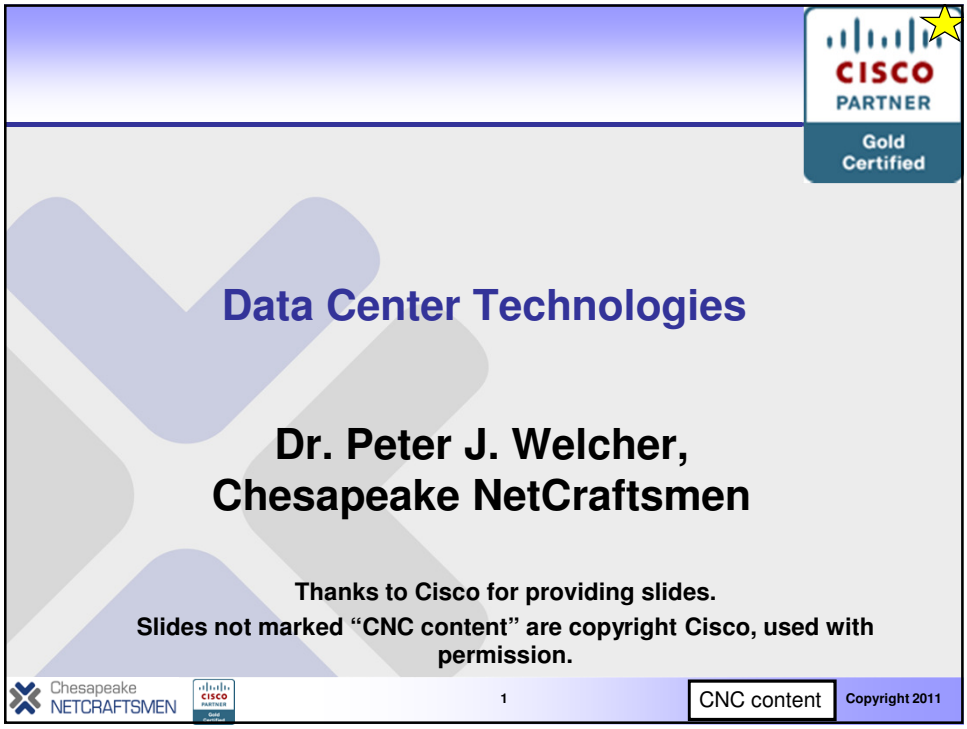
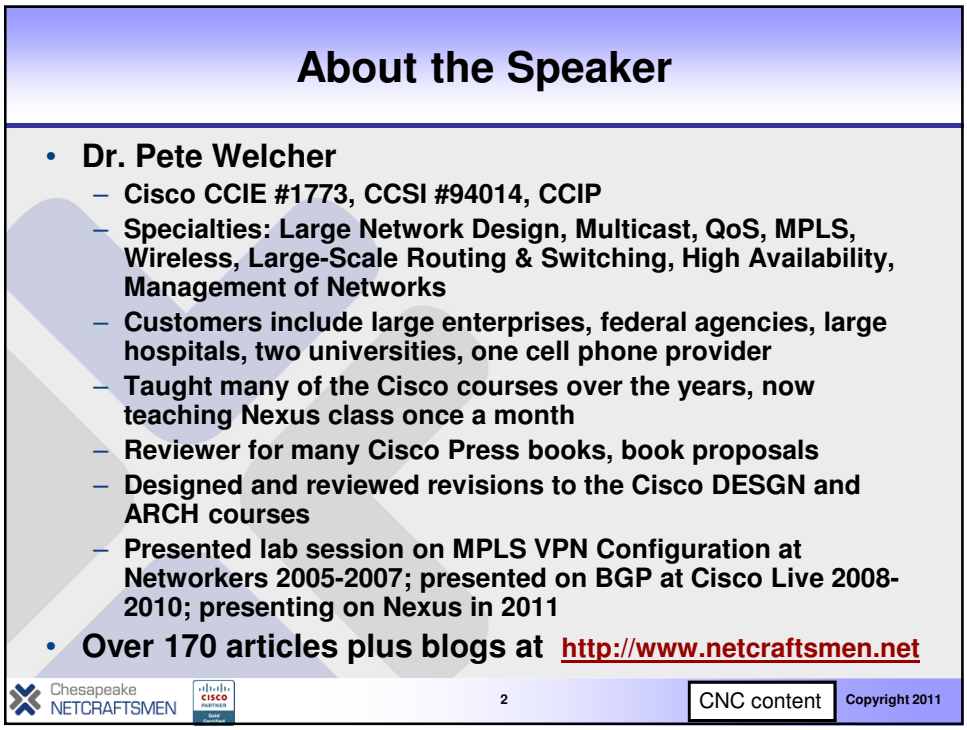




Objective

- Share new developments and insights into some data center-related topics:
  - FabricPath
  - Fiber Channel / SAN for Network Engineers
  - Fiber Channel over Ethernet (FCoE)
  - Recent Cisco Announcements

Motivation

- FabricPath is Cisco's version of TRILL
  - Many L2 interconnects
  - No blocked links
  - Enables huge "east-west" bandwidth in the datacenter
  - Flatten the data center: safe VLANs anywhere
  - Contrast OTV (prior talk, posted): data center to data center L2 over L3
- Fiber Channel / SAN provides some FC basics for network engineers
- Fiber Channel over Ethernet is about how to save money and cabling by combining Data and SAN traffic onto a 10 G links

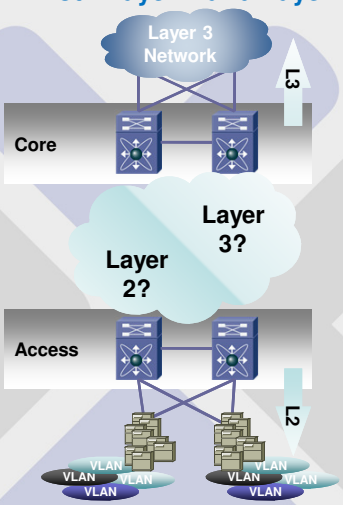
Agenda

- Introduction
- FabricPath
  - Why FabricPath
  - FabricPath Forwarding Details
  - Monitoring and Troubleshooting
- Fiber Channel / SAN for Network Engineers
- Fiber Channel over Ethernet (FCoE)
  - IO Consolidation
  - FCoE Technology
  - Basic Designs
- Recent Announcements
- Summary, References, and Q&A

Eternal Debates on Network Design

Layer 2 or Layer 3?

- Both Layer 2 and Layer 3 are required for any network design



- Subnet provide fault isolation
- Scalable control planes with inherent provision of multi-pathing and multi-topology
- HA with fast convergence
- Additional loop-mitigation mechanism in the data plane (e.g. TTL, RPF check, etc.)

Cisco has solutions for both Layer 2 and Layer 3 to satisfy Customers' requirements

- Simplicity (no planning/configuration required for either addressing or control plane)
- Single control plane protocol for unicast, broadcast, and multicast
- Easy application development

L2 Network Requirements inside DC

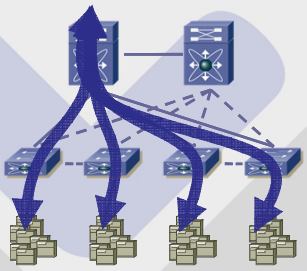
- Maximize Bi-Sectional Bandwidth
- Scalable Layer 2 domain
- High Availability
  - Resilient control-plane
  - Fast convergence upon failure
  - Fault-domain isolation
- Facilitate Application Deployment
  - Workload mobility, Clustering, etc.
- Multi-Pathing/Multi-Topology

Limitations of Spanning Tree Protocol

- Sub-optimal path selection
  - Single path between any 2 bridges in the same L2 network
  - Shortest path only from Root Bridge's perspective
- Under-utilized bandwidth
  - Ensure loop-free L2 logical topologies by blocking redundant links
  - Increased waste of available bandwidth as link-speed getting faster and faster
- No control plane security
  - Root election purely based on switch-ID, which is prone to problems caused by operator errors
- Slow and unreliable reconvergence upon link failure
  - Up to seconds of service disruption even with RSTP

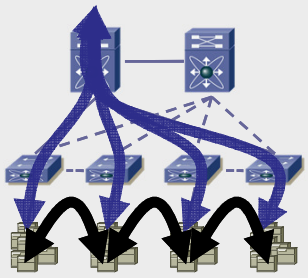
Is Over-subscription Acceptable?

Campus Network



- Mostly North-South traffic flows
- Over-subscription acceptable for client-server type of applications

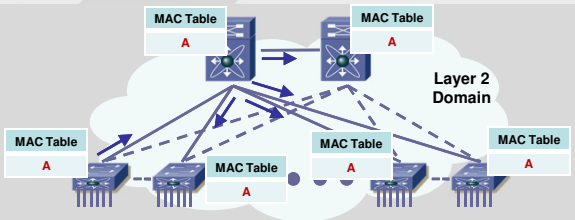
Data Center



- Mix of North-South and East-West traffic flows
- Often demands special design consideration to minimize bandwidth limitation imposed by over-subscription

Nature of Layer 2 Bridging

- Transparent – act like "shared media" to end devices
- Plug-N-Play – No user configuration is required to build forwarding database
- Data plane learning – Forwarding database built based on frame contents
- Flooding – Default forwarding behavior for frames with unknown unicast destination is to flood the whole broadcast domain
- Every MAC, Everywhere!!! – All unicast MACs need be learn by all bridges in the same bridge domain to minimize flooding



The Next Era of Layer 2 Network
What Can Be Improved?

- Network Address Scheme: Flat ➔ *Hierarchical*
  - Additional header is required to allow L2 "Routing" instead of "Bridging"
  - Provide additional loop-prevention mechanism like TTL
- Address Learning: Data Plane ➔ *Control Plane*
  - Eliminate the needs to program all MACs on every switches to avoid flooding
- Control Plane: Distance-Vector ➔ *Link-State*
  - Improve scalability, minimize convergence time, and allow multipathing inherently

The ultimate solution needs to take both control and data plane into consideration this time!!!

Introducing Cisco FabricPath
NX-OS software innovation for multi-pathing Ethernet

"FabricPath brings Layer 3 routing benefits to flexible Layer 2 bridged Ethernet networks"

Switching

- Easy Configuration
- Plug & Play
- Provisioning Flexibility

Routing

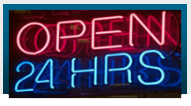
- Multi-pathing (ECMP)
- Fast Convergence
- Highly Scalable

