements of the enterprise.

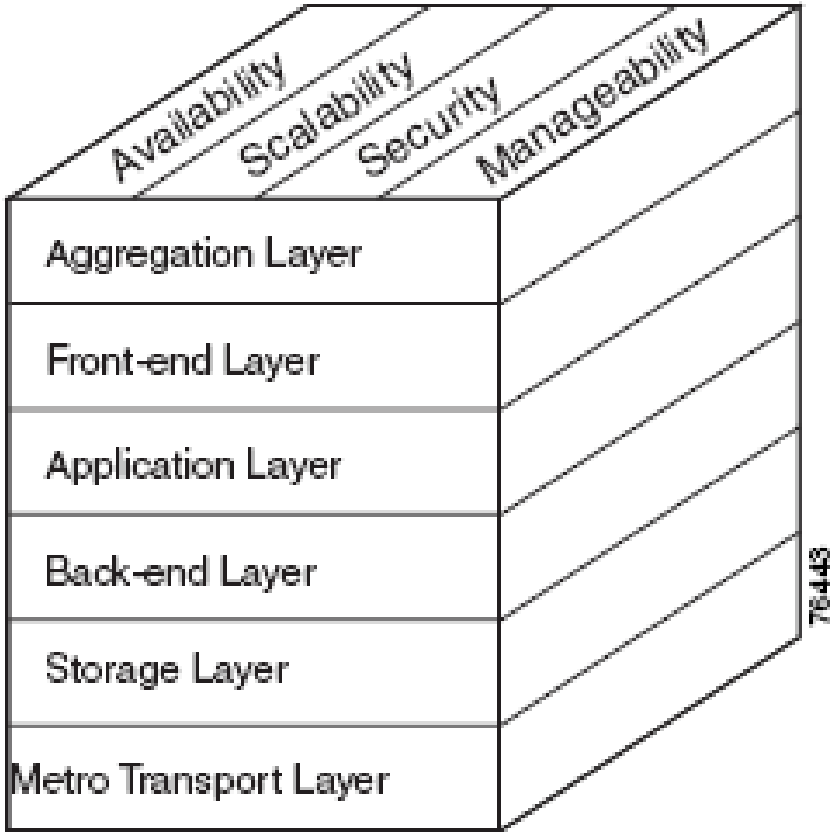
Enterprise Network Infrastructure

- A single data center may use a shared infrastructure, resources such as servers, firewalls, routers, switches, etc., for multiple server farm types.

- Another data center may require that the infrastructure for server farms be physically dedicated.
  - Enterprises make these choices according to business drivers and their own particular needs.

Data Center Architecture

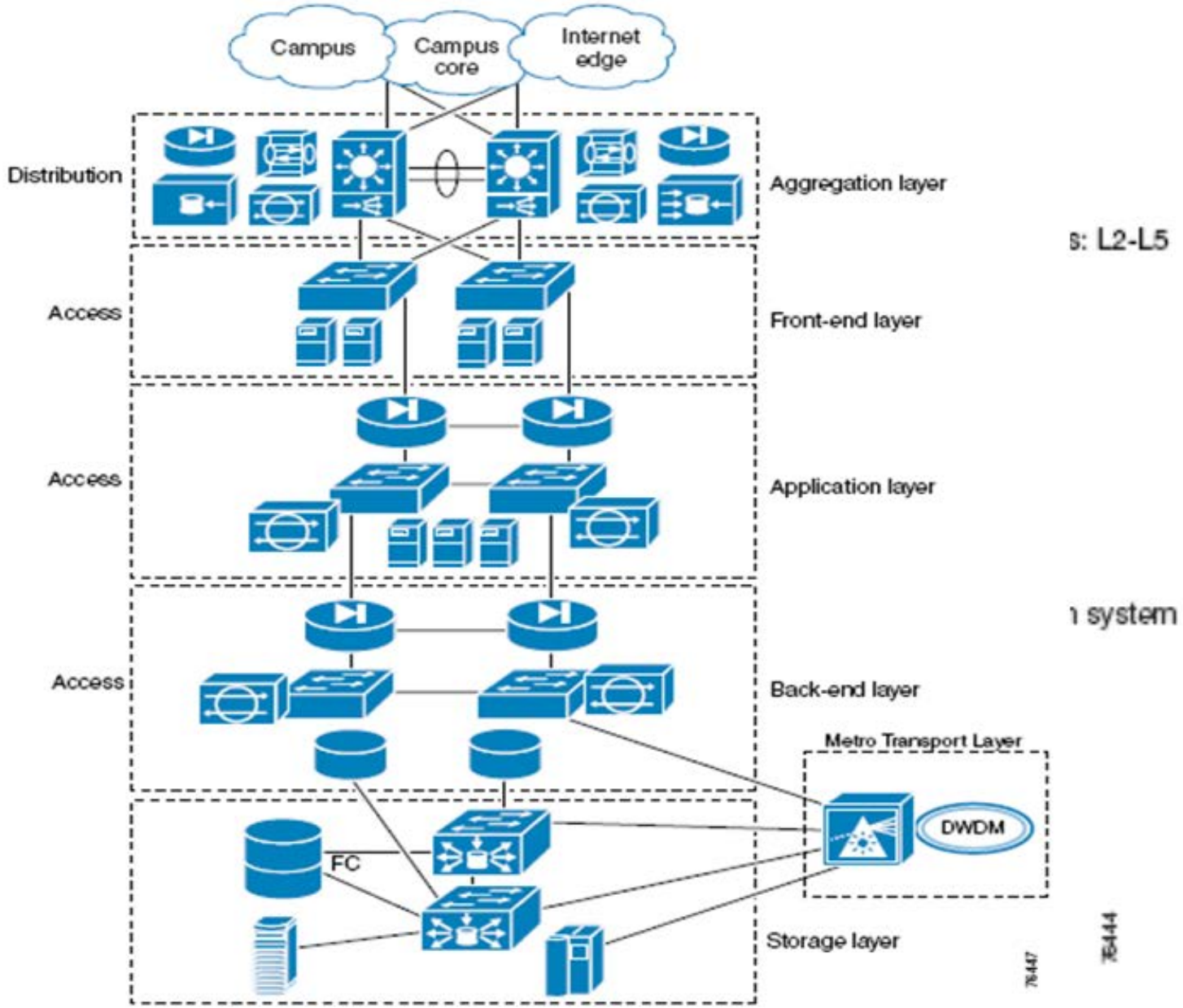
- **Four** key design criteria is used in this translation process that help you produce design goals.
- These criteria are:
  - availability
  - scalability
  - security
  - management



Data Center Architecture

- A layered approach to the data center design that supports N-Tier applications yet it includes other components related to other business trends.
- The layers of the architecture include:
  - Aggregation
  - Front-end
  - Application
  - Back-end
  - Storage
  - Metro Transport

