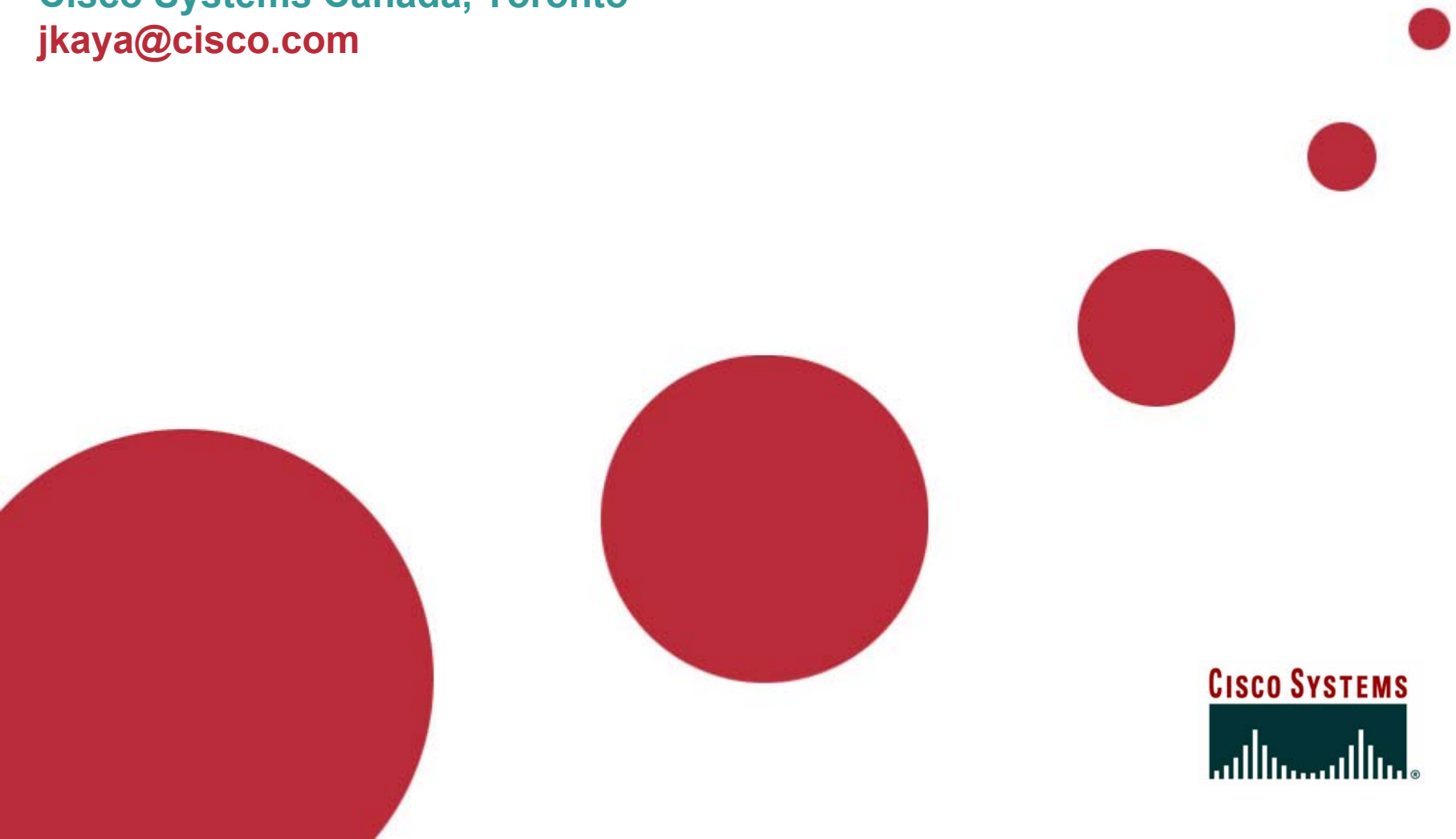
Technical Symposium

Next-Generation Intelligent Storage Area Networks

Joshua Kaya, Consulting Systems Engineer
Cisco Systems Canada, Toronto
jkaya@cisco.com



Cisco Systems

Agenda:

- Review: Storage Architectures
- Scalable Networks
- High Availability
- Security
- Management
- Inter-VSAN Routing
- FCIP & iSCSI
- Cisco iSCSI Network Boot

Storage Architectures

Cut-through vs. Store and Forward

- **“Unlike a cut-through switch, which starts to transmit a frame before it has completely received it, a store-and-forward switch waits until it has received a whole packet into its buffer before forwarding it.”**

BYTE.com

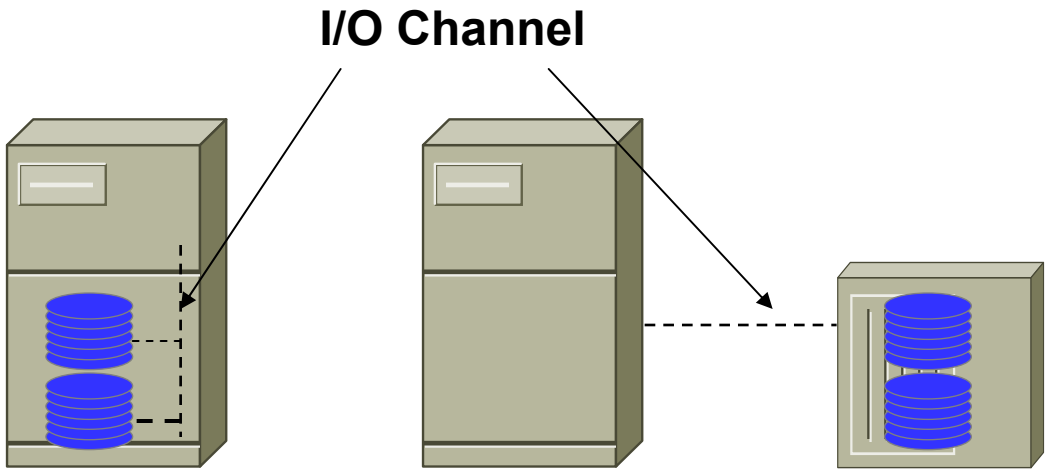
July 1995 / BYTE Lab Product Report / Store-And-Forwards

What is DAS?

- DAS = Direct Attached Storage
- DAS solutions provide:
  - Low-cost, slow to medium speed storage for use in home computers and small businesses
  - Can provide higher-cost, high-performance storage, which can be used in solutions needing high-speed access

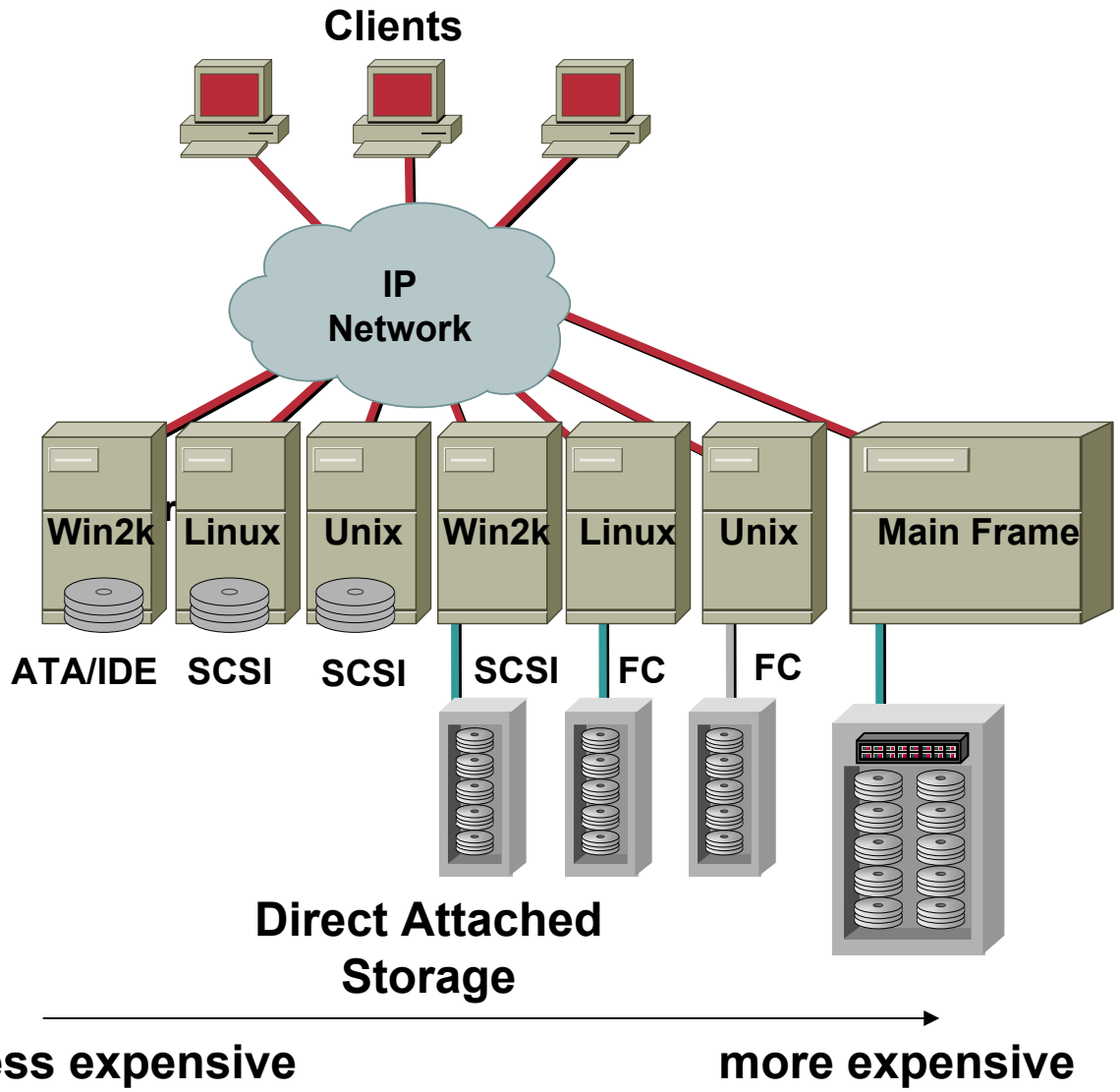
DAS Architecture

- DAS uses an I/O Channel architecture, which resides between a computer (initiator) and the device (target) used to store its data.
- Storage device is only accessible by attached host computer.
- Block level access to data.



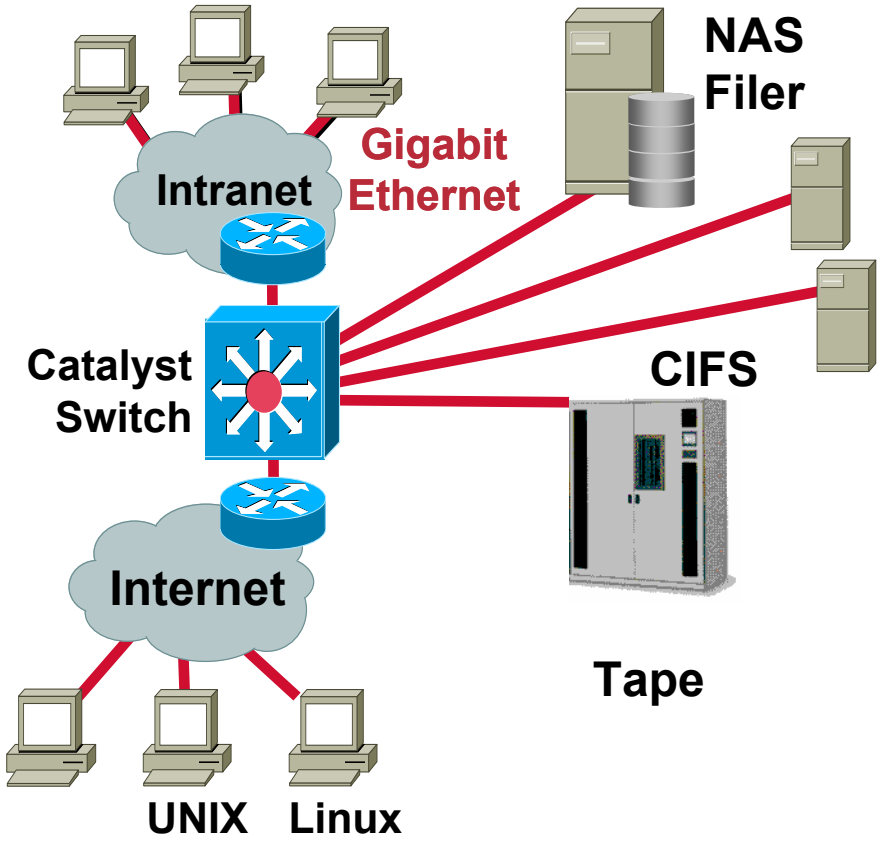
DAS Options

- The "typical" storage environment

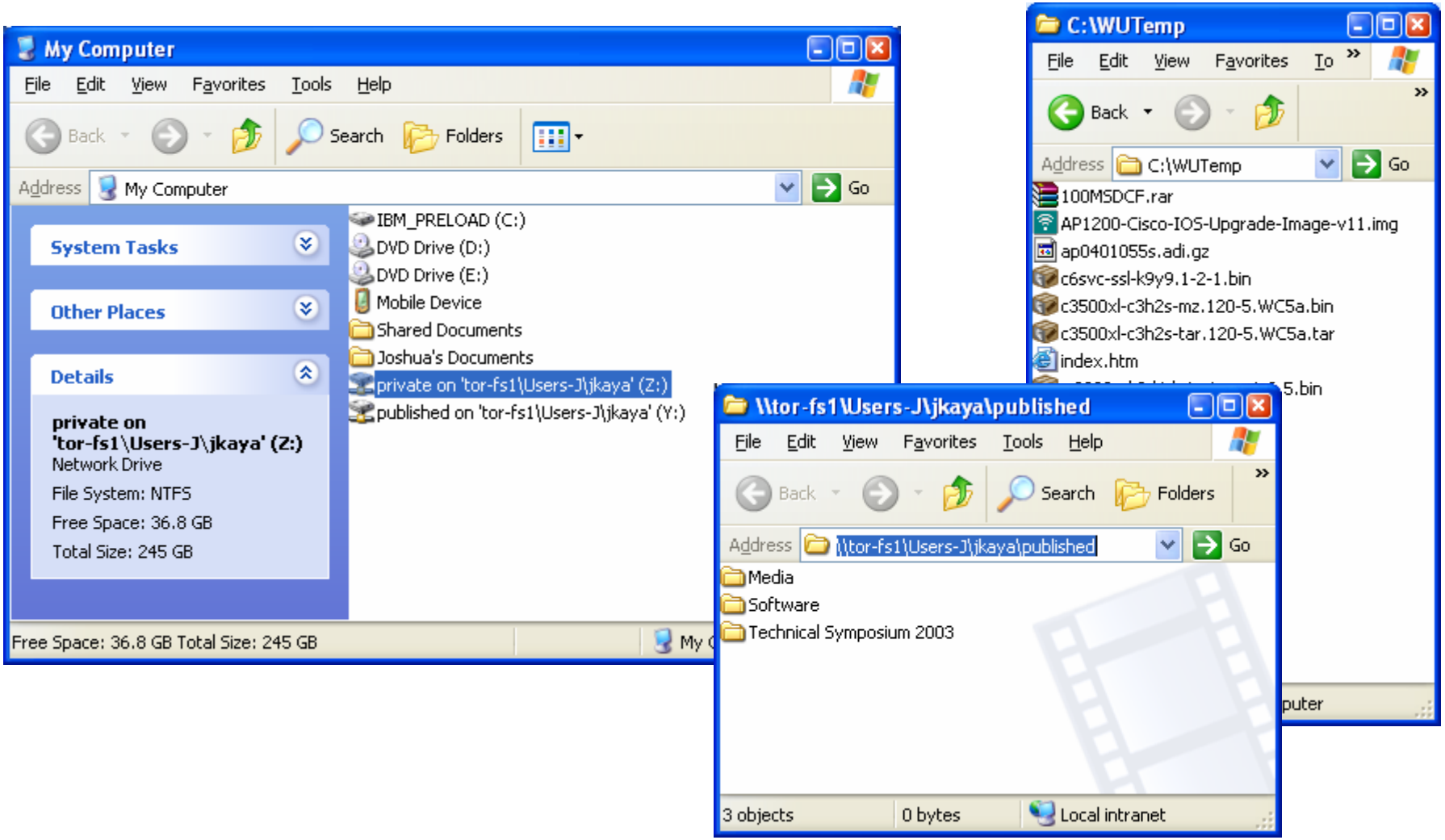


What is Network Attached Storage?

- NAS = Network Attached Storage
- NAS devices are network attached "appliances"
- NAS is the attachment of storage devices to the Local Area Network (LAN)



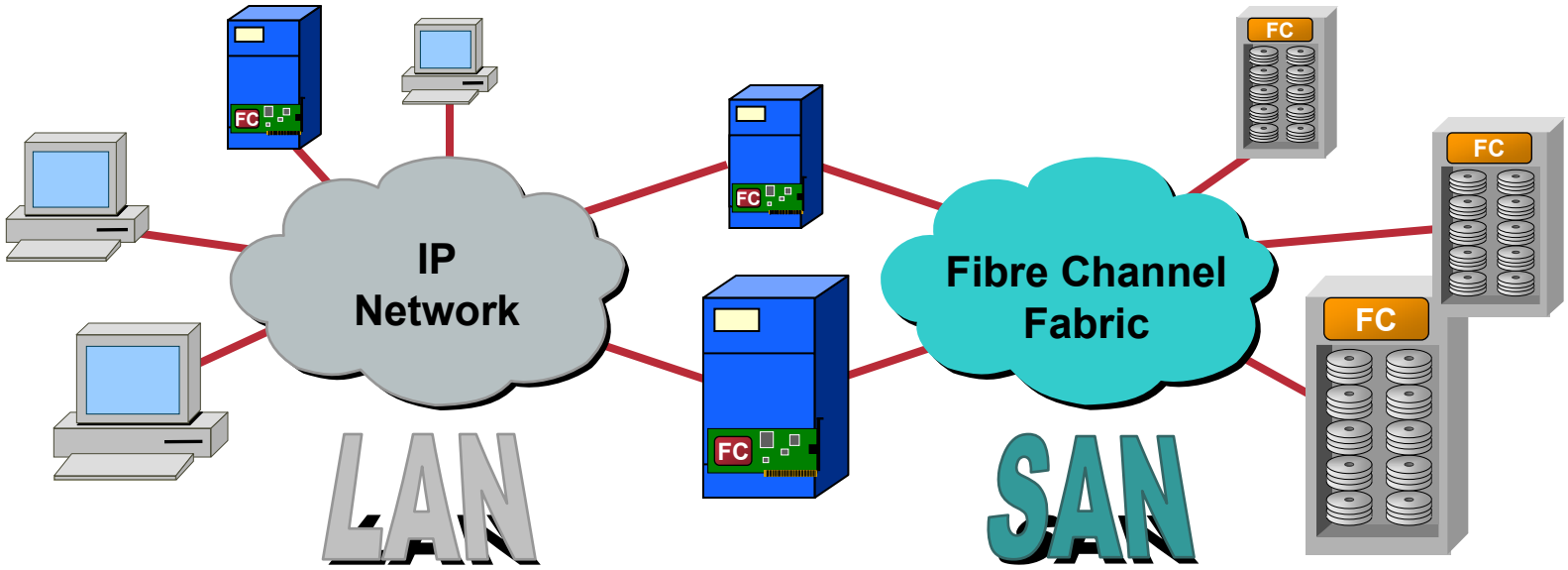
Example: DAS vs. NAS



What is Fibre Channel?

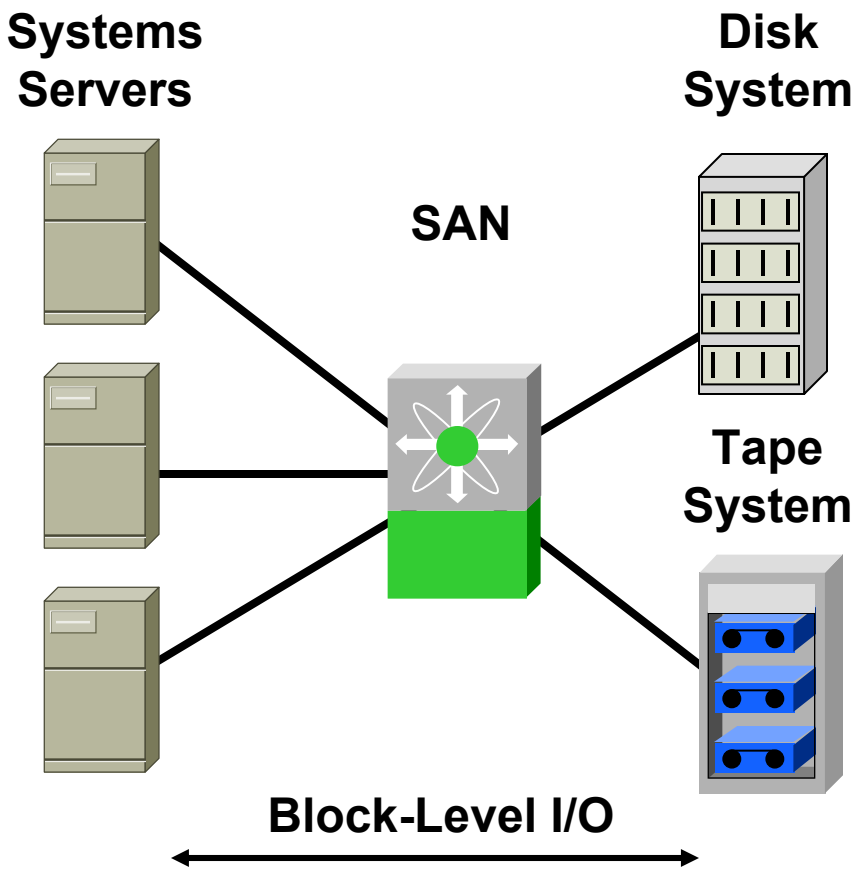
Fibre Channel is a protocol for transporting data between devices.

Fibre Channel is the transport technology most commonly used for SANs today.



SAN Components

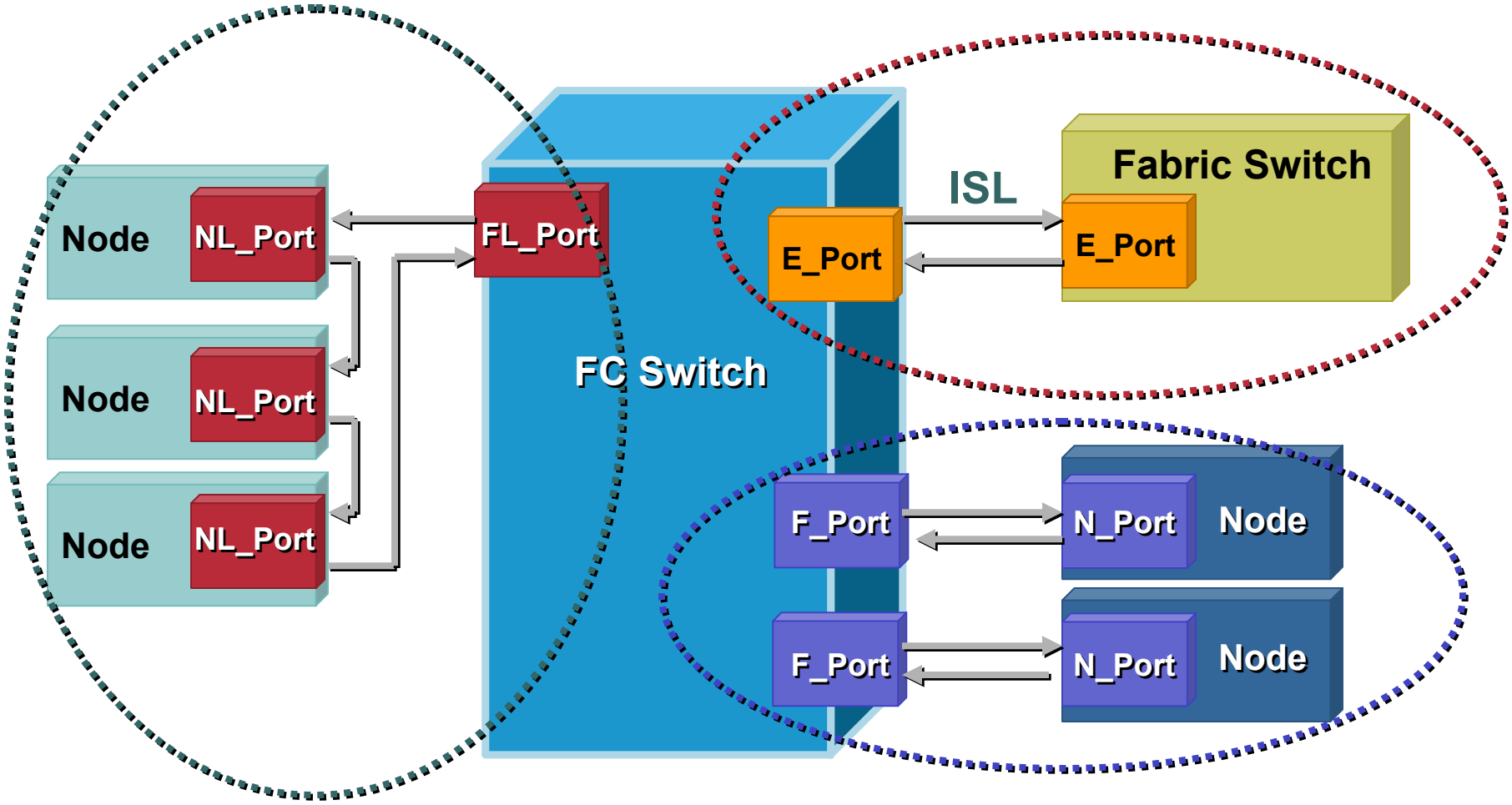
- **A SAN consists of:**
  - **Host systems with host bus adapters and drivers**
  - **An interconnection network with switches and hubs**
  - **Disk and tape storage subsystems**
  - **Block-level I/O protocols used to access the storage devices**



