
TOMORROW
starts here.
CISCO
Cisco live!

Real World Data Centre Deployments and Best Practices

BRKDCT-2334

Conrad Bullock CCIE #10767

Consulting Systems Engineer

#clmel

Cisco live!

Abstract

- This breakout session will discuss real world NX-OS deployment scenarios to ensure your Nexus based network will meet your demands for performance and reliability. We will provide you with up-to-date information on Cisco Data Centre network architecture and best practices around those designs. This will include areas such as spanning tree, vPC, Fabric Path, QOS, routing and service insertion, covering the data centre network from the core to the host. This session will not cover all of the possible options just the best practices to ensure the best outcome.

Cisco live!

Cisco Live Melbourne Related Sessions

BRKDCT-2048	Deploying Virtual Port Channel (vPC) in NXOS
BRKDCT-2049	Data Centre Interconnect with Overlay Transport Virtualisation
BRKDCT-2218	Small to Medium Data Centre Designs
BRKDCT-2404	VXLAN Deployment Models - A Practical Perspective
BRKDCT-2615	How to Achieve True Active-Active Data Centre Infrastructures
BRKDCT-3640	Nexus 9000 Architecture
BRKDCT-3641	Data Centre Fabric Design: Leveraging Network Programmability and Orchestration
BRKARC-3601	Nexus 7000/7700 Architecture and Design Flexibility for Evolving Data Centres

Cisco Live Melbourne Related Sessions

BRKACI-2000	Application Centric Infrastructure Fundamentals
BRKACI-2001	Integration and Interoperation of Existing Nexus Networks into an ACI Architecture
BRKACI-2006	Integration of Hypervisors and L4-7 Services into an ACI Fabric
BRKACI-2601	Real World ACI Deployment and Migration
BRKVIR-2044	Multi-Hypervisor Networking - Compare and Contrast
BRKVIR-2602	Comprehensive Data Centre & Cloud Management with UCS Director
BRKVIR-2603	Automating Cloud Network Services in Hybrid Physical and Virtual Environments
BRKVIR-2931	End-to-End Application-Centric Data Centre
BRKVIR-3601	Building the Hybrid Cloud with Intercloud Fabric - Design and Implementation

Agenda

- Data Centre Design Evolution
- Fundamental Data Centre Design
- Small Data Centre/Colo Design
- Scalable Data Centre Design
- Scaling the Scalable 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
- **VXLAN** – Virtual Network Local Area Network, UDP based overlay
- **OTV** - Overlay Transport Virtualisation
- **FEX** - Fabric Extender
- **UDLD** - Unidirectional Link Detection
- **LACP** - Link Aggregation Control Protocol
- **SVI** - Switch Virtual Interface
- **MCEC** - Multi-chassis EtherChannel

Data Centre Design Evolution

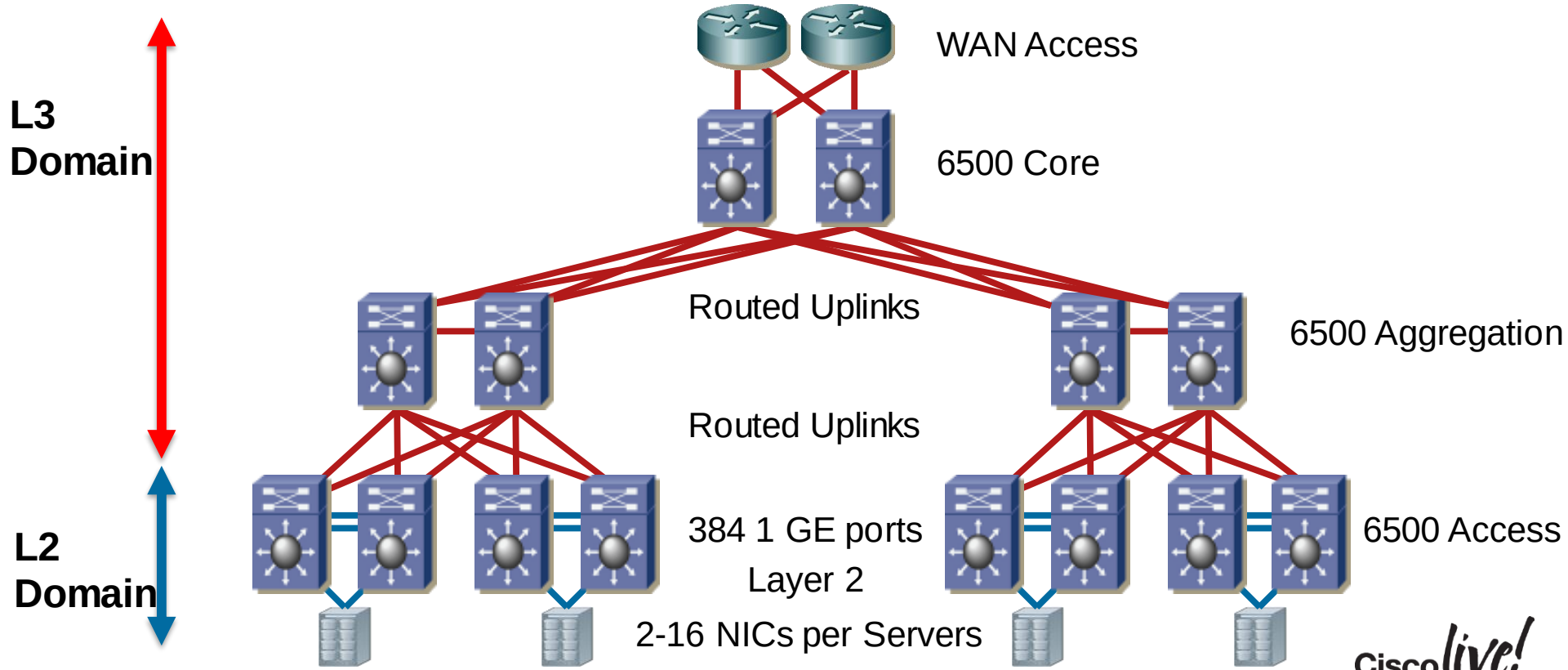
Cisco live!

What Makes Designing Networks for the Data Centre Different?

- Extremely high density of end nodes and switching
- Power, cooling, and space management constraints
- Mobility of servers a requirement, without DHCP
- The most critical shared end-nodes in the network, high availability required with very small service windows
- Multiple logical multi-tier application architectures built on top of a common physical topology
- Server load balancing, firewall, other services required

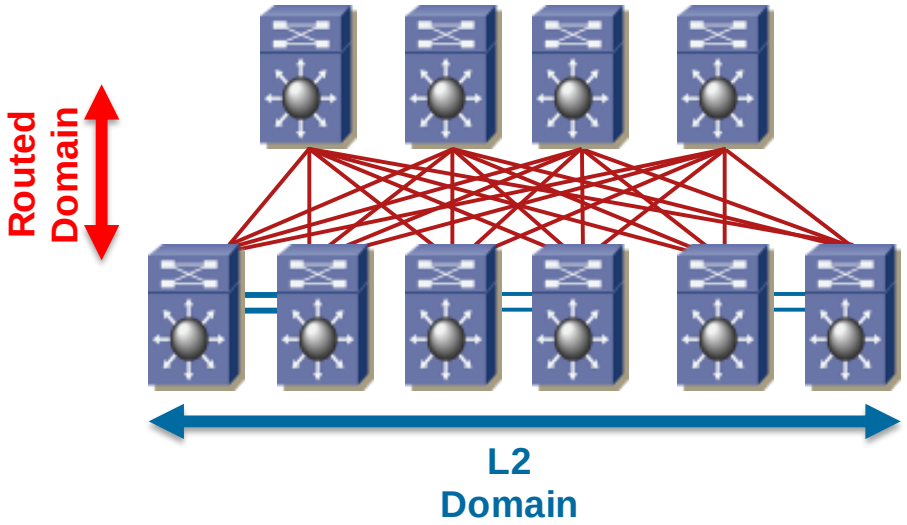
Data Centre Design Circa 2000

Original Design



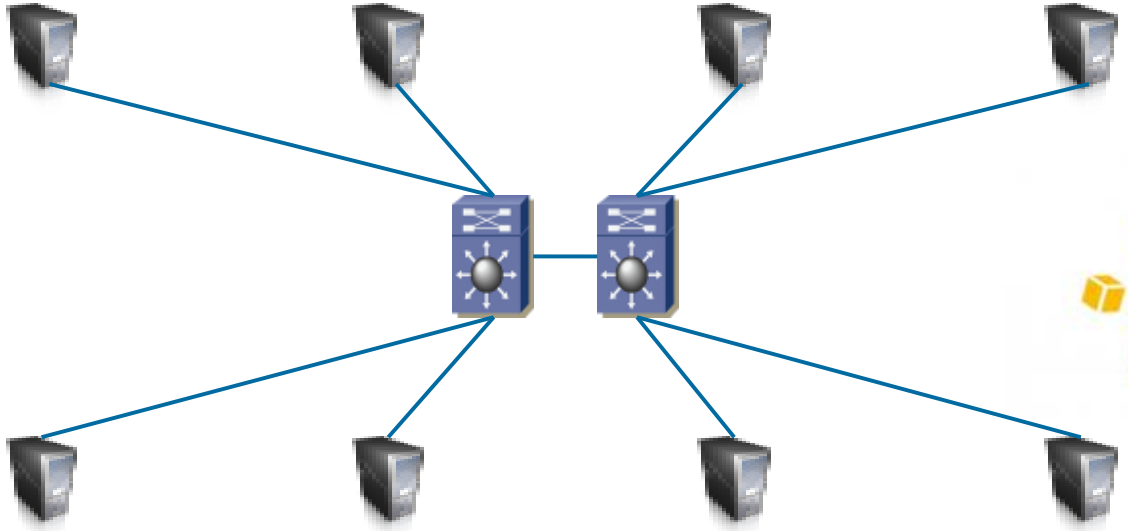
Data Centre Design Circa 2014 and Beyond Design Evolution

- Moving to Spine/Leaf construct
- No Longer Limited to two aggregation boxes
- Created Routed Paths between “access” and “core”
  - Routed based on MAC, IP, or VNI
- Layer 2 can be anywhere even with routing
- Automation/Orchestration, removing human error.



Data Centre Design Requirements

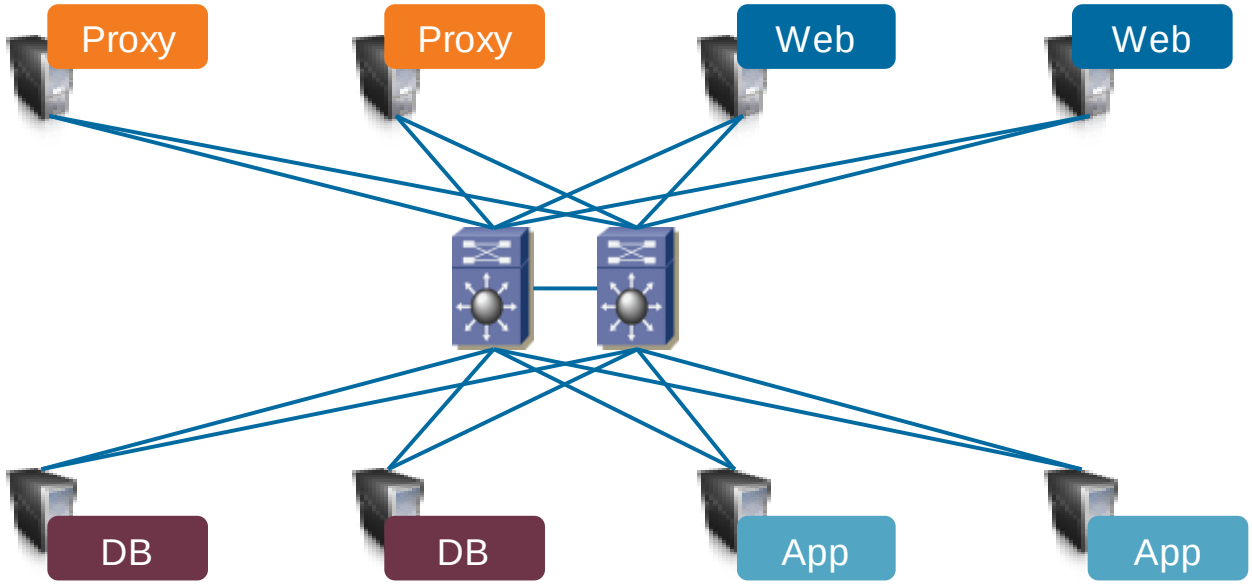
Things will fail, so how can we protect ourselves



* http://techblog.netflix.com/2012/07/chaos-monkey-released-into-wild.html

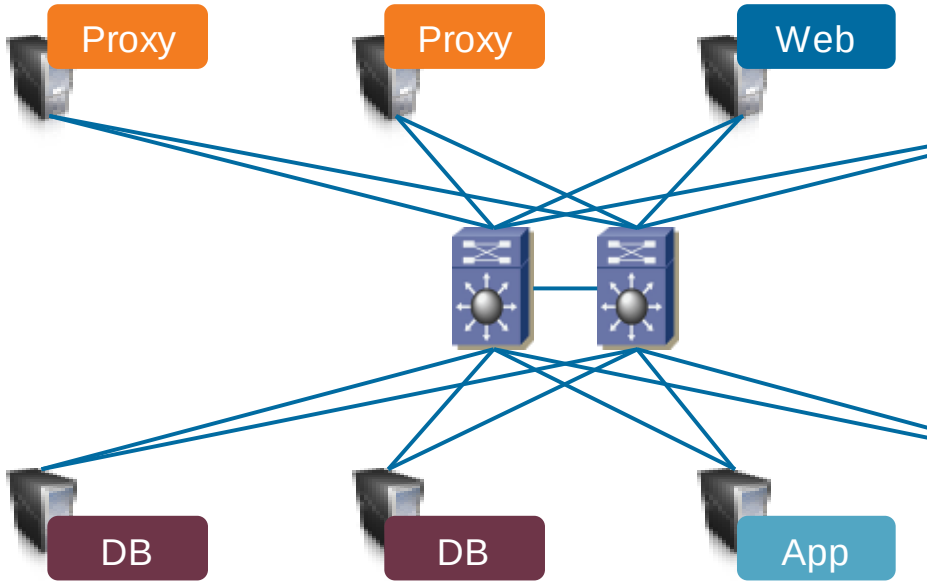
Data Centre Design Requirements

The networks needs to be flexible



Data Centre Design Requirements

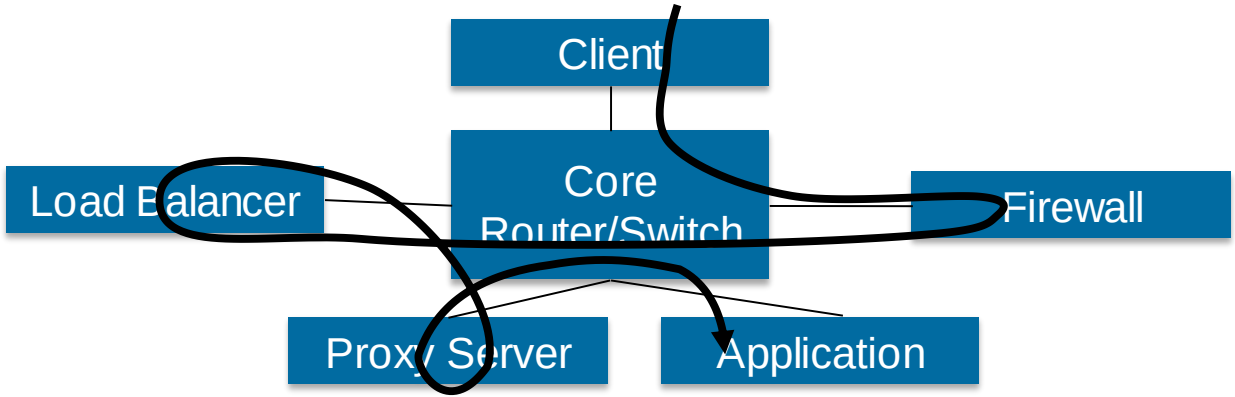
Enterprise Application Requirements Layer 2 and or Layer 3



- Layer 2 reachability
  - Layer 2 keepalives
  - Cluster Messages
  - Microsoft NLB
  - Vmotion
- Layer 3 reachability
  - “Think Cloud”
  - “New Applications”
  - IP only
  - Overlay options