Data Center Layer Architecture



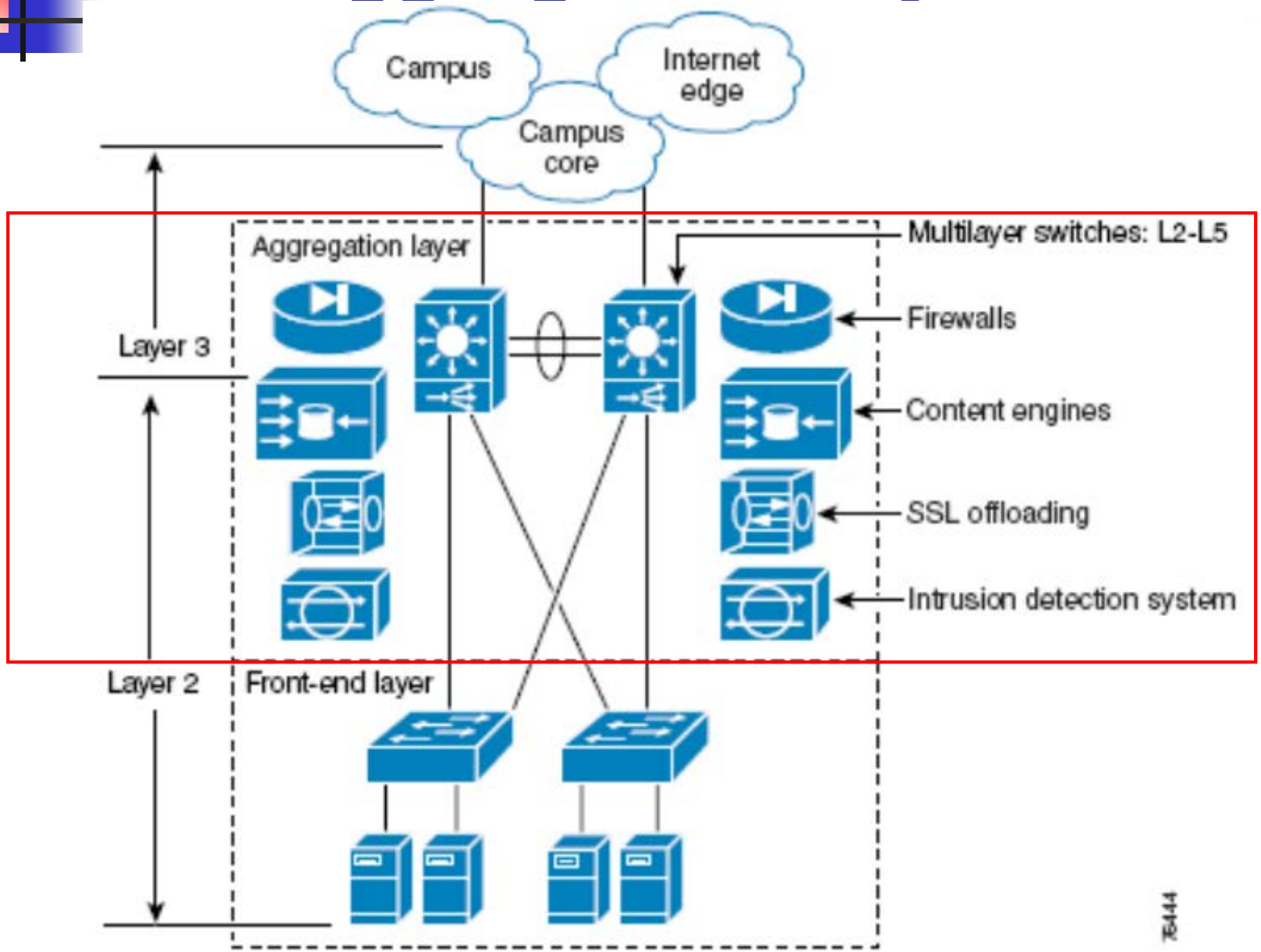
Aggregation Layer

- Provides network connectivity between the server farms and the rest of the enterprise network.
- Provides network connectivity for data center service devices.
- Supports fundamental layer 2 and layer 3 functions.

Aggregation Layer

- The aggregation layer is analogous to the campus network distribution layer.
- Data center services that are common to servers in the front-end or other layers should be centrally located in the aggregation layer for
  - predictability
  - consistency
  - manageability
- The aggregation layer includes:
  - Multilayer switches (aggregation switches) that provide the layer 2 and layer 3 functionality
  - Content switches
  - Firewalls
  - IDSs
  - Content engines
  - SSL offloaders

Aggregation Layer



Front-end Layer

- The front-end layer is analogous to the campus access layer in its functionality, and provides connectivity to the first tier of servers.

- The front-end server farms typically include
  - FTP
  - Telnet
  - TN3270
  - SMTP
  - Web servers
    - Social services
  - other business application servers
  - network-based application servers, such as
    - IPTV broadcast servers
    - Content distribution managers
    - Call managers

Front-end Layer Functionality

- **Multicast** and **QoS** that may be required, depend on the servers and their functions.
  - E.g., live video streaming over IP is supported, **multicast** must be enabled
  - E.g., voice over IP is supported, **QoS** must be enabled.
- Layer 2 connectivity through VLANs is required between
  - servers with backup servers supporting the same application on different layer 2 switches
  - servers and service devices to process content switches.
- Other requirements may be used
  - IDSs or host IDSs to detect intruders
  - PVLANs to segregate servers in the same subnet from each other.

PVLAN (Private-VLAN)

- Provide layer 2 isolation between ports within the same broadcast domain.
- There are three types of PVLAN ports:
  - **Promiscuous**— can communicate with all interfaces, including the isolated and community ports within a PVLAN.
  - **Isolated**— has complete layer 2 separation from the other ports within the same PVLAN, but not from the promiscuous ports.
    - PVLANs block all traffic to isolated ports except traffic from promiscuous ports.
    - Traffic from isolated port is forwarded only to promiscuous ports.
  - **Community**— communicate among themselves and with their promiscuous ports.
    - The interface is separated at layer 2 from all other interfaces in other communities or isolated ports within their PVLAN.

Application Layer

- It provides connectivity to the servers supporting the business logic, which are grouped under the application servers tag.
- Applications servers
  - run a portion of the software used by business applications
  - provide the communication logic between front-end and the back-end, which is typically referred to as the middleware or business logic
  - translate user requests to commands the back-end database systems understand.