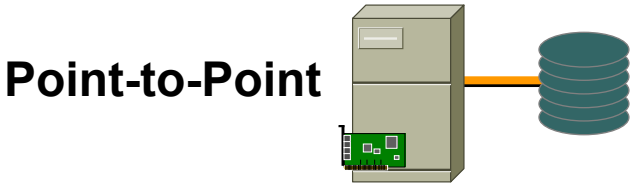
DHA1 change the diagrams

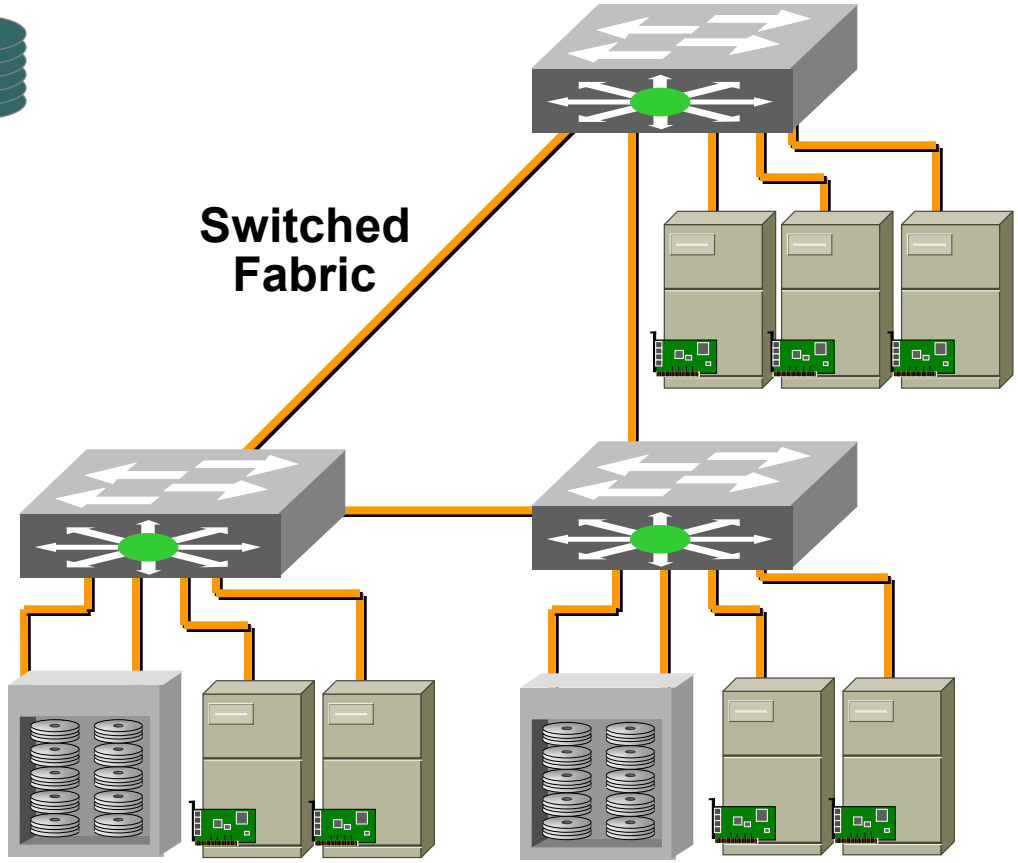
Duncan Anderson, 10/23/2002

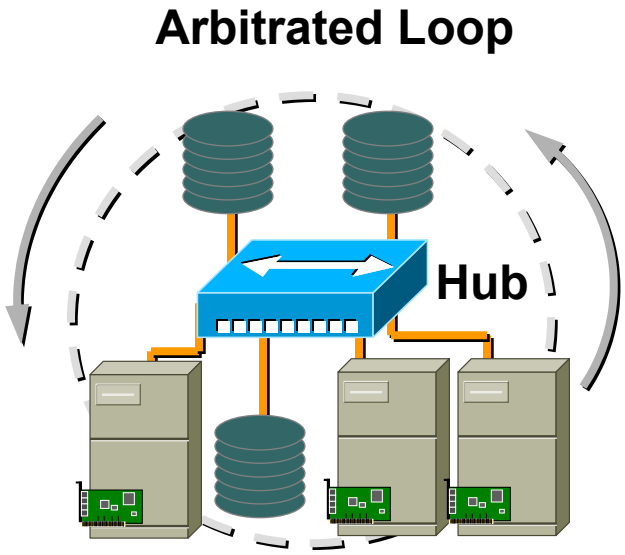
FC Port Types



Fibre Channel Topologies

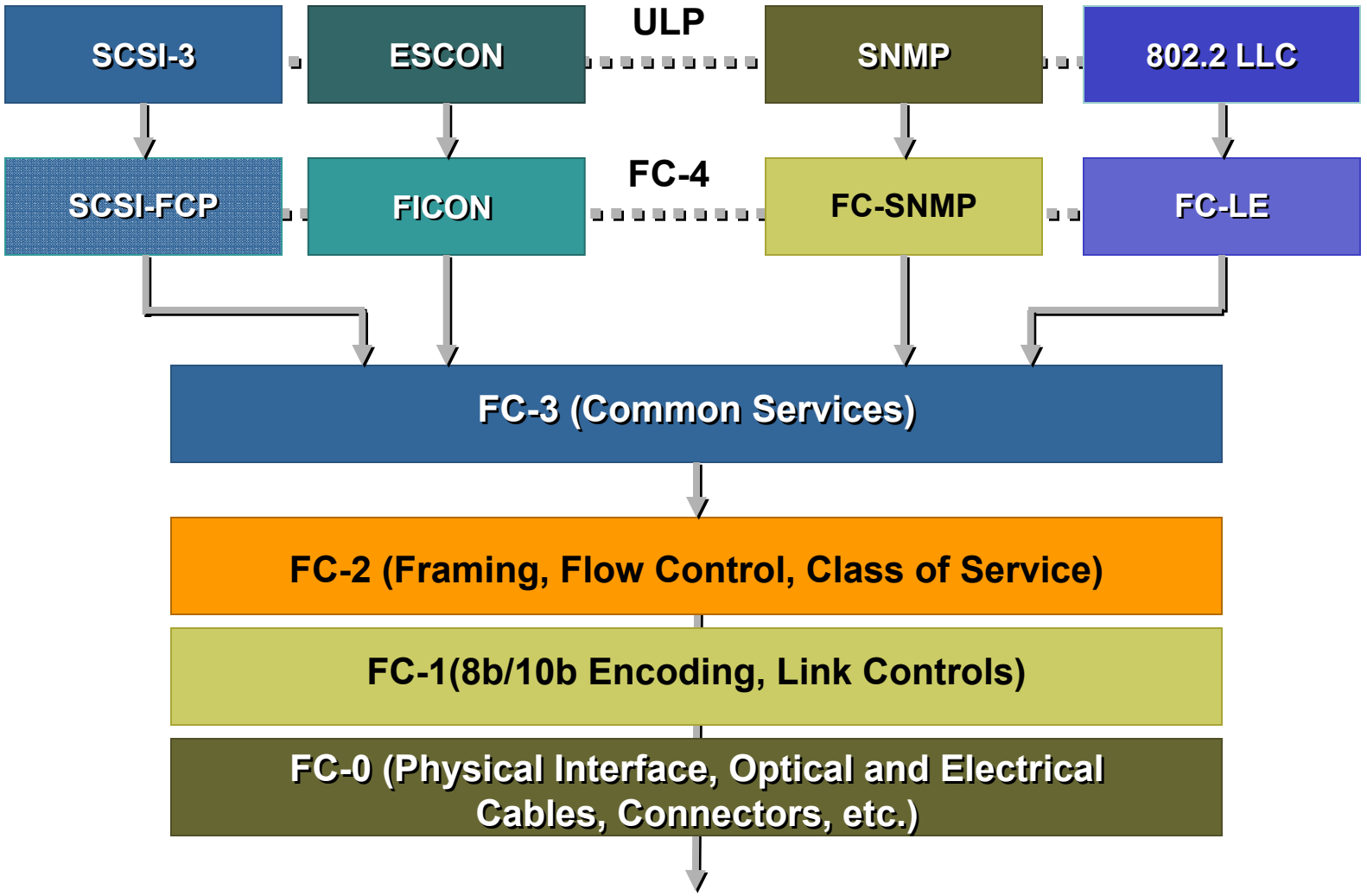




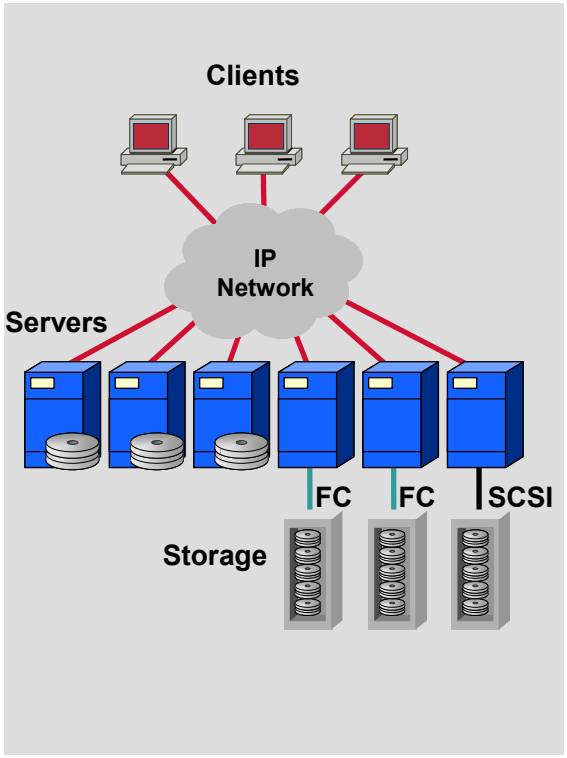


Cisco Storage Vision

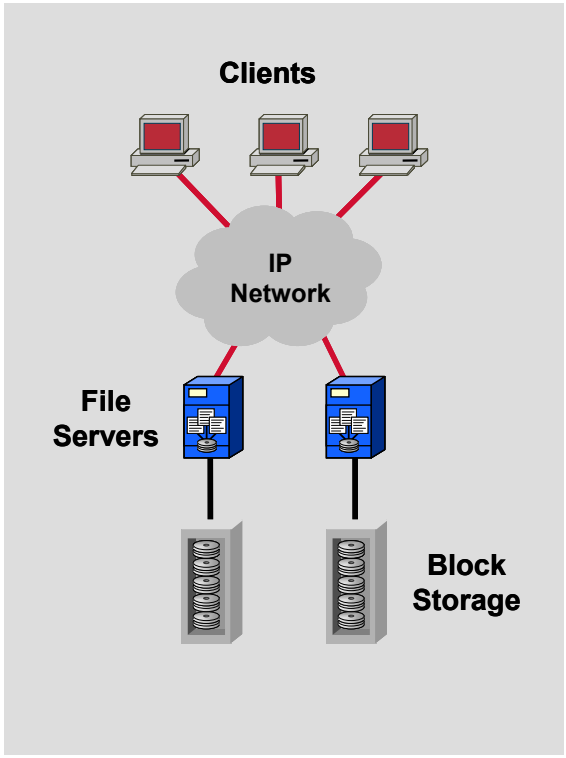
Fibre Channel Architecture Model



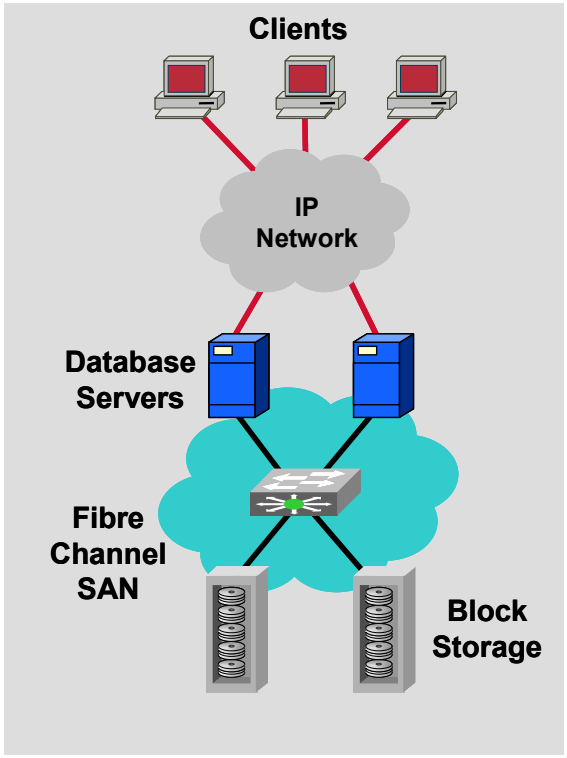
Storage Architectures



Direct-Attached Storage (DAS)



Network-Attached Storage (NAS)



Storage Area Network (SAN)

Scalable Network

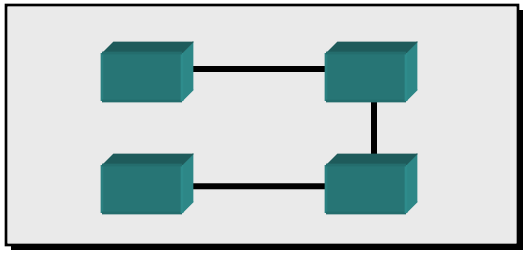
Comparing Fabric Designs

- **Cascade:**

  Very limited inter-switch bandwidth

  No resilience to failures

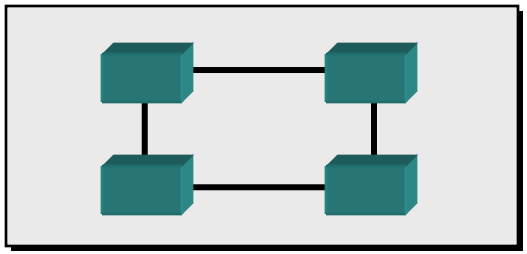
  Suitable for 2- or 3- switch fabrics where performance and availability are less of a concern than cost



- **Cascade ring:**

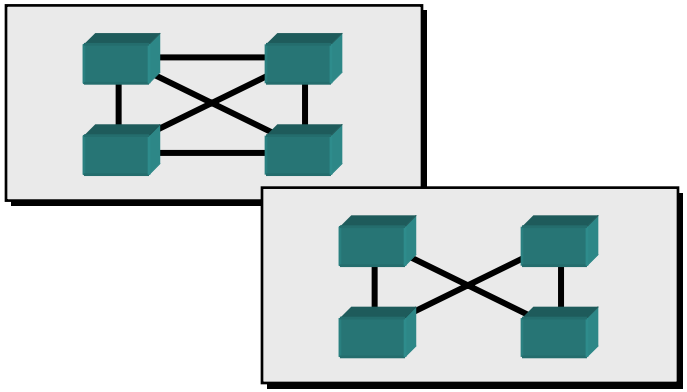
  Better performance and availability than cascade design

  Suitable for 3- to 5-switch fabrics with limited scalability, performance, and availability requirements

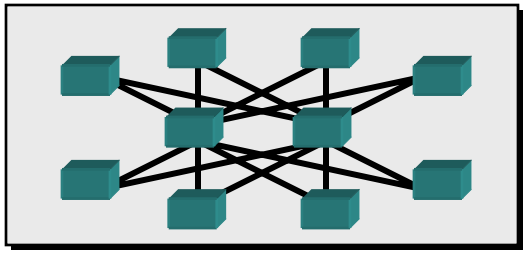


Comparing Fabric Designs

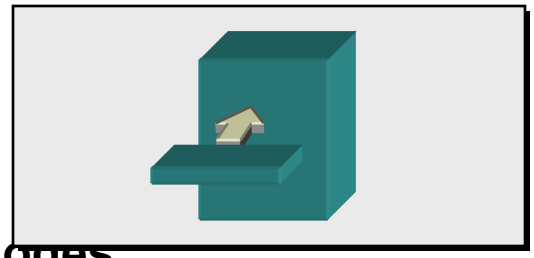
- **Mesh:**
  - Highly available
  - Performance varies according to full/partial configuration
  - Suitable for 4- to 8-switch fabrics with limited scalability requirements



- **Core-edge:**
  - Highly scalable solution for large fabrics
  - Strong performance
  - Relatively easy to manage



Collapsed Fabric Design

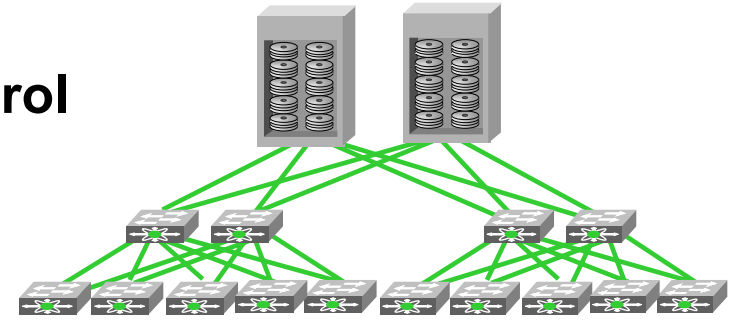


- **Collapsed Architecture:**
  - **Lack of ISLs means:**
    - **All purchased ports are available for nodes**
    - **Increased reliability**
    - **Simplified management**
  - **Scales easily (hot-swap blade architecture)**
  - **Fixed latency between ports = highest performance**
  - **Single management interface**
  - **Cost-effective for large SANs when ISL ports and management costs are added up**
  - **Not all “director-class switches” are the same**

Scalability—Current vs. Multilayer

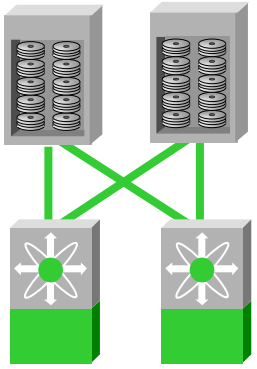
- 62% efficient port utilization
  - List/port = $2600 x 832 ÷ 512 = $4225
- Over-subscribed but no congestion control
- Limited port aggregation
- Only one set of fabric services
- No isolation of traffic

512 Port SAN Built w/ 64-Port Directors

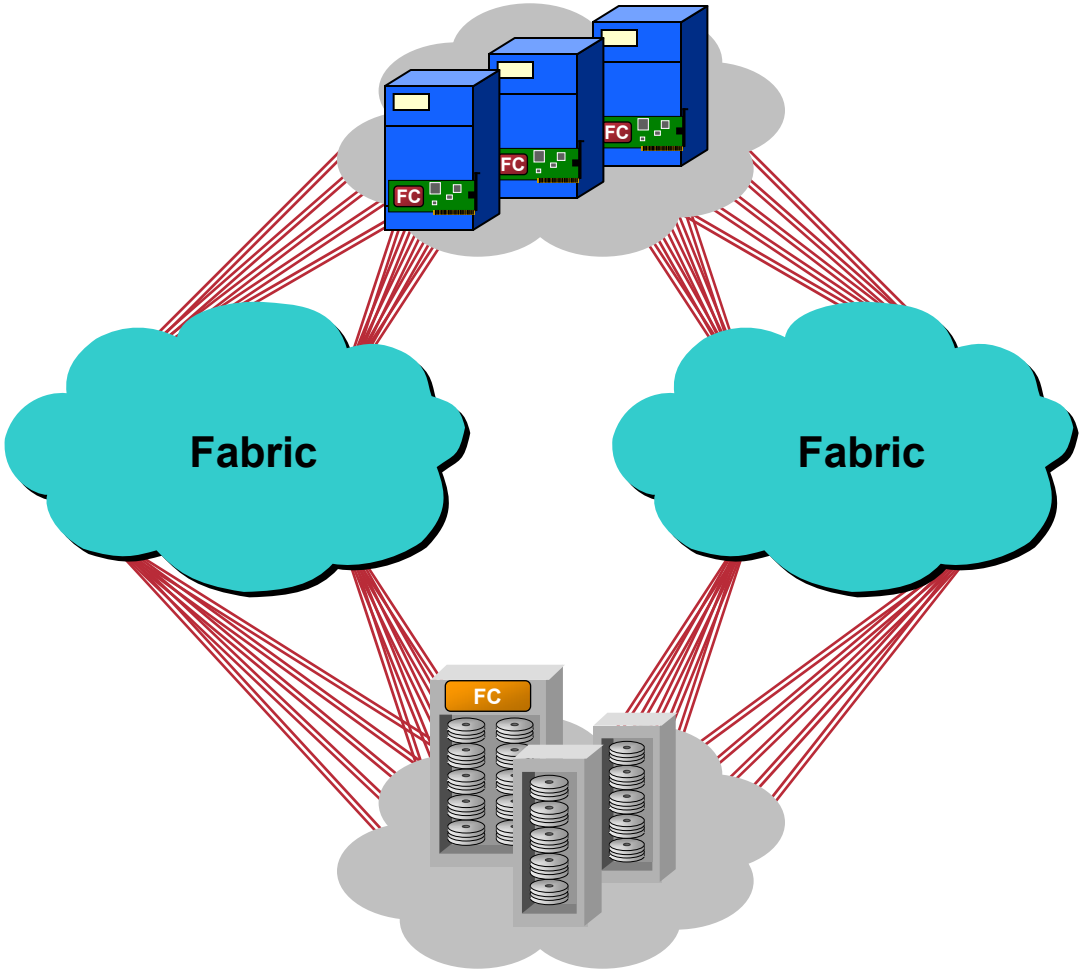


- 100% efficient port utilization
  - List/port = $2600 x 512 ÷ 512 = $2600
- High performance switches, 1.44 Tbps
- Forward congestion control for ISLs
- Scalable port aggregation
- Virtual SANs provide isolation and per VSAN fabric services

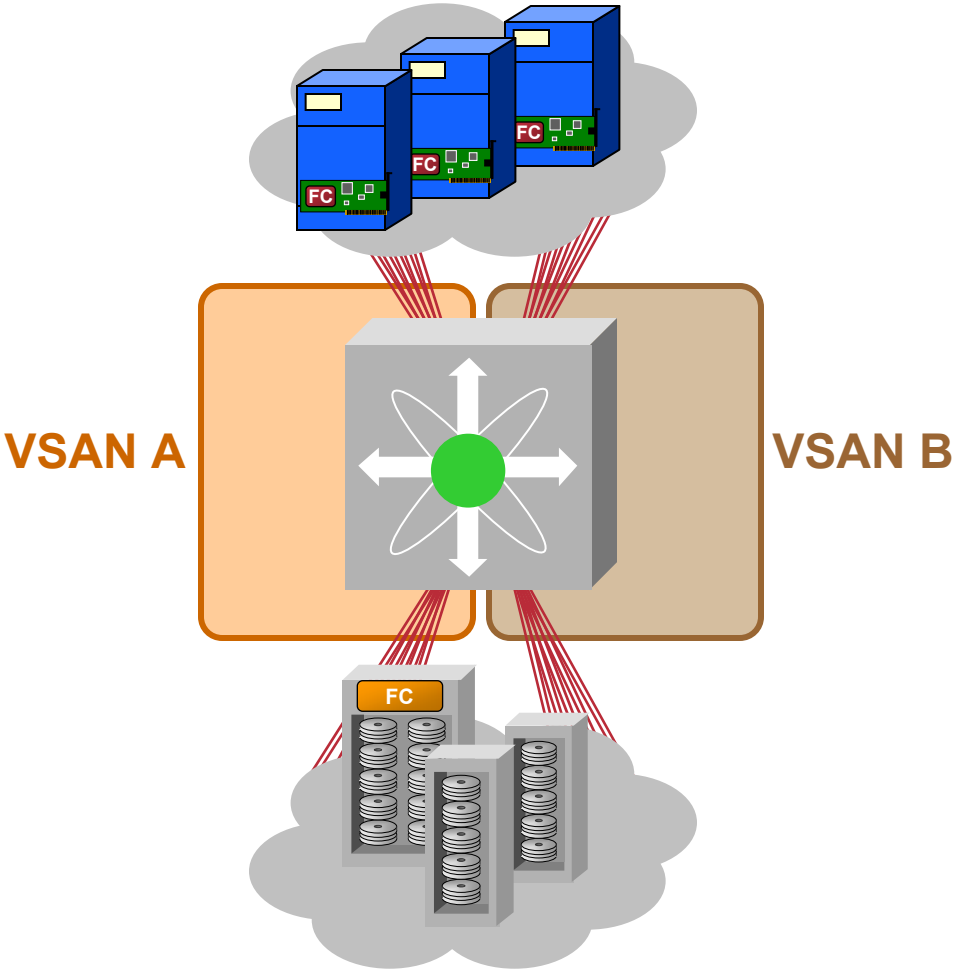
512 Port SAN Built w/ 256-Port Directors



What is a Redundant Fabric?

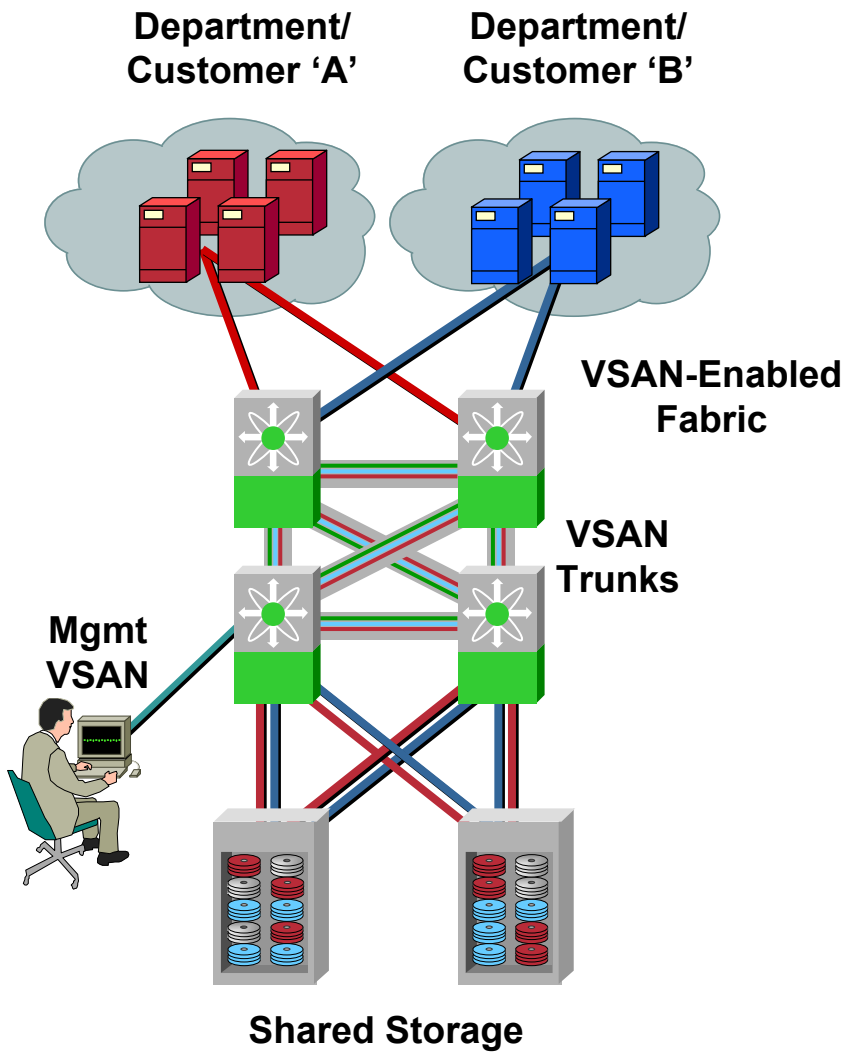


What is a Redundant Fabric?

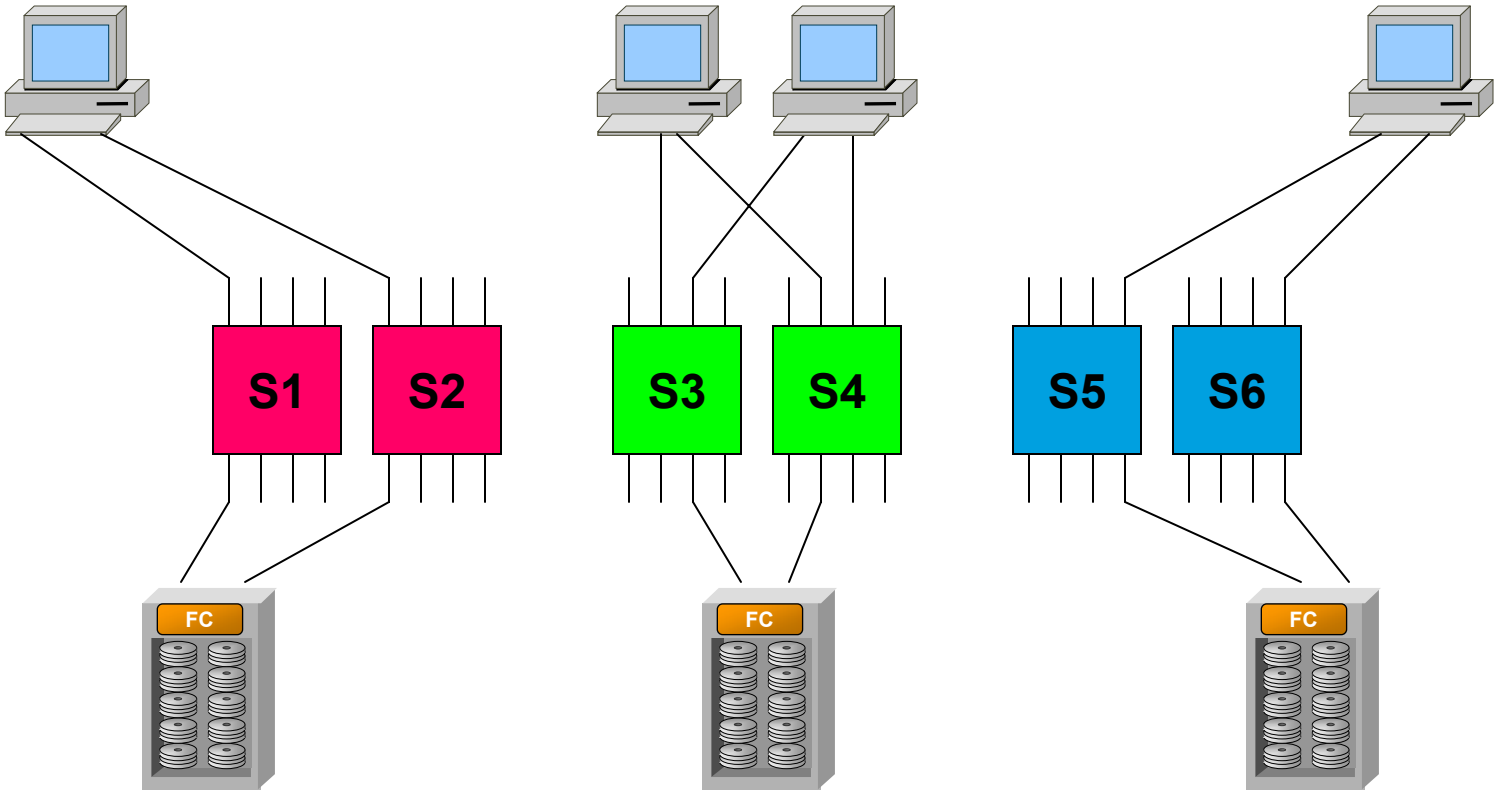


Virtual SANs (VSANs)

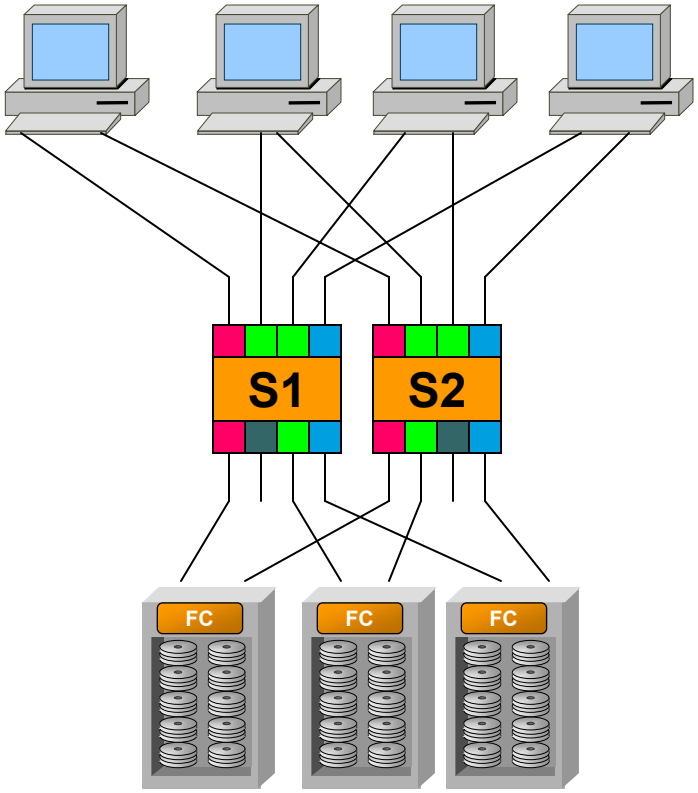
- **Overlay isolated virtual fabrics on same physical infrastructure**
  - Each VSAN contains zones and separate (replicated) fabric services
  - VSAN membership determined by port or WWN
- **Eliminates costs associated with separate physical fabrics**
- **VSANs for availability**
  - Isolate virtual fabrics from fabric-wide faults/reconfigurations
- **Security**
  - Complete hardware isolation
- **Scalability**
  - Replicated fabric services
  - Thousands of VSANs per storage network



