
TOMORROW starts here.
cisco
Cisco live!

Real World Data Centre Deployments and Best Practice Session

BRKDCT-2334

Conrad Bullock

CCIE #10767

Consulting Systems Engineer

Cisco live!

Cisco Live Melbourne Related Sessions

BRKDCT-2218 - Scalable Midsize Data Centre Designs
BRKDCT-2615 - How to Achieve True Active-Active Data Centre Infrastructures
BRKDCT-2081 - Cisco FabricPath Technology and Advanced Fabric Architectures
BRKDCT-2640 - How to Integrate Virtual and Physical Network Services in the Data Centre
BRKDCT-2641 - Designing Storage Infrastructure in a Multi-protocol Data Centre
BRKDCT-2642 - Migration from Classic Design to Insieme Fabric
BRKDCT-2643 - VxLAN Deployment - Use Cases and Best Practices
BRKAPP-2033 - Designing Big Data Clusters with Cisco UCS and Nexus
BRKDCT-3060 - Deployment Challenges with Interconnecting Data Centres
BRKDCT-2328 - Evolution and Challenges of Data Centre Network and Host-Based Overlays

Cisco Live Melbourne Related Sessions

BRKARC-3470 - Cisco Nexus 7000 Switch Architecture
BRKDCT-2048 - Deploying Virtual Port Channel in NXOS
BRKDCT-2049 - Overlay Transport Virtualisation
BRKVIR-2012 - Inside the Nexus 1000v
BRKVIR-2023 - How to Interconnect with Hybrid Cloud Architectures and Approaches
BRKVIR-3013 - Deploying and Troubleshooting the Nexus 1000v Virtual Switch
BRKAPP-9000 - Introduction to Application Centric Infrastructure
BRKAPP-9001 - Policy Driven Data Centre Design
BRKAPP-9005 - Integration of Multi-Hypervisors with the Application Centric Infrastructure
BRKDCT-3640 - Nexus 9000 Architecture
TECDCT-2001 - Next Generation Data Centre Infrastructure

Abstract

- The seminar will discuss real world Nexus Deployment scenarios to make sure your network will meet the demands for performance and reliability. This session will provide and equip you with the latest information on Cisco® data centre network architecture and best practices around those designs. This session will focus on STP, vPC, Fabric Path, QOS, routing and service node insertion from the core of the network to the host. This session will not cover all of the possible options just the best practices to make sure we are all successful.

Agenda

- Fundamental Data Centre Design
- Small Data Centre/Colo Design
- Scalable Data Centre Design
- Scaling the Scalable Data Centre
- Managing the Data Centre
- Overlays

Acronym Slide

- **VPC** - Virtual Port Channel
- **VPC+** - Virtual Port Channel using Fabric Path as the protocol between the peer nodes
- **Fabric Path** - enable highly scalable Layer 2 multipath networks without Spanning Tree Protocol
- **Leaf** - access node used for end node attachment
- **Spine** - used to interconnect all leaf nodes
- **OTV** - Overlay Transport Virtualisation
- **FEX** - Fabric Extender
- **UDLD** - Unidirectional Link Detection
- **LACP** - Link Aggregation Control Protocol
- **SVI** - Switch Virtual Interface
- **MCEC** - Multi-chassis EtherChannel



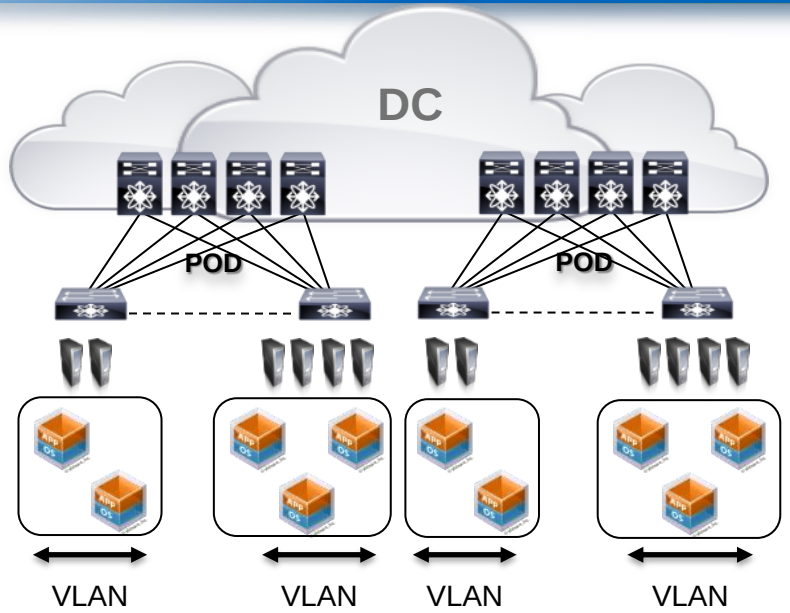

Learning from History?

Problem Statement:
Evolving Network Requirements ….

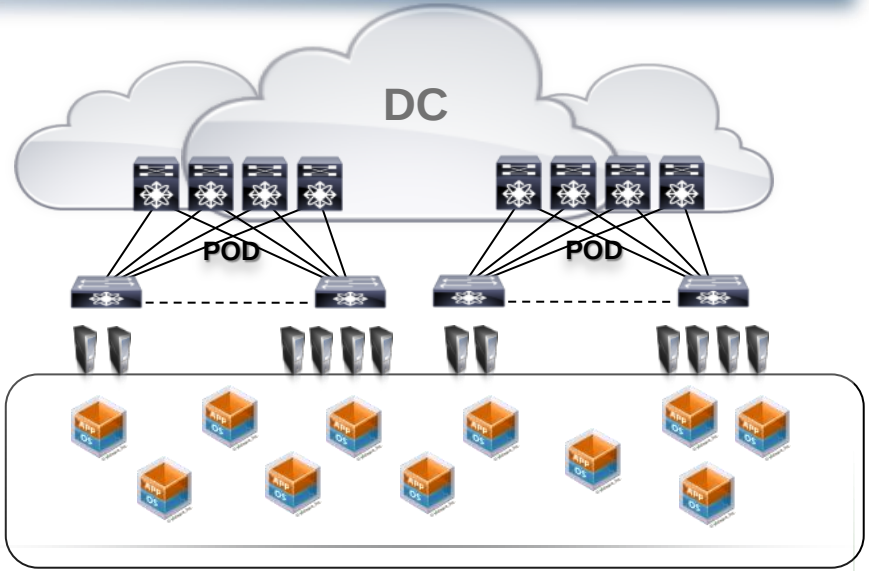
- Changing Traffic Flow Requirements
- Physical to Virtual Resources
- Oversubscription Ratios
- Insertion of services {load-balancing, firewalling, IPS/IDS, WAAS, tenant routing}
- Optical Requirements

VLAN Ubiquity Intra Data Centre

Rack-Wide VM Mobility



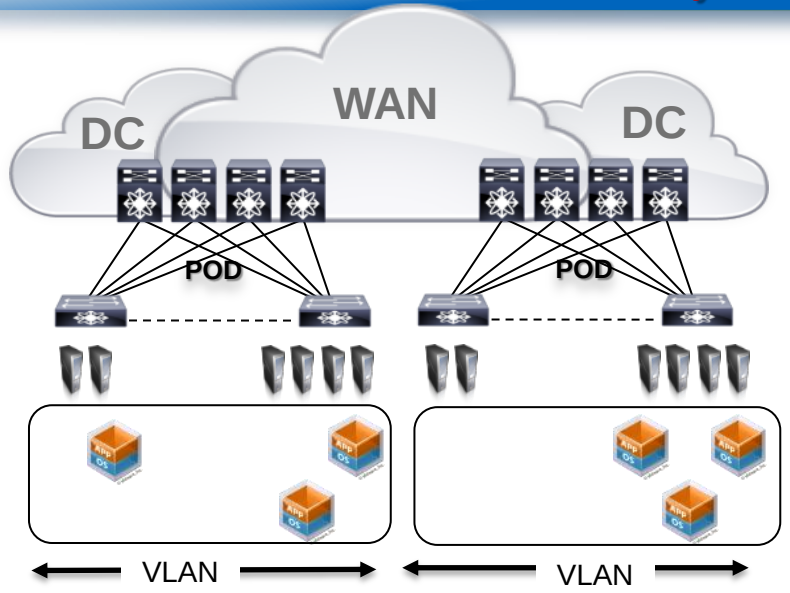
DC-Wide VM Mobility



- Network protocols enable broader VM Mobility
- Implementable on Virtual and Physical
- Examples: VPC, FabricPath/TRILL, VXLAN

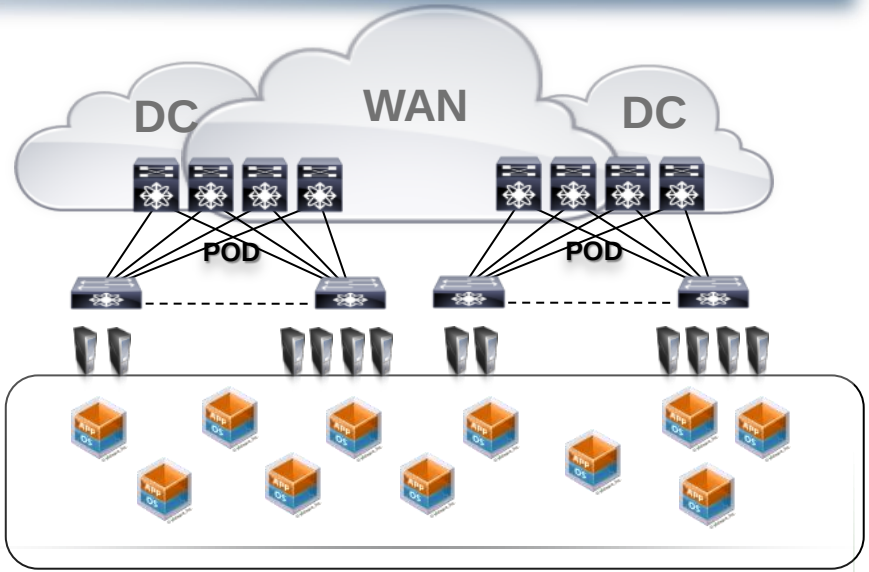
VLAN Ubiquity Inter Data Centre

Data Centre-Wide VM Mobility



- Network protocols enable broader VM Mobility
- Implementable on Virtual and Physical
- Examples: VPC, FabricPath/TRILL, VXLAN

Seamless Layer 2 between DC



- L2 Extension between DC enable broader VM Mobility
- Implementable on Virtual and Physical
- Examples: VPLS, MPLS, OTV, LISP, InterCloud

Services, Services, Services ….

- Insertion of services fundamentally drives the logical design (both L2 and L3) within the data centre
  - Firewall, Application Firewalls
  - Application Delivery Controllers
  - IDS/IPS
  - Client Routing
- Industry transition underway from physical resources to virtualised resources
- How should we provide logical connectivity between data centres with services.

Data Centre Aggregation Layer Design

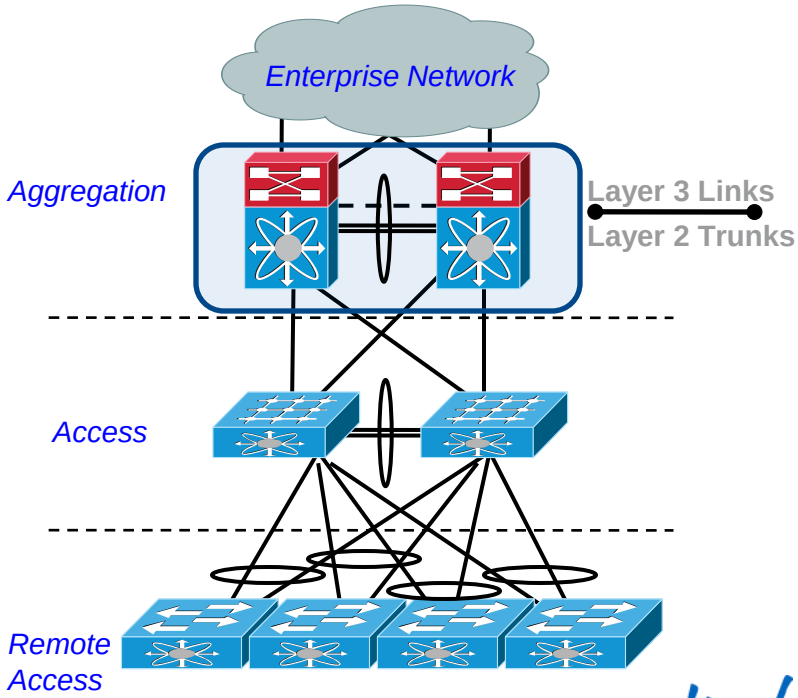
Function & Key Considerations

AS / Area boundary

Routing table scale

Fast routing convergence

Virtualised Aggregation Layer provides
- L2 / L3 boundary
- Access layer connectivity point: STP root, loop-free features
- Service insertion point
- Network policy control point: default GW, DHCP Relay, ACLs

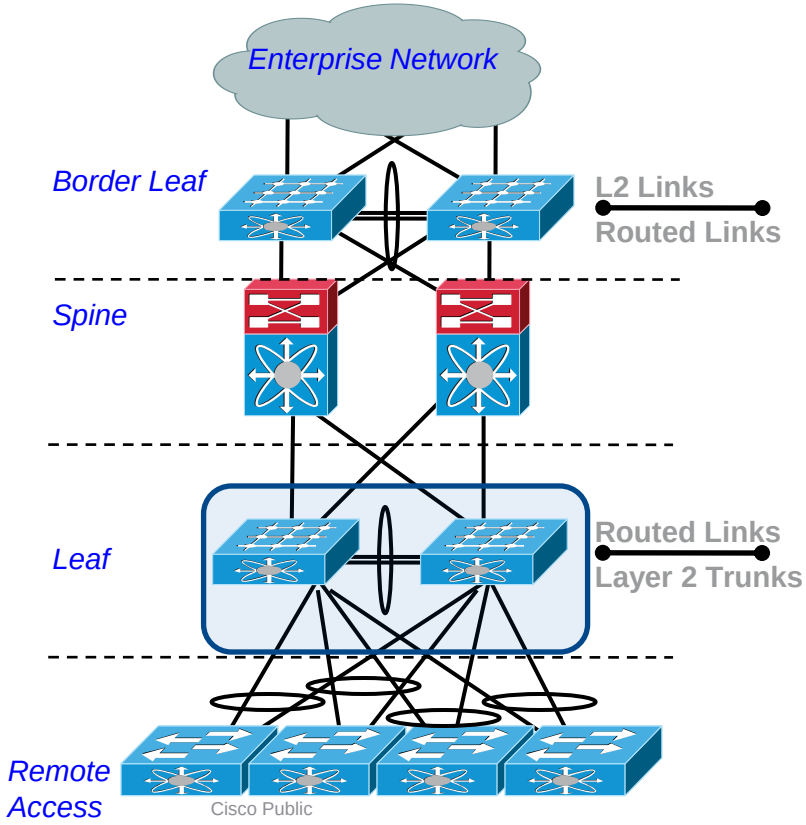


Fat Tree Data Centre Design Evolution

Function & Key Considerations

As L3 pushes to the edge so do the services that used to be at the aggregation layer potentially.

- Fast convergence
- L2 / L3 boundary
- Access layer connectivity point: STP root, loop-free features
- Service insertion point
- Network policy control point: default GW, DHCP Relay, ACLs



Oversubscription Ratio

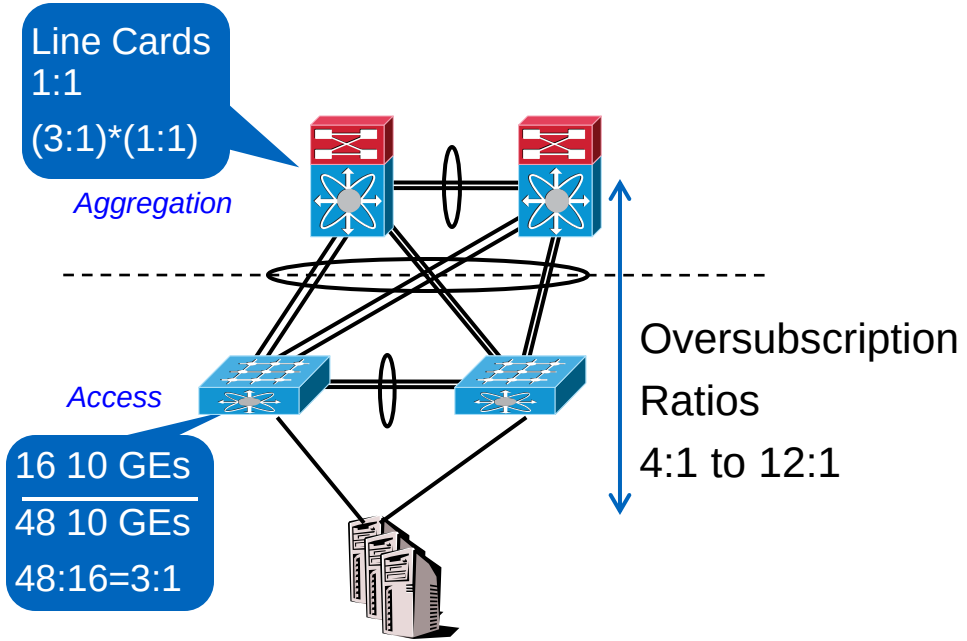
Access to Core/Aggregation

- Large layer 2 domain with collapsed Access and Core
- Worse Case Calculation

  Assume all the traffic is north-south bound

  Assume 100% utilisation from the Access Switches

  All the ports operated in dedicated mode



Oversubscription Ratio

Goal 12:1 to 4:1 and … ↓

Line Cards 1:1
(4:1)*(6:1)*(1:1)
24:1 Oversubscription

Aggregation

8 10 GEs
48 10 GEs
48:8=6:1

Access

8 10 GEs
32 10 Gs
32:8=4:1

Line Cards 1:1
(4:1)*(12:1)*(1:1)
48:1 Oversubscription

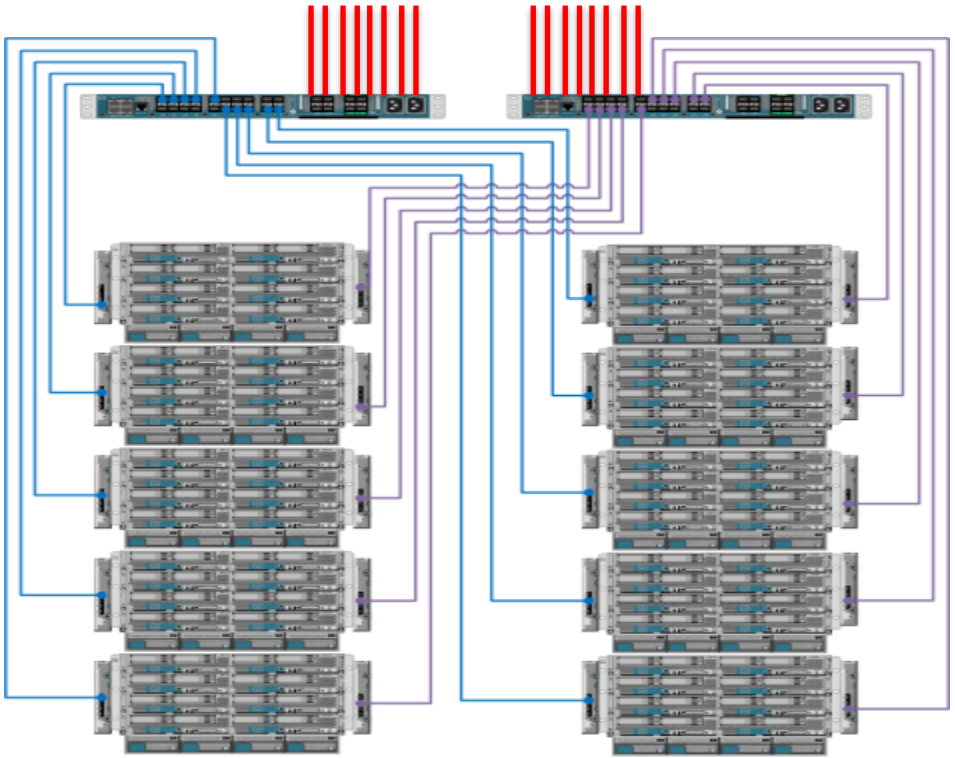
4 10 GEs
48 10 GEs
48:4=12:1

Access

16 Servers
8 10 GEs Possible
Using 4
4 10 GEs
16 10 Gs
16:4=4:1

Oversubscription with Cisco UCS

With UCS, Databases Get the Defined Latency They Require for Optimal Performance, With Full Redundancy.

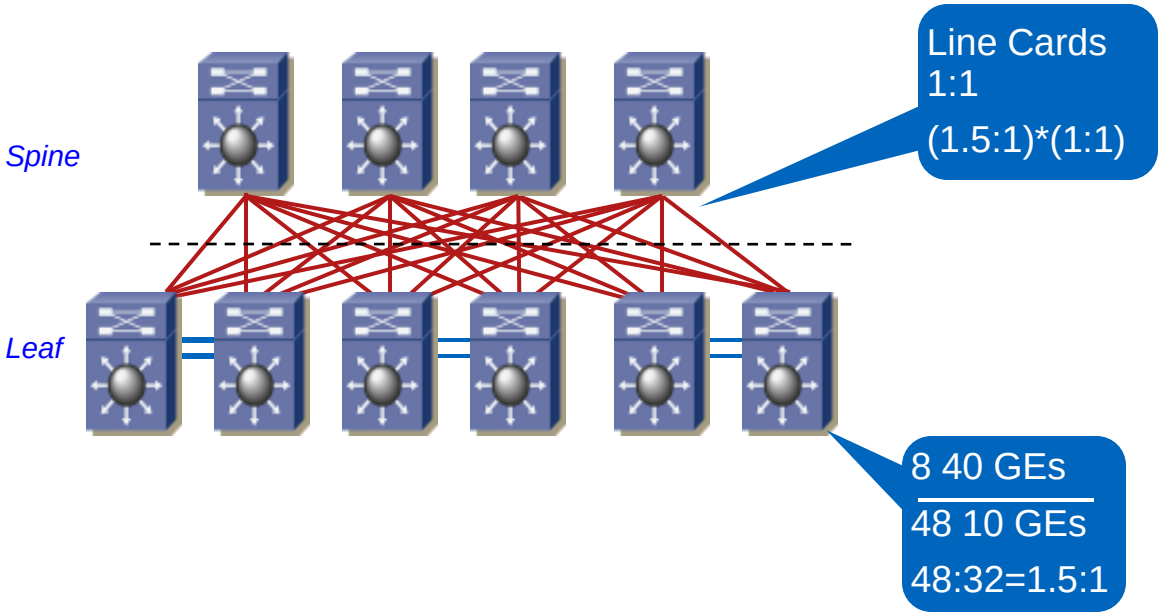


Consistent Latency

Cisco UCS enclosuure

- 15 UCS 5108s with 8 servers installed in each
- 4 Ethernet Modules Per IOM, 80 Gigs out of each server
- Each server has 10 GE line rate access to all other servers in UCS domain
- Server to Server over subscription 8:8 * 8:8= 1
- Servers to Core 120:32=3.75
- Chassis 1 Blade 1, to Chassis 15 Blade 8 = 1 switch hop

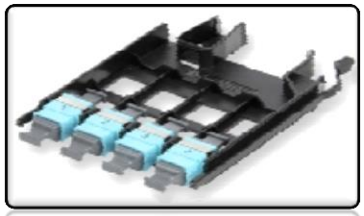
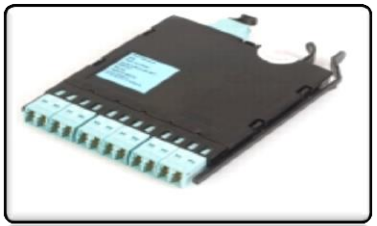
Clos Fabric, Fat Trees

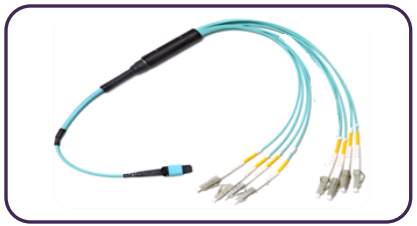
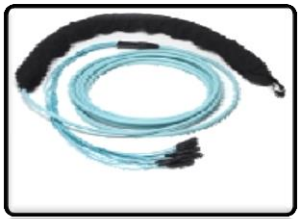


- Changing Traffic Flow Requirements
- Services are deployed at the leaf nodes
- Oversubscription Ratios defined by number of spines and uplink ports
- True horizontal scale

Cabling Infrastructure Parts

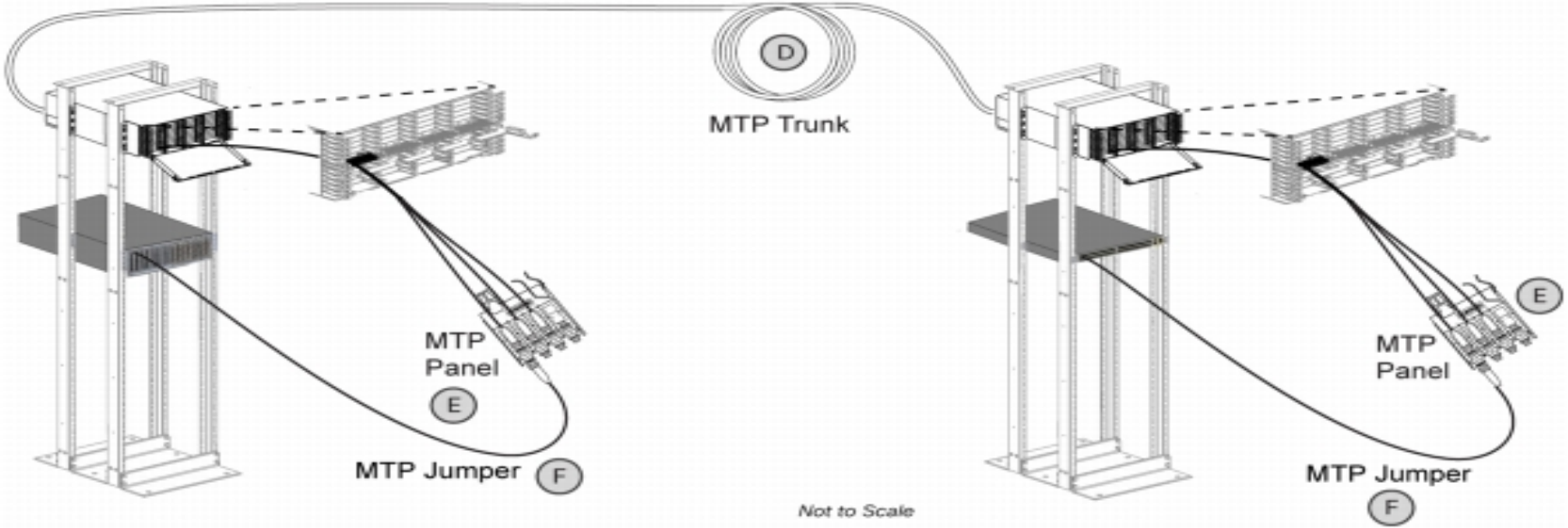
- Joint whitepaper with Corning for Nexus 6000 Cabling Infrastructure
  - http://www.cisco.com/en/US/prod/collateral/switches/ps9441/ps12806/guide_c07-726164.html
- Joint whitepaper with Panduit for Nexus 6000 Cabling Infrastructure
  - http://www.cisco.com/en/US/prod/collateral/switches/ps9441/ps12806/guide_c07-726163.pdf

QSFP to QSFP Connection with Structured Cable

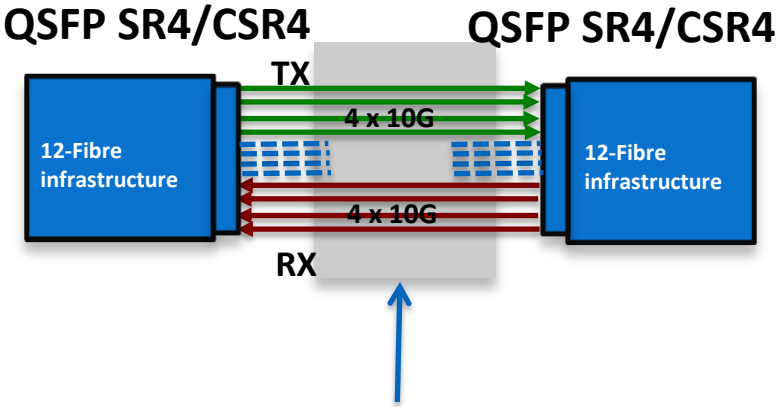
- MTP Trunk provide fibres for multiple 40G connection



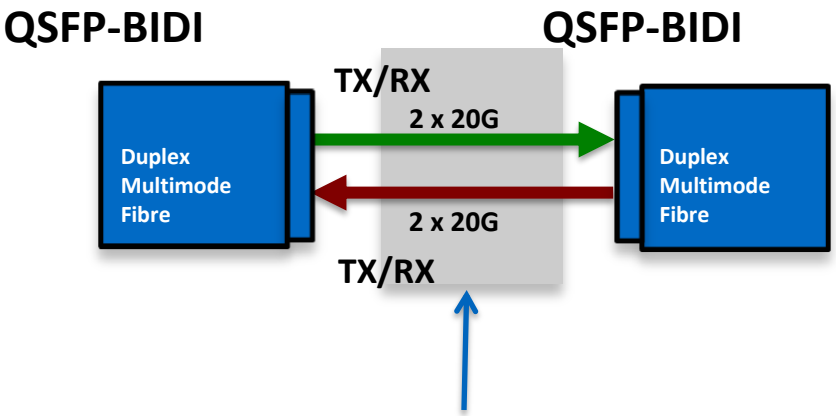
- With CSR4 and SR4 you lose 33% of the fibre capacity.