FabricPath

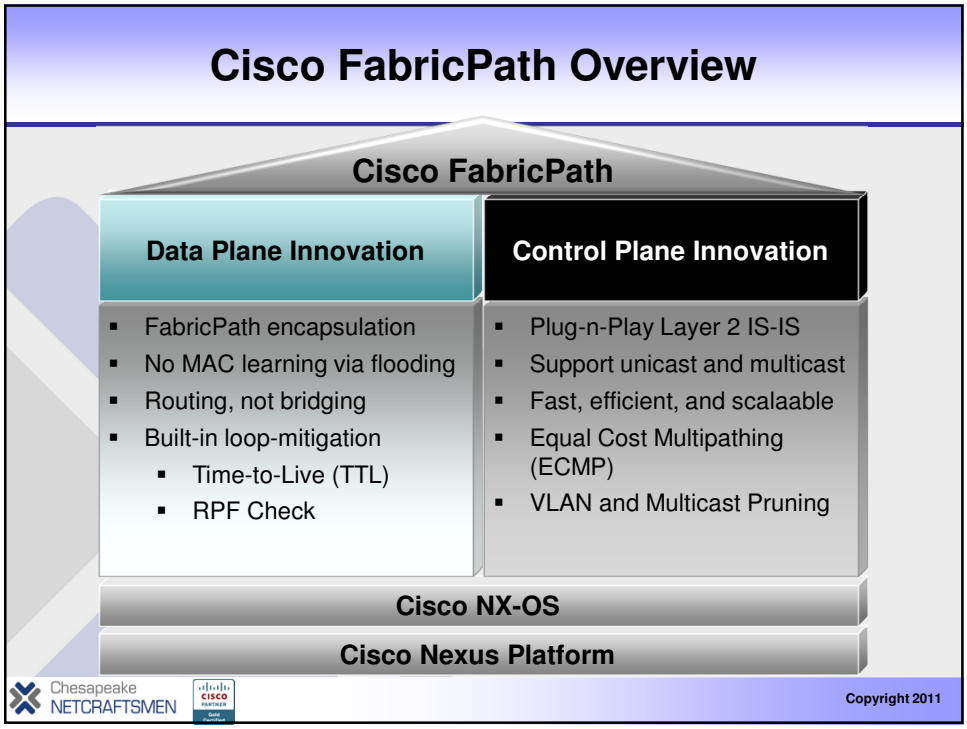
•Auto-Discovery
•Simple Operation

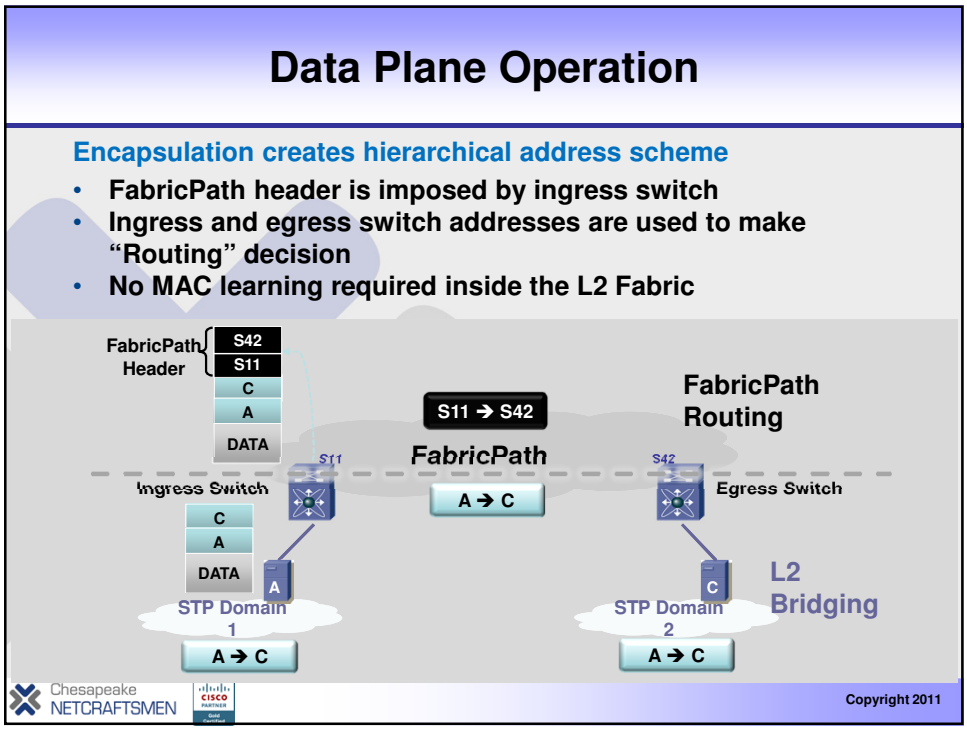
•Flexible Workload Mobility

•Scalable Bandwidth



•High-availability

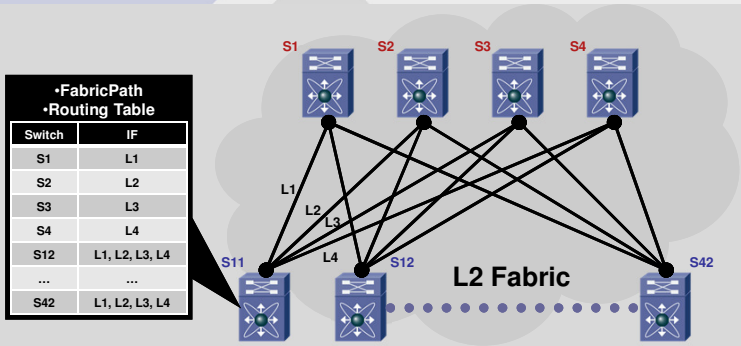




Control Plane Operation

Plug-N-Play L2 IS-IS is used to manage forwarding topology

- Assigned switch addresses to all FabricPath enabled switches automatically (no user configuration required)
- Compute shortest, pair-wise paths
- Support equal-cost paths between any FabricPath switch pairs

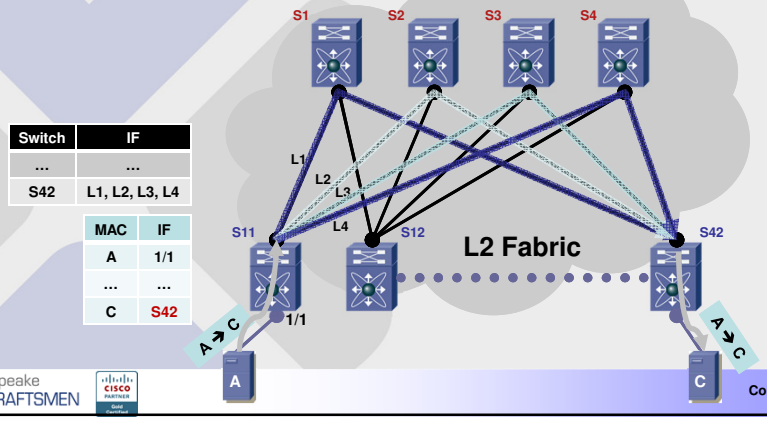


Chesapeake NETCRAFTSMEN

Unicast with FabricPath

Forwarding decision based on 'FabricPath Routing Table'

- Support more than 2 active paths (up to 16) across the Fabric
- Increase bi-sectional bandwidth beyond port-channel
- High availability with N+1 path redundancy

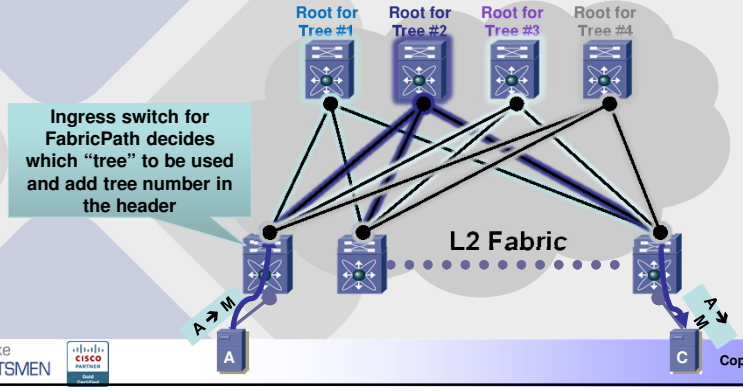


Chesapeake NETCRAFTSMEN

Multicast with FabricPath

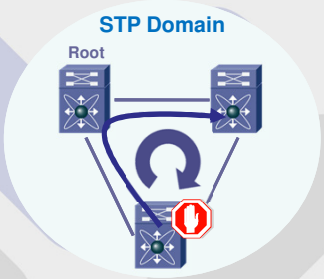
Forwarding through distinct 'Trees'

- Several 'Trees' are rooted in key location inside the fabric
- All Switches in L2 Fabric share the same view for each 'Tree'
- Multicast traffic load-balanced across these 'Trees'

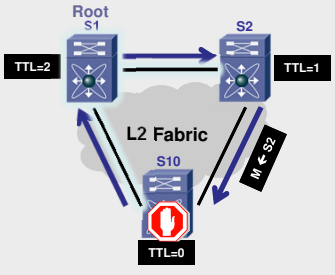


Loop Mitigation with FabricPath

Minimize impact of transient loop with TTL and RPF Check

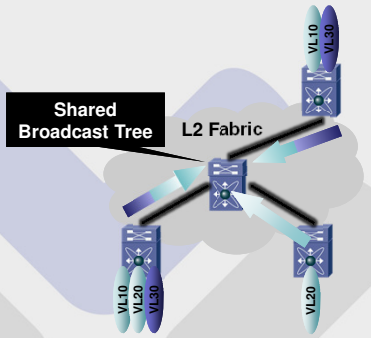


- Block redundant paths to ensure loop-free topology
- Frames loop indefinitely if STP failed
- Could results in complete network melt-down as the result of flooding

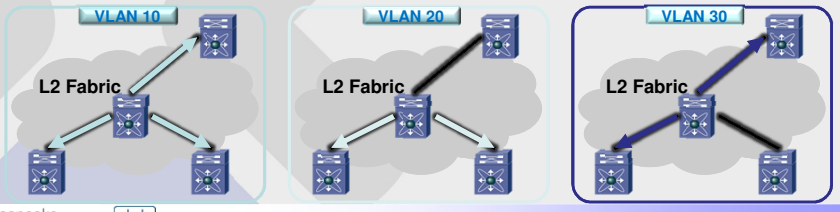


- TTL is part of FabricPath header
- Decrement by 1 at each hop
- Frames are discarded when TTL=0
- RPF check for multicast based on "tree" info

VLAN Pruning in L2 Fabric



- Switches indicate *'locally interested VLANs'* to the rest of the L2 Fabric
- Broadcast traffic for any VLAN only sent to switches that have requested for it



Chesapeake NETCRAFTSMEN

FabricPath Summary

- **FabricPath is simple, keeps the attractive aspects of Layer 2**
  - **Transparent to L3 protocols**
  - **No addressing, simple configuration and deployment**
- **FabricPath is scalable**
  - **Can extend a bridged domain without extending the risks generally associated to Layer 2 (frame routing, TTL, RPFC)**
- **FabricPath is efficient**
  - **High bi-sectional bandwidth (ECMP)**
  - **Optimal path between any two nodes**

Chesapeake NETCRAFTSMEN

Agenda

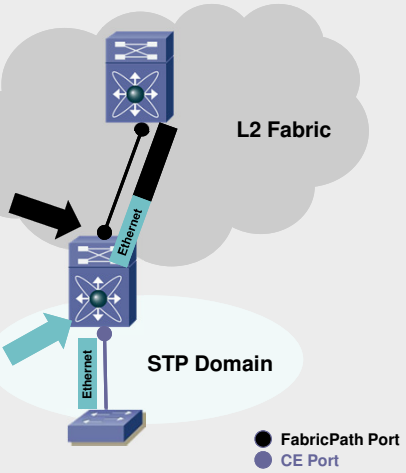
- Introduction
- FabricPath
  - Why FabricPath
  - FabricPath Forwarding Details
  - Monitoring and Troubleshooting
- Fiber Channel / SAN for Network Engineers
- Fiber Channel over Ethernet (FCoE)
  - IO Consolidation
  - FCoE Technology
  - Basic Designs
- Recent Announcements
- Summary, References, and Q&A

FabricPath Interfaces

Configuration used to determine if FabricPath should be run on an interface

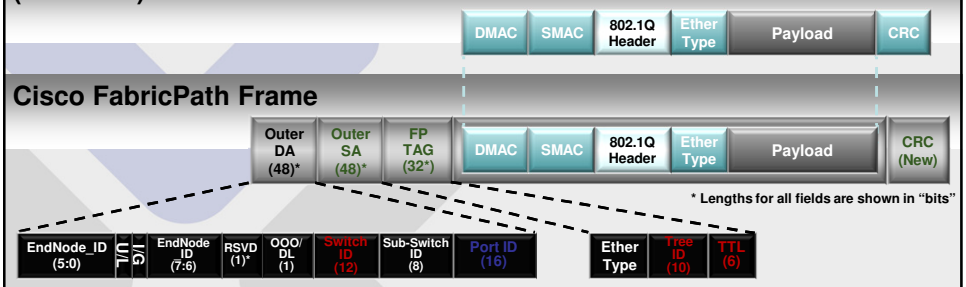
FabricPath Port
- Interfaces connected to another FabricPath Port
- Send/Receive traffic with *FabricPath header*
- *No Spanning-Tree!!!*
- *No 'MAC Learning'*
- Exchange topology info through *L2 ISIS* Adjacency
- Forwarding based on *'Switch Table'*

Classic Ethernet (CE) Port
- Interfaces connected to all existing NICs and Network Devices
- Send/Receive traffic in 802.3 Ethernet frames format
- Participated in STP domain
- Forwarding based on MAC Table



FabricPath Encapsulation

16-byte header provides fields to help create hierarchical L2 address space and facilitate feature enhancements

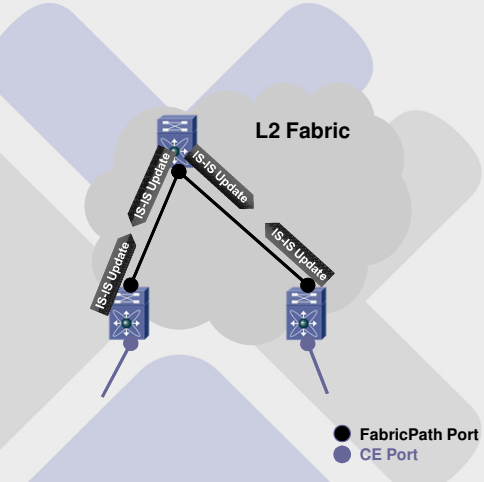


- *Switch ID:* Unique number assigned to help identify each device that is part of L2 Fabric
- *Port ID:* Used to provide information about MAC-to-Interface association at L2 Fabric boundary
- *Tree ID:* Unique number assigned to help identify each distribution "Tree"
- *TTL:* Decrement at each hop to prevent frames looping infinitely in the fabric in case of unexpected failure

Chesapeake NETCRAFTSMEN

Layer 2 IS-IS

Control-Plane Protocol for FabricPath to Replace STP

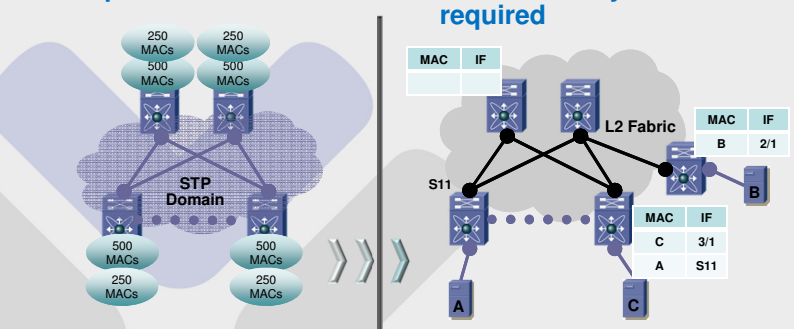


- **Minimal knowledge required with no user configuration by default**
- **➔ Maintain PnP nature of Layer 2**
- **Based on ISO/IEC 10589**
- **Extensible protocol design allows Layer 2 info to be exchanged through IS-IS**
- **Single-level IS-IS with support for P2P links**
- **Calculate forwarding information for traffic forwarding**
- **Link-state protocol with support for ECMP improves failure detection, network reconvergence, and high-availability**