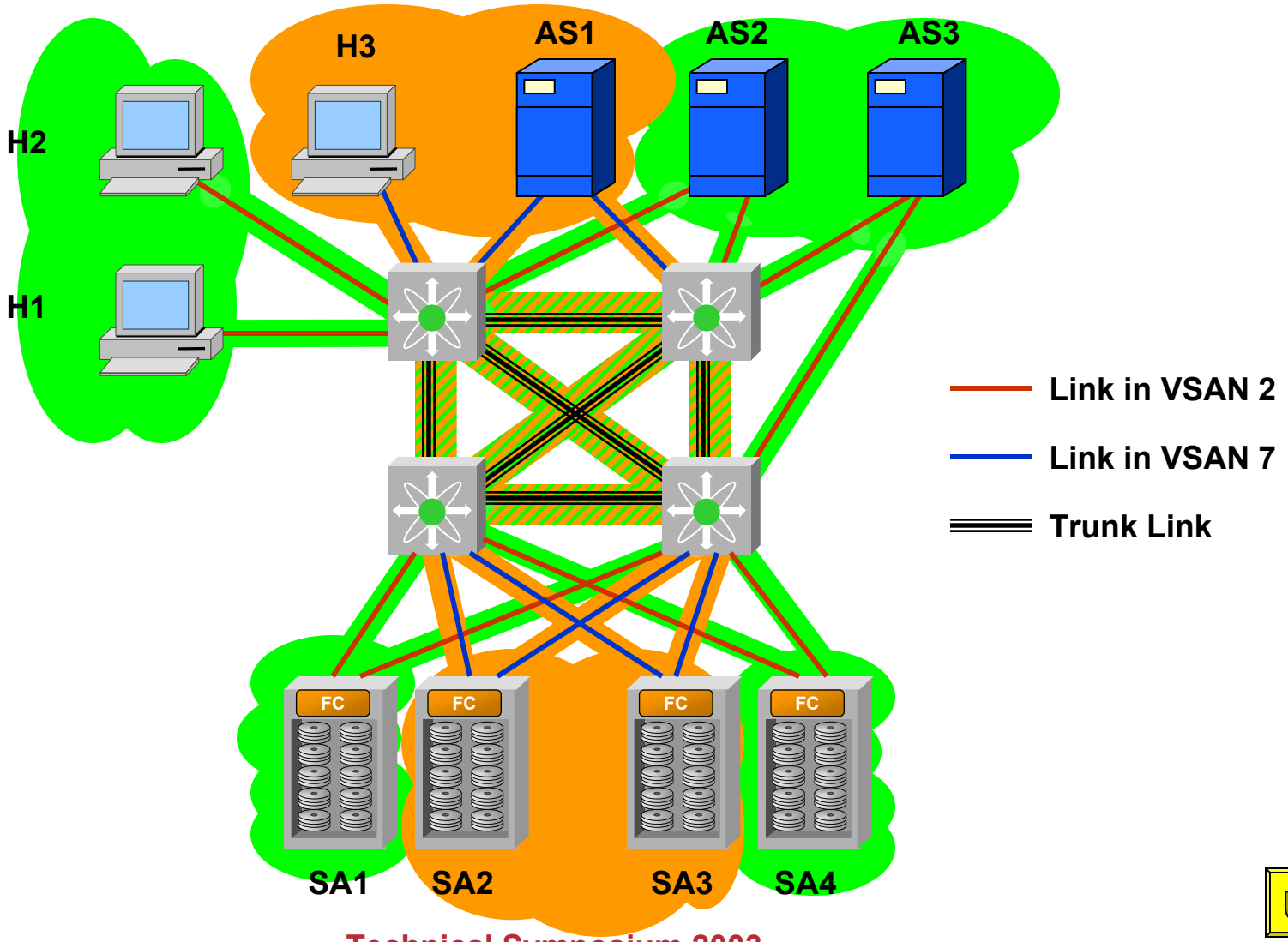
Without VSANs



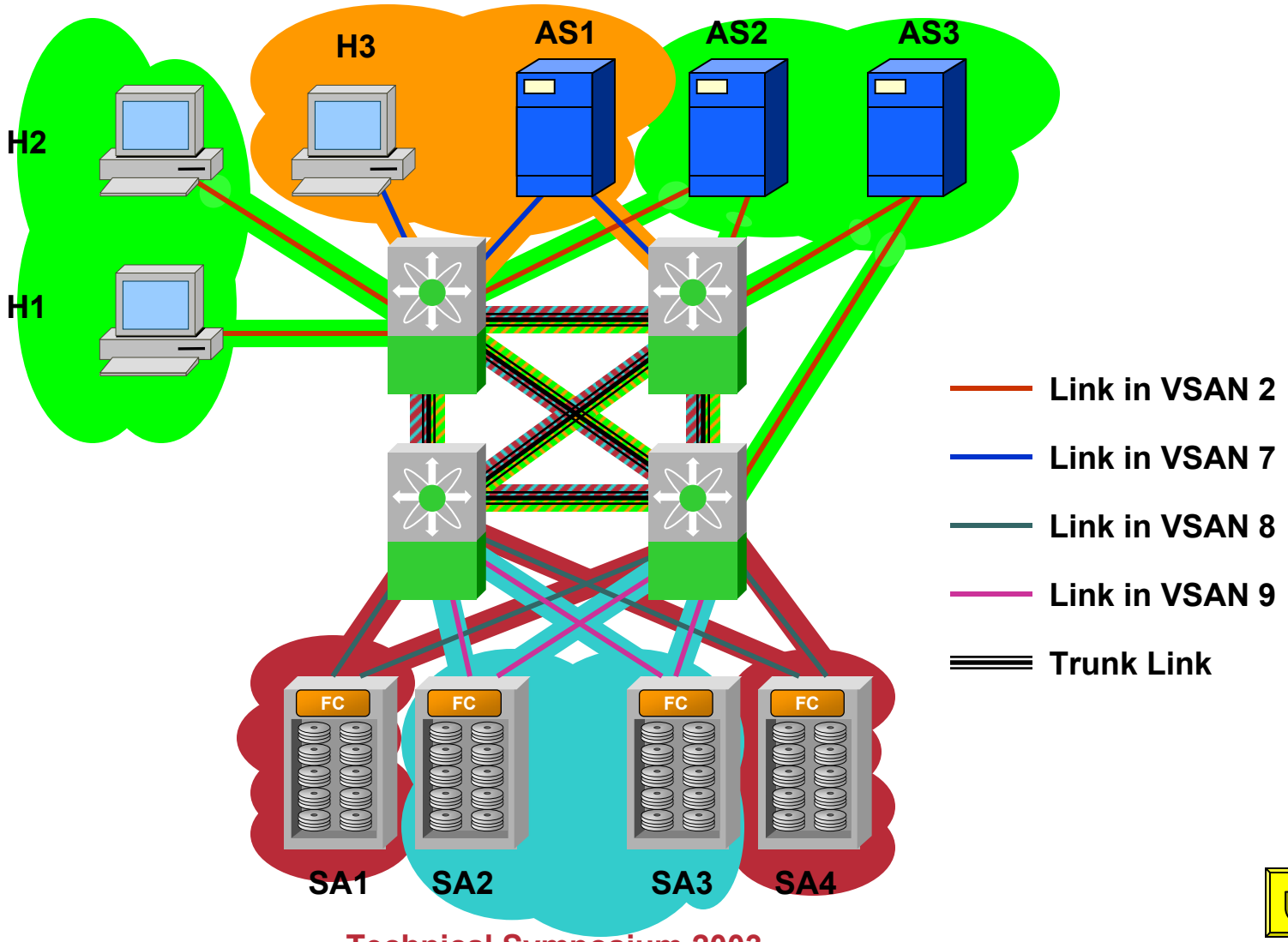
With VSANs



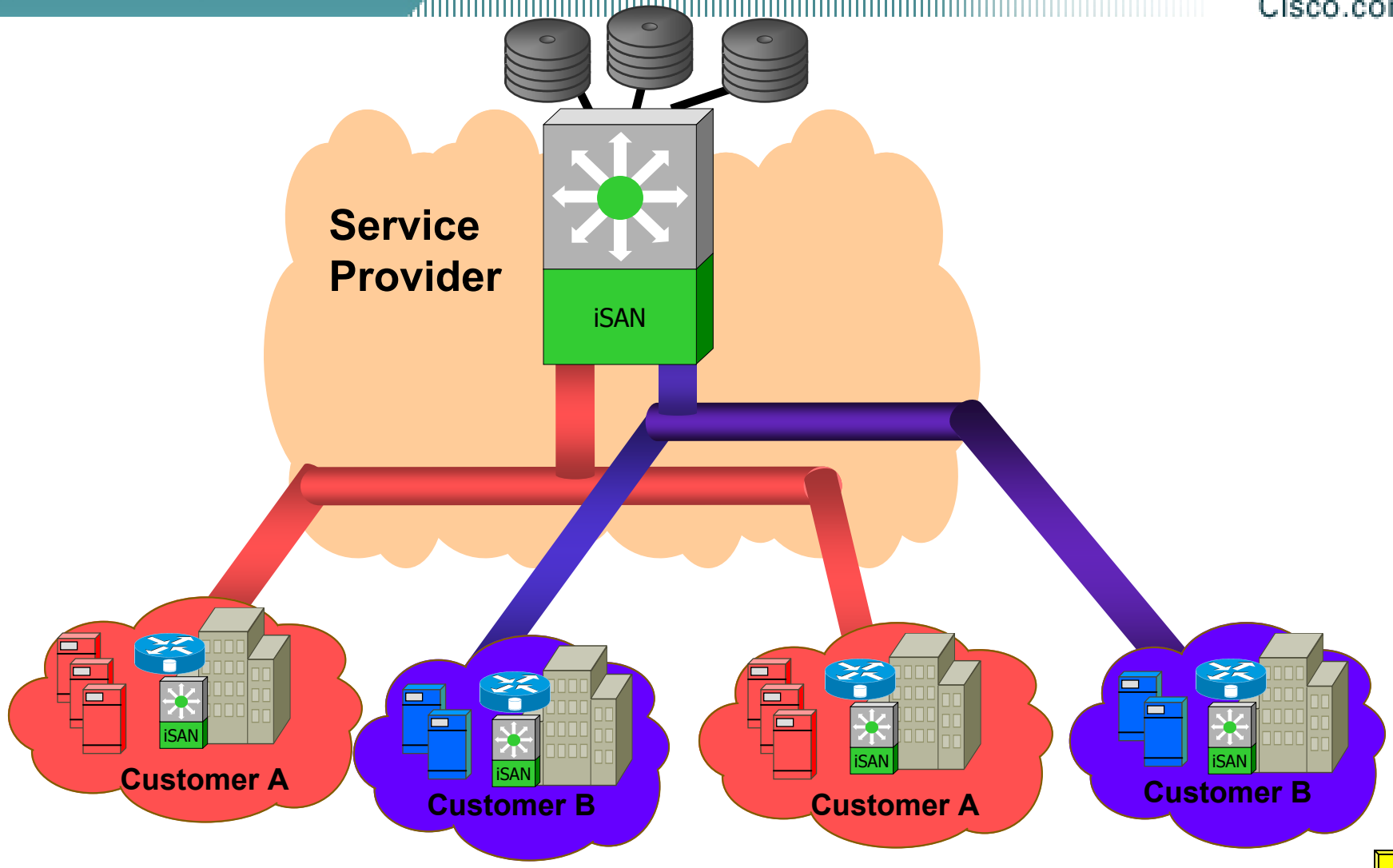
Multiple Switches with VSANs



With Virtualization



Powerful for Service Providers

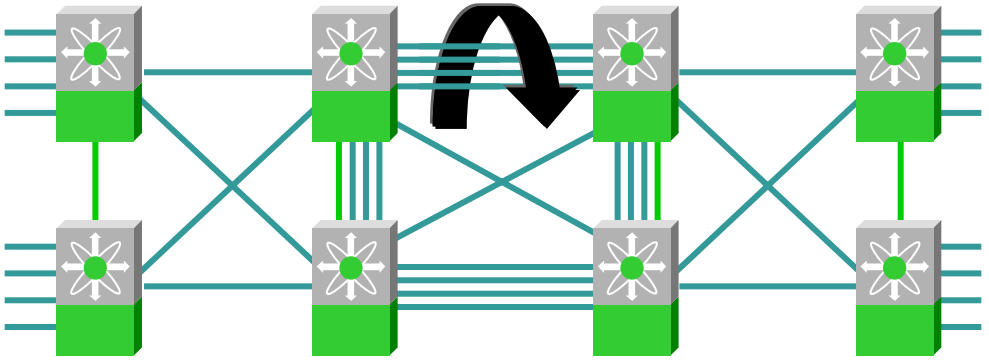


High Availability

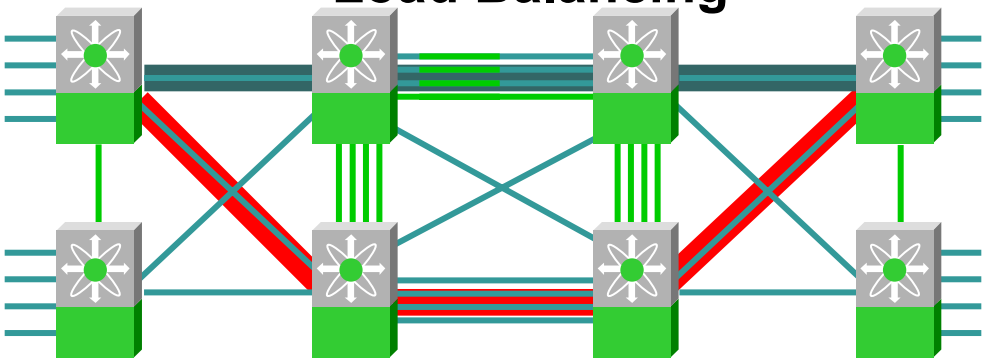
Intelligent Network Services—Multipath Forwarding and PortChannel

- Optimize use of fabric
- Bundle up to 16 links for aggregate of 32Gbps
- Utilize up to 16 equal-cost paths
- Use of unequal-cost paths via FSPF
- Reduce costs of adding more bandwidth to fabric
- Hardware-based intelligent load distribution

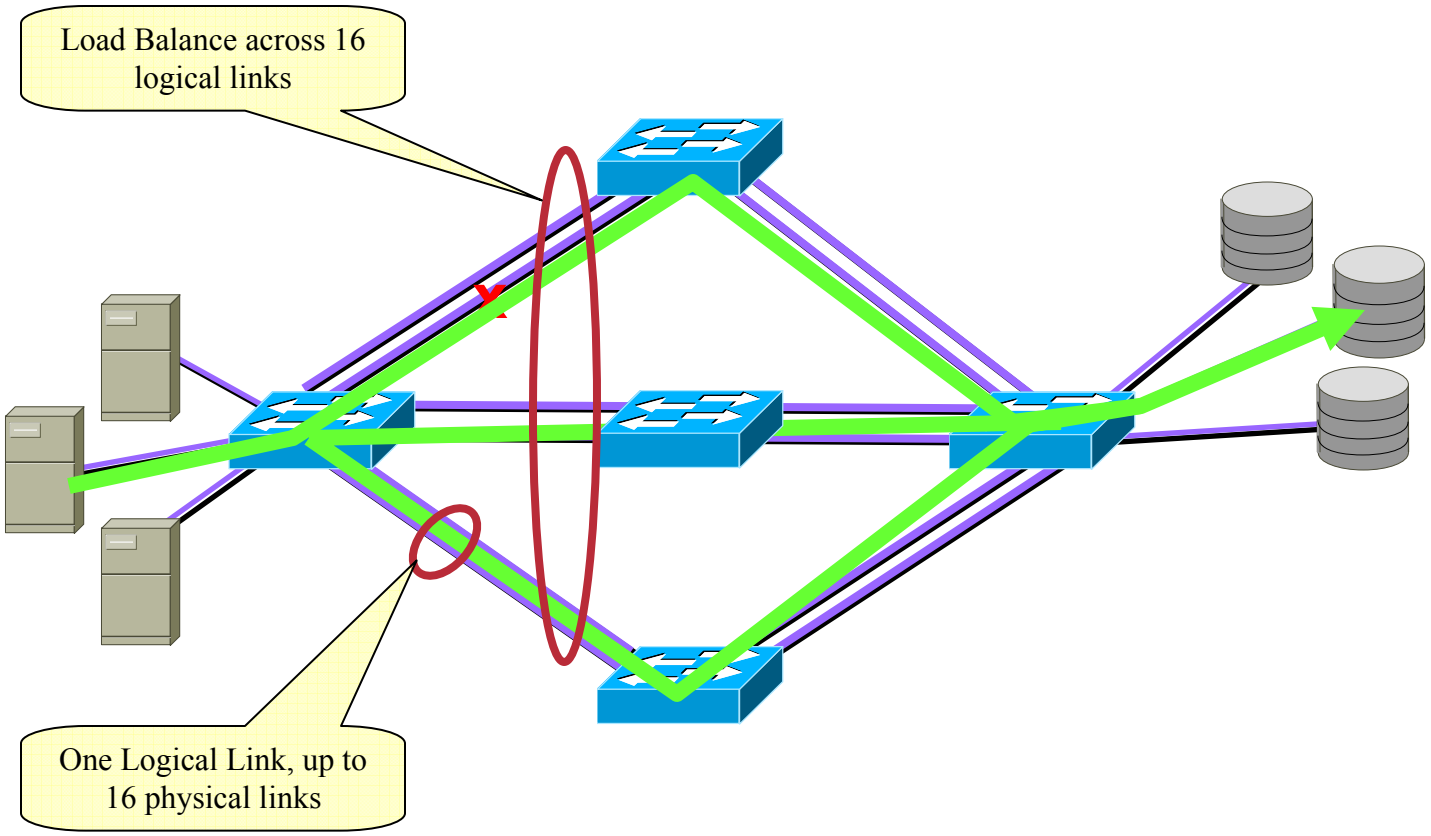
PortChannel



Load Balancing

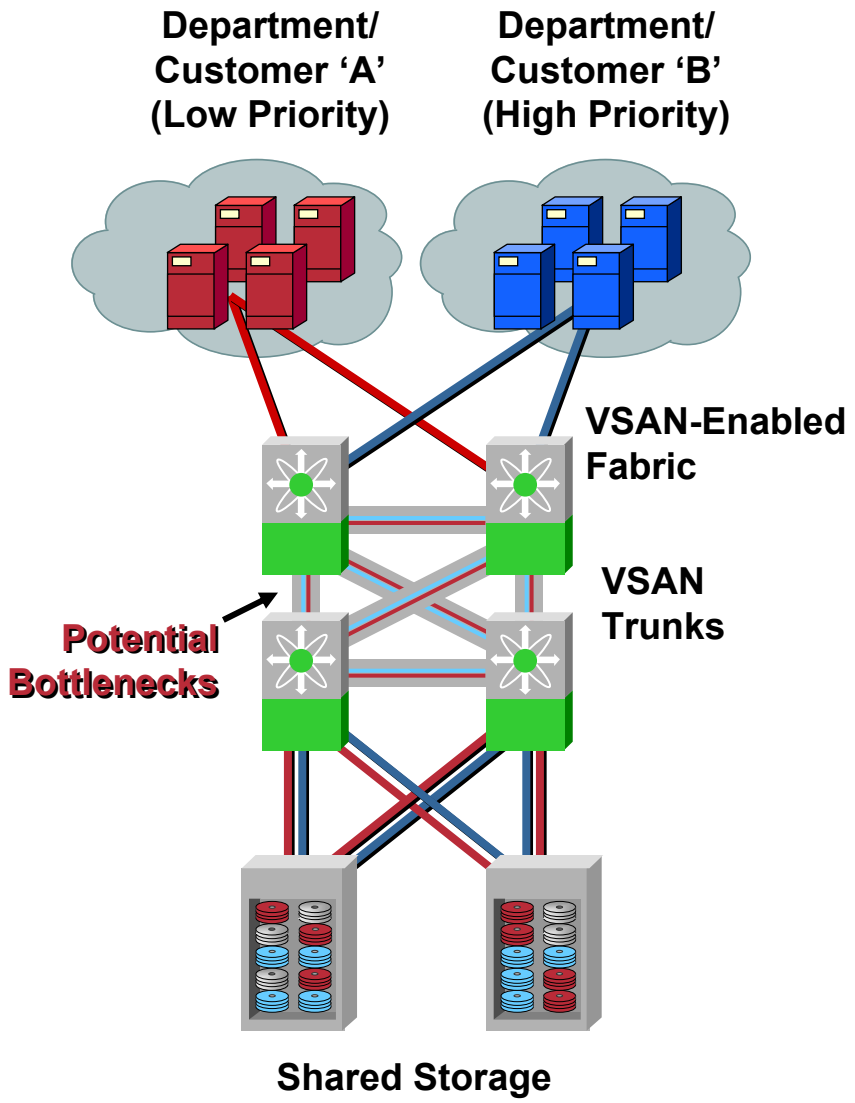


Multi-path Load Balancing

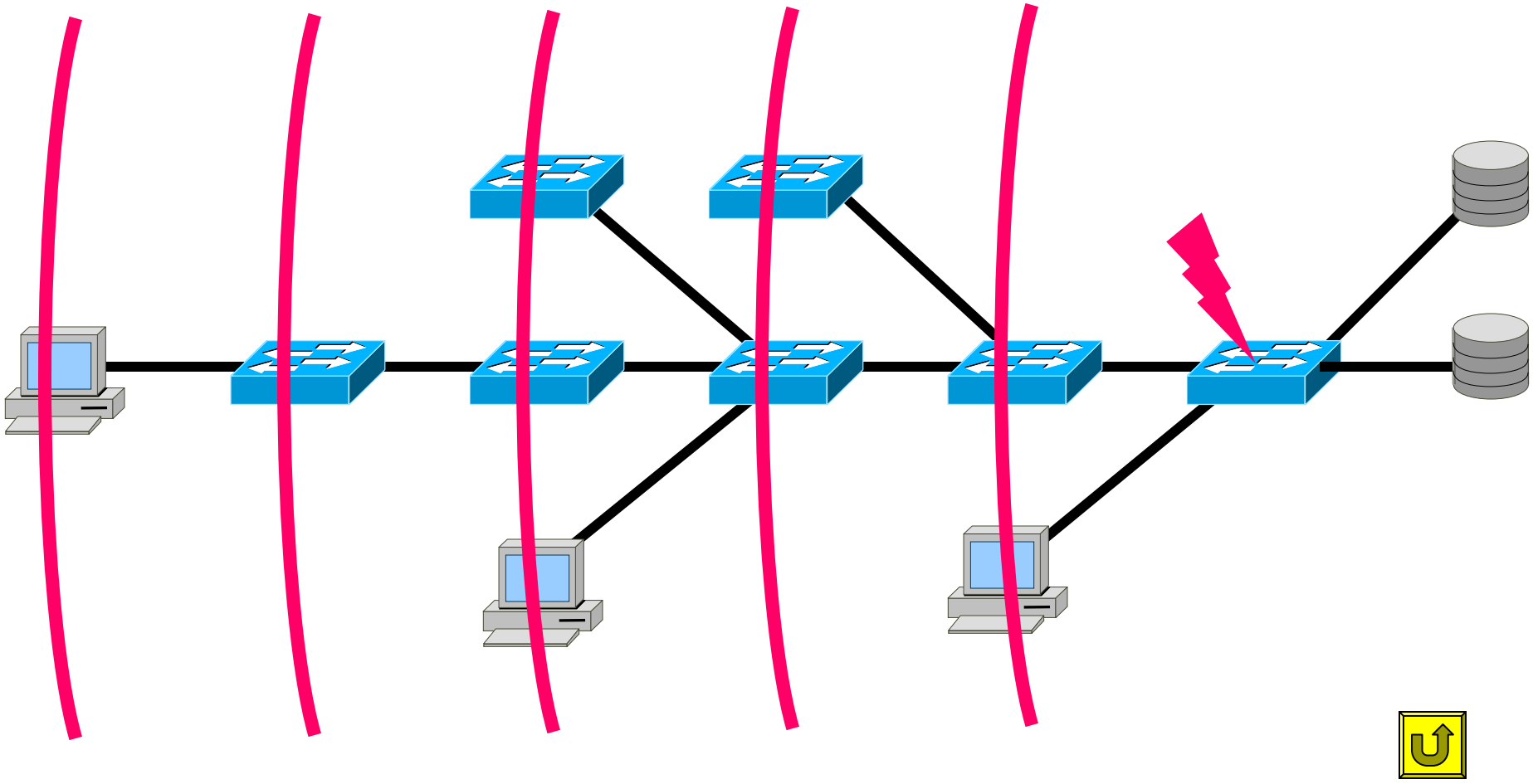


Advanced Traffic Management

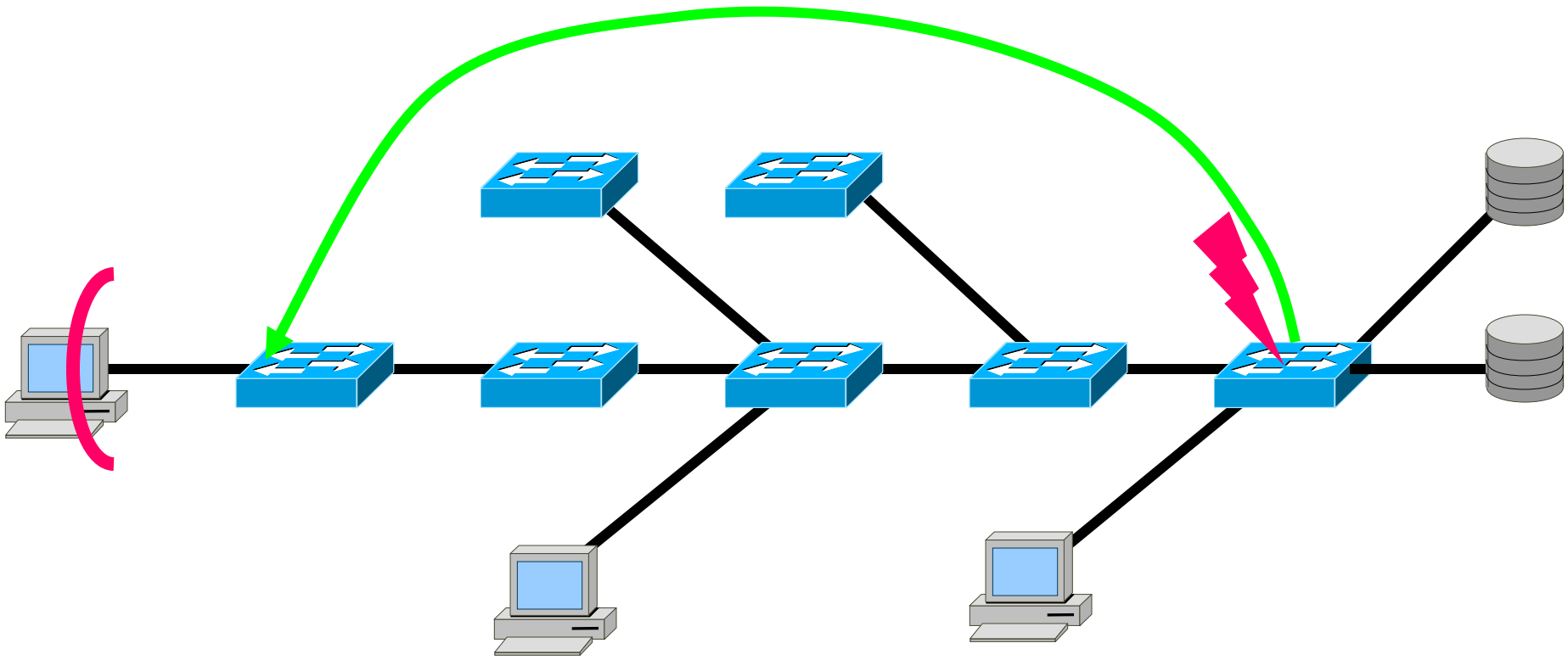
- Forward Congestion Control (FCC) mechanism can throttle back traffic at its origin
- QoS allows traffic to be intelligently managed
  - Low-priority traffic throttled at source
  - High-priority traffic not affected
  - Minimizes impact of oversubscription
  - Allows more economical topologies
- Virtual Output Queuing for optimal crossbar performance
- MPLS for advanced traffic engineering capability



