LET’S
BUILD
TOMORROW
TODAY
Fibre Channel Networking for the IP Network Engineer and SAN Core Edge Design Best Practices

Chad Hintz
Technical Solutions Architect

BRKVIR-1121
Fibre Channel Networking for the IP Network Engineer and SAN Core-Edge Design Best Practices

This session gives non-storage-networking professionals the fundamentals to understand and implement storage area networks (SANs). This curriculum is intended to prepare attendees for involvement in SAN projects and I/O Consolidation of Ethernet & Fibre Channel networking. You will be exposed to the introduction of Storage Networking terminology & Designs. Specific topics covered include Fibre Channel (FC), FCoE, FC services, FC addressing, fabric routing, zoning, virtual SANs (VSANs). The session includes discussions on Designing Core-Edge Fibre Channel Networks and the best practice recommendations around them. This is an introductory session and attendees are encouraged to follow up with other SAN breakout sessions and labs to learn more about specific advanced topics.
Goals of This Session…
Session Agenda

- History of Storage Area Networks
  - Fibre Channel Technology
  - Designing Fibre Channel Networks
  - Storage Topologies
  - Introduction to NPIV/NPV
  - Core-Edge Design Review
  - Recommended Core-Edge Designs for Scale and Availability
  - Fibre Channel over Ethernet
  - Next Generation Core-Edge Designs
History of Storage Area Networks
It starts with the relationship...

The #1 Rule
Not all storage is local

• Must maintain that 1:1 relationship, especially if there's a network in-between

• All storage networking is based on this fundamental principle
SCSI

Must maintain that 1:1 relationship, especially if there's a network in-between

Otherwise, you get data loss
Very Bad Things™ happen
This Is Why Storage People Are Paranoid

• Losing data means:
  • Losing job
  • Running afoul of government regulations
  • Possibly SEC filings
  • Lost revenue
Direct Attached Storage (DAS)

- Hosts directly access storage as block-level devices
- As Hosts need more disk externally connected storage gets added
- Storage is captive behind the server, limited mobility
- Limited scalability due to limited devices
- No efficient storage sharing possible
- Costly to scale; complex to manage
Network Attached Storage (NAS)

- Clients connect to a network-capable storage server
- Serves files to other devices on the network
  - Called "File-Based" storage
- SMB - successor of CIFS protocol with enhanced functionality, Microsoft
- NFS – Network File System,
- Different from DAS, which uses block-based storage (a block may only contain a part of a single file)
- TCP 'or' UDP based
- Almost all NFS clients perform local caching of both file and directory data
Storage Area Networks

Extending the SCSI Relationship across a network

• Same SCSI protocol carried over a network transport via serial implementation

• Transport must not jeopardize SCSI payload (security, integrity, latency)

• Two primary transports to choose from today, namely IP and Fibre Channel
  • Fibre Channel provides high-speed transport for SCSI payload via Host Bus Adapter (HBA)
  • Fibre Channel overcomes many shortcomings of parallel I/O and combines best attributes of a channel and a network together
Storage Area Networks

iSCSI

- A SCSI transport protocol that operates on top of TCP
- Encapsulates SCSI commands and data into TCP/IP byte-streams
SAN Protocols for Block I/O

- **Fibre Channel**
  - A gigabit-speed network technology primarily used for storage networking

- **Fibre Channel over Ethernet (FCoE)**
  - An encapsulation of FibreChannel frames over Ethernet networks. This allows Fibre Channel to use 10 Gigabit Ethernet networks while preserving the Fibre Channel protocol

- **iSCSI**
  - A TCP/IP-based protocol for establishing and managing connections between IP-based storage devices, hosts and clients

- **FCIP**
  - A provides a standard way of encapsulating FC frames within TCP/IP, allowing islands of FC SANs to be interconnected over an IP-based network
<table>
<thead>
<tr>
<th>Layer</th>
<th>Protocol 1</th>
<th>Protocol 2</th>
<th>Protocol 3</th>
<th>Protocol 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SCSI</strong></td>
<td><strong>iSCSI</strong></td>
<td><strong>FCP</strong></td>
<td><strong>FCIP</strong></td>
<td><strong>FCoE</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>FC</strong></td>
<td></td>
<td><strong>FC</strong></td>
</tr>
<tr>
<td></td>
<td><strong>TCP</strong></td>
<td></td>
<td><strong>IP</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Ethernet</strong></td>
<td><strong>Ethernet</strong></td>
<td><strong>Ethernet</strong></td>
</tr>
<tr>
<td><strong>PHYSICAL WIRE</strong></td>
<td><strong>SCSI</strong></td>
<td><strong>iSCSI</strong></td>
<td><strong>FCIP</strong></td>
<td><strong>FCoE</strong></td>
</tr>
</tbody>
</table>
Session Agenda