Chesapeake NETCRAFTSMEN

Conversational MAC Learning

Optimize Resource Utilization – Learn only the MAC addresses required



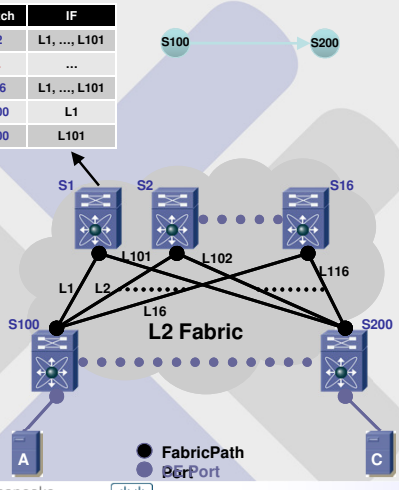
- ALL MACs needs to be learned on EVERY Switch
- Large L2 domain and virtualization present challenges to MAC Table scalability

- **Local MAC:** Source-MAC Learning only happens to traffic received on *CE Ports*
- **Remote MAC:** Source-MAC for traffic received on *FabricPath Ports* are only learned if Destination-MAC is already known as Local

FabricPath 'Routing Table'

Contains Information for Forwarding L2 Unicast Traffic inside the L2 Fabric

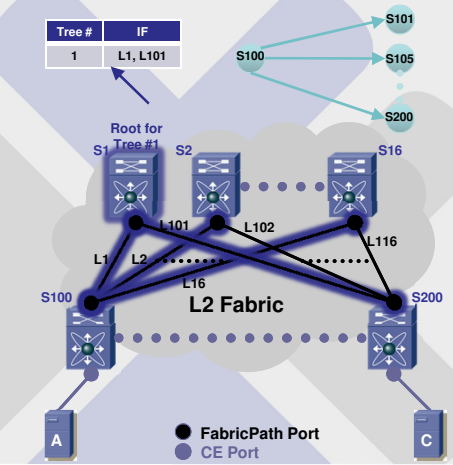
Switch	IF
S2	L1, …, L101
…	…
S16	L1, …, L101
S100	L1
S200	L101



- Known Unicast Destination
  - *One Ingress Switch ➔ One Egress Switch*
- Forwarding path selection based on destination Switch-ID inside FabricPath encapsulation imposed by Ingress Switch
- Each switch that is part of a L2 Fabric calculate it's local Switch Table based on the information received from L2 IS-IS
- Switch Table basically contains information about *{Switch-ID, Output Interfaces}*
- Up to 16 'Next-hop Interfaces' (i.e. L2 ECMP) can be programmed for a given Switch-ID

FabricPath 'Tree'

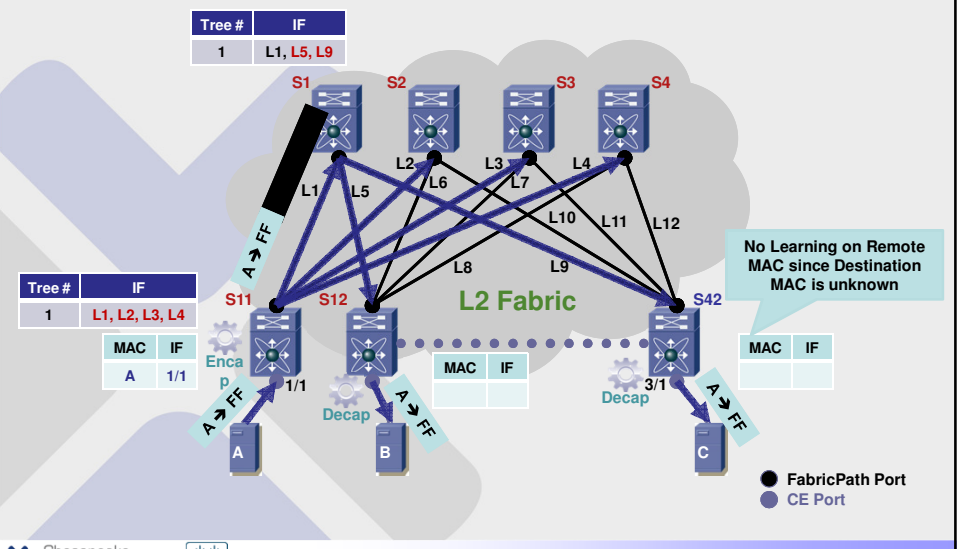
Used for forwarding L2 multi-destination traffic (Unknown Unicast, Broadcast, and Multicast) inside the L2 Fabric



- 'Tree' topology is required to forward multi-destination traffic properly
  - *One Ingress Switch ➔ Many Egress Switches*
- Same method is also used by L3 (e.g. PIM Source Tree/Shared Tree)
- One or more 'Root' devices are first elected for the L2 Fabric
- A 'Tree' spanning from each 'Root' is then formed and a network-wide unique ID is assigned to it
- Support for multiple 'Trees' allows Cisco FabricPath to support multipathing even for multi-destination traffic
- Ingress Switch determines the 'Tree' for each traffic flow

Chesapeake NETCRAFTSMEN

FabricPath Forwarding: Broadcast



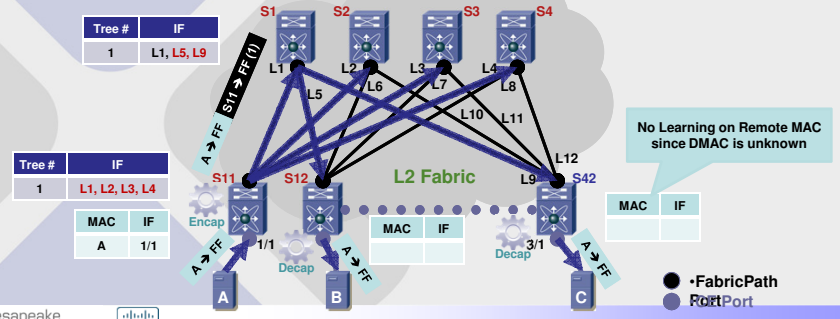
Chesapeake NETCRAFTSMEN

Detailed Version of Previous Slide

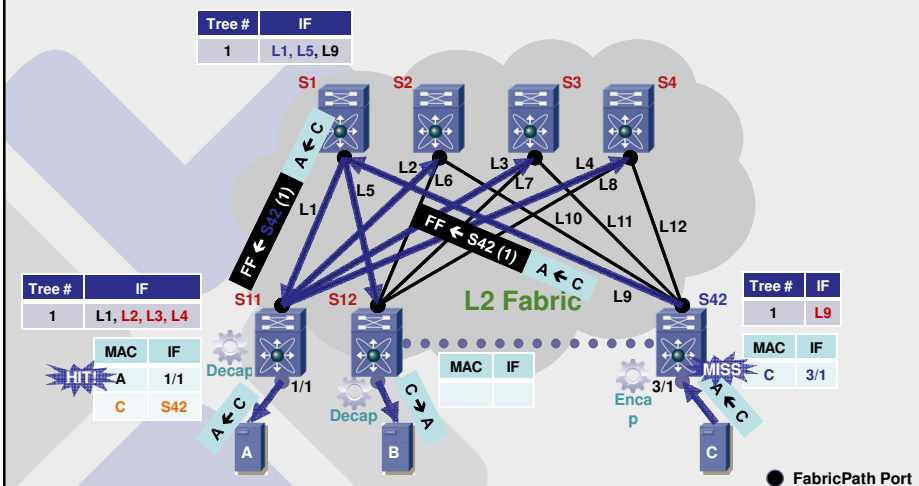
FabricPath Forwarding: Broadcast

Step-by-Step Details

1. Host A communicate to Host C for the first time. Send ARP Request to C
2. S11 add A into MAC Table as the result of new source learning on CE Ports
3. Since destination MAC is 'All F', S11 flood this frame out all CE Ports
4. Meanwhile, S11 select 'Tree 1', marks this in FabricPath header and floods this frame out all FabricPath ports (L1 ~ L4) that are part of 'Tree 1'
5. S1 flood this frame further (L5, L9) based on local info about 'Tree 1'
6. S12 and S42 remove FabricPath header and flood the frame out all local CE Ports



FabricPath Forwarding: Unknown Unicast

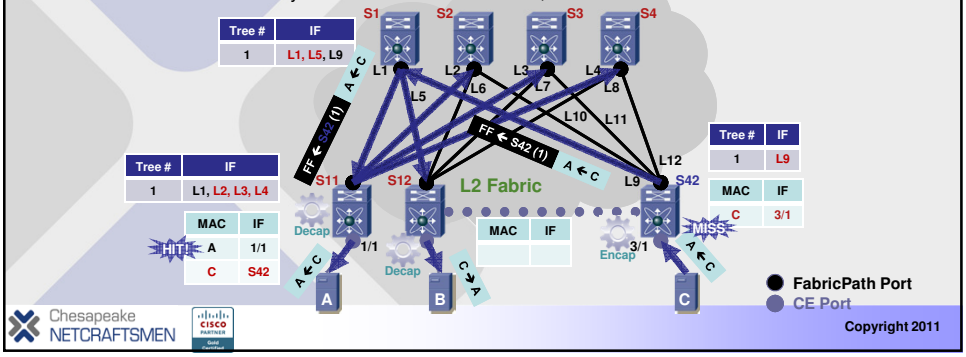


Detailed Version of Previous Slide

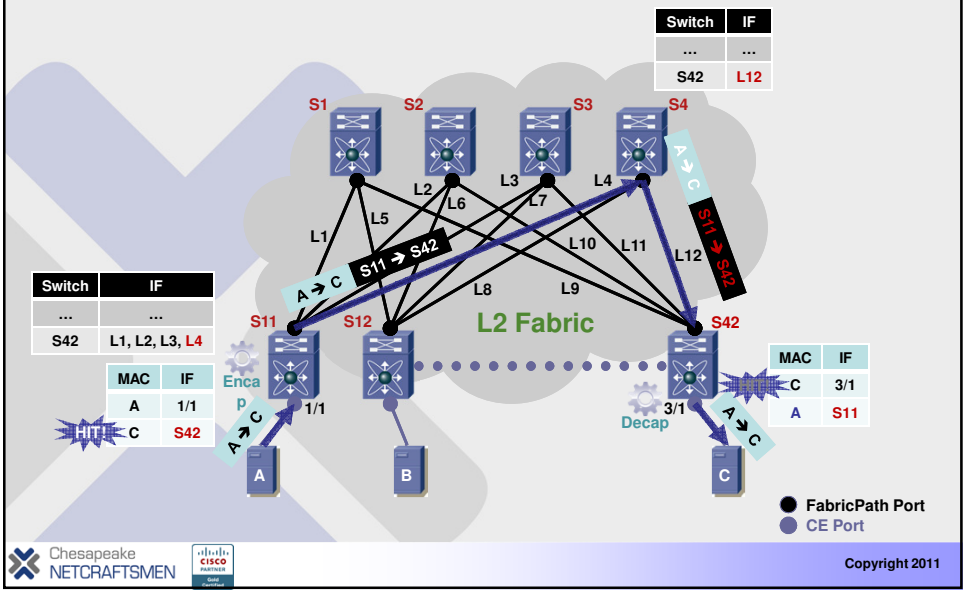
FabricPath Forwarding: Unknown Unicast

Step-by-Step Details

1. Host C sends ARP Reply back to Host A
2. S42 adds C into MAC Table from source learning on CE Port
3. Since A is unknown, S42 floods this frame out all CE Ports
4. Meanwhile, S42 selects 'Tree 1', marks this in FabricPath header and floods this frame out all FabricPath ports (L9) that are part of 'Tree 1'
5. S1 floods this frame further (L1, L5) along 'Tree 1'
6. S11 floods this frame further (L2~L4) along 'Tree 1'. Also, upon removing FabricPath header, S11 finds A was learned locally. Therefore adds C as remote, associated with S42.



FabricPath Forwarding: Known Unicast

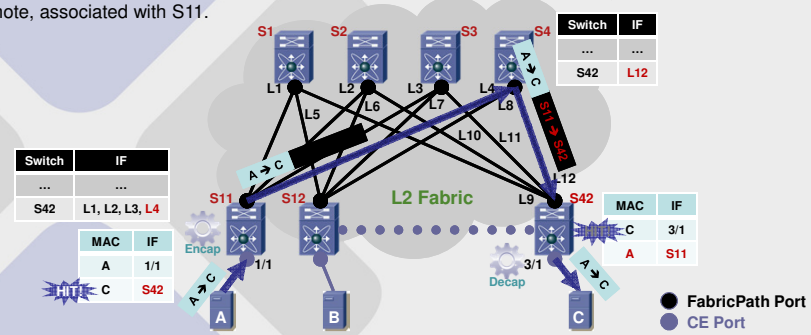


Detailed Version of Previous Slide

FabricPath Forwarding: Known Unicast

Step-by-Step Details

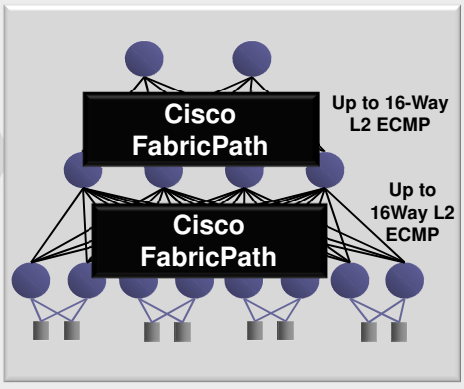
1. Host A starts sending traffic to Host C after ARP resolution
2. S11 finds C was learned as remote, associated with S42. Encap all subsequent frames to C with S42 as destination in FabricPath header
3. S11's Routing Table indicates multiple paths to S42. Runs ECMP hash and selects L4 as next-hop
4. Routing Table lookup at S4 indicates L12 as next hop for S42
5. S42 finds itself as destination in FabricPath header and C is also known locally. Therefore, adds A as remote, associated with S11.