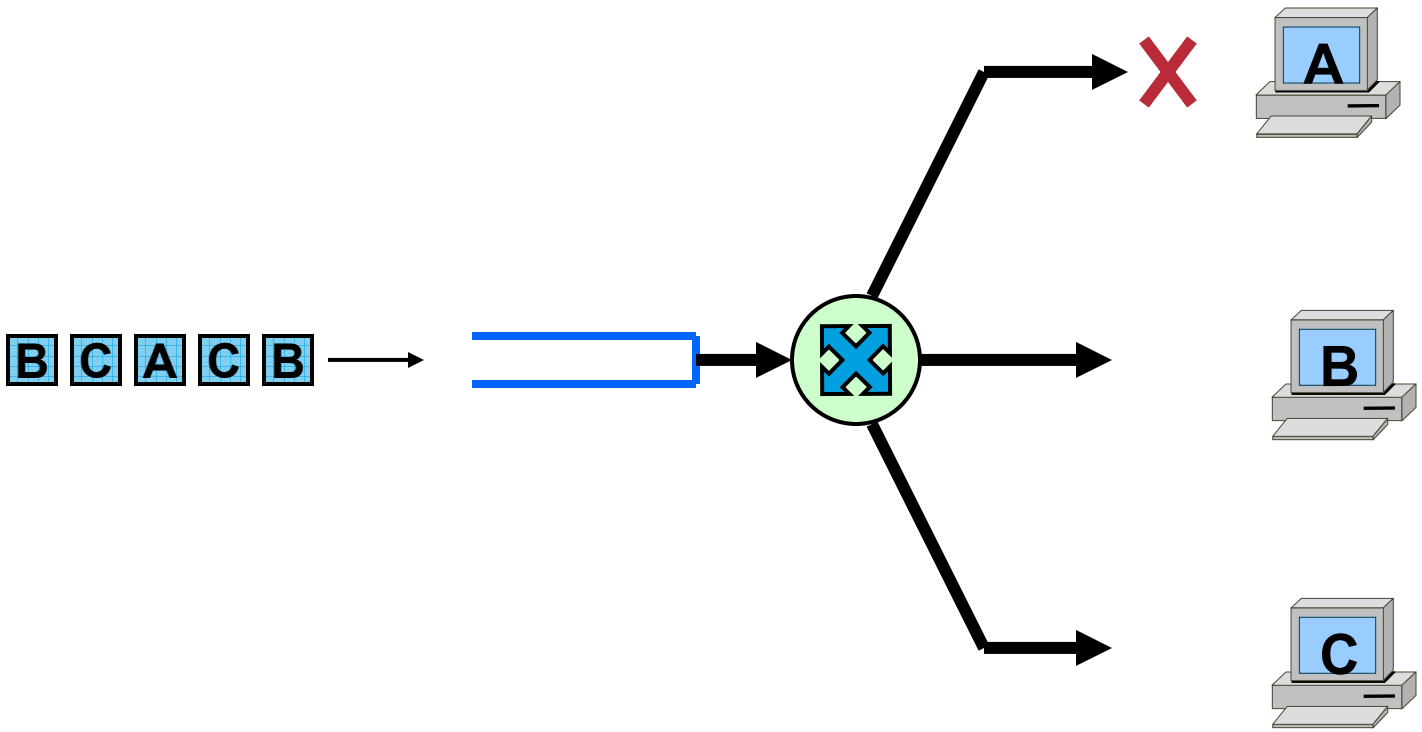
B2B Congestion Control



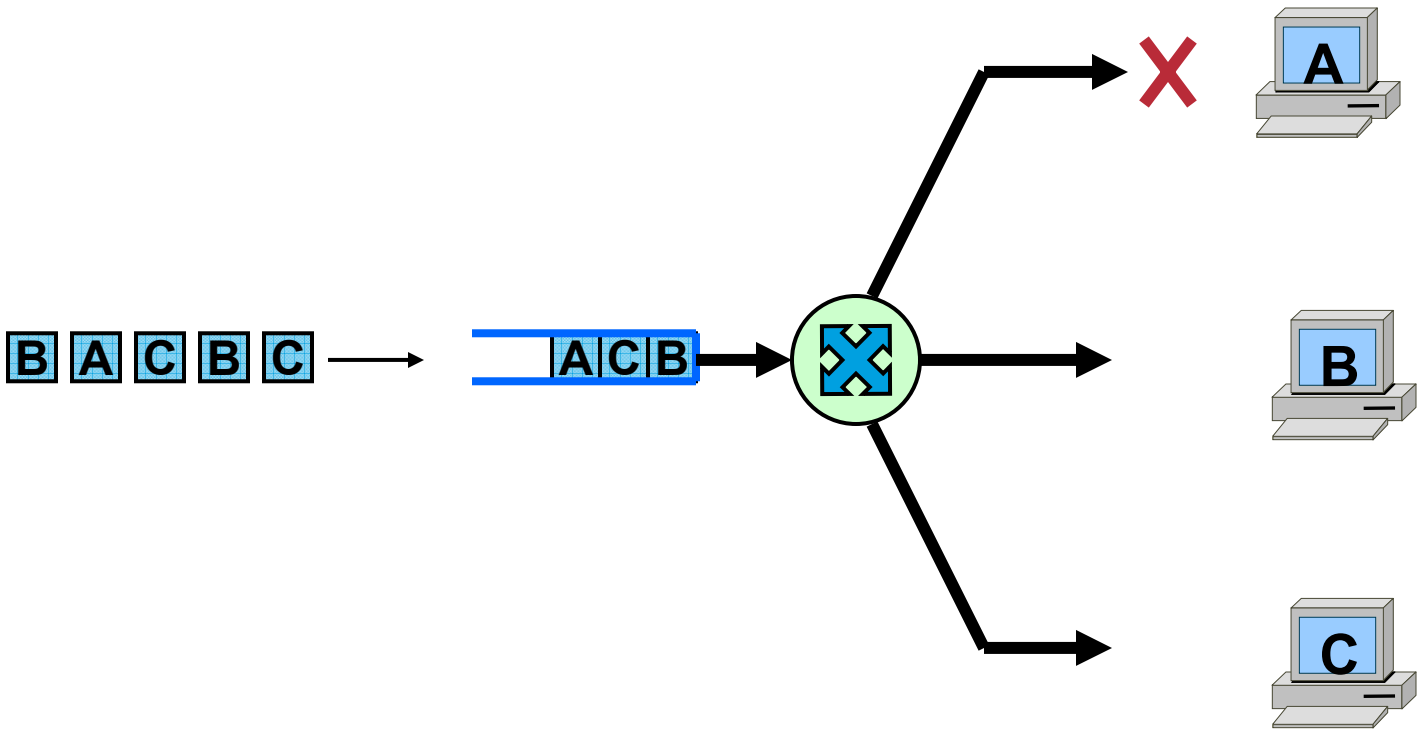
FCC – FC Congestion Control



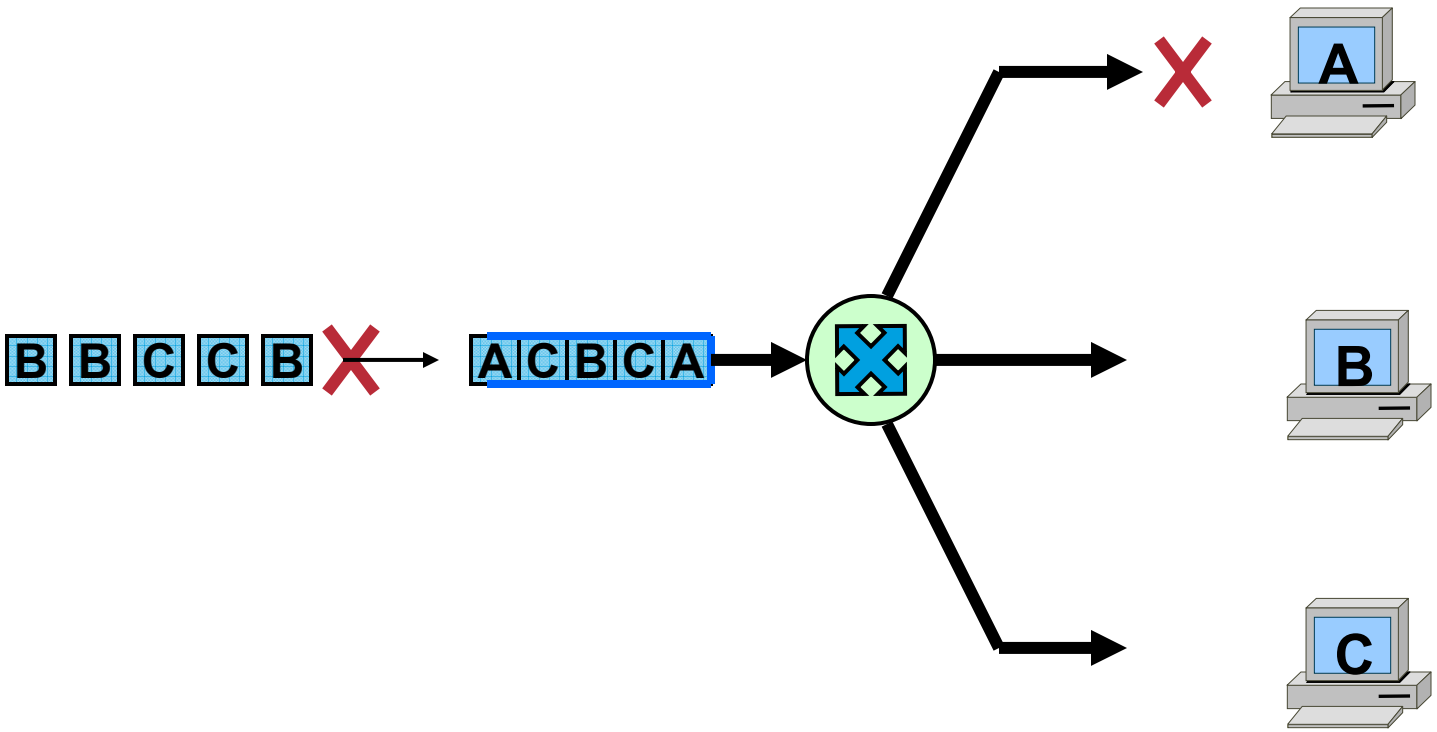
Head Of Line Blocking



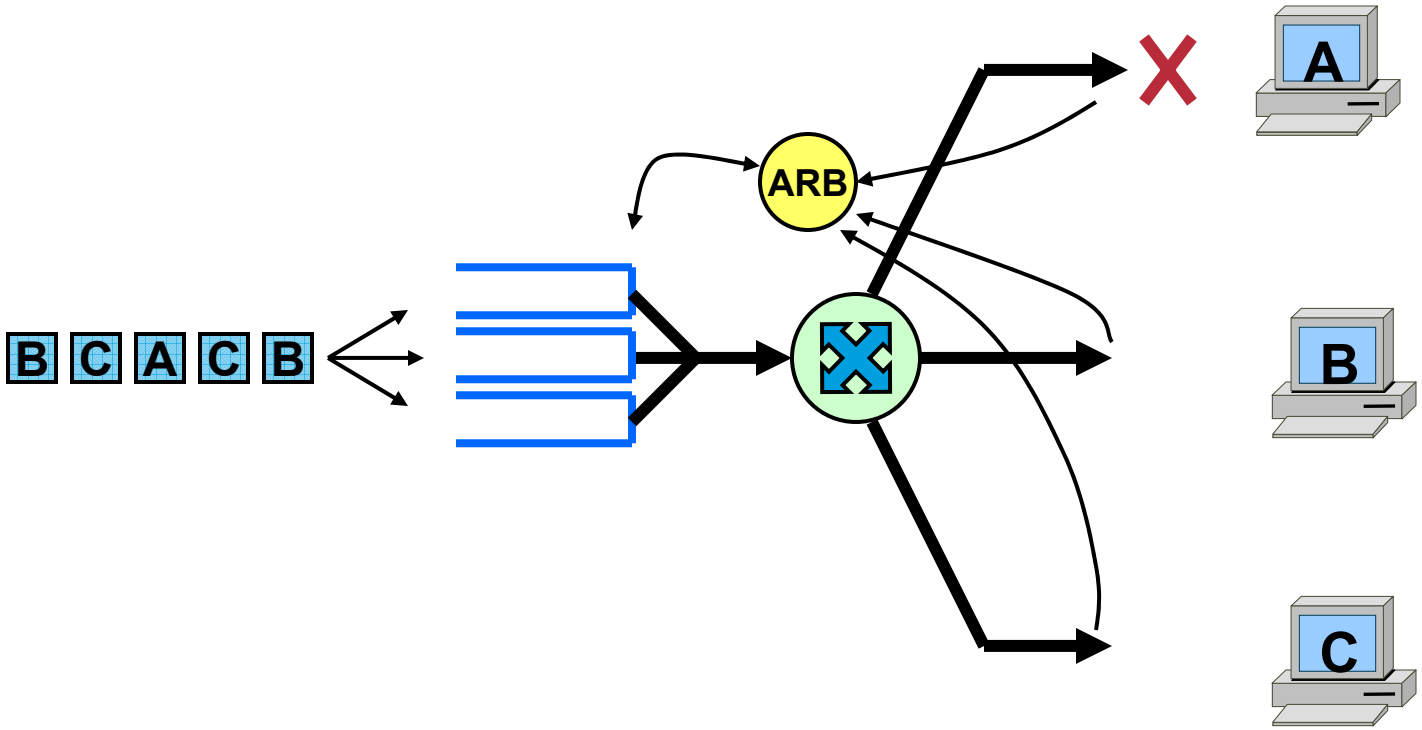
Head Of Line Blocking



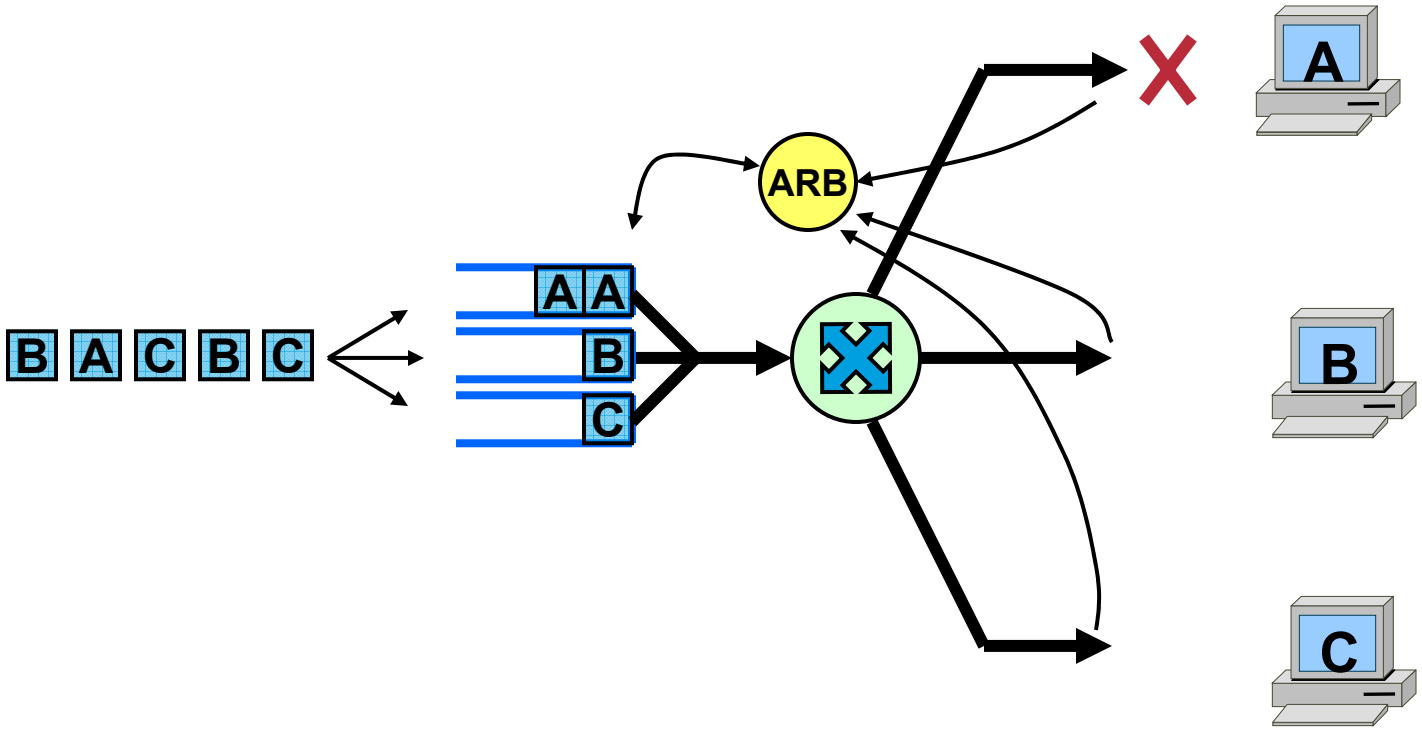
Head Of Line Blocking



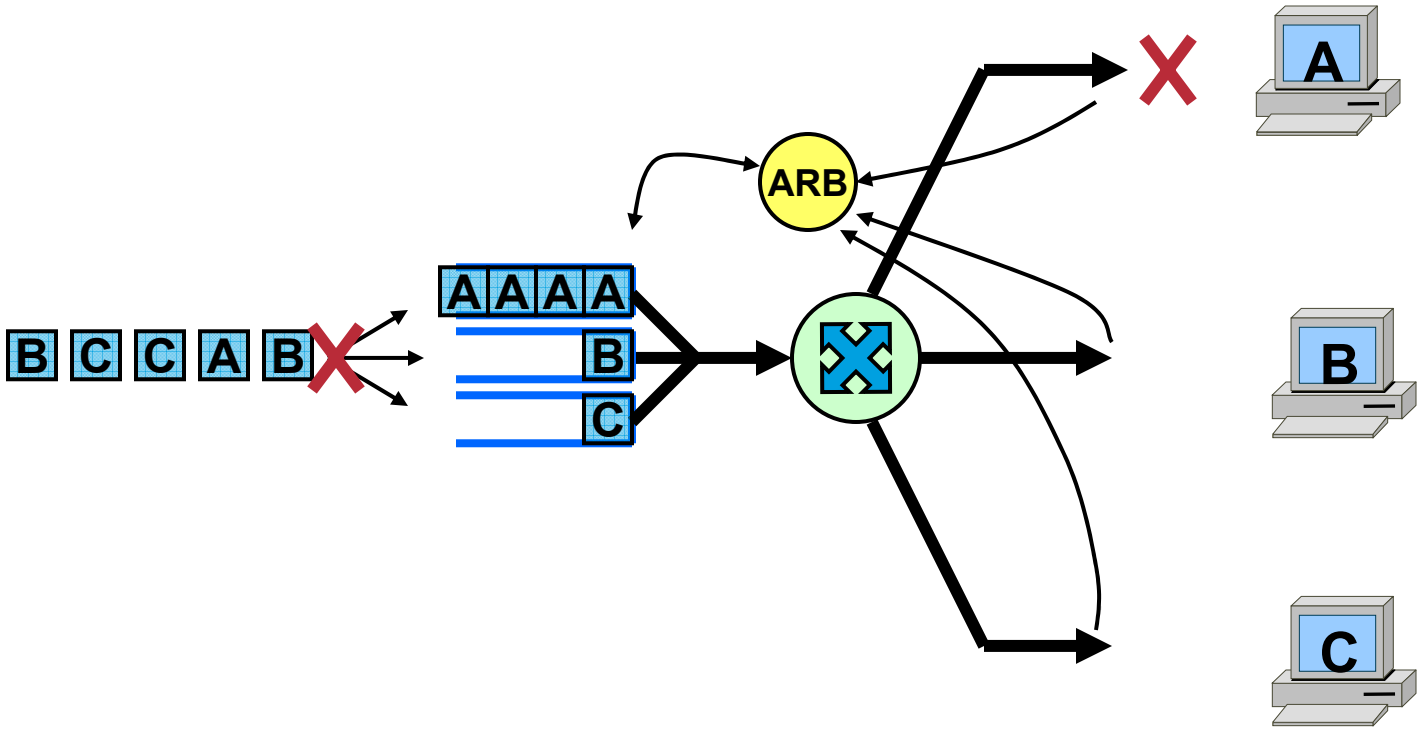
VOQ – Virtual Output Queues



VOQ – Virtual Output Queues

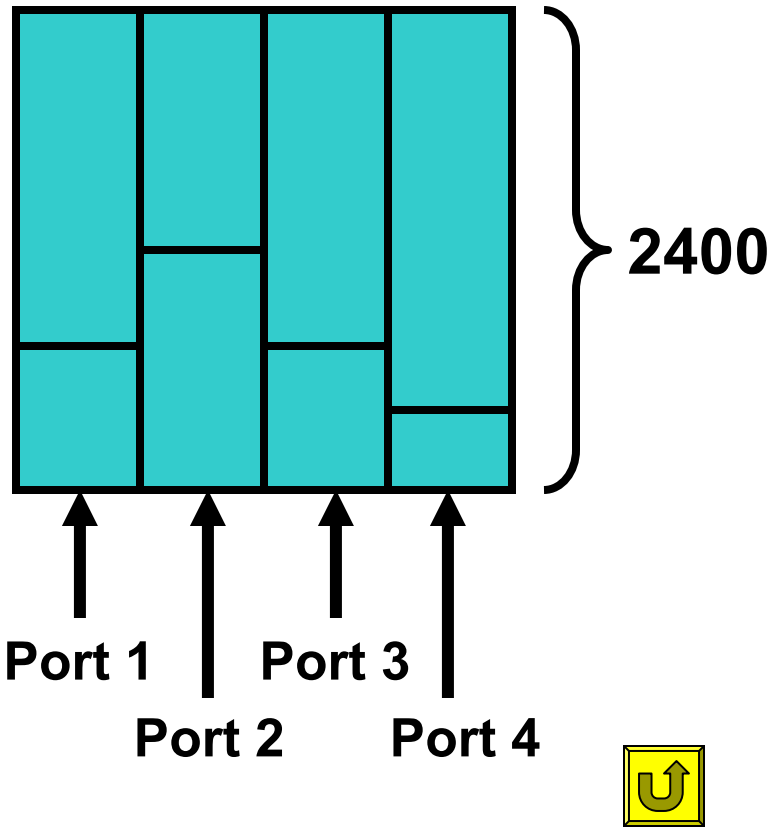


VOQ & Head Of Line Blocking



Lots of Flexible Buffering

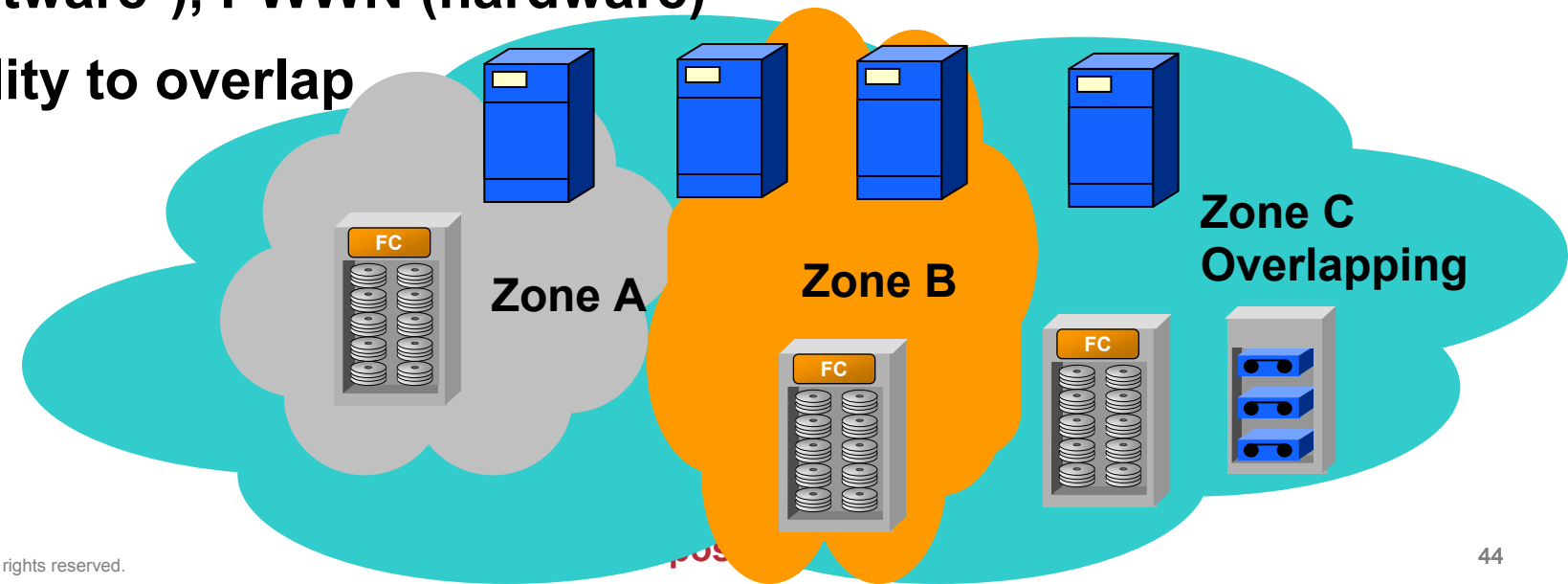
- Pool of Buffers Shared Across 4 Ports
- 2400 Buffers total in pool
- Flexible Carving of Buffers
- User settable credit amounts up to 255 w/ default
- Each buffer is 1 Maximum frame size



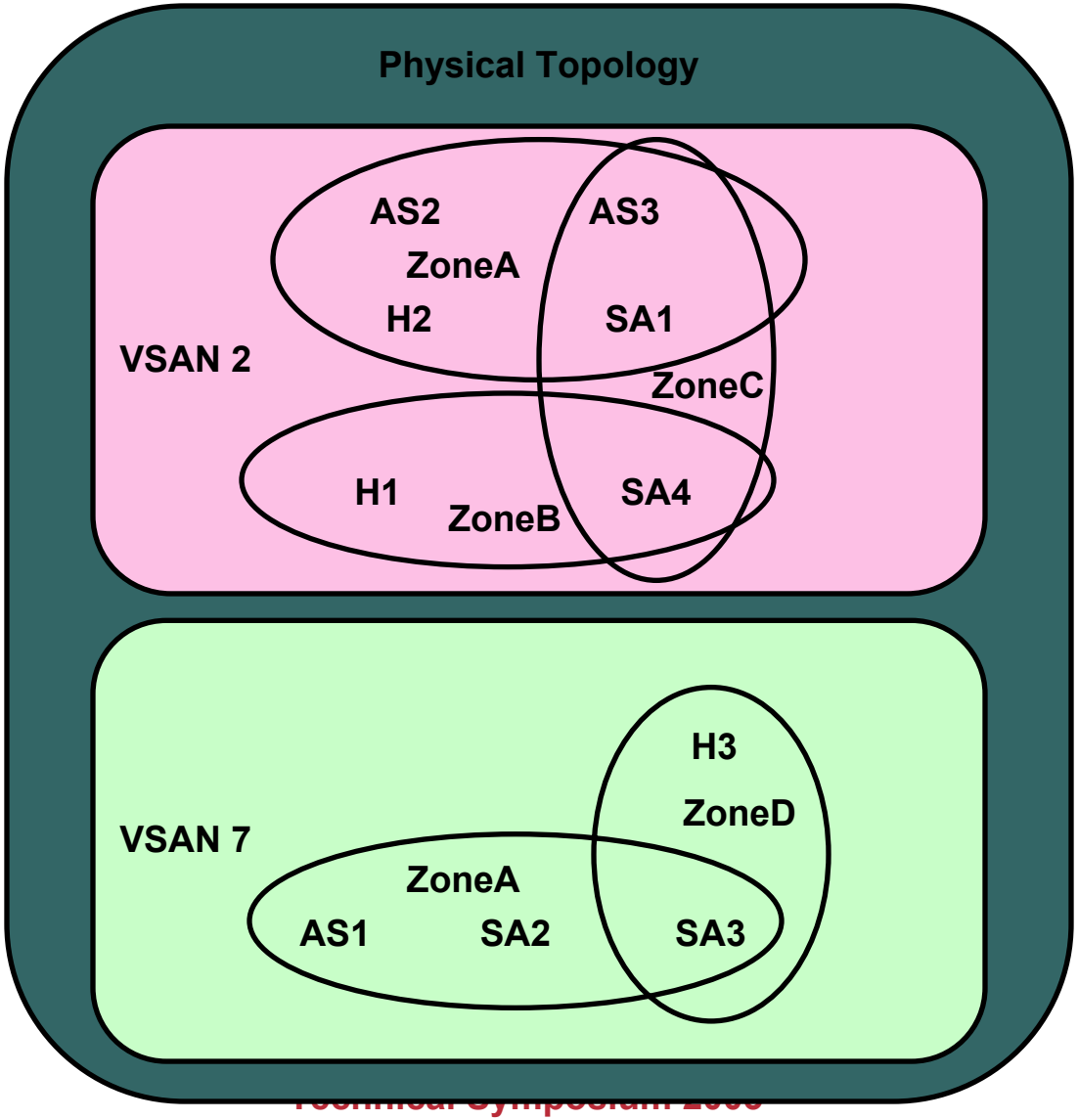
Security

Storage Features

- **Zones**
  - **Why were they invented?**
    - **OS's accessing blocks that do not belong to them**
    - **Security – First Attempt**
  - **How are they assigned – WWN (software*), PWWN (hardware)**
  - **Ability to overlap**

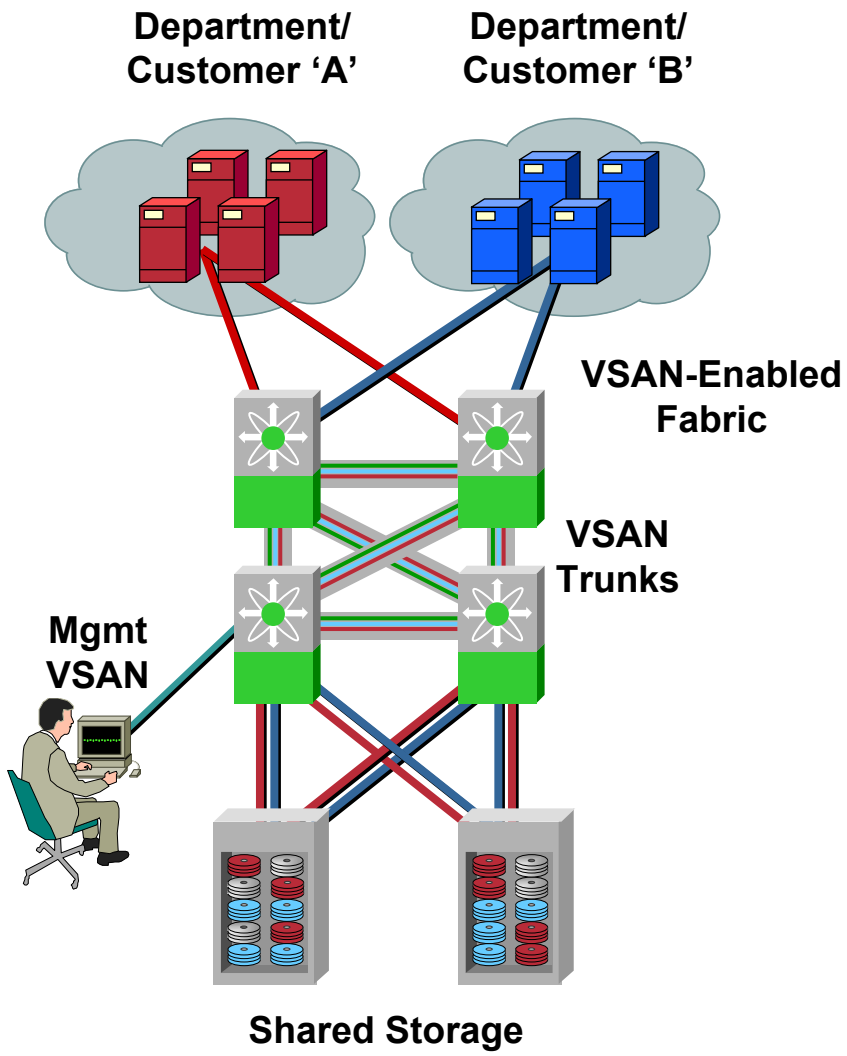


Zones vs VSANs

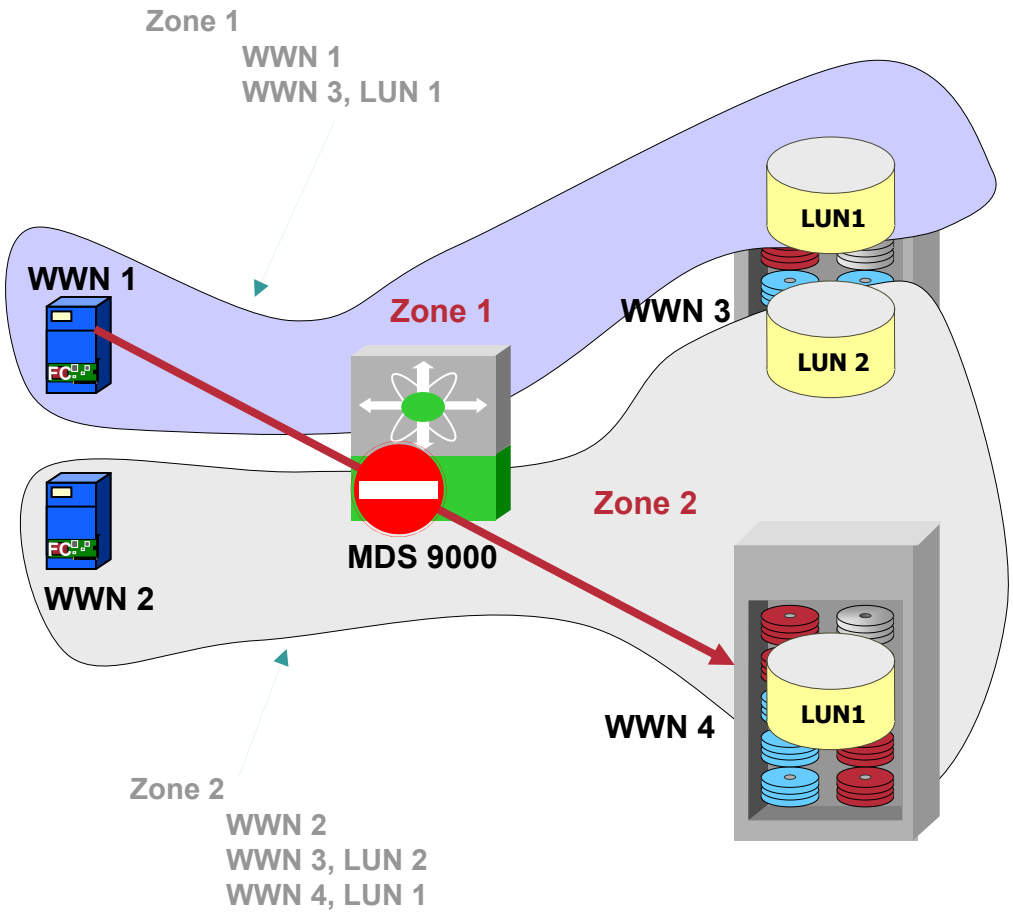


Virtual SANs (VSANs)