✓ History of Storage Area Networks
✓ Fibre Channel Technology
  • Designing Fibre Channel Networks
  • Storage Topologies
  • Introduction to NPIV/NPV
  • Core-Edge Design Review
  • Recommended Core-Edge Designs for Scale and Availability
  • Fibre Channel over Ethernet
  • Next Generation Core-Edge Designs
Fibre Channel Technology
Fibre Channel Basic Building Blocks

- Components
- Naming Conventions
- Connection Fundamentals
Fibre Channel SAN Components

- Servers with host bus adapters
- Storage systems
  - RAID
  - JBOD
  - Tape
- Switches
- SAN management software
Types of MDS/Nexus DC Switches

SAN
- Cisco MDS 9500, 9700 Series
- Cisco MDS 9200 Series
- Cisco MDS 9100 Series

LAN
- Cisco Nexus 3000
- Cisco Nexus 2000
- Cisco Nexus 1000V
- Cisco Nexus 1010

LAN/SAN
- Cisco Nexus 6000
- Cisco Nexus 5500/5600
- Cisco Nexus 4000
- Cisco Nexus 7000

Cisco MDS 9500, 9700 Series
Cisco MDS 9200 Series
Cisco MDS 9100 Series
Cisco Nexus 3000
Cisco Nexus 2000
Cisco Nexus 1000V
Cisco Nexus 1010
Cisco Nexus 6000
Cisco Nexus 5500/5600
Cisco Nexus 4000
Cisco Nexus 7000

CISCO NX-OS: From Hypervisor to Core
CISCO DCNM: Single Pane of Management
FC Switch Port Types

- "N" for Node Port
  - Host or target port facing a switch
FC Switch Port Types

- "N" for Node Port
  - Host or target port facing a switch
- "F" for Fabric
  - Switch port facing the end Nodes
FC Switch Port Types

- "N" for Node Port
  - Host or target port facing a switch
- "F" for Fabric
  - Switch port facing the end Nodes
- "E" for "Extender"
  - Switch port facing another switch
FC Switch Port Types

- "N" for Node Port
  - Host or target port facing a switch
- "F" for Fabric
  - Switch port facing the end Nodes
- "E" for "Extender"
  - Switch port facing another switch

These names are **very** important!
(we will be talking about them a lot)
Fabric Name and Addressing, Part 1: WWN-Like a MAC address

- Every Fibre Channel port and node has a hard-coded address called World Wide Name (WWN)

- During FLOGI the switch identifies the WWN in the service parameters of the accept frame and assigns a Fibre Channel ID (FCID)

- Switch Name Server maps WWNs to FCID
  - WWNN uniquely identify devices
  - WWPN uniquely identify each port in a device

![Example WWN and WWNs from a Dual-Ported Device]
Fibre Channel Addressing, Part 2: Domain ID

- Domain ID is native to a single FC switch
- Limitation of domain IDs in a single fabric
- Forwarding decisions made on domain ID found in first 8 bits of FCID
Start from the Beginning…

- Start with the host and a target that need to communicate
  - Host has 2 HBAs (one per fabric) each with a WWN
  - Target has multiple ports to connect to fabric

- Connect to a FC Switch
  - Port Type Negotiation
  - Speed Negotiation

- FC Switch is part of the SAN “fabric”

- Most commonly, dual fabrics are deployed for redundancy
My Port Is Up…Can I Talk Now?

FLOGIs/PLOGIs

• Step 1: Fabric Login (FLOGI)
  • Determines the presence or absence of Fabric
  • Exchanges Service Parameters with Fabric
  • Switch identifies the WWN in service parameters of the accept frame and assigns a Fibre Channel ID (FCID)
  • Initializes the buffer-to-buffer credits

• Step 2: Port Login (PLOGI)
  • Required between nodes that want to communicate
  • Similar to FLOGI – transports a PLOGI frame to the designation node port
  • In p2p topology (no fabric present), initializes buffer-to-buffer credits
Buffer to Buffer Credits

Fibre Channel Flow Control

- B2B Credits used to ensure that FC transport is lossless
- # of credits negotiated between ports when link is brought up
- # Credits decremented with each packet placed on the wire
  - Independent of packet size
  - If # credits == 0, no more packet transmission
- # of credits incremented with each “transfer ready” received
- B2B Credits need to be taken into consideration as distance and/or bandwidth increases
Fabric Channel ID (FCID)-Like an IP address