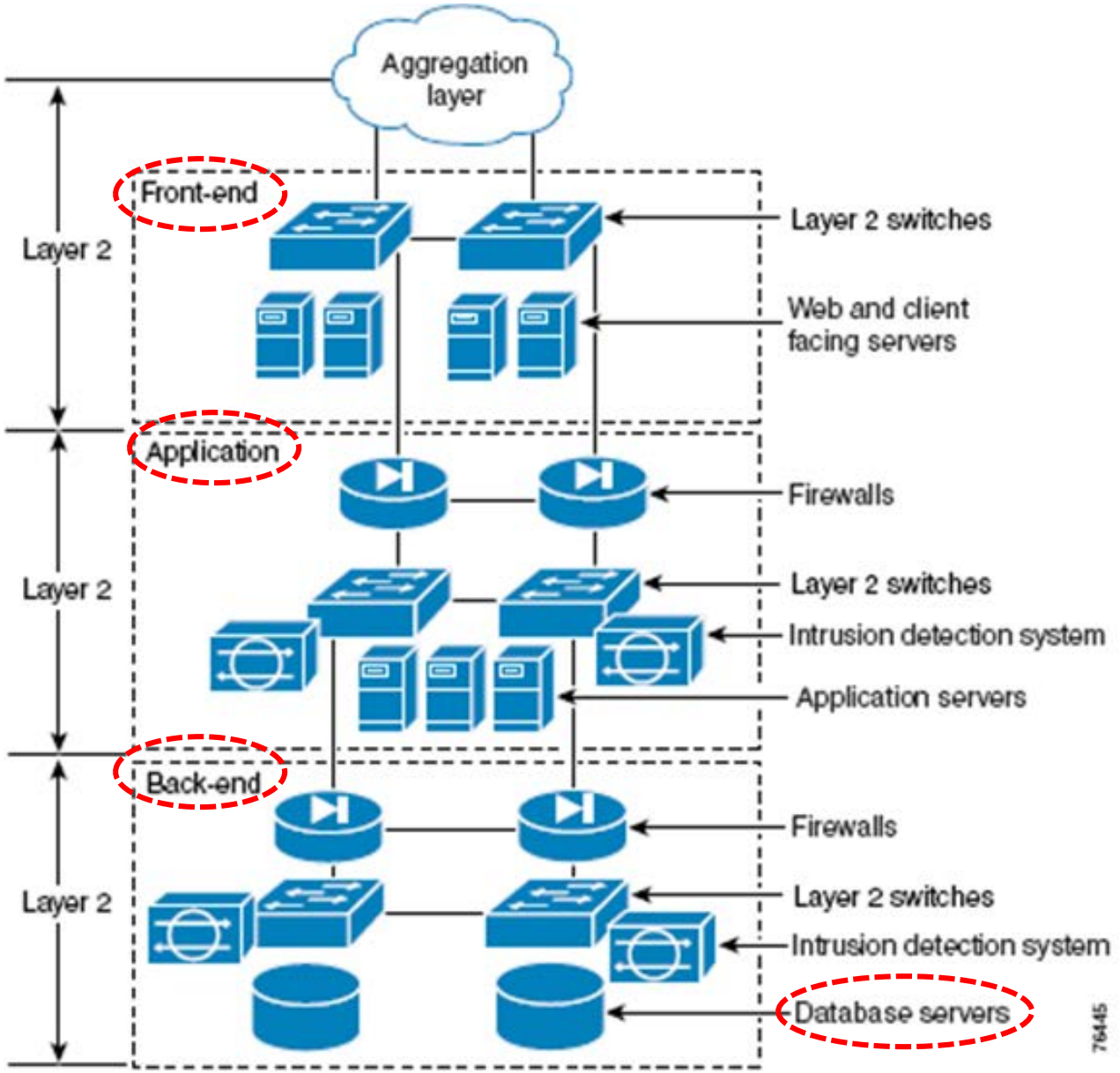
Application Layer

- The features required are almost identical to those needed in the front-end layer.
- Additional **security** is typically used to tighten security between servers that face users and the next layer of servers.
  - using firewall in between.
- Additional IDSs may also be deployed to monitor different kinds of traffic types.
- Additional services may require load balancing between the web and application servers typically based on layer 5 information (front-end), or SSL if the server-to-server communication is done over SSL.

Back-End Layer

- Provides connectivity to the database servers.
  - the relational database systems that provide the mechanisms to access the enterprise's information, which makes them highly critical.
- The hardware supporting the relational database systems range from medium sized servers to mainframes, some with locally attached disks and others with separate storage.
- The security considerations are more stringent and aimed at protecting the enterprise data.

Front-End, Application and Back-End Layers



Storage Layer

- Using **Fibre-Channel (FC)** or **iSCSI** connects devices in the storage network
- Through FC switches is used for storage-to-storage communications between devices.
  - such as attached server and disk subsystems of tape units.
- iSCSI
  - provides SCSI connectivity to servers over an IP network
  - is supported by iSCSI routers, port adaptors, and IP services modules.
- FC is typically used for **block** level access, whereas iSCSI is used for **file** level access.

Metro Transport Layer

- The metro transport layer is used to provide a high speed connection between distributed data centers.
  - high speed campus-to-campus connectivity.

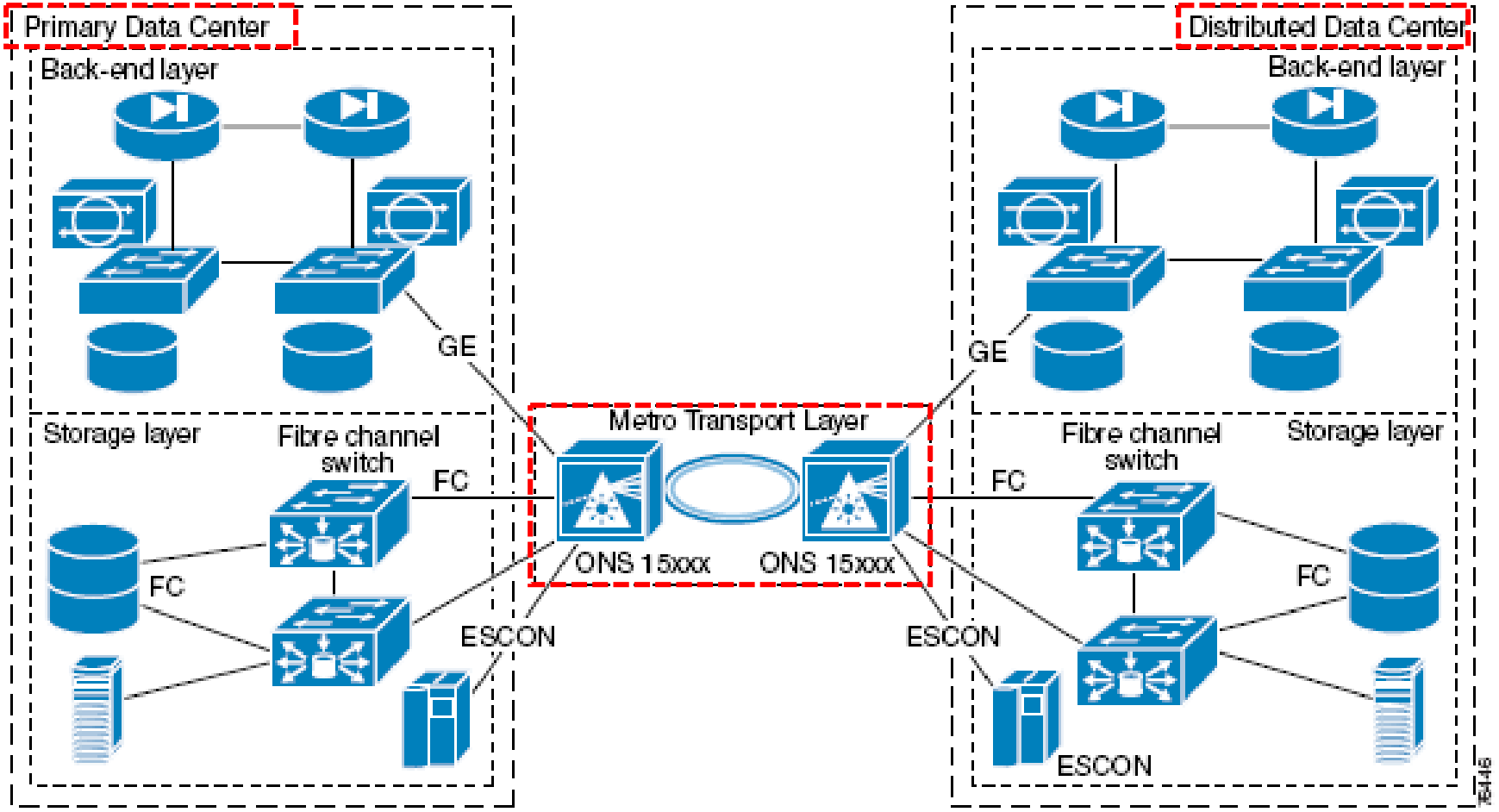
- Distributed data centers use metro optical technology to provide transparent transport media, which is typically used for database or storage mirroring and replication.