QSFP-BIDI vs. QSFP-40G-SR4

12-Fibre vs. Duplex Multimode Fibre



12-Fibre ribbon cable with MPO connectors at both ends

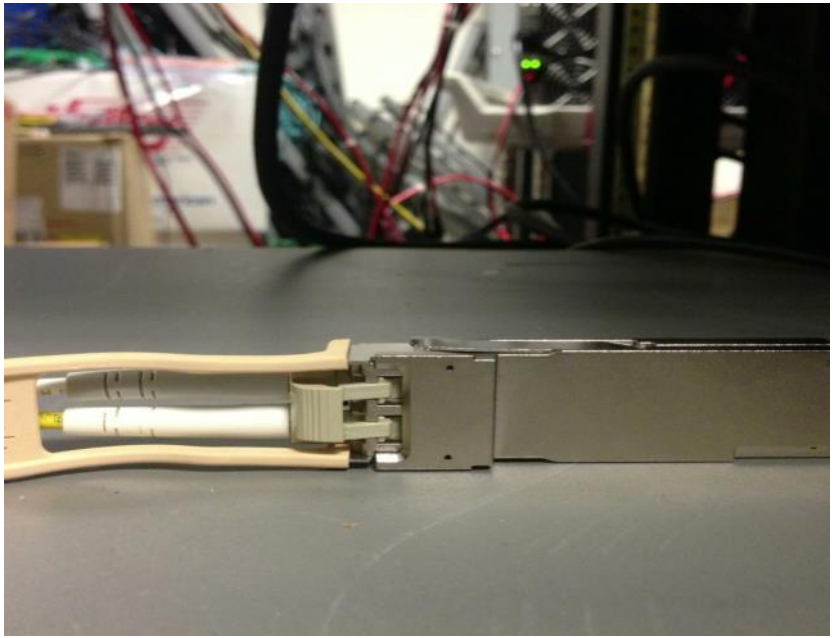


Duplex multimode fibre with Duplex LC connectors at both ends

Higher cost to upgrade from 10G to 40G due to 12-Fibre infrastructure

Use of duplex multimode fibre lowers cost of upgrading from 10G to 40G by leveraging existing 10G multimode infrastructure

QSFP BIDI Overview



- Short reach transceiver with 2 channels of 20G, each transmitted and received over single multi-mode fibre
- 100m with OM3 grade fibre Corning OM4 125m. Panduit OM4 fibre 150m

QSFP+ SKU	Centre Wavelength (nm)	Cable Type	Cable Distance (m)
QSFP-40G-SR-BD	850nm	LC Duplex	100m (OM3) 125m (OM4)

Product	Code Version
Nexus 9000	FCS
Nexus 7700	6.2.6 F3-24 Module
Nexus 7000	6.2.6 for the M2-06 and F3-12
Nexus 600X	7.0.1N1
Nexus 3100	6.0.2A




Fundamental Data Centre Design

UDLD Behaviour

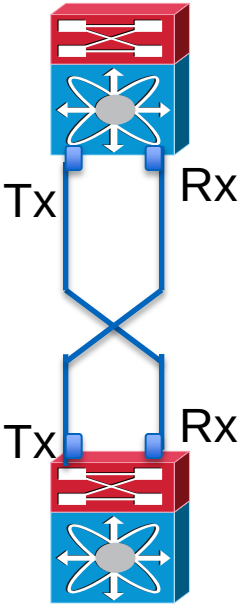
- UDLD is running as a conditional feature, it needs to be enabled:

```
NEXUS(config)# feature udld
```

- UDLD has 2 mode of operations : normal (default) or aggressive mode
- Once UDLD feature is enabled, it will be running on all enabled fibre ethernet interfaces globally as default.
- For copper Ethernet interfaces. UDLD will be globally disabled and needs to be enabled/disabled on per interface (interface config will override the global config):

```
NEXUS(config)# int eth1/1
NEXUS(config-if)# udld enable
```

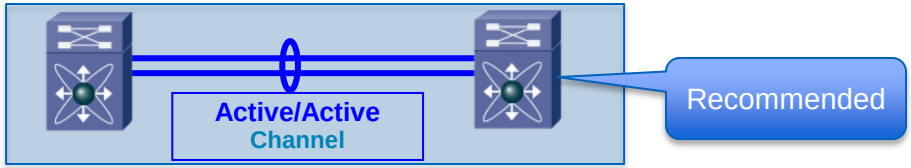
- UDLD needs to be configured on both sides of the line



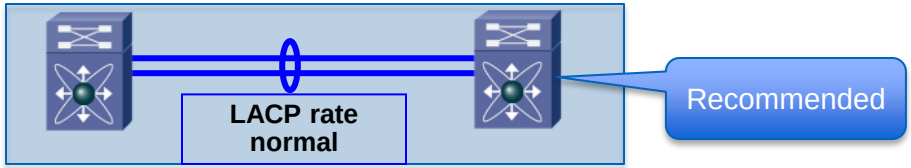
UDLD less important when using bi directional protocols like LACP

Port-Channel

Link Aggregation - IEEE 802.3ad




```
interface eth1/1
  channel-group 1 mode <Active|Passive|On>
```




```
interface eth1/1
  channel-group 1 mode active
  lacp rate <normal|fast>
```

- Recommendation:
  - Use LACP when available for graceful failover and misconfiguration protection
  - Configure port-channel with mode Active/Active

- Recommendations:
  - Use LACP rate normal. It provides capability to use ISSU.
  - If fast convergence is a strong requirement, enable LACP rate fast (however, ISSU and stateful switchover cannot be guaranteed).

NX-OS - Spanning Tree
STP Best Practices For Data Centre

- Implementing STP long path-cost method
  - RSTP default is short and MST default is long

```
NX-OS(config)# spanning-tree pathcost method long
```

- Protect STP root switch by enforcing root guard on its physical ports
  - Spanning Tree costs without pathcost method long may provide unusual results

```
NX-OS(config)# spanning-tree guard root
```

- Block STP BPDU if not needed as soon as it enters the network

```
NX-OS(config)# spanning-tree port type edge
  --- or ---
NX-OS(config)# spanning-tree port type edge trunk
```

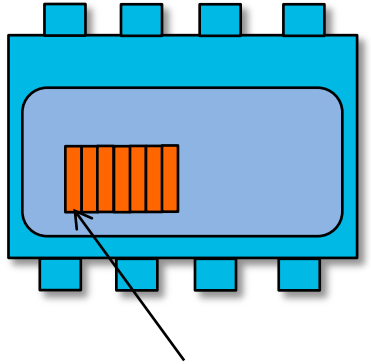
If *switchport mode trunk* and without the "*trunk*" keyword command has no effect

Jumbo Frames on N6K/N5K/N2K

- Nexus 5000 / 3000 supports different MTU for each system class
- MTU is defined in network-qos policy-map
- L2: no interface level MTU support on Nexus 5000

```
policy-map type network-qos jumbo
  class type network-qos class-default
    mtu 9216

system qos
   service-policy type network-qos jumbo
```

```
Nexus 6000
Interface ethernet 1/x
  Mtu 9216
```



Each qos-group on the Nexus 5000/3000 supports a unique MTU

Jumbo Frame Configuration on N7k

- Nexus 7000 all Layer 2 interfaces **by default support Jumbo** frames
- Use system jumbomtu command to change Layer 2 MTU,
  - default 9216
- Layer 3 MTU changed under Interface

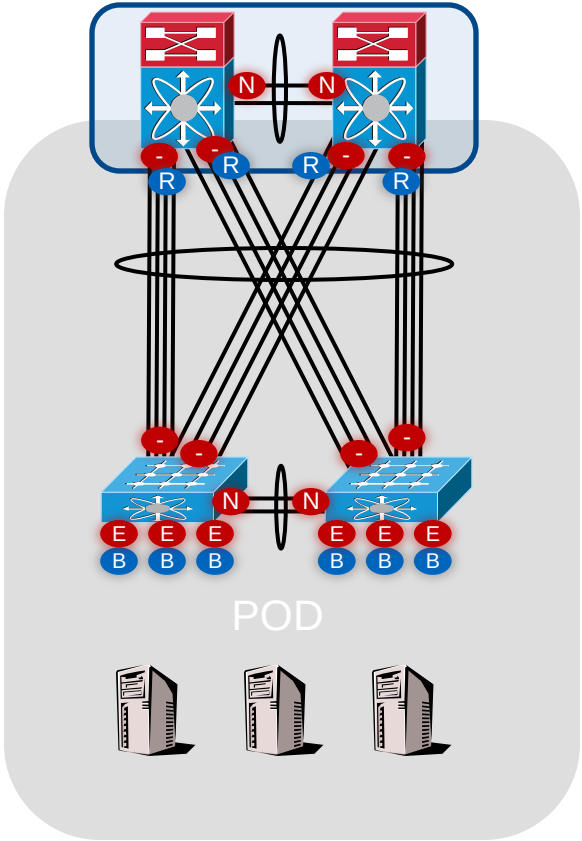
```
show run all | grep jumbomtu
  system jumbomtu 9216

interface Vlan10
  ip address 10.87.121.28/27
  mtu 9216
```

- Nexus 7000 **FCoE policy sets MTU lower** per policy-map than jumbomtu
- Interface MTU overrides network-qos

```
policy-map type network-qos default-nq-4e-policy
  class type network-qos c-nq-4e-drop
    mtu 1500
  class type network-qos c-nq-4e-ndrop-fcoe
    mtu 2112
```

Spanning Tree Recommendations

- Define Peer Switch on Aggregation layer, Both switches have same priority
  - Switch/Port Failure will not cause Spanning Tree recalculation
- Normal Ports down to access Layer
- Network ports for VPC Peer link
- Edge or Edge Trunk going down to access layer
- Define ***Spanning-tree path cost long***



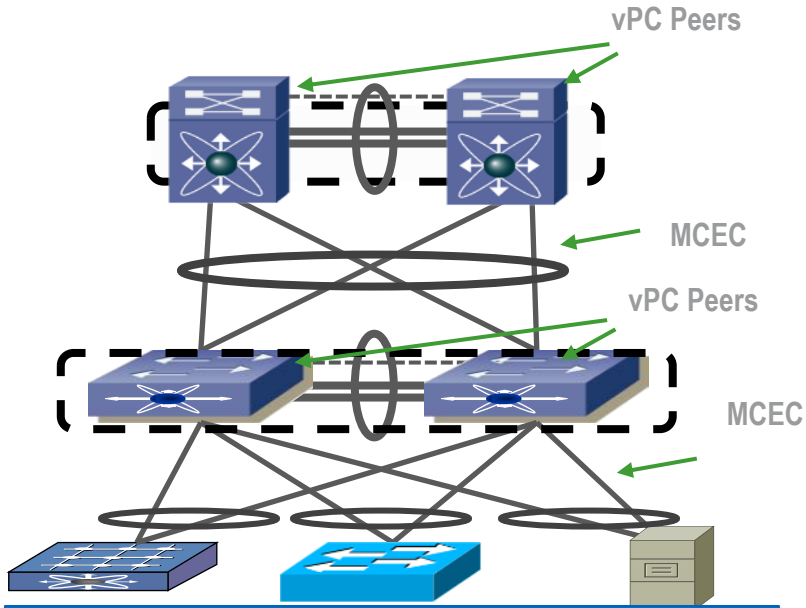
N Network port
E Edge or Edge Trunk
- Normal port type
B BPDUguard
R Rootguard
L Loopguard

vPC – Virtual Port Channel

Multi-Chassis EtherChannel (MCEC)

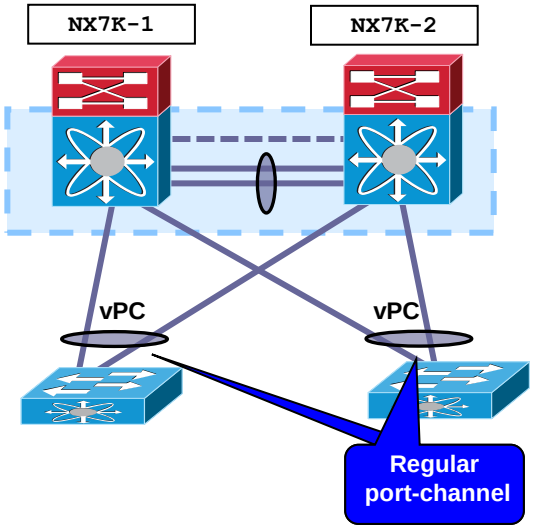
- vPC allows a single device to use a port channel across two neighbour switches (vPC peers) (Layer 2 port channel only)
- Eliminate STP blocked ports & reduces STP Complexity

  (Do not disable STP)
- Uses all available uplink bandwidth - enables dual-homed servers to operate in active-active mode
- Provides fast convergence upon link/device failure
- If HSRP enabled, both vPC devices are active on forwarding plane




```
! Enable vpc on the switch
NX-OS(config)# feature vpc

! Check the feature status
NX-OS(config)# show feature | include vpc
vpc                   1         enabled
```

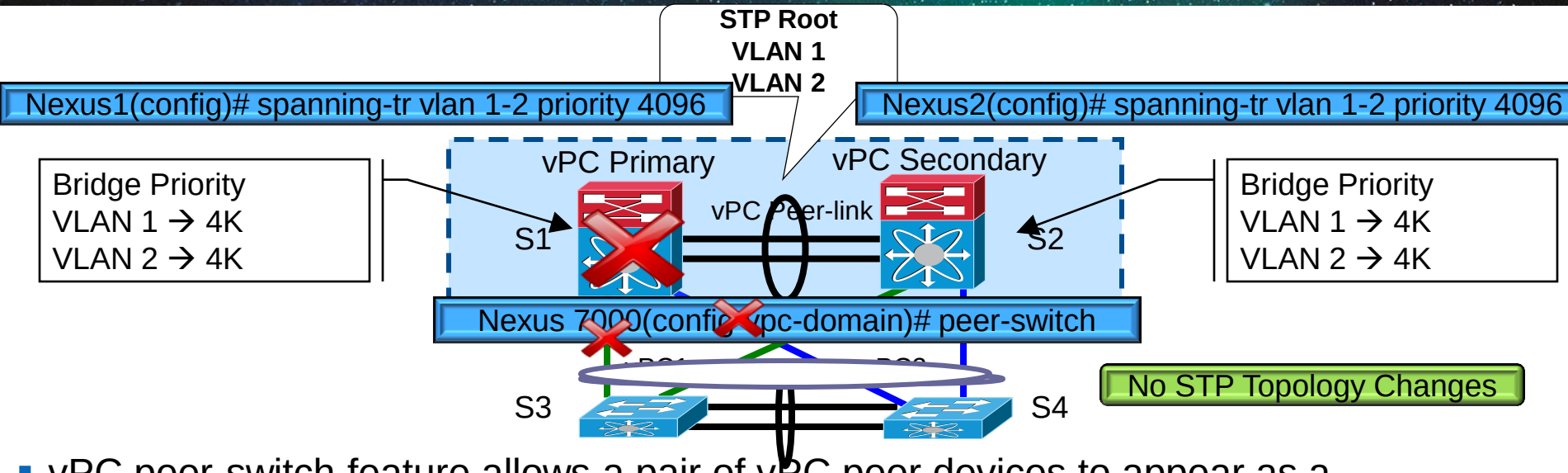
Attaching to vPC Domain



- The Most Important Rule:
  - **Always Dual Attach Devices to a vPC Domain**
- Recommendations:
  - Access device to support STANDARD 802.3ad capability LACP (**mode Active**) or static port-channels (mode on)
  - Use LACP when available for graceful failover and misconfiguration protection
  - MAC pinning for active active NIC redundancy with Hypervisor

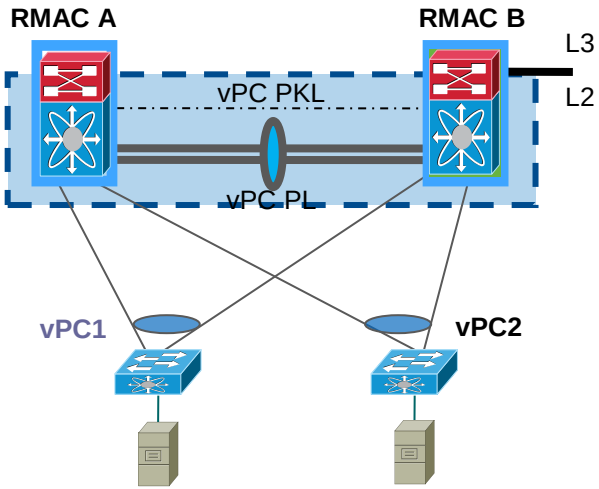
vPC Peer-switch

Unified STP Root with vPC: Improving Convergence



- vPC peer-switch feature allows a pair of vPC peer devices to appear as a single STP Root in the L2 topology (same bridge-id)
- Improves convergence during vPC primary switch failure/recovery avoiding Rapid-STP Sync
- Why doesn't the access switch need Peer-Switch? Not Root…

vPC Peer-Gateway



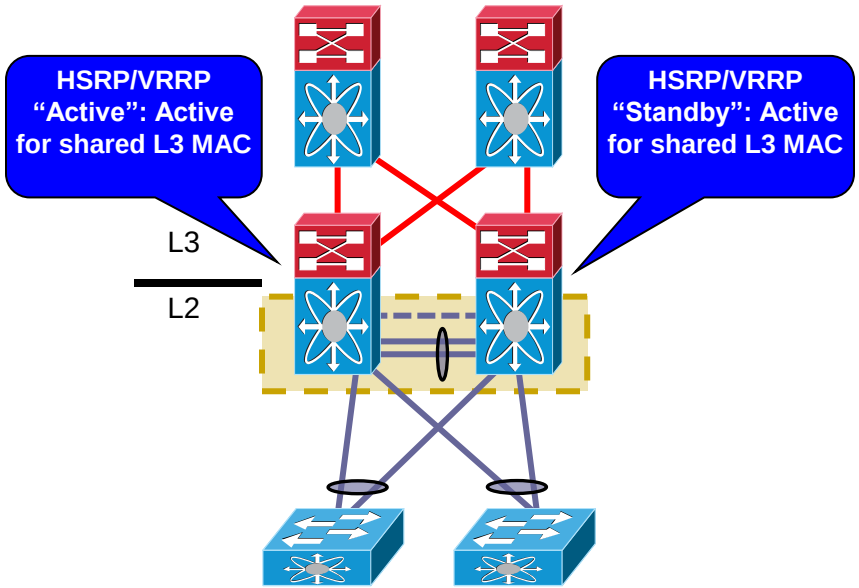

```
Nexus7K(config-vpc-domain)# peer-gateway

Note: Disable IP redirects on all interface-
vlans of this vPC domain for correct
operation of this feature
```

- Allows a vPC peer device to act as the active gateway for packets addressed to the other peer device MAC
  - Necessary for devices which reply to sender's mac-address instead of HSRP virtual mac-address
  - Traffic forwards locally and does not traverse the peer-link
- Keeps forwarding of traffic local to the vPC node and avoids use of the peer-link.
- Allows Interoperability with features of some NAS or load-balancer devices.
- Recommendation:

  Enable vPC peer-gateway in vPC domain

  Disable IP redirects on all SVIs associated with vPC VLANs (Default with NX-OS 5.1)

HSRP with vPC
FHRP Active/Active



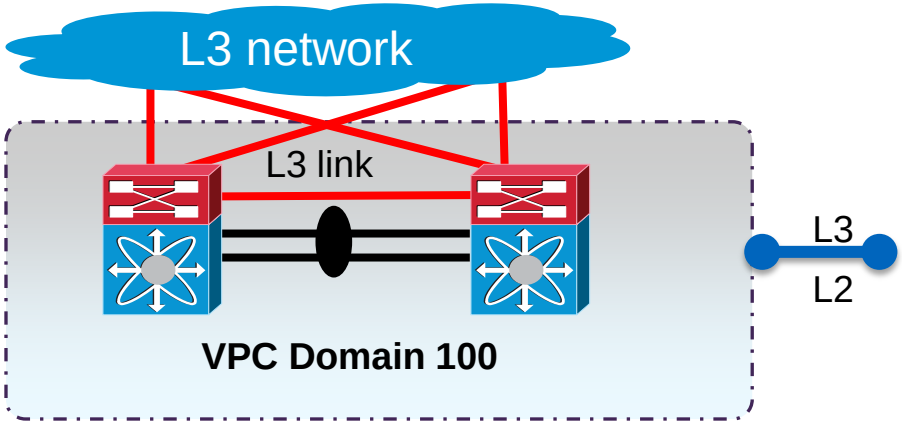
- Support for HSRP/VRRP protocols in Active/Active mode with vPC
  - HSRP or VRRP operate in Active/Active mode from data plane standpoint
  - HSRP or VRRP operate in Active/Standby mode from control plane standpoint (Active instance responds to ARP requests)
- Recommendations:
  - Do not tune HSRP timers (use default ones)
  - One vPC peer can be configured as HSRP active router for all VLANs since both vPC devices are active forwarders
  - Define SVIs as passive interfaces
  - Disable ip redirect on the interface VLAN where HSRP/VRRP is configured

Nexus7k-1# show mac address-t vlan 10 | inc 0000.0c9f.
G 10 **0000.0c9f.f000** static - F F **sup-eth1(R)**
Nexus7k-2# show mac address-t vlan 10 | inc 0000.0c9f.
G 10 **0000.0c9f.f000** static - F F **sup-eth1(R)**

N7K VPC Topology with L3

Backup routing path between N7k

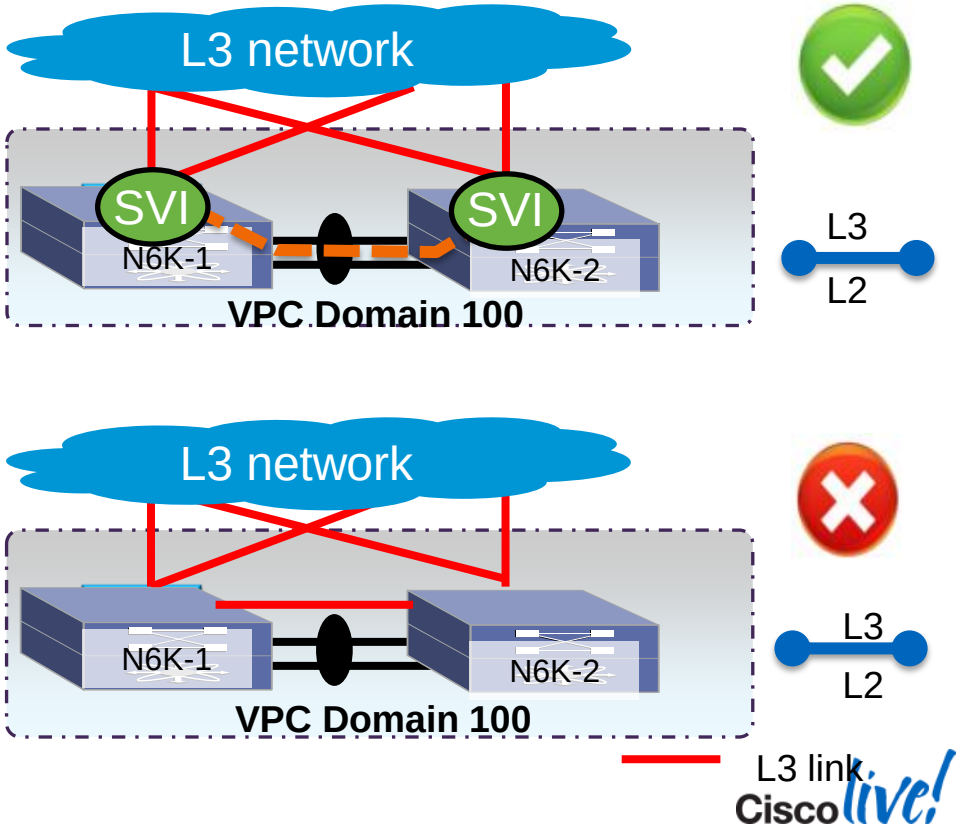
- Peering between two N7k for alternative path in case uplinks fail
- Recommend to have dedicated L3 interface and run routing protocol over L3 interconnect
- Alternately can use SVI over L2 link or VPC as alternate secondary option.
- Unique VPC Domain ID per pair.
  - VPC Domain ID is used at the VPC virtual Bridge ID so it can not be duplicated per L2 domain



N6K VPC Topology with L3

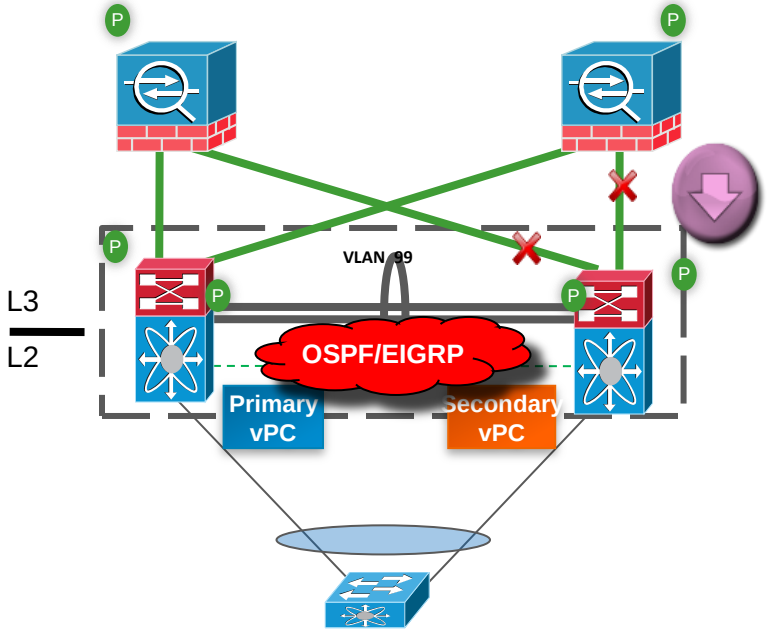
Backup routing path between N6k

- Peering between two N6k for alternative path in case uplinks fail
- Recommend to have dedicated VLAN trunked over peer-link and run routing protocol over SVI
- No support for the topology
  - with additional L3 link between N6k
  - Or additional L2 link with SVI between two N6k running protocol



Backup Routing with SVIs

vPC Peer-Gateway Exclude-Vlan (NX-OS 5.1.3)



- When Backup Routing is formed via SVIs peer-gateway could block traffic
- vPC peer-gateway exclude-vlan knob will be needed for point-to-point VLAN/SVI (backup routing path/VLAN) defined over vPC peer-link.

```
vpc domain 10
...
  peer-gateway exclude-vlan 40,201
```

Command will overwrite previous command

P Routing Protocol Peer

QOS, Why Bother? You Have Tons of Bandwidth …

- Customers have a global QOS policy, do we need to match that in the DC?
- Dedicated appliances are moving to Virtual Machines
- What is more important;

  Moving a Virtual Machine or the Storage that allows the Machine to run?
- Processors and Applications can drive 10 GE and beyond!
- Speed change = Buffering
- What about existing Multi-Tier applications and DCI?
- Incast issues?
- TCP was defined for Low Speed/High Latency Networks; not what we have today!

Impact of Video Compression on Packet Loss Tolerance

1920 lines of Vertical Resolution (Widescreen Aspect Ratio is 16:9)

1080 lines of Horizontal Resolution

1080 x 1920 lines =

2,073,600 pixels per frame

x 3 colours per pixel

x 1 Byte (8 bits) per colour

x 30 frames per second

= 1,492,992,000 bps

or **1.5 Gbps Uncompressed**

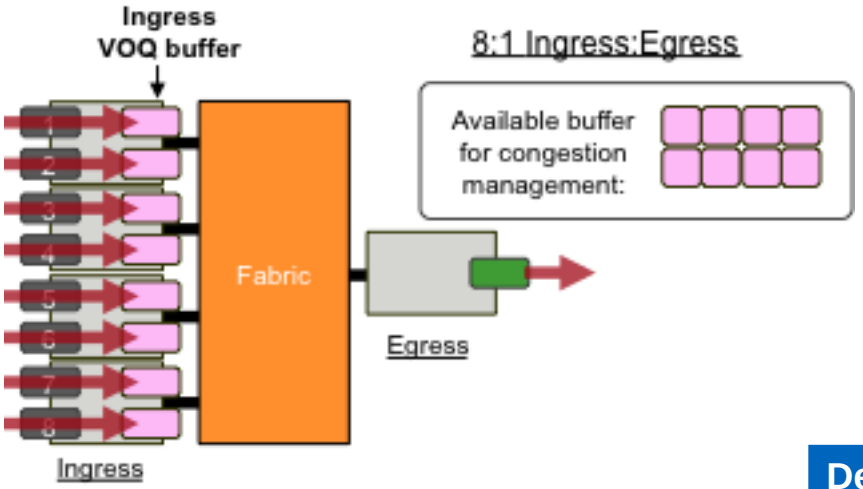
Cisco H.264-based HD Codecs transmit 3-5 Mbps per 1080p image
which represents *over 99.67% compression (300:1)*
Therefore packet loss is proportionally magnified in overall video quality
Users can notice a single packet lost in 10,000—Making HD Video
One Hundred Times More Sensitive to Packet Loss than VoIP!