Fibre Channel Addressing Scheme

- FCID assigned to every WWPN corresponding to an N_Port
- FCID made up of switch domain, area and device
- Domain ID is native to a single FC switch
- Limitation of domain IDs in a single fabric
- Forwarding decisions made on domain ID found in first 8 bits of FCID
Session Agenda
✓ History of Storage Area Networks
✓ Fibre Channel Technology
✓ Designing Fibre Channel Networks
  • Storage Topologies
  • Introduction to NPIV/NPV
  • Core-Edge Design Review
  • Recommended Core-Edge Designs for Scale and Availability
• Fibre Channel over Ethernet
• Next Generation Core-Edge Designs
Designing Fibre Channel Networks
Storage Criteria Themes

• Fibre Channel Fabric Name Server
• Frame Forwarding
• Security/Segmentation
• Load Balancing
• Bandwidth Considerations
• HA and Reliability
Fabric Shortest Path First (like OSPF)

Fibre Channel Forwarding

- FSPF “routes” traffic based on destination domain ID found in the destination FCID
- For FSPF a domain ID identifies a single switch
  - The number of Domains IDs are limited to 239/75 (theoretical/qualified) within the same fabric (VSAN)
- FSPF performs hop-by-hop routing
- FSPF uses total cost as the metric to determine most efficient path
- FSPF supports equal cost load balancing across links
FC vs. IP routing
Directory Server/Name Server (Like DNS)

- Repository of information regarding the components that make up the Fibre Channel network
- Located at address FF FF FC (some readings call this the name server)
- Components can register their characteristics with the directory server
- An N_Port can query the directory server for specific information
  - Query can be the address identifier, WWN and volume names for all SCSI targets

```
show fcons database
```

<table>
<thead>
<tr>
<th>FCID</th>
<th>TYPE</th>
<th>PWWN</th>
<th>(VENDOR)</th>
<th>FC4-TYPE:FEATURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0xc70000</td>
<td>N</td>
<td>20:47:00:0d:ec:f0:85:40 (Cisco)</td>
<td>npv</td>
<td></td>
</tr>
<tr>
<td>0xc70001</td>
<td>N</td>
<td>20:fa:00:25:b5:aa:00:01</td>
<td>scsi-fcp:init fc-gs</td>
<td></td>
</tr>
<tr>
<td>0xc70002</td>
<td>N</td>
<td>20:fa:00:25:b5:aa:00:00</td>
<td>scsi-fcp:init fc-gs</td>
<td></td>
</tr>
<tr>
<td>0xc70008</td>
<td>N</td>
<td>20:fe:00:25:b5:aa:00:01</td>
<td>scsi-fcp:init fc-gs</td>
<td></td>
</tr>
<tr>
<td>0xc7000f</td>
<td>N</td>
<td>20:fe:00:25:b5:aa:00:00</td>
<td>scsi-fcp:init fc-gs</td>
<td></td>
</tr>
<tr>
<td>0xc70010</td>
<td>N</td>
<td>20:fe:00:25:b5:aa:00:02</td>
<td>scsi-fcp:init fc-gs</td>
<td></td>
</tr>
<tr>
<td>0xc70029</td>
<td>N</td>
<td>20:00:00:25:b5:20:00:0e</td>
<td>scsi-fcp:init fc-gs</td>
<td></td>
</tr>
<tr>
<td>0xc70100</td>
<td>N</td>
<td>50:06:0b:00:00:65:e2:0e (HP)</td>
<td>scsi-fcp:target</td>
<td></td>
</tr>
<tr>
<td>0xc70200</td>
<td>N</td>
<td>20:20:54:7f:ee:2a:14:40 (Cisco)</td>
<td>npv</td>
<td></td>
</tr>
<tr>
<td>0xc70201</td>
<td>N</td>
<td>20:00:54:7b:1a:86:c6:65</td>
<td>scsi-fcp:init fc-gs</td>
<td></td>
</tr>
</tbody>
</table>
Login Complete…
Almost There Fabric Zoning

- Zones are the basic form of data path security
- Zone members can only “see” and talk to other members of the zone
- Devices can be members of more than one zone
- Default zoning is “deny”
- Zones belong to a zoneset
- Zoneset must be “active” to enforce zoning
- Only one active zoneset per fabric or per VSAN

Cisco Live!
SAN Islands

Production SAN

Tape SAN

Test SAN

SAN A
DomainID=1
DomainID=7

SAN B
DomainID=2
DomainID=8

SAN C
DomainID=3

SAN D
DomainID=4

SAN E
DomainID=5

SAN F
DomainID=6
Virtual SANs (VSANs)-Where have I heard this before?