Metro Transport Layer

- The high speed connectivity needs are for synchronous communications or asynchronous communications, which one depends on the recovery time expected when the primary data location fails.

- The most common business drivers to use distributed data centers and their connectivity is for
  - disaster recovery plans
  - business continuance plans

Metro Transport Layer



ESCON:Enterprise Systems Connection

Data Center Services

- These services include:
  1. Infrastructure service: layer 2, layer 3, intelligent network services and data center transport
  2. Application optimization services: content switching, caching, SSL offloading, and content transformation
  3. Storage: consolidation of local disks, network attached storage, storage area networks
  4. Security: access control lists, firewalls, and intrusion detection systems
  5. Management: management devices applied to the elements of the architecture

Infrastructure Services

- All core features for the functions and services of data center infrastructure.
- The infrastructure features are organized as follows:
  - Metro
  - Layer 2
  - Layer 3
  - Intelligent Network Services

Metro Services

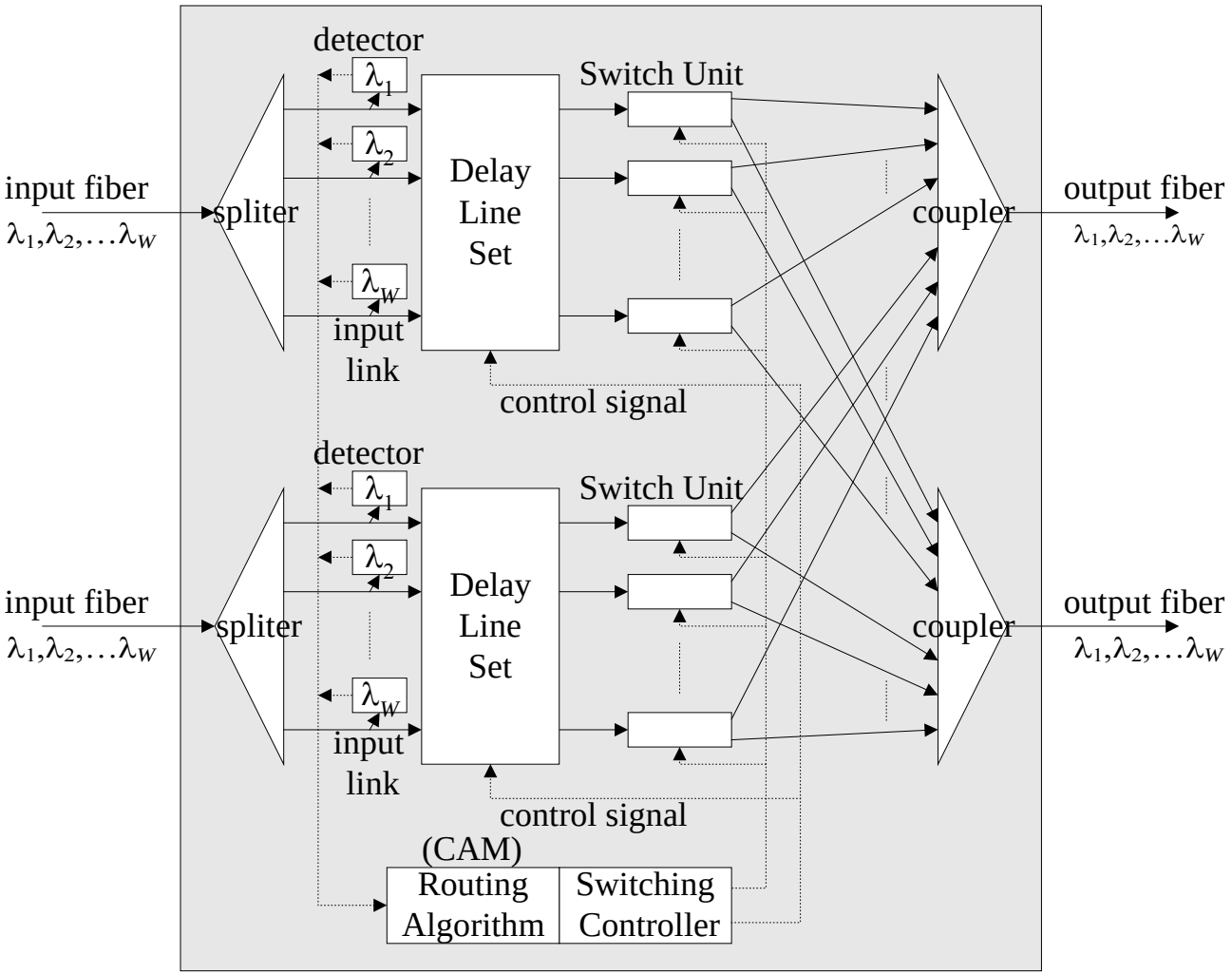
- Metro services include a number of physical media access, such as
  - Fibre-Channel
  - iSCSI
  - Metro transport technologies such as
    - Dense wave division multiplexing (DWDM)
    - Coarse wave division multiplexing (CWDM)
    - Synchronous Optical Networking (SONET)
    - 10GE.

Metro Services

- Metro transport technologies enable a number of applications that require high bandwidth and low predictable delay.
- DWDM provides physical connectivity for a number of different physical media concurrently such as
  - Gigabit Ethernet
  - Asynchronous Transfer Mode (ATM)
  - Fibre Channel
- Some instances where this connectivity is required are for
  - long-haul storage area networks (SAN) extension over SONET or IP
  - short-haul SAN extension over DWDM/CWDM, SONET, or IP (Ethernet)

The DWDM Switch Element Architecture

- The detail architecture of a 2×2 DWDM Switch with W wavelengths in each input fiber.



Layer 2 Services

- Support the layer 2 adjacency between the server farms and the service devices
- Layer 2 domain supports
  - a fast convergence
  - loop free
  - fault tolerance
  - scalable
- LAN media access
  - Gigabit Ethernet
  - ATM
  - Packet over SONET (PoS)
  - IP over optical media

Spanning Tree Protocol (STP)

- Layer 2 domain features ensure the spanning tree protocol (STP) convergence time for deterministic topologies is in the single digit seconds, and the failover and fallback scenarios are predictable.

- The list of features includes:
  - 802.1s + 802.1w (Multiple Spanning-Tree)
  - PVST+802.1w (Rapid Per VLAN Spanning-Tree)
  - 802.3ad (Link Aggregate Control Protocol)
  - 802.1q (trunking)
  - Loop guard
  - Uni-directional link detection (UDLD)
  - Broadcast suppression

Layer 3 Services

- Layer 3 services enable fast convergence and a resilient routed network, including redundancy, for basic layer 3 services, such as default gateway support.
  - The network operation is predictable under normal and failure conditions.
- The list of available features includes:
  - Static routing
  - Border gateway protocol (BGP)
  - Interior gateway protocols (IGPs): OSPF and EIGRP
  - HSRP, MHSRP & VRRP (fault-tolerant default gateway)

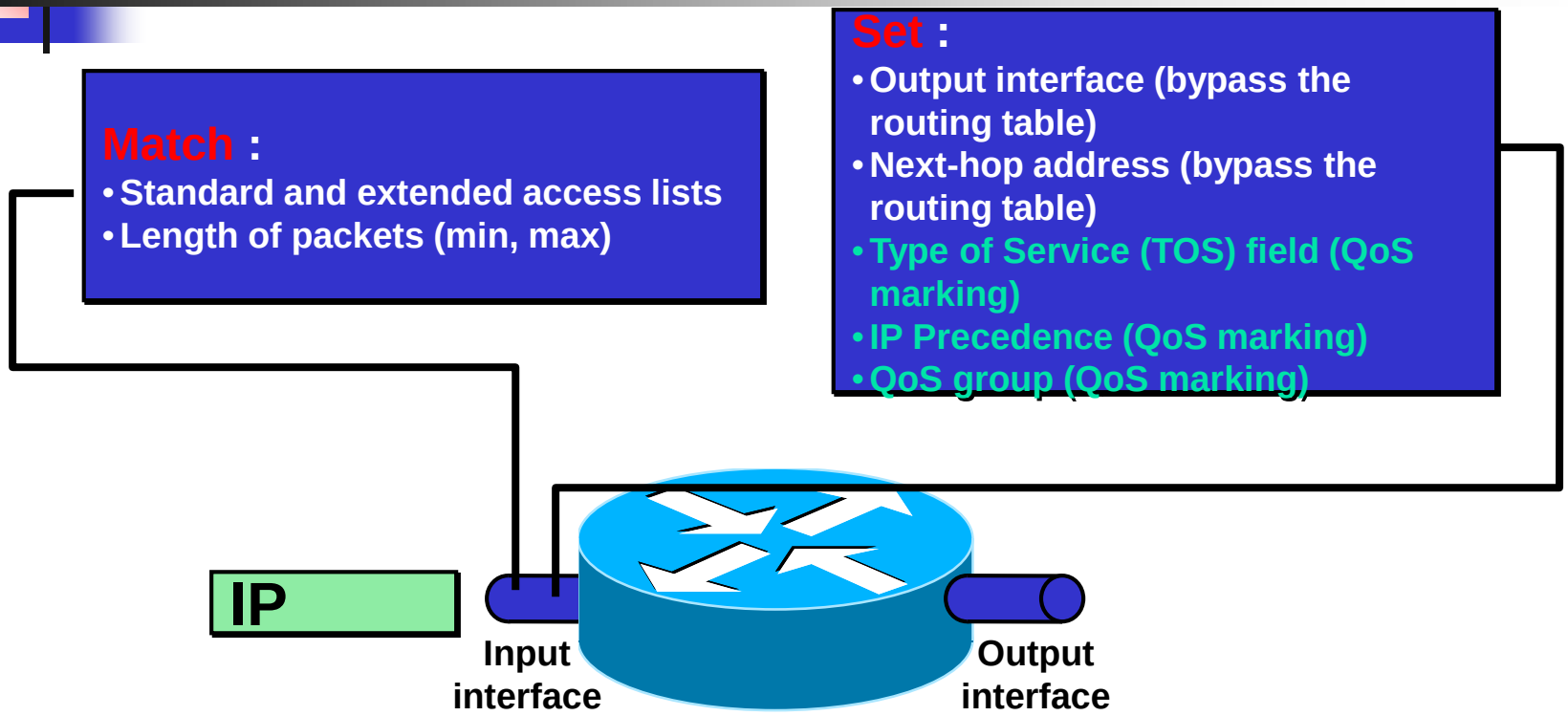
Inteligent Network Services

- It include a number of features that enable application services.
- The most common features are **QoS** and **Multicast**.
  - live or on demand video streaming and IP telephony.
  - the classic set of enterprise applications.
    - Social services
- Other important intelligent network services include
  - Private VLANs (PVLANs)
  - Policy based routing (PBR).

Policy-based Routing

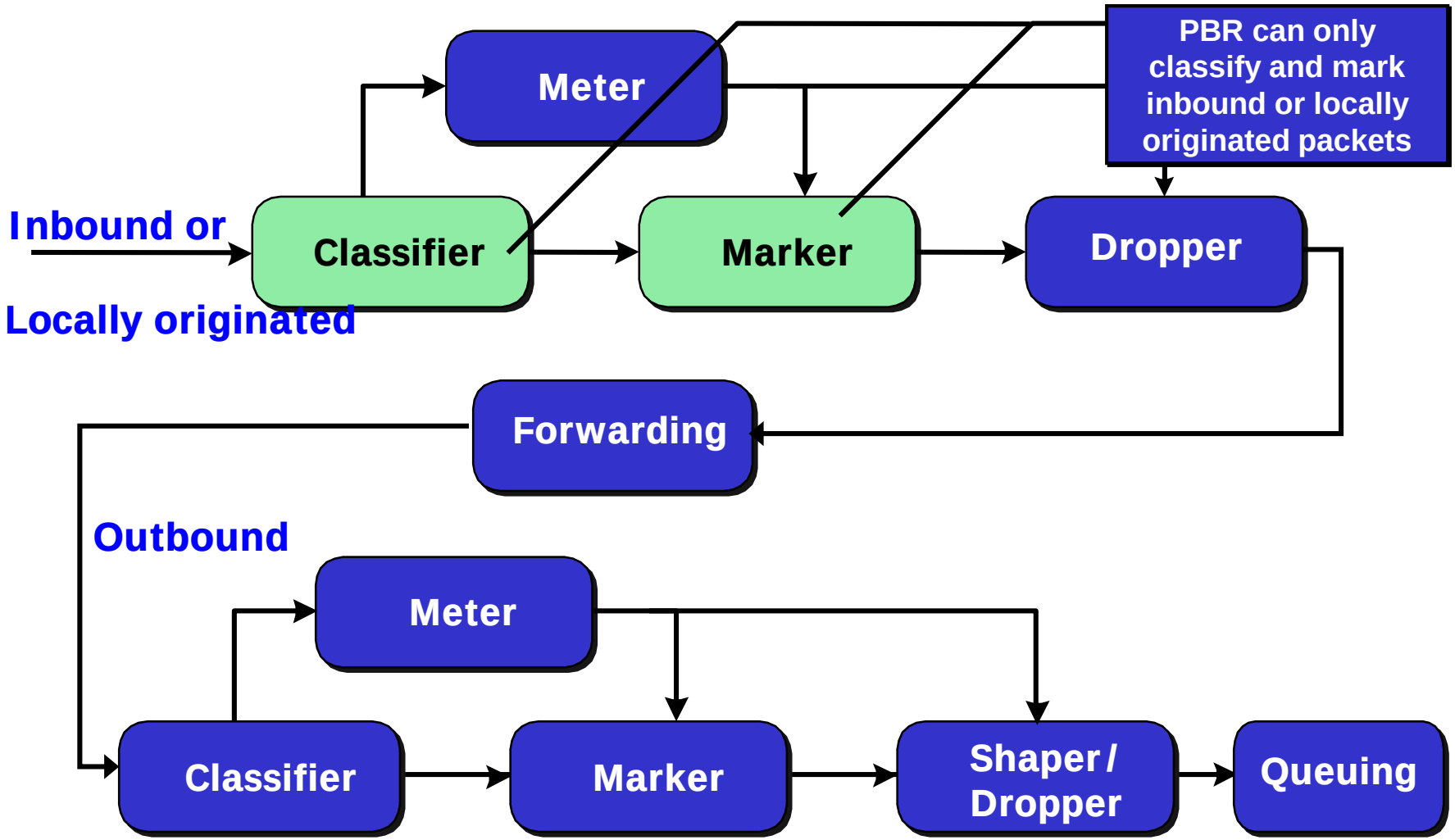
- Policy-based Routing (PBR) is a mechanism that can be used to bypass the default destination-based forwarding functionality of routers
- PBR is implemented using a route map
  - **match commands** are used to classify packets
  - **set commands** are used to process packets
- Route maps are applied to interfaces for processing of inbound packets (forwarding and/or marking)

PBR Match and Set Options



- PBR has two primary applications:
  - Implementation of more complex routing paradigms than a simple destination-based forwarding
  - Classification and marking of packets for QoS purposes

PBR Capabilities



Inteligent Network Services

- QoS is important for two reasons:
  - application traffic and port based rate limiting capabilities that enforces a proper QoS service class as traffic leaves the server farms
- Multicast enables the capabilities needed to reach multiple users concurrently or servers to receive information concurrently (cluster protocols).

Application Optimization Services

- It include a number of features that provide intelligence to the server farms.
- These features permit the scaling of applications supported by the server farms and packet inspection beyond layer 3 (layer 4 or layer 5).
- The application services are:
  - server load balancing or content switching
  - caching
  - SSL offloading
  - web services

Application Optimization Services

- Content switching scales application services by front ending servers and load balancing of the incoming requests to those available servers.

- The load balancing mechanisms could be based on layer 4 or layer 5 information, thus allowing the partitioning of the server farms by the content.
  - A group of servers supporting video streaming could be partitioned on those that support MPEG versus the ones that support Quicktime or Windows Media.

  - The content switch is able to determine the type of request, by inspecting the URL, and forwards it to the proper server.
    - This process simplifies the management of the video servers and allows you to deal with scalability at a more granular level, per type of video server.

Application Optimization Services

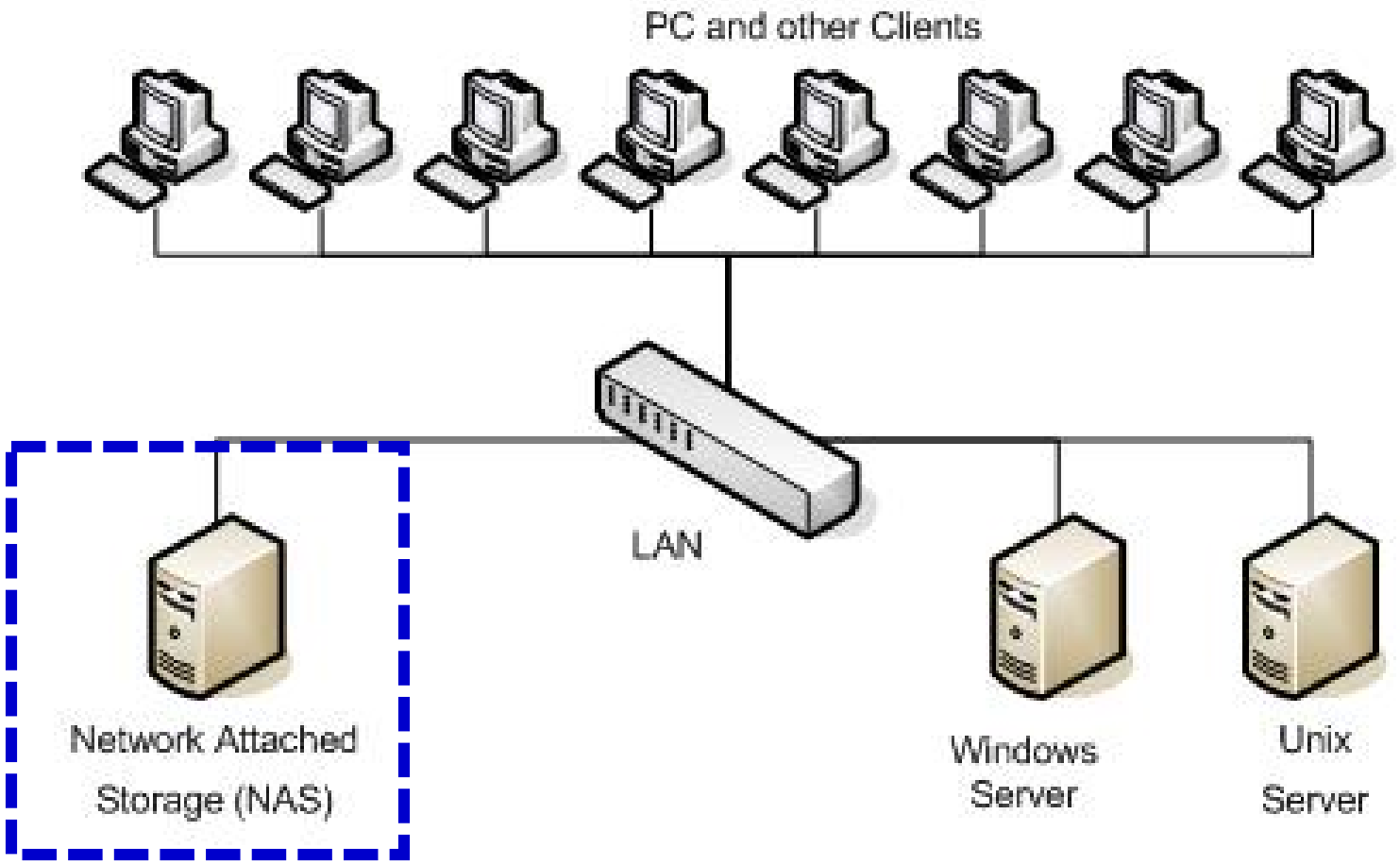
- The process of offloading occurs transparently for both the user and the server farm.
- SSL offloading also offloads CPU capacity from the server farm by processing all the SSL traffic.
- The two key advantages:
  - The centralized management of SSL services on a single device.
  - The capability of content switches to balance load.

Storage Services

- Storage services include the storage network connectivity required for user-to-server and storage-to-storage transactions.
- The major features could be classified in the following categories:
  - Network attached storage (NAS)
  - Storage area networks (SAN) to IP: Fibre Channel and SCSI over IP
  - Localized SAN fabric connectivity (Fibre Channel or iSCSI)
  - Fibre Channel to iSCSI Fan-out

Network-Attached Storage(NAS)

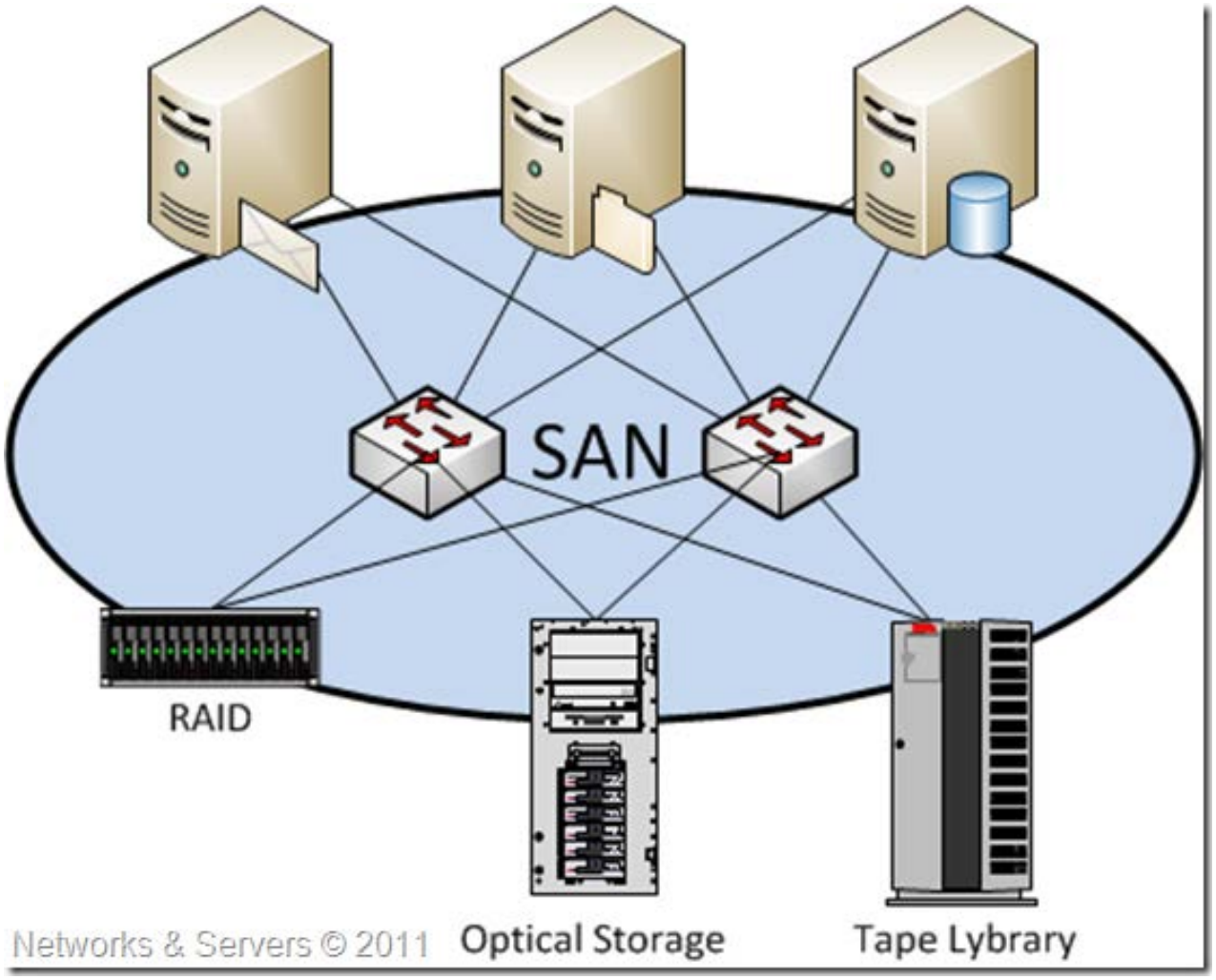
Typical Network Architecture Incorporating NAS Data Storage



NAS

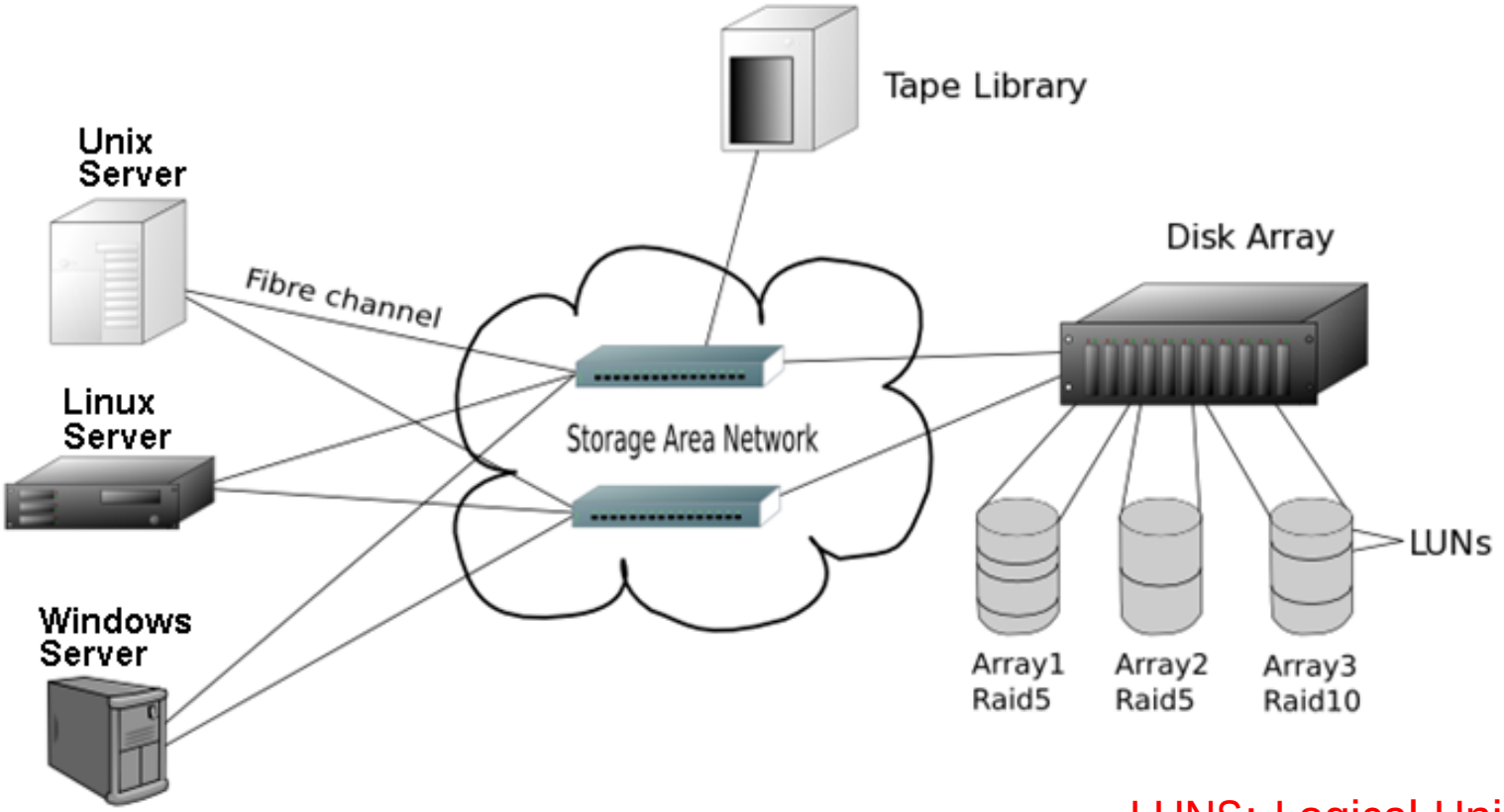
- Scalability: good
- Availability: as long as the LAN and NAS device work, generally good
- Performance: bandwidth limited by speed of LAN, traffic conflicts, inefficient protocol
- Management: easy
- Connection: homogeneous vs. heterogeneous

Storage Area Network (SAN)



Storage Area Network (SAN)

- SAN is created by using the Fibre Channel to link peripheral devices such as disk storage and tape libraries



LUNS: Logical Unit Number

Storage Services

- NAS relies on the IP infrastructure and, in particular, features such as QoS to ensure the proper file over the IP network to the NAS servers.
- SAN:
  - commonly found in data centers,
  - uses FC to connect servers to the storage device
  - transmits SCSI commands between them.
- The SAN environments need to be accessible to the NAS and the larger IP Network.

SAN compare with NAS

- Dedicated Fibre Channel network for storage
- More efficient protocol
- Higher availability
- Reduce traffic conflict
- Longer distance (up to 10 km)

Storage Services

- FC over IP (FCIP) and SCSI over IP (iSCSI) are the emerging IETF standards
  - SCSI access and connectivity over IP.
  - The transport of SCSI commands over IP enables storage-to-IP and storage-to-storage over an IP infrastructure.
- SAN remains prevalent in data center environment
- The localized SAN fabric becomes important to permit storage-to-storage block access communication at FC speeds.
- There are other features focused on enabling FC to iSCSI fan-out for both storage-to-IP and storage-to-storage interconnects.

Security Services

- Server farms suffer from external threats but also internal attacks.

- It needs to have
  - a tight security perimeter around the server farms
  - a plan to keep the security policies applied in a manner consistent with the risk and impact if the enterprise data was compromised.

- Since different portions of the enterprise's data is kept at different tiers in the architecture, it is important to consider deploying security between tiers.
  - the specific tier has its own protection mechanisms according to likely risks.

Security Services

- Utilizing a layered security architecture provides a scalable modular approach to deploying security for the multiple data center tiers.
  - The layered architecture uses the various security services and features to enhance security.

Security Services

- The goal of security services is to mitigate against threats, such as:
  - Unauthorized access
  - Denial service
  - Network reconnaissance
  - Viruses and worms
  - IP spoofing
  - Layer 2 attacks

Security Services

- The security services offered in the data center include:
  - Access control lists (ACLs)
  - Firewalls
  - Intrusion detection systems (IDS, Host IDS)
  - Authentication mechanism
  - Authorization mechanism
  - Accounting mechanisms
  - A number of other services that increase security in the data center.

ACLs

- ACLs can be applied at various points in the data center infrastructure
- ACLs prevent:
  - unwanted access to infrastructure devices
  - protect server farm services
- ACLs come in different types:
  - Router ACLs (RACLs)
  - VLAN ACLs (VACLs)
  - QoS ACLs.
- An important feature of ACLs is the ability to perform packet inspection and classification without causing performance bottlenecks.
- This lookup process is possible when done in **hardware**, in which case the ACLs can operate at the speed of the media, or at wire speed.

Firewalls

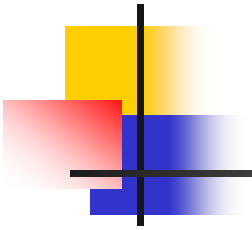
- The placement of firewalls marks a clear delineation between highly secured and loosely secured network perimeters.
- The typical location for firewalls remains the Internet edge and the edge of the data center
- They are also used in multi-tier server farm environments to increase security between the different tiers.

Intrusion Detection Systems(IDS)

- IDSs proactively address security issues intruder detection and the subsequent notification is a fundamental step to highly secure data centers.

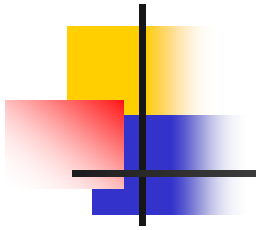
- Host IDSs enable real-time analysis and reaction to hacking attempts on applications or web servers.

- The host IDS is able to identify the attack and prevent access to server resources before any unauthorized transactions occur.



AAA

- AAA provides one more layer of security by
  - preventing user access unless authorized
  - ensuring controlled user access to the network and network devices by a predefined profile.
- The transactions of all authorized and authenticated users are logged for accounting purposes, for billing, or for postmortem analysis.



AAA

- Additional security considerations may include the use of the following features or templates:
  - One time passwords (OTPs)
  - SSH or IPSEC from user-to-device
  - Cisco discovery protocol (CDP) to discover neighboring Cisco devices
  - Securing virtual terminal (VTY) security
  - Default security templates for data center devices, such as
    - Routers
    - Switches
    - Firewalls
    - Content switches

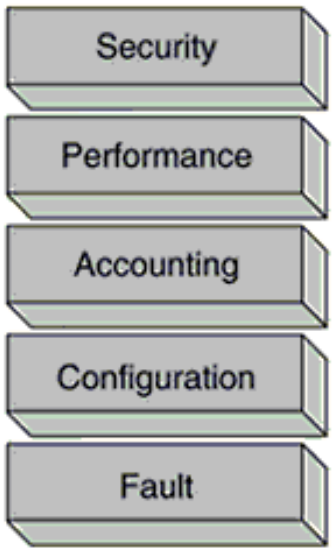
Management Services

- It includes service provisioning, which depending on the specific service, management considerations.

- Each service is also likely supported by different organizational entities or even by distinct functional groups whose expertise is in the provisioning, monitoring, and troubleshooting of such service.

Management Services

- Managing data center services should follow a consistent and comprehensive approach.
- The **FCAPS OSI** management standard and uses its management categories to provide management functionality.
  - FCAPS is a model commonly used in defining network management functions.
- The management features focus on the following categories:
  - Fault management
  - Configuration management
  - Accounting management
  - Performance management
  - Security management



FCAPS Model