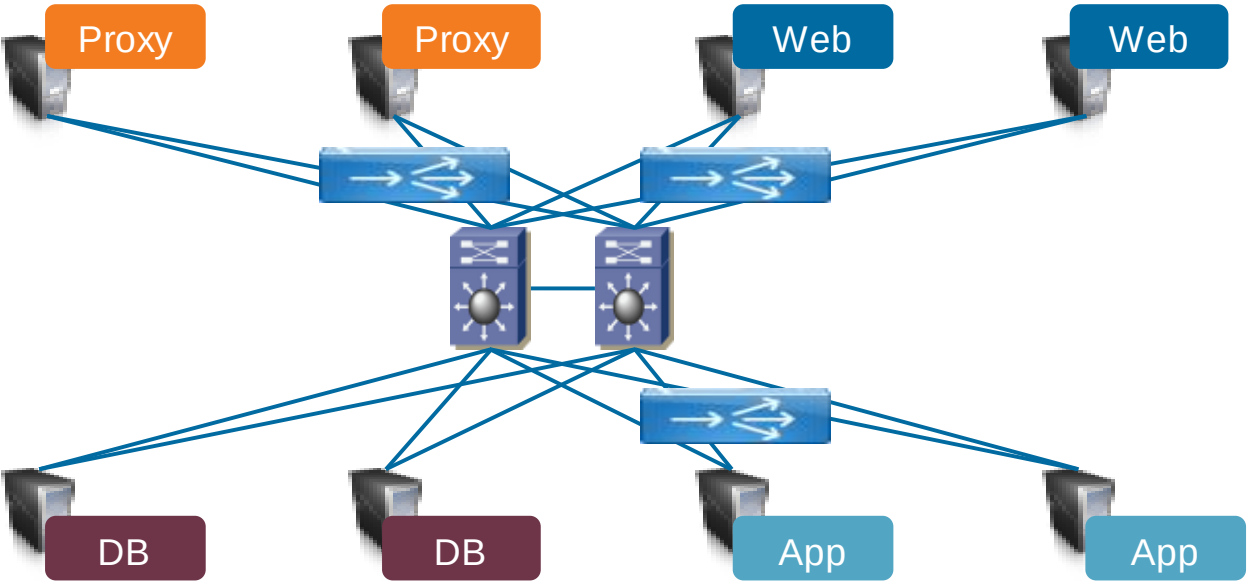
Application Requirements for Network Services

- Current generation network capabilities are driven by physical network topology.
- Many resources participate in the delivery of an application
- Full chain of services starts with the user/client and ends with the data
- Chain is multivendor
- Any resource may negatively impact the experience or availability
- Service Chain may include Physical and/or Virtual Services



Data Centre Design Requirements

ADC Services Insertion

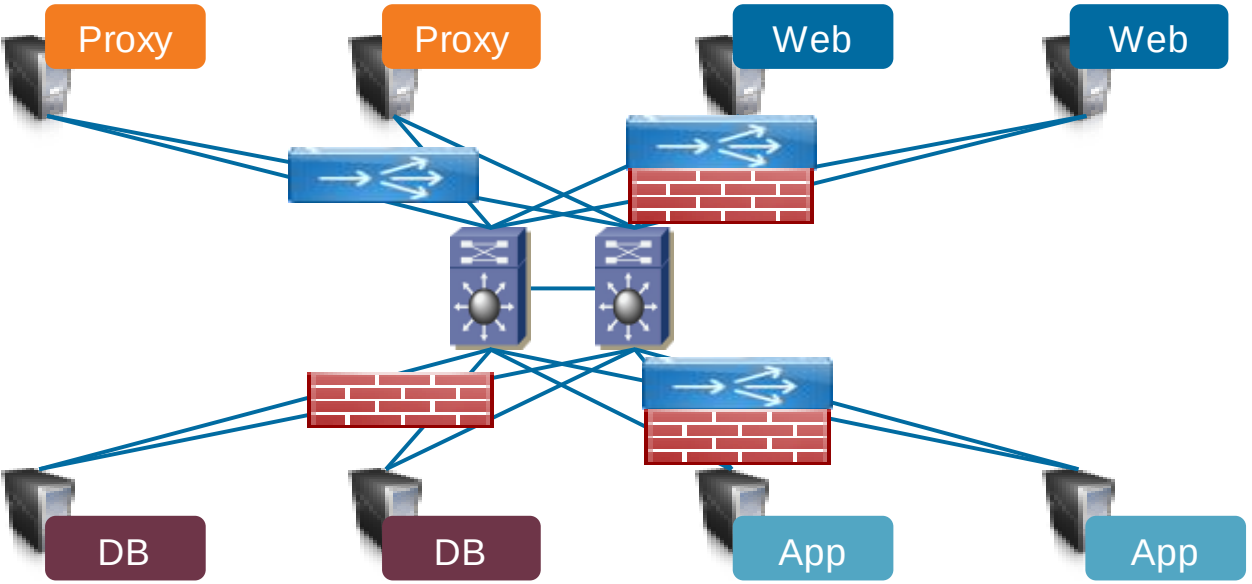


- Load Balancers
  - Performance Limits
  - Physical and/or Virtual
  - Internet Facing
  - Between App Tiers
  - Routed
  - Bridged
  - One Armed
  - Source NAT
  - PBR

NFV -> Network Function Virtualisation

Data Centre Design Requirements

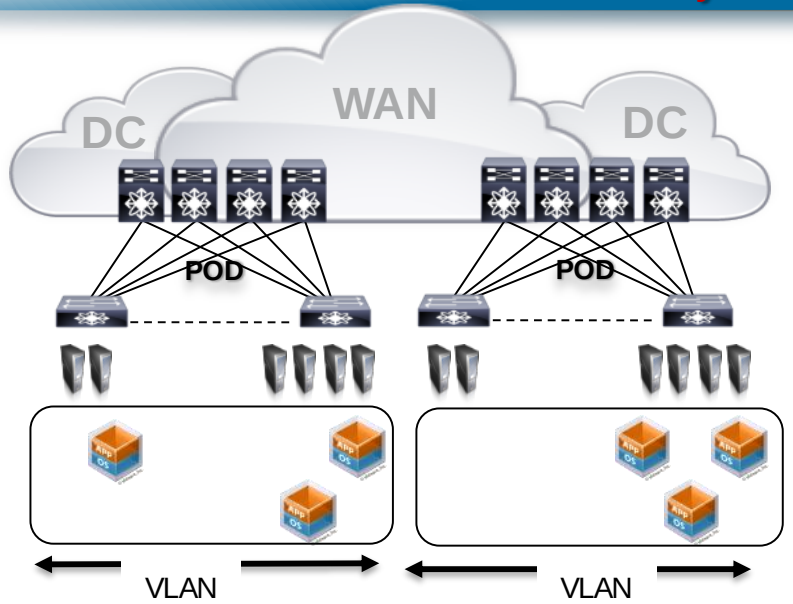
Firewall Services Insertion



- Firewalls
  - Performance Limits
  - Physical and/or Virtual
  - Transparent
  - Routed
  - VRFs
  - Between Tiers
  - Internet Facing
  - IPS/IDS
  - Service Chaining

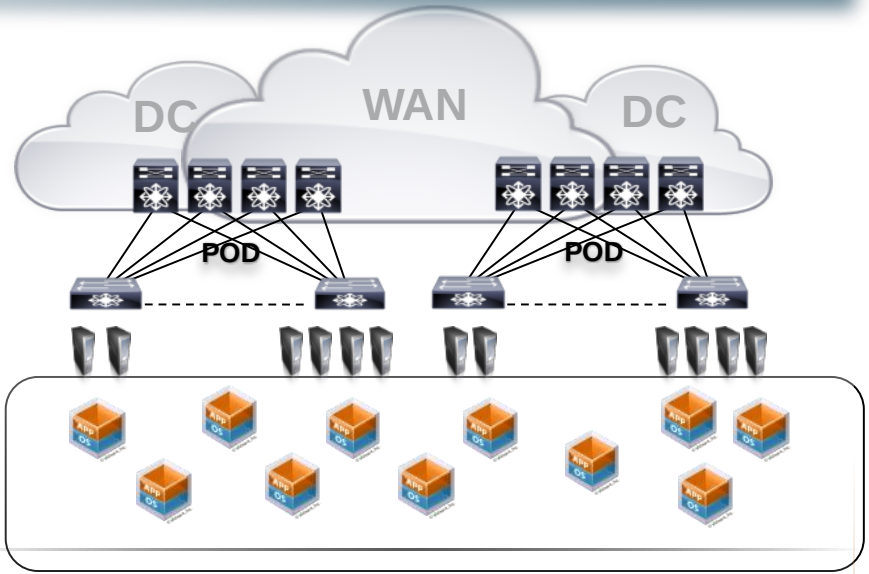
VLAN Ubiquity Inter Data Centre

Data Centre-Wide VM Mobility



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

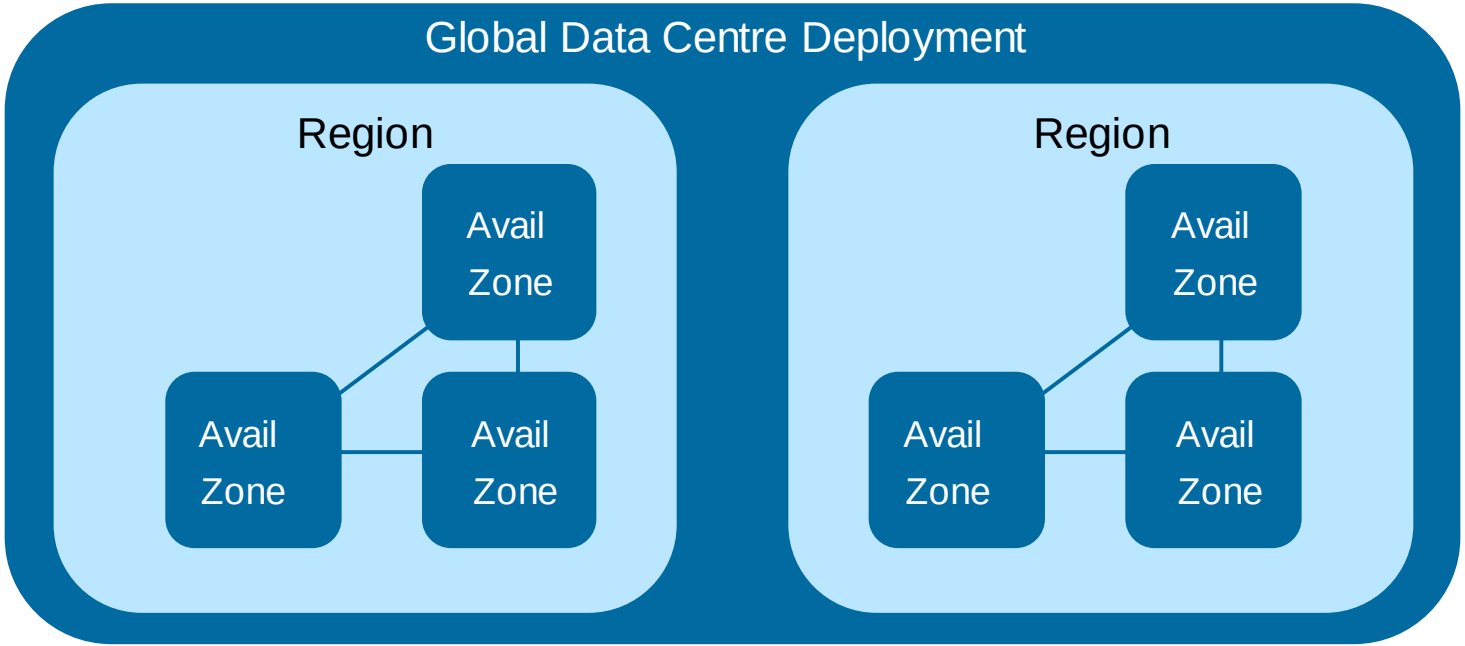
Seamless Layer 2 between DC



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

Availability Zones

Using Amazon Web Services terms



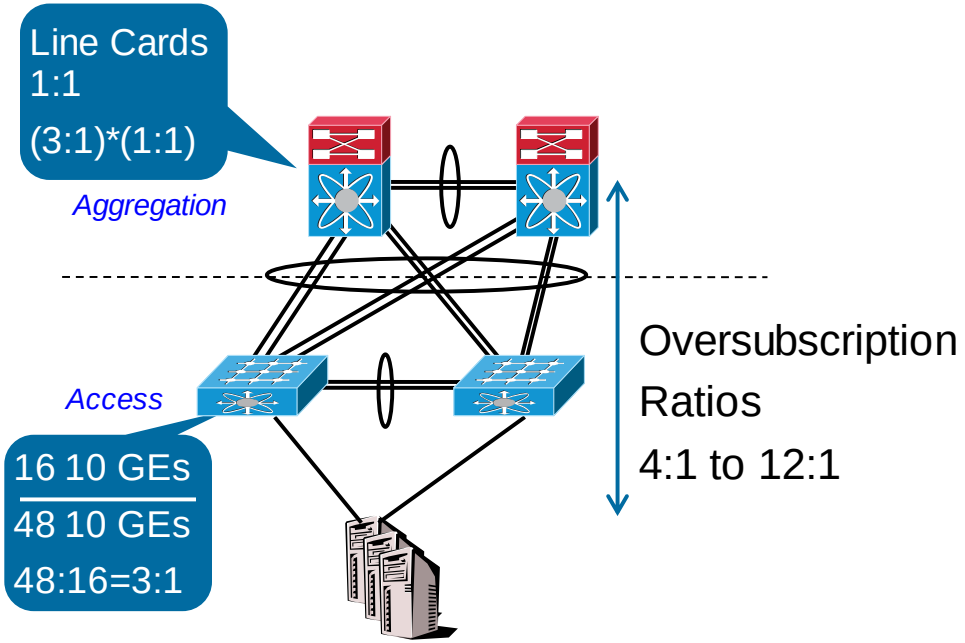
Example of Constrained Resource

Feature	Parameter	Verified Limit (Cisco NX-OS 6.2)		
		Sup 1	Sup 2	Sup 2E
ARP/ND	Number of entries in ARP table	128,000	128,000	128,000
	Number of ARP packets per second	1500	1500	5000
	Number of ARP glean packets per second	1500	1500	5000
	Number of IPv6 ND packets per second	1500	1500	5000
	Number of IPv6 glean packets per second	1500	1500	5000

Oversubscription Ratio

Access to Core/Aggregation

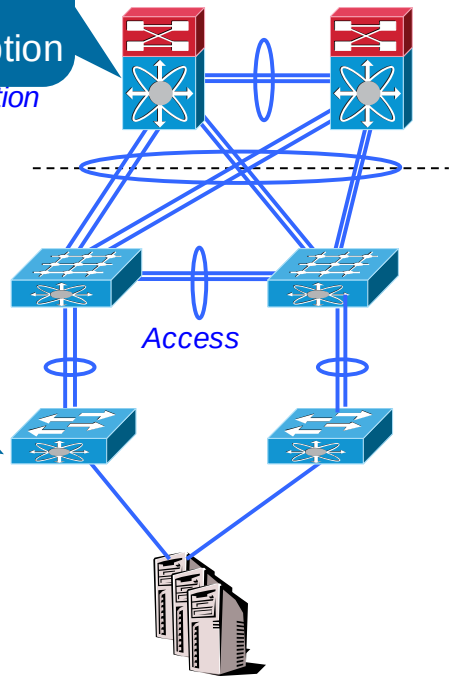
- Large layer 2 domain with collapsed Access and Core
- Worse Case Calculation
  - Assume all the traffic is north-south bound
  - Assume 100% utilisation from the Access Switches
  - All the ports operated in dedicated mode