• Analogous to VLANs in Ethernet

• Virtual fabrics created from larger cost-effective redundant physical fabric

• Reduces wasted ports of a SAN island approach

• Fabric events are isolated per VSAN which gives further isolation for High Availability

• Statistics can be gathered per VSAN

• Each VSAN provides Separate Fabric Services
  • FSPF, Zones/Zoneset, DNS, RSCN
SAN Islands – with Virtual SANs

Production SAN

Tape SAN

Test SAN

SAN A
DomainID=1

SAN B
DomainID=2

SAN C
DomainID=3

SAN D
DomainID=4

SAN E
DomainID=5

SAN F
DomainID=6
VSANs and Zones—Complimentary

Virtual SANs and Fabric Zoning Are Very Complementary

- Hierarchical relationship—
  - First assign physical ports to VSANs
  - Then configure independent zones per VSAN

- VSANs divide the physical infrastructure

- Zones provide added security and allow sharing of device ports

- VSANs provide traffic statistics

- VSANs only changed when ports needed per virtual fabric

- Zones can change frequently (e.g. backup)

- Ports are added/removed non-disruptively to VSANs

---

**Relationship of VSANs to Zones**

- **Physical Topology**

- **VSAN 2**
  - ZoneA: Host1, Disk1
  - ZoneB: Host2, Disk4
  - ZoneC: Host1, Disk3

- **VSAN 3**
  - ZoneA: Host3, Disk5
  - ZoneB: Host3, Disk6
  - ZoneD: Host4

---

**Cisco Live!**
Over-Subscription FAN-OUT Ratio

Over-subscription (or fan-out) ratio for sizing ports and links

- Factors used
  - Speed of Host HBA interfaces
  - Speed of Array interfaces
  - Type of server and application

- Storage vendors provide guidance in the process
- Ratios range between 4:1 - 20:1

Example:

- 10:1 O/S ratio
- 60 Servers with 4 Gb HBAs
- 3 x 8G ISL ports
- 6 x 4G Array ports
- 240 G
- 24 G
- 24 G
Inter-Switch Link PortChanneling

A PortChannel Is a Logical Bundling of Identical Links

• Criteria for forming a PortChannel
  • Same speed links
  • Same modes (auto, E, etc.) and states
  • Between same two switches
  • Same VSAN membership

• Treated as one logical ISL by upper layer protocols (FSPF)

• Can use up to 16 links in a PortChannel

• Can be formed from any ports on any modules—HA enabled

• Exchange-based in-order load balancing
  • Mode one: based on src/dst FC_IDs
  • Mode two: based on src/dst FC_ID/OX_ID

• Much faster recovery than FSPF-based balancing

• Given logical interface name with aggregated bandwidth and derived routing metric
PortChannel vs. Trunking

- ISL = inter-switch link
- PortChannel = E_Ports and ISLs
- Trunk = ISLs that support VSANs
- Trunking = TE_Ports and EISLs
Session Agenda

- History of Storage Area Networks
- Fibre Channel Technology
- Designing Fibre Channel Networks
- Storage Topologies
  - Introduction to NPIV/NPV
  - Core-Edge Design Review
  - Recommended Core-Edge Designs for Scale and Availability
  - Fibre Channel over Ethernet
  - Next Generation Core-Edge Designs
Storage Topologies
SAN Design

Key Requirements

• High Availability - Providing a Dual Fabric (current best practice)
• Meeting oversubscription ratios established by disk vendors
• Effective zoning
• Providing Business Function/Operating System Fabric Segmentation and Security
• Fabric scalability (FLOGI and domain-id scaling)
• Providing connectivity for virtualized servers
• Providing connectivity for diverse server placement and form factors
The Design Requirements

Classical Fibre Channel

• Fibre Channel SAN
  • Transport and Services are on the same layer in the same devices
  • Well defined end device relationships (initiators and targets)
  • Does not tolerate packet drop – requires lossless transport
  • Only north-south traffic, east-west traffic mostly irrelevant

• Network designs optimized for Scale and Availability
  • High availability of network services provided through dual fabric architecture
  • Edge/Core vs Edge/Core/Edge
  • Service deployment