Key Concepts – Common Points

Nexus 7000 (F-Series) compared to Nexus 5000/6000 QoS

- Nexus 5000/6000 & Nexus 7000 F-Series I/O Modules are sharing the Ingress Buffer Model
- Ingress buffering and queuing (as defined by ingress queuing policy) occurs at VOQ of each ingress port
  - Ingress VOQ buffers are primary congestion-management point for arbitrated traffic
- Egress scheduling (as defined by egress queuing policy) enforced by egress port
  - Egress scheduling dictates manner in which egress port bandwidth made available at ingress
  - Per-port, per-priority grants from arbiter control which ingress frames reach egress port

NEXUS F2 Module Buffer Structure



Distributed Ingress Buffer

Gbps Line Rate: 10 Gbps = 1,250 MB/s
or 1,250 KB/ms

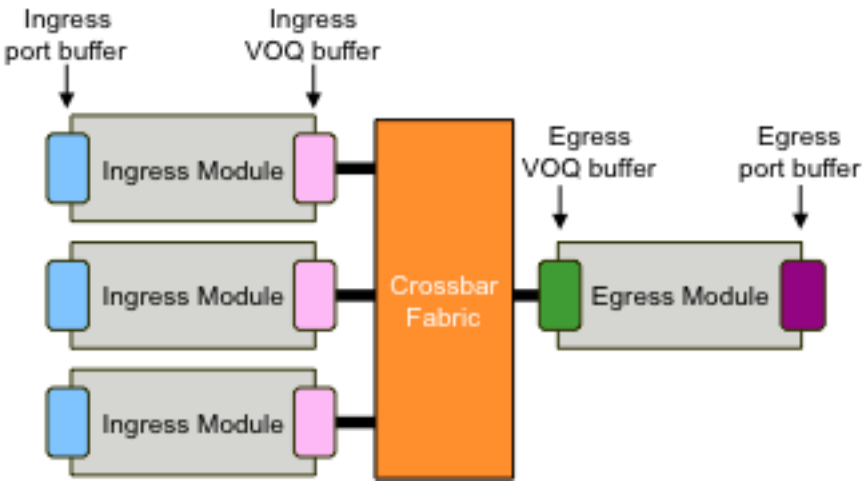
Total Per-Port Buffer (1:1): 1.5 MB
Total Per-Port Buffer (8:1): 12MB

Total Port Buffering Capacity (1:1): ~1.2 ms
Total Port Buffering Capacity (8:1): ~9.6 ms

Default Queue Mapping	COS Values	Buffer Allocated
Queue 0	COS 0 to COS 4	90% 1.35 MB
Queue 1	COS 5 to COS 7	10% 0.15 MB

http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/MSDC/1.0/MSDC1_C.html#wp1120251

NEXUS 7000 M2 Module Buffer Structure



Default Queue Mapping	COS Values	Buffer Allocated
Queue 0	COS 0 to COS 3	50% 2.6 MB
Queue 1	COS 4 to COS 7	50% 2.6 MB

Gbps Line Rate: 10 Gbps = 1,250 MB/s
or 1,250 KB/ms

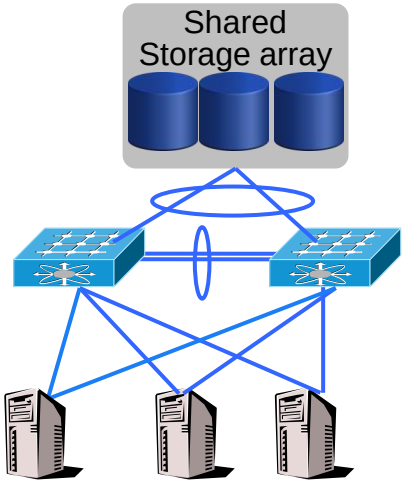
Per Port Ingress Buffer: 5.2 MB
Queue 0 default Buffer 2.6 MB
Queue 1 default Buffer 2.6 MB
Per Port Ingress VoQ Buffer: 4.5 MB
Total Ingress Per-Port Buffer: 9.7 MB

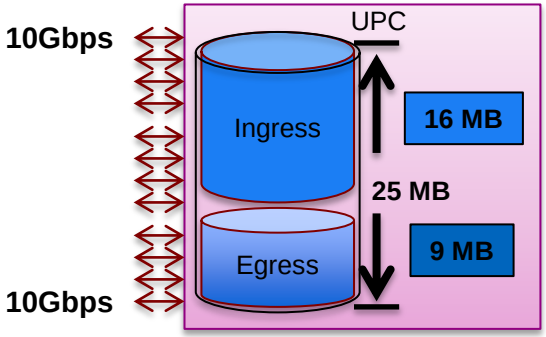
Per Port Egress Buffer: 5 MB
Per Port Egress VoQ Buffer: 380 Kb
Total Egress Per-Port Buffer: 5.3MB
Total Ingress+Egress Per-Port Buffer: 15MB

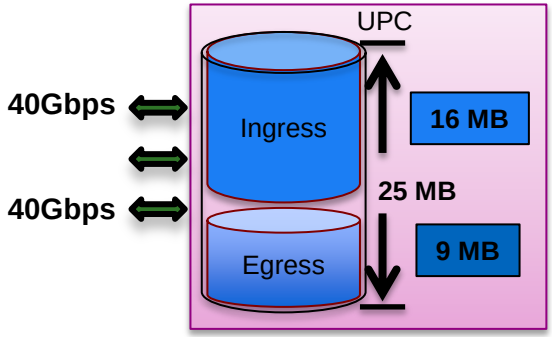
Total Queue 0 or 1 Buffering Capacity: ~ 2.1 ms
Total Ingress Port Buffering Capacity: ~10 ms
Total Ingress+Egress Buffering Capacity: ~12 ms

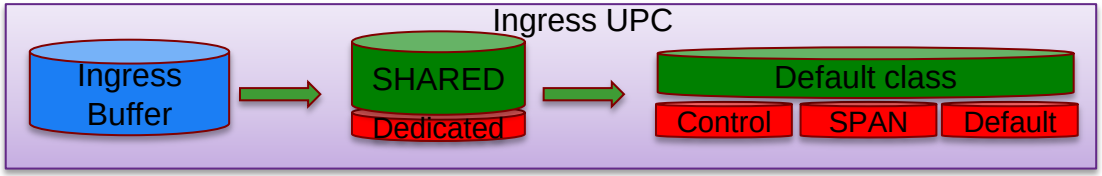
Ingress Buffering and Queuing Model

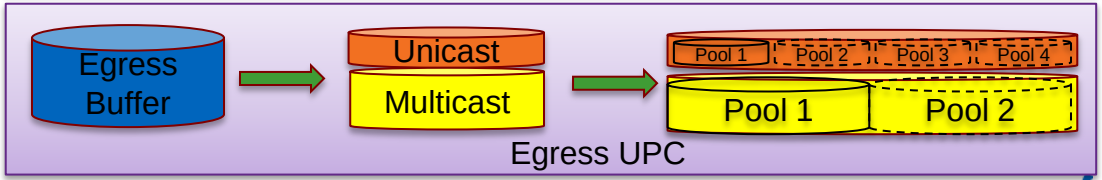
Nexus 6000 Example





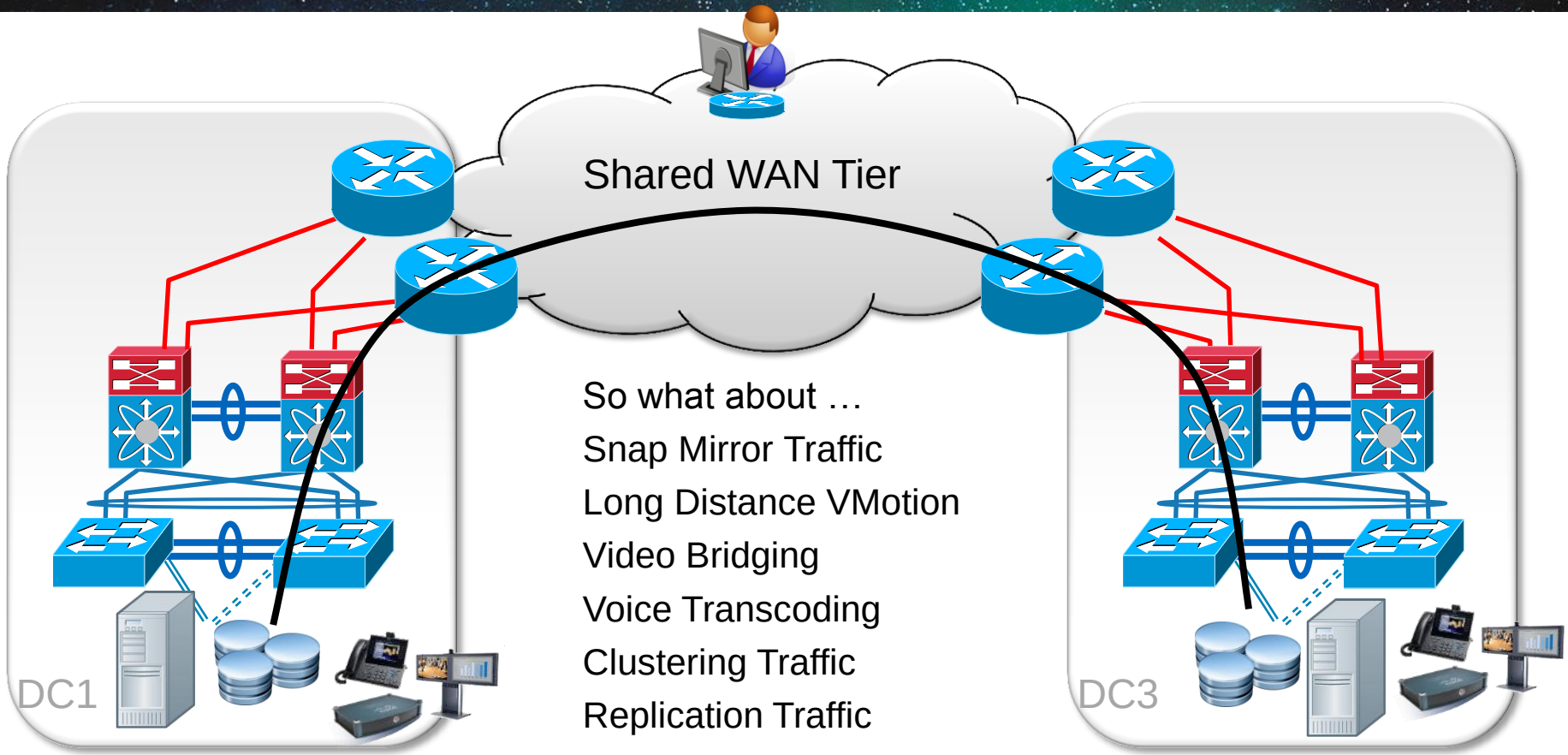






All Tenants use COS = 0

Dual DC Reference Topology

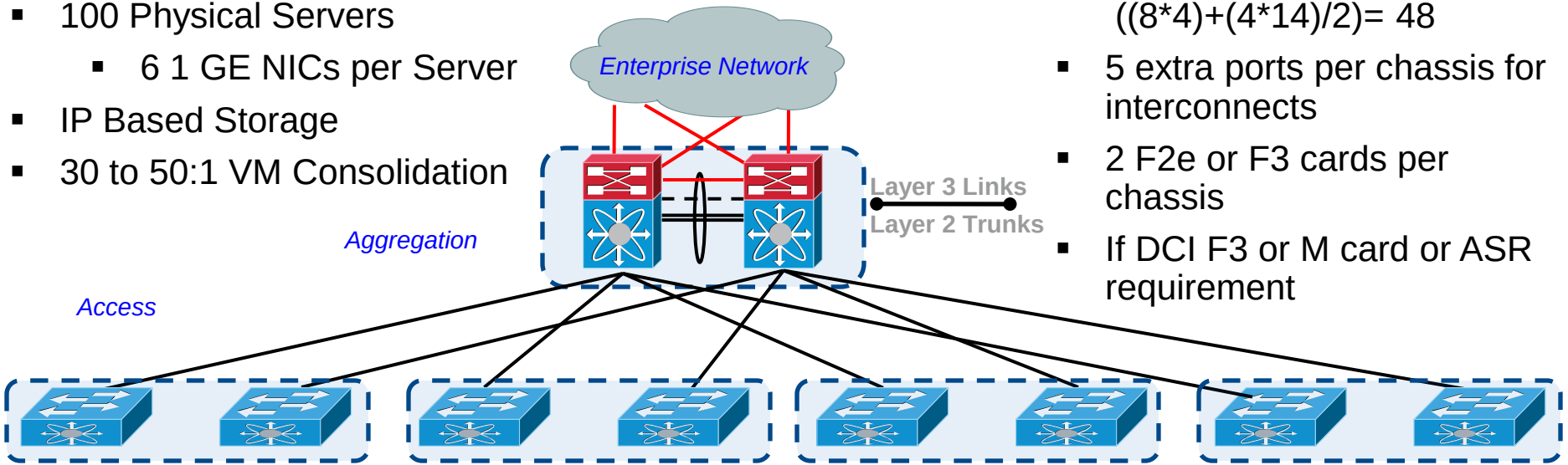


Small Data Centre/Colo Design

Data Centre Building Blocks

Petite Data Centre/ CoLo facility

- 50 Physical Servers
  - 2 10GEs per Server
- 100 Physical Servers
  - 6 1 GE NICs per Server
- IP Based Storage
- 30 to 50:1 VM Consolidation

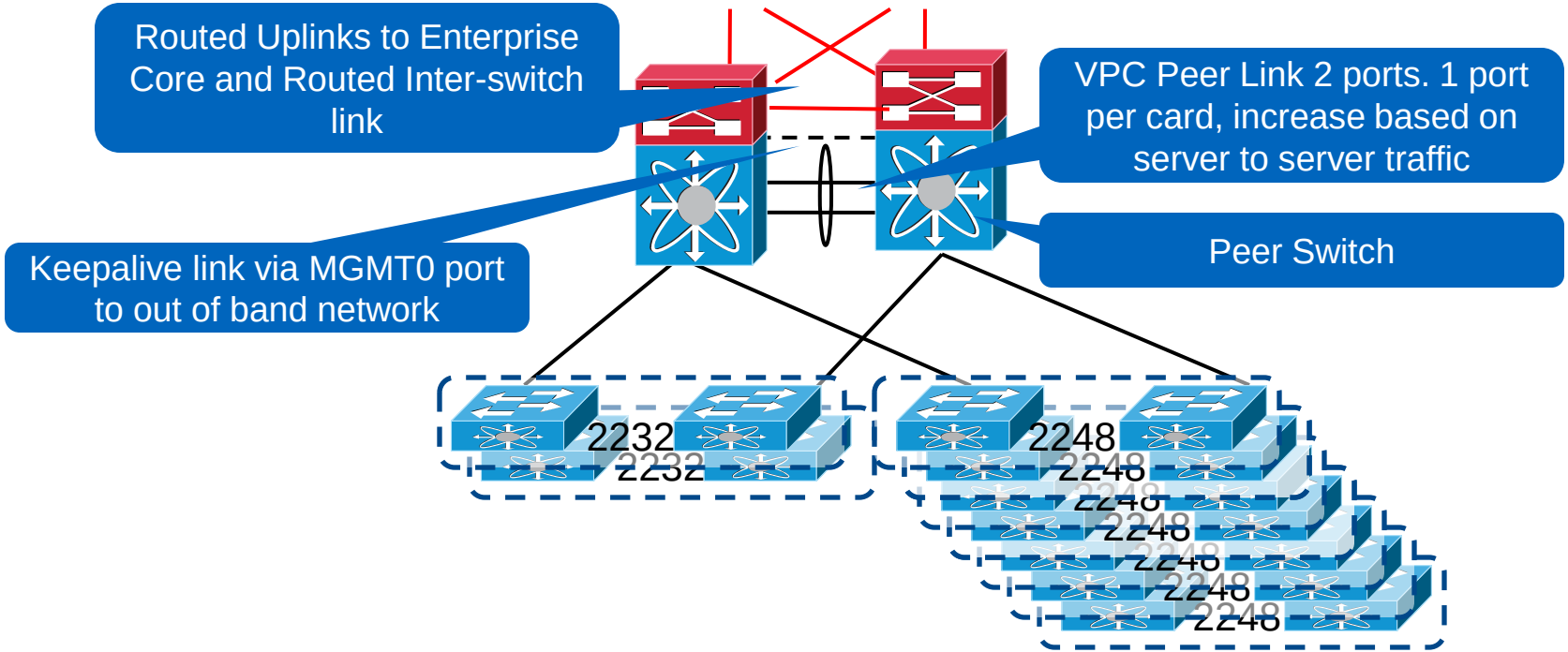


- 4 2232s (8 uplinks per)
- 14 2248s (4 uplinks per)

  ((8*4)+(4*14)/2)= 48
- 5 extra ports per chassis for interconnects
- 2 F2e or F3 cards per chassis
- If DCI F3 or M card or ASR requirement

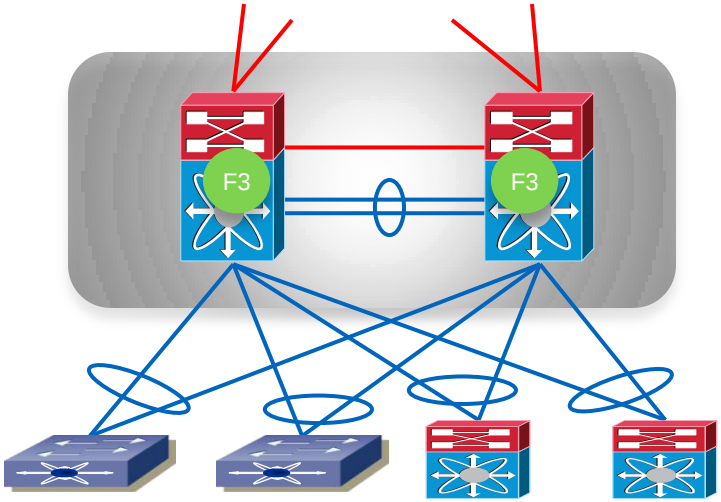
Data Centre Building Blocks

Function & Key Considerations



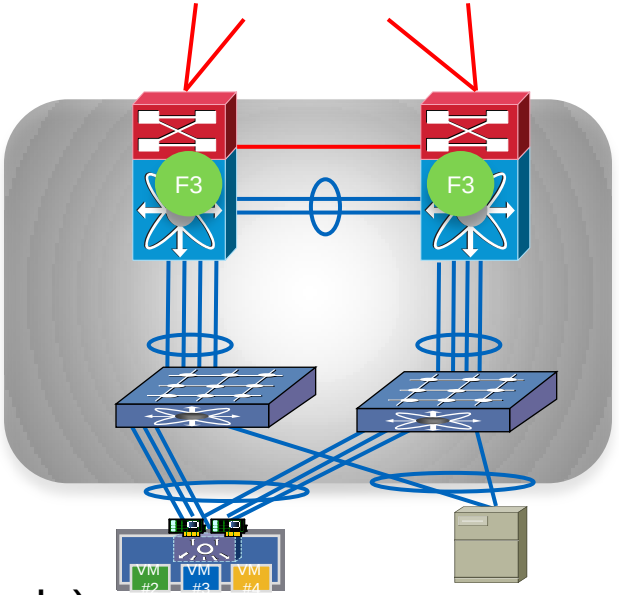
F3 in the Aggregation Layer

- Design requirements
  - High performance
  - 10G and 40G Mix
  - Full L2 and L3 Feature Set
- Topology Description
  - Mix of 10G and 40G vPCs southbound
  - vPC Peer Link
  - Dedicated L3 link
  - Could be backup VLAN
- Scale Attributes
  - 64K MAC Table size (4 times F2e scale)
  - 64K FIB Table size (2 times F2e scale)



F3 in the L2 Access Layer with FEX

- Design requirements
  - High performance
  - 100M/1G/10G/40 ports
- Topology Description
  - vPC Peer Link
  - Full range of FEX support
  - FCoE not supported today
- Scale Attributes
  - 64 FEX with SUP2E
  - 64K MAC Table size (4 times F2e scale)
  - 64K FIB Table size (2 times F2e scale)

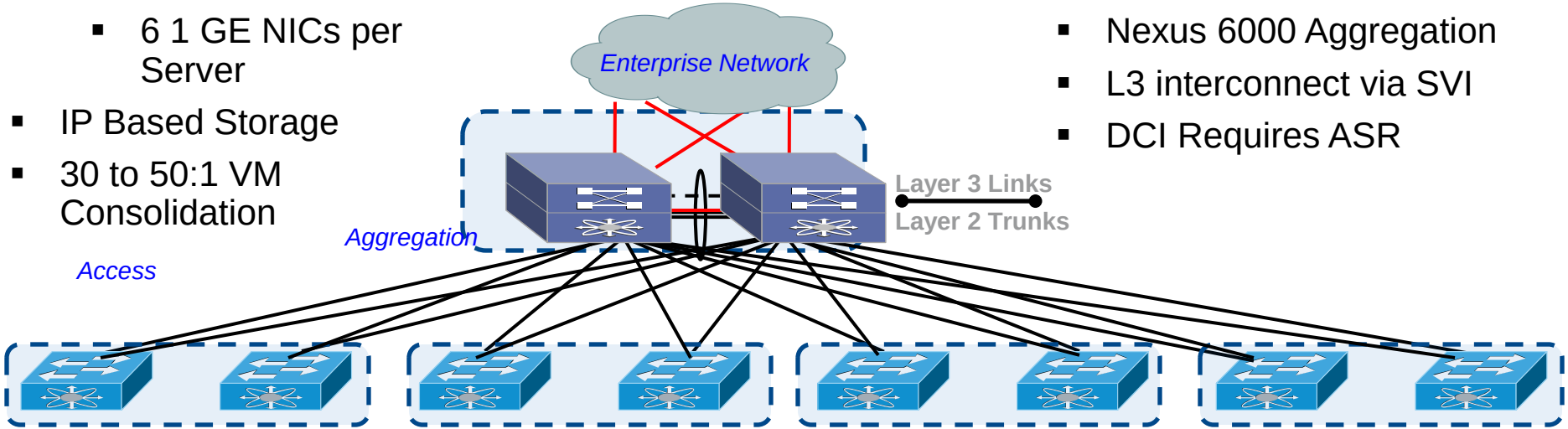


Data Centre Building Blocks

Petite Data Centre/CoLo facility EvPC based 6000 Design

- 50 Physical Servers
  - 2 10GEs per Server
- 100 Physical Servers
  - 6 1 GE NICs per Server
- IP Based Storage
- 30 to 50:1 VM Consolidation

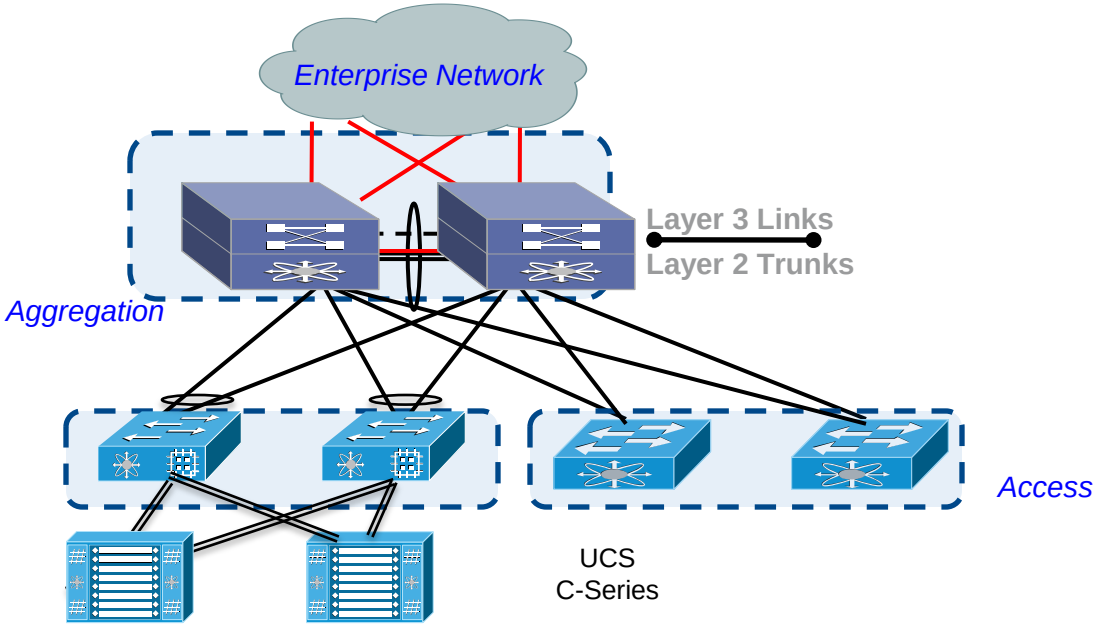
- 4 2232s (8 uplinks per)
- 14 2248s (4 uplinks per)
  ((8*4)+(4*14)/2)= 48
- Nexus 6000 Aggregation
- L3 interconnect via SVI
- DCI Requires ASR



Data Centre Building Blocks

Petite Data Centre/CoLo facility EvPC based and UCS

- Nexus 6000 or Nexus 7000 Aggregation
- ISSU, L3 Interconnect and DCI match previous slide
- Access Mix of FEX and UCS Fabric Interconnect
  - Do not connect UCS FI into FEX.
- VPC from UCS FI to Aggregation



Data Centre Core Considerations

Modular, High-End Solution

Recommended when:

- ✓ ISSU with Features
- ✓ Highest Availability (HA)
- ✓ Investment Protection
- ✓ Multi-Protocol / Services
  - ✓ FEX, VDC, OTV, MPLS/VPLS, LISP
- ✓ Scale and Flexibility 10G/40G/100G

Nexus 7x00 Series

17 Tbps

Fixed, Mid-Range Solution

Recommended when:

- ✓ High density compact 10/40G*
- ✓ Low footprint & low power
- ✓ Low latency & jitter
- ✓ Layer 2 only ISSU*
- ✓ Advanced FEX Features

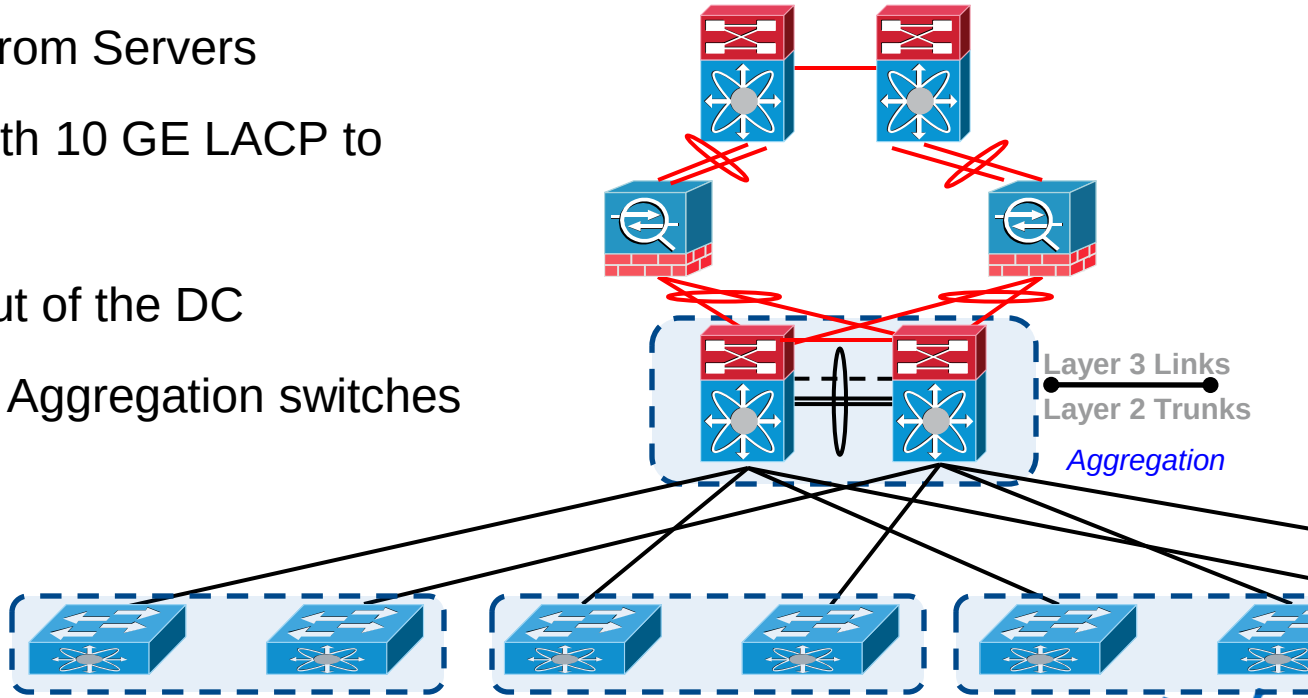
Nexus 6004

Up to 7.68 Tbps

Data Centre Building Blocks

Security Requirements

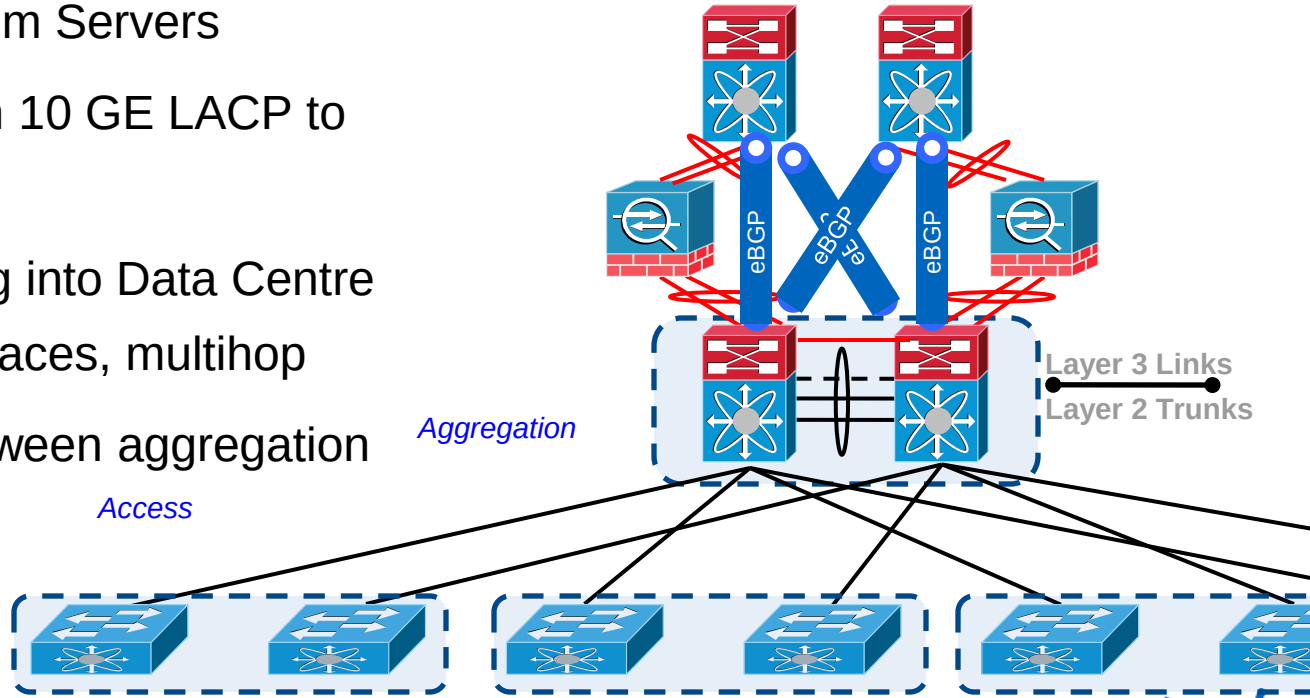
- Firewall off all Users from Servers
- Deployed Firewalls with 10 GE LACP to Aggregation tier.
- Static routes in and out of the DC
- Routed Link Between Aggregation switches



Data Centre Building Blocks

Routing with Security Requirements

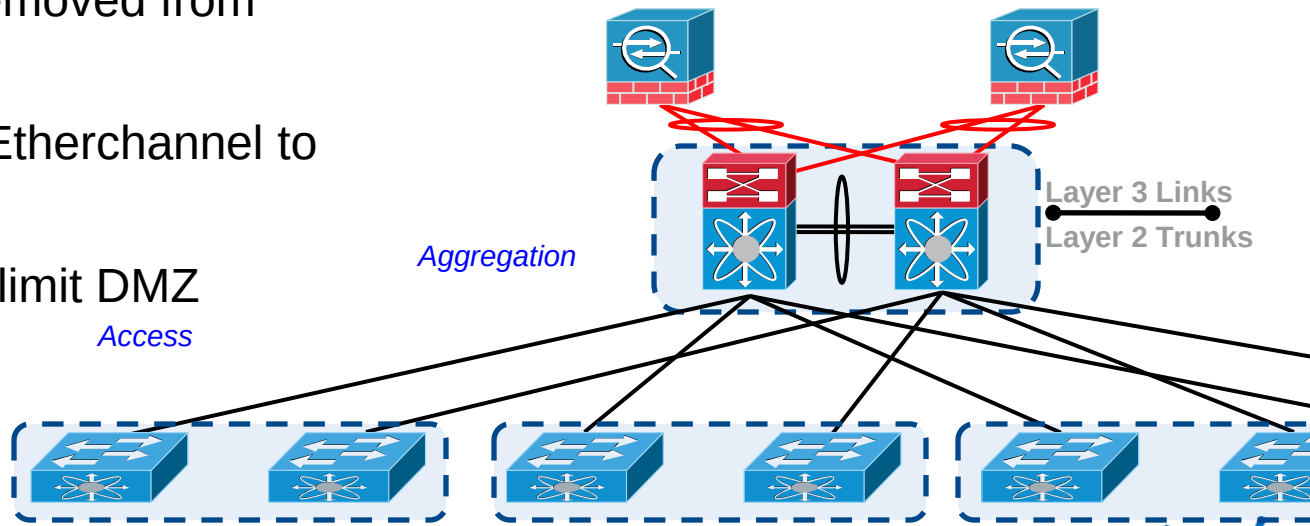
- Firewall off all Users from Servers
- Deployed Firewalls with 10 GE LACP to Aggregation tier.
- eBGP to provide routing into Data Centre against Loopback interfaces, multihop
- Define Routed Link between aggregation 7ks for Routing



Data Centre Building Blocks

User Isolation… What about VDI?