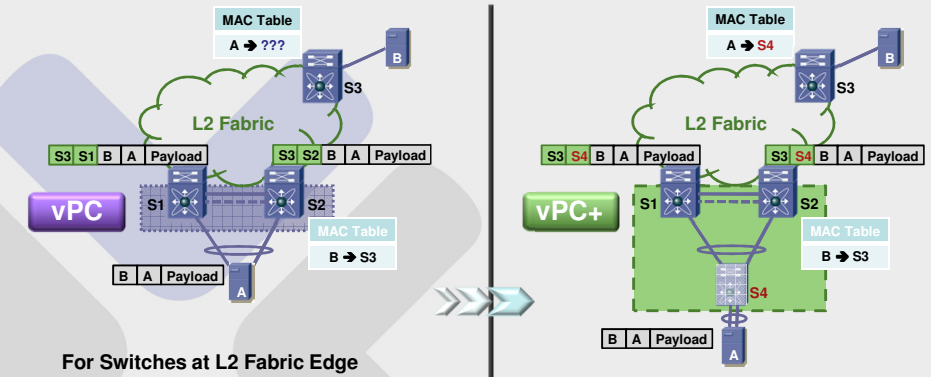
Cisco FabricPath Feature Set

Value-Add Enhancements

- **16-Way Equal Cost Multipathing (ECMP) at Layer 2**
- **FabricPath Header**
  - **Hierarchical Addressing with built in loop mitigation (RPF,TTL)**
- **Conversational MAC Learning**
  - **Efficient use of hardware resource by learning only MACs for interested hosts**
- **Interoperability with existing classic Ethernet networks**
  - ***VPC +* allows VPC into a L2 Fabric**
  - ***STP Boundary Termination***
- **Multi-Topology – providing traffic engineering capabilities**



vPC Enhancement for FabricPath



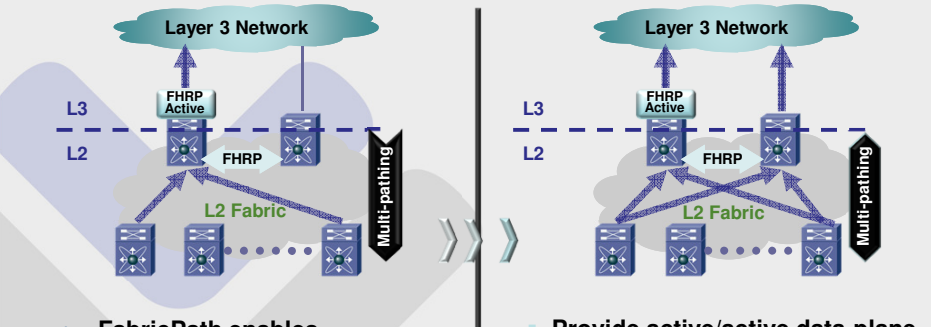
For Switches at L2 Fabric Edge

- vPC is still required to provide active/active L2 paths for dual-homed CE devices or clouds
- However, MAC Table only allows 1-to-1 mapping between MAC and Switch ID

- Each vPC domain is represented by an unique 'Virtual Switch' to the rest of L2 Fabric
- Switch ID for such 'Virtual Switch' is then used as Source in FabricPath encapsulation

Chesapeake NETCRAFTSMEN

Connect L3 or Services to L2 Fabric

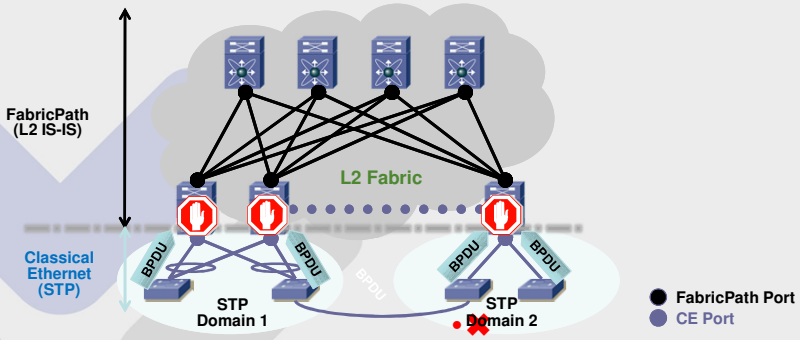


- FabricPath enables multipathing for bridged traffic
- However, FHRP allows only 1 active gateway for each host, therefore prevent traffic that needs to be routed to take advantage of multi-pathing

- Provide active/active data-plane for FabricPath with no change to existing FHRP
- Allow multi-pathing even for routed traffic
- Same feature can be leveraged by service nodes as well

Chesapeake NETCRAFTSMEN

STP Boundary Termination

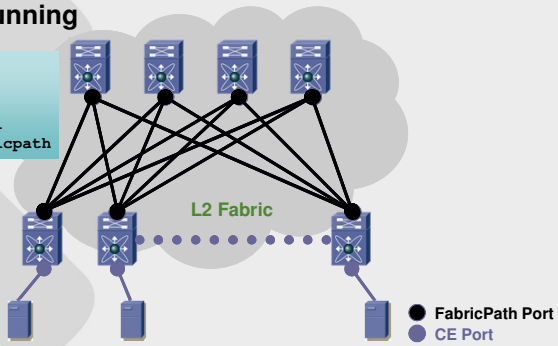


- A L2 Fabric is presented as a single bridge to all connected CE devices
- STP BPDUs are processed and terminated by CE Ports
- CE devices not interconnected will form separate STP domains
- Loops outside L2 Fabric will be blocked within each STP domain
- *L2 Fabric should be the root for all connected STP domain.* CE ports will be put into blocking state when 'superior BPDU' is received

FabricPath Configuration

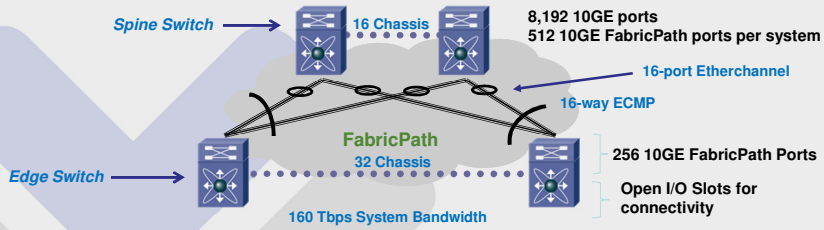
- No L2 IS-IS configuration required
- New 'feature-set' keyword allows multiple conditional services required by FabricPath (e.g. L2 IS-IS, LLDP, etc.) to be enabled in one shot
- Simplified operational model – only 3 CLIs to get FabricPath up and running

```
N7K(config)# feature-set fabricpath
N7K(config)# vlan 10-19
N7K(config-vlan)# mode fabricpath
N7K(config)# interface port-channel 1
N7K(config-if)# switchport mode fabricpath
```



Use Case: High Performance Compute

Building Large Scalable Compute Clusters



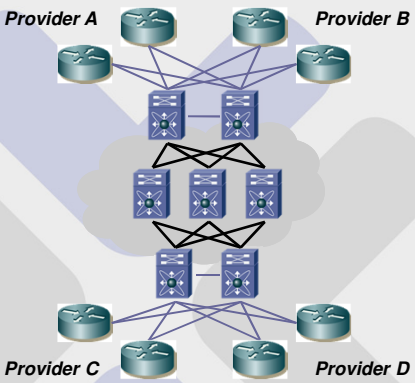
HPC Requirements

- **HPC Clusters require high-density of compute nodes**
- **Minimal over-subscription**
- **Low server to server latency**

FabricPath Benefits for HPC

- **FabricPath enables building a high-density fat-tree network**
- **Fully non-blocking with FabricPath ECMP & port-channels**
- **Minimize switch hops to reduce server to server latencies**

Use Case: L2 Internet Exchange Point

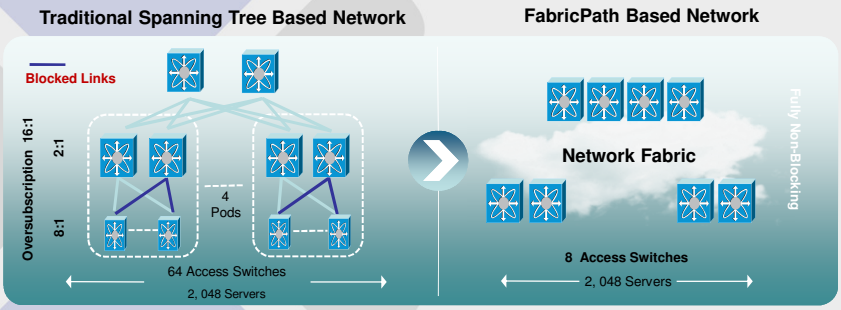


- ***IXP Requirements***
- Layer 2 Peering enables multiple providers to peer their internet routers with one another
- 10GE non-blocking fabric
- Scale to thousands of ports
- ***FabricPath Benefits for IXP***
- Transparent Layer 2 fabric
- Scalable to thousands of ports
- Bandwidth not limited by chassis / port-channel limitations
- Simple to manage, economical to build

Scaling Bandwidth with FabricPath

Example: 2,048 X 10GE Server Design

- 16X improvement in bandwidth performance
- From 74 managed devices to 12 devices
- 2X+ increase in network availability
- Simplified IT operations

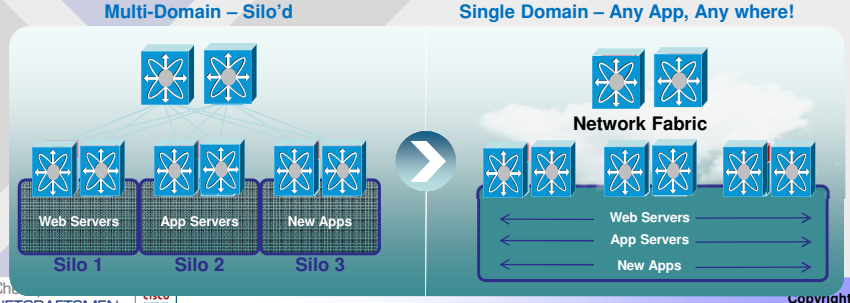


Chesapeake NETCRAFTSMEN

Workload Flexibility with FabricPath

Example: Removing Data Center Silos

- **Single domain**
  - Pooled compute resources
- **Increased agility**
  - Seamless data center wide workload mobility
- **Responsive**
  - Virtualized Applications move within minutes vs. days
- **Capex and Opex savings**
  - Maximize resource utilization, simplify IT operations



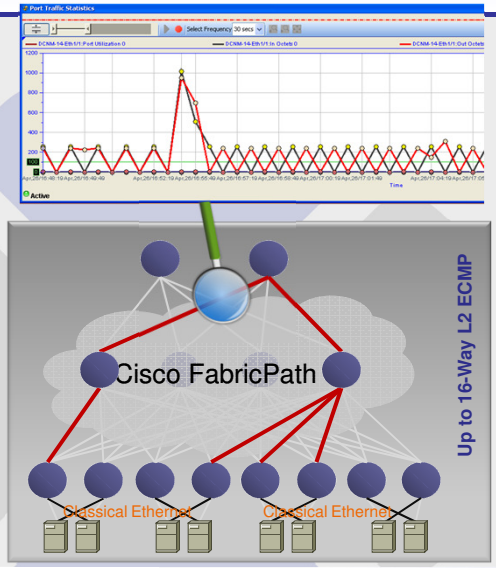
Agenda

- Introduction
- FabricPath
  - Why FabricPath
  - FabricPath Forwarding Details
  - Monitoring and Troubleshooting
- Fiber Channel / SAN for Network Engineers
- Fiber Channel over Ethernet (FCoE)
  - IO Consolidation
  - FCoE Technology
  - Basic Designs
- Recent Announcements
- Summary, References, and Q&A

Troubleshooting FabricPath

- Leverage the same tooling for L3 technologies
  - ✓ Routing table
  - ✓ Link-state database
  - ✓ Distribution trees
  - ✓ ECMP path selection
- Pong – L2 Ping + Traceroute
  - ✓ Provide info on all devices on a given path in L2 Fabric
  - ✓ Check on link health
- Performance Profiling across FabricPath
  - Through IEEE 1588 timestamp and pong to help estimate average end-to-end latency

FabricPath: In Control with DCNM

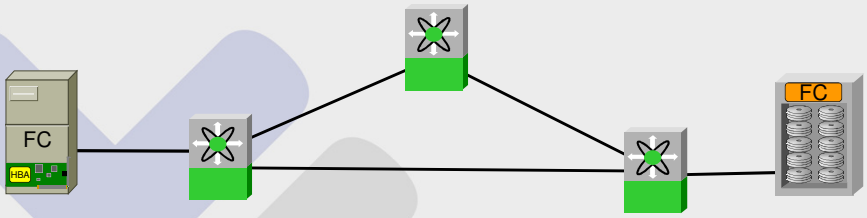


- Abstracted Fabric View
  - ✓ Identify fabric 'hot-spots'
  - ✓ FabricPath state awareness
- Traffic Monitoring
  - ✓ Frames distribution visibility
  - ✓ Threshold crossing alerts for bandwidth management
- Troubleshooting
  - ✓ Visualize unicast, multicast and broadcast paths
  - ✓ Check reachability between source and destination nodes
- Configuration Expert
  - ✓ Manage FabricPath topologies with Wizard tools
  - ✓ Simplify fine-tuning FabricPath