Fabric topology, services and traffic flows are structured

Client/Server Relationships are pre-defined
SAN Design – Single Tier Topology

- Collapsed Core Design
- Servers connect to the Core switches
- Storage devices connect to one or more core switches
- Core switches provide storage services
- Large amount of blades to support Initiator (Host) and Target (Storage) ports
- Single Management per Fabric
- Normal for Small SAN environments
- HA achieved in two physically separate, but identical, redundant SAN fabrics
How Do We Avoid This?
SAN Design – Two Tier Topology

- “Core-Edge” Topology- Most Common
- Servers connect to the edge switches
- Storage devices connect to one or more core switches
- Core switches provide storage services to one or more edge switches, thus servicing more servers in the fabric
- ISLs have to be designed so that overall fan-in ratio of servers to storage and overall end-to-end oversubscription are maintained
- HA achieved in two physically separate, but identical, redundant SAN fabrics
SAN Design – Three Tier Topology

- “Edge-Core-Edge” Topology
- Servers connect to the edge switches
- Storage devices connect to one or more edge switches
- Core switches provide storage services to one or more edge switches, thus servicing more servers and storage in the fabric
- ISLs have to be designed so that overall fan-in ratio of servers to storage and overall end-to-end oversubscription are maintained
- HA achieved in two physically separate, but identical, redundant SAN fabrics
Session Agenda

✓ History of Storage Area Networks
✓ Fibre Channel Technology
✓ Designing Fibre Channel Networks
✓ Storage Topologies
✓ Introduction to NPIV/NPV
  • Core-Edge Design Review
  • Recommended Core-Edge Designs for Scale and Availability
  • Fibre Channel over Ethernet
  • Next Generation Core-Edge Designs
Introduction to NPIV/NPV
What Is NPIV? and Why?

- N-Port ID Virtualization (NPIV) provides a means to assign multiple FCIDs to a single N_Port
  - Limitation exists in FC where only a single FCID can be handed out per F-port. Therefore an F-Port can only accept a single FLOGI

- Allows multiple applications to share the same Fiber Channel adapter port

- Usage applies to applications such as Virtualization

Note: you may hear of it as the “NPIV upstream” switch
What Is NPV? and Why?

• N-Port Virtualizer (NPV) utilizes NPIV functionality to allow a “switch” to act like a server performing multiple logins through a single physical link

• Physical servers connected to the NPV switch login to the upstream NPIV core switch

• No local switching is done on an FC switch in NPV mode

• FC edge switch in NPV mode does not take up a domain ID
  • Helps to alleviate domain ID exhaustion in large fabrics
NPV Auto Load Balancing

- Uniform balancing of server loads on NP links
  - Server loads are not tied to any uplink

- Benefit
  - Optimal uplink bandwidth utilization
NPV Auto Load Balancing

• Automatically moves the failed servers to other available NP links
  • Servers is forced to re-login immediately after experiencing “short” traffic disruption.

• Benefit
  • Downtime greatly reduced

• Automatic failover of loads on NP links
F-Port Port Channel

- **F-Port PortChannels**
  - Bundle multiple ports into 1 logical link
  - Similar to ISL portchannels in FC and EtherChannels in Ethernet

- **Benefits**
  - High-Availability- no disruption if cable, port, or line cards fail
  - Optimal bandwidth utilization & higher aggregate bandwidth with load balancing

**Enhance NPV uplink Resiliency**
NPV F-Port Port Channel

- Link failures do not affect the server connectivity
- No application disruption

No traffic disruption
F-Port Trunking

- F-Port Trunking
  - Uplinks carry multiple VSANs
- Benefits
  - Extend VSAN benefits to Blade servers
  - Separate management domains
  - Traffic Isolation and ability to host differentiated services on blades
- Extend VSAN Benefits to Blades
Session Agenda

- History of Storage Area Networks
- Fibre Channel Technology
- Designing Fibre Channel Networks
- Storage Topologies
- Introduction to NPIV/NPV
- Core-Edge Design Review
  - Recommended Core-Edge Designs for Scale and Availability
  - Fibre Channel over Ethernet
  - Next Generation Core-Edge Designs
Core-Edge Design Review
SAN Design – Two Tier Topology

• “Edge-Core” Topology- Most Common

• Servers connect to the edge switches

• Storage devices connect to one or more core switches

• Core switches provide storage services to one or more edge switches, thus servicing more servers in the fabric

• ISLs have to be designed so that overall fan-in ratio of servers to storage and overall end-to-end oversubscription are maintained

• HA achieved in two physically separate, but identical, redundant SAN fabrics
Blade Switch Explosion Issues

• **Scalability**
  • Each Blade Switch uses a single Domain ID
  • Theoretical maximum number of Domain IDs is 239 per VSAN
  • Supported number of domains is quite smaller (depends on OSM)
    • EMC: 80 domains
    • Cisco Tested: 90