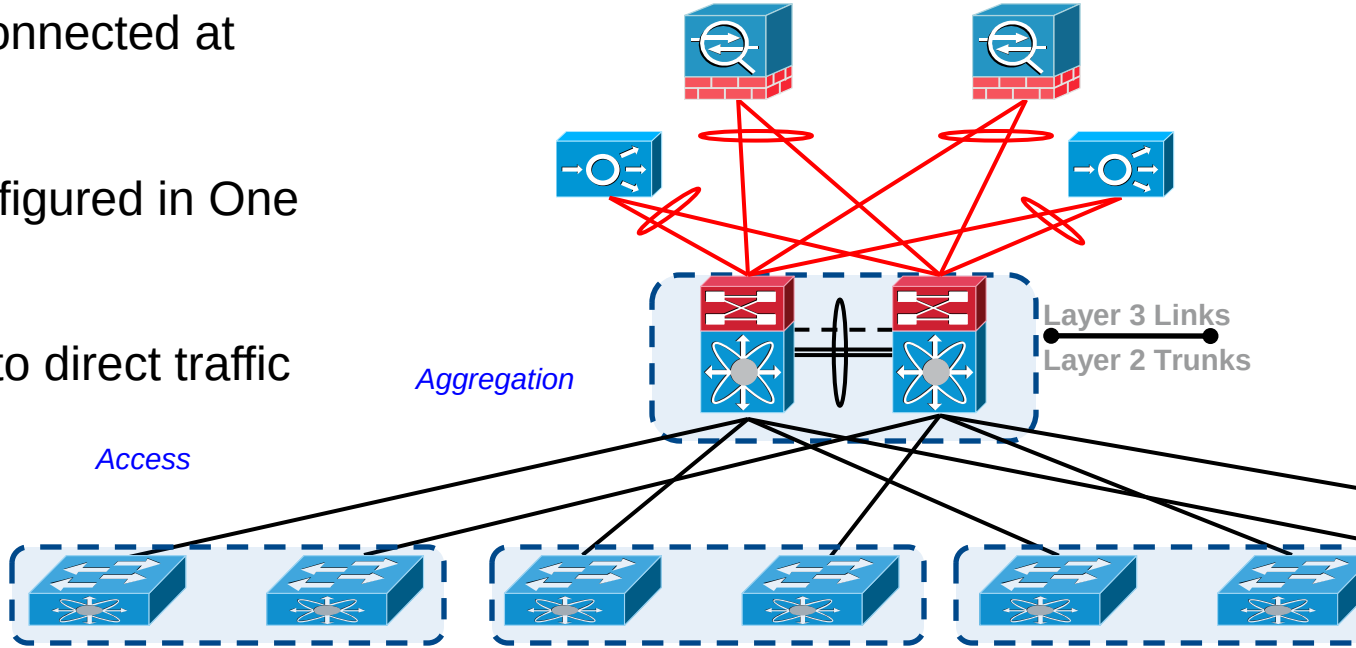
- All VDI servers in Dedicate VLAN range (2)
- SVIs for VLAN range removed from Agg layer
- DMZ VLANs added to Etherchannel to Agg layer
- VRFs was an option to limit DMZ VLANs going to firewall



Data Centre Building Blocks

Load Balancing Requirements

- Load Balancers Connected at Aggregation layer
- Load balancer configured in One Armed mode.
- Source NAT used to direct traffic back to LB

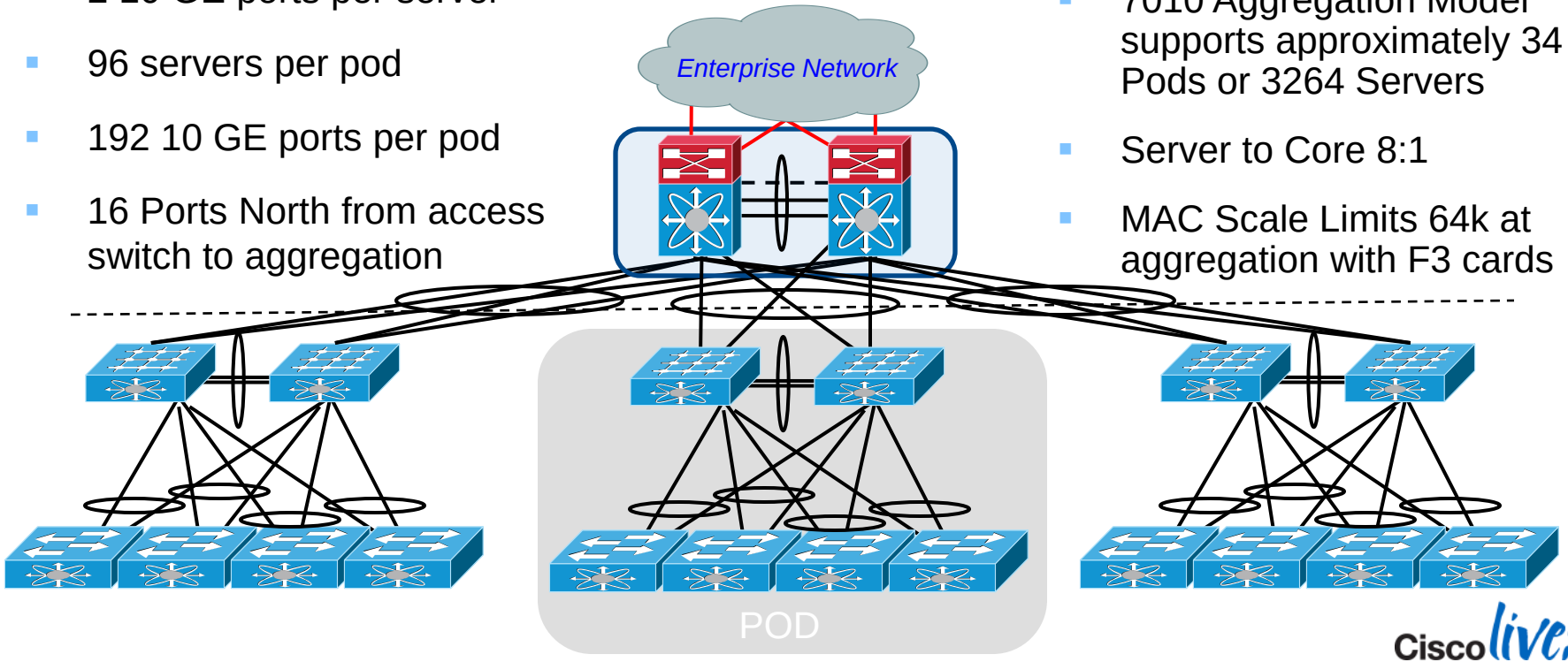


Scalable Layer 2 Data Centre with vPC

Data Centre Building Blocks Larger Design

Repeatable Building Blocks

- 2 10 GE ports per server
- 96 servers per pod
- 192 10 GE ports per pod
- 16 Ports North from access switch to aggregation

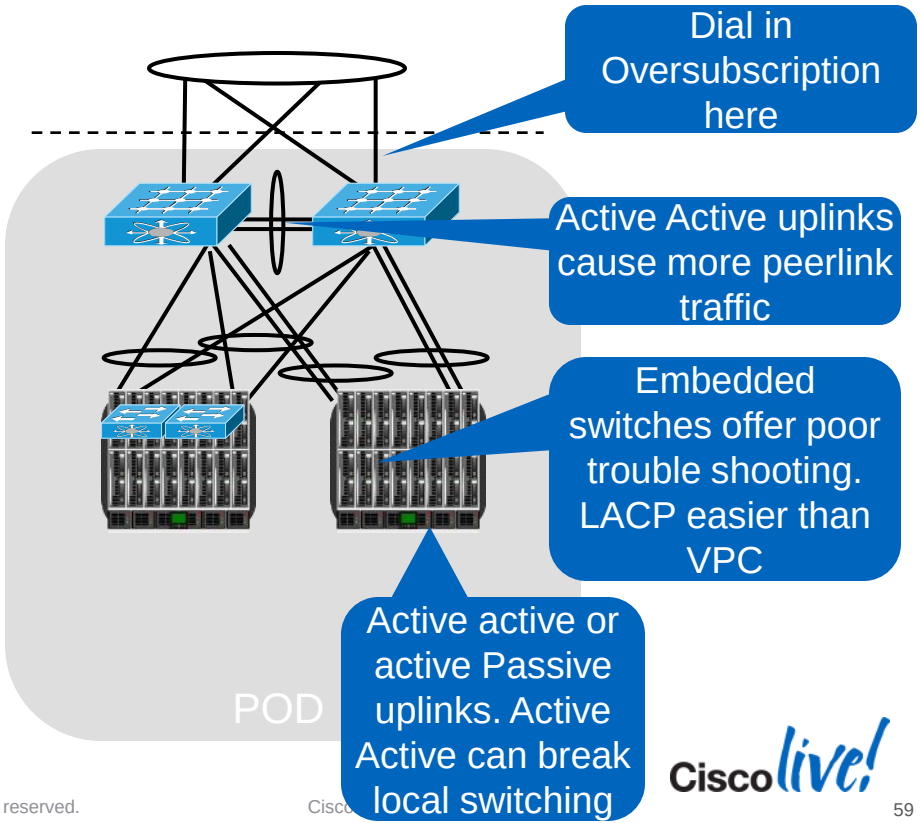


- 7010 Aggregation Model supports approximately 34 Pods or 3264 Servers
- Server to Core 8:1
- MAC Scale Limits 64k at aggregation with F3 cards

Data Centre Building Blocks Larger Design

3rd Party Blade Enclosures

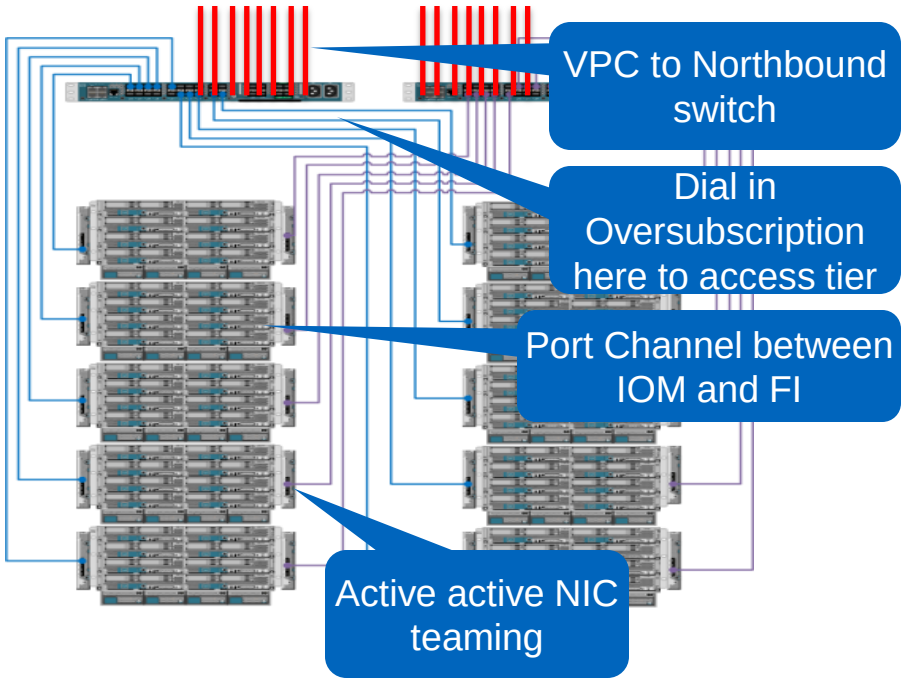
- Straight through to embedded switch or to FEX
- VPC to FEX embedded in Blade enclosure
- 6 Blade enclosures per Access switch Pair based on oversubscription numbers
- 8 uplinks per switch
- 4 ports for peer link without embedded FEX



Data Centre Building Blocks Larger Design

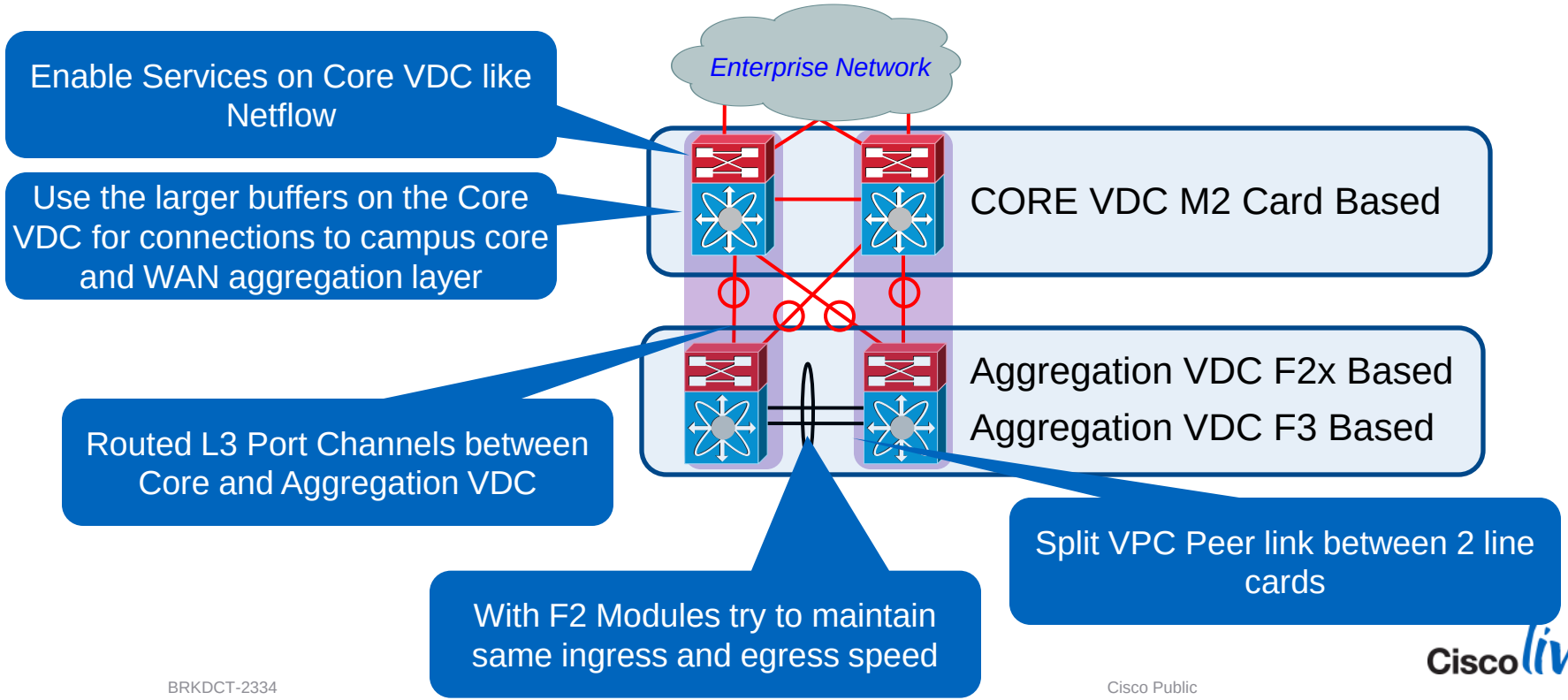
Cisco Blade Enclosures Pod

- 15 Blade enclosures per Mini Pod, 120 servers
- 8 Uplinks per enclosure
- 32 10 GEs north bound
- Aggregate 2 UCS Mini pods per Access tier switch.



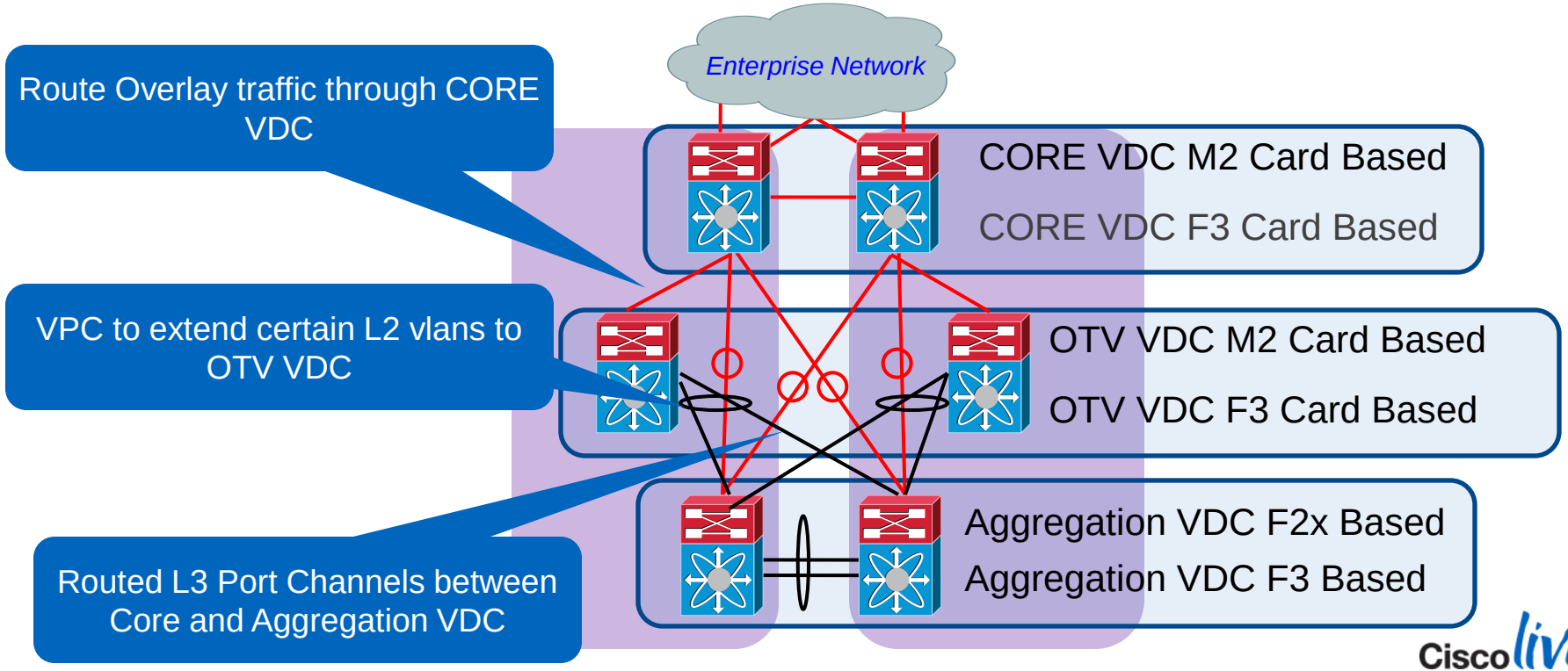
Data Centre Building Blocks Larger Design

Aggregation Layer Detailed Break out



Data Centre Building Blocks Larger Design

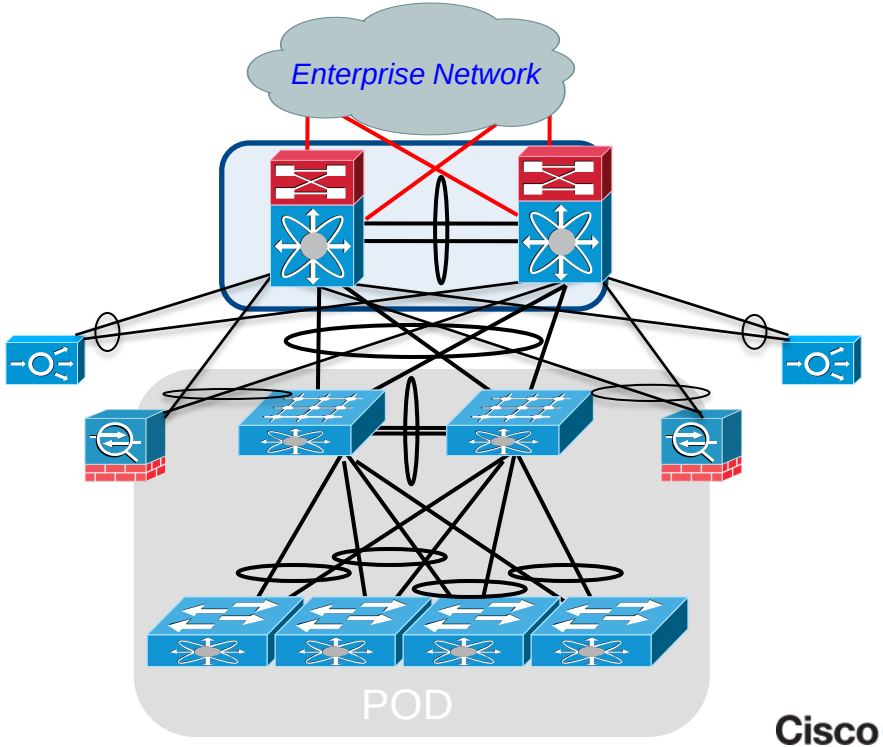
Aggregation Layer Detailed Break out



Data Centre Building Blocks Larger Design

Services Location

- If Appliance supports LACP
  - connect appliance to both core switches
  - Use LACP negotiation
- If Appliance does not support LACP,
  - Connect appliance to single Core switch
  - Decide on Failure behaviour
  - Determine proper orphan behaviour required



Scaling Points of VPC Design

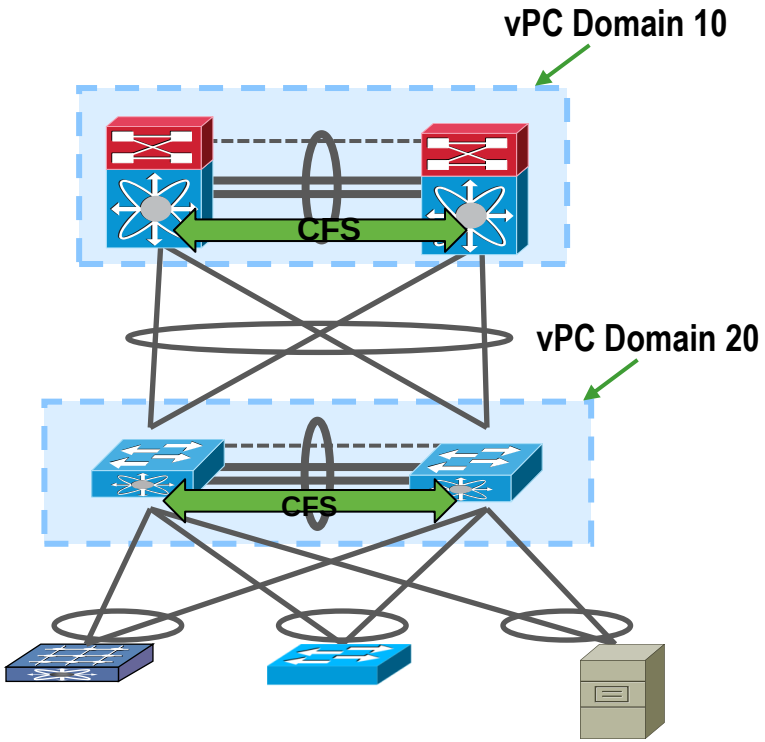
- Configuration Complexity
  - VPC Configuration needs to be replicated to both nodes
  - Failures could isolate orphan ports
- Scaling Limitations
  - F3 Modules today support 64K MAC addresses
  - F2 and F2e Modules today support 16k MAC addresses
    - F2e Proxy functionality to scale MAC address table
    - Move to M2 cards for high MAC scalability
  - Buffers Oversubscription
- Trouble shooting Layer 2 issue complexity

EtherChannel/vPC Maximums

Feature	Nexus 7000 Verified Limit (Cisco NX-OS 6.2)	Nexus 7000 Verified Limit (Cisco NX-OS 6.1)	Nexus 7000 Verified Limit (Cisco NX-OS 6.0)	Nexus 7000 Verified Limit (Cisco NX-OS 5.2)
Port Channels Per System	744	528	528	384
Virtual Port Channels (vPCs) (total) per system	744	528	528	244
Number of vPCs (FEX) per system	528	528	528	244
Number of vPC+s (total) per system	244	244	244	244

Feature	Nexus 6000 Verified Topology	Nexus 6000 Verified Maximum	Nexus 5548 Verified Maximum	Nexus 5596 Verified Maximum
Number of Switchport Etherchannels	48	96 (Single member port-channel for 40G ports) 384 (Single member port-channel for 10G ports) 64 (Multi member port-channel)	48	96
Number of HIF FEX port channels/vPCs (across the maximum number of FEXs)	576	576	576	576

vPC Consistency Check

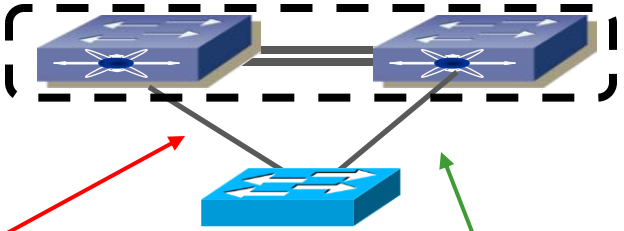


- Both switches in the vPC Domain maintain distinct control planes
- CFS provides for protocol state synchronisation between both peers (MAC table, IGMP state, …)
- Currently a manual process with an automated consistency check to ensure correct network behaviour
- Two types of interface consistency checks
  - Type 1 – Will put interfaces into suspend. With Graceful Consistency check only suspend on secondary switch
  - Type 2 – Error messages to indicate potential for undesired forwarding behaviour

Virtual Port Channel - vPC

vPC Control Plane – Type 2 Consistency Check

- Type 2 Consistency Checks are intended to prevent undesired forwarding
- vPC will be modified in certain cases (e.g. VLAN mismatch)



```
5020-1# sh run int po 201
 interface port-channel201
  switchport mode trunk

  switchport trunk allowed vlan [redacted]-105
  vpc 201
  spanning-tree port type network
```

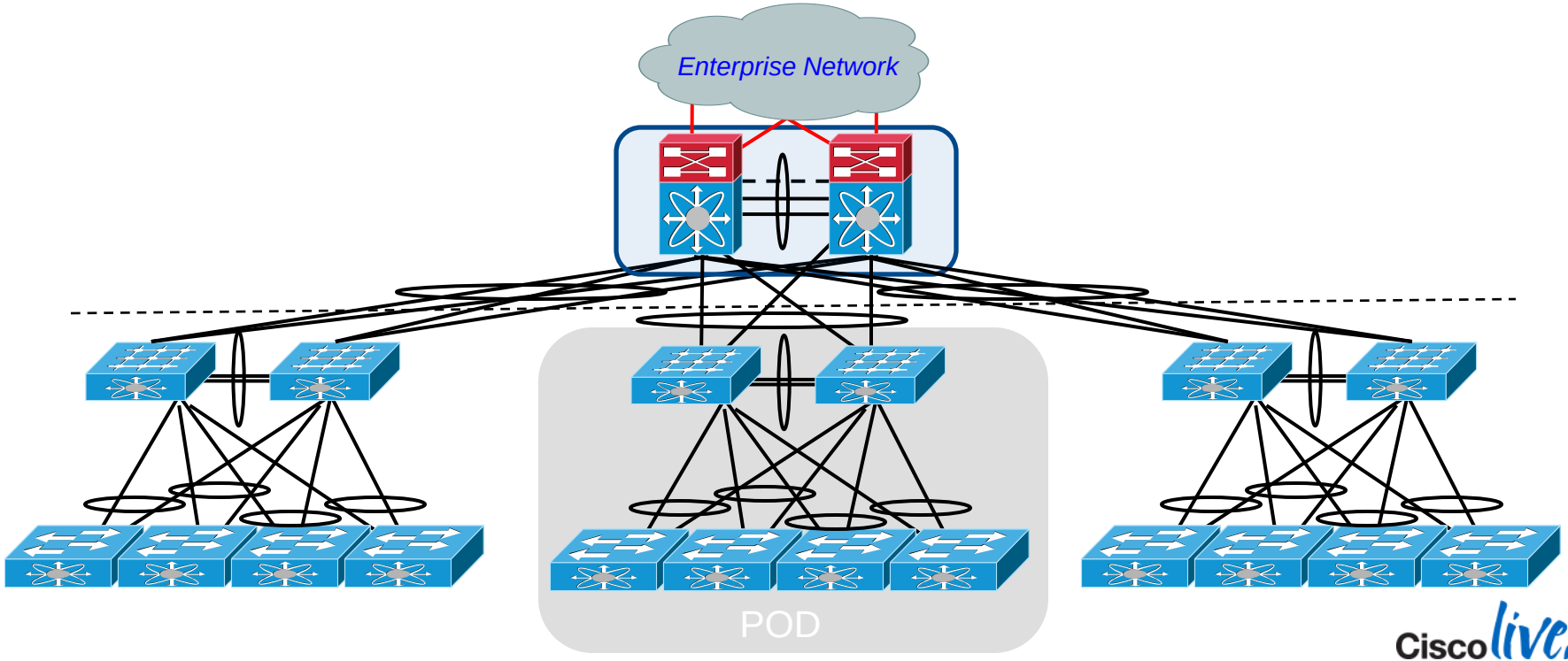
```
5020-2# sh run int po 201
interface port-channel201
  switchport mode trunk
  switchport trunk native vlan 100
  switchport trunk allowed vlan 100-104
  vpc 201
  spanning-tree port type network
```

```
5020-1# show vpc brief vpc 201
vPC status
----------------------------------------------------------------------------
id     Port        Status Consistency Reason                     Active vlans
------ ----------- ------ ----------- ------------------------- -----------
201    Po201       up     success     success                   100-104
2009 May 17 21:56:28 dc11-5020-1 %ETHPORT-5-IF_ERROR_VLANS_SUSPENDED: VLANs 105 on Interface
port-channel201 are being suspended. (Reason: Vlan is not configured on remote vPC interface)
```