Oversubscription Ratio

Lower is better, The old goal was 12:1 to 4:1 and …

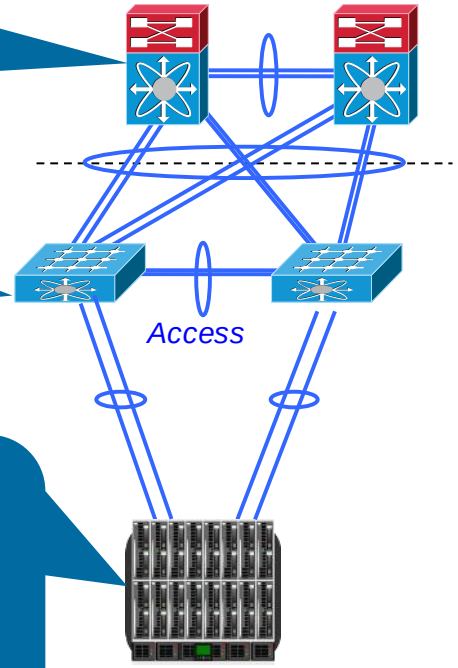
Line Cards 1:1
(4:1)*(6:1)*(1:1)
24:1 Oversubscription



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

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

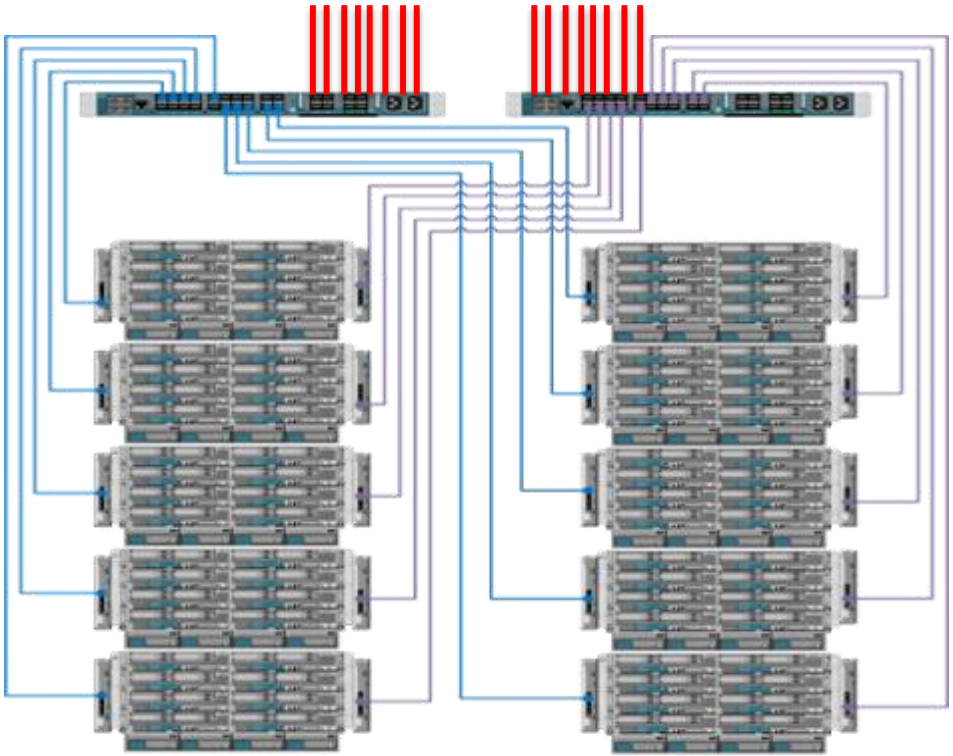
Line Cards 1:1
(4:1)*(12:1)*(1:1)
48:1 Oversubscription



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

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

Oversubscription with Cisco UCS

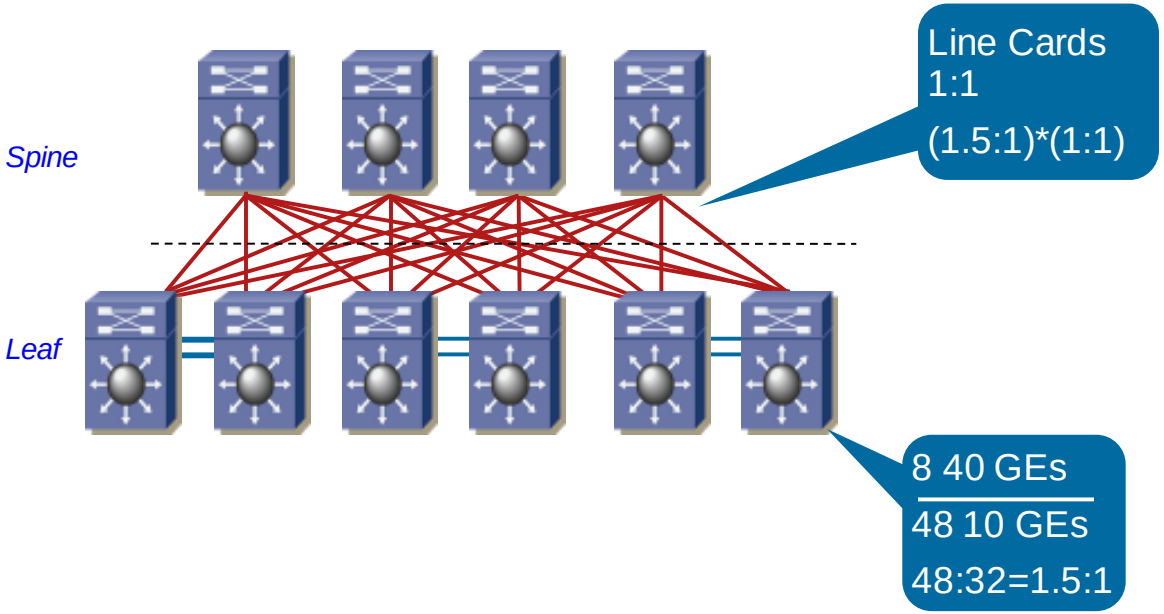


Consistent Latency

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

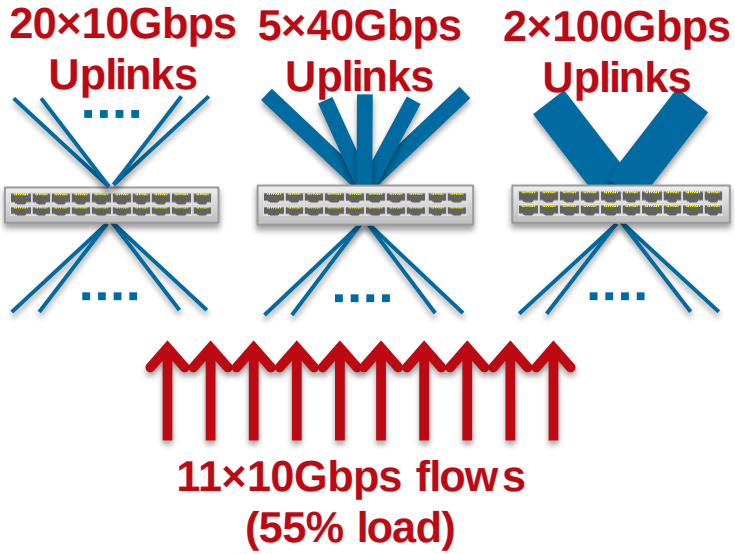


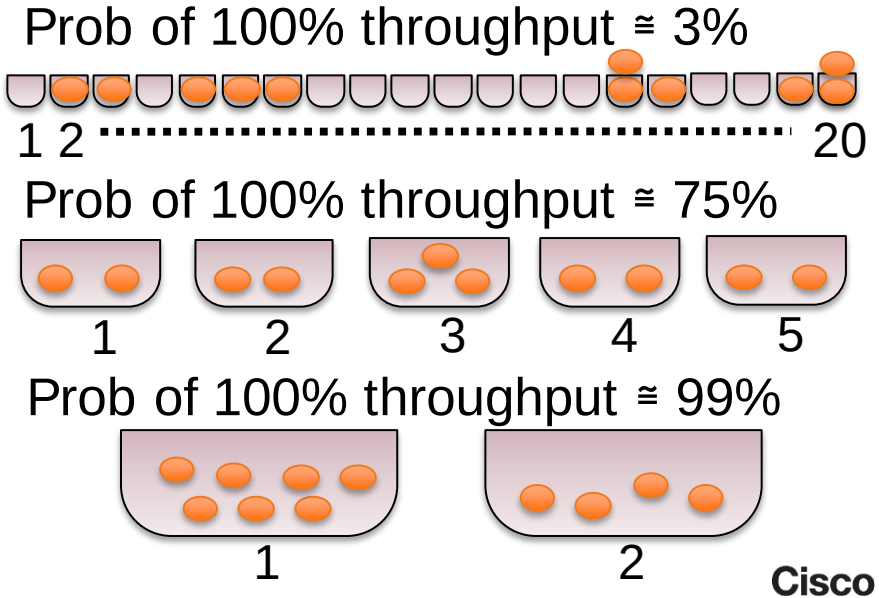
- Changing Traffic Flow Requirements
- Services are deployed at the leaf nodes
- Oversubscription Ratios defined by number of spines and uplink ports
- True horizontal scale

Statistical Probabilities…

Intiution: Higher speed links improve ECMP efficiency

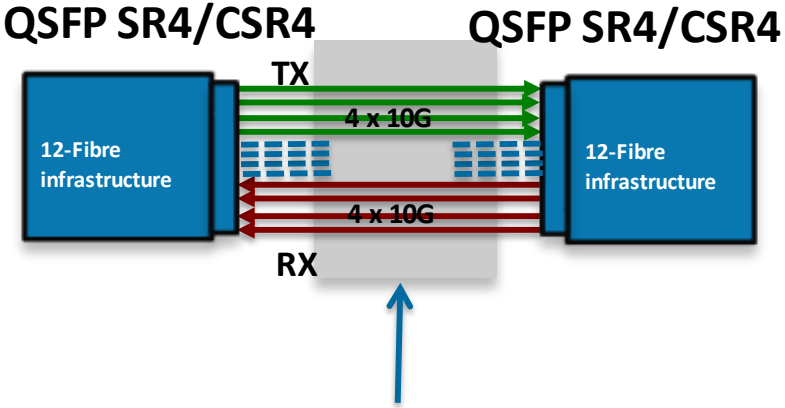
- Assume 11 10G source flows, the probability of all 11 flows being able to run at full flow rate (10G) will be almost impossible with 10G (~3%), much better with 40G (~75%) & 100G (~99%)



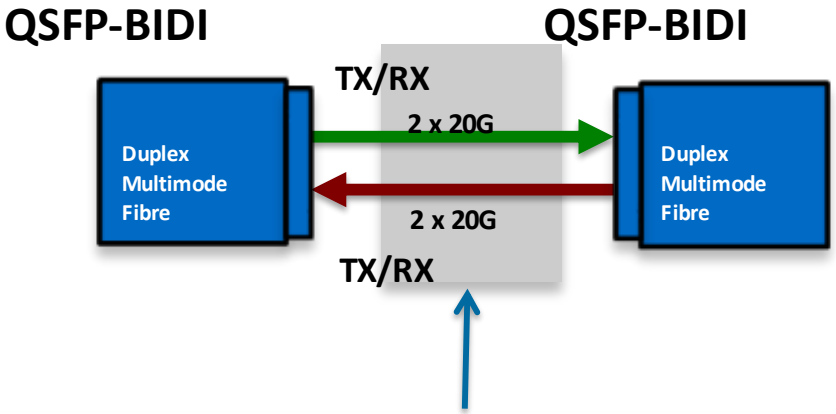


QSFP-BiDi vs. QSFP-40G-SR4

12-Fibre vs. Duplex Multimode Fibre



12-Fibre ribbon cable with MPO connectors at both ends

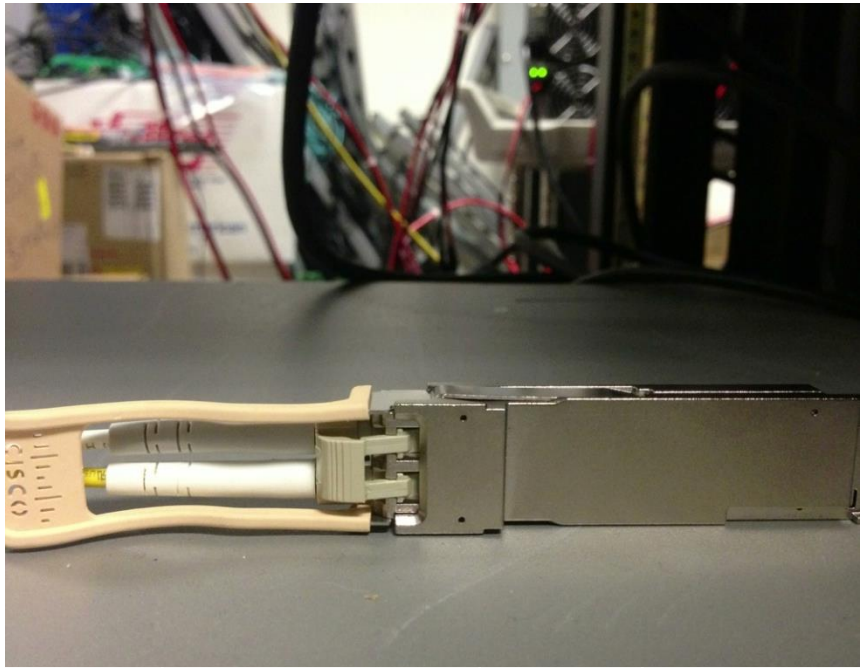


Duplex multimode fibre with Duplex LC connectors at both ends

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 5600	FCS
Nexus 3100	6.0.2A

Fundamental Data Centre Design

Cisco live!

UDLD Behaviour

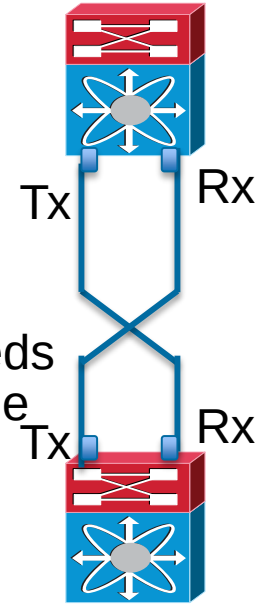
- UDLD is running as a conditional feature, it needs to be enabled:

```
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 and 10GE

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

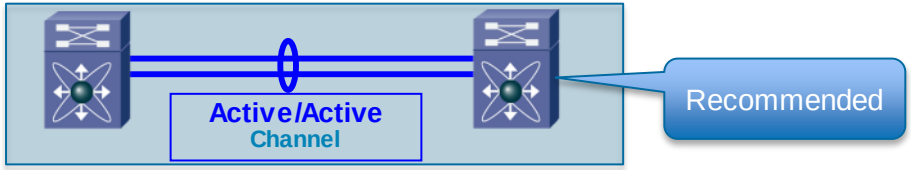
```
NX-OS(config)# spanning-tree port type edge
  --- or ---
NX-OS(config)# spanning-tree port type edge trunk

NX-OS(config)# spanning-tree port type edge bpduguard default
```

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

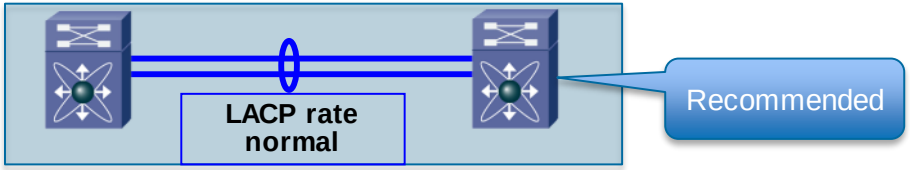
Cisco live!

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).

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

```
show run all | grep jumbomtu
  system jumbomtu 9216
interface Vlan10
  ip address 10.87.121.28/27
  mtu 9216
```

- Layer 3 MTU changed under Interface
- Nexus 7000 FCoE policy sets MTU lower per policy-map than jumbomtu
- Interface MTU overrides network-qos

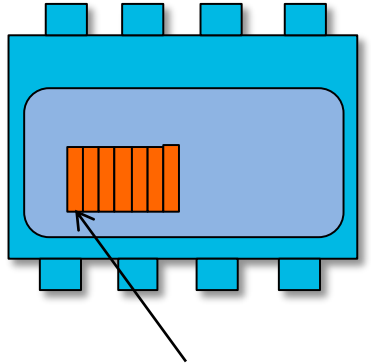
```
policy-map type network-qos default-nq-4e-policy
  class type network-qos c-nq-4e-drop
    mtu 1500
  class type network-qos c-nq-4e-ndrop-fcoe
    mtu 2112
```

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 6000/5600
Interface ethernet 1/x
  Mtu 9216
```

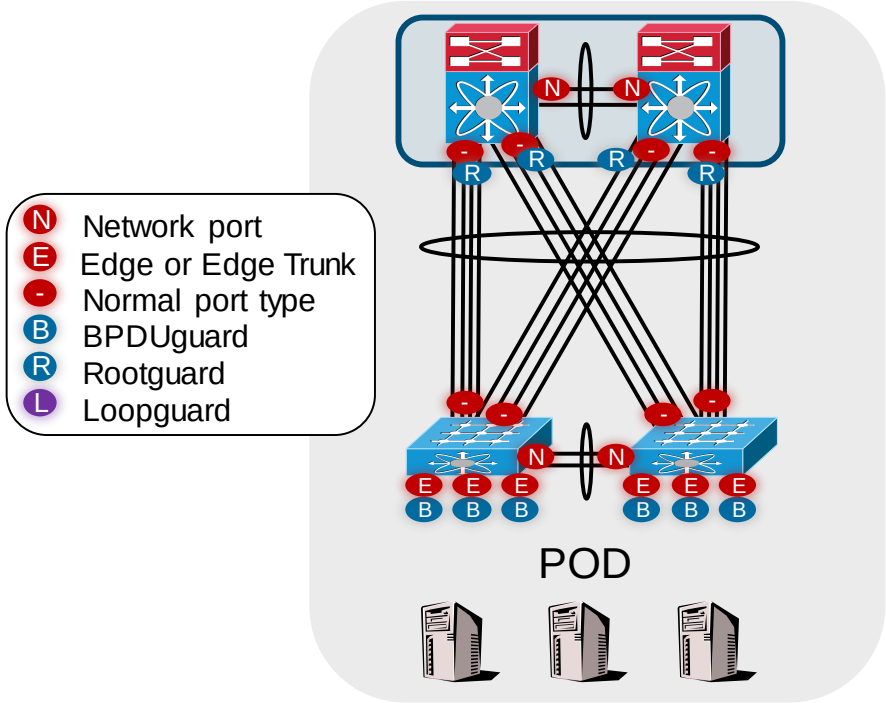


Each qos-group on the Nexus 5000/3000 supports a unique MTU

Cisco live!