Agenda

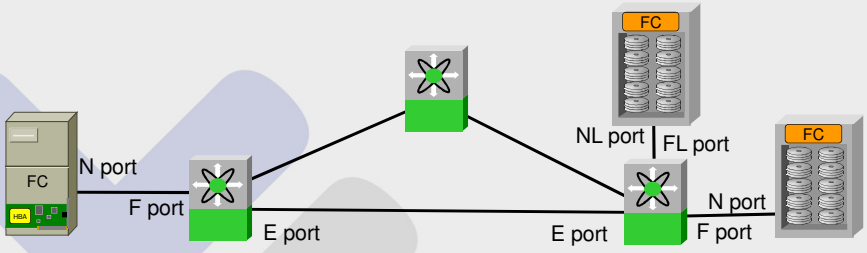
- Introduction
- FabricPath
  - Why FabricPath
  - FabricPath Forwarding Details
  - Monitoring and Troubleshooting
- Fiber Channel / SAN for Network Engineers
- Fiber Channel over Ethernet (FCoE)
  - IO Consolidation
  - FCoE Technology
  - Basic Designs
- Recent Announcements
- Summary, References, and Q&A

Fiber Channel Introduction



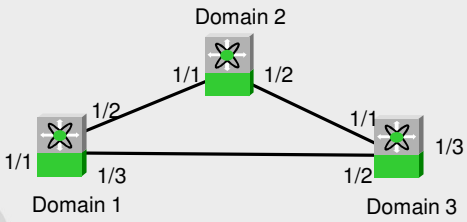
- Why do FC / SAN?
  - Consistent storage management and backup
  - Reduce size of storage distributed inside server chassis
  - Separate highly reliable environment
    - If SAN goes down, your business is down!
  - VMotion requires SAN

Port Naming

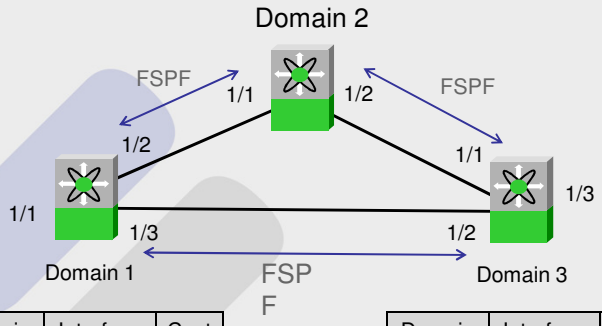


FC Domains

- Principal Switch Selection (PSS)
  - BF SW_ILS type frame sent by each SAN switch, wait while flood
  - Flood EFP (Exchange Fabric Parameters) SW_ILS, lowest wins, flood revised EFP…
  - Domain_ID Distribution process
  - PS transmits DIA SW_ILS, immediate recipients reply, PS assigns Domain ID ("DID")
  - Work outwards in "rings" (hops): PS sends DIA, passed along, receiver sends back request, gets DID
- Domain ID
  - Max of 239
  - Usually 2, 40 or so is a huge SAN
  - Local to a SAN fabric
- ISL

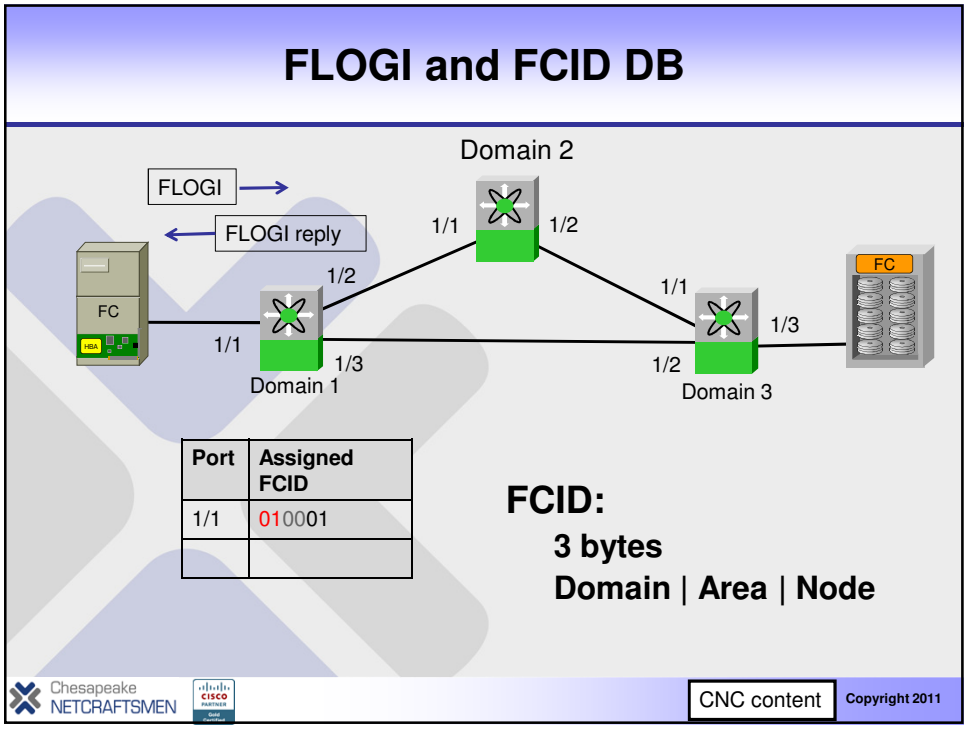


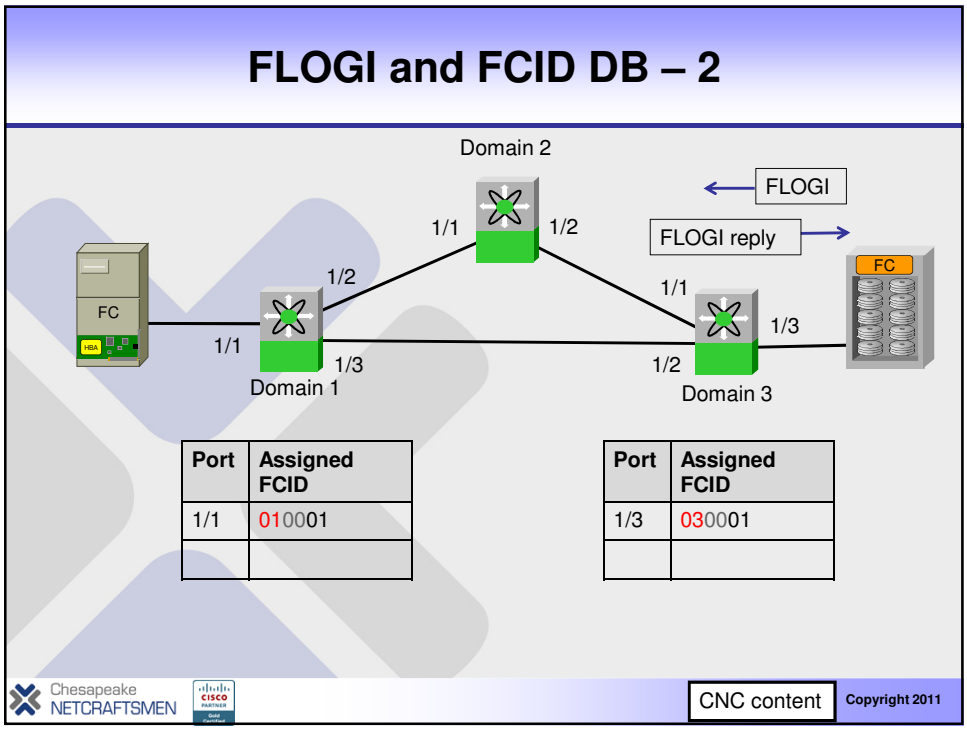
FC FSPF Routing



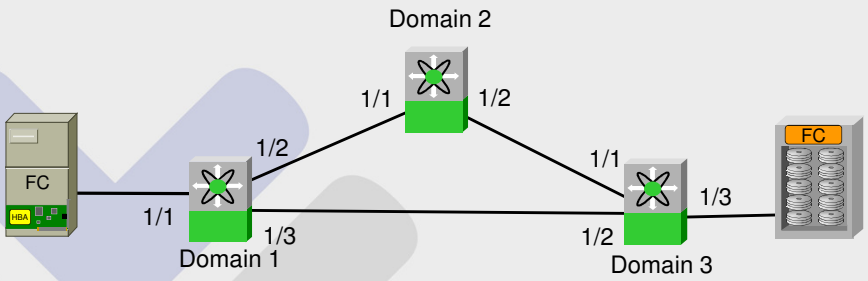
Domain	Interface	Cost
2	1/2	1
3	1/3	1

Domain	Interface	Cost
1	1/2	1
2	1/1	1



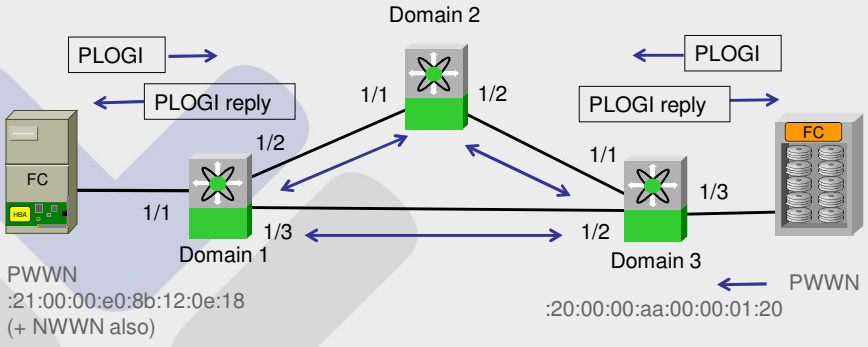


WWN's and FCNS DB



- Each FC port has a Port World Wide Name (PWWN or WWPN)
- Each FC node (device) has a Node WWN (NWWN or WWNN) to uniquely ID it
- While WWN may be burned in, it behaves more like a NAME (DNS name) than a BIA MAC address

WWN's and FCNS DB – 2



FCNS DB (aka Name Server functionality)

Assigned FCID	PWWN
010001	:21:00:00:e0:8b:12:0e:18
030001	:20:00:00:aa:00:00:01:20

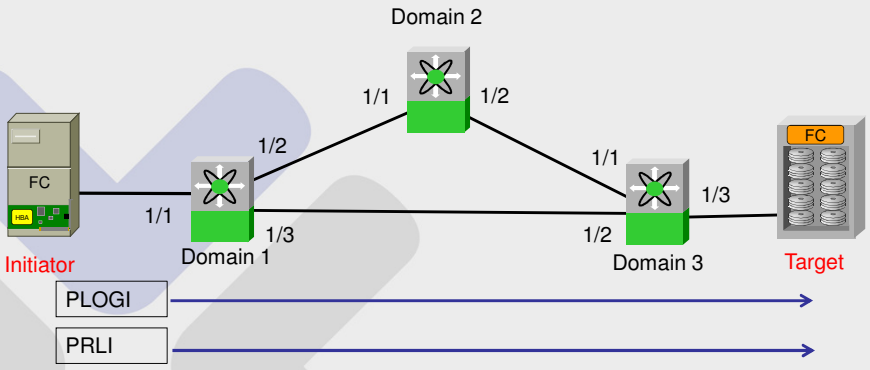
What is Zoning?