• **Manageability**
  • More switches to manage
  • Shared management of blade switches between storage and server administrators
Server Consolidation with Top-of-Rack Fabric Switches

Top of Rack Design

Ports Deployed:
- Storage Ports (4 G Dedicated): 1200
- Host Ports (4 G Shared): 192
- Disk Oversubscription (Ports): 896
- Number of FC switches in the fabric: 9.3 : 1
- Number of FC switches in the fabric: 30
Server Consolidation with Blade servers

Blade Servers

Blade Server Design Using 2 x 4 G ISL per Blade Switch;

Ports Deployed:
- Storage Ports (4 G Dedicated): 1608
- Host Ports (4 G Shared): 240
- Disk Oversubscription (Ports): 480
- Number of FC switches in the fabric: 62

2X ISL to Core at 4G
16 Host Ports at 4G
120 Storage Ports at 2 G
60 ISL to Edge at 4 G

Five Racks
96 Dual Attached Blade Servers per Rack
Session Agenda

✓ History of Storage Area Networks
✓ Fibre Channel Technology
✓ Designing Fibre Channel Networks
✓ Storage Topologies
✓ Introduction to NPIV/NPV
✓ Core-Edge Design Review
✓ Recommended Core-Edge Designs for Scale and Availability
  • Fibre Channel over Ethernet
  • Next Generation Core-Edge Designs
Recommended Core-Edge Designs for Scale and Availability
SAN Design

Key Requirements

- High Availability - Providing a Dual Fabric (current best practice)
- Fabric scalability (FLOGI and domain-id scaling)
- Providing connectivity for diverse server placement and form factors
  - Meeting oversubscription ratios established by disk vendors
  - Effective zoning
  - Providing Business Function/Operating System Fabric Segmentation and Security
- Providing connectivity for virtualized servers
N-Port Virtualizer (NPV) Reduces Number of FC Domain IDs

Top of Rack Design
Fabric Switches in NPV mode

Ports Deployed: 1200
Storage Ports (4 G Dedicated): 192
Host Ports (4 G Shared): 896
Disk Oversubscription (Ports): 9.3:1
Number of FC switches in the fabric 2

MDS 91xx in NPV mode
14 Racks
32 Dual Attached Servers per Rack
NPV Blade Switch

Blade Servers

Blade Server Design
Using 2 x 4 G ISL per Blade Switch;

Less cables/power

Ports Deployed: 1608
Storage Ports (4 G Dedicated): 240
Host Ports (4 G Shared): 480
Disk Oversubscription (Ports): 8 : 1
Number of fabric switches to manage 2
F-Port Port Channel

Enhance NPV Uplink Resiliency

- F-Port PortChannels
  - Bundle multiple ports into 1 logical link
  - Similar to ISL portchannels in FC and EtherChannels in Ethernet
- Benefits
  - High-Availability- no disruption if cable, port, or line cards fail
  - Optimal bandwidth utilization & higher aggregate bandwidth with load balancing
Control and monitor VMs in the SAN using NPIV

- NPIV gives Virtual Servers SAN identity
  - Designed for virtual server environments

- Allows SAN control of VMs
  - Zoning and LUN Masking at VM level

- Multiple applications on the same port can use different IDs
  - Better utilization of the server connectivity
Summary of Recommendations

• High Availability - Provide a Dual Fabric
• Use of Port-Channels and F-Port Channels with NPIV to provide the bandwidth to meet oversubscription ratios
• Use NPIV/NPV to provide Domain ID scaling and ease of management
• Use of host level NPIV and Nested NPV to provide visibility to Virtualized servers
Session Agenda

- History of Storage Area Networks
- Fibre Channel Technology
- Designing Fibre Channel Networks
- Storage Topologies
- Introduction to NPIV/NPV
- Core-Edge Design Review
- Recommended Core-Edge Designs for Scale and Availability
- Fibre Channel over Ethernet
  - Next Generation Core-Edge Designs
FCoE
Traditional Data Center Design

Ethernet LAN and Fibre Channel SAN

• Physical and Logical separation of LAN and SAN traffic

• Additional Physical and Logical separation of SAN fabrics
Network Behavior

• Ethernet is non-deterministic
  • Flow control is destination-based
  • Relies on TCP drop-retransmission / sliding window

• Fibre-Channel is deterministic
  • Flow control is source-based (B2B credits)
  • Services are fabric integrated (no loop concept)
“Fabric vs. Network” or “Fabric & Network”

SAN Dual Fabric Design

- SAN and LAN have different design assumptions
  - SAN leverages dual fabric with multi-pathing failover initiated by the client
  - LAN leverages single fully meshed fabric with higher levels of component redundancy
Fibre Channel over Ethernet

What Enables It?