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



vPC Terms

For Your Reference

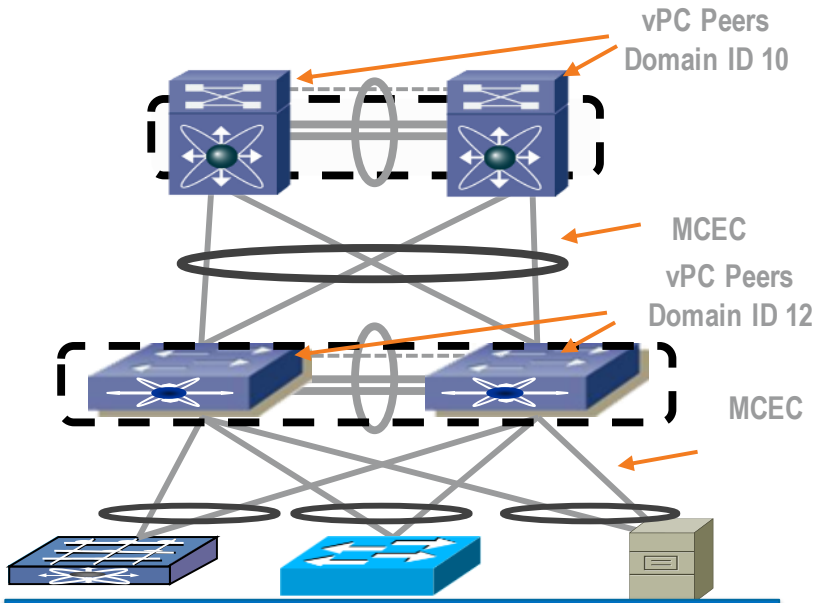
Term	Meaning
vPC	The combined port-channel between the vPC peers and the downstream device. A vPC is a L2 port type: switchport mode trunk or switchport mode access
vPC peer device	A vPC switch (one of a Cisco Nexus 7000 Series pair).
vPC domain	Domain containing the 2 peer devices. Only 2 peer devices max can be part of same vPC domain. Domain ID needs to be unique per L2 domain
vPC member port	One of a set of ports (that is, port-channels) that form a vPC (or port-channel member of a vPC).
vPC peer-link	Link used to synchronise the state between vPC peer devices. It must be a 10-Gigabit Ethernet link. vPC peer-link is a L2 trunk carrying vPC VLAN.
vPC peer-keepalive link	The keepalive link between vPC peer devices; this link is used to monitor the liveness of the peer device.
vPC VLAN	VLAN carried over the vPC peer-link and used to communicate via vPC with a third device. As soon as a VLAN is defined on vPC peer-link, it becomes a vPC VLAN
non-vPC VLAN	A VLAN that is not part of any vPC and not present on vPC peer-link.
Orphan port	A port that belong to a single attached device. vPC VLAN is typically used on this port.
Cisco Fabric Services (CFS) protocol	Underlying protocol running on top of vPC peer-link providing reliable synchronisation and consistency check mechanisms between the 2 peer devices.

Cisco live!

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
- 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
```

vPC Best Practice Summary

- Use LACP Protocol when connecting devices
- Use multiple interfaces for Peer-Link
- Enable Auto Recovery
- IP Arp sync
- Use Peer-Switch with appropriate spanning tree priorities set
- IPv6 ND synchronisation
- Peer Gateway
  - with exclude VLAN where required
- Fabric Path Multicast Loadbalance.

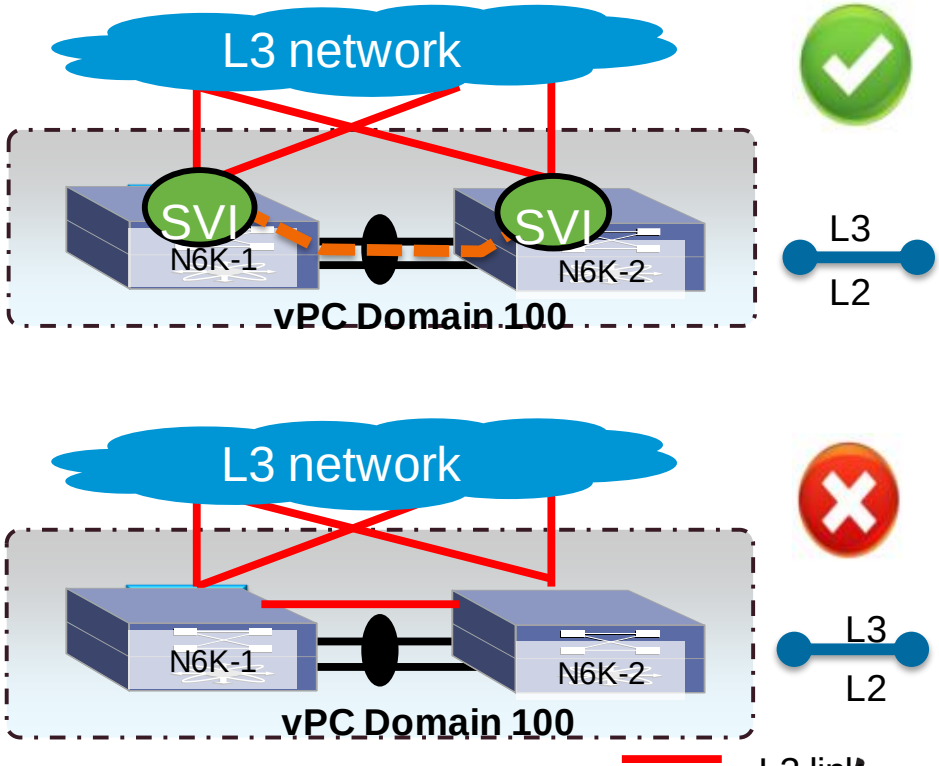
http://www.cisco.com/c/dam/en/us/td/docs/switches/datacenter/sw/design/vpc_design/vpc_best_practices_design_guide.pdf

N6K/N5600 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

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



Example of Constrained Resource

Feature	Parameter	Verified Limit (Cisco NX-OS 6.2)		
		Sup 1	Sup 2	Sup 2E
ARP/ND	Number of entries in ARP table	128,000	128,000	128,000
	Number of ARP packets per second	1500	1500	5000
	Number of ARP glean packets for second	1500	1500	5000
	Number of IPv6 ND packets per second	1500	1500	5000
	Number of IPv6 glean packets per second	1500	1500	5000

128,000ARPs/1500(ARPs Per Second)=85.3 seconds

COPP Policy Monitoring

Control Plane Policy Exceeded

Customer 1 (5.2.5 code)

```
show policy-map interface control-pla class copp-system-p-class-normal | inc
violate prev 4 | inc module|violated
  module 3
        violated 0 bytes; action: drop
  module 8
        violated 1152074225 bytes; action: drop (approximately 18 Million ARPs)
  module 9
        violated 2879379238 bytes; action: drop (approximately 45 Million ARPs)
```

Customer 2 (6.2.10 code)

```
show policy-map interface control-plane class copp-system-p-class-normal | inc
violate
        violate action: drop
        violated 8241085736 bytes, (approximately 128 Million ARPs in 123 Days)
          5-min violate rate 0 bytes/sec
        violated 0 bytes,
          5-min violate rate 0 bytes/sec
        violated 0 bytes,
```

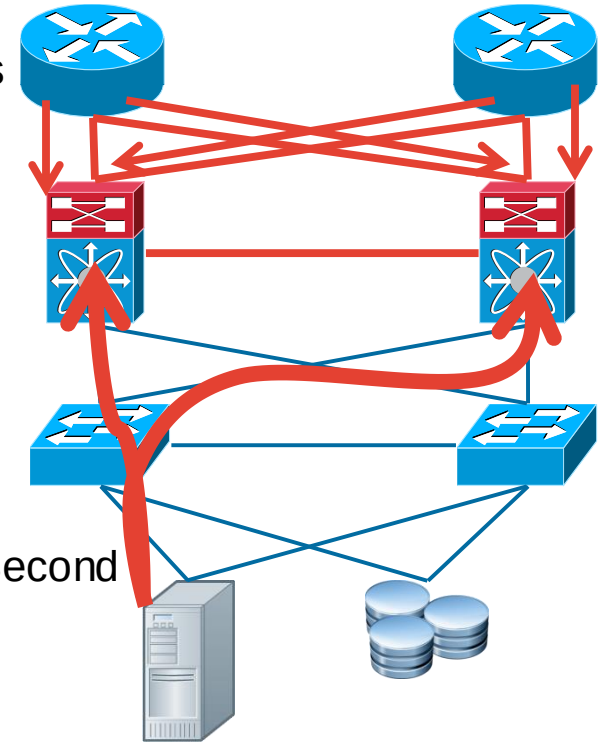
Effects of an ARP Flood

Arp Time out 4 Hours

Routing Peers

25 Minute ARP Time out

30 Minute CAM Time out



2000 ARPs/Second

Arp Time out 4 Hours

Routing Peers

25 Minute ARP Time out

30 Minute CAM Time out

Which ARP is better?

- Server ARP
- Storage ARP
- Router ARP
- Switch ARP

Cisco live!

Control Plane Policing

```
show copp diff profile strict profile moderate
'+' Line appears only in profile strict, version 6.2(6a)
'-' Line appears only in profile moderate, version 6.2(6a)
    -policy-map type control-plane copp-system-p-policy-moderate
reduced
    -  class copp-system-p-class-normal
    -    set cos 1
    -    police cir 680 kbps bc 310 ms conform transmit violate drop
reduced
    +  class copp-system-p-class-normal
    +    set cos 1
    +    police cir 680 kbps bc 250 ms conform transmit violate drop
```

- 680 kbps / (64 byte Arp Frames * 8 bits) = 1328 ARPs per second
- BC = TC * CIR or 310 msec *680,000 = 204000 this means approximately another 400 ARPs per second are allowed for burst.

show policy-map interface control-plane class copp-system-p-class-normal | inc violate

Control Plane Protection

Notification about Drops

- Configure a syslog message threshold for CoPP
  - in order to monitor drops enforced by CoPP.
- The logging threshold and level can be customised within each traffic class with use of the **logging drop threshold <packet-count> level <level>** command.

```
logging drop threshold 100 level 5
```

Example syslog output

```
%COPP-5-COPP_DROPS5: CoPP drops exceed threshold in class:
copp-system-class-critical,
check show policy-map interface control-plane for more info.
```

Control Plane Tuning

- Do not disable CoPP. Tune the default CoPP, as needed.
- Create Custom Policy to match your environment

Nexus# copp copy profile strict prefix LAB

```
monitor session 1
  source exception all
  destination interface Eth1/3
  no shut
```

```
nexus7k(config-monitor)# show monitor session 1
source exception   : fabricpath, layer3, other
filter VLANs       : filter not specified
destination ports : Eth1/3
Feature        Enabled   Value   Modules Supported
-----------------------------------------------------
L3-TX          -         -       1  6  8
ExSP-L3        -         -       1
ExSP-FP        -         -       8
ExSP-OTHER     -         -       1
RB span        No
```

Control Plane Protection

Good ARPs versus Bad ARPs

```
Nexus(config)# arp access-list LAB-copp-arp-critical
Nexus(config-arp-acl)# 10 permit ip 10.1.2.1 255.255.255.255 mac any
Nexus(config-arp-acl)# 20 permit ip 10.1.2.5 255.255.255.255 mac any
Nexus(config-arp-acl)# class-map type control-plane match-any LAB-copp-class-arp-critical
Nexus(config-cmap)# match access-group name LAB-copp-arp-critical
Nexus(config-cmap)# policy-map type control-plane LAB-copp-policy-strict
Nexus(config-pmap)# class LAB-copp-class-arp-critical insert-before LAB-copp-class-normal
Nexus(config-pmap-c)# set cos 6
Nexus(config-pmap-c)# police cir 100 kbps bc 250 ms conform transmit violate drop
Nexus(config)# control-plane
Nexus(config-cp)# service-policy input LAB-copp-policy-strict
```

High CPU?

Use EEM to Determine Source of CPU Spike

Use 1.3.6.1.4.1.9.9.305.1.1.1 from Cisco-system-ext-mib to determine [XX]

```
ENTITY-MIB::entPhysicalDescr.22 = STRING: 1/10 Gbps Ethernet Module
ENTITY-MIB::entPhysicalDescr.25 = STRING: 10/40 Gbps Ethernet Module
ENTITY-MIB::entPhysicalDescr.26 = STRING: Supervisor Module-2
ENTITY-MIB::entPhysicalDescr.27 = STRING: Supervisor Module-2
```

event manager applet highcpu

```
  event snmp oid 1.3.6.1.4.1.9.9.109.1.1.1.1.6.[XX] get-type exact entry-op ge
entry-val 70 exit-op le exit-val 30 poll-interval 1
  action 1.0 syslog msg High CPU DETECTED 'show process cpu sort' written to
bootflash:highcpu.txt
  action 2.0 cli enable
  action 3.0 cli show clock >> bootflash:highcpu.txt
  action 4.0 cli show process cpu sort >> bootflash:highcpu.txt
```

NX-API Developer Sandbox

Learning Python via the API

REQUEST:

Copy

Python

```
[
  {
    "jsonrpc": "2.0",
    "method": "cli",
    "params": {
      "cmd": "config t",
      "version": 1
    },
    "id": 1
  },
  {
    "jsonrpc": "2.0",
    "method": "cli",
    "params": {
      "cmd": "vlan 1234",
      "version": 1
    },
    "id": 2
  }
]
```


Cisco live!

Dynamic Python Program

```
import requests
import json
url='http://YOURIP/ins'
switchuser='USERID'
switchpassword='PASSWORD’
myheaders={'content-type':'application/json-rpc’}
payload=[
 {
   "jsonrpc": "2.0",
   "method": "cli",
   "params": {
   "cmd": "config t",
     "version": 1
   },
   "id": 1
   },
   {
     "jsonrpc": "2.0",
     "method": "cli",
     "params": {
       "cmd": "vlan 1234",
       "version": 1
     },
     "id": 2
   }
]
response = requests.post(url,data=json.dumps(payload),
headers=myheaders,auth=(switchuser,switchpassword)).json()
```

Python Consolidation