- A zone is a list of FC devices that are allowed to communicate over FC
  - Think "these servers can talk to these disks"
  - Might be based on WWN's, or on many of the other attributes available
- A zoneset is a list of zones, e.g.
  - Servers A, B, C can talk to disks D, E, F
  - Servers G, H can talk to disks D, I, J, K, and L
  - Etc.
- Analogy: zone = ACE (ACL rule), zoneset = ACL

LUN Masking

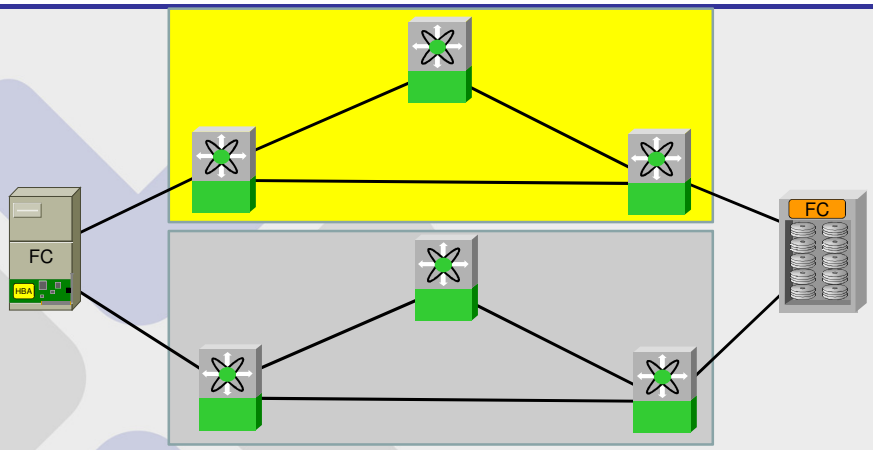
- Masking is which LUN's are visible where, i.e. perhaps the server masks as a way of limiting the LUN's that it is informed about

Connecting to Storage



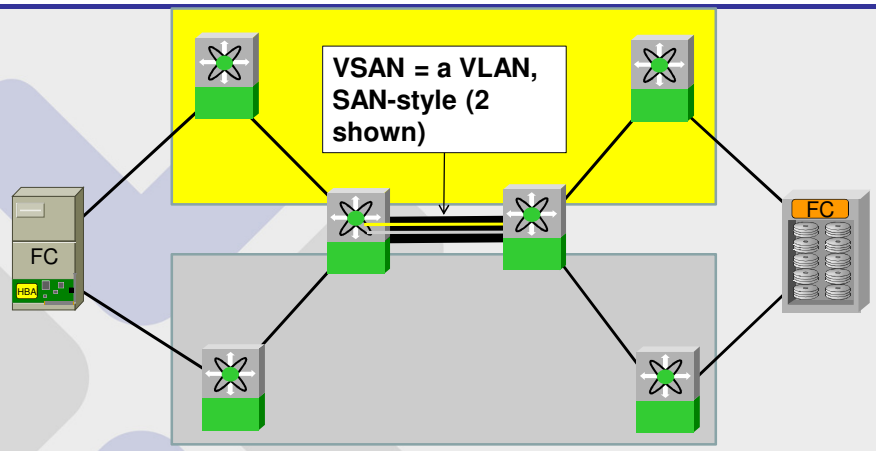
- Initiator configured w/ NWWN of Target (manual or pick list or targets in **zone**)
- Name resolution provides FCID of Target
- Packets FC routed between Initiator and Target
- PLOGI, PRLI

FC Fabric



- **FC needs no STP – routing!**
- **If interconnect two fabrics, you have fabric MERGE, it becomes one big (unhappy?) fabric**

VSAN, SAN Fabric, EISL



- EISL = Extended Inter-Switch Link = SAN trunk carrying VSANs
- Fabric = interconnected SAN switches with common PS and shared routing (using dedicated or EISL links)

Terminology

- LUN: basically a chunk of disk space
  - What the Initiator is trying to access on the Target
- IVR: routing between VSANs
  - Allows failure domain control and security by using smaller FC fabrics
  - Works by passing permitted traffic between fabrics, e.g. to a shared backup fabric
- JBOD: Just a Bunch of Disks

Terminology – 2

- NPIV: server / VMware technique, WWN's per VM, and several FCID's on shared N-port via FLOGI then FDISC – allows per-VM FC security controls
- NPV: N5K no longer a FC switch, reduces domain usage, leaves control with FC admins, improves scalability
  - N5K behaves like a VMware chassis doing NPV, in effect, proxy for attached VMware server(s)
  - NPiV and NPV require enabling NP(i)V "server" function on MDS FC switch in the FC network

SNIA = Storage Networking Industry Association

- Good source of generic SAN training & certs
  - www.snia.org
  - FCoE presentation: http://www.snia.org/education/tutorials/2008/spring/networking/Hufferd-J_Fibre_Channel_Over_Ethernet.pdf
- From www.t11.org/ftp/t11/pub/fc/bb-5/08-264v3.pdf:

Table 1 – Defined Constants

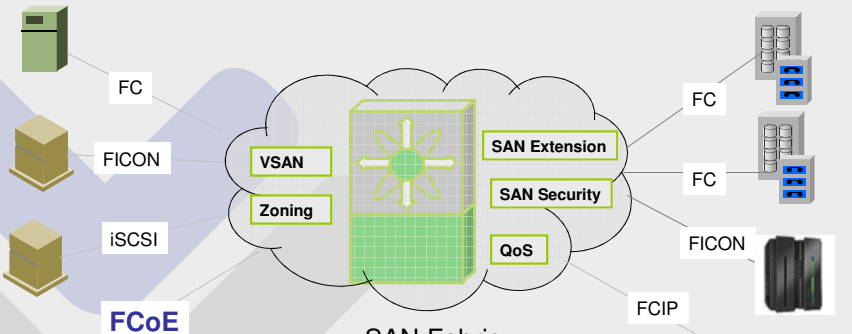
Constant	Value	Description
FIP_TYPE	8914h	Value to be used in the Type field of the 802.3 frame to indicate a FIP payload
FCoE_Type	8906h	Value to be used in the type field of the 802.3 frame to indicate an FCoE payload
ALL_FCF_MACS	01-10-18-01-00-02	Group address for all FCFs
ALL_ENODE_MACS	01-10-18-01-00-01	Group address for all ENodes

SAN Book / Protocol Details

Agenda

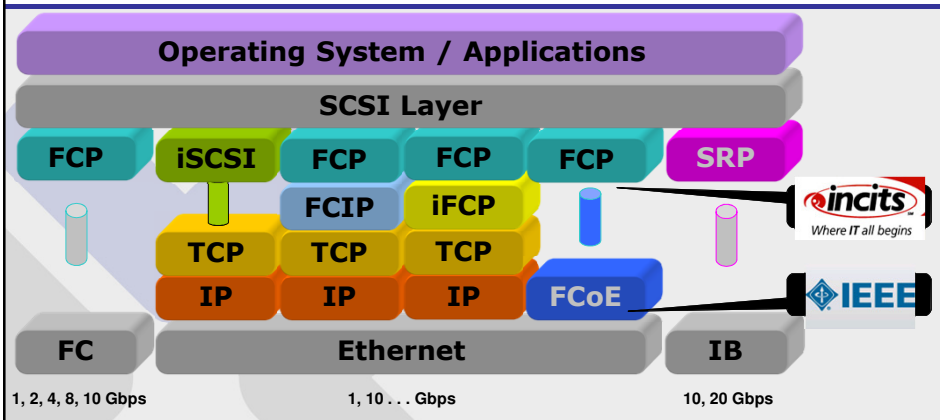
- Introduction
- FabricPath
  - Why FabricPath
  - FabricPath Forwarding Details
  - Monitoring and Troubleshooting
- Fiber Channel / SAN for Network Engineers
- Fiber Channel over Ethernet (FCoE)
  - IO Consolidation
  - FCoE Technology
  - Basic Designs
- Recent Announcements
- Summary, References, and Q&A

FCoE Connectivity Extends FC SANs



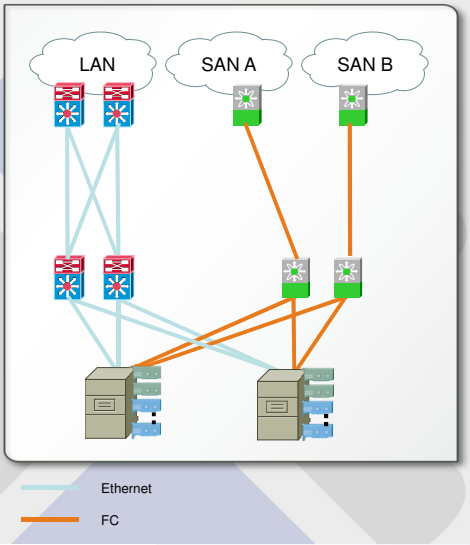
- Preserves FC investments
- Simplifies SAN-attach of servers

Encapsulation Technologies



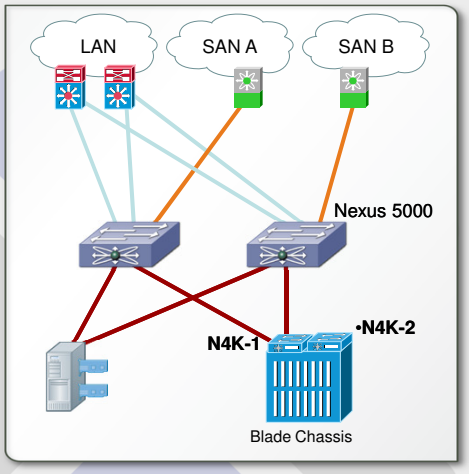
•FCoE is non routable, localised DC transport solution with lower protocol overhead than FCIP or iSCSI

Before I/O Consolidation



- Parallel LAN/SAN Infrastructure
- Inefficient use of Network Infrastructure
- 5+ connections per server—higher adapter and cabling costs
  - Adds downstream port costs; cap-ex and op-ex
  - Each connection adds additional points of failure in the fabric
- Longer lead time for server provisioning
- Multiple fault domains—complex diagnostics
- Management complexity

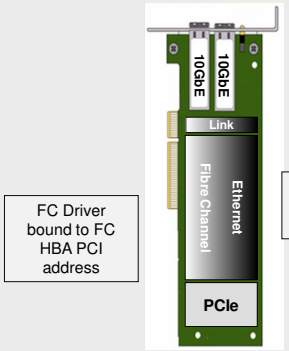
I/O Consolidation



- Reduction of server adapters
- Simplification of access layer and cabling
- Gateway free implementation—fits in installed base of existing LAN and SAN
- L2 Multipathing Access—Distribution
- Lower total cost of ownership
- Fewer cables
- Investment protection (LANs and SANs)
- Consistent operational model

Converged Network Adapters (CNA)

- Replaces multiple adapters per server, consolidating both Ethernet and FC on a single interface
- Appears to the operation system as individual interfaces (NICs and HBAs)



FC Driver bound to FC HBA PCI address

Ethernet Driver bound to Ethernet NIC PCI address

Fibre Channel Drivers

Ethernet Drivers

Operating System

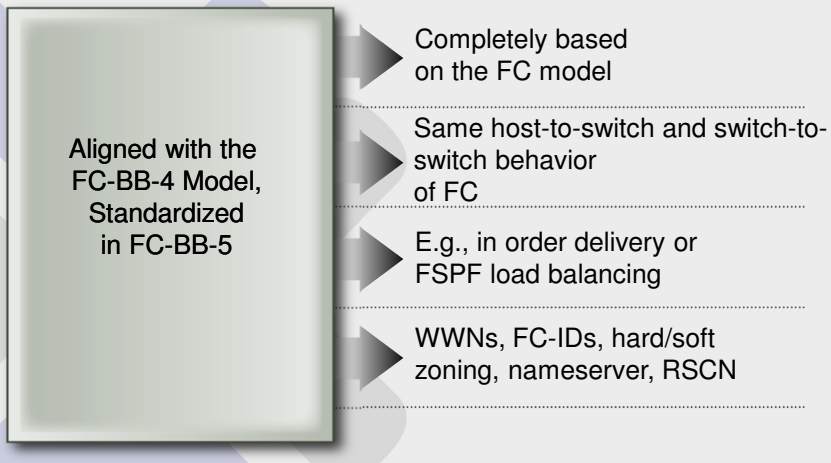
Why Virtualize with Converged Fabric?

- Consolidates separate LAN, SAN, and server cluster network environments into a single unified fabric
- Same operational model as today
- Wire once infrastructure – the road to stateless computing
- Lower IT total cost of ownership
- Unified fabric with Fibre Channel over Ethernet (FCoE) for I/O consolidation, reducing power and cabling requirements and simplifying data center networks, especially for SAN consolidation of Fibre Channel

Unified Fabric

Fibre Channel over Ethernet (FCoE)

FCoE *is* Fibre Channel.



Standards for I/O Consolidation

- **Developed by IEEE 802.1 Data Center Bridging Task Group (DCB)**
- **All technically stable**
- **Final standards expected by mid 2010**

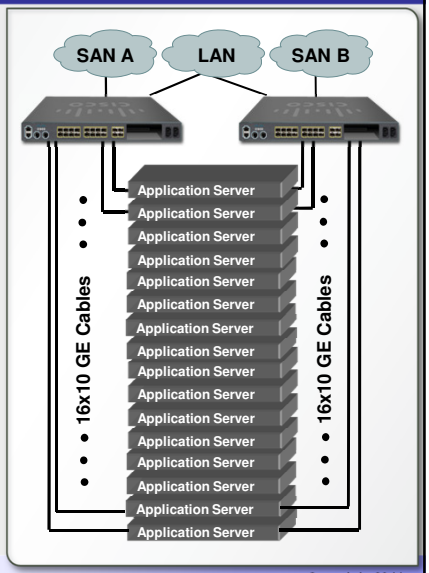
Standard / Feature	Status of the Standard
IEEE 802.1Qbb Priority-based Flow Control (PFC)	Completed first sponsor ballot. Draft 2.3 available. Basically done.
IEEE 802.3bd Frame Format for PFC	Completed first sponsor ballot. Draft 2.1 available. Basically done.
IEEE 802.1Qaz Enhanced Transmission Selection (ETS) and Data Center Bridging eXchange (DCBX)	Completed WG recirculation ballot. Draft 1.4 available. Going in sponsor ballot after the July plenary.
T11 FC-BB-5 FCoE Standard	Approved on June 3rd 2009.

Why 10 Gigabit Ethernet to the Server?