• 10Gbps Ethernet

• Lossless Ethernet
  • Matches the lossless behavior guaranteed in FC by B2B credits

• Ethernet jumbo frames
  • Max FC frame payload = 2112 bytes
Unified Fabric

Why?

- Fewer CNAs (Converged Network adapters) instead of NICs, HBAs and HCAs
- Limited number of interfaces for Blade Servers

![Diagram showing traffic flows]

All traffic goes over 10GE
Standards for I/O Consolidation

www.T11.org

Fibre Channel on network media

FC-BB-5

Completed June 2009
Published by ANSI in May 2010

IEEE 802.1

DCB

PFC

IEEE 802.1Qbb
Priority-based Flow Control

ETS

IEEE 802.1Qaz
Priority Grouping
Enhanced Transmission Selection

DCBx

IEEE 802.1Qaz
Configuration Verification

Completed March 2011
Forwarded to RevCom for publication

FCoE
FCoE, Same Model as FC

- Same host to target communication
  - Host has 2 CNA’s (one per fabric)
  - Target has multiple ports to connect to fabric
- Connect to a DCB capable switch
  - Port Type Negotiation
  - Speed Negotiation
  - DCBX Negotiation
- Access switch is a Fibre Channel over Ethernet Forwarder (FCF) in this directly connected model
- Dual fabrics are still deployed for redundancy
IEEE 802.1Qbb Priority Flow Control

- PFC enables Flow Control on a Per-Priority basis
- Therefore, we have the ability to have lossless and lossy priorities at the same time on the same wire
- Allows FCoE to operate over a lossless priority independent of other priorities
- Other traffic assigned to other CoS will continue to transmit and rely on upper layer protocols for retransmission
- Not only for FCoE traffic
Fibre Channel over Ethernet Forwarder (FCF)

**Fibre Channel over Ethernet Forwarders (FCF)**

Allows switching of FCoE frames across multiple hops

Creates Standards based FCoE ISL

Necessary for Multi-Hop FCoE

Nothing Further Required

(No TRILL or Spanning Tree)
Introducing FCoE Initialization Protocol

• "FIP" discovers other FCoE capable devices within the lossless Ethernet Network

• Enables FCoE adapters (CNAs) to discover FCoE switches (FCFs) on the FCoE VLAN

• Establishes a virtual link between the adapter and FCF or between two FCFs

• "FIP" uses a different Ethertype from FCoE: allows for "FIP"-Snooping by DCB capable Ethernet bridges

• Building foundation for future Multitier FCoE topologies
"FIP" - Login Flow Ladder

- ENode
  - VLAN Discovery
  - FCF Discovery
  - FLOGI/FDISC
  - FC Command

- FCoE Switch
  - VLAN Discovery
  - FCF Discovery
  - FLOGI/FDISC Accept
  - FC Command responses

"FIP": FCoE Initialization Protocol

SAN Protocols
FCoE Behind the Scenes

• Data Center Bridging eXchange (DCBX):
  • DCBX, is a protocol that extends the Link Layer Discovery Protocol (LLDP) defined in IEEE802.1Qaz. DCBX allows the FCF to provide Link Layer configuration information to the CNA and allows both the CNA and FCF to exchange status.

• FCoE Protocols:
  • FCoE Data plane protocol.
    FCoE data protocol requires lossless Ethernet and is typically implemented in hardware, and is used to carry the FC frames with SCSI.
  • "FIP" (FCoE Initialization Protocol)
    • Used to discover FCoE capable devices connected to an Ethernet network and to negotiate capabilities.
Session Agenda

✓ History of Storage Area Networks
✓ Fibre Channel Technology
✓ Designing Fibre Channel Networks
✓ Storage Topologies
✓ Introduction to NPIV/NPV
✓ Core-Edge Design Review
✓ Recommended Core-Edge Designs for Scale and Availability
✓ Fibre Channel over Ethernet
✓ Next Generation Core-Edge Designs
Next Generation Core-Edge Designs
Single Hop FCoE Design