```
payload=[
  {
    "jsonrpc": "2.0",
    "method": "cli",
    "params": {
      "cmd": "config t",
      "version": 1
    },
    "id": 1
  },
  {
    "jsonr
    "method": "cli",
    "params": {
      ....
```

```
payload=[
      {"jsonrpc": "2.0","method": "cli","params": {"cmd": "config t","version": 1},"id": 1},
      {"jsonrpc": "2.0","method": "cli","params": {"cmd": "vlan 1234","version": 1},"id": 2},
      {"jsonrpc": "2.0","method": "cli","params": {"cmd": "exit","version": 1},"id": 3}
    ]
```

Adding a Loop to Python

```
1  import requests
2  import json
3
4  ip = [
5            '161.44.45.9',
6            '161.44.45.10'
7          ]
8  username = "admin"
9  password = "         "
10
11 print "enter vlan to be configured"
12 vlanId=raw_input()
13
14 for address in ip:
15     myheaders = {'content-type': 'application/json-rpc'}
16     url = "http://"+address+"/ins"
17     print url
18
19     payload=[
20       {"jsonrpc": "2.0","method": "cli","params": {"cmd": "conf t","version": 1},"id": 1},
21       {"jsonrpc": "2.0","method": "cli","params": {"cmd": "vlan "+vlanId,"version": 1},"id": 2},
22       {"jsonrpc": "2.0","method": "cli","params": {"cmd": "exit","version": 1},"id": 3}
23     ]
24
25     response = requests.post(url,data=json.dumps(payload), headers=myheaders,auth=(username,password)).json()
26   # print payload
```

Python Programing for Networking Engineers
@kirkbyers
http://pynet.twb-tech.com

Encrypting Script Traffic to the Devices

```
1  import requests
2  import json
3
4  ip = [
5            '161.44.45.9',
6            '161.44.45.10'
7          ]
8  username = "admin"
9  password = "        "
10
11 print "enter vlan to be configured"
12 vlanId=raw_input()
13
14 for address in ip:
15     myheaders = {'content-type': 'application/json-rpc'}
16     url = "https://"+address+"/ins"
17     print url
18
19     payload=[
20       {"jsonrpc": "2.0","method": "cli","params": {"cmd": "conf t","version": 1},"id": 1},
21       {"jsonrpc": "2.0","method": "cli","params": {"cmd": "vlan "+vlanId,"version": 1},"id": 2},
22       {"jsonrpc": "2.0","method": "cli","params": {"cmd": "exit","version": 1},"id": 3}
23     ]
24
25     response = requests.post(url,data=json.dumps(payload), headers=myheaders,auth=(username,password), verify=False).json()
```

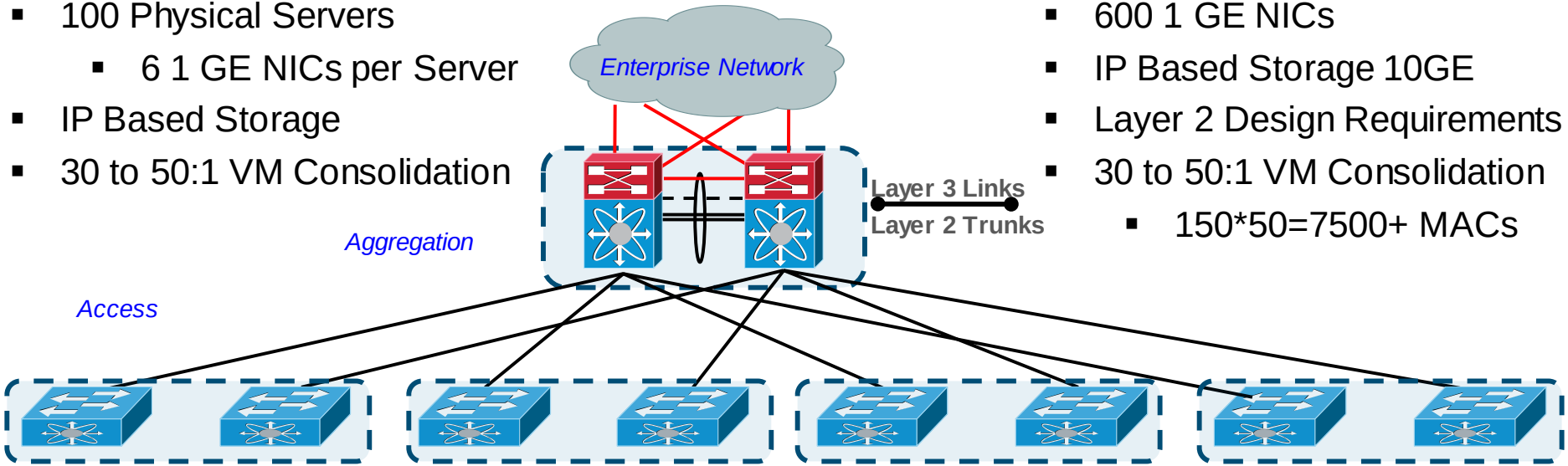
Small Data Centre/Colo Design

Cisco live!

Data Centre Building Blocks

Small Data Centre/CoLo facility

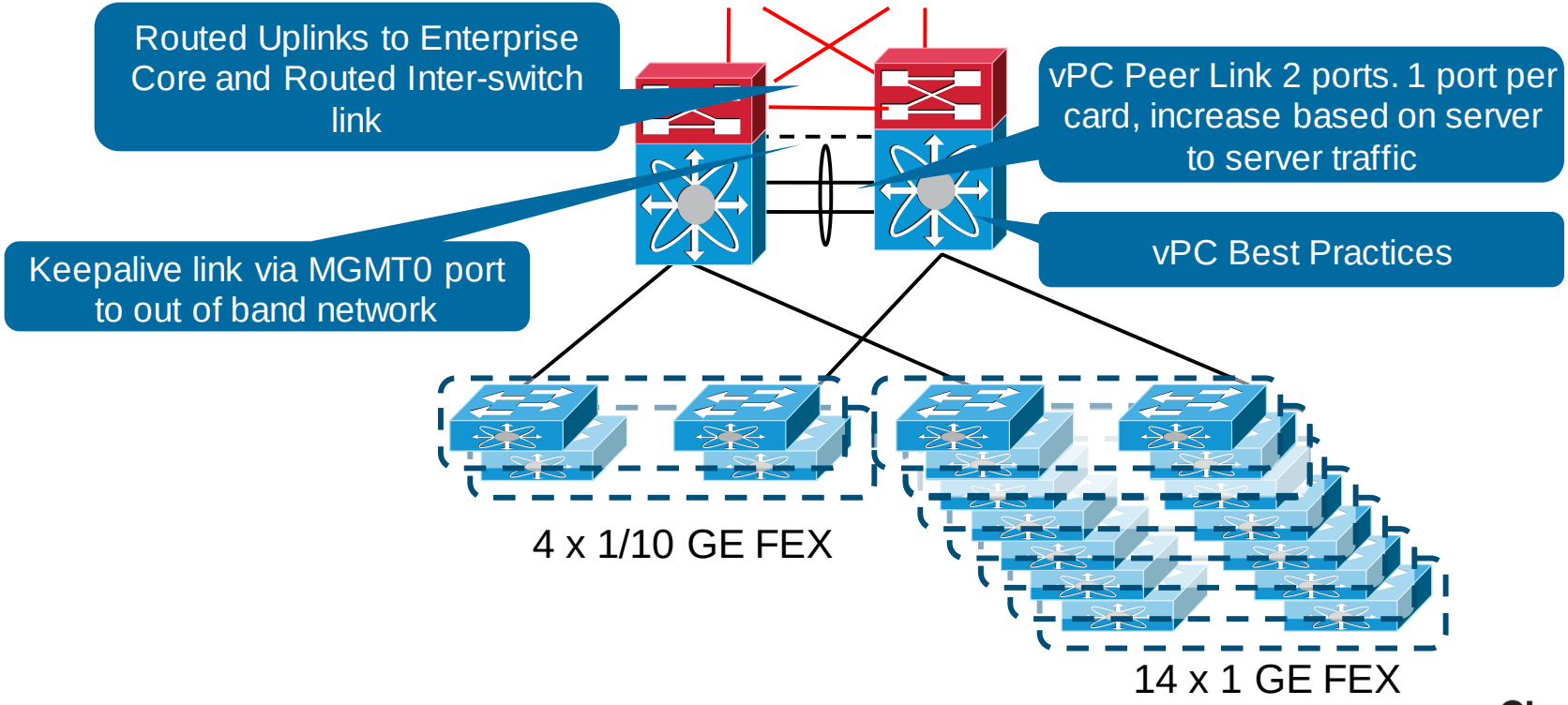
- 50 Physical Servers
  - 2 10GEs per Server
- 100 Physical Servers
  - 6 1 GE NICs per Server
- IP Based Storage
- 30 to 50:1 VM Consolidation



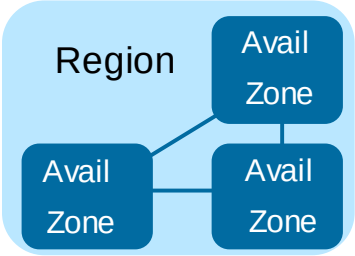
- Dual Attached Servers+
- 100 10GE interfaces
- 600 1 GE NICs
- IP Based Storage 10GE
- Layer 2 Design Requirements
- 30 to 50:1 VM Consolidation
  - 150*50=7500+ MACs

Data Centre Building Blocks

Function & Key Considerations



Data Centre Building Blocks



Scaling Concerns

- Control Plane Scale
  - ARP learning
  - MAC addresses
  - CPU Traffic Level, any packets that get punted to the CPU
- Spanning Tree Scale
  - RSTP -> 16k Logical Ports, logical port limit is equal (# of ports)*(Vlans per ports)
  - MST -> 25K Logical Ports, logical port limit is equal (# of ports)*(# of MST instances allowed per port)
- Port Channel Scaling Numbers
- Buffer Oversubscription
- Failure Domain Size (Availability Zones)
  - ISSU

Multicast Example

NXOS Best Practices

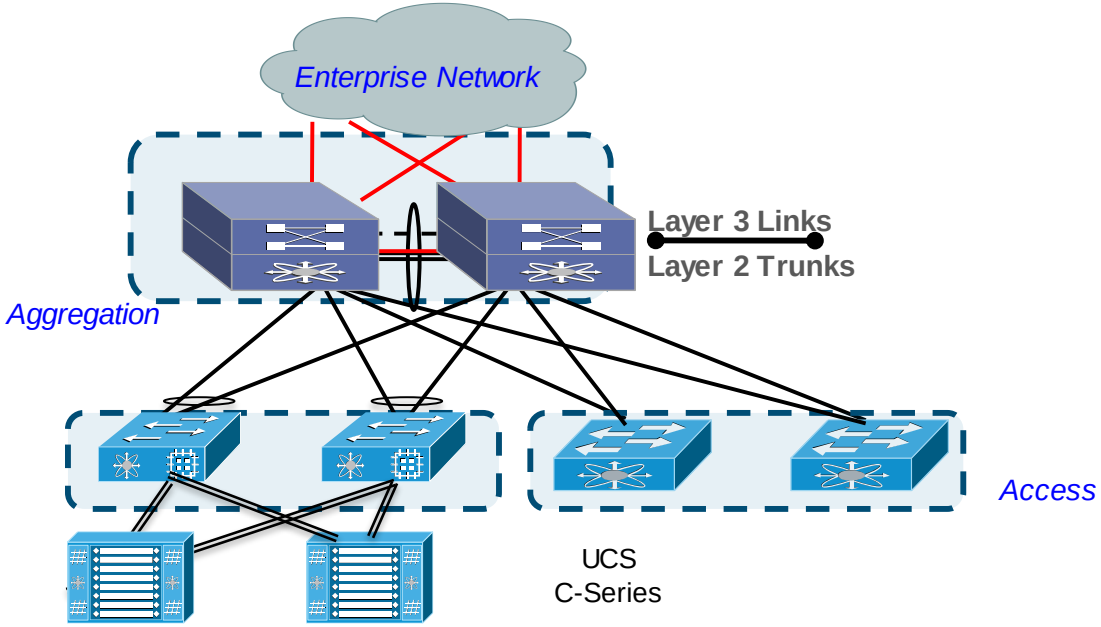
```
Anycast-RP 1:
feature pim
feature eigrp
interface loopback0
  ip address 10.1.1.6/32
  ip router eigrp 10
  ip pim sparse-mode

interface loopback1
  ip address 10.10.10.50/32
  ip router eigrp 10
  ip pim sparse-mode

router eigrp 10
ip pim rp-address 10.10.10.50 group-list 224.0.0.0/4
ip pim ssm range 232.0.0.0/8
ip pim anycast-rp 10.10.10.50 10.1.1.4
ip pim anycast-rp 10.10.10.50 10.1.1.6
```

```
Anycast-RP 2:
feature pim
feature eigrp
interface loopback0
  ip address 10.1.1.4/32
  ip router eigrp 10
  ip pim sparse-mode

interface loopback1
  ip address 10.10.10.50/32
  ip router eigrp 10
  ip pim sparse-mode