- **Overlay isolated virtual fabrics on same physical infrastructure**
  - Each VSAN contains zones and separate (replicated) fabric services
  - VSAN membership determined by port or WWN
- **Eliminates costs associated with separate physical fabrics**
- **VSANs for availability**
  - Isolate virtual fabrics from fabric-wide faults/reconfigurations
- **Security**
  - Complete hardware isolation
- **Scalability**
  - Replicated fabric services
  - Thousands of VSANs per storage network

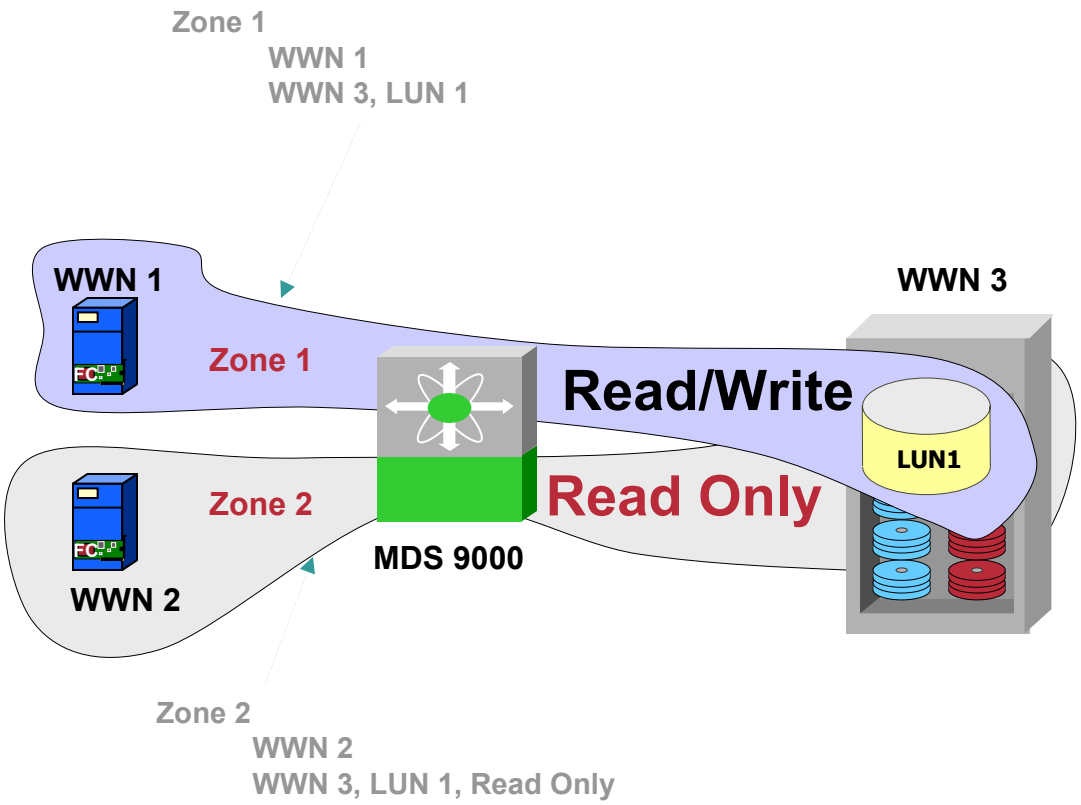


Fabric-Based LUN Zoning



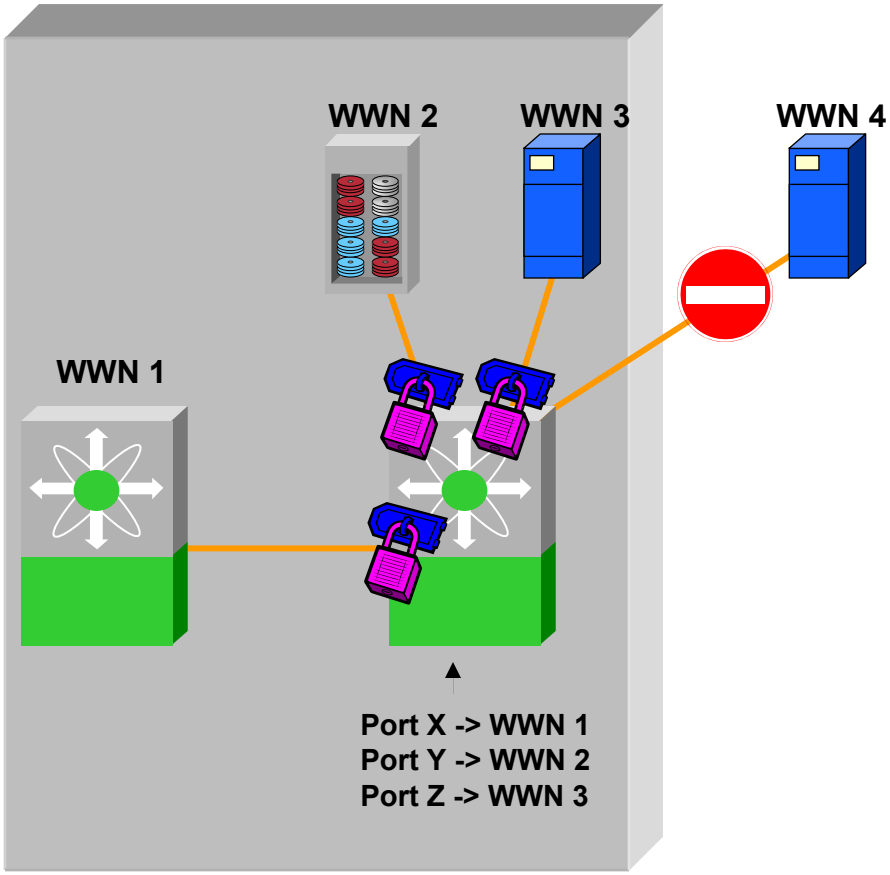
- Extension of fabric Zoning capability
- Provides an additional level of security in the fabric
- Hardware enforcement of LUN Zoning in MDS switch

Read-Only Zones



- **Enables a volume to be shared across different application servers**
  - **Data Warehousing servers**
  - **Web Servers**
  - **Backup server**
- **Zoning based on SCSI I/O operation type:**
  - **Read only**
  - **Read and Write**

Port Security and Fabric Binding



- **Port Security**

  Security feature that locks down who (remote entity) can connect to a switch port

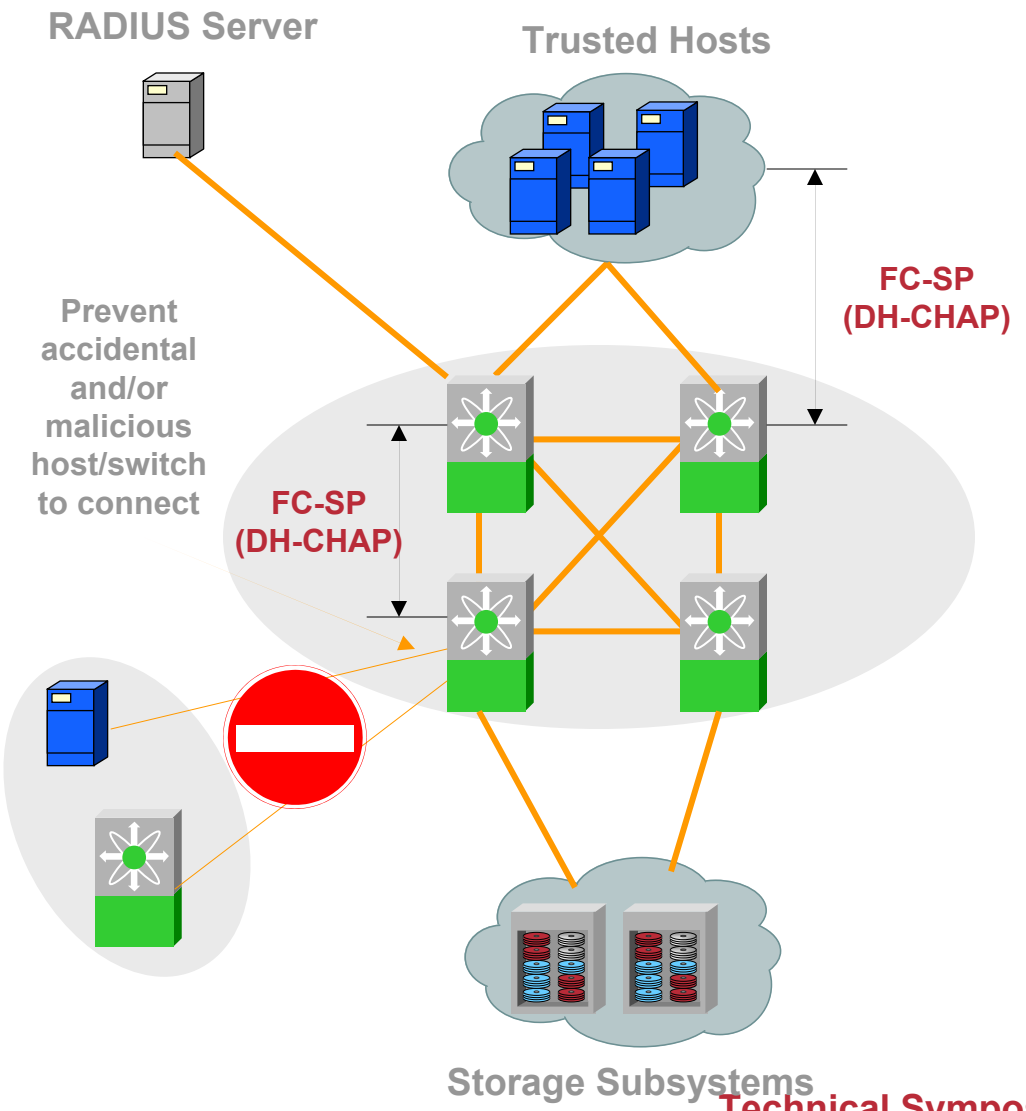
  The remote entity can be a host, target or a switch

  Remote entities are identified via WWN

- **Fabric Binding**

  Only authorized switches can connect to a secure fabric

Switch-Switch and Host-Switch Authentication



- **Prevents accidental and/or malicious devices from joining a secure SAN**
  - **Hosts**
  - **Switches**
- **Based on FC-SP security protocol**
  - **Initial phase focused on authentication**
  - **DH-CHAP between devices**
  - **Centralized RADIUS server**

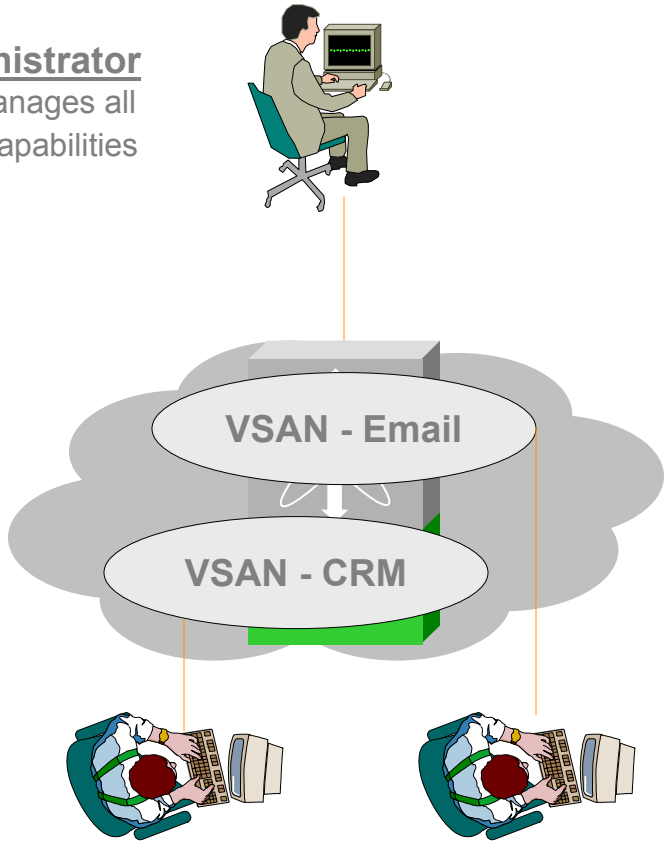
Fabric-Wide Security (cont.)

- **AES for SNMP and SSH**
  - Advanced Encryption Standard (AES) enhances the encryption capabilities of SNMP v3 and SSH
  - Increased security with larger key sizes (128*, 192 and 256 bits)
- **TACACS+**
  - Complements RADIUS server to provide centralized AAA services
  - Reliable (TCP/IP), and secure connection between MDS and TACACS+ server

Management

VSAN Based Roles

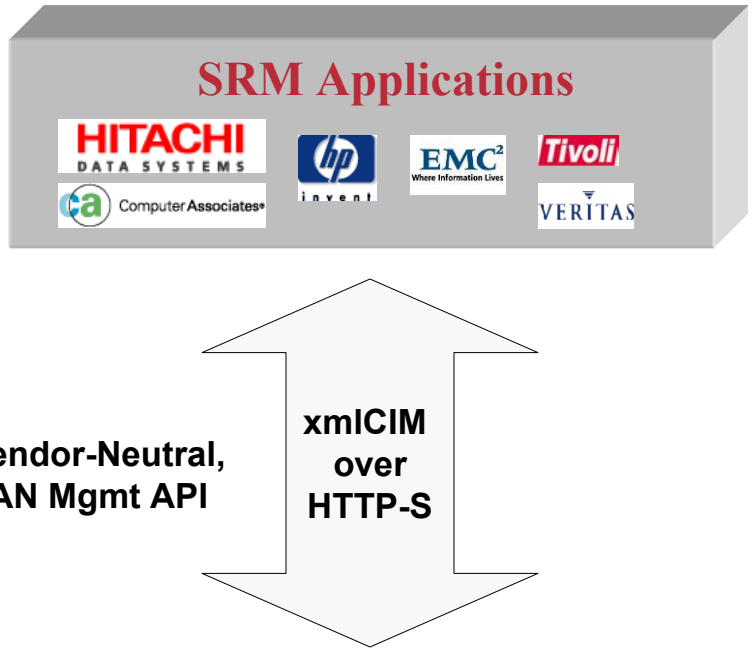
Network Administrator
Configures and manages all platform-specific capabilities



VSAN Administrators
Configure and manages only their VSANs

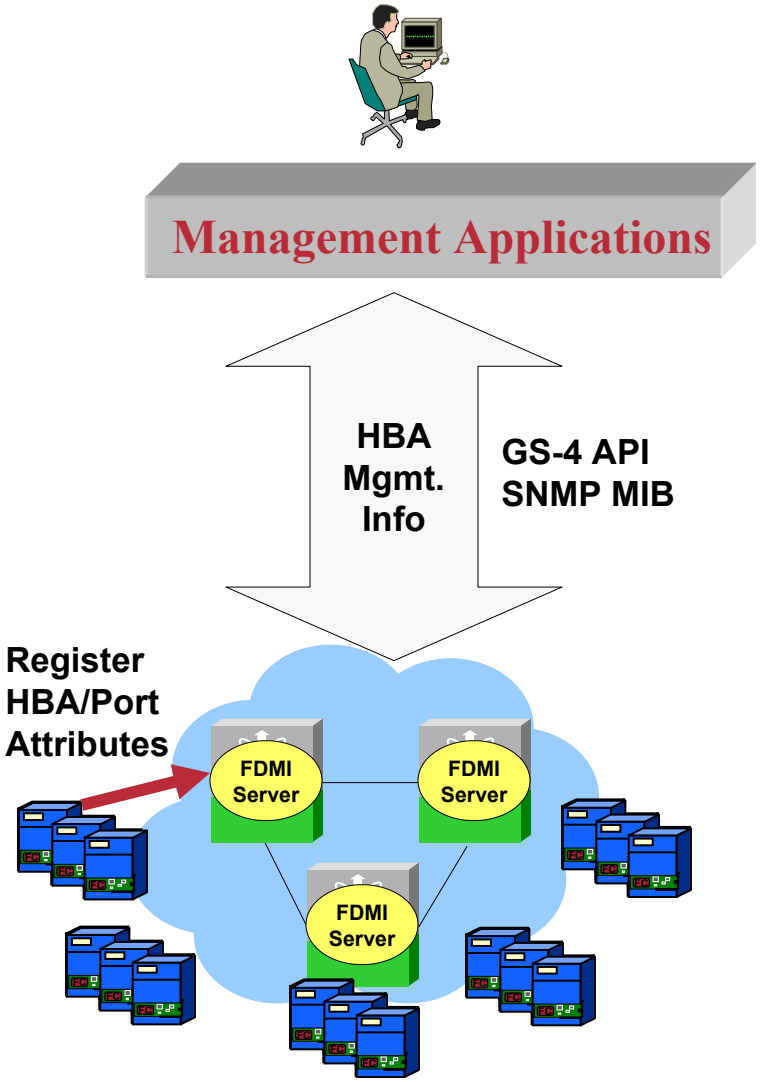
- **Enables deployment of VSANs that fit existing operational models**
  - **Network-admin configures all platform-specific capabilities**
  - **VSAN-admin(s) configure and manage their own VSANs**
- **The existing "role" definition is enhanced to include VSAN(s)**

SMI-S (Bluefin) For Heterogeneous Storage Management



- **Storage Management Interface Specification [SMI-S], a SNIA initiative, simplifies heterogeneous storage management**
- **A vendor-neutral API for discovering, monitoring and managing devices on a SAN**
- **"Embedded CIM Provider" based on SMI-S version 1.0**
  - Discovery of storage subsystems, HBA and switches
  - Topology view
  - Switch, port, and interface view
  - Zoning management
  - Extensions to include VSANs

Centralized HBA Management

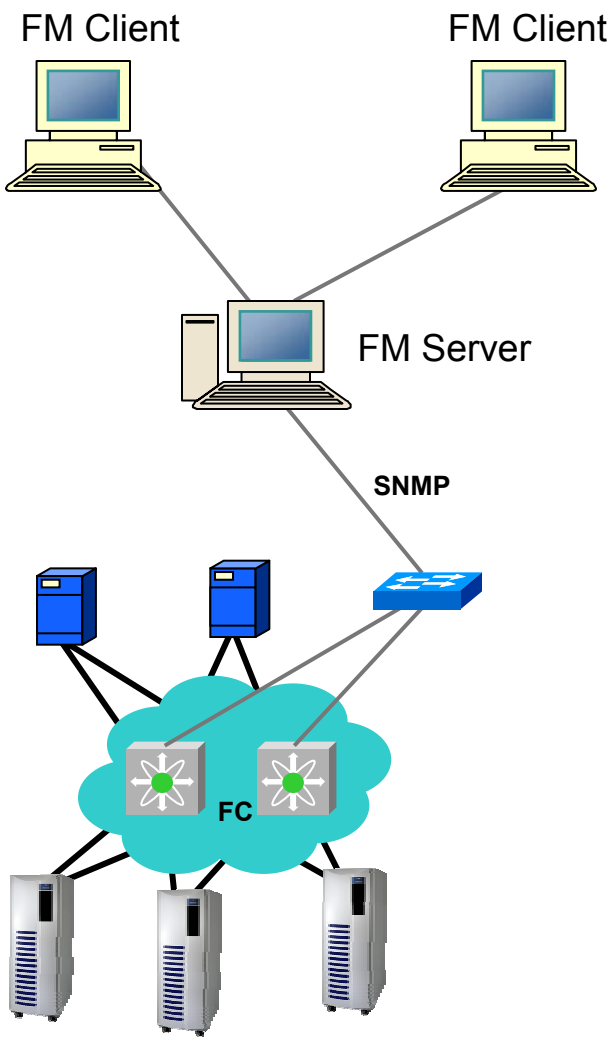


- Fabric Device Management Interface (FDMI) for centralized asset, inventory management of HBAs from fabric
- Eliminates the need of proprietary per-server agents
- FDMI Server on MDS manages the following HBA and Port Attributes:

  Manufacturer, Serial Number, Model, Hardware Version, Driver versions, ROM version, Firmware version

  OS Name and version, Max CT payload etc.

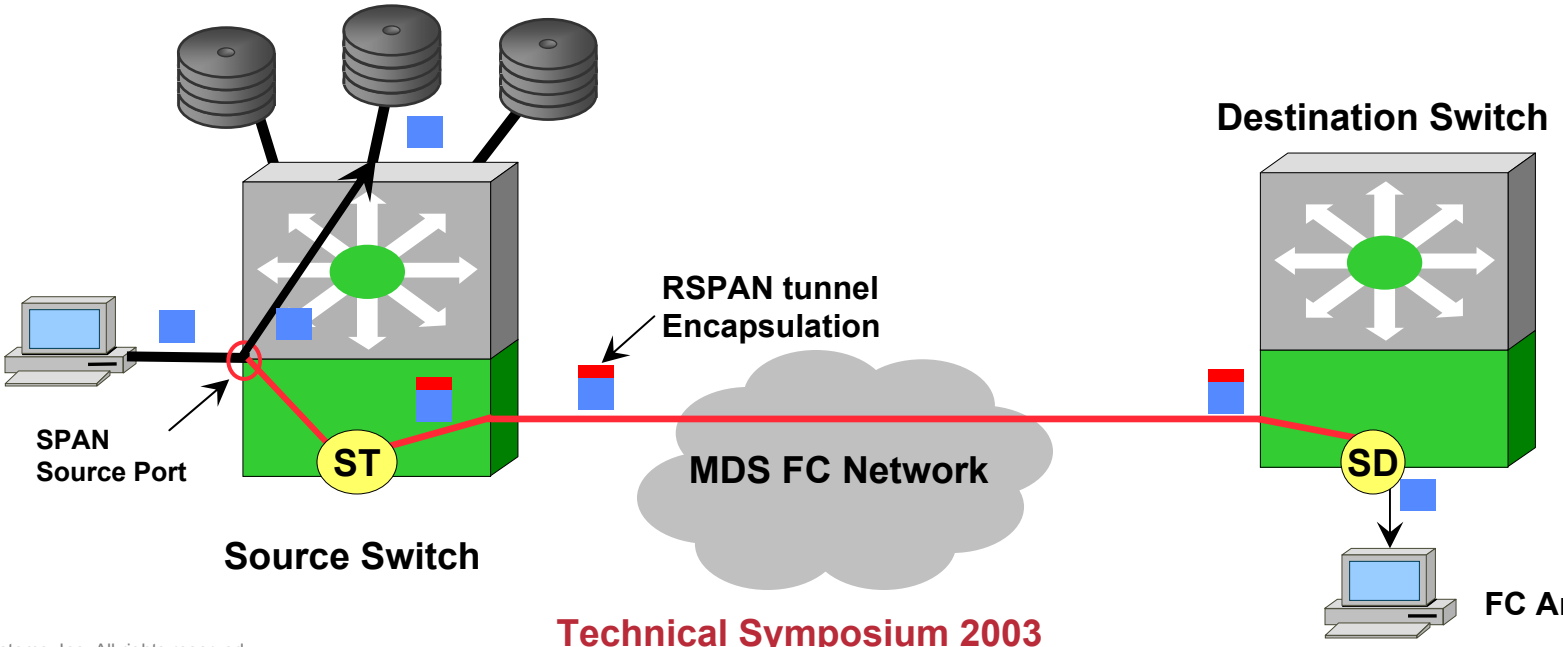
Enterprise Fabric Manager (EFM)



- **Server-Based, Centralized Management Server**
  - **Continuous health & event monitoring**
  - **Background discovery**
  - **Multiple, remote client access**
- **Historic performance monitoring for hot spot analysis**
  - **Statistics for ISLs, Ports PortChannels, Route Flows**
  - **SCSI statistics**

Remote SPAN (RSPAN)

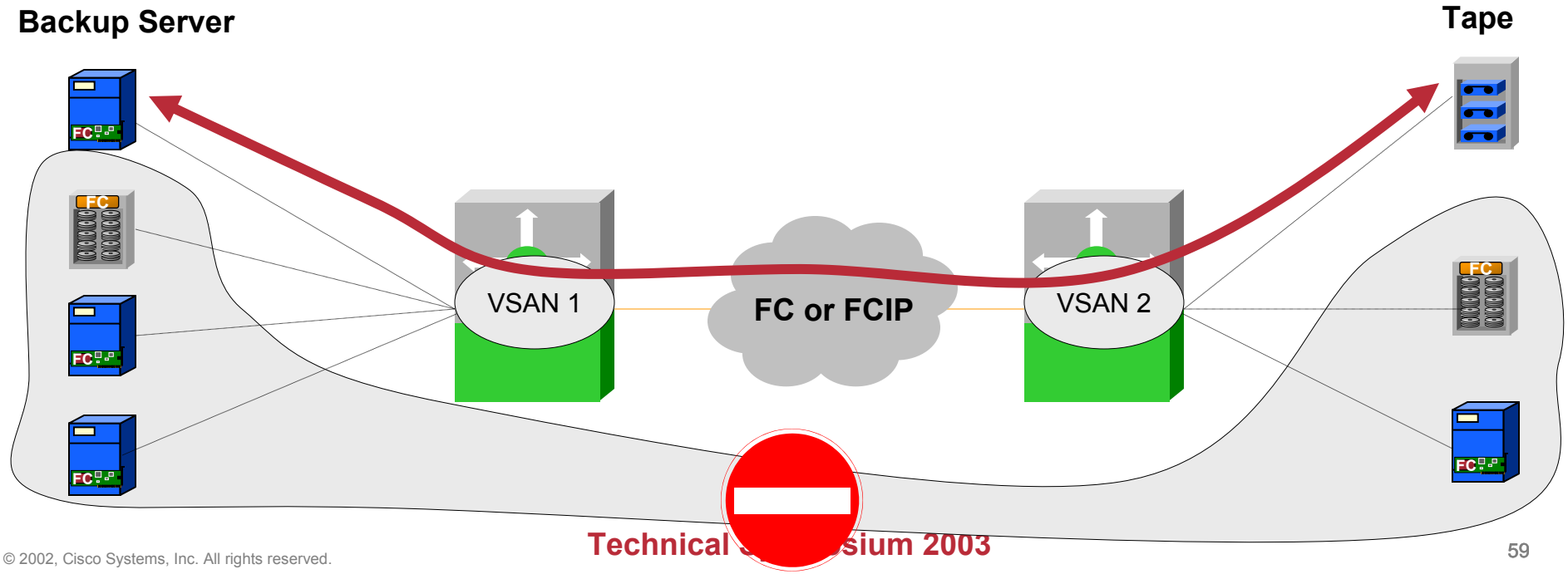
- Non-intrusively monitor FC ports at a remotely located switch using a SD port
- RSPAN traffic is tunneled through network using FC-tunnels
- RSPAN traffic is compatible with off-the-shelf FC analyzers and MDS 9000 port adapters



Inter-VSAN Routing (IVR):

What is Inter-VSAN Routing (IVR)?

- **Preserve the benefits of VSANs yet selectively allow traffic between VSANs**
- **Works for both FC and FCIP links**



Why Inter-VSAN Routing (IVR)?