- Single Hop (Directly Connected) FCoE from the CNA to FCF, then broken out to Native FC
- A VLAN can be dedicated for every Virtual Fabric in the SAN
- "FIP" discovers the FCoE VLAN and signals it to the hosts
- Trunking is not required on the host driver – all FCoE frames are tagged by the CNA
- FCoE VLANs can be pruned from Ethernet links that are not designate for FCoE
- Maintains isolated edge switches for SAN ‘A’ and ‘B’ and separate LAN switches for NIC 1 and NIC 2 (standard NIC teaming or Link Aggregation)
FCoE and VPC Together

- VPC with FCoE are ONLY supported between hosts and FCF pairs...AND they must follow specific rules
  - A ‘vfc’ interface can only be associated with a single-link channel bundle
  - While the link aggregation configurations are the same on FCF-A and FCF-B, the FCoE VLANs can be different
- FCoE VLANs are ‘not’ carried on the VPC peer-link
- FCoE and "FIP" ethertypes are ‘not’ forwarded over the VPC peer link either
- VPC carrying FCoE between two FCF’s is NOT recommended to keep SAN A/B separation
UCS Core-Edge Design

N-Port Virtualization Forwarding with MDS

- F_Port Channeling and Trunking from MDS to UCS
- FC Port Channel can carry all VSANs (Trunk)
- UCS Fabric Interconnects remains in NPV end host mode
- Server vHBA pinned to an FC Port Channel
- Server vHBA has access to bandwidth on any link member of the FC Port Channel
- Load balancing based on FC Exchange_ID
  - Per Flow
- Loss of Port Channel member link has no effect on Server vHBA (hides the failure)
  - Affected flows to remaining member links
  - No FLOGI required
Extending FCoE with "FIP" Snooping

- What does a FIP Snooping device do
  - FIP solicitations (VLAN Discovery, FCF Discovery and FLOGI) sent out from the CNA and FIP responses from the FCF are “snooped”

- How does a FIP Snooping device work?
  - The FIP Snooping device will be able to know which FCFs hosts are logged into
  - Will dynamically create an ACL to make sure that the host to FCF path is kept secure

- A FIP Snooping device has NO intelligence or impact on FCoE traffic/path selection/load balancing/login selection/etc

- Mentioned in the Annex of the FC-BB-5 (FCoE) standard as a way to provide security in FCoE environments
Fibre Channel Aware Device

FCoE NPV

- "FCoE NPV bridge" improves over a "FIP snooping bridge" by intelligently proxying FIP functions between a CNA and an FCF
- Active Fibre Channel forwarding and security element
- FCoE-NPV load balances logins from the CNAs evenly across the available FCF uplink ports
- FCoE NPV will take VSAN into account when mapping or ‘pinning’ logins from a CNA to an FCF uplink
- Emulates existing Fibre Channel Topology (same mgmt, security, HA)
Multi-Tier with FCoE-NPV

NPV Considerations

- Considered Multi Hop because of intelligence of edge.
- Designed the Same way as FC NPV
- No Local Switching on a NPV edge device
  - Usual traffic is North South from Initiator to Target, but east to west traffic could be an issue in this design
- Still want to use FCoE dedicated uplinks to meet or allocate enough bandwidth for oversubscription rates at 10G

\[\text{NPV Considerations}\]

\[\begin{align*}
\text{FCF} & \quad \text{NPIV} \\
\text{Core} & \quad \text{VF} \\
\text{FCoE} & \quad \text{NPV} \\
\text{VF} & \quad \text{VNP} \\
\text{FCF} & \quad \text{NPV} \\
\text{VF} & \quad \text{VN} \\
\end{align*}\]
Recommended Reading


Available Onsite at the Cisco Company Store
Participate in the “My Favorite Speaker” Contest
Promote Your Favorite Speaker and You Could be a Winner

• Promote your favorite speaker through Twitter and you could win $200 of Cisco Press products (@CiscoPress)

• Send a tweet and include
  • Your favorite speaker’s Twitter handle @chadh0517
  • Two hashtags: #CLUS #MyFavoriteSpeaker

• You can submit an entry for more than one of your “favorite” speakers

• Don’t forget to follow @CiscoLive and @CiscoPress

• View the official rules at http://bit.ly/CLUSwin
Complete Your Online Session Evaluation

- Give us your feedback to be entered into a Daily Survey Drawing. A daily winner will receive a $750 Amazon gift card.

- Complete your session surveys though the Cisco Live mobile app or your computer on Cisco Live Connect.

Don’t forget: Cisco Live sessions will be available for viewing on-demand after the event at CiscoLive.com/Online
Continue Your Education

• Demos in the Cisco Campus
• Walk-in Self-Paced Labs
• Table Topics
• Meet the Engineer 1:1 meetings
Thank you
TOMORROW starts here.