Simplifying the Scalable Data Centre

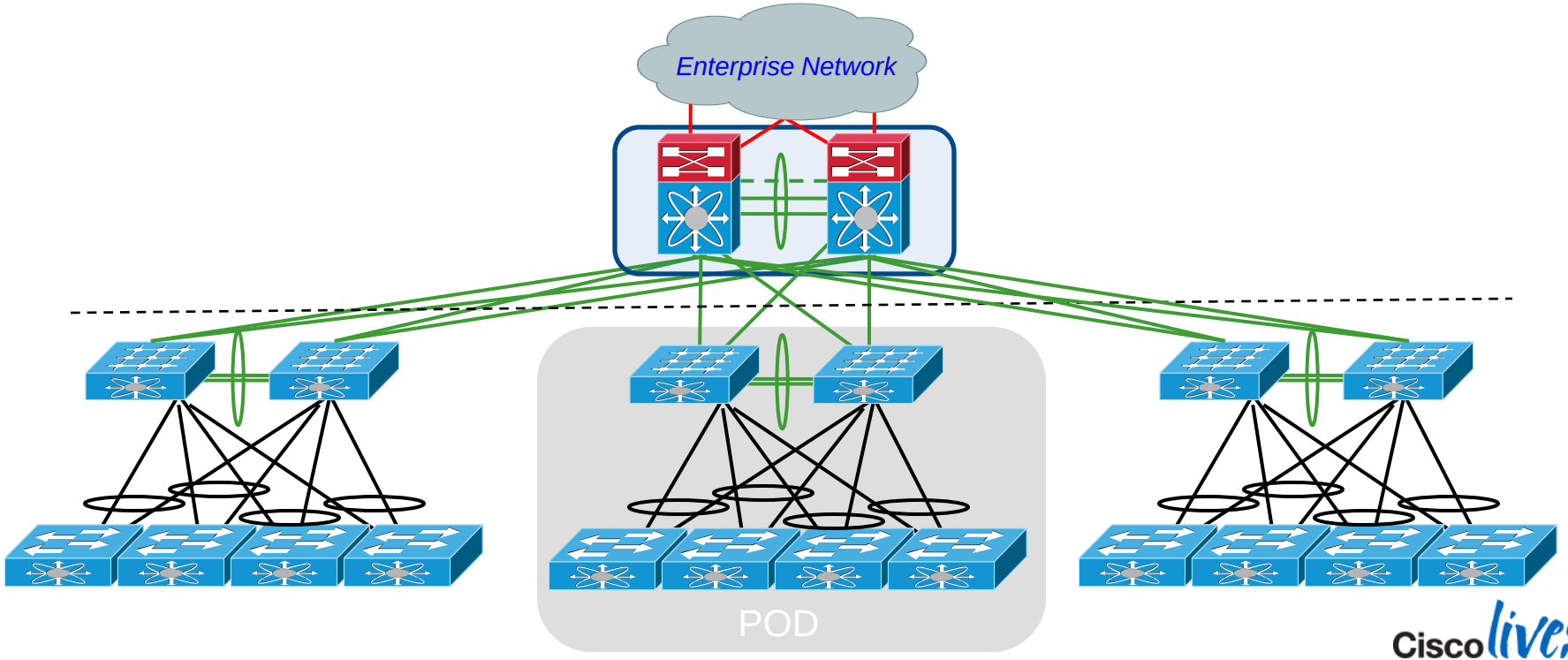
2/3 Tier Data Centre Building Blocks

Needing to scale



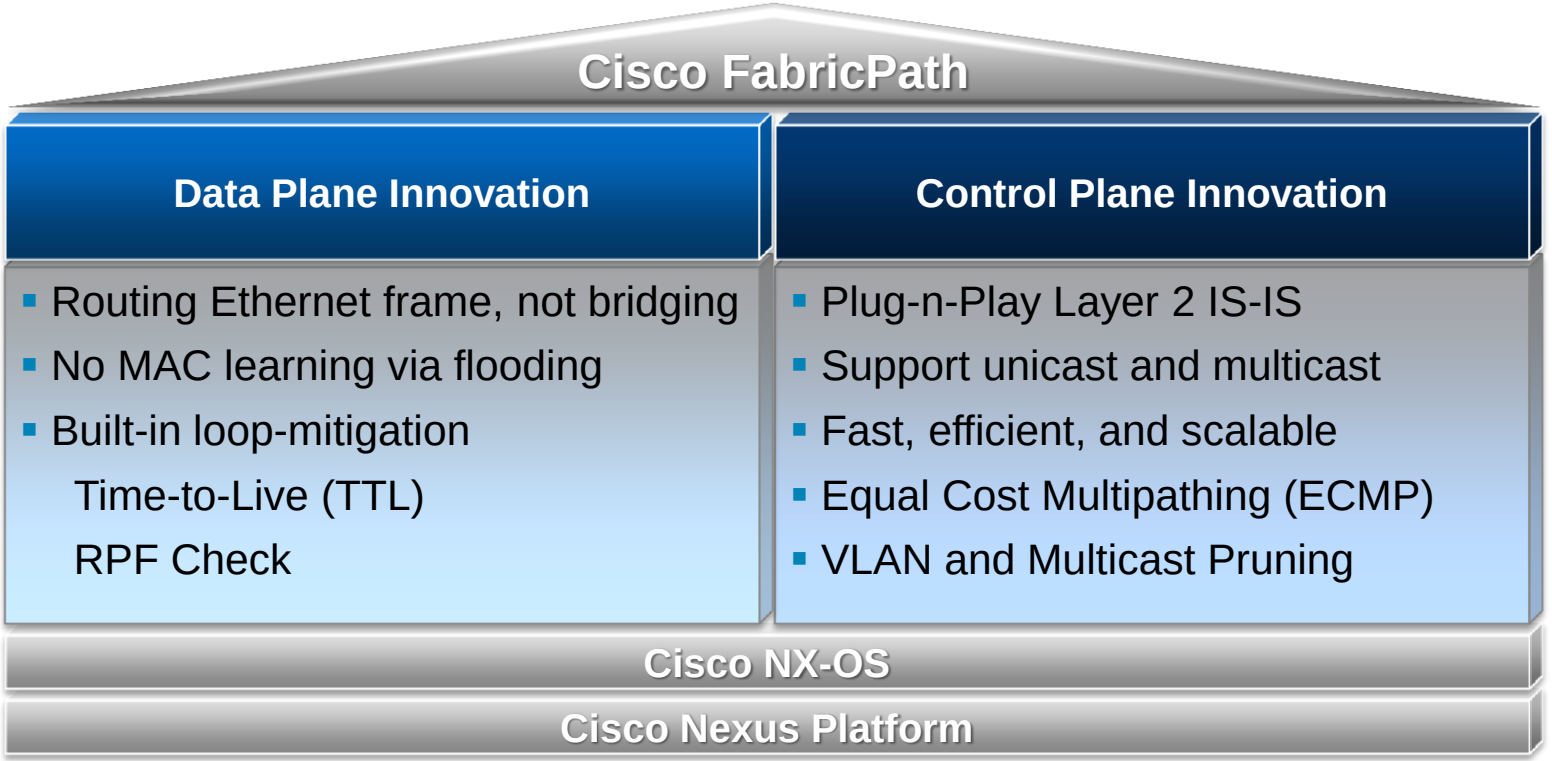
2/3 Tier Data Centre Building Blocks

Moving to Fabric Path



Introduction to Cisco Fabric Path

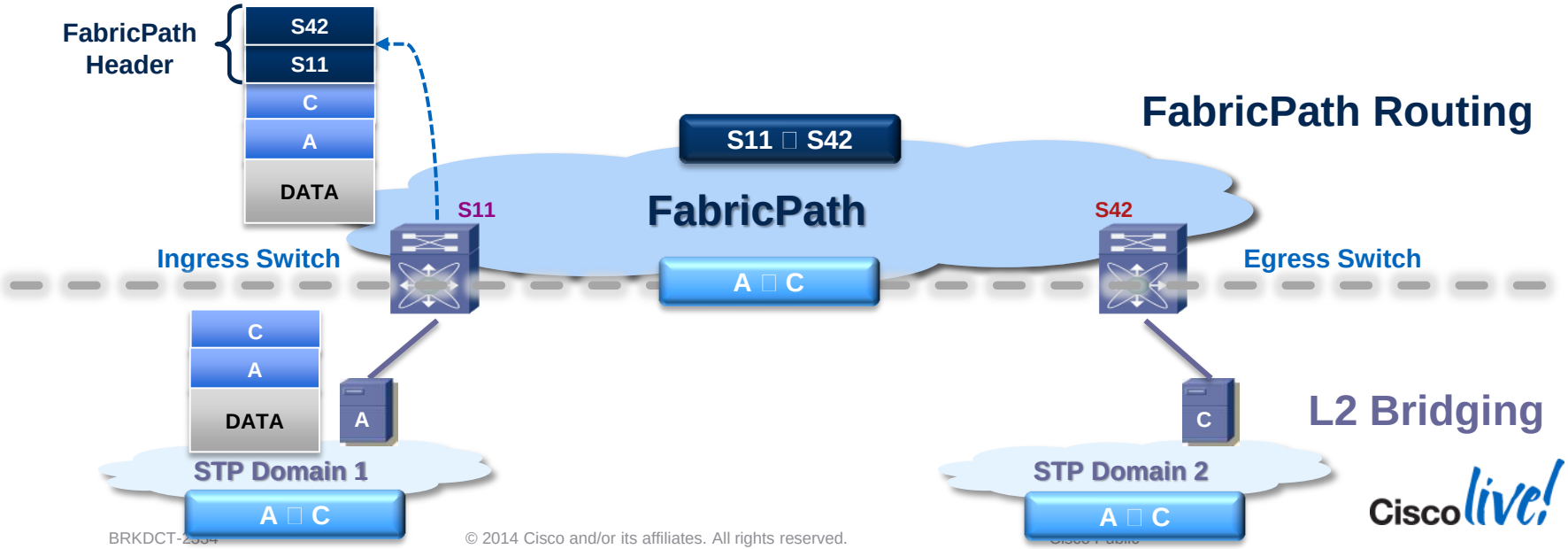
An NX-OS Innovation Enhancing L2 with L3



Data Plane Operation

Encapsulation to creates hierarchical address scheme

- FabricPath header is imposed by ingress switch
- Ingress and egress switch addresses are used to make “Routing” decision
- No MAC learning required inside the L2 Fabric

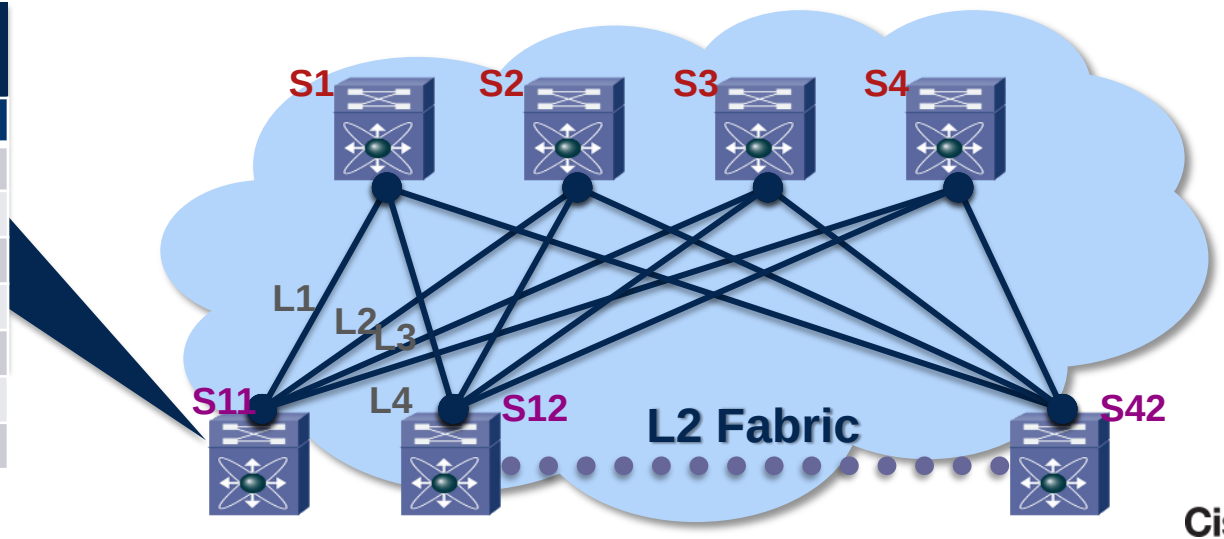


Control Plane Operation

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

FabricPath Routing Table	
Switch	**IF**
S1	L1
S2	L2
S3	L3
S4	L4
S12	L1, L2, L3, L4
…	…
S42	L1, L2, L3, L4



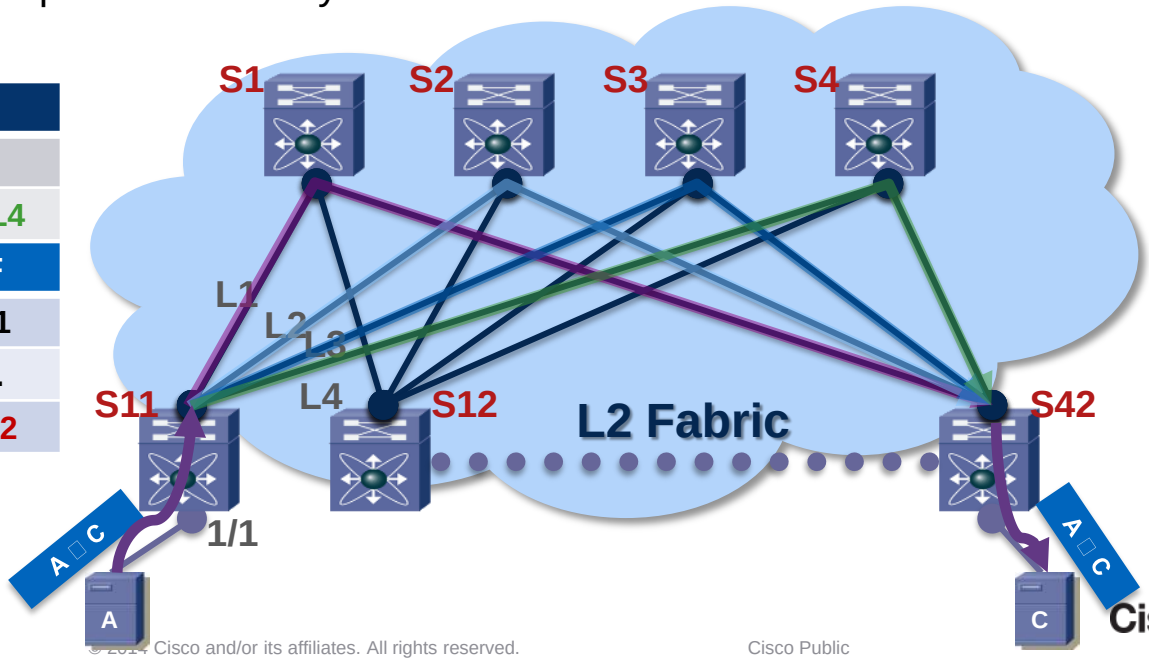
Unicast with FabricPath

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

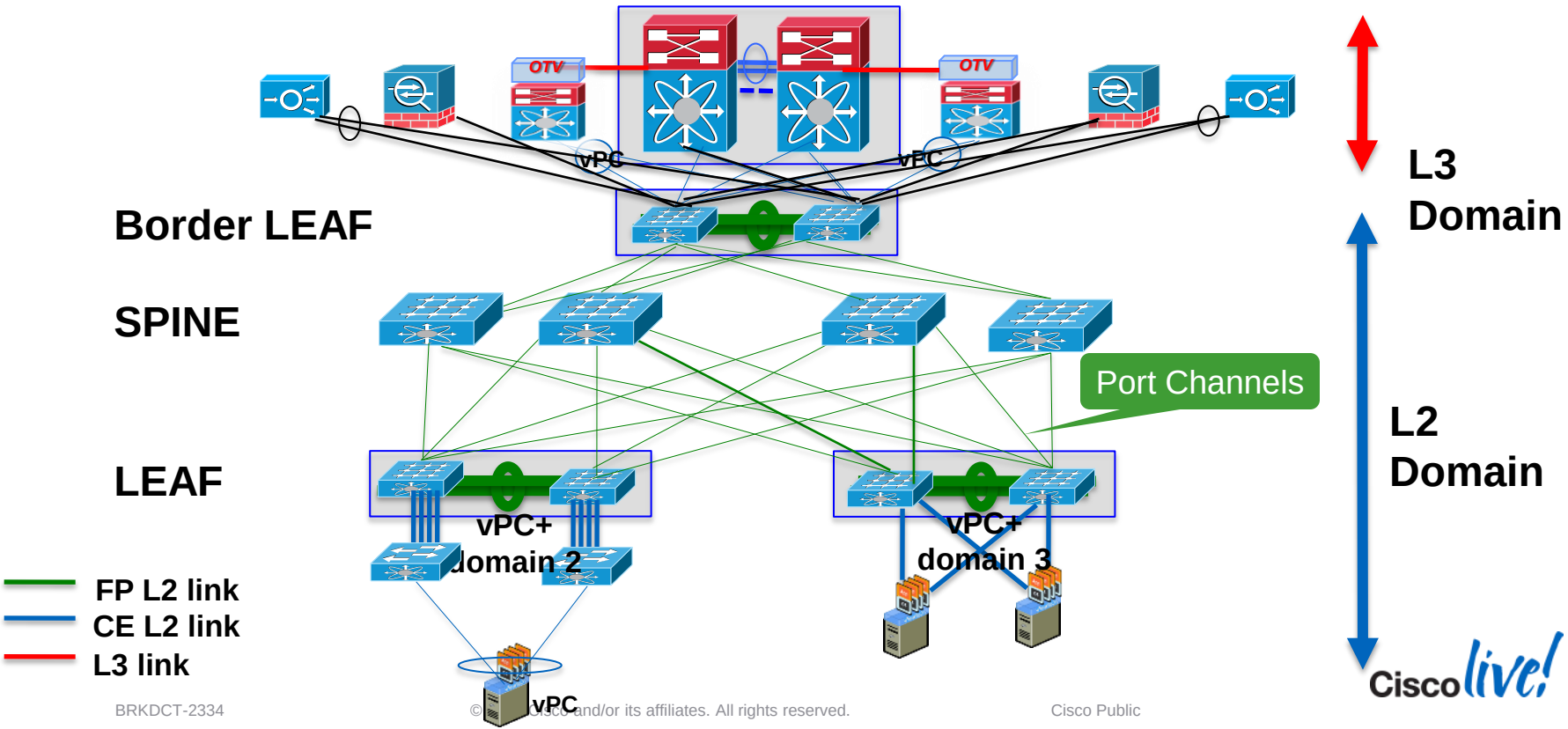
Switch	IF
…	…
S42	L1, L2, L3, L4

MAC	IF
A	1/1
…	…
C	S42

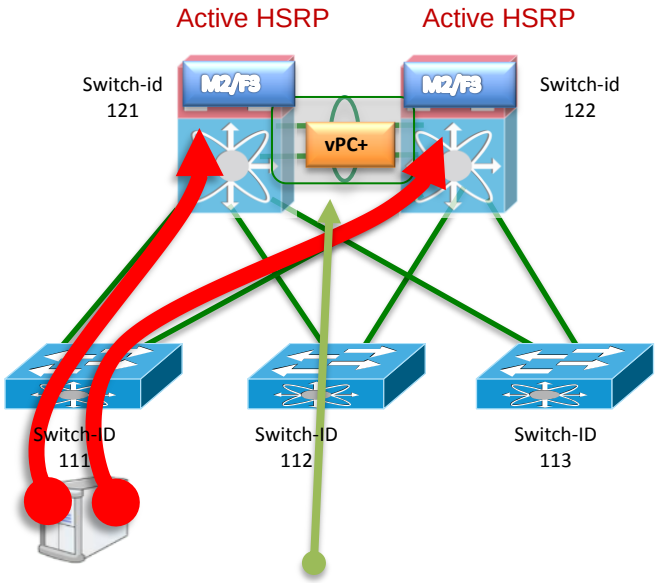


Fabric Path Based Data Centre

Classical Ethernet Isolation



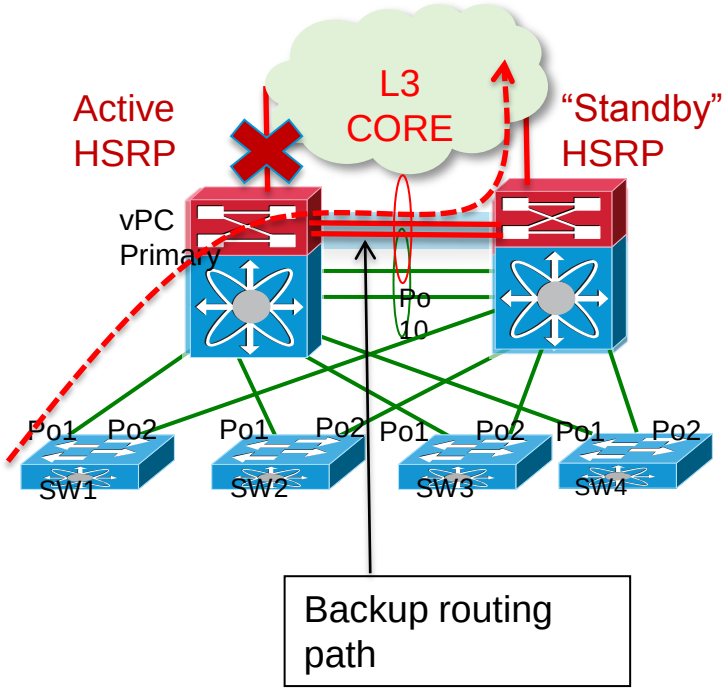
FabricPath - vPC+ at SPINE Layer



Peer-link on F1/F2 cards
Peer-link is a FabricPath link

- It is possible to distribute routed traffic to both spines by using vPC+
- With vPC+ the HSRP MAC is advertised with the same Emulated Switch ID to all edge devices
- Edge switches will have a vMAC entry pointing to Emulated Switch ID
- Each edge switch has an equal cost path to the Emulated Switch ID (via both spine devices)
- All you need to do is to configure a vPC domain and a peer-link
- NO NEED FOR VPC+ PORTS

FabricPath - vPC+ at SPINE Layer
L3 Backup Routing Path

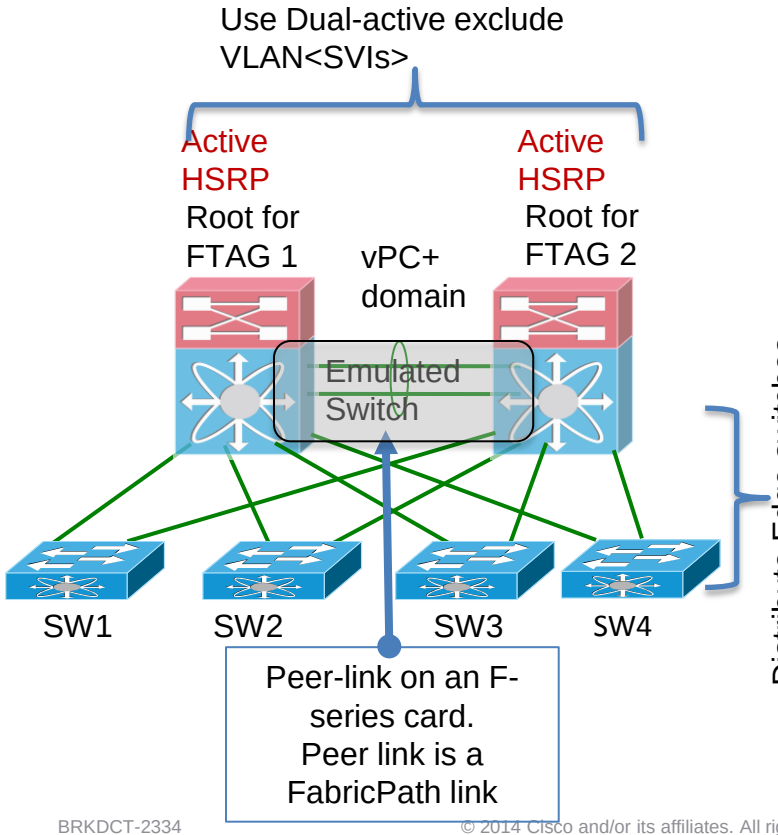


- Like vPC best practices, it is recommended to implement L3 backup routing path with vPC+
- Recommendation:

  Always implement L3 backup routing path.
  Descending order of preference:

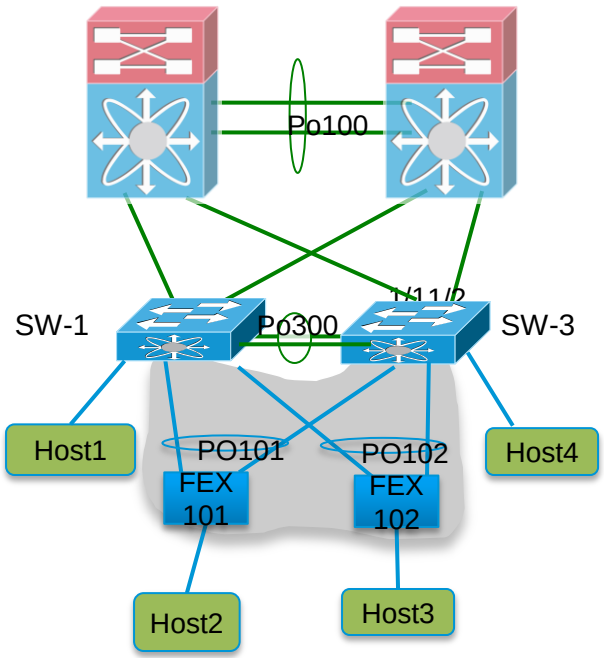
  - 1 – L3 port-channel
  - 2 – L2 port-channel trunk.
    Use a dedicated VLAN (in CE mode) as transit VLAN inside this L2 trunk link.
  - 3 – Use vPC+ Peer-Link. Use a dedicated VLAN (in FP mode) as transit VLAN (N6K required Mode)

FabricPath - vPC+ at SPINE Layer
Conclusion



- No links are normally in vPC mode
- The vPC domain is configured exclusively for these two purposes:
- The HSRP MAC would appear with the G flag for routing on BOTH spines
- The HSRP MAC would be announced with the same emulated switch-id from both spines
- Traffic destined to the HSRP vMAC can go to either SPINES and be routed directly
- Improves failover time

FabricPath - vPC+ at LEAF Layer



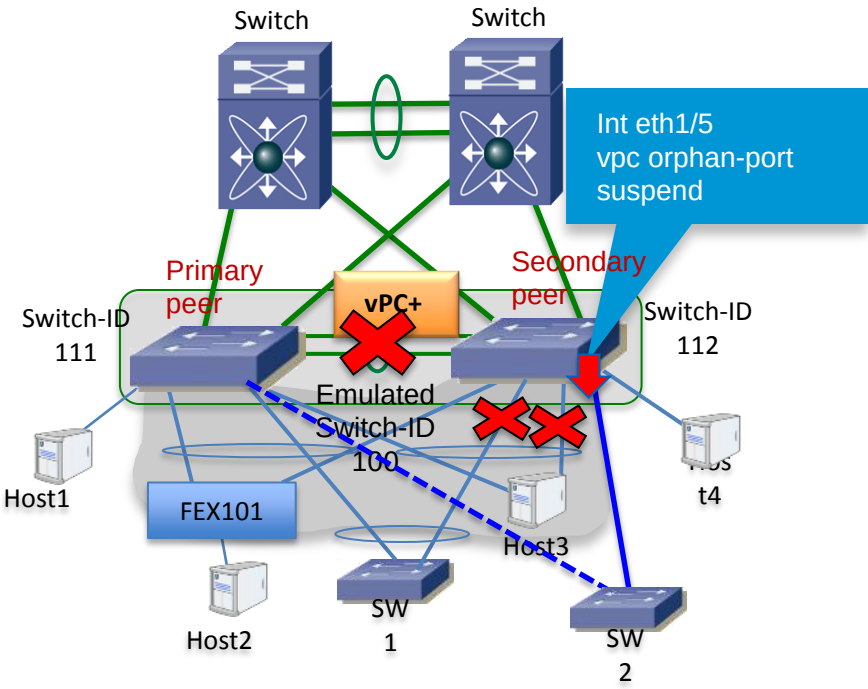
- Using vPC+ at the edge enables several options:
  - Attaching servers with Port-channels
  - Attaching other Classic Ethernet Switches in vPC mode
  - Attaching FEX in Active/Active mode (like in the picture)
- Each pair of Nexus 5500/6000 configured for vPC+ runs Fabricpath with the spine AND on the vPC peer-link too
- The vPC domain is advertised as a different Switch-id (emulated switchID)
- The area highlighted (i.e. its MAC addresses) are advertised as belonging to the “emulated

FabricPath - vPC+ at LEAF Layer

vPC+ Peer-Link Failure : Orphan-Port Suspend

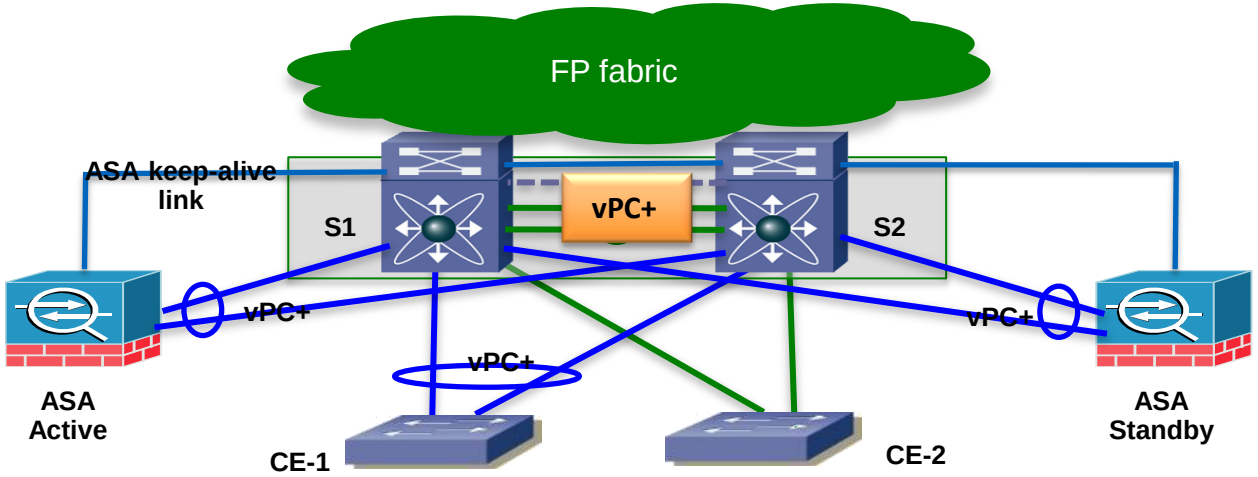
- vPC+ orphan-port suspend is useful to trigger failover for active/standby devices connected to vPC+ domain when peer-link goes down.
  - Orphan-port suspend feature works seamlessly with vPC+
  - Allows to shut down desired physical interface(s) on secondary peer device when vPC+ peer-link fails
- Knob must be configured on individual physical ports:
  - To suspend a port-channel, apply the knob to all member ports
- Recommendation:

  Use vPC+ orphan-port suspend with active/standby type of device (server, switch, FW, LB) to signal Peer-Link down event



FabricPath - vPC+

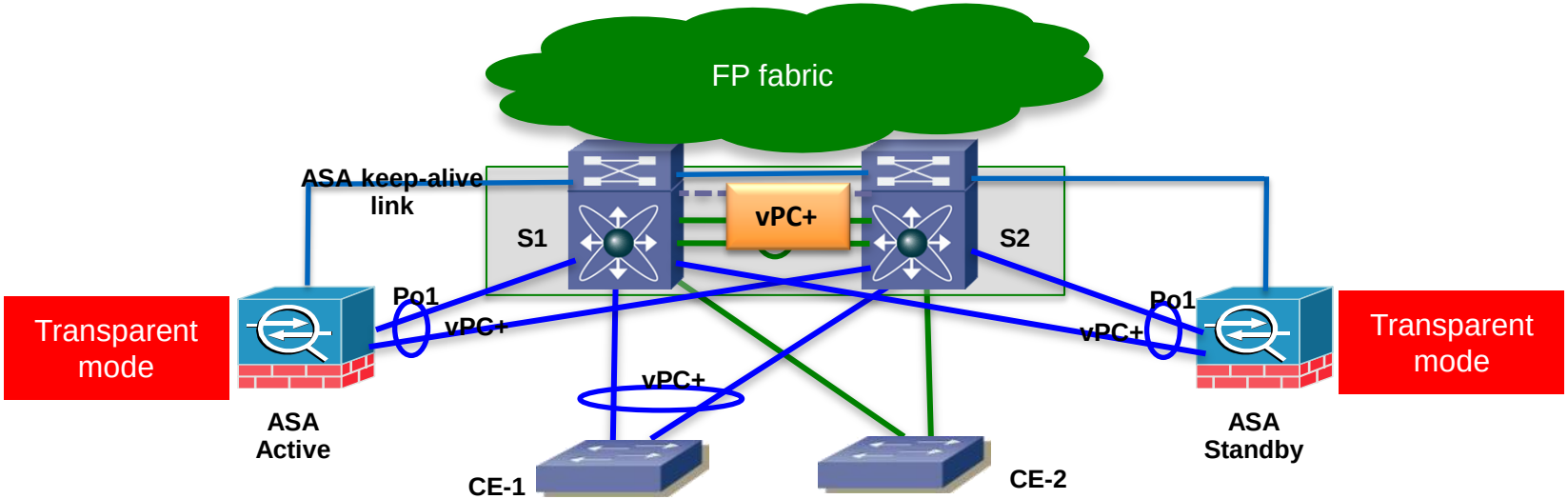
vPC+ And Network Services



- Network services device like FW or LB can be attached to vPC+ domain.
- Network services device can be configured in transparent mode or in routed mode.
- Deploy a dedicated inter-switch link for network services device keep-alive messages.
- Recommendation: Always dual-attach service device to vPC+ domain using port-channel. Active network services device is connected to vPC+ id 1 while Standby network services device is connected to vPC+ id 2.