router eigrp 10
ip pim rp-address 10.10.10.50 group-list 224.0.0.0/4
ip pim ssm range 232.0.0.0/8
ip pim anycast-rp 10.10.10.50 10.1.1.4
ip pim anycast-rp 10.10.10.50 10.1.1.6
```

Data Centre Building Blocks

Small Data Centre/CoLo facility EvPC based and UCS

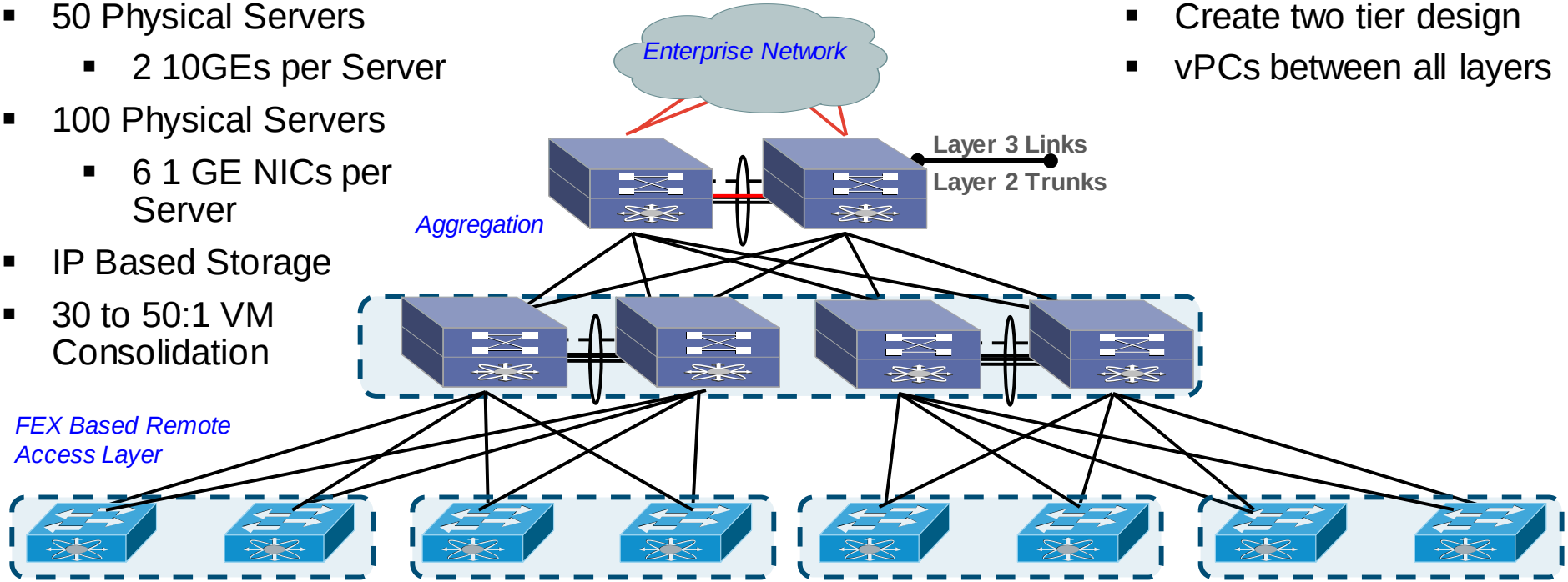
- Nexus 5600 or Nexus 7000 Aggregation
- ISSU, L3 Interconnect and DCI match previous slide
- Access Mix of FEX and UCS Fabric Interconnect
  - Do not connect UCS FI into FEX.
- vPC from UCS FI to Aggregation



Data Centre Building Blocks

Small Data Centre/CoLo facility EvPC based 5600 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

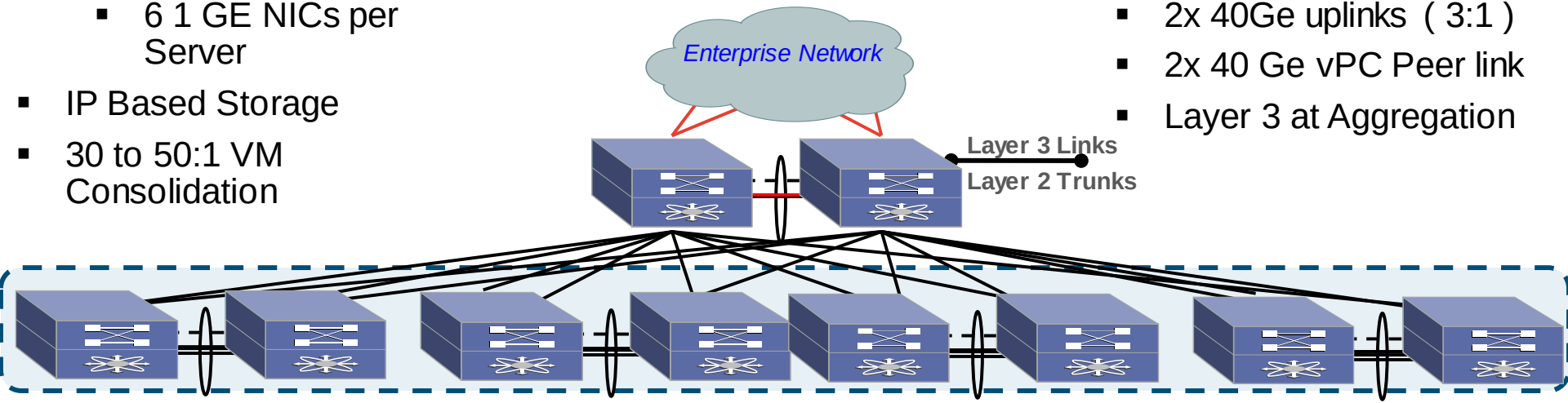


- Create two tier design
- vPCs between all layers

Data Centre Building Blocks

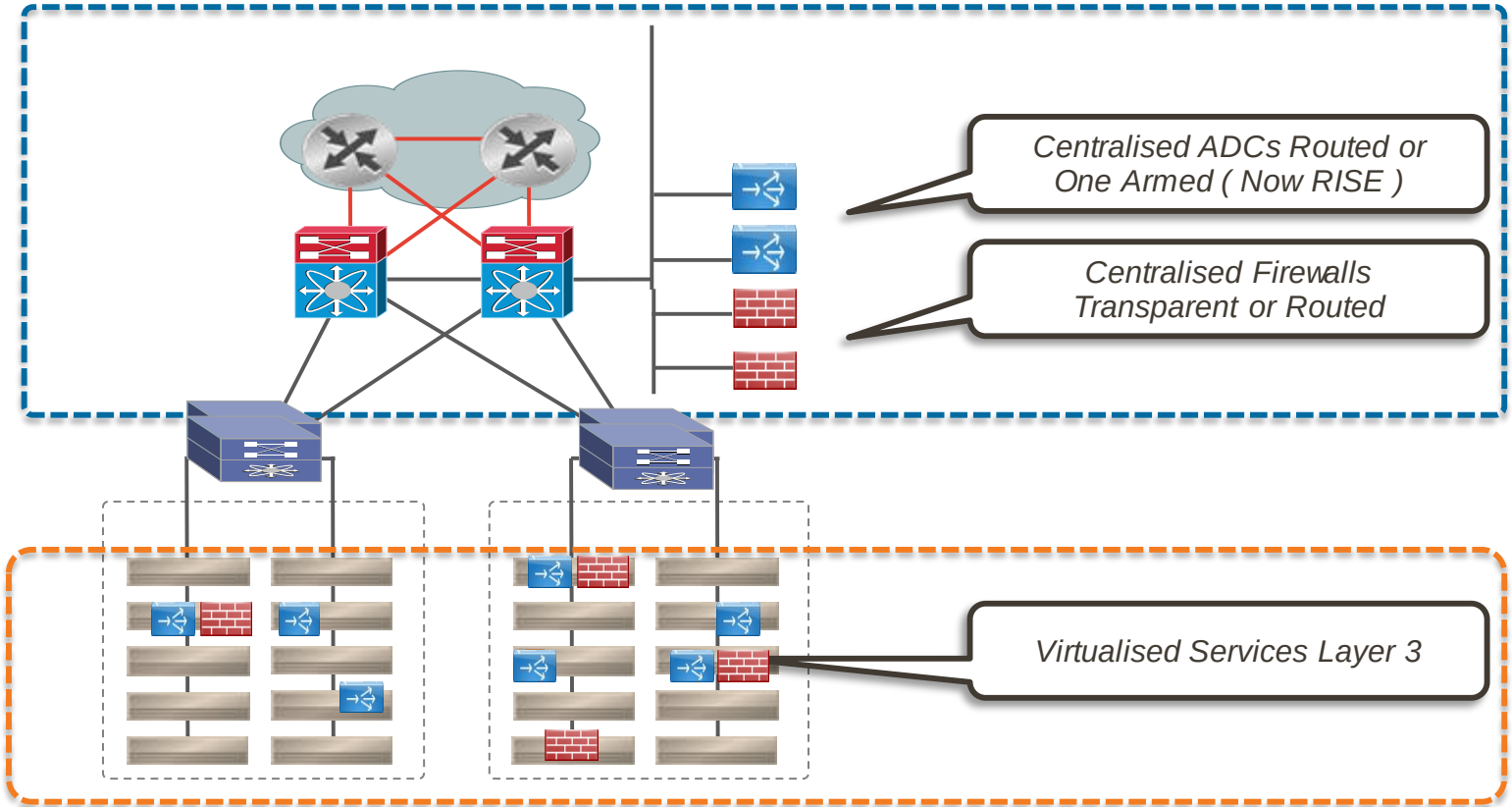
Small Data Centre/CoLo without FEX Design

- 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



- No FEXes
- Create two tier design
- vPCs between all layers
- 2x 40Ge uplinks (3:1)
- 2x 40 Ge vPC Peer link
- Layer 3 at Aggregation

Services Deployment Models

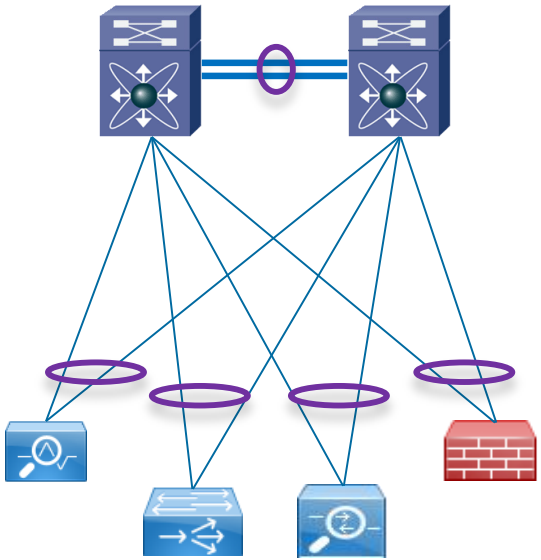


Dedicated Solutions have ability to use Hardware Acceleration Resources like SSL Offload

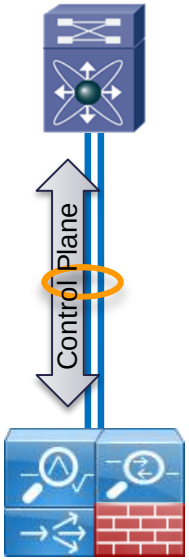
Cisco Remote Integrated Service Engine (RISE)

Challenge: Services and switching are deployed independently which increases the complexity for deploying and maintaining networks

Physical Topology



Logical RISE Topology

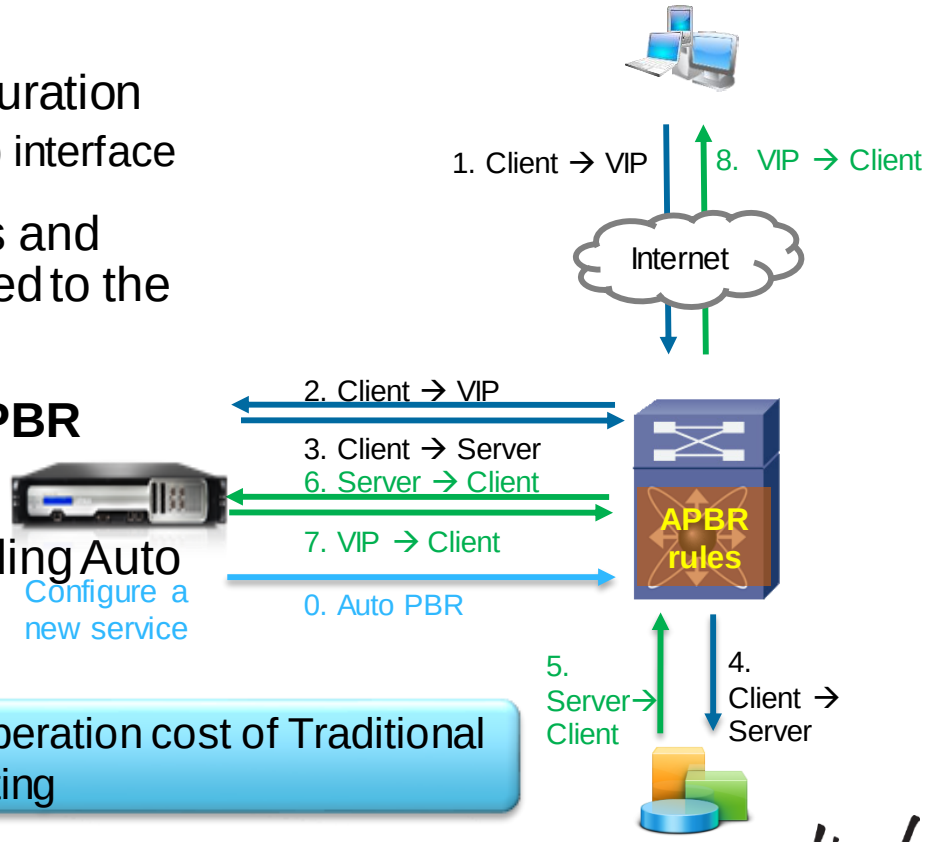


RISE Overview:

- Logical integration of a service appliance with Nexus 7000 and 7700 platforms
- Enables staging to streamline initial deployment of the service appliance
- Allows ongoing configuration updates to drive flows to and from the service appliance
- Allows data path acceleration and increased performance
- Integrated with N7K VDC architecture

Cisco Solution: Use RISE for Auto PBR

- NS adds redirection rules as per configuration
  - Sends the list of servers and the next hop interface
- N7K applies to rules for its local servers and propagates the rules for servers attached to the neighbouring N7K
- **No need for Source-NAT or manual PBR configuration**
- Uses the RISE control channel for sending Auto PBR messages

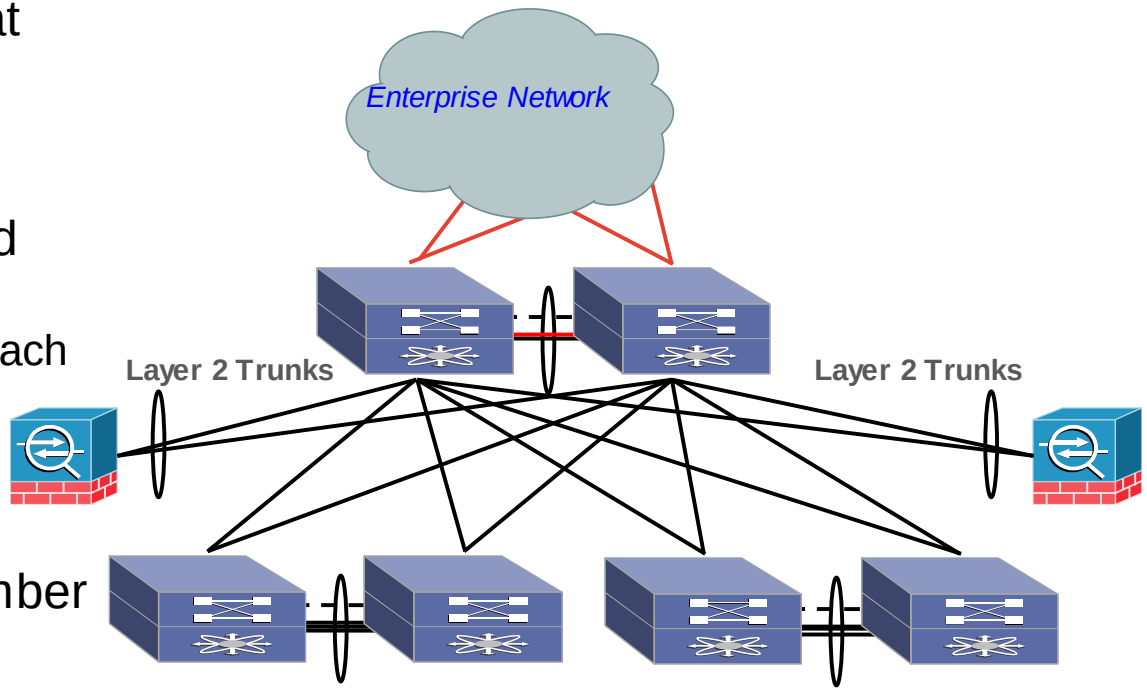


Preserve Client IP Visibility without the operation cost of Traditional Policy Based Routing

Securing the vPC Based Design

Transparent Firewall Insertion

- Insert Transparent Firewall at Aggregation Point
- vPC connect firewalls
- Trunk only Security Required Vlans
  - Due to Transparent Firewall each vlan will actually be two
  - Creating a numbering system
  - Watch CoPP for ARP
- Firewall tier is scaled by number of VLANs per pair
- Leave L3 on Aggregation



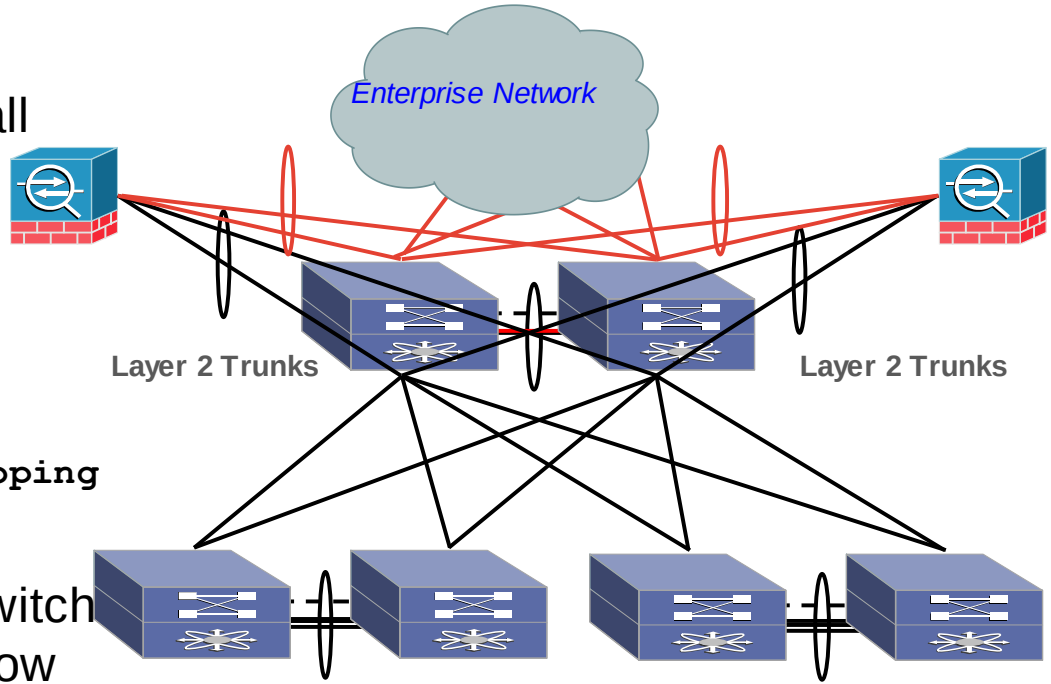
Layer 3 Firewalls Insertion

Servers Default Gateway located on Firewall

- vPC connect firewalls
- Server Default Gateway on Firewall
- If Clustering or L2 Heartbeats required you need to handle igmp

```
N5k# configure terminal
N5k(config)# vlan 5
N5k(config-vlan)# no ip igmp snooping
```

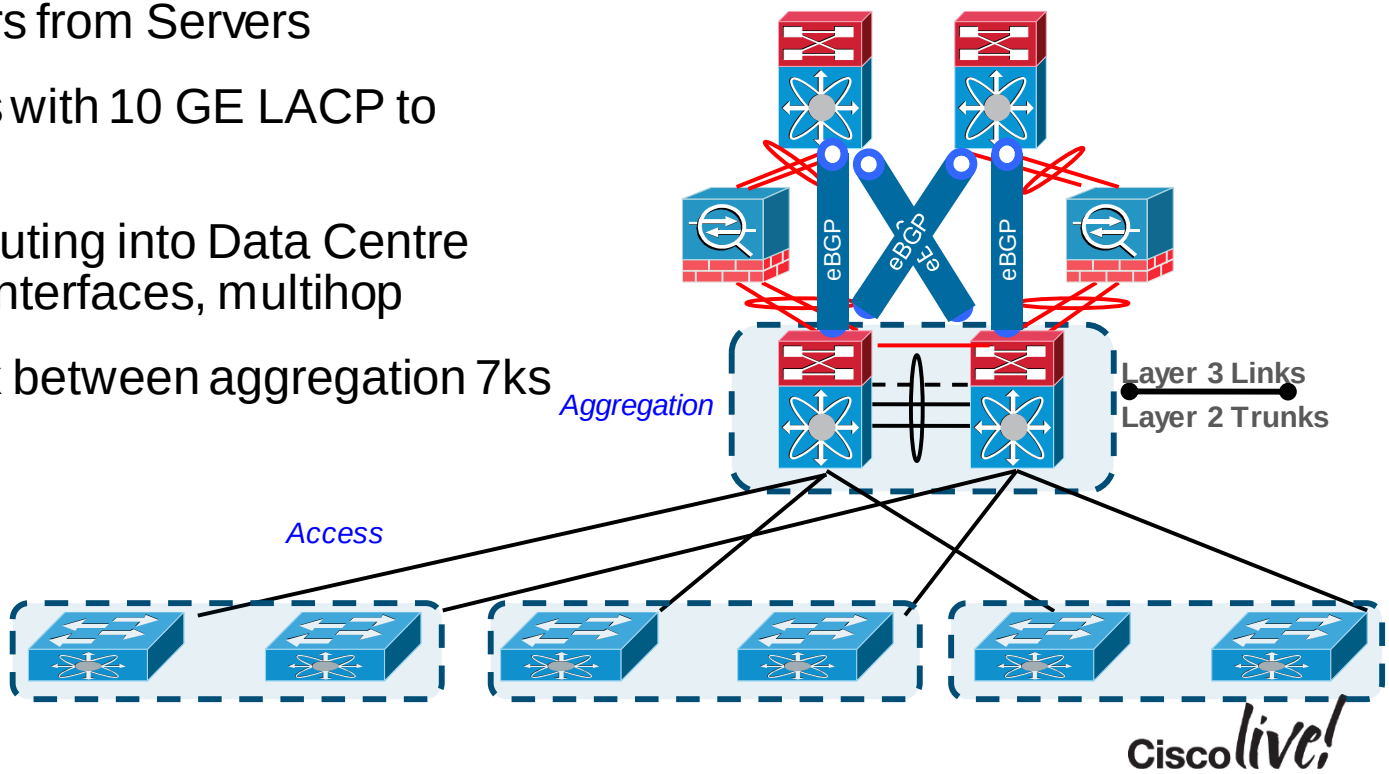
- Look at Moving Layer 3 back to Switch with VRFs to create isolation to allow for more flexibility



Data Centre Building Blocks

Routing with Security Requirements

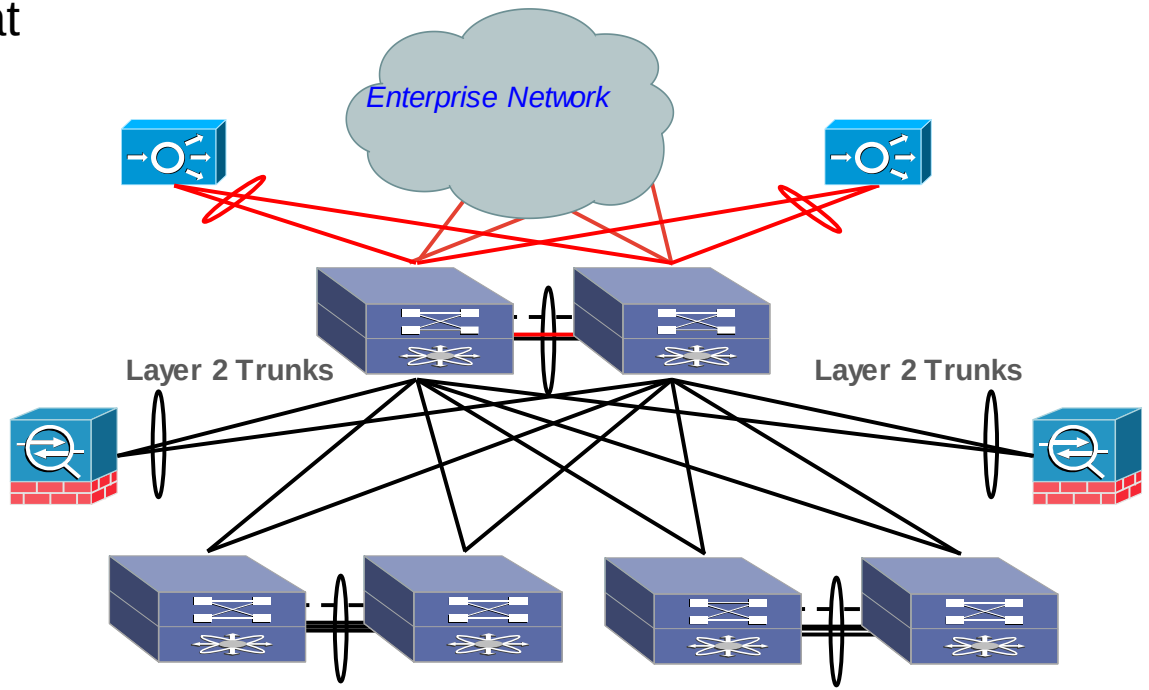
- Firewall off all Users from Servers
- Deployed Firewalls with 10 GE LACP to Aggregation tier.
- eBGP to provide routing into Data Centre against Loopback interfaces, multihop
- Define Routed Link between aggregation 7ks for Routing



Securing the vPC Based Design

Adding in ADC, Application Delivery Controllers

- Insert Transparent Firewall at Aggregation Point
- vPC connect ADCs and Firewalls
- Load balancer configured in One Armed, Routed or via RISE
- Source NAT used to direct traffic back to LB or RISE



If Routing on Services nodes, use standard Etherchannel not vPCs

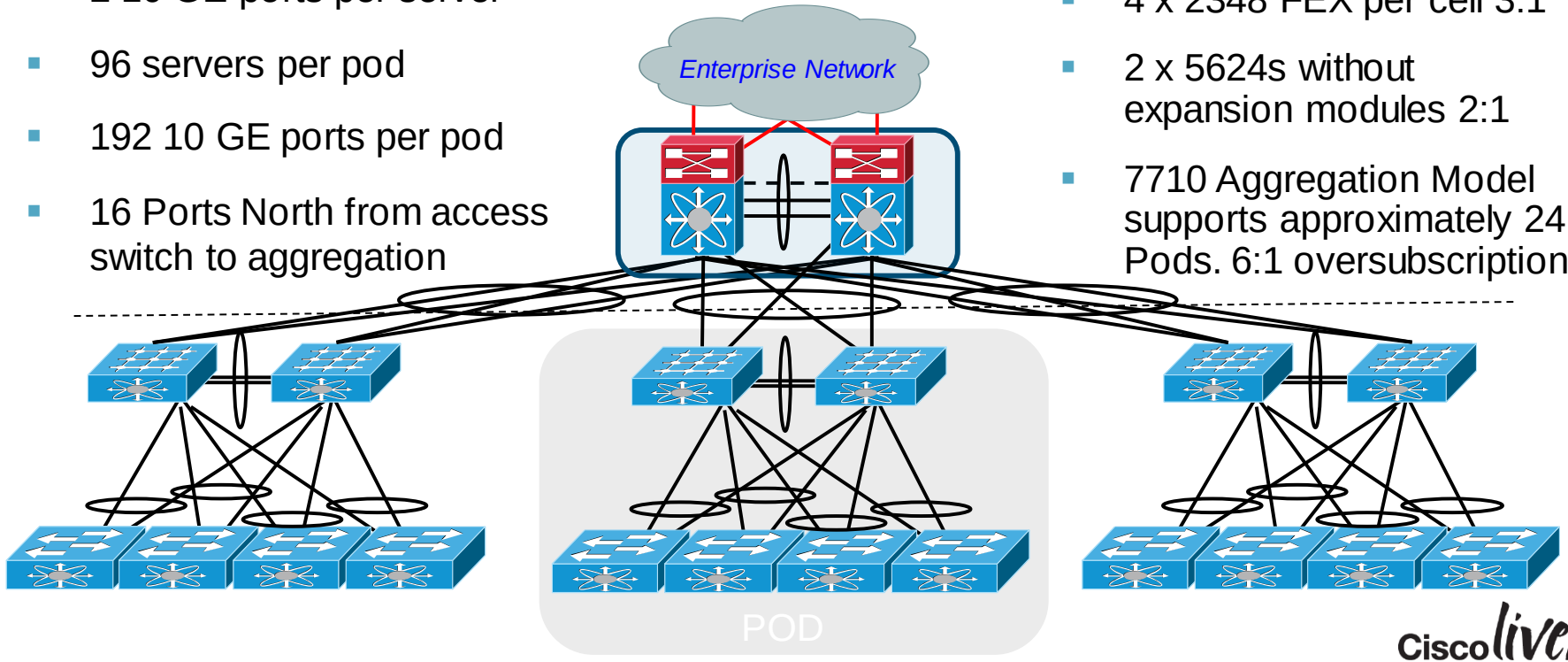
Scalable Layer 2 Data Centre with vPC

Cisco live!

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

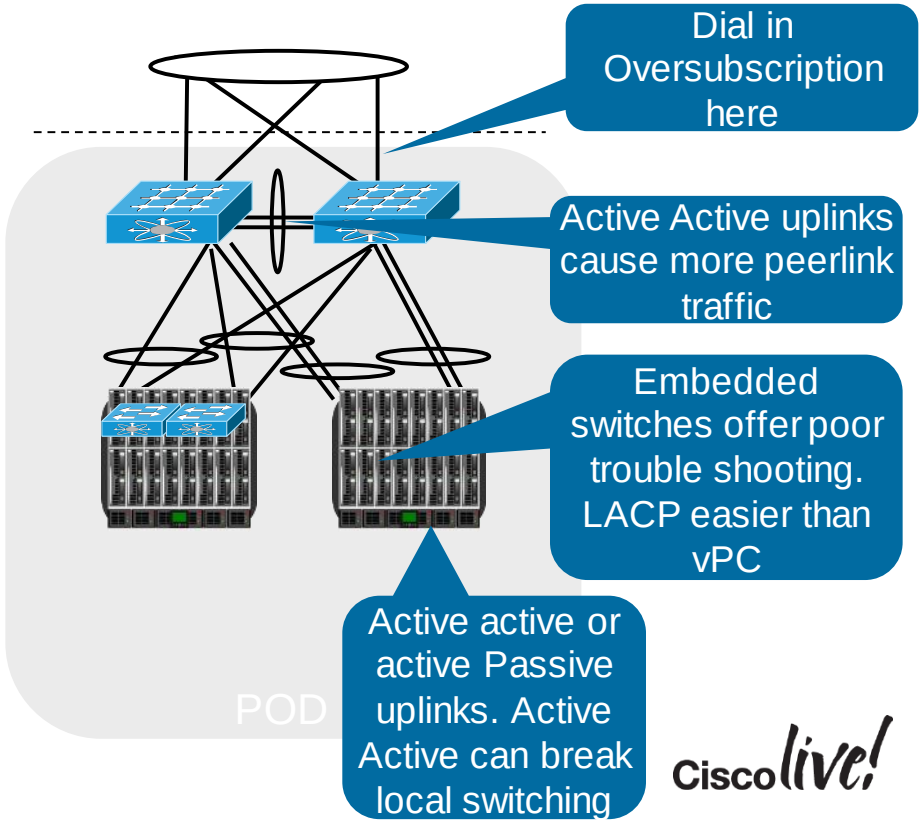


- 4 x 2348 FEX per cell 3:1
- 2 x 5624s without expansion modules 2:1
- 7710 Aggregation Model supports approximately 24 Pods. 6:1 oversubscription

Data Centre Building Blocks Larger Design

3rd Party Blade Enclosures

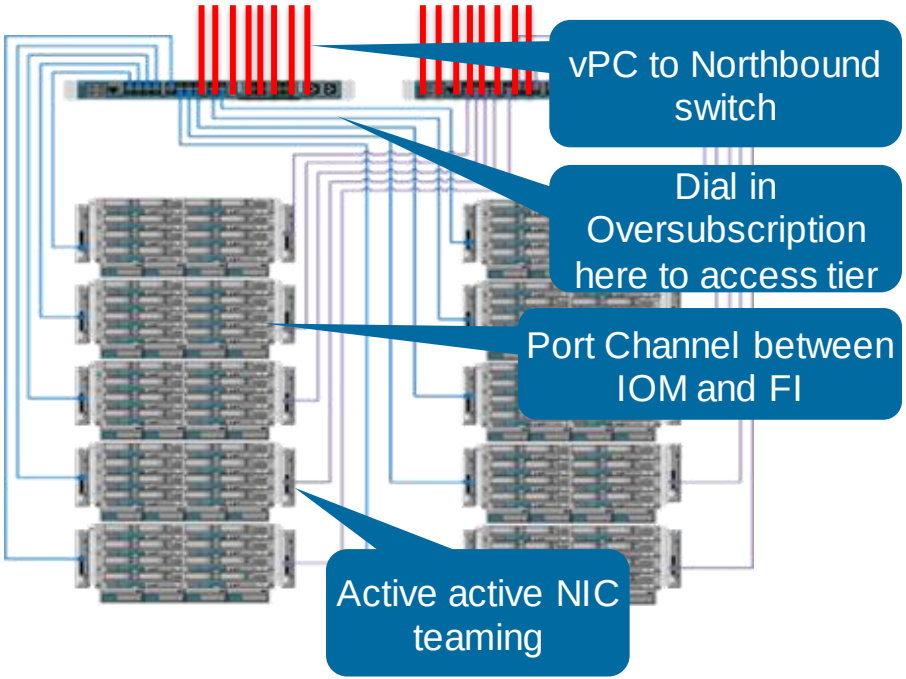
- Straight through to embedded switch or to FEX
- vPC to FEX embedded in Blade enclosure, HP, Dell, Fujitsu
- 6 Blade enclosures per Access switch Pair based on oversubscription numbers
- 8 uplinks per switch
- 4 ports for peer link without embedded FEX



Data Centre Building Blocks Larger Design

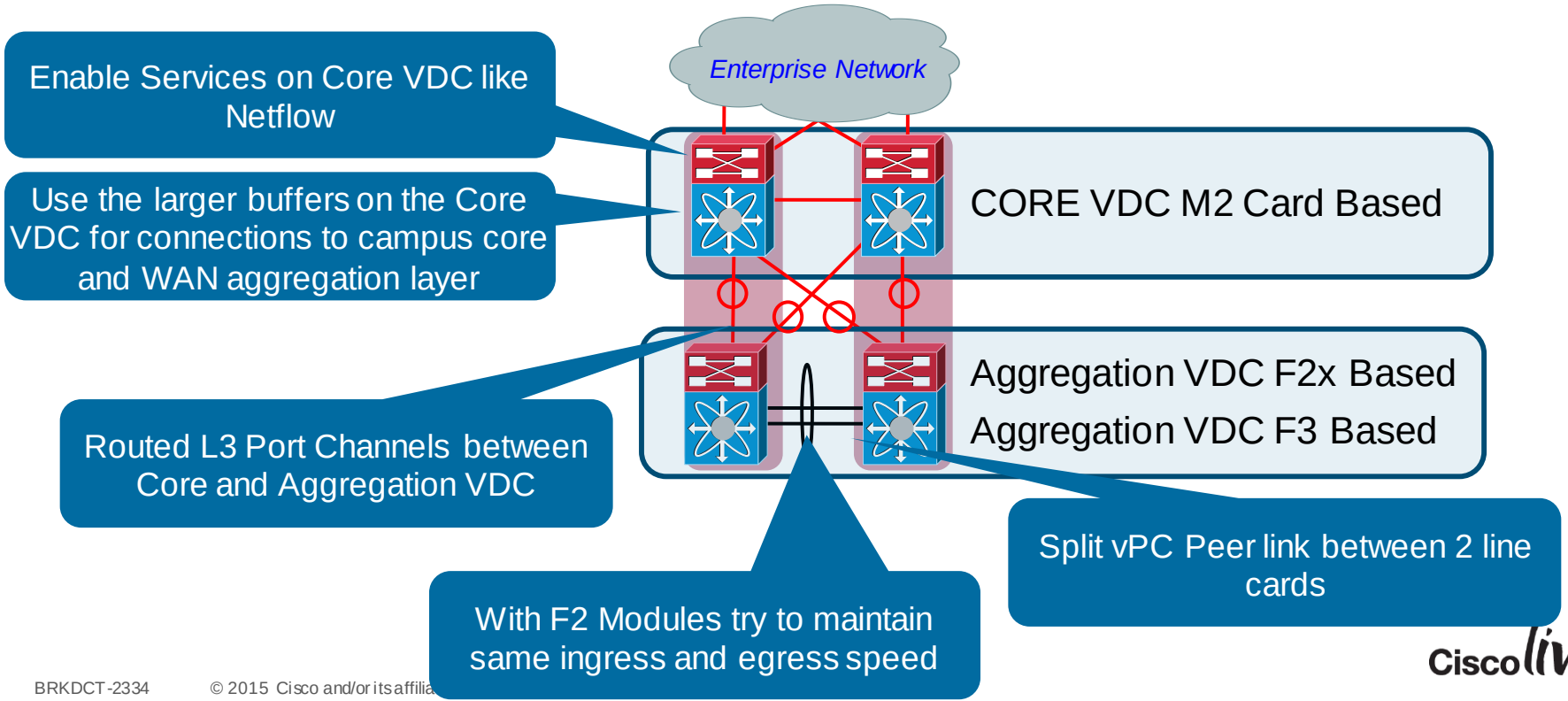
Cisco Blade Enclosures Pod

- 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.



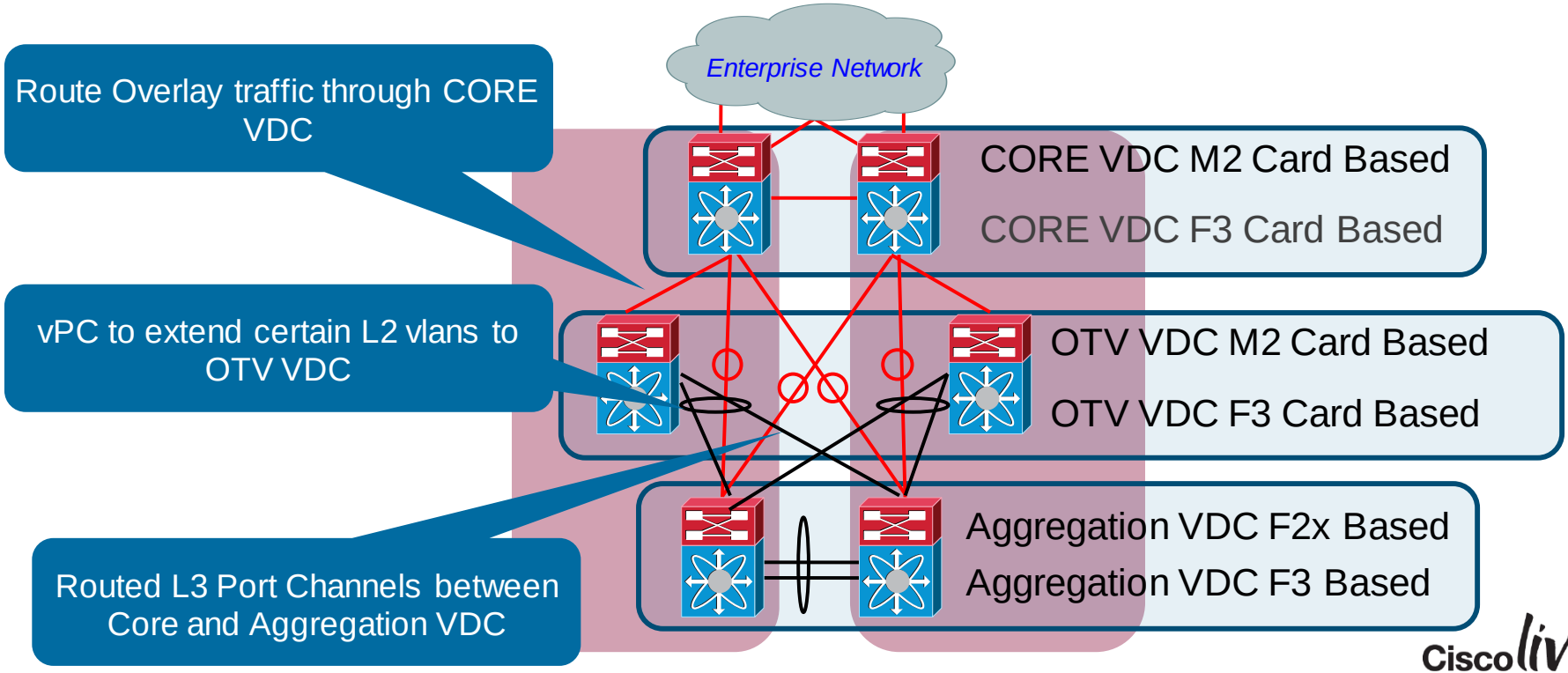
Data Centre Building Blocks Larger Design

Aggregation Layer Detailed Break out



Data Centre Building Blocks Larger Design

Aggregation Layer Detailed Break out



Scaling Points of vPC Design

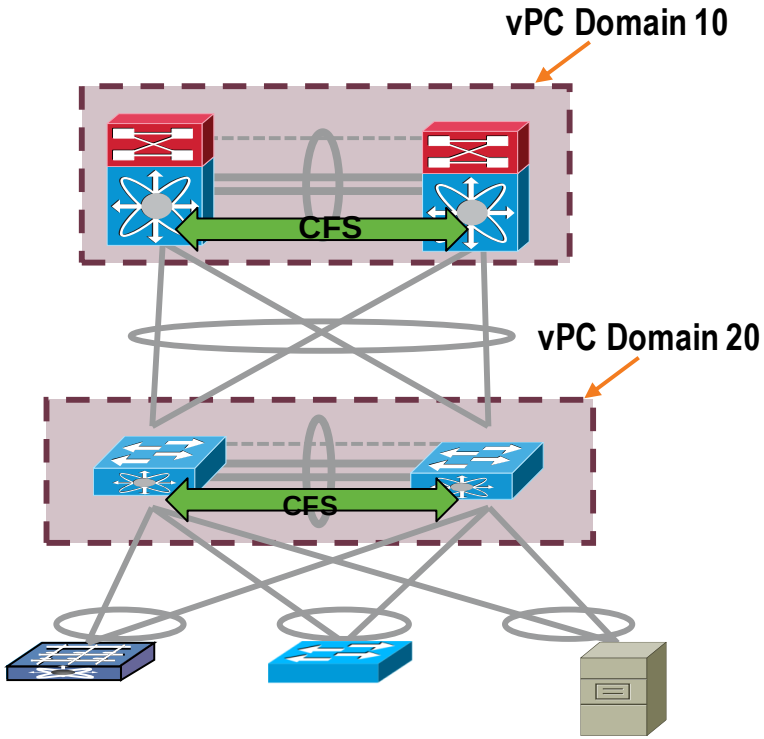
- Configuration Complexity
  - vPC Configuration needs to be replicated to both nodes
  - Failures could isolate orphan ports
- Scaling Limitations
  - F3 Modules today support 64K MAC addresses
  - F2 and F2e Modules today support 16k MAC addresses
    - F2e Proxy functionality to scale MAC address 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

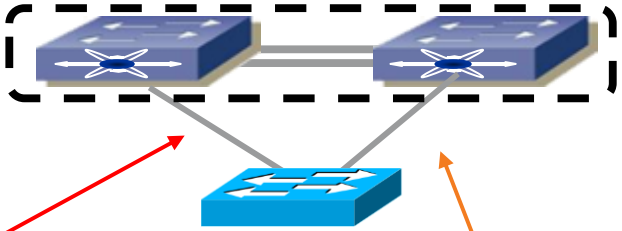


- Both switches in the vPC Domain maintain distinct control planes
- CFS provides for protocol state synchronisation between both peers (MAC table, IGMP state, …)
- Currently a manual process with an automated consistency check to ensure correct network behaviour
- Two types of interface consistency 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 Checks

- 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 native vlan 100
  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
```

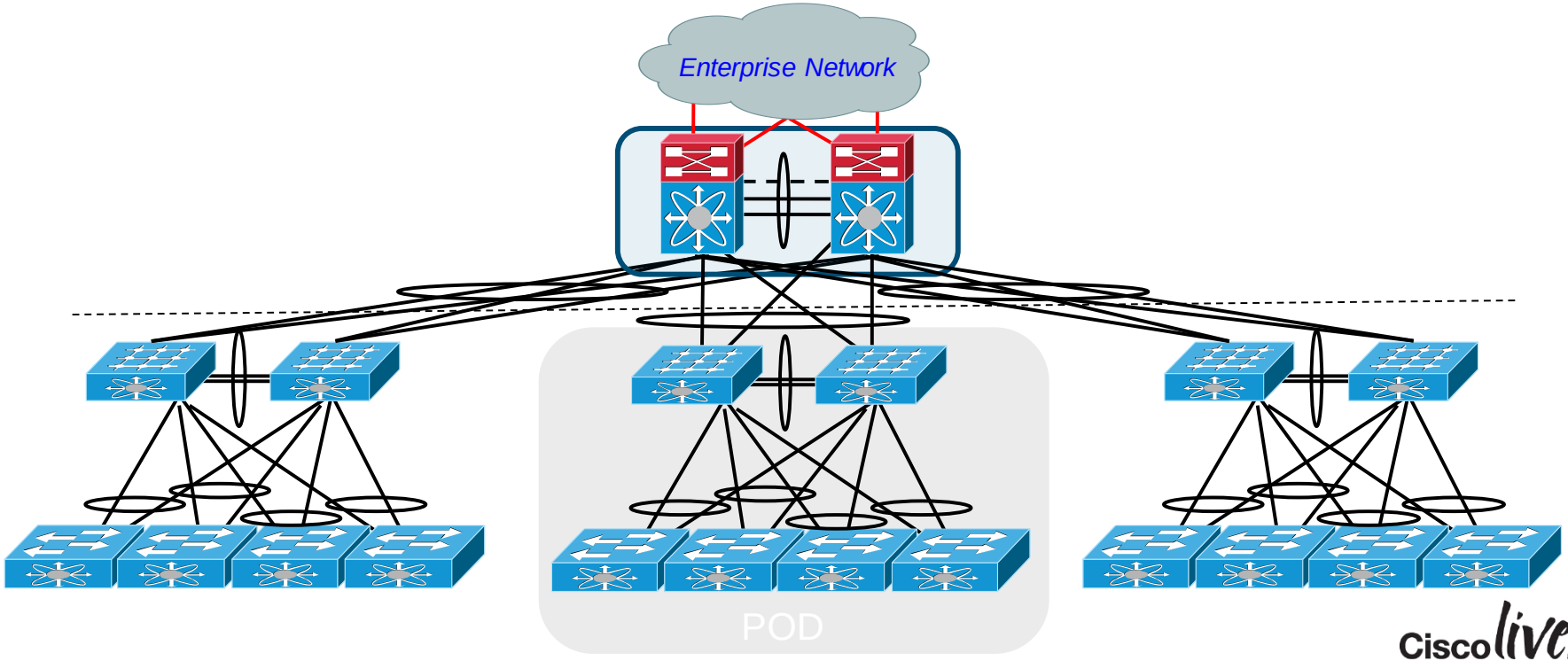
```
5020-1# show vPC brief vPC 201
vPC status
----------------------------------------------------------------------------
id     Port        Status Consistency Reason                     Active vlans
------ ----------- ------ ----------- -------------------------- -----------
201    Po201       up     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

Cisco live!

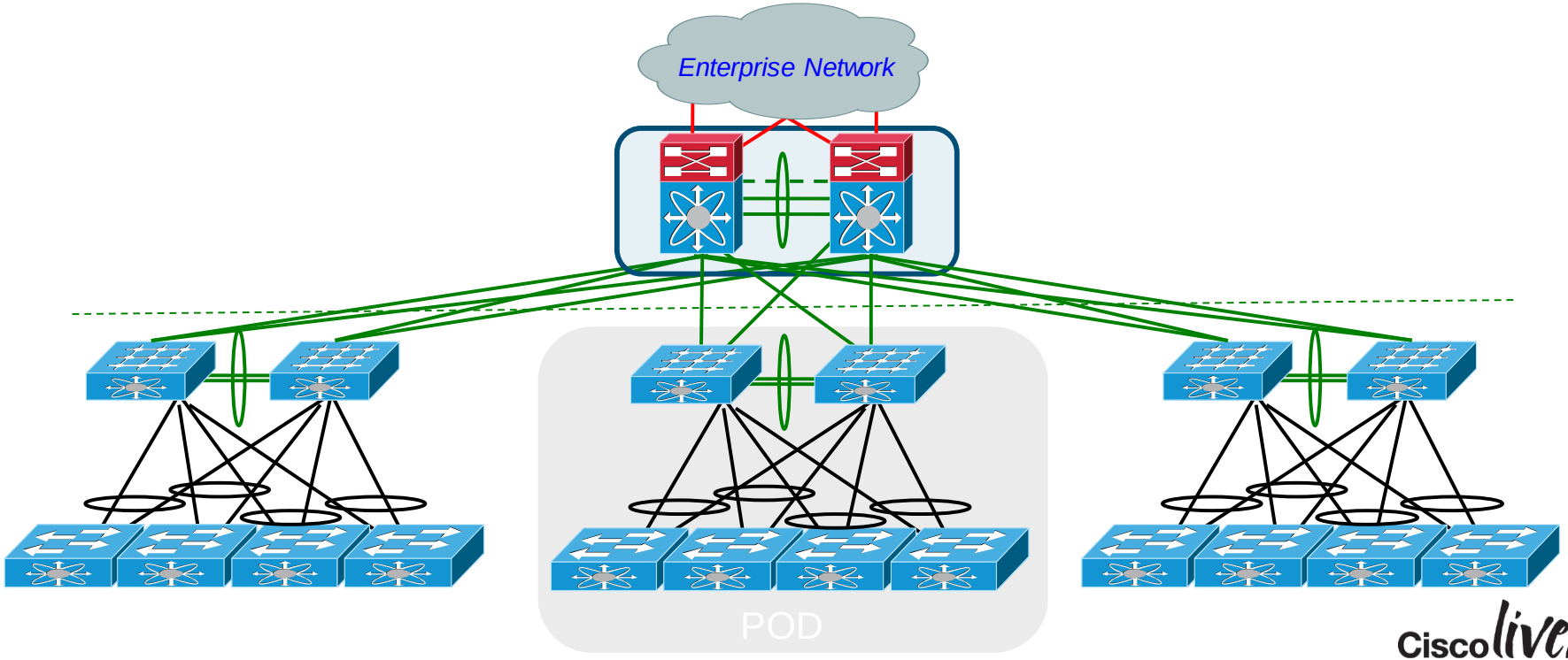
2/3 Tier Data Centre Building Blocks

Needing to scale