- **Reason # 1: Enable customers to leverage expensive centralized storage services such as tape libraries and disks**
- **Reason #2: Address deployment issues related to geographically disparate fabrics**

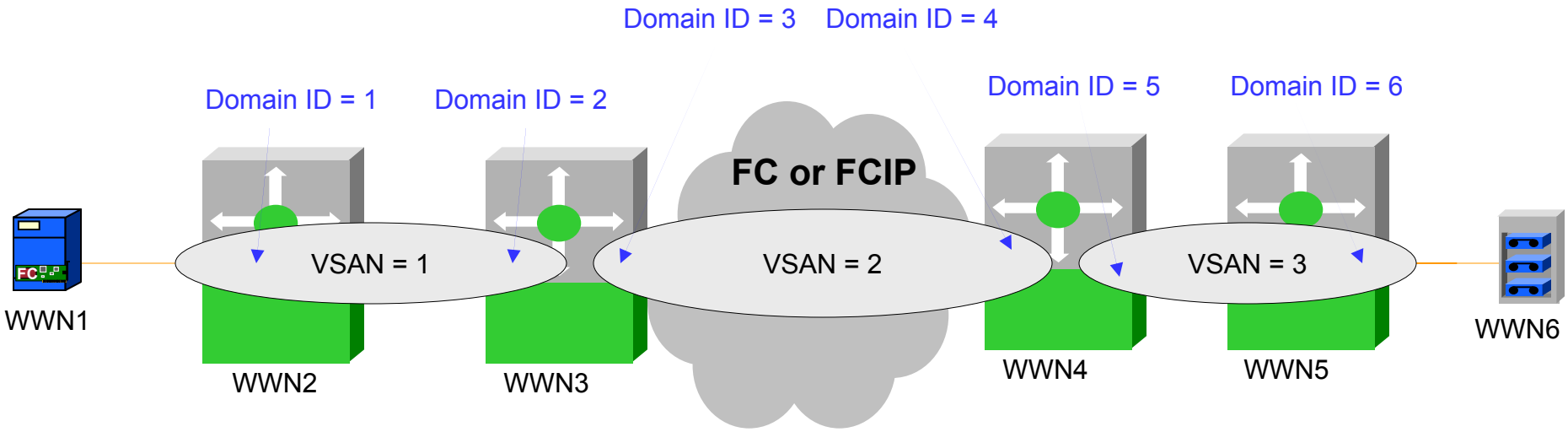
  Operational challenges because each location is managed autonomously by separate storage administrators. Change in a remote location should not impact the local administrator

  Minimize the impact of change in fabric services (Zoning, Name Server, Domain ID Assignment etc.) in one location on all other locations

  Minimize the impact of link state changes(FSPF) in one location on all other locations

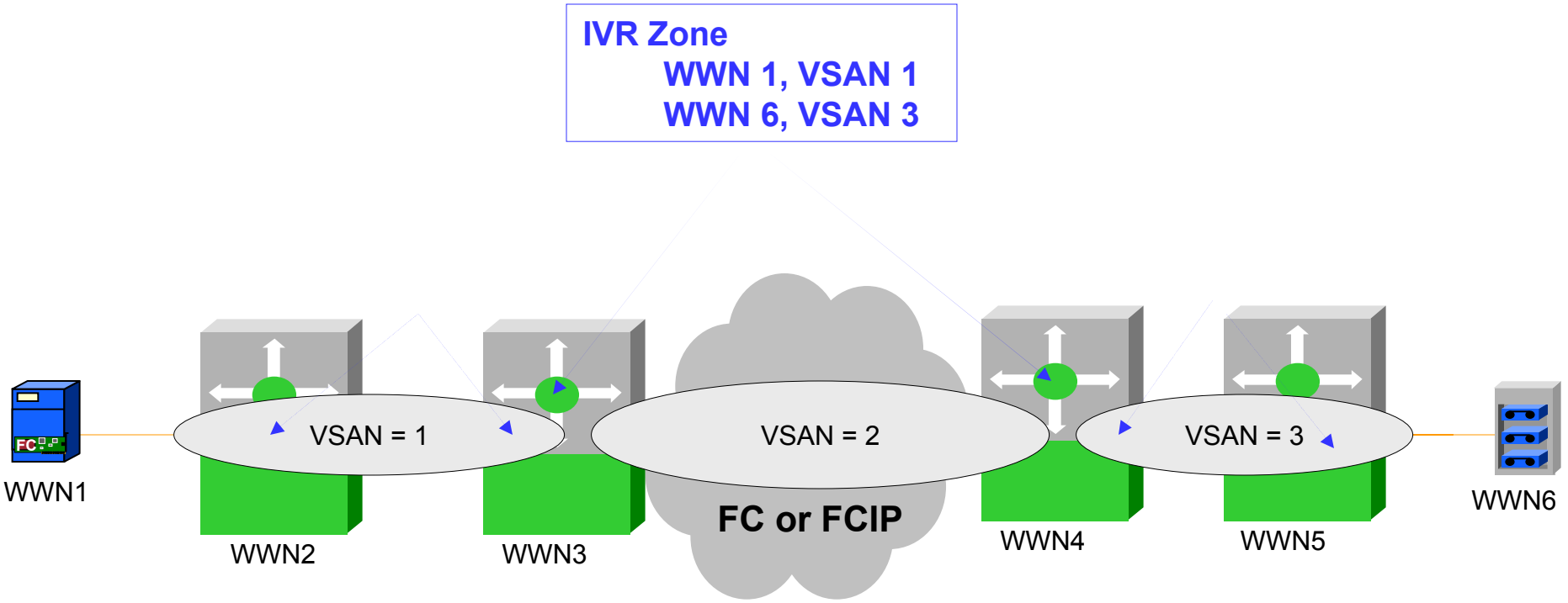
Assumptions

- **Domain IDs must be unique for each: {Switch: VSAN}. This ensures that FC-IDs are unique in the entire fabric.**



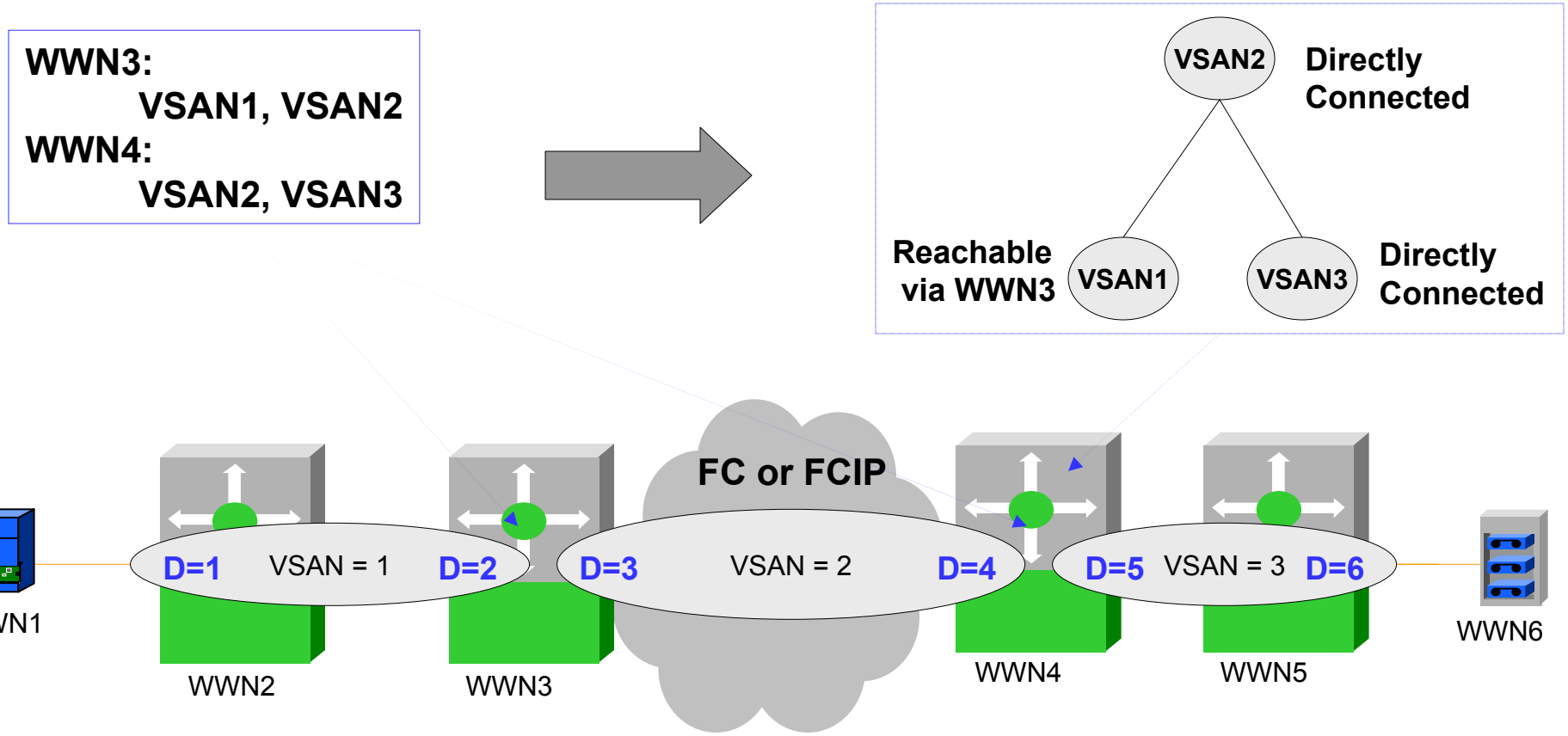
Step #1: Configure IVR Zone

- Create a IVR Zone that spans across VSANs.
- Zone membership is limited to devices of interest



Step #2: Configure Reachability

- Configure and activate topology on each Gateway Switch



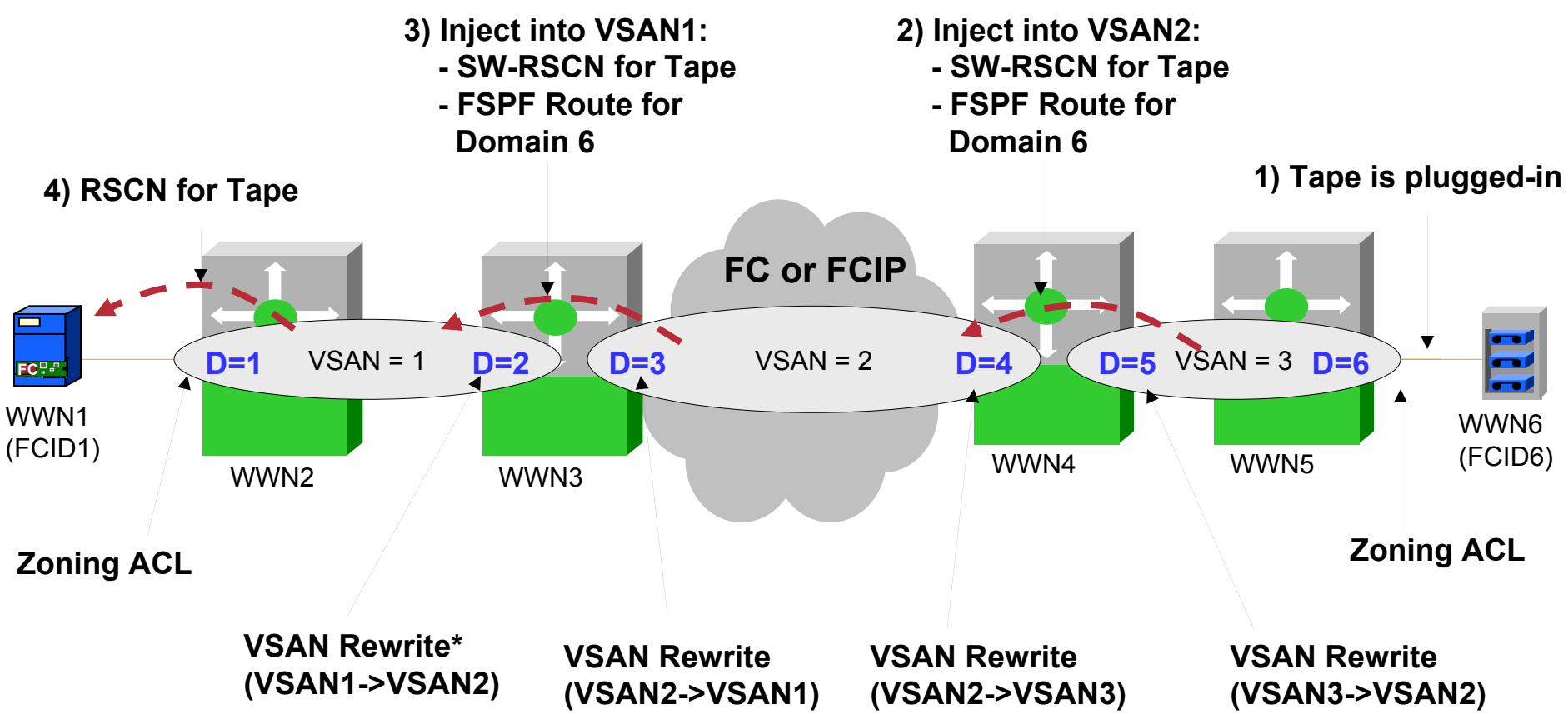
Step #3: Plug in Host (Assume no Tape for Now)

- First, Host sends FLOGI to the switch and gets a FCID1 in VSAN1
- Next, Host sends PLOGI to the switch. It does not find any Target

  This is because the Tape is not plugged in. As a result, it does not appear in Name Server of VSAN1

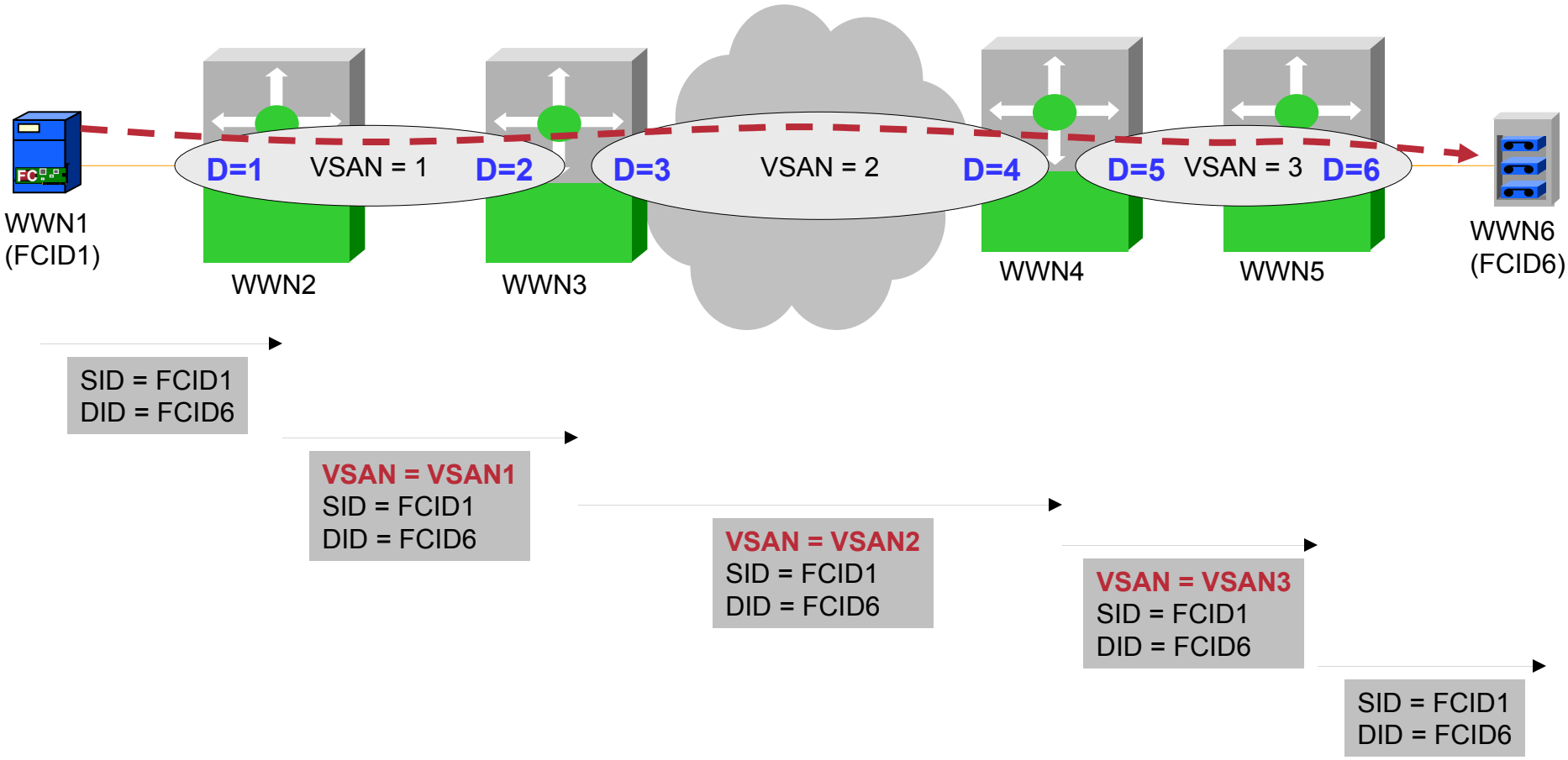
- Host (WWN and FCID) is propagated to Name Server in VSAN2 and VSAN3

Step #4: Now Plug in Tape



** VSAN Rewrite are installed on ingress ports only.*

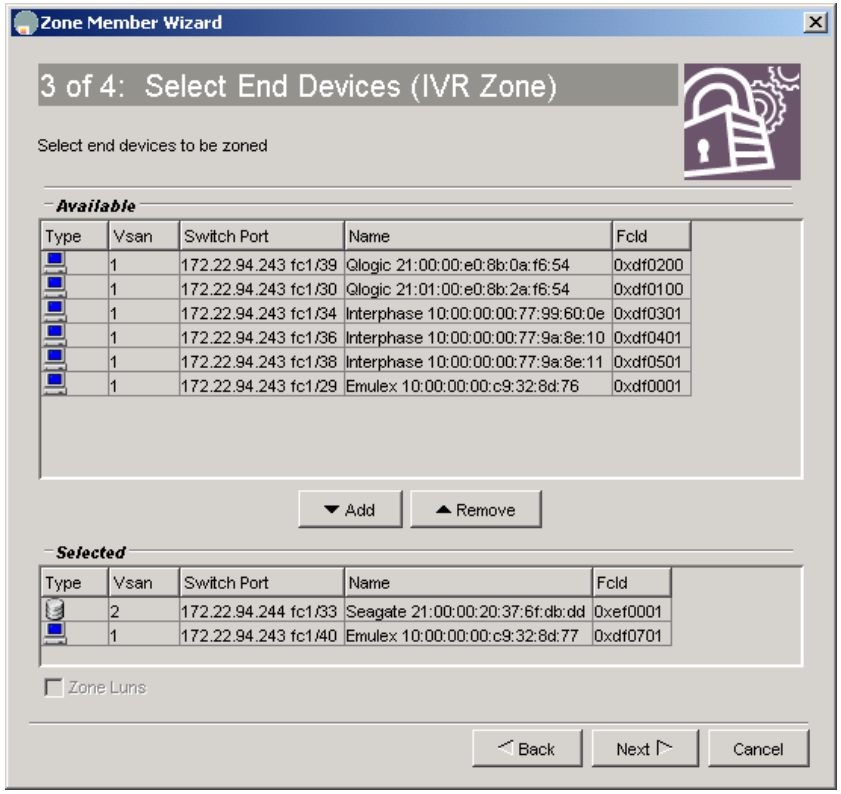
Step #5: Inter-VSAN Frame Switching



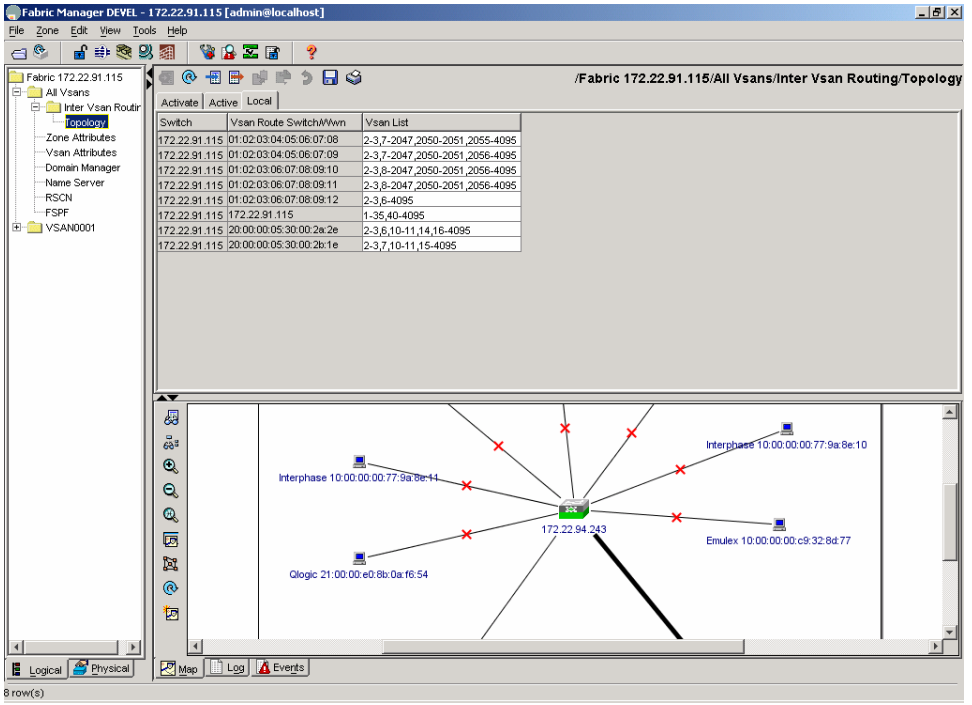
Fabric Manager for IVR Management

Cisco.com

IVR Zone Mgmt



IVR Topology Mgmt



FCIP & iSCSI

IP Protocol Encapsulation

- IP SANs carry block I/O traffic on top of IP

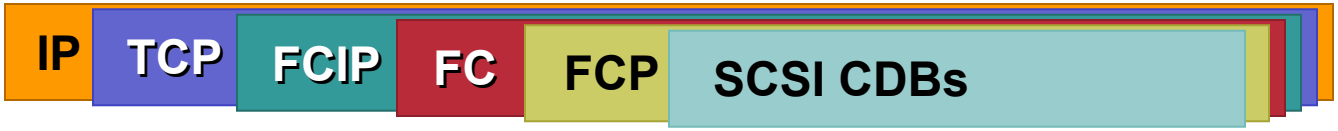
  Leverage Gigabit Ethernet performance for local traffic

  Use TCP: A reliable transport for delivery in MAN/WANs

- Two primary protocols:

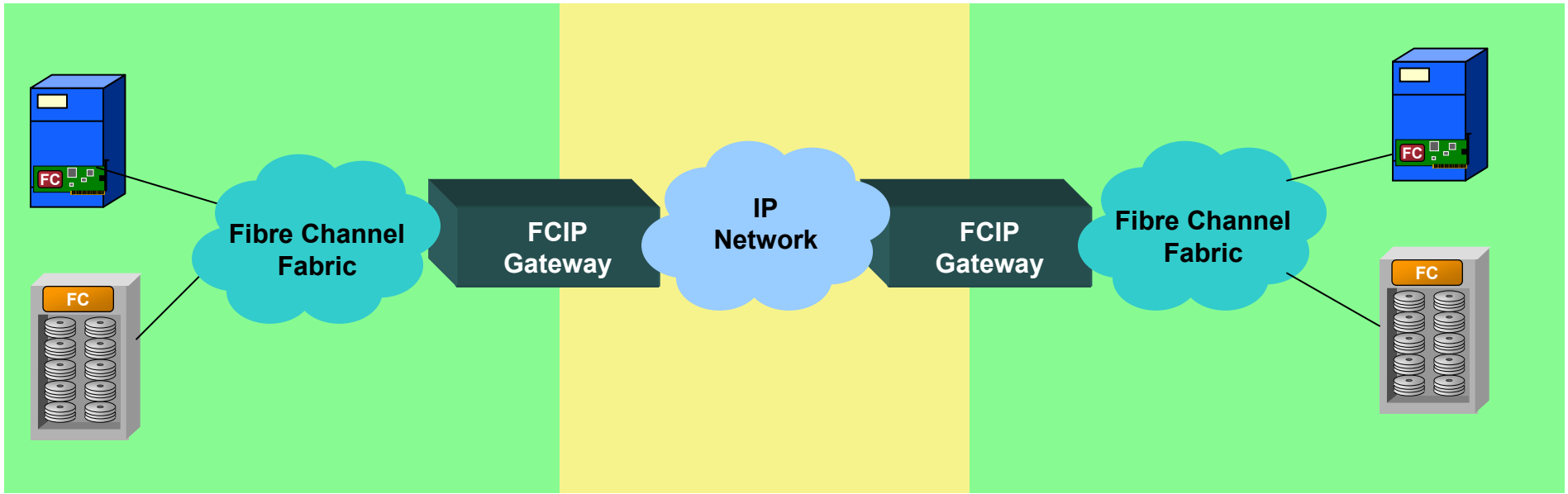
  iSCSI—"IP-SCSI" IP-native transport of SCSI CDBs and data within TCP/IP connections

  FCIP—"Fibre-Channel-over-IP"— Tunneling of Fibre Channel frames within TCP/IP connections, including FC fabric management frames



What is FCIP (Fibre Channel over IP)?

IT creates one logical fabric between remote SANS, and the switches think they are connected. IP is only used for tunneling through the WAN.

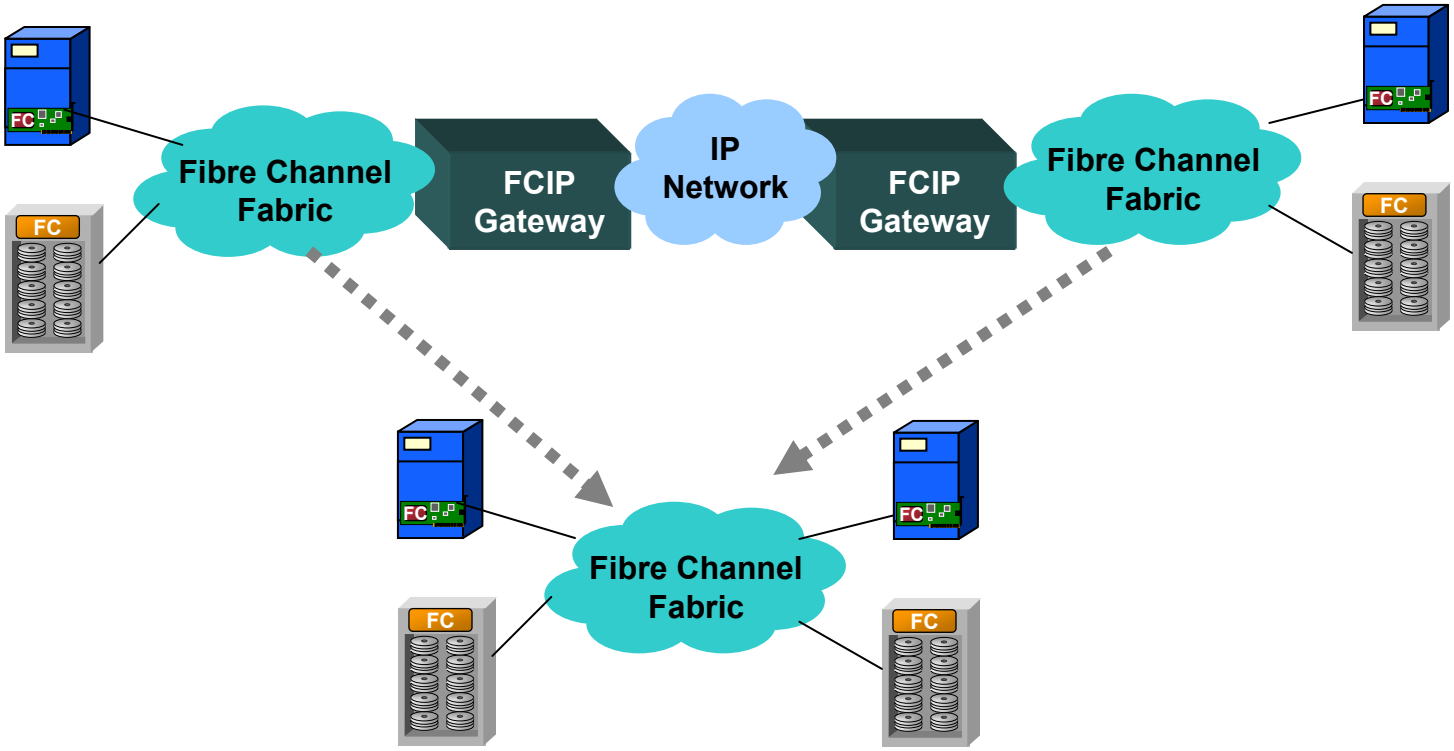


What is FCIP (Fibre Channel over IP) cont…

Remote FC resources are viewed as local

FCIP creates a Virtual FC Inter-Switch Link (ISL)

Fabric service information is extended across the FCIP ISLs

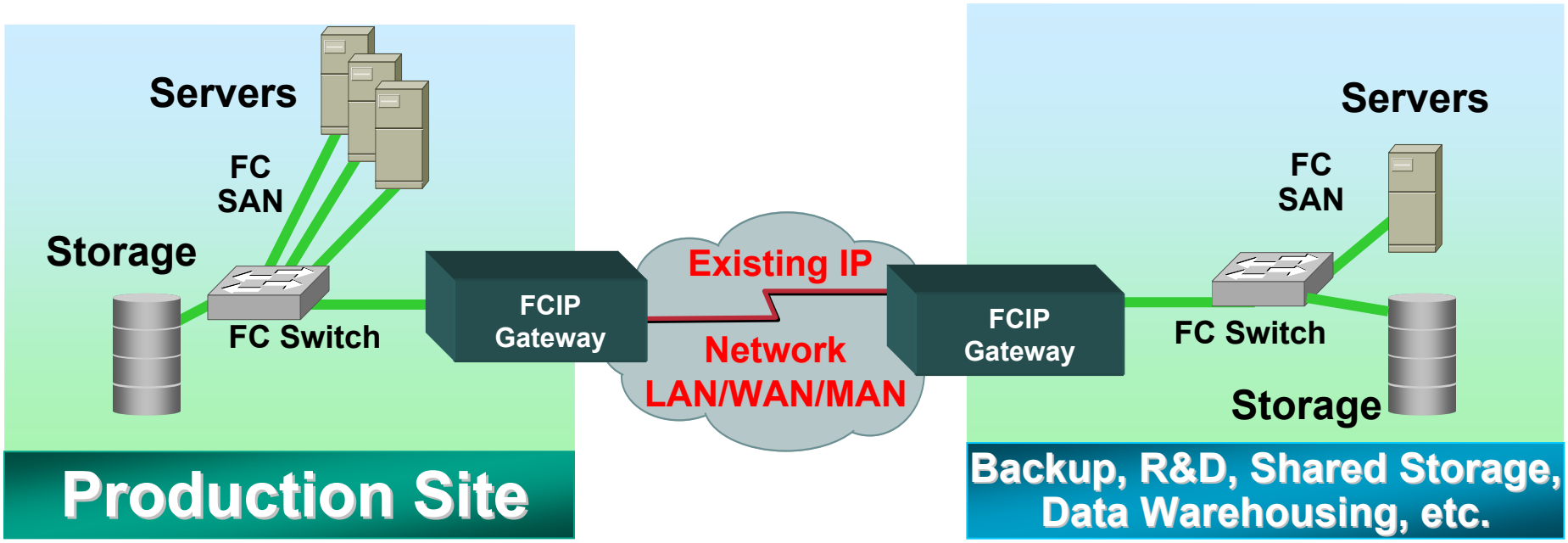


An FCIP Application Topology

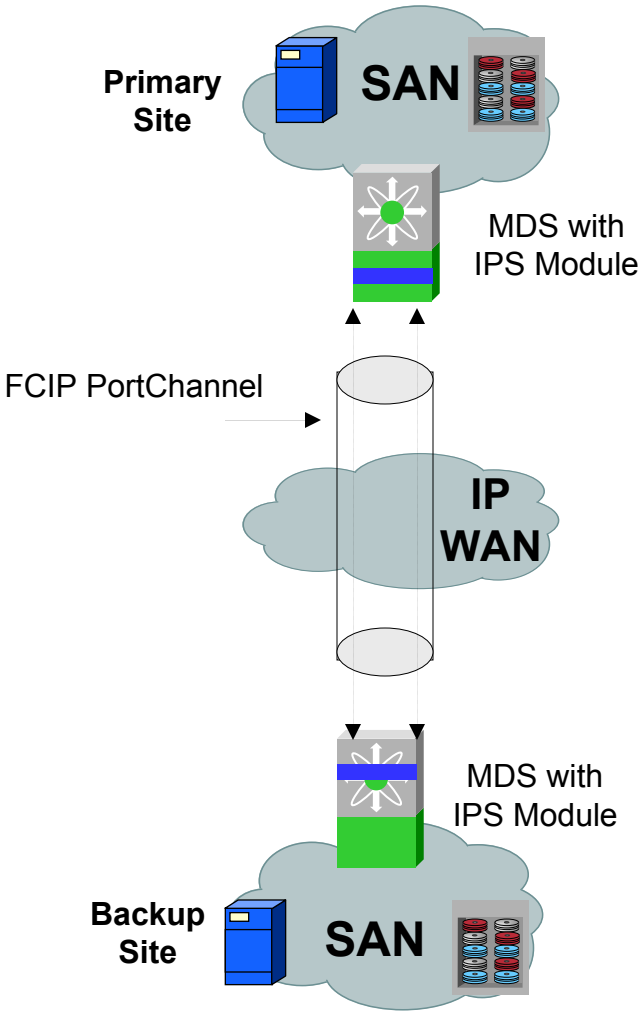
FCIP Gateways perform Fibre Channel encapsulation process into IP Packets and reverse that process at the other end