FabricPath - vPC+

vPC+ And ASA in Transparent Mode



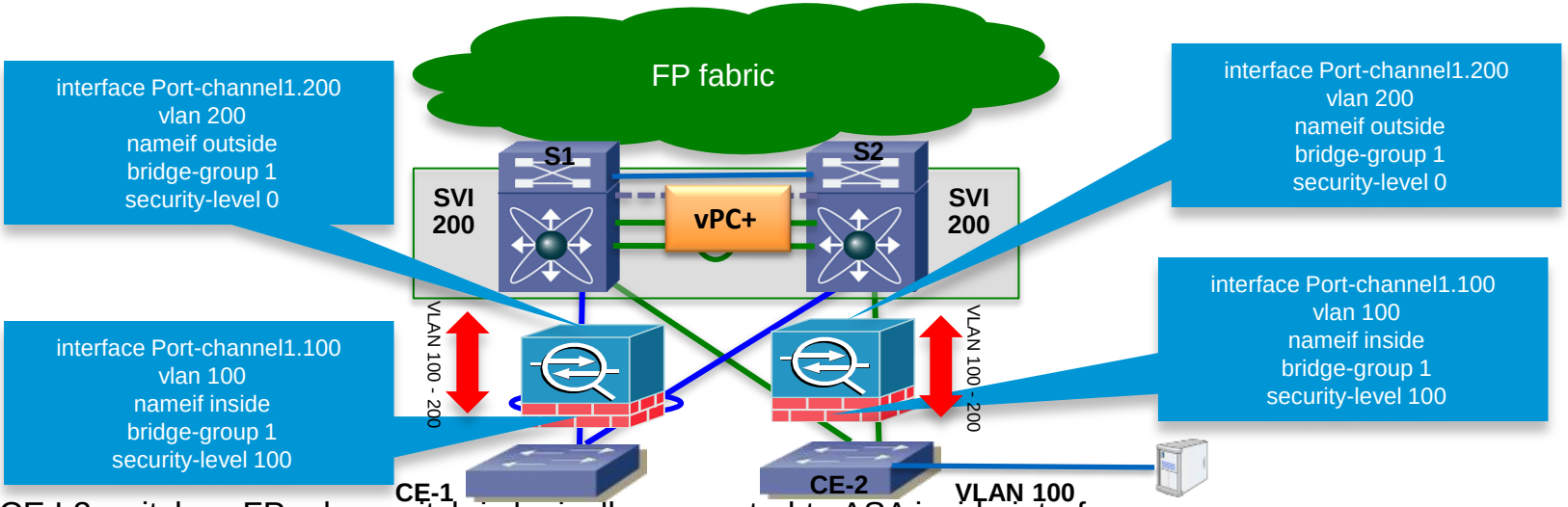
- ASA in transparent mode is dual-attached to vPC+ domain using port-channel
- 2 sub-interfaces are created under the port-channel:
  - Po1.100 for inside interface. Carry VLAN 100
  - Po1.200 for outside interface. Carry VLAN 200.
- Po1.100 and Po1.200 belong to same bridge-group. ASA performs automatic VLAN translation between the 2 sub-interfaces.
- ASA is acting like a bump on the wire; ASA must filter BPDU received from vPC+ domain using access-list.

CAVEAT, Secondary Dual Active protection on the Services Node REQUIRED

FabricPath - vPC+

vPC+ And ASA in Transparent Mode

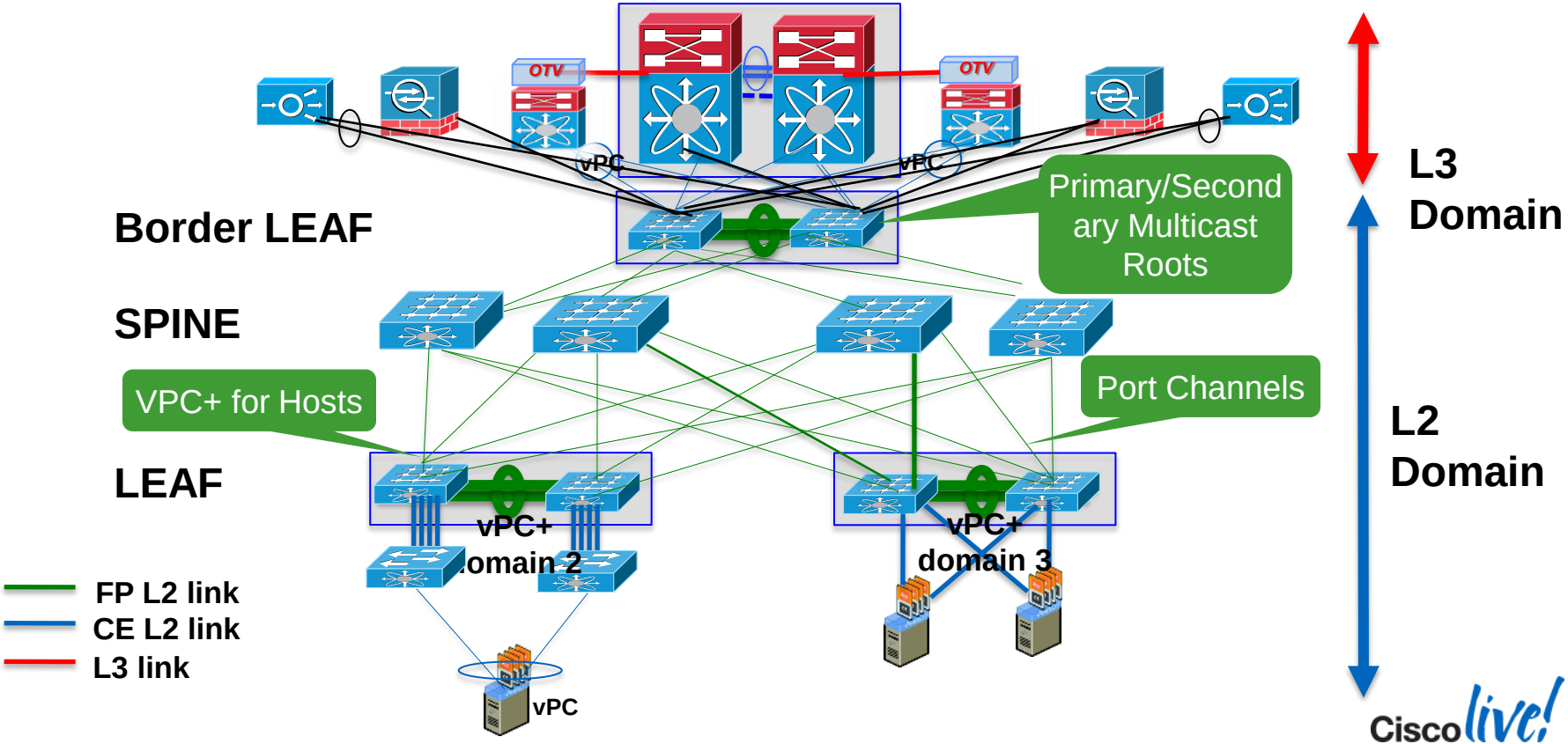
Logical topology



- CE L2 switch or FP edge switch is logically connected to ASA inside interface
- vPC+ domain is logically connected to ASA outside interface
- Default gateway for servers located behind ASA inside interface is configured on interface VLAN located out of ASA outside interface. In our topology:
- Servers located on VLAN 100
- Default gateway for these servers is configured on interface VLAN

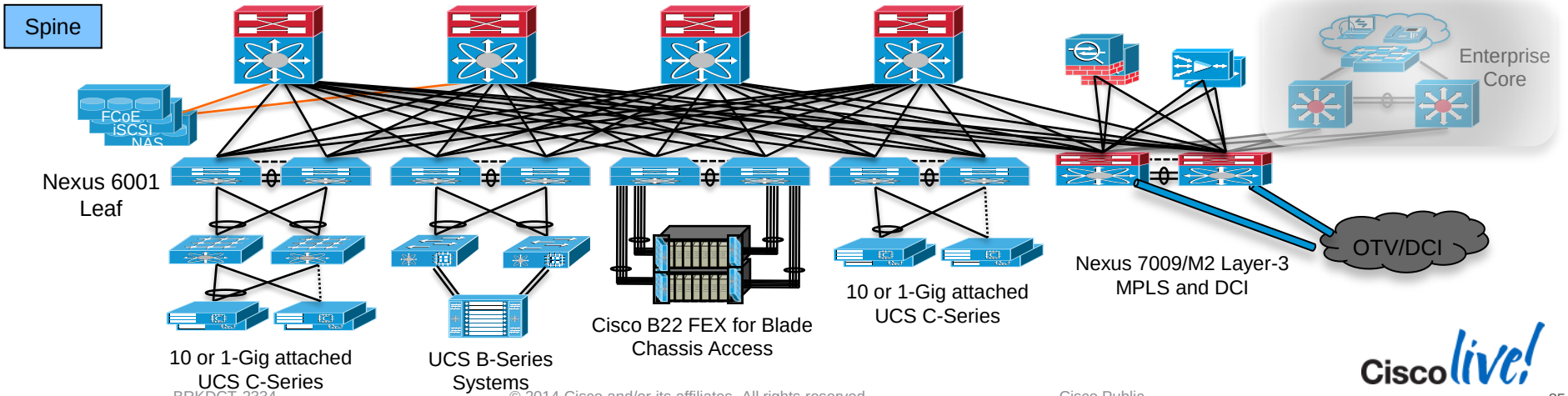
Fabric Path Based Data Centre

Classical Ethernet Isolation



Scalable Leaf/Spine with Border Leaf for Layer-3 with DCI

- Nexus 6004/7000 Spine layer creating Layer-2 FabricPath domain
- Nexus 6001 Leaf switches deployed in vPC+ pairs for edge link redundancy
- FEX, UCS, 3rd-party blade, and direct attach server models supported
- Nexus 7009-M2 Leaf switch pair acting as Layer-3 border for the FabricPath domain
- No MAC learning required on Spine switches with Border Leaf model
- Nexus 7009 with M2 also supports Overlay Transport Virtualisation and MPLS services for DCI

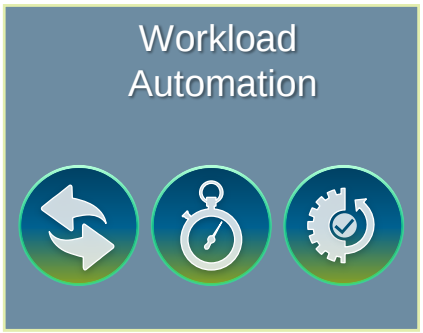


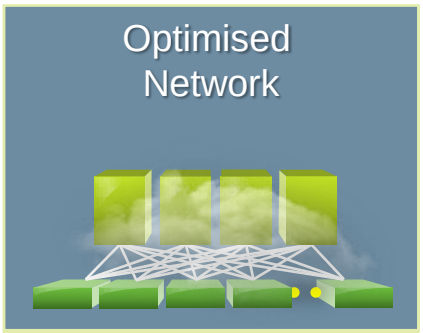



Managing the Data Centre

Dynamic Fabric Automation Architecture
Innovative Building Blocks

Bundled functions are modular and simplified for scale and automation

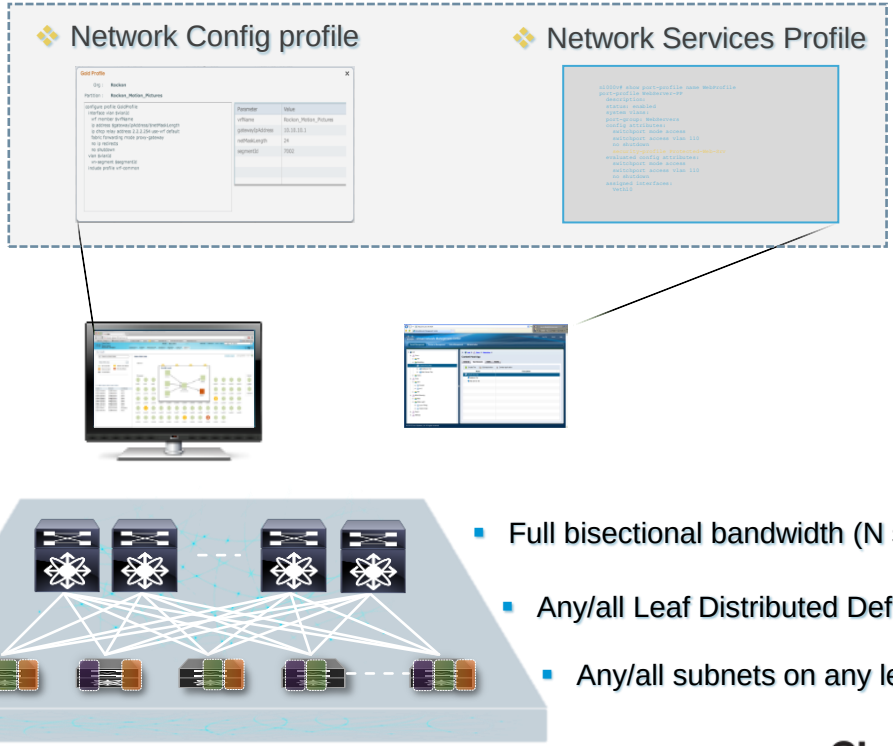




Cisco Dynamic Fabric Automation
Scale, Resiliency and Efficiency

Advantages

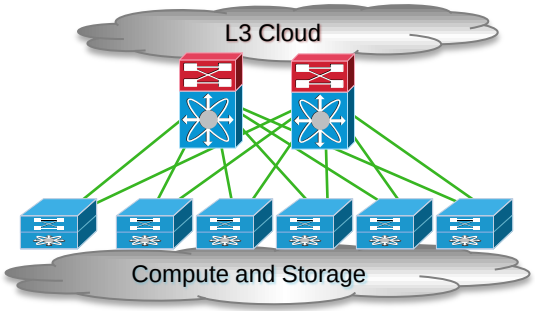
- Any subnet, anywhere, rapidly
- Reduced Failure Domains
- Extensible Scale & Resiliency
- Profile Controlled Configuration



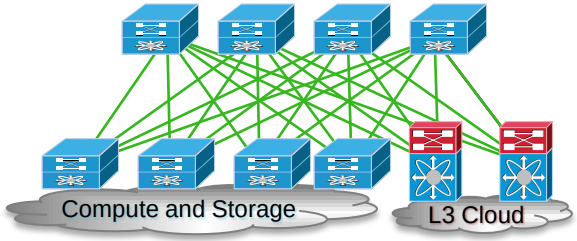
- Full bisectional bandwidth (N spines)
- Any/all Leaf Distributed Default Gateways
- Any/all subnets on any leaf

Cisco Dynamic Fabric Automation
Flexible Topologies Support

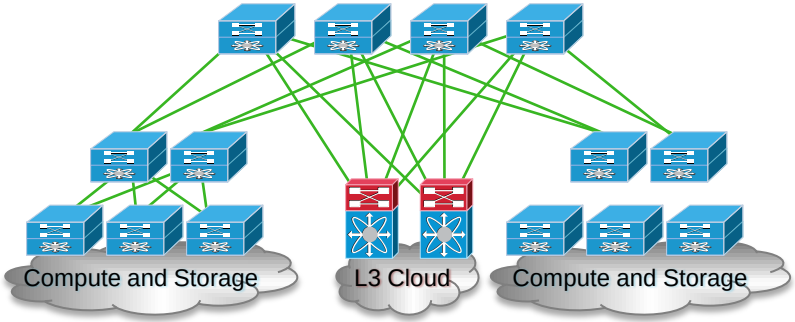
Traditional Access/Aggregation



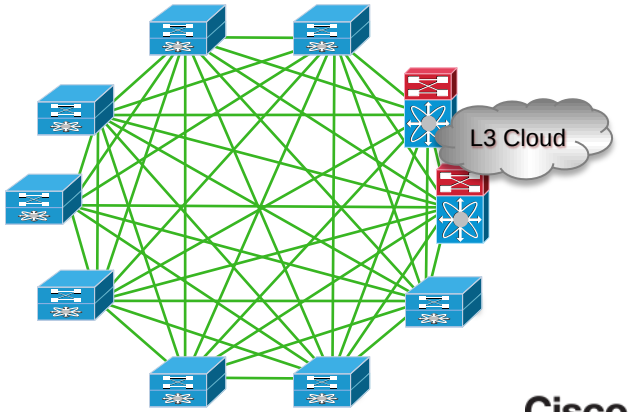
Folded CLOS



Three Tiers (Fat Tree)



Full Mesh



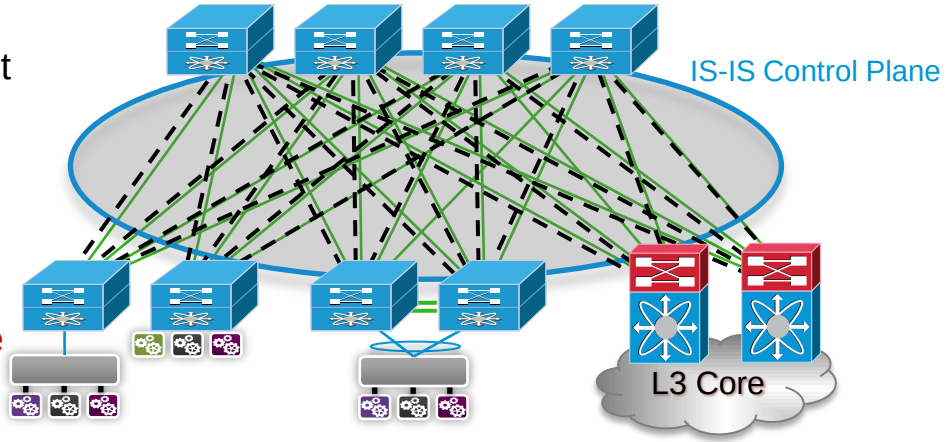
Control Plane
1 - IS-IS as Fabric Control Plane

ISIS for fabric link state distribution

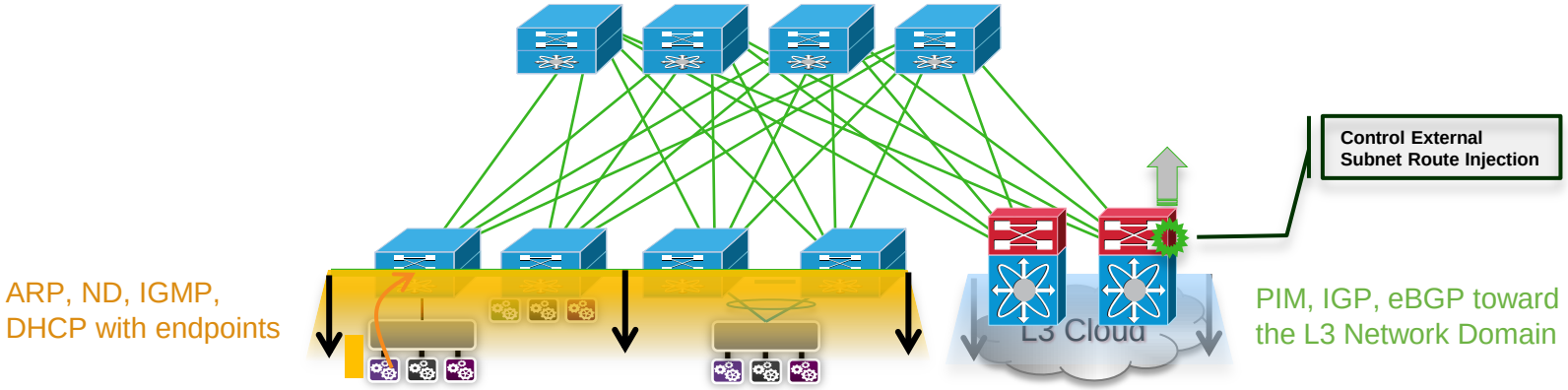
- Fabric node reachability for overlay encap
- Building multi-destination trees for multicast and broadcast traffic
- Quick reaction to fabric link/node failure
- Enhanced for mesh topologies

Fabric Control Protocol doesn't distribute

- Host Routes
- Host originated control traffic
- Server subnet information

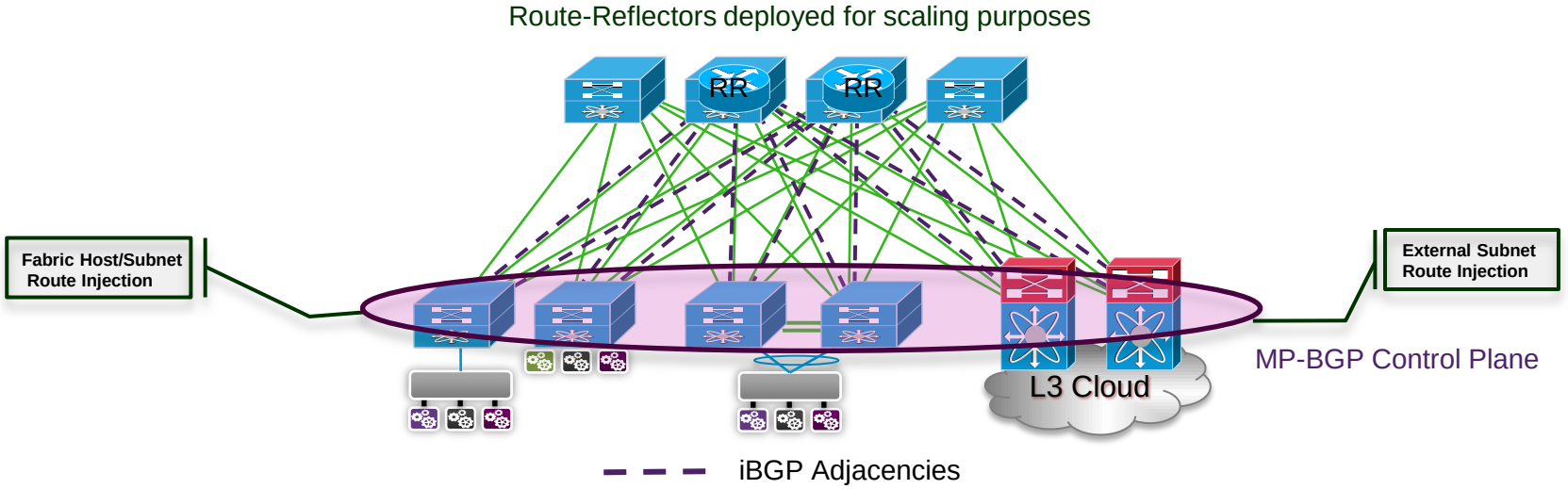


Control Plane
2 – Host Originated Protocols Containment



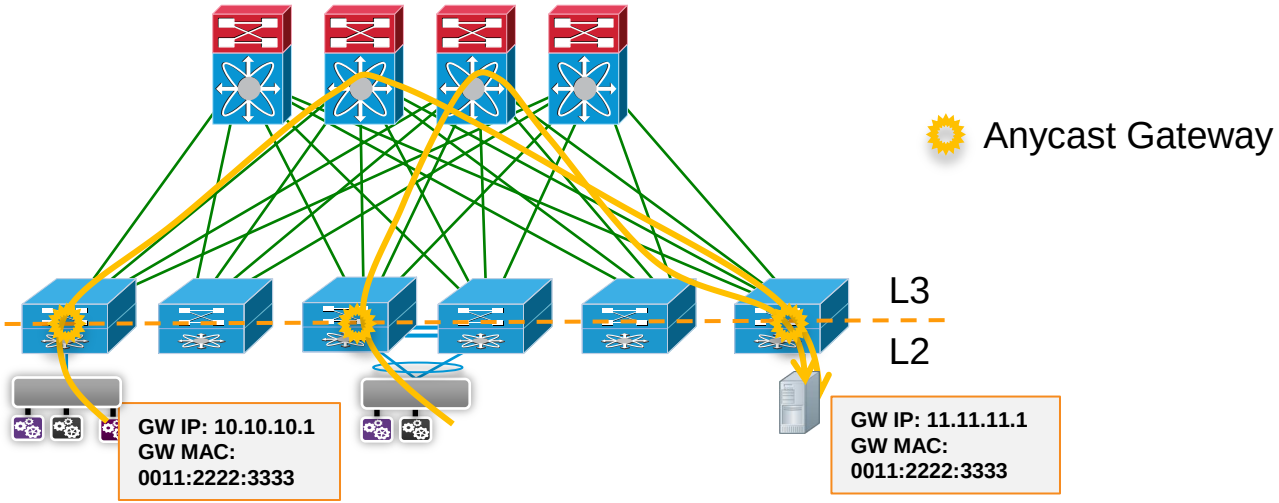
- ARP, ND, IGMP, DHCP originated on servers are terminated on Leaf nodes
- Contain floods and failure domains, distribute control packet processing
- Terminate PIM, OSPF, eBGP from external networks on Border Leafs

Control Plane
3 – Host and Subnet Route Distribution



- Host Route Distribution decoupled from the Fabric link state protocol
- Use MP-BGP on the leaf nodes to distribute internal host/subnet routes and external reachability information
- MP-BGP enhancements to carry up to 100s of thousands of routes and reduce convergence time

Optimised Network
Distributed Gateway at the Leaf



- Any subnet anywhere => Any leaf can instantiate any subnet

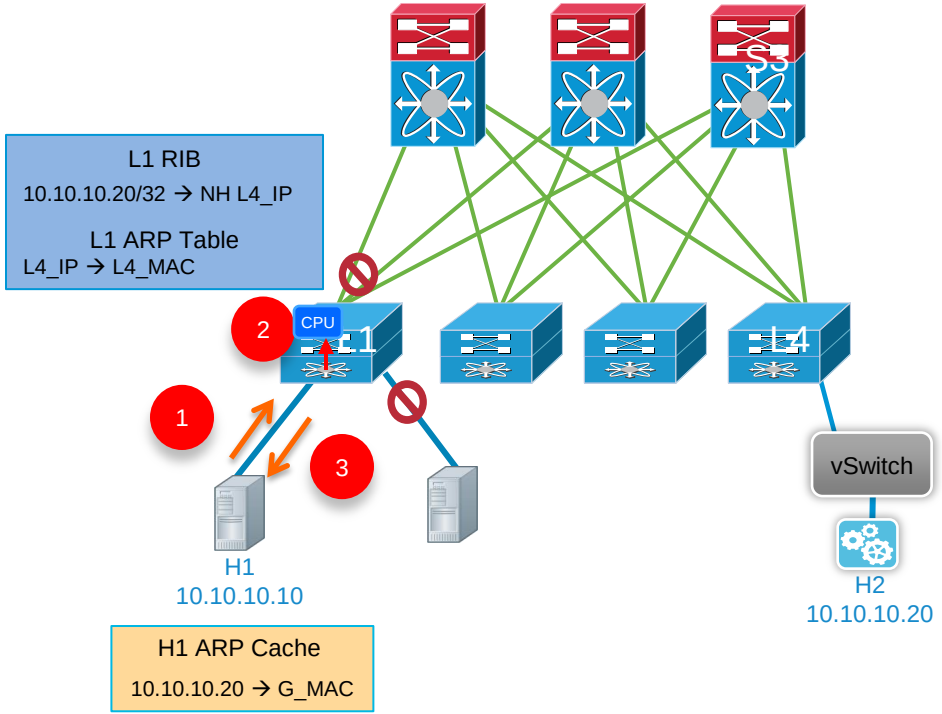
  All leafs share gateway IP and MAC for a subnet (No HSRP)

  ARPs are terminated on leafs, No Flooding beyond leaf
- Facilitates VM Mobility, workload distribution, arbitrary clustering
-

Optimised Network
IP Forwarding within the Same Subnet

1. H1 sends an ARP request for H2 – 10.10.10.20
2. The ARP request is intercepted at the leaf L1 and punted to the Sup
3. Assuming a valid route to H2 does exist in the Unicast RIB, L1 sends the ARP reply with the G_MAC so that H1 can build its ARP cache

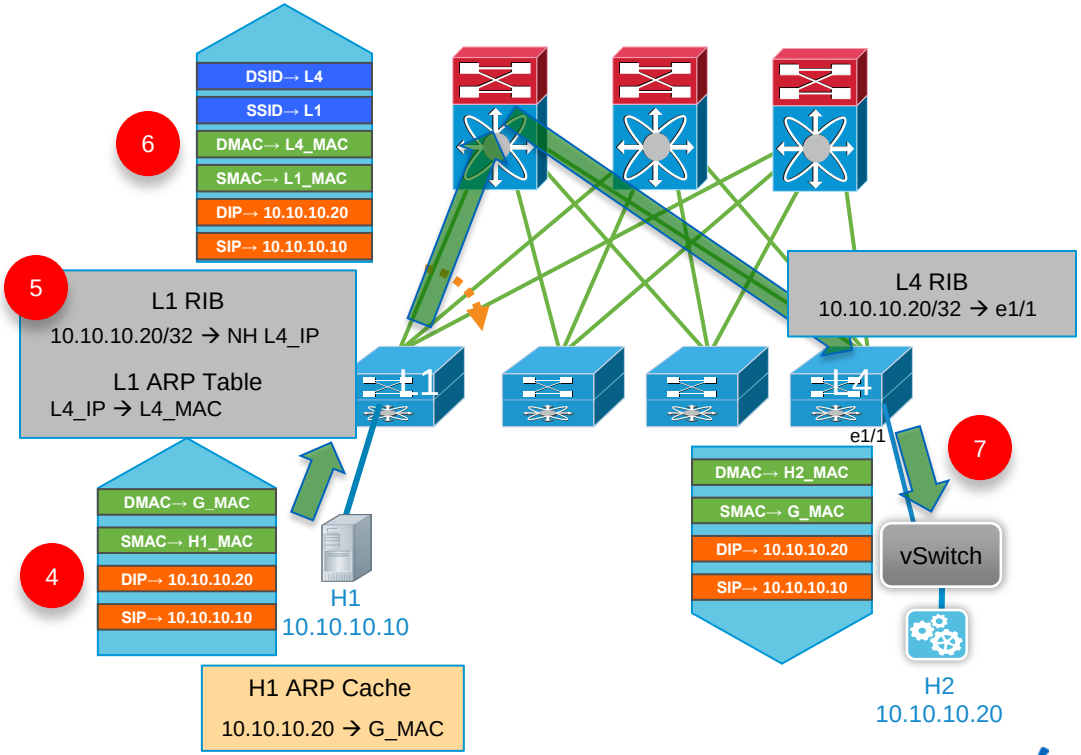
Note: the ARP request is NOT flooded across the Fabric, nor out of other local interfaces belonging to the same L2 domain