- Multi core CPU architectures driving increased network bandwidth demands
- Virtual Machines driving increased I/O connections and I/O bandwidth per server
- Low latency 10GE affordability (even optics…)
- Increased adoption of NAS (NFS/CIFS) and iSCSI
- Consolidation of networks
  - Unified Fabrics with ethernet
- Ubiquity of large scale Ethernet networks
  - Extensive range of management, diagnostic and troubleshooting tools
  - Massive base of manufacturers, suppliers and integrators
    - → competition, price, services and innovation

Twinax Copper Cable

- Low power consumption
- Low cable cost
  - SFP+ CX-1 Copper (SFF 8431)
- Low transceiver latency
- Low error rate (10^{17})
- Thinner cable with higher bend radius
  - Supports 10GE passive direct attached up to 10 meters
- Easier to manage cabling solution reduces deployment time
- All copper cables are contained within rack



10 Gigabit Ethernet Media

10G Options

In-rack & Cross-rack

Connector (Media)	Cable	Distance	Power (each side)	Transceiver Latency (link)	Standard
SFP+ CU* copper	Twinax	<10m	~ 0.1W	~ 0.1µs	SFF 8431**
X2 CX4 copper	Twinax	15m	4W	~ 0.1µs	IEEE 802.3ak
SFP+ USR MMF, ultra short reach	MM OM2 MM OM3	10m 100m	1W	~ 0	none
SFP+ SR MMF,short reach	MM OM2 MM OM3	82m 300m	1W	~ 0	IEEE 802.3ae
RJ45 10GBASE-T copper	Cat6 Cat6a/7 Cat6a/7	55m 100m 30m	~ 6W*** ~ 6W*** ~ 4W***	2.5µs 2.5µs 1.5µs	IEEE 802.3an

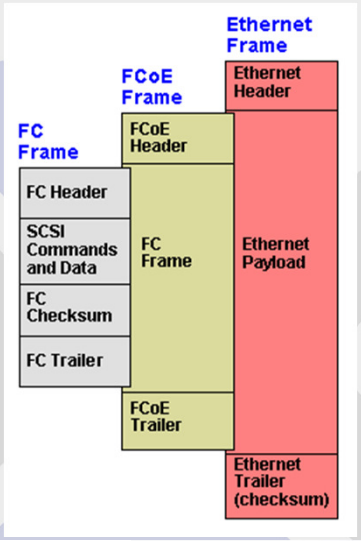
* Terminated cable ** Draft *** As of 2008; expected to decrease over time

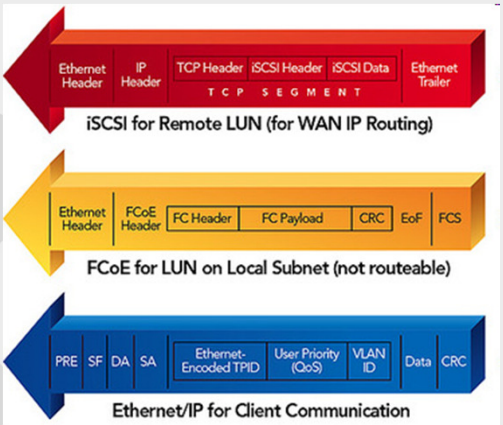
Across racks

Agenda

- Introduction
- FabricPath
  - Why FabricPath
  - FabricPath Forwarding Details
  - Monitoring and Troubleshooting
- Fiber Channel / SAN for Network Engineers
- Fiber Channel over Ethernet (FCoE)
  - IO Consolidation
  - FCoE Technology
  - Basic Designs
- Recent Announcements
- Summary, References, and Q&A

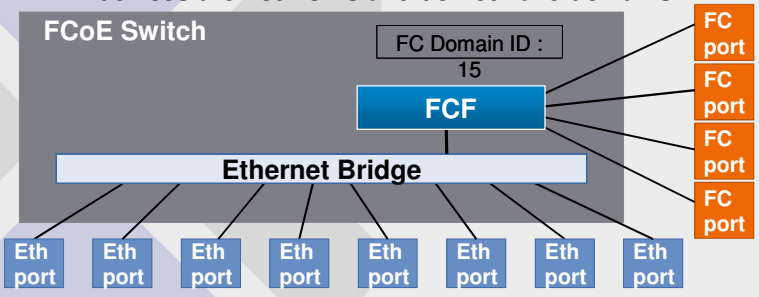
FCoE Frame versus iSCSI and Ethernet





FCoE Forwarding

- FCF (Fibre Channel Forwarder) is the forwarding entity inside an FCoE switch
  - Fibre Channel login happens at the FCF
  - contains an FCF-MAC address
  - consumes a Domain ID
- FCoE encap/decap happens within the FCF
- NPV devices are *not* FCF's and do *not* have domains



Protocol Organization

FCoE is really two different protocols:

FCoE itself

- Is the data plane protocol
- It is used to carry most of the FC frames and all the SCSI traffic

FIP (FCoE Initialization Protocol)

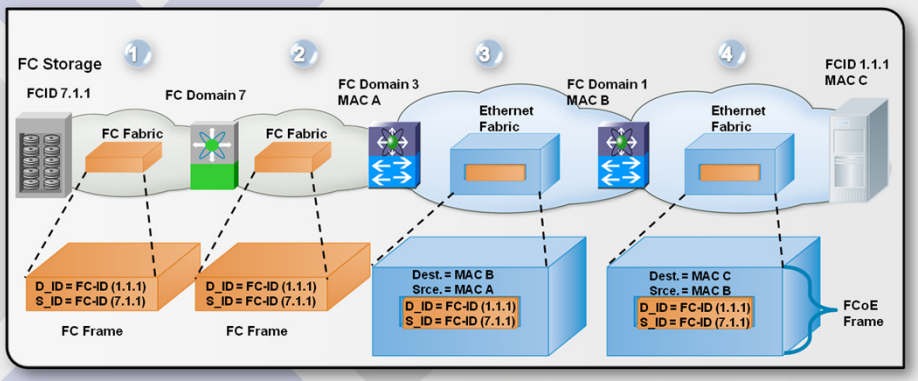
- It is the control plane protocol
- It is used to discover the FC entities connected to an Ethernet cloud
- It is also used to login to and logout from the FC fabric
- Uses unique BIA on CNA for MAC

- ***The two protocols have:***
  - Two different Ethertypes
  - Two different frame formats
  - Both are defined in FC-BB-5

http://www.cisco.biz/en/US/prod/collateral/switches/ps9441/ps9670/white_paper_c11-560403.html

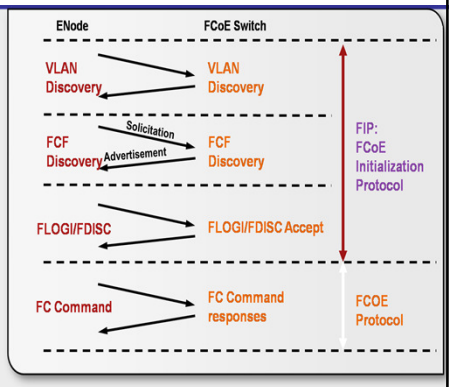
FCoE Addressing and Forwarding

- **FCoE frames have:**
  - **MAC addresses (hop-by-hop)**
  - **FC addresses (end-to-end)**



FIP: FCoE Initialization Protocol

- FCoE VLAN discovery
  - Automatic discovery of FCoE VLANs
- Device discovery
  - ENodes discover VF_Port capable FCF-MACs for VN_Port to VF_Port Virtual Links
  - VE_Port capable FCF-MACs discover other VE_Port capable FCF-MACs for VE_Port to VE_Port Virtual Links
  - The protocol verifies the Lossless Ethernet network supports the required Max FCoE Size

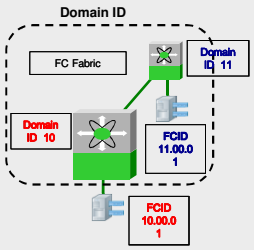


- **Virtual Link instantiation**
  - Builds on the existing Fibre Channel Login process
  - Fabric Provided MAC Address (FPMA),
- **Virtual Links maintenance**
  - Timer based

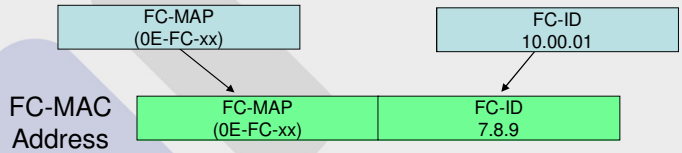
Enode MAC Address

Fibre Channel over Ethernet Addressing Scheme

- Enode MAC assigned for each FCID
- Enode MAC composed of a FC-MAP and FCID
  - FC-MAP is the upper 24 bits of the Enode's MAC
  - FCID is the lower 24 bits of the Enode's MAC
- FCoE forwarding decisions still made based on FSPF and the FCID within the Enode MAC



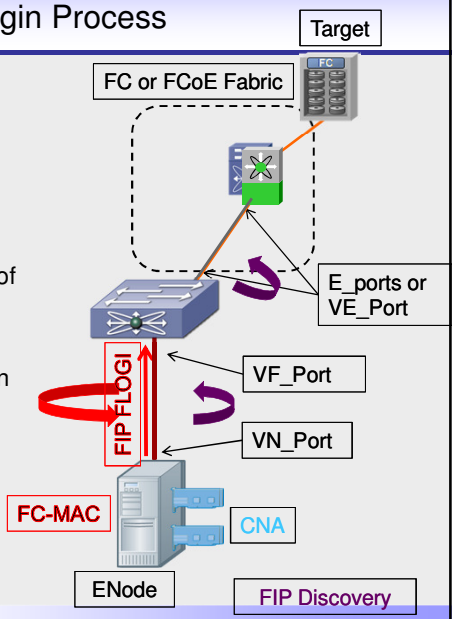
Fibre Channel FCID Addressing



My port is up…can I talk now?

FIP and FCoE Login Process

- Step 1: FIP Discovery Process
  - Enables FCoE adapters to discover which VLAN to transmit & receive FCoE frames
  - Enables FCoE adapters and FCoE switches to discovers other FCoE capable devices
  - Verifies Lossless Ethernet is capable of FCoE transmission
- Step 2: FIP Login Process
  - Similar to existing Fibre Channel Login process - sends FLOGI to upstream FCF
  - Adds the negotiation of the MAC address to use Fabric Provided MAC Address (FPMA)
  - FCF assigns the host a Enode MAC address to be used for FCoE forwarding

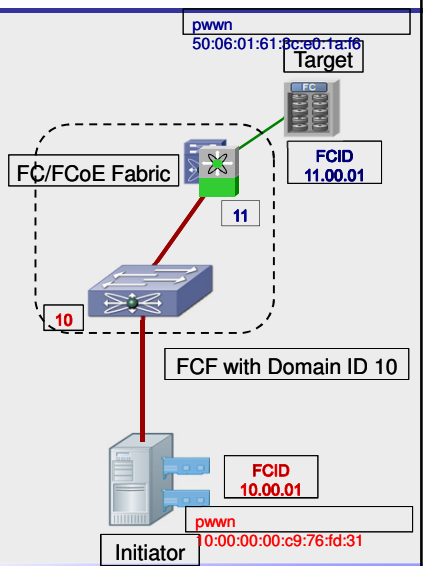


Login complete…almost there

Fabric Zoning

- FCoE fabric zoning done the same as FC fabric zoning
- Zoning is enforced at the FCF
- Zoning can be configured on the Nexus 5000 using the CLI or Fabric Manager
- If Nexus 5000 is in NPV mode, zoning will be configured on the upstream core switch and pushed to the Nexus 5000

fcid 0x10.00.01 [pwwn 10:00:00:00:c9:76:fd:31] [initiator]
fcid 0x11.00.01 [pwwn 50:06:01:61:3c:e0:1a:f6] [target]



Agenda

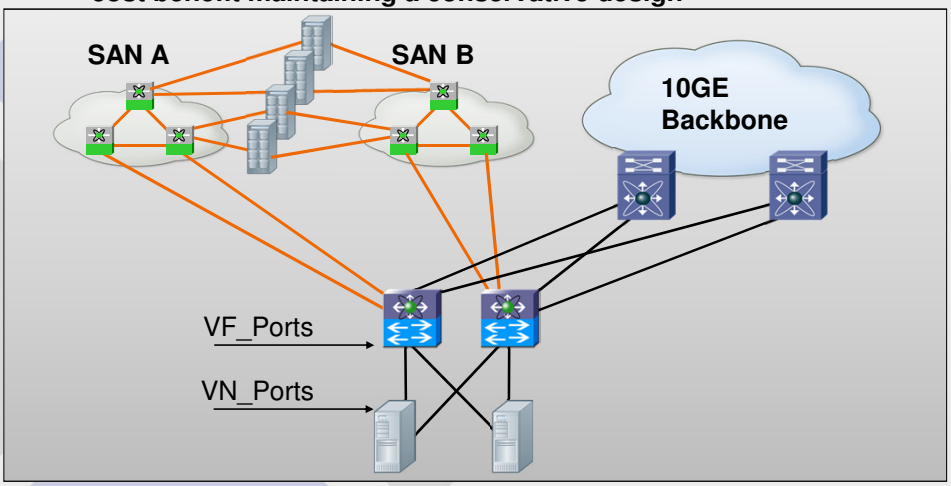
- Introduction
- FabricPath
  - Why FabricPath
  - FabricPath Forwarding Details
  - Monitoring and Troubleshooting
- Fiber Channel / SAN for Network Engineers
- Fiber Channel over Ethernet (FCoE)
  - IO Consolidation
  - FCoE Technology
  - Basic Designs
- Recent Announcements
- Summary, References, and Q&A

Nexus 5000 FC Module Modes

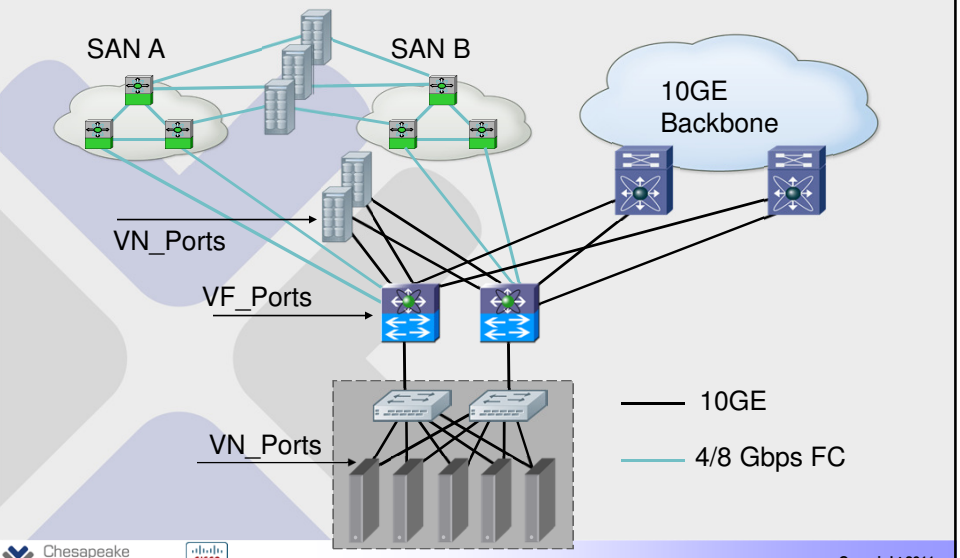
- Nexus 5000 supports two modes of operation
  - "Switch" mode
  - "N-Port Virtualization (NPV)" mode
- Modes of operation describe how native FC traffic is handled and therefore requires an FC module
- Modes are *independent* of the Nexus 5000 acting as an FCF or FCoE Pass-Through for FCoE traffic
- Modes are consistent with current FC edge switches (MDS 9124/9134/9222i)

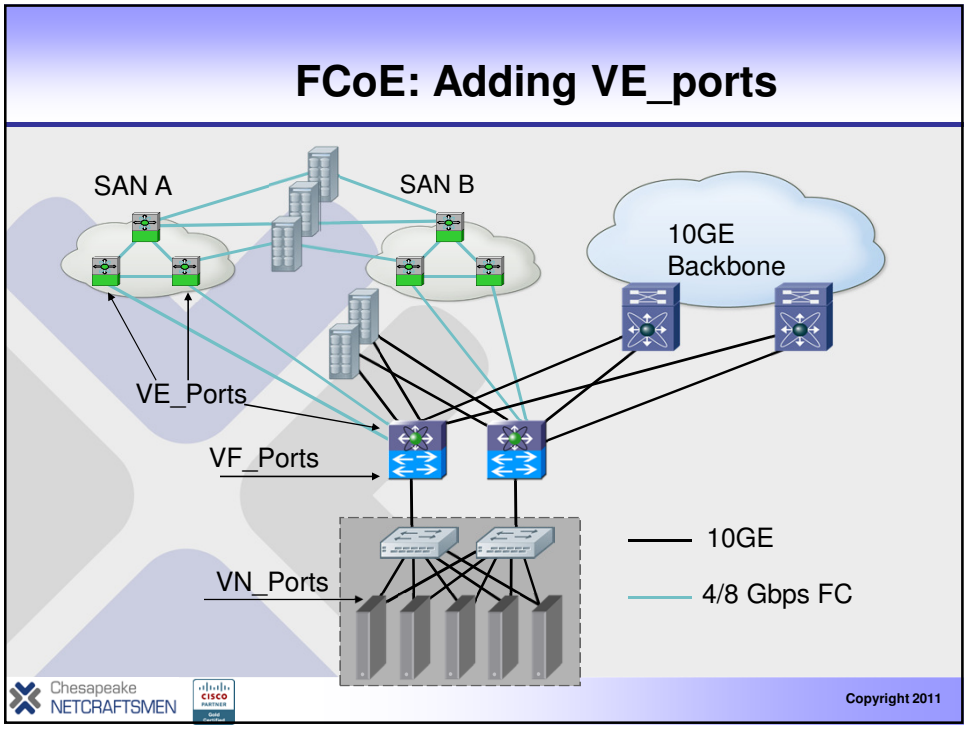
FCoE: Initial Deployment

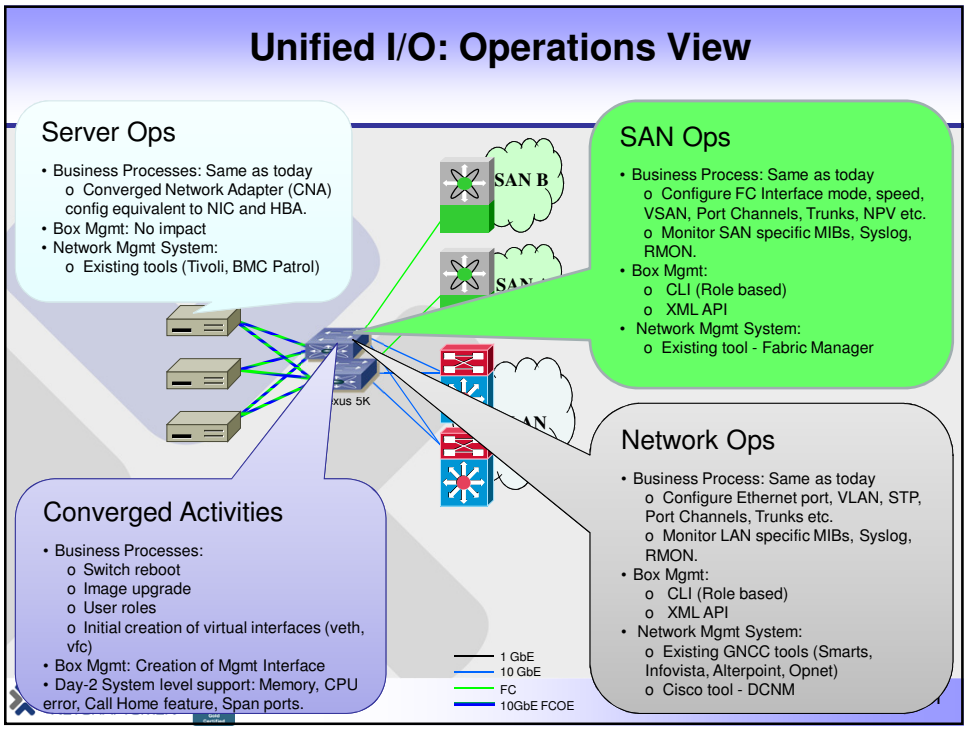
- The first step is "access consolidation": it provides significant cost benefit maintaining a conservative design



FCoE: Adding Native FCoE Storage







Agenda

- Introduction
- FabricPath
  - Why FabricPath
  - FabricPath Forwarding Details
  - Monitoring and Troubleshooting
- Fiber Channel / SAN for Network Engineers
- Fiber Channel over Ethernet (FCoE)
  - IO Consolidation
  - FCoE Technology
  - Basic Designs
- Recent Announcements
- Summary, References, and Q&A

Cisco Announcements, March 31, 2011

- 5548 UP and 5596 UP switches
  - UP = unified port, 1/10 Gbps or FC (2/4/8 Gbps), or FCoE
  - L2 or L3, FabricPath ready (pending software)

Cisco Announcements, March 31, 2011 – 2

- 3000 series switches
  - 1 RU, 1/10 Gbps ports
  - < 1 microsecond latency port to port
  - L2 / L3, NX-OS
  - For high-volume financial, etc.
  - See http://www.cisco.com/en/US/partner/prod/collateral/switches/ps9441/ps11541/data_sheet_c78-651097.html

- FEX features:
  - FEX w/ N7K (2232 PP)
  - Adapter FEX (Cisco VIC, aka "Palo")
  - VM FEX
  - Basically, N5K configures VIC, switching offloaded to N5K hardware (as with N2K)

Cisco Announcements, March 31, 2011 – 3

- Multi-hop FCoE
  - N7K (via F1 card), MDS 9500
  - Storage VDC
  - See http://www.cisco.com/en/US/partner/solutions/collateral/ns340/ns517/ns224/ns945/ns1060/guide_c07-648629_ns1122_Networking_Solutions_White_Paper.html

Multi-hop FCoE Implications

- New sales angle:
  - N7K as "Director class" SAN switch combining FC and FCoE
  - MDS can now do both as well (interoperation, migration)
  - Can do multi-hop VE-VE trunks to tie the two together
- Reliability and manageability is part of the recommended design approaches:
  - Separate storage VDC splits FC off FCoE from Ethernet, which goes to a different VDC, different admin
  - Separate uplinks for FCoE above the access layer

Multi-Hop FCoE: HP versus Cisco

- HP is throwing stones at Cisco for "not fully supporting end-end congestion notification". QCN, or 802.1Qau.
- Cisco envisions all nodes in multi-hop being FCF's, i.e. de-encapsulating and routing and doing local per-priority pause to avoid drops
  - End-end controls don't work in that environment (source MAC not available, it's routing hops)
  - Cisco claims (probably correctly) that nobody (in their right mind) wants to mix FCoE and STP
  - Cisco's recommended approach only mixes Ethernet and FCoE traffic at the access layer
- HP envisions cheaper DCB-capable switches being used in between FCoE devices, which is a bridged approach
  - Needing end-end controls
  - Little visibility, FCoE is just more Ethernet traffic
- Meanwhile Brocade is claiming that interoperability with Cisco reduces functionality…

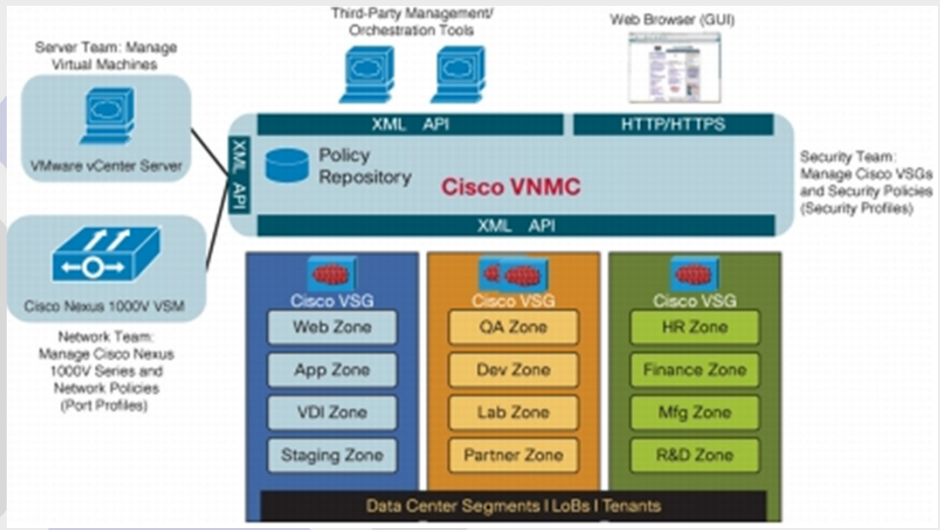
FCoE Status

- Apparently most if not all major storage vendors will be shipping FCoE connectivity soon
- NetApp has been doing so for a while

Other Cisco Announcements, March 31, 2011

- DCNM end-end management, can license LAN and SAN separately if desired
- LISP, MPLS VPN on N7K
- Nexus 1010 functions:
  - Cisco Virtual Security Gateway is a virtual appliance providing VM-level segmentation and zones for compliance and multi-tenant datacenters
  - Cisco Virtual NMC (Virtual Network Management Center) is a virtual appliance GUI to control VSG devices and policies

VNMC/VSG



New 6500 Announcements

- ACE30 for 6500, adds OTV to 6500
  - Comparable to ACE 4710
  - Up to 16 Gbps depending on licensing
- New ES40 for 40 Gbps DCI
- ASA service module (FWSM Next Generation)
  - Twice the performance and four times the session count of competitive network security modules:
    - 20 Gbps maximum firewall throughput (max)
    - 16 Gbps of maximum firewall throughput (multi-protocol)
    - 300,000 connections per second
    - 10 million concurrent connections
    - 250 security contexts
    - 1,000 VLANs

Agenda

- Introduction
- FabricPath
  - Why FabricPath
  - FabricPath Forwarding Details
  - Monitoring and Troubleshooting
- Fiber Channel / SAN for Network Engineers
- Fiber Channel over Ethernet (FCoE)
  - IO Consolidation
  - FCoE Technology
  - Basic Designs
- Recent Announcements
- Summary, References, and Q&A

Summary

- FabricPath provides safe scalable routed MAC L2 technology for massive bandwidth and L2 scalability in the datacenter
- Fibre Channel / SAN basics can be comprehended by networking engineers
- FCoE transports FC across Data Center Bridging-capable Ethernet, allows consolidation of FC and Ethernet

References

- FabricPath Config Guide
  - http://www.cisco.com/en/US/docs/switches/datacenter/sw/5_x/nx-os/fabricpath/configuration/guide/fp_cli_Book.html
- Intro to SAN etc.: see the tutorials at
  - www.snia.org
- Intro to FCoE
  - https://learningnetwork.cisco.com/docs/DOC-3926

Any Questions?

- For a copy of the presentation, email me at pjw@netcraftsmen.net
- About Chesapeake Netcraftsmen:
  - Cisco Gold Partner
  - Cisco Customer Satisfaction Excellence rating
  - We rewrote the DESGN / ARCH (CCDA / CCDP courses) for Cisco
  - Cisco Advanced Specializations:
    - Advanced Route & Switch (10+ CCIEs on staff)
    - Advanced Unified Communications (and IP Telephony)
    - Advanced Wireless
    - Advanced Security (4 double R&S/Sec CCIE's now)
    - Advanced Data Center
  - Deep expertise in Routing and Switching, done some major designs and deployments
  - We've done some large and very large data center assessments, designs, and deployments
  - Hospital wireless deployment, 650 AP's, 9 controllers, 200 Cisco VoWLAN phones