FC Switches connect to the FCIP gateways through an E_Port for SAN fabric extension to remote location

A tunnel connection is set up through the existing IP network routers and switches across LAN/WAN/MAN

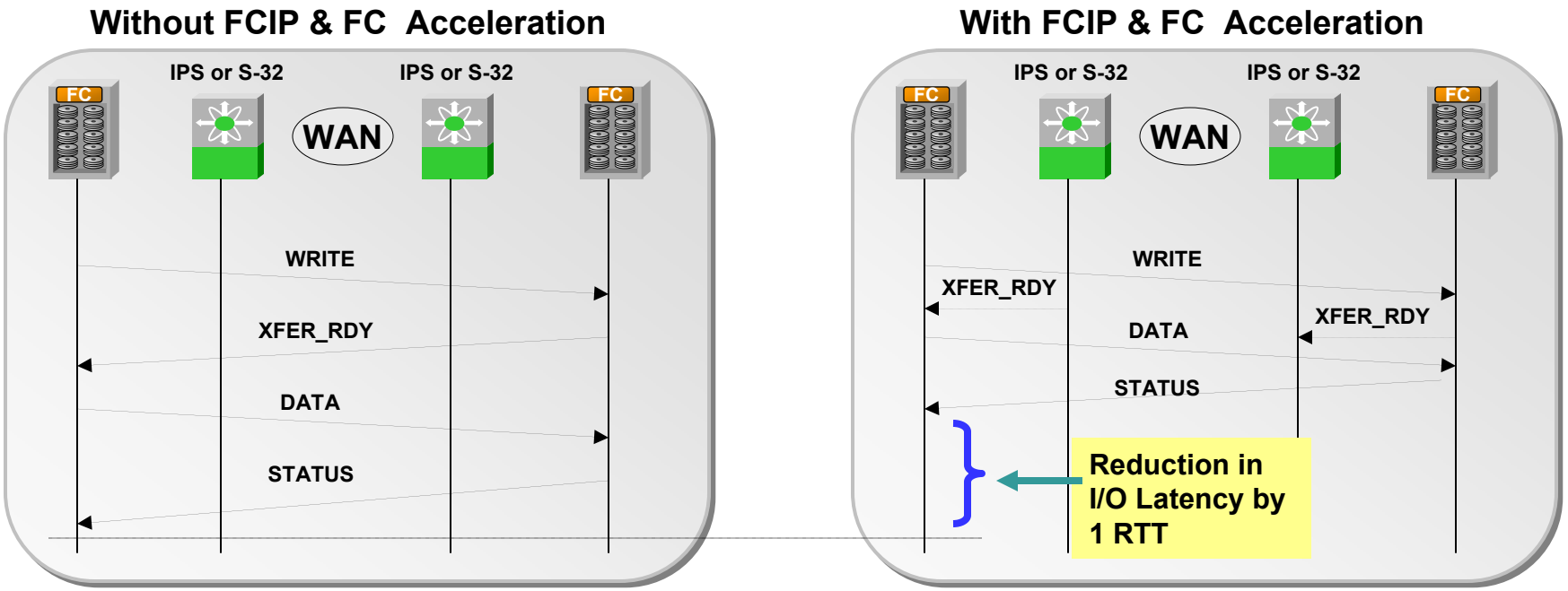


FCIP Compression



- **FCIP IP Payload Compression**
  - **LZS (Lempel-Zif-Stac) compression algorithm**
  - **Throughput depends on type of app and compression ratio**
- **Each IPS port has compression capability**
  - **Throughput can be shared between 3 FCIP tunnels**
- **FCIP PortChannels for:**
  - **Bandwidth**
  - **Resiency**

FCIP & FC Write Acceleration



- Improves response time for the storage app
- Extended distance for DR and BC apps

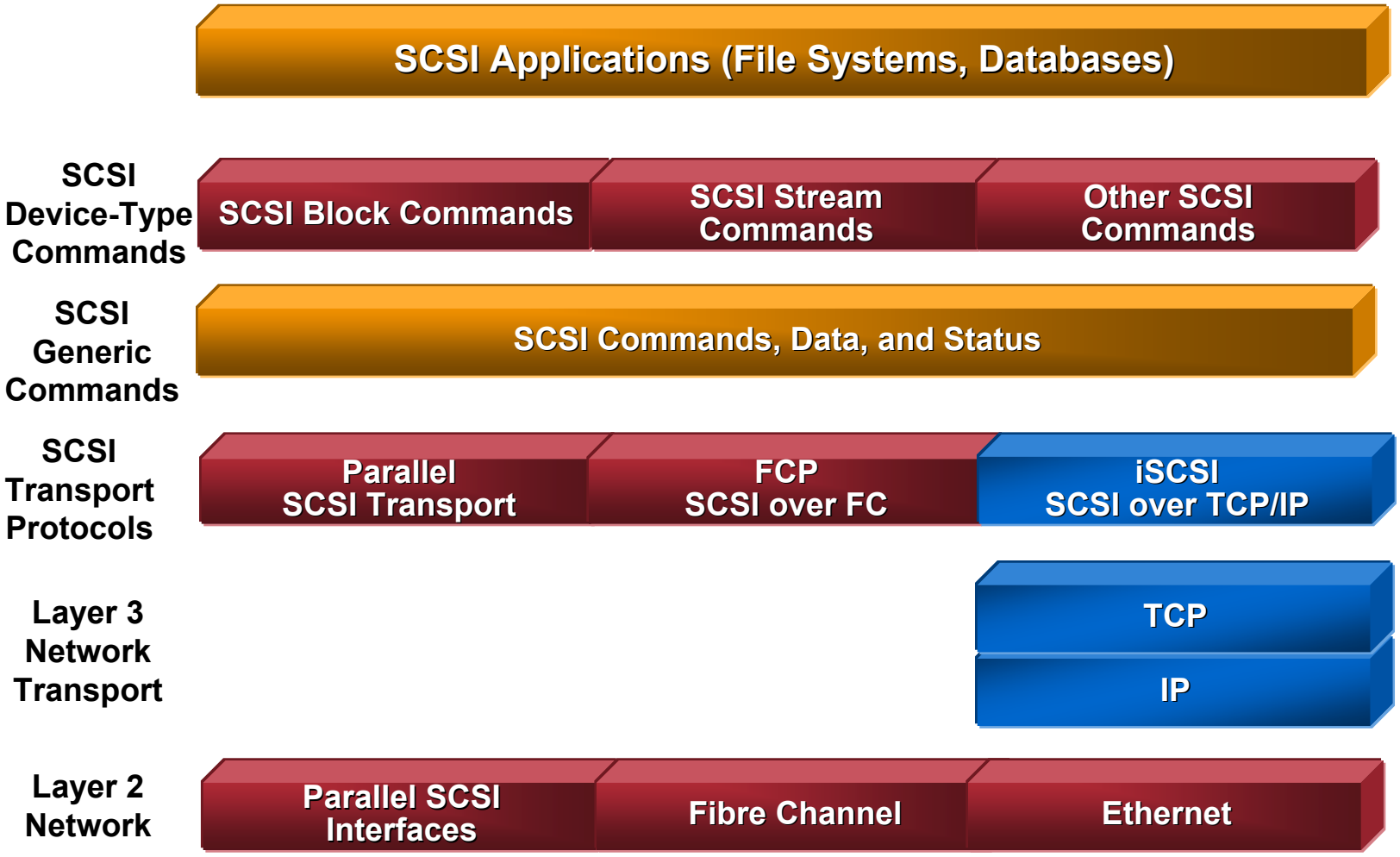
What is iSCSI?

- A way to access storage across an IP network as though it was locally attached.
- Transports SCSI protocol commands and data across an IP network
- Standards status update

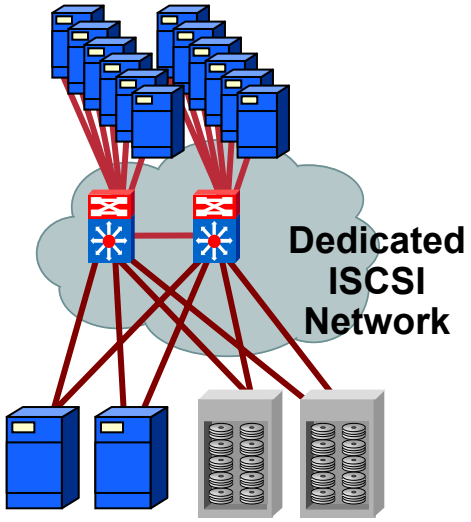
  Ratified Standard (IETF)

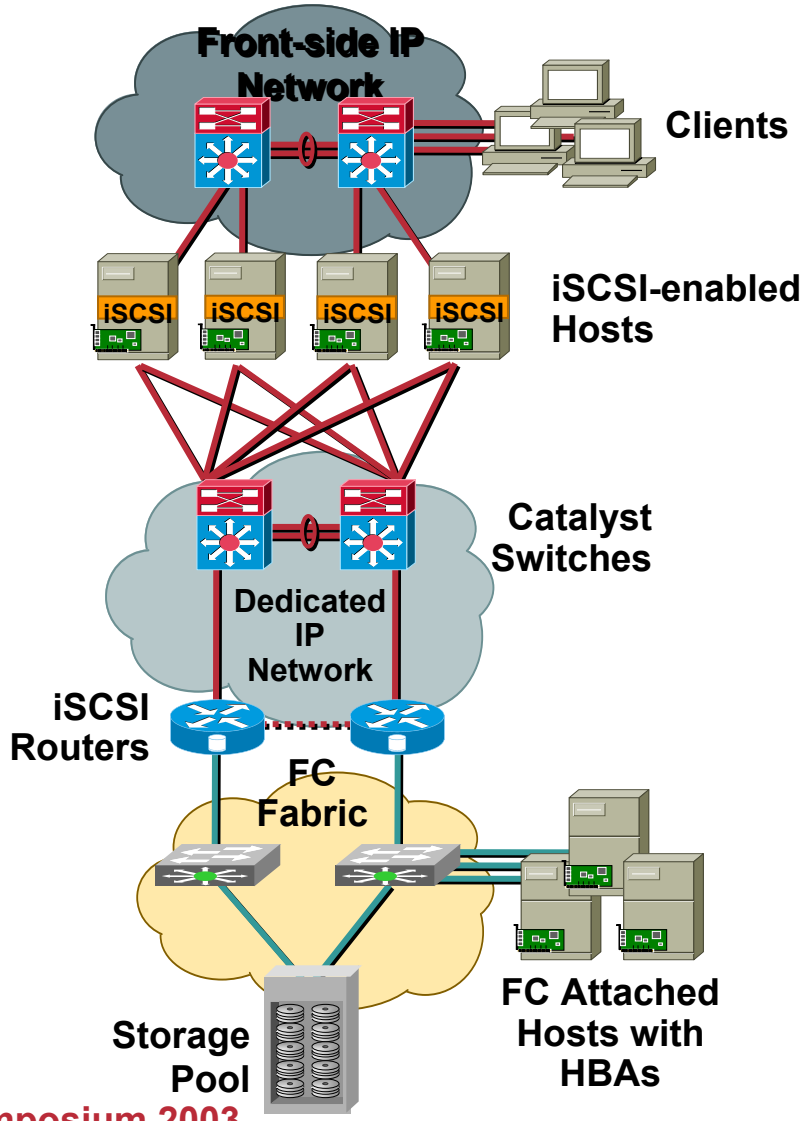
  Major industry support (Cisco, IBM, EMC, Microsoft, Intel..)

iSCSI Architectural Model



IP Storage Networks

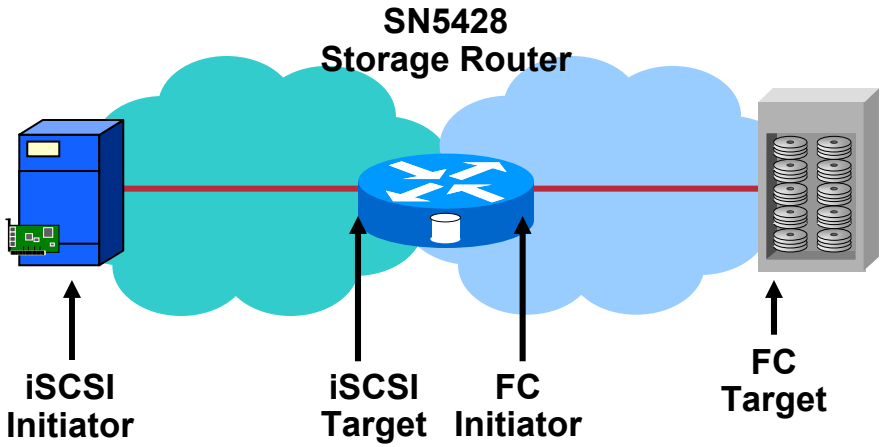




Initiator and Target Model for iSCSI

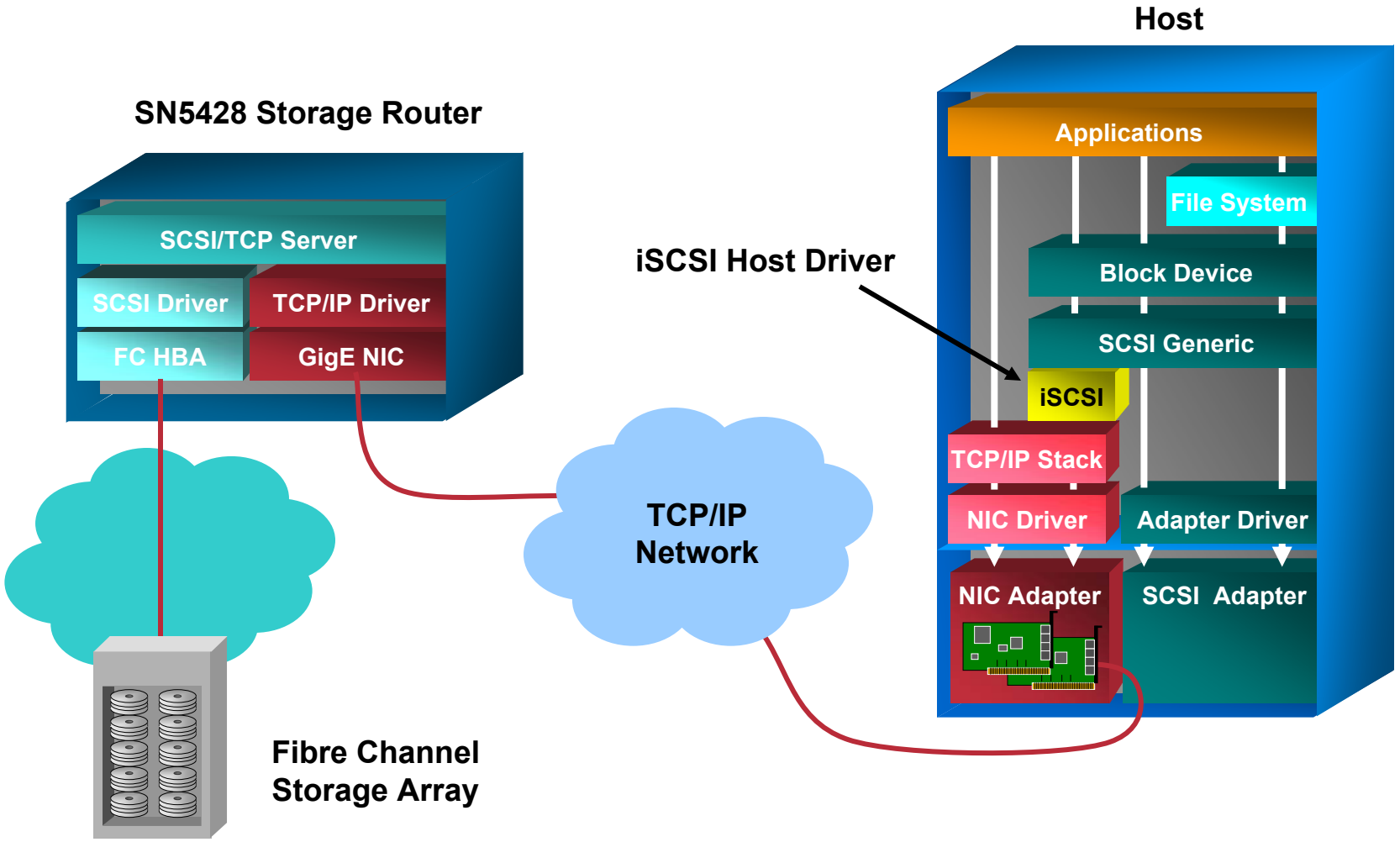
Initiator - SCSI device which is capable of originating SCSI commands and task management requests

Target - SCSI device which is capable of executing SCSI commands and task management requests



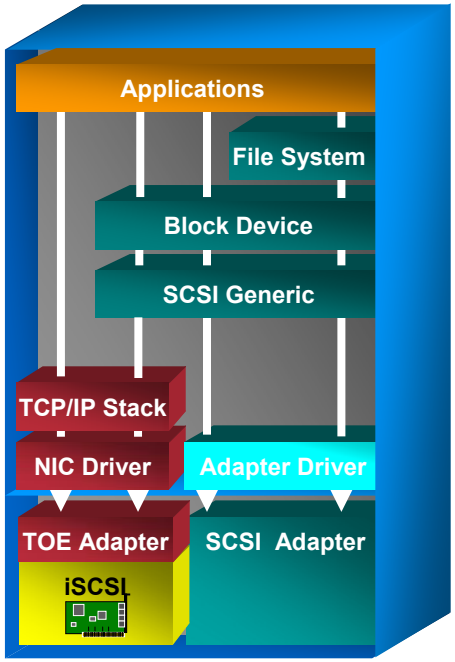
Initiators and Targets in a Multi-Protocol SAN

Cisco iSCSI Solution Architecture

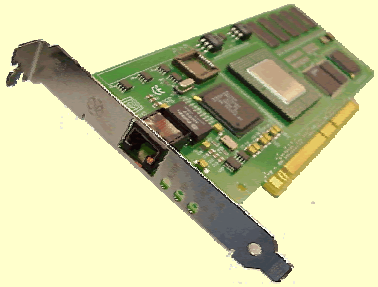


iSCSI + TCP Offload Engine (TOE)

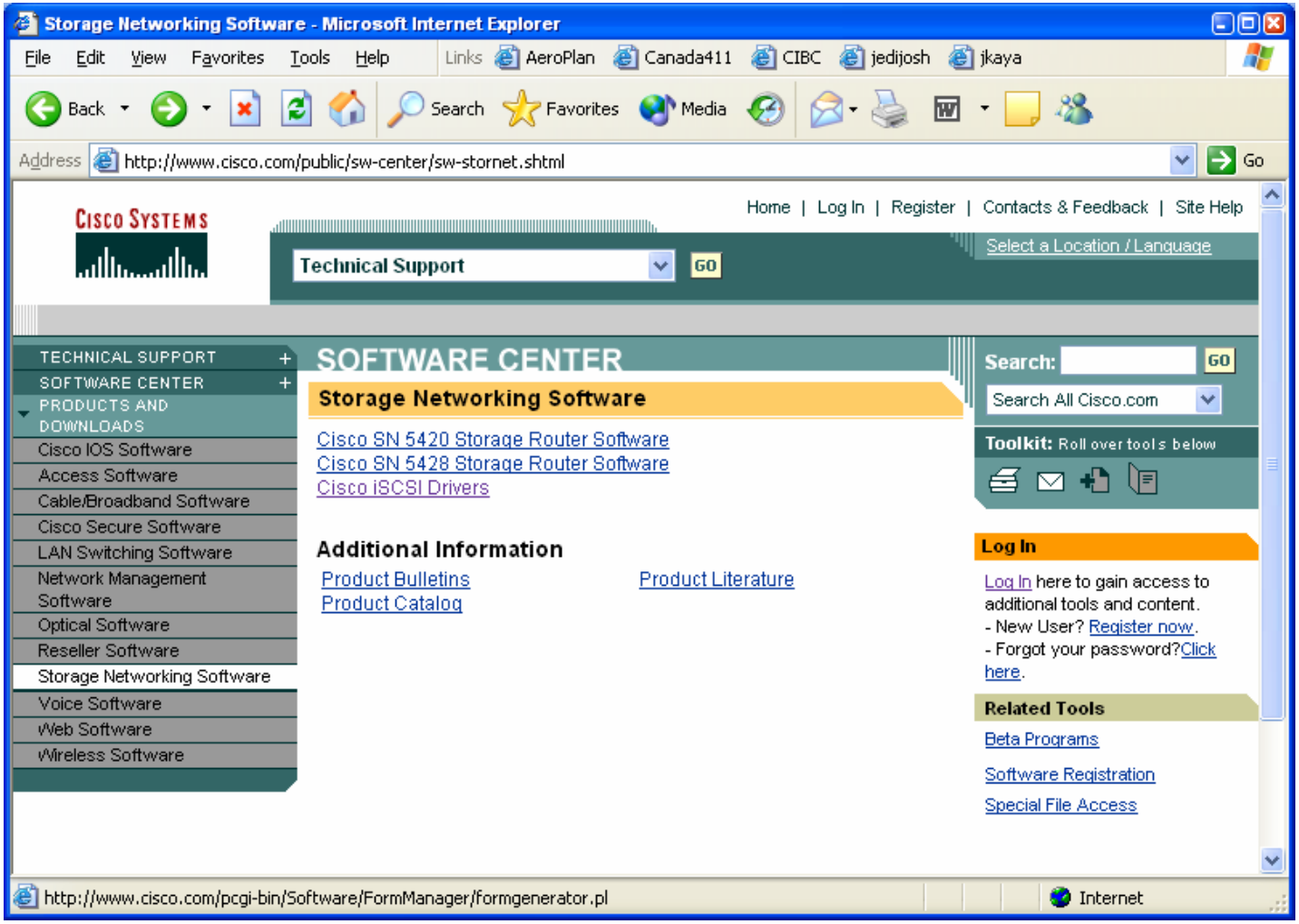
- Hardware implementation of iSCSI and TCP/IP on a NIC
- Relieves host CPU from iSCSI and TCP processing, reducing CPU processing
- Two forms of TOEs
  - Offloads storage traffic only
  - Offloads TCP/IP for storage and data traffic
- Wire-rate iSCSI performance



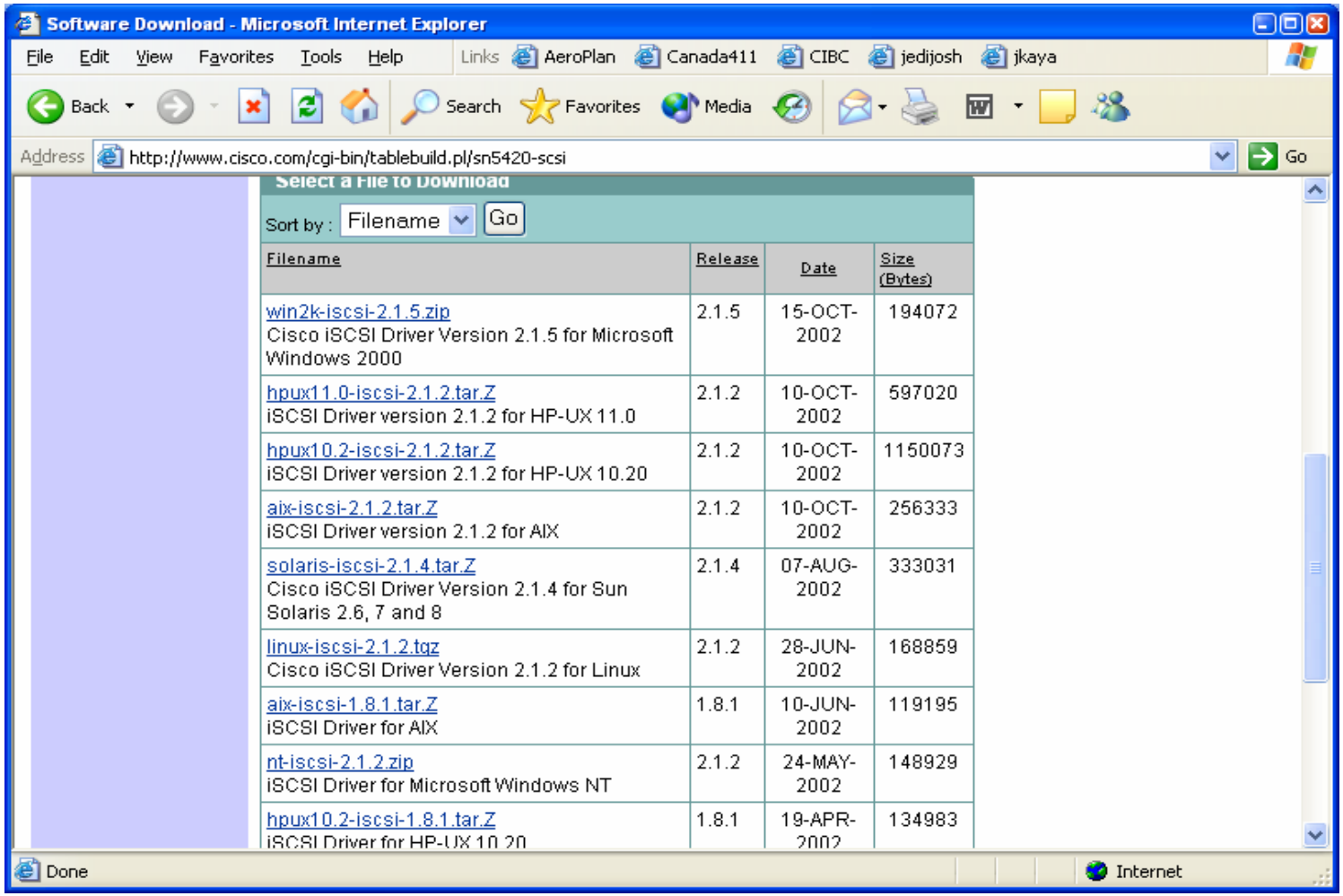
- Alacritech
- Adaptec
- Emulex
- Intel
- Qlogic



Cisco iSCSI Drivers

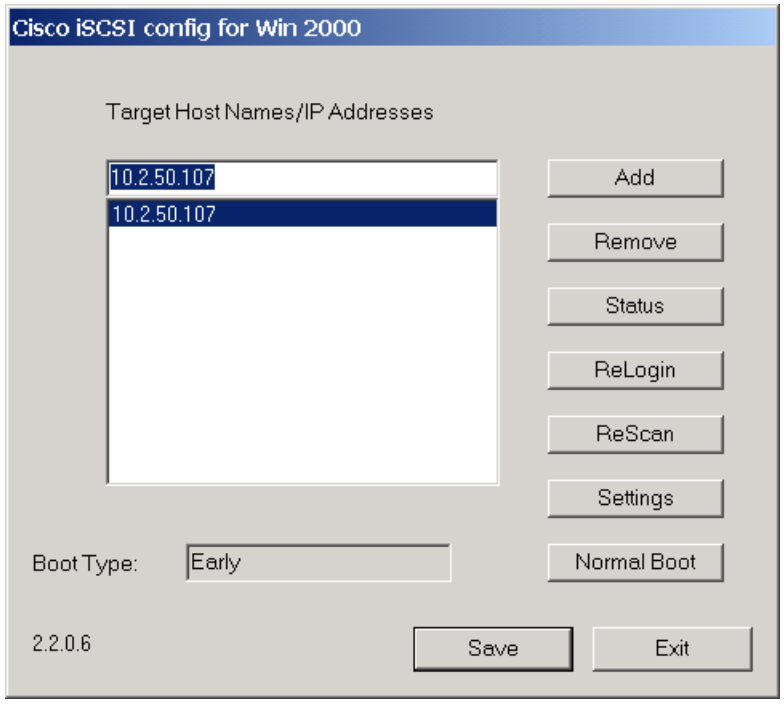


Cisco iSCSI Driver: Download



Filename	Release	Date	Size (Bytes)
win2k-iscsi-2.1.5.zip Cisco iSCSI Driver Version 2.1.5 for Microsoft Windows 2000	2.1.5	15-OCT-2002	194072
hpux11.0-iscsi-2.1.2.tar.Z iSCSI Driver version 2.1.2 for HP-UX 11.0	2.1.2	10-OCT-2002	597020
hpux10.2-iscsi-2.1.2.tar.Z iSCSI Driver version 2.1.2 for HP-UX 10.20	2.1.2	10-OCT-2002	1150073
aix-iscsi-2.1.2.tar.Z iSCSI Driver version 2.1.2 for AIX	2.1.2	10-OCT-2002	256333
solaris-iscsi-2.1.4.tar.Z Cisco iSCSI Driver Version 2.1.4 for Sun Solaris 2.6, 7 and 8	2.1.4	07-AUG-2002	333031
linux-iscsi-2.1.2.tgz Cisco iSCSI Driver Version 2.1.2 for Linux	2.1.2	28-JUN-2002	168859
aix-iscsi-1.8.1.tar.Z iSCSI Driver for AIX	1.8.1	10-JUN-2002	119195
nt-iscsi-2.1.2.zip iSCSI Driver for Microsoft Windows NT	2.1.2	24-MAY-2002	148929
hpux10.2-iscsi-1.8.1.tar.Z iSCSI Driver for HP-UX 10.20	1.8.1	19-APR-2002	134983