Optimised Network

IP Forwarding within the Same Subnet (2)

4. H1 generates a data packet with G_MAC as destination MAC
5. L1 receives the packet, remove the L2 header and performs Layer 3 lookup for the destination
6. L1 adds the Layer 2 and the FP headers and forwards the FP frame across the Fabric, picking one of the 3 equal cost paths available via S1, S2 and S3
7. L4 receives the packet, strips off the FP and L2 headers and performs L3 lookup and forwarding toward H2



Optimised Network
Introducing L3 Conversational Learning

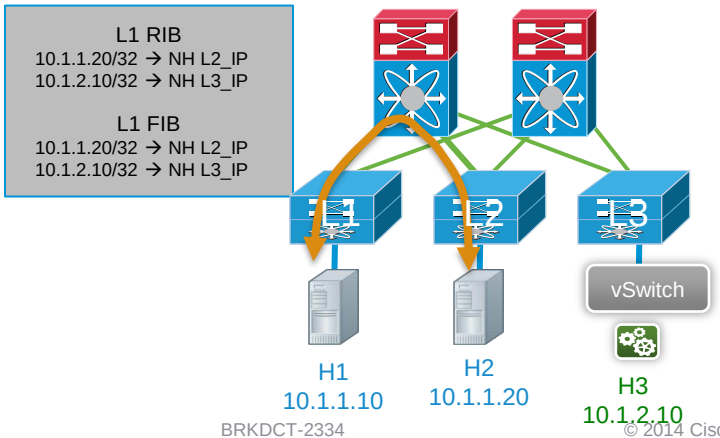
- Use of /32 host routes may lead to scaling issues if all the routes are installed in the HW tables of all leaf nodes

  L3 conversational learning is introduced to alleviate this concern

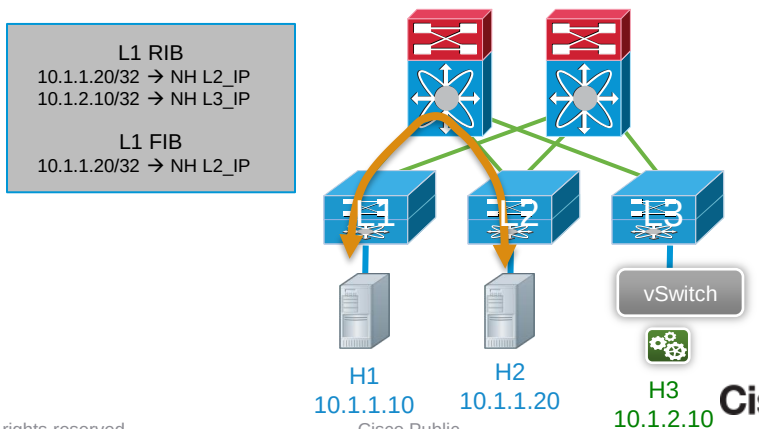
  Disabled by default → all host routes are programmed in the HW

- With L3 conversational learning, host routes for remote endpoints will be programmed into the HW FIB (from the SW RIB) upon detection of an active conversation with a local endpoint

Default Behaviour (No L3 Conversational Learning)



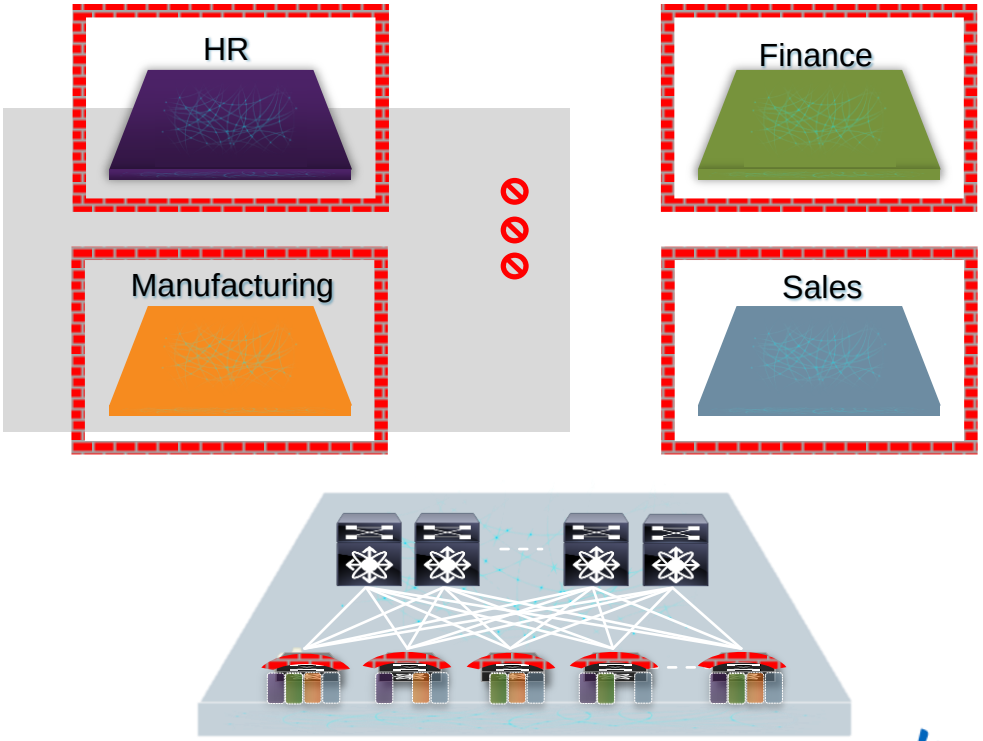
After Enabling L3 Conversational Learning



Virtual Fabrics for Public or Private Cloud Environments

Advantages

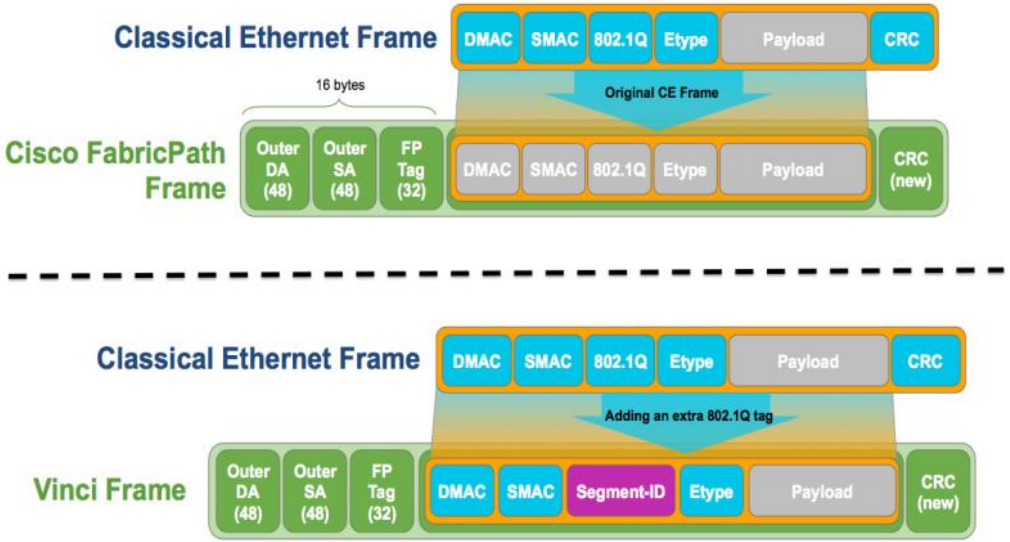
- Any workload, any vFabric, rapidly
- Scalable Secure vFabrics
- vFabric Tenant Visibility
- Routing/Switching Segmentation



Virtual Fabrics
Introducing Segment-ID Support

- Traditionally VLAN space is expressed over 12 bits (802.1Q tag)
  - Limits the maximum number of segments in a data centre to 4096 VLANs
- DFA leverages a double 802.1Q tag for a total address space of 24 bits
  - Support of ~16M L2 segment (10K targeted at FCS)
- Segment-ID is hardware-based innovation offered by leaf and spine nodes part of the Integrated Fabric

FabricPath Frame Format



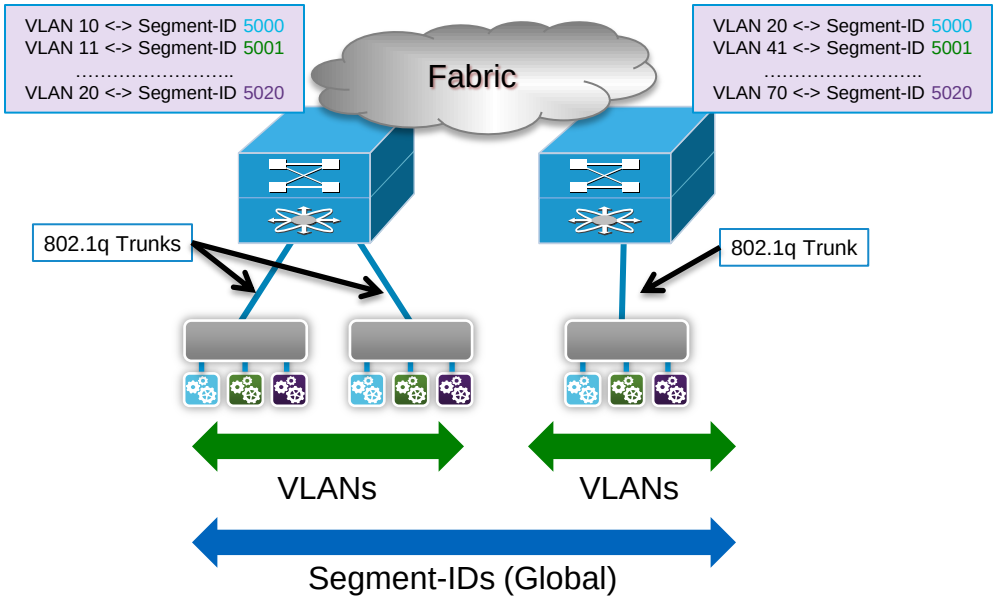
Integrated Fabric Frame Format

Segment-ID = 802.1Q 802.1Q

Virtual Fabrics
802.1Q Tagged Traffic to Segment-ID Mapping

- Segment-IDs are utilised for providing isolation at L2 and L3 across the Integrated Fabric
- 802.1Q tagged frames received at the leaf nodes from edge devices must be mapped to specific Segments
- The VLAN-Segment mapping can be performed on a leaf device level

  VLANs become locally significant on the leaf node and 1:1 mapped to a Segment-ID
- Segment-IDs are globally significant, VLAN IDs are locally significant



Workload Automation & Open Environment

Orchestration Stack
UCS Director, OpenStack, vCloud Director

Compute & Storage Orchestration

Network & Services Orchestration

Auto-config Triggers

VDP

DHCP/ARP-ND

Data Packet Driven

Programmatic

N1kv/OVS

Virtual Machines

Physical Machines

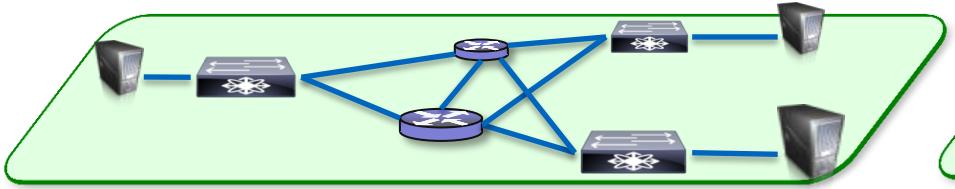
DCNM (CPoM)

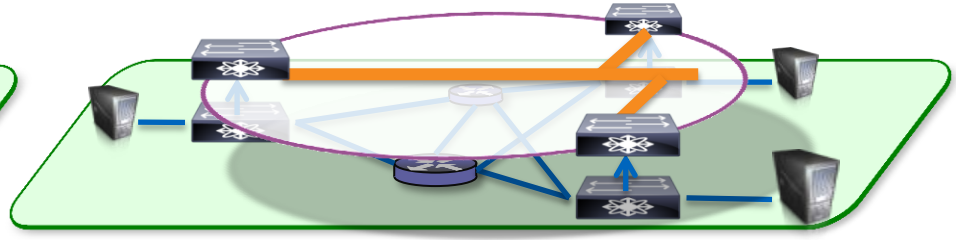
*VDP (VSI Discovery and Configuration Protocol) is IEEE 802.1Qbg Clause 41




Overlays

What about an Overlay?





Robust Underlay/Fabric

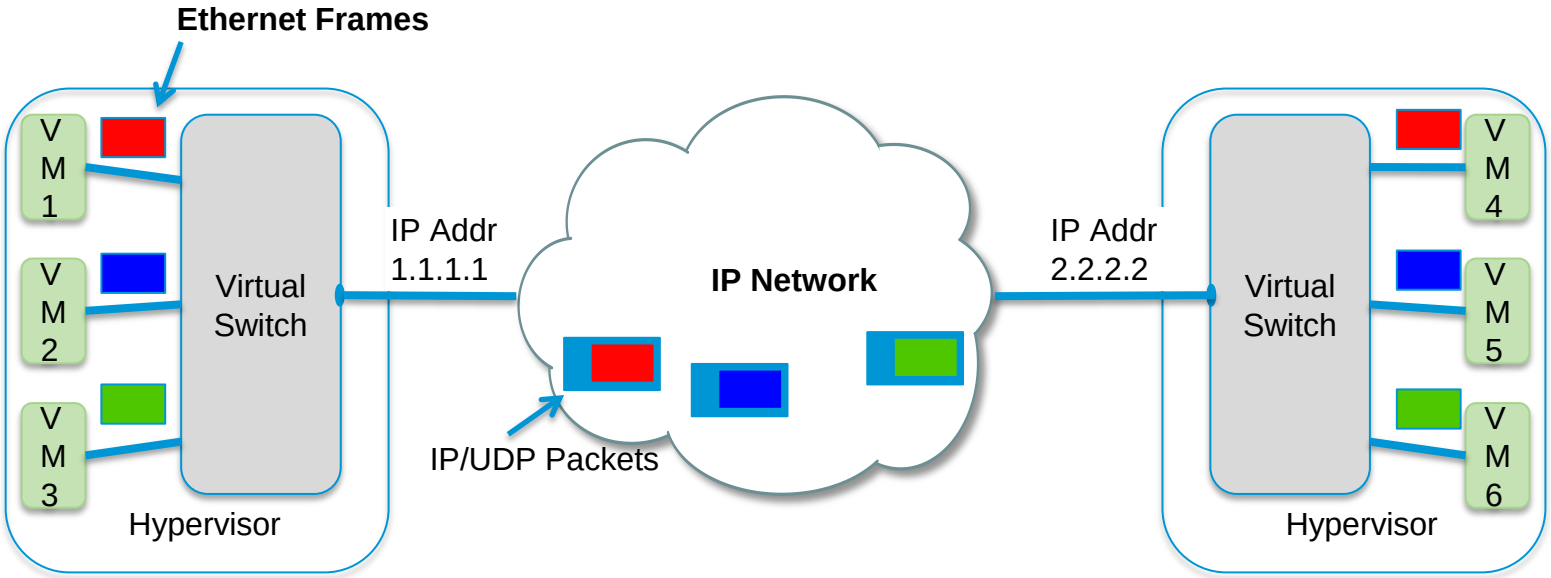
- High Capacity Resilient Fabric
- Intelligent Packet Handling
- Programmable & Manageable

Flexible Overlay Virtual Network

- Mobility – Track end-point attach at edges
- Scale – Reduce core state
  - Distribute and partition state to network edge
- Flexibility/Programmability
  - Reduced number of touch points

What is a Virtual Overlay Technology ?

- Servers perform data encapsulation and forwarding
- SW based virtual switches instantiate customer topologies



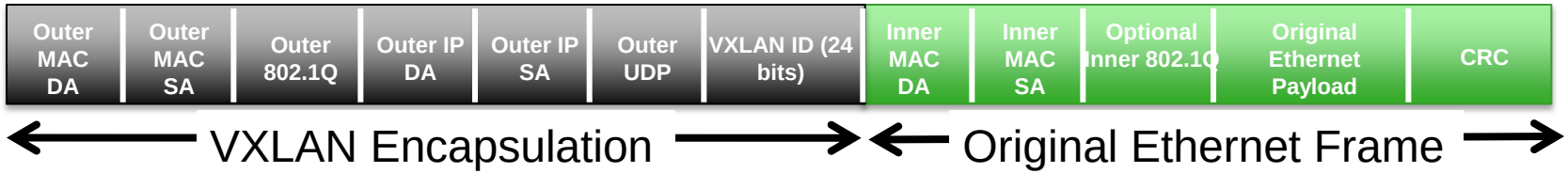
Virtual Overlay Encapsulations and Forwarding

- Ethernet Frames are **encapsulated** into an IP frame format
- **New control logic** for learning and mapping VM identity (MAC address) to Host identity (IP address)
- **Three Hypervisor based Overlays**
  - VXLAN Virtual Extensible Local Area Newtork
  - NVGRE, Network Virtualisation Generic Router Encapsulation
  - STT Stateless Transport Tunnelling
- **Network Based Overlays**
  - OTV, Overlay Transport Virtualisation
  - VPLS, EVPN-PBB Ethernet VPN-Provider Backbone Bridging
  - FabricPath
  - VXLAN and NVGRE

Virtual Extensible Local Area Network (VXLAN)

- Ethernet in IP overlay network
  - Entire L2 frame encapsulated in UDP
  - 50 bytes of overhead
- Include 24 bit VXLAN Identifier
  - 16 M logical networks
  - Mapped into local bridge domains
- VXLAN can cross Layer 3

- Tunnel between VEMs
  - VMs do NOT see VXLAN ID
- IP multicast used for L2 broadcast/multicast, unknown unicast
- Technology submitted to IETF for standardisation
  - With Cisco, Arista, VMware, Citrix, Red Hat and Others



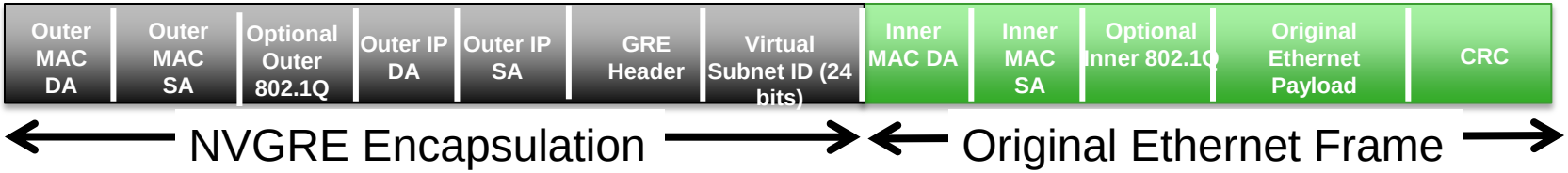
NVGRE, Network Virtualisation GRE

- https://datatracker.ietf.org/doc/draft-sridharan-virtualisation-nvgre/
- Generic Routing Encapsulation (GRE) header for Network Virtualisation (NVGRE) in multi-tenant data centres
- 24 Bit Segment ID
- NVGRE Encapsulation 42 bytes
- Port Channel Load Distribution will be polarised
  - Most current switches do not hash on the GRE header
- Firewall ACL will need to allow GRE protocol.
- Forwarding Logic
  - NVGRE: IETF draft assumes end points knows destination via management plane provisioning, control plane distribution, or data plane learning

NVGRE

- Ethernet in IP overlay network
  - Entire L2 frame encapsulated in GRE
  - 42 bytes of overhead
- Include 24 bit Virtual Subnet Identifier, VSID
  - 16 M logical networks
  - Mapped into local bridge domains
- NVGRE can cross Layer 3

- Tunnel between End Points
  - VMs do NOT see NVGRE Encapsulation Hypervisor removes.
- IP multicast used for L2 broadcast/multicast, unknown unicast
- Technology submitted to IETF for standardisation
  - With Microsoft, Intel, Broadcom and Others

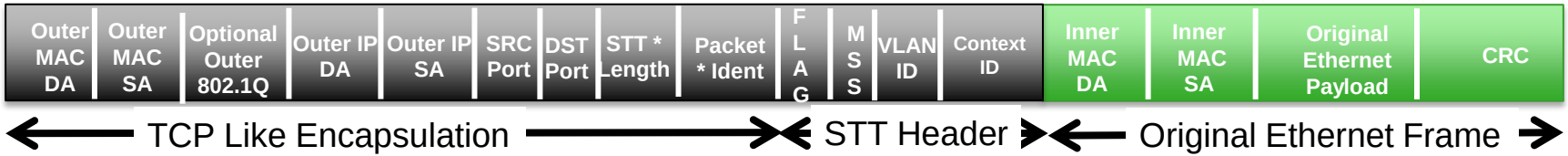


STT, Stateless Transport Tunnelling

- http://www.ietf.org/id/draft-davie-stt-03.txt
- STT (Stateless Transport Tunnelling), a tunnel encapsulation that enables overlay networks to be built in virtualised networks.
- Leverage TCP Segment Offload (TSO) in server NIC for improved performance (NOT A REAL TCP FLOW)
- Any Stateful device that is not STT aware will drop the traffic
- No public implementation exists,
- Node Types/Functions based on implementation requirements
  - Transport Nodes, end point
  - Controller, management/provisioning
  - Gateway Nodes: provide connectivity for Physical and non-NSX-OVS Nodes

Stateless Transport Tunnelling

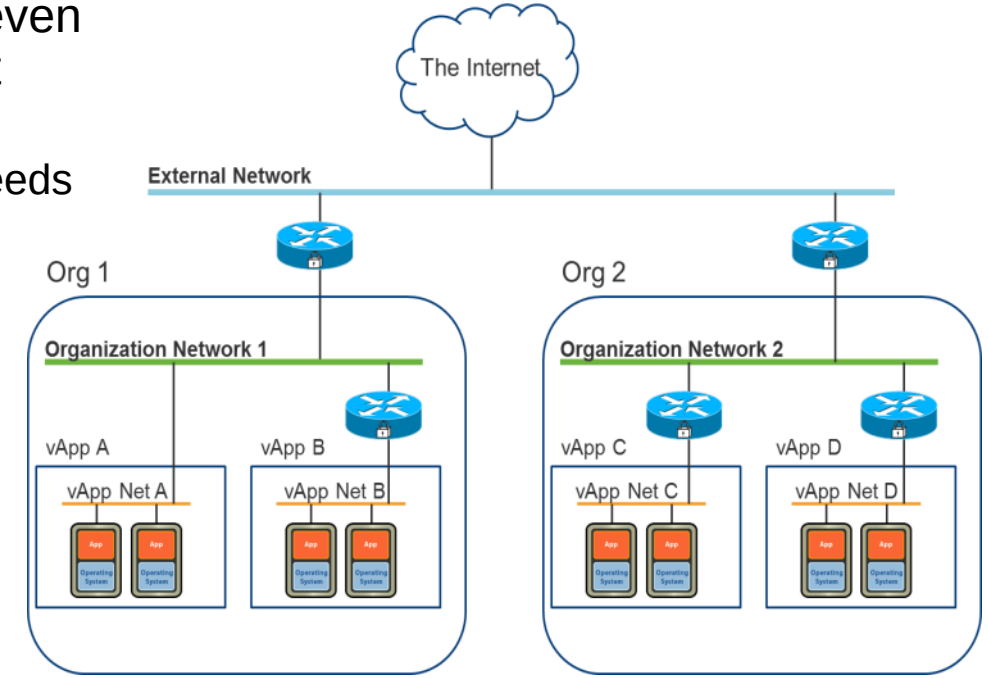
- Ethernet in IP overlay network
  - Entire L2 frame encapsulated in TCP like header
- The STT header
  - Only the first packet of a large segment has header, therefore reassembly is required.
  - Include 12 bit VLAN ID and 64 bit Context ID
- Service node acts as a Broadcast and multicast Server (BUS)
- 50 active VNICs supported in 2.x
- Multicast becomes broadcast in logical network
- IETF draft submitted
  - By Nicira



Multi-Tenancy and vApps Drive Layer2 Segments

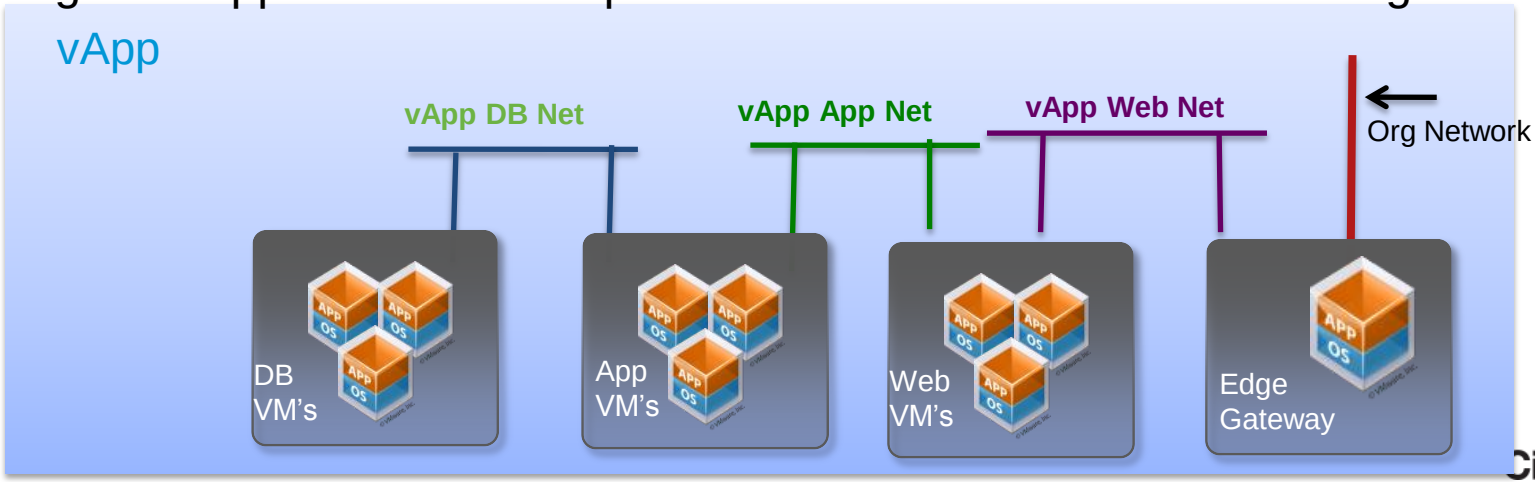
- Both MAC and IP addresses could overlap between two tenants, or even within the same tenant in different vApps.
  - Each overlapping address space needs
  - a separate segment

- VLANs uses 12 bit IDs = 4K
- VXLANs use 24 bit IDs = 16M
- NVGRE uses 24 bit IDs = 16M
- STT uses 64 bit Context IDs 12 bit VLAN ID
- DFA uses 24 bit Segment-ID



What is a vApp?

- A Cloud Provider using vCloud Director offers catalogs of vApps to their Users
- When cloned, new vApps retain the same MAC and IP addresses
- Duplicate MACs within different vApps requires L2 isolation
- Duplicate IP addresses requires L2/L3 isolation (NAT of externally facing IP addresses)
- Usage of vApps causes an explosion in the need for isolated L2 segments

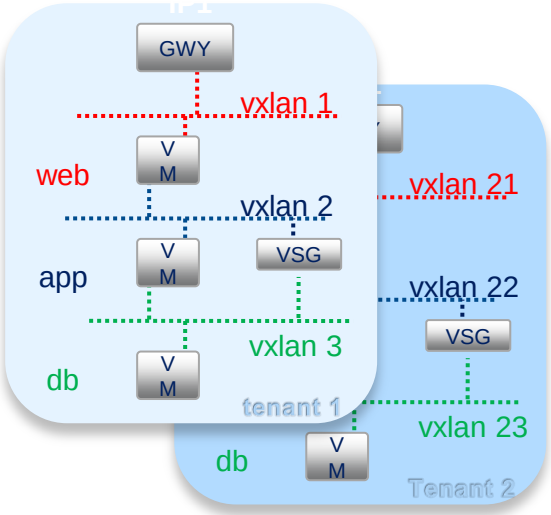


L2 Host Overlays and Virtualisation – VXLAN

Creating virtual segments

Multi-tier Virtual App = VMs + Segments + GWY

Application: Cloud Services

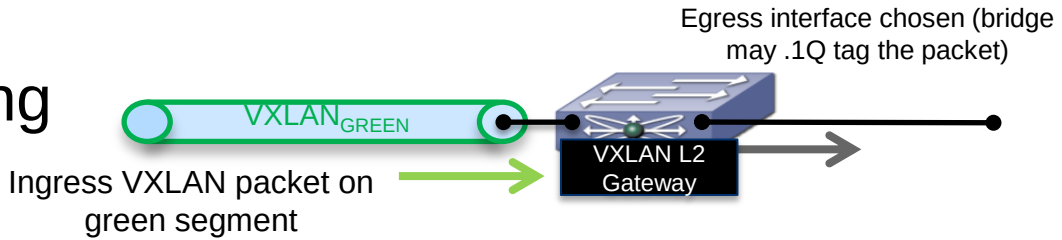


VXLAN elastic creation of **virtual Segments**

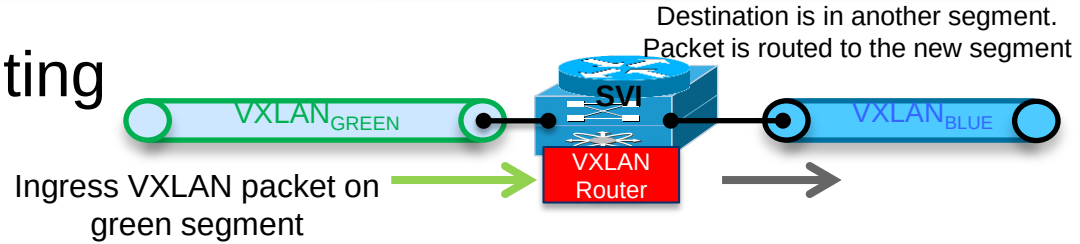
- Small L2 Segments
  - Usually don't stretch outside of a POD
- Very large number of segments
  - Do not consume resources in the network core
- Isolated, not reachable from the IP network
  - Front-end segment must be handled by the fabric
- Host overlays are initiated at the hypervisor virtual switch ➔ Virtual hosts only
- GWY to connect to the non-virtualised world
- VXLAN shipping since 2011 on Cisco Nexus 1000v, other variants: NVGRE, STT

VXLAN L2/L3 Gateways

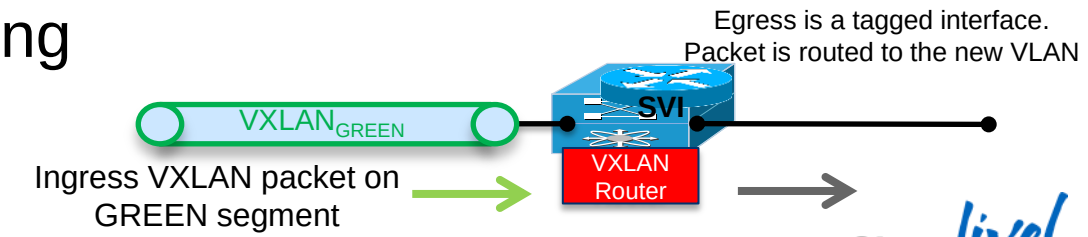
- VXLAN to VLAN Bridging (L2 Gateway)



- VXLAN-to-VXLAN Routing (L3 Gateway)

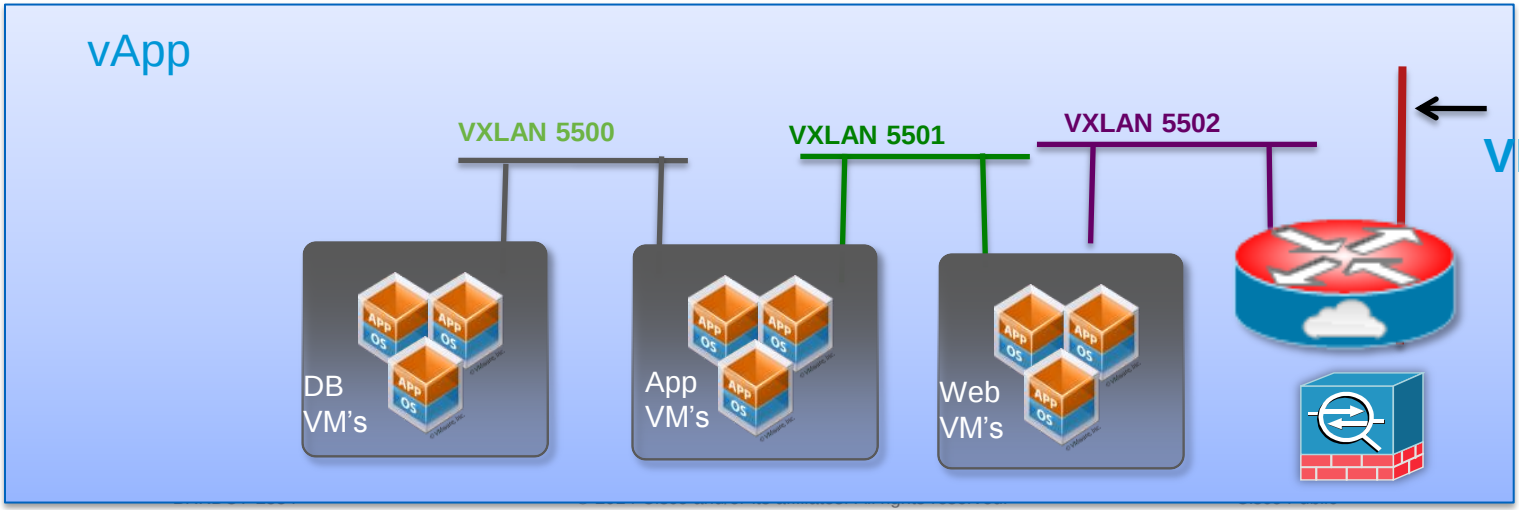


- VXLAN-to-VLAN Routing (L3 Gateway)



Possible vApp Instantiation

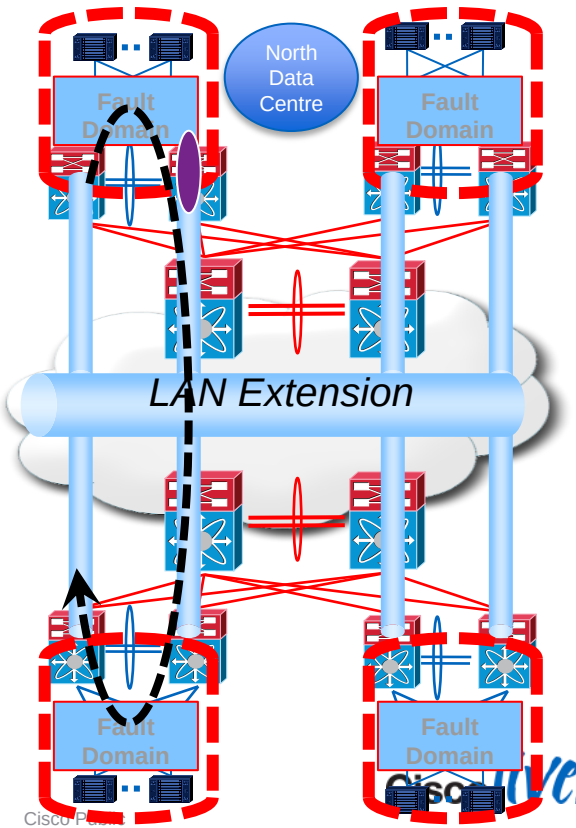
- Edge Gateway options:
  - ASA 1000V
  - CSR 1000V
  - Nexus 1x10 VXLAN Gateway
- Edge Gateway performs NAT or VPN to remote location
- VXLANs are perfect candidates for vApp Networks … conserving VLANs



L2 Network Overlays for Data Centre Interconnect
OTV/VPLS

OTV/VPLS resilient geo-extension of segments

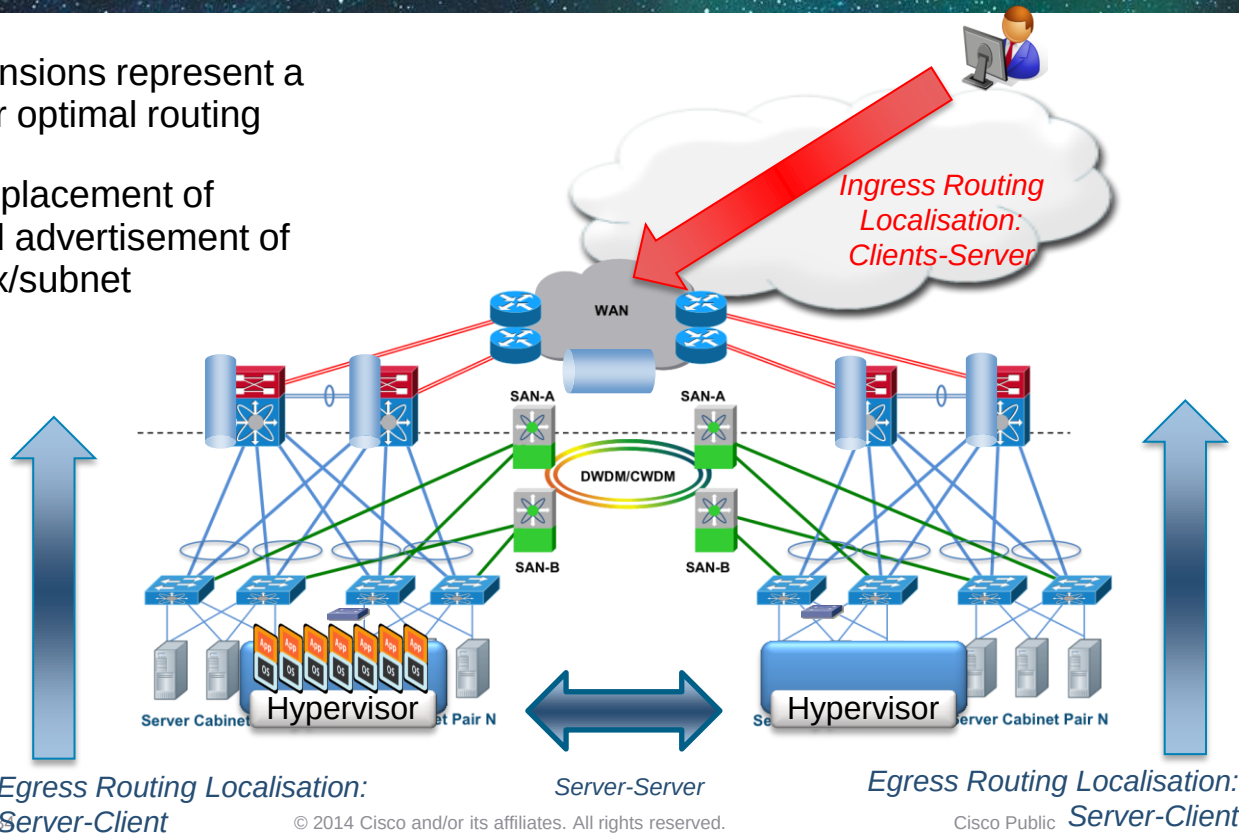
- Preserve failure isolation between locations
- Network resiliency and multi-pathing
- Built in loop handling
- Optimal traffic handling
- Streamlined operations
- Egress routing optimisation
- HW Accelerated high performance connectivity



Path Optimisation

Optimal Routing Challenges

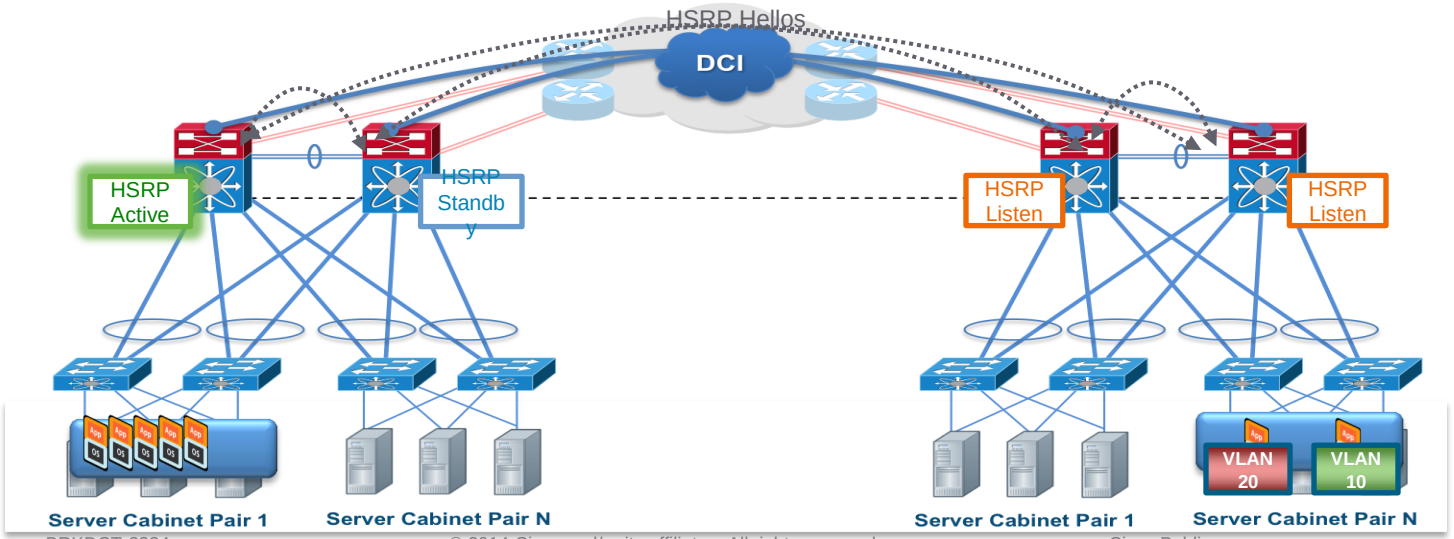
- Layer 2 extensions represent a challenge for optimal routing
- Challenging placement of gateway and advertisement of routing prefix/subnet



Path Optimisation

Egress Routing with LAN Extension

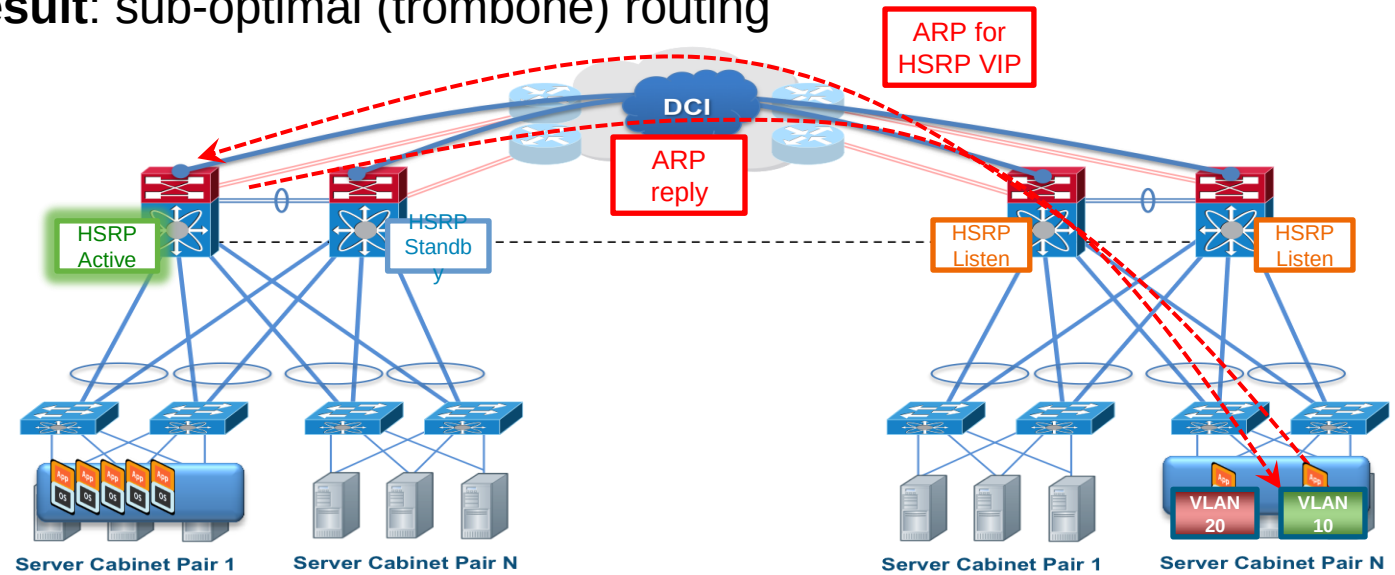
- Extended VLANs typically have associated HSRP groups
- By default, only one HSRP router elected active, with all servers pointing to HSRP VIP as default gateway
- **Result**: sub-optimal (trombone) routing



Path Optimisation

Egress Routing with LAN Extension

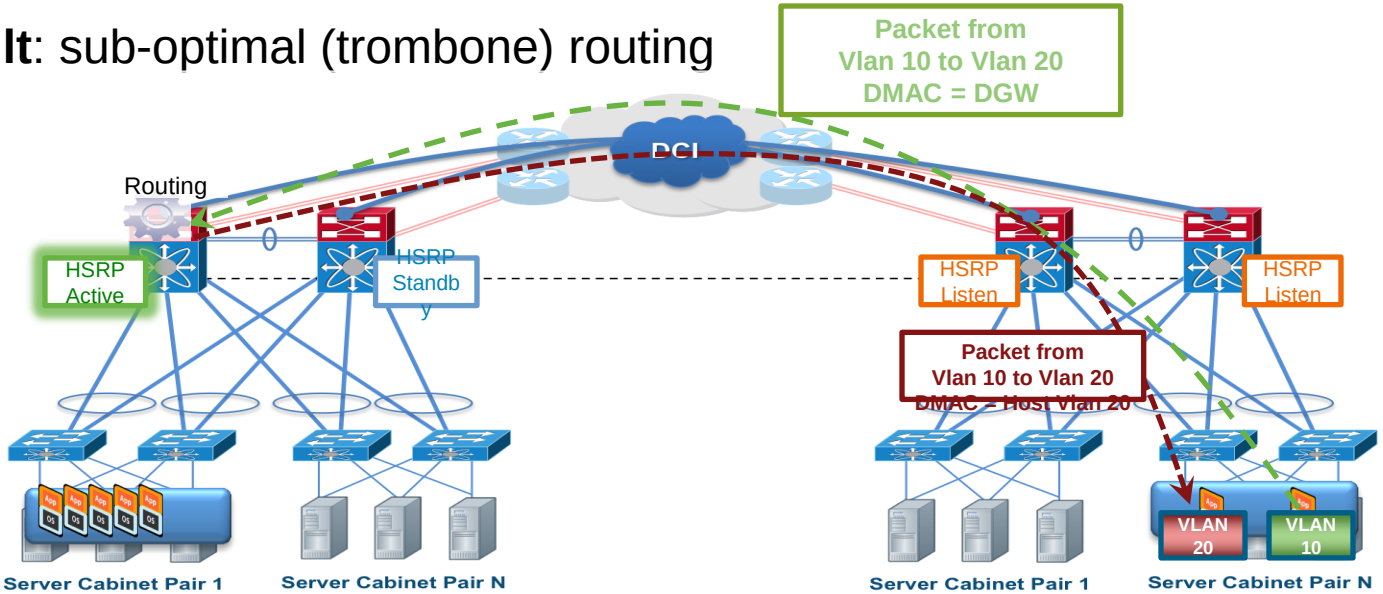
- Extended VLANs typically have associated HSRP groups
- By default, only one HSRP router elected active, with all servers pointing to HSRP VIP as default gateway
- **Result**: sub-optimal (trombone) routing



Path Optimisation

Egress Routing with LAN Extension

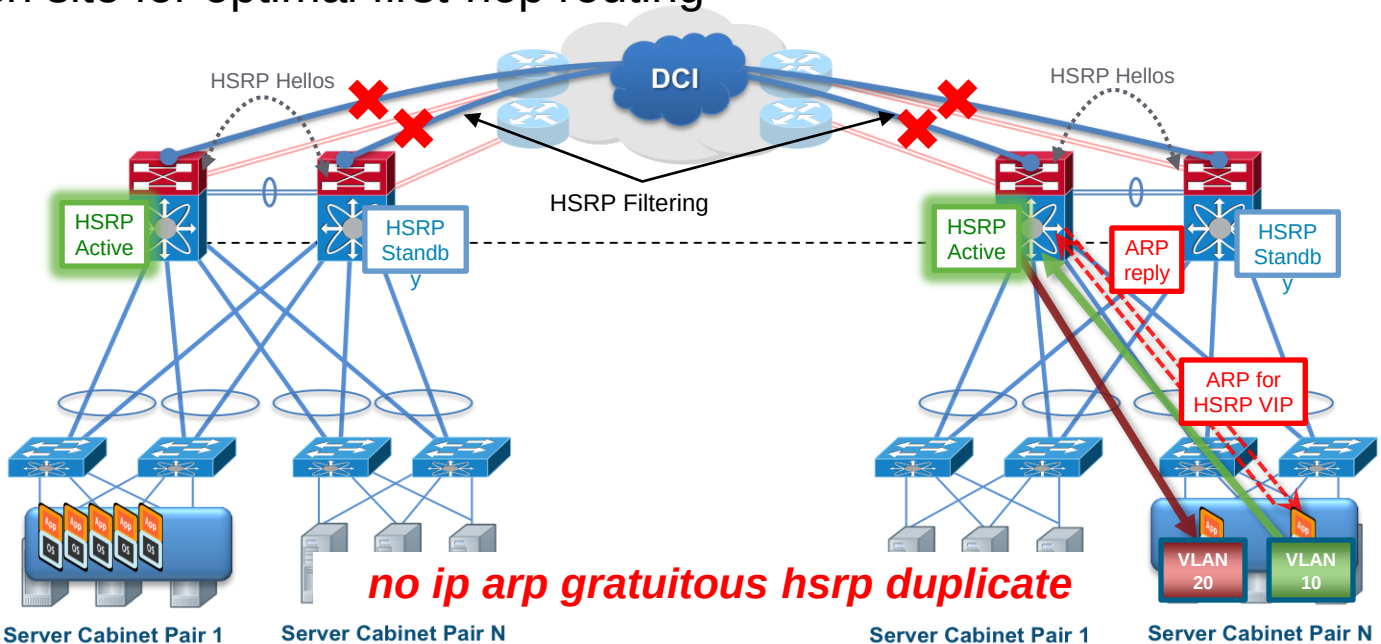
- Extended VLANs typically have associated HSRP groups
- By default, only one HSRP router elected active, with all servers pointing to HSRP VIP as default gateway
- **Result**: sub-optimal (trombone) routing



Egress Routing Localisation

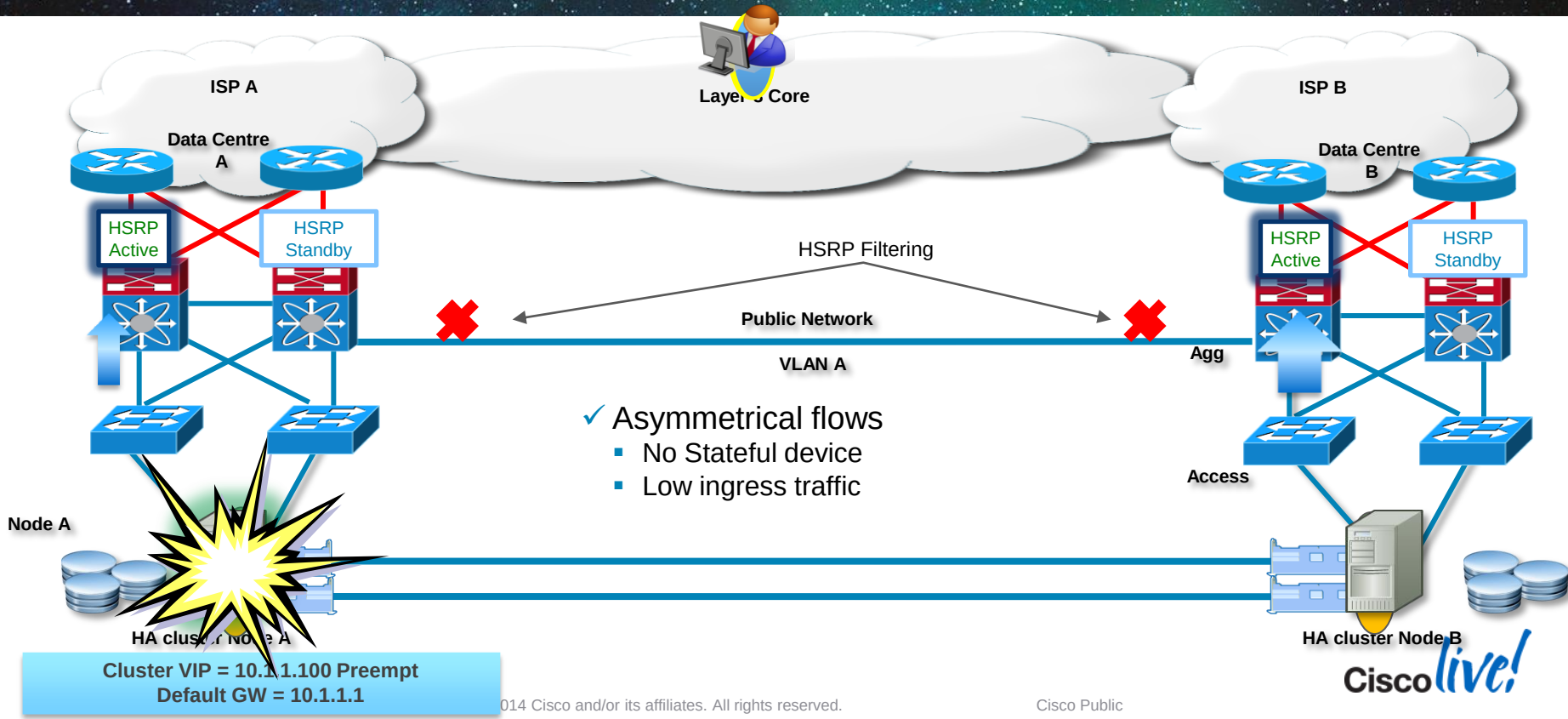
FHRP Filtering Solution

- Filter FHRP with combination of VACL and MAC route filter
- **Result: Still have one HSRP group with one VIP**, but now have active router at each site for optimal first-hop routing



Sample Cluster - Primary Service in Left DC

FHRP Localisation – Egress Path Optimisation

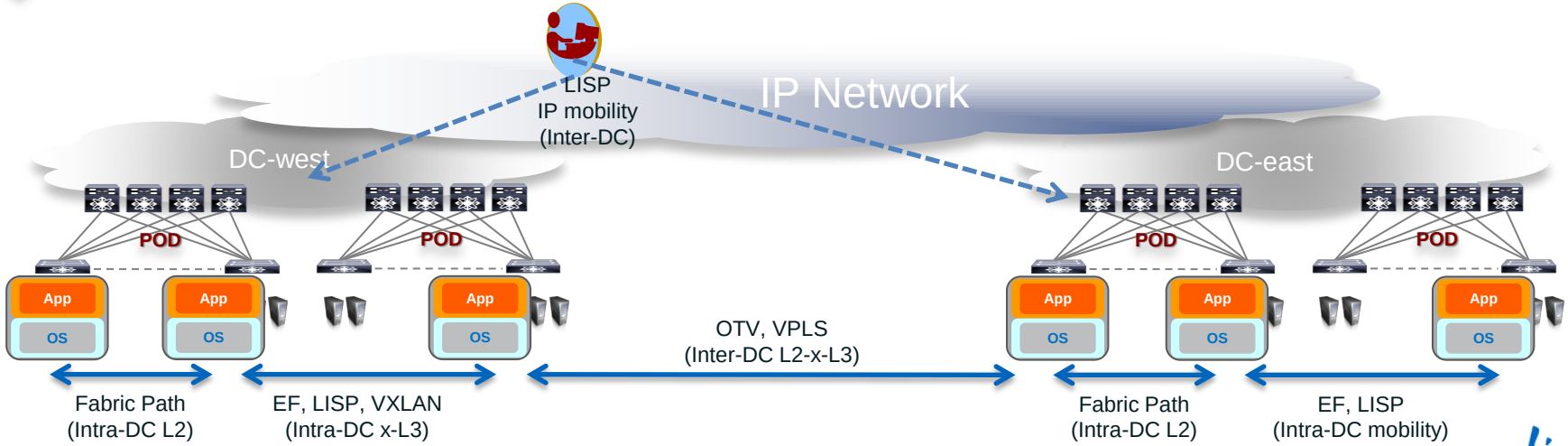


- ✓ Asymmetrical flows
  - No Stateful device
  - Low ingress traffic

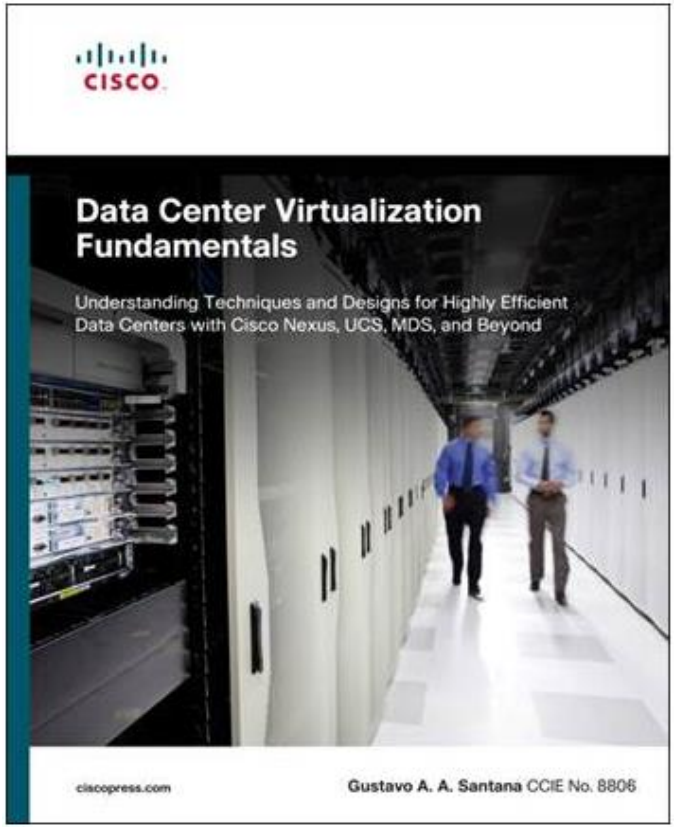
Technologies Intra-DC and Inter-DC

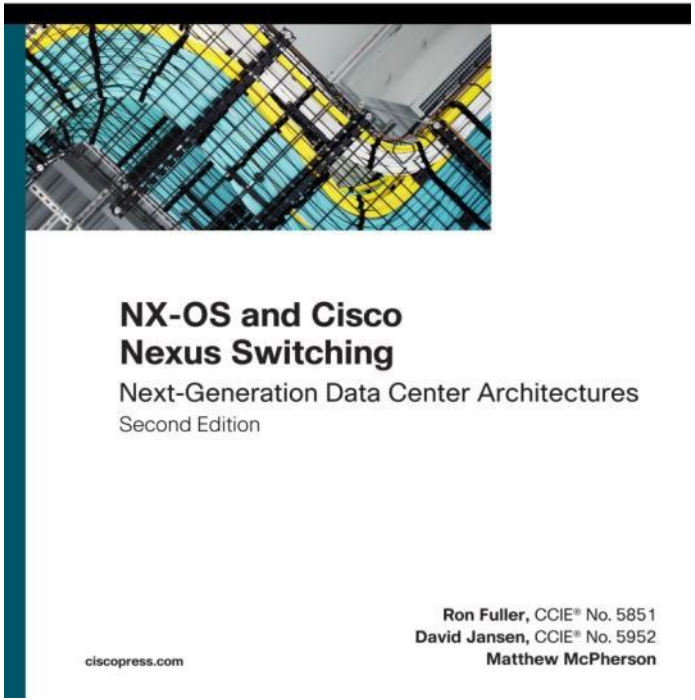
Scale

Requirement	Intra-DC	Inter-DC
Layer 2 connectivity	FabricPath, VXLAN	OTV, VPLS
IP Mobility	LISP, FP, Enhanced Forwarding	LISP, OTV
Secure Segmentation	VXLAN / Segment-ID	LISP, MPLS-IP-VPNs



Recommended Reading








Q & A

Complete Your Online Session Evaluation

Give us your feedback and receive a Cisco Live 2014 Polo Shirt!

Complete your Overall Event Survey and 5 Session Evaluations.

- Directly from your mobile device on the Cisco Live Mobile App
- By visiting the Cisco Live Mobile Site www.ciscoliveaustralia.com/mobile
- Visit any Cisco Live Internet Station located throughout the venue

Polo Shirts can be collected in the World of Solutions on Friday 21 March 12:00pm - 2:00pm

Learn online with Cisco Live!

Visit us online after the conference for full access to session videos and presentations.
www.CiscoLiveAPAC.com


CISCO
TM