Configuration Applet

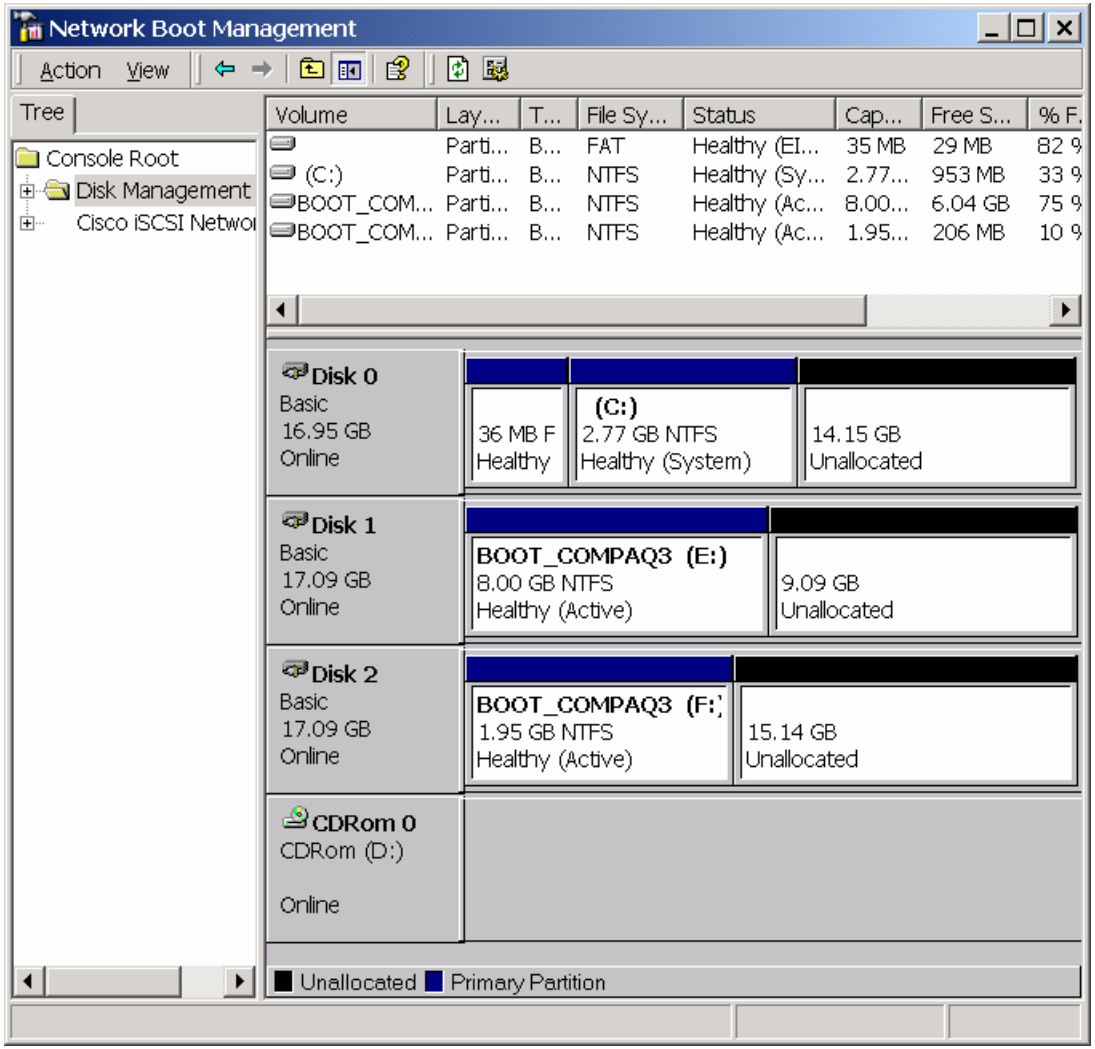


iSCSI Configuration Applet

iSCSI Connection Status

- T:0 means TargetID 0
- Correlate TargetID with Disk Manager

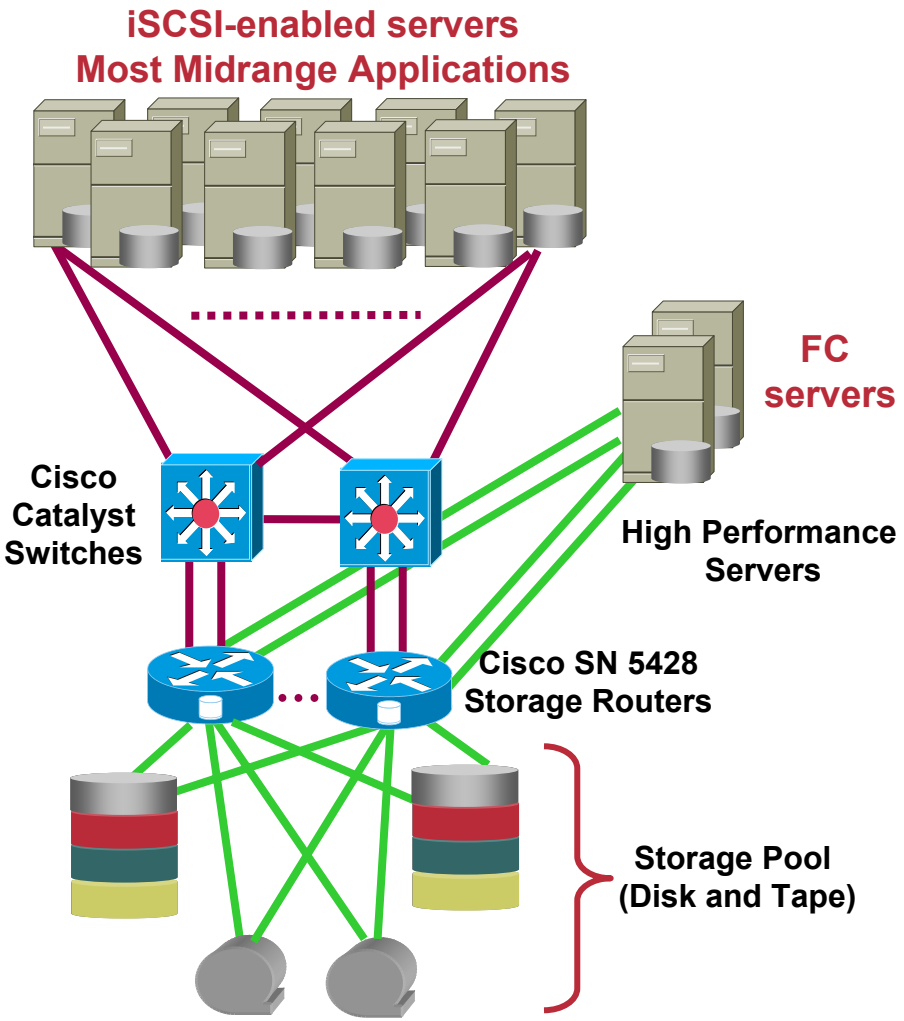
iSCSI Disks



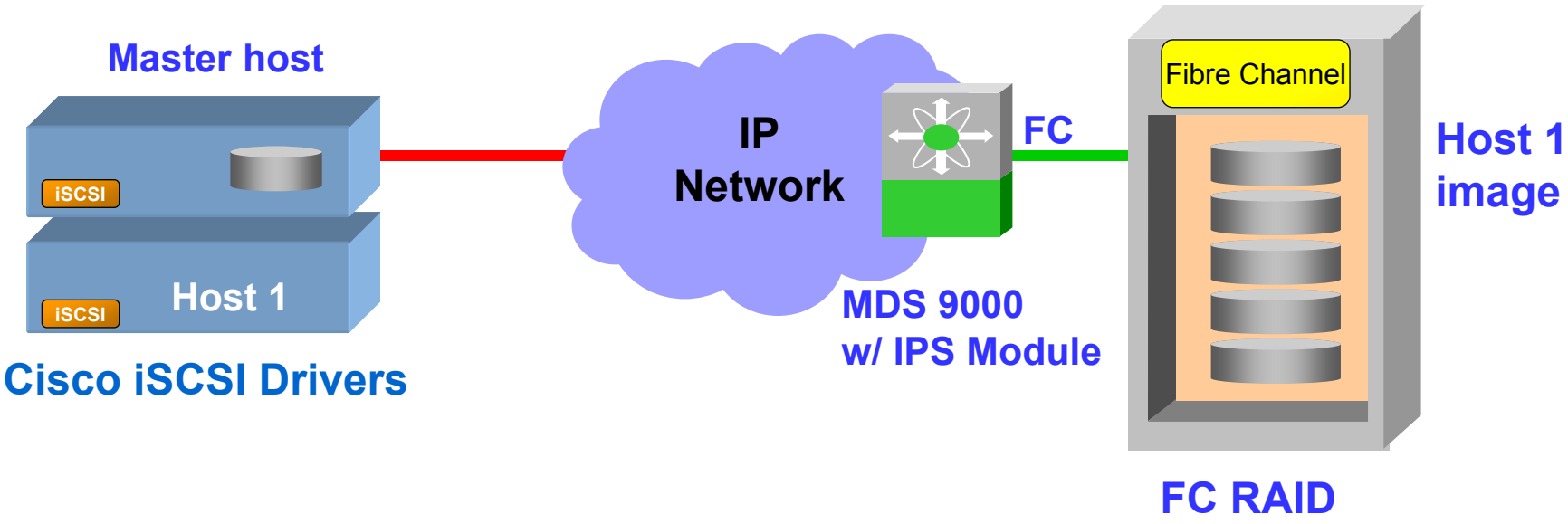
- **iSCSI disks appear in normal disk manager app**
- **Disk 0 is internal**
- **Disks 1 and 2 are iSCSI disks**

The Workgroup SAN Overview

- Workgroup SAN Environments
  - Enterprise Department
  - Small-Medium Business
- 10 - 40 Servers with Midrange Applications
  - Email – Exchange, Notes
  - Database – SQL, Oracle
  - Financials – Great Plains, Lawson, Oracle
  - Web Servers – IIS
  - File Servers
  - Print Servers
  - Customer Developed
- Most servers cost under $10,000
- The network/system/storage admin are the same person

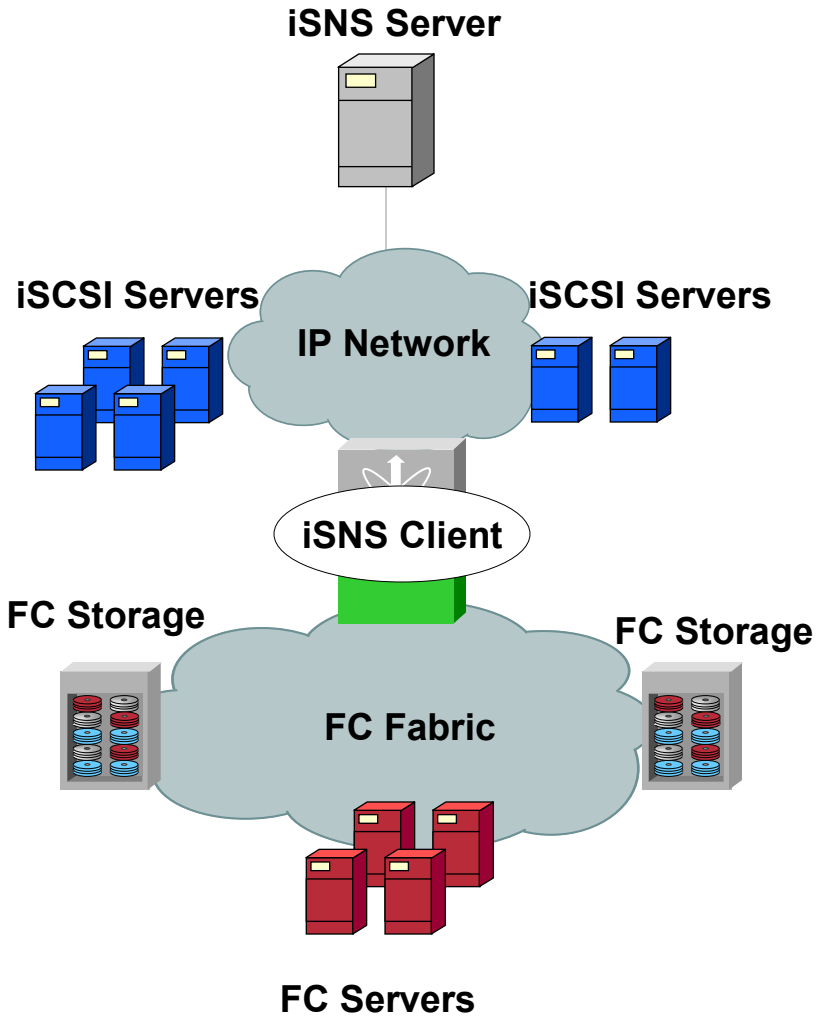


iSCSI Network Boot (iNBP)



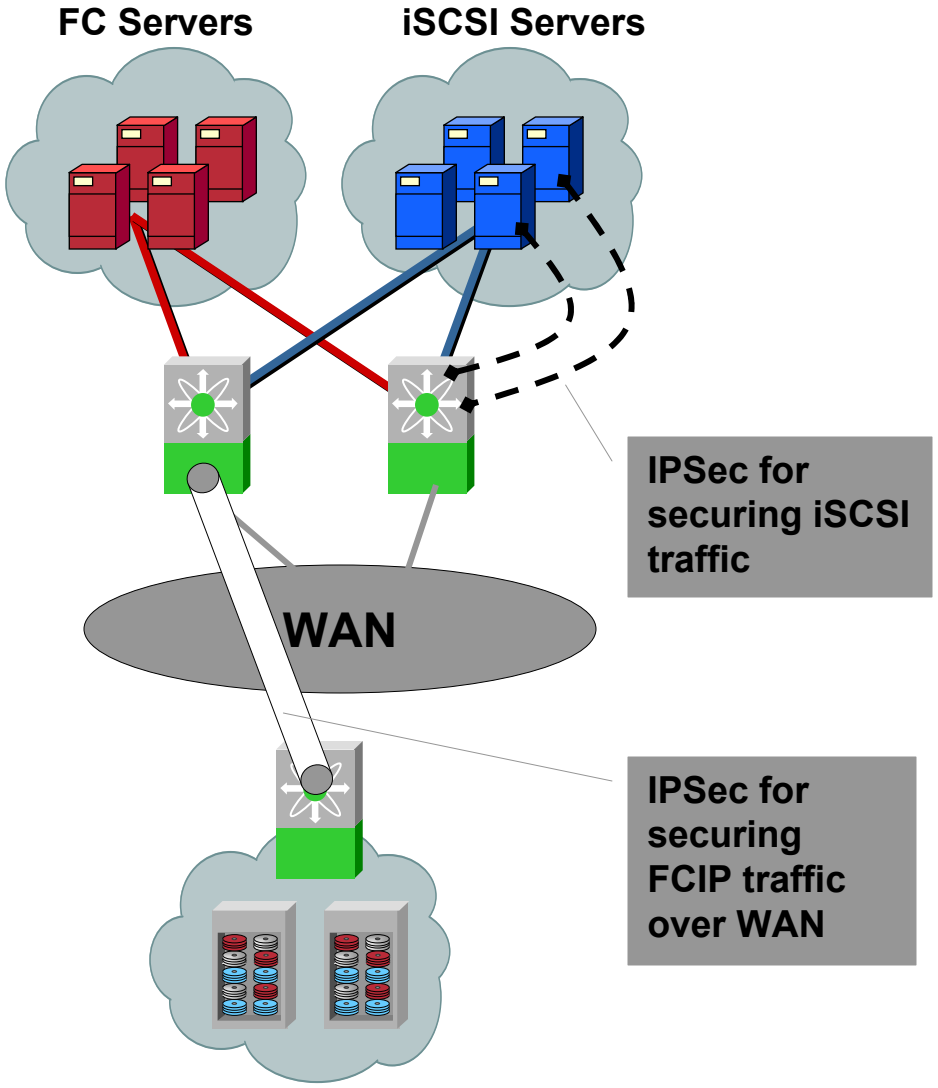
- ✓ **Boot diskless Windows/Linux Servers over iSCSI**
- ✓ **Cost-effective storage networking for Blade Servers**
- ✓ **No special hardware required. Leverages existing infrastructure**

Internet Storage Name Server (iSNS)



- **iSNS complements the existing target discovery mechanism**
- **iSNS is a directory-based protocol for management of IPS devices**
  - Device registration
  - Device discovery
  - Zoning (Discovery Domains)
  - State change notification

Security for IP Storage



- **Standards-based IPSec protocol for securing IP Storage**

  Authentication, data integrity and encryption for iSCSI and FCIP protocols

- **IPSec protocol is implemented on IPS Module**

  Each Gigabit Ethernet port provides security for both iSCSI and FCIP traffic

IP Storage Features (cont.)

- **IP Access Control List (ACLs)**

  Provides an additional level of security by allowing only trusted iSCSI Initiators and FCIP Gateways to connect to IPS Gateway

  Minimizes the impact of Denial of Service (DOS) attacks on IPS Gateways

- **Proxy-iSCSI Initiator Mode**

  Complements "Transparent Mode" by presenting all iSCSI Initiators to a storage subsystem as a single initiator

  Minimizes the configuration tasks required to enable large scale iSCSI deployment

Cisco iSCSI Network Boot

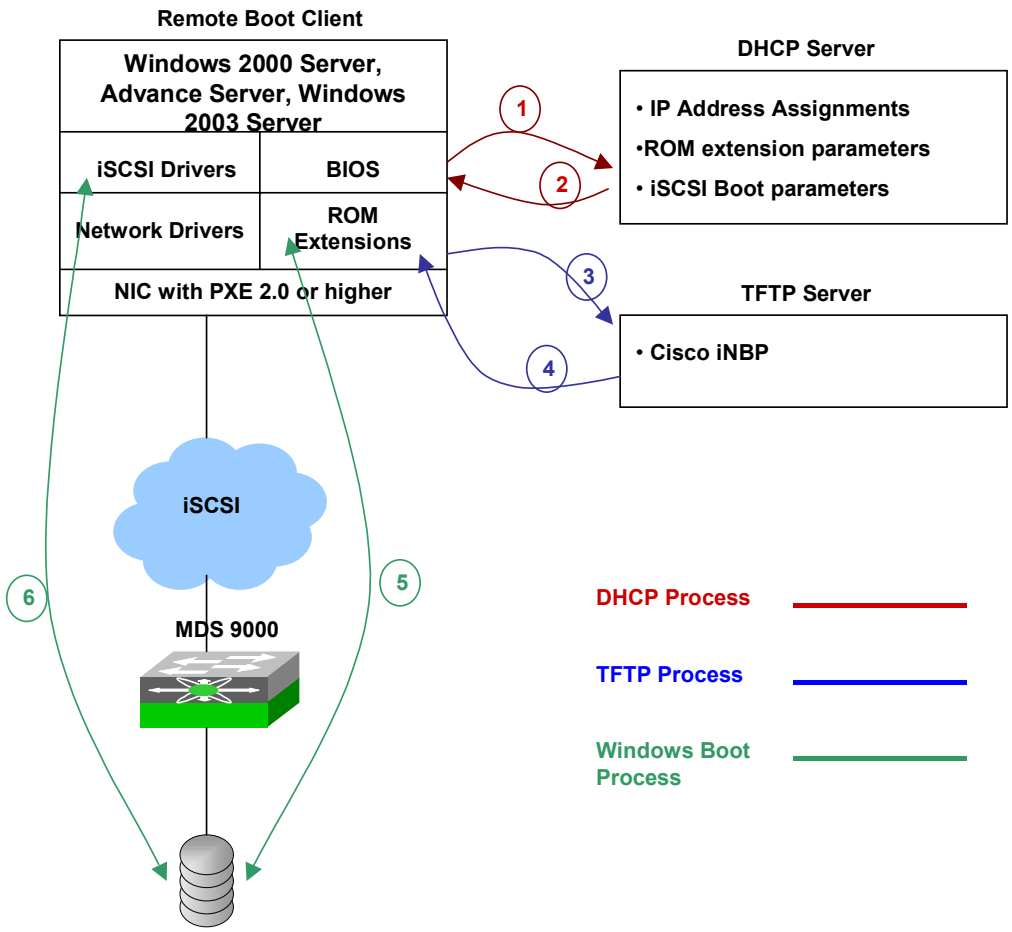
Cisco Network Boot

- Enables diskless Windows hosts to boot over iSCSI
- Compelling storage networking solution for blade and rack hosts
  - Cost-effective – Leverages existing infrastructure
  - Operationally simple – Re-deploy and/or replace hosts quickly

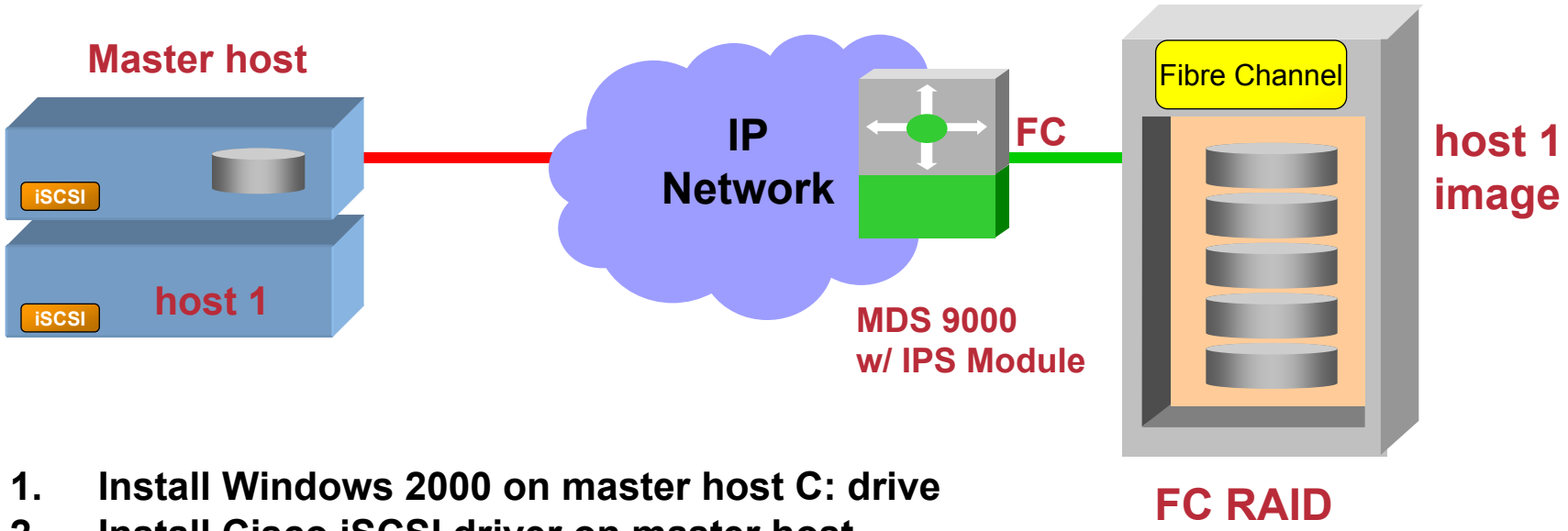
Network Boot Mechanics

Remote Boot Process

1) System BIOS does a DHCP request
2) DHCP Server returns:
   - System IP Address and Router Information
   - TFTP Server information and ROM extension filename
   - iSCSI boot information
3) BIOS downloads through TFTP Cisco's iNBP
4) Cisco's iNBP is downloaded into BIOS of the Remote Boot Client
5) BIOS connects to iSCSI device with a mini-iscsi driver (within iNBP) to load Windows
6) With "Early Boot" option for the normal iSCSI driver enabled, the mini-iSCSI driver (from iNBP) is released and normal iSCSI driver is loaded. Normal Windows operation and iSCSI Remote Access has completed.

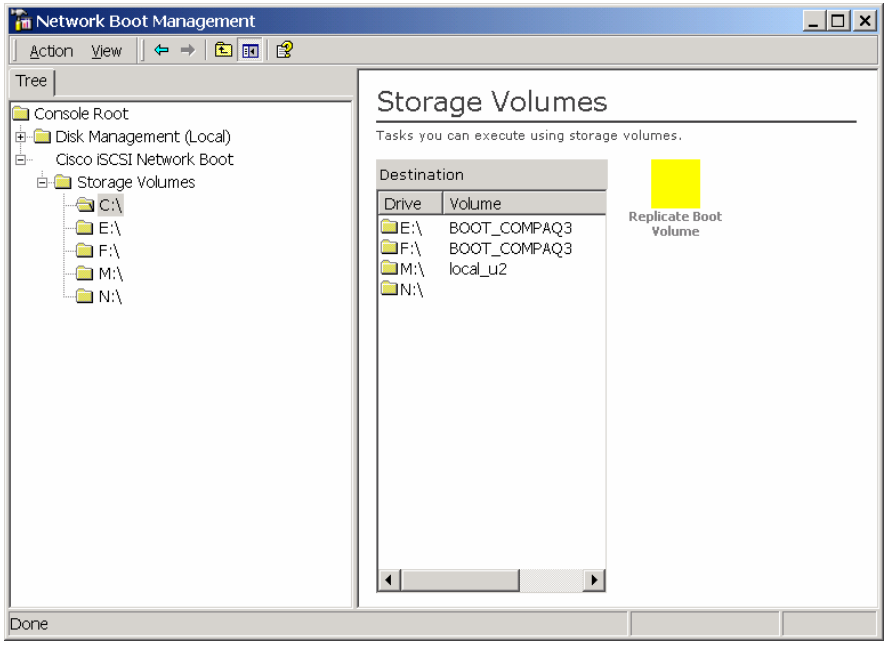


Installation & Setup



1. Install Windows 2000 on master host C: drive
2. Install Cisco iSCSI driver on master host
3. Configure & mount iSCSI target
4. Run Cisco copy utility which does the following:
   - Copy boot image from Master host to storage
   - Configure Windows for the new host
5. Set PXE boot option in host 1
6. Repeat for additional hosts

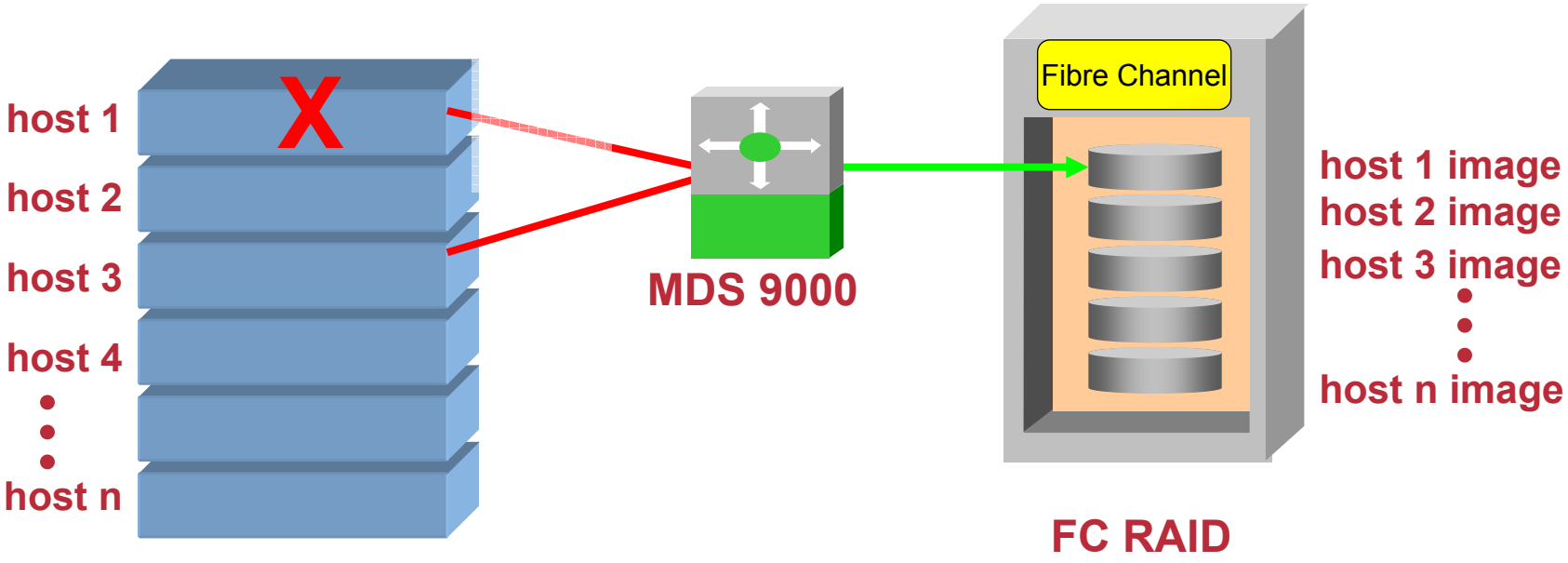
Cisco Copy Utility



- Runs as MMC Snap-in
- Copies DAS or iSCSI boot volumes
- File-level copy
- Changes IP address, host name, drive signature
  - New host can be booted without modifying IP address
  - Also forces new iSCSI initiator name to be created

Network Boot Operation - Failover

1. System fully operational
2. Failure on host
3. Connect new host
4. Point host 3 at host 1 Boot Image (DHCP host change)
5. Reboot host 3



Cisco Network Boot Requirements

- **Host**
  - PXE 2.0 or 2.1 (available on most Intel-based platforms)
  - Windows 2000 or 2003 host or Advanced host (XP Pro also supported). Linux support in Oct 03
  - Cisco iSCSI driver
  - Cisco inbp.com (mini-iSCSI Driver)
- **Storage**
  - Cisco SN5400 or MDS IPS blade
  - One (master) boot image per unique host and OS
  - One boot image (LUN) per network boot host
- **Network**
  - Windows 2000 or Linux DHCP and TFTP host

Useful Information

- **Cisco Storage Networking**
  http://www.cisco.com/go/storagenetworking
- **Cisco AVVID Storage Networking Partner Program**
  http://www.cisco.com/warp/public/779/largeent/partner/esap/storage.html
- **Cisco Storage Router Product Information**
  http://www.cisco.com/go/storagenetworking
- **Cisco Metro Optical Product Information**
  http://www.cisco.com/go/comet
- **Storage Network Industry Association (SNIA)**
  http://wwwin.snia.org
- **Internet Engineering Task Force – IP Storage**
  http://www.ietf.org/html.charters/ips-charter.html
- **ANSI T11 – Fibre Channel**
  http://www.t11.org/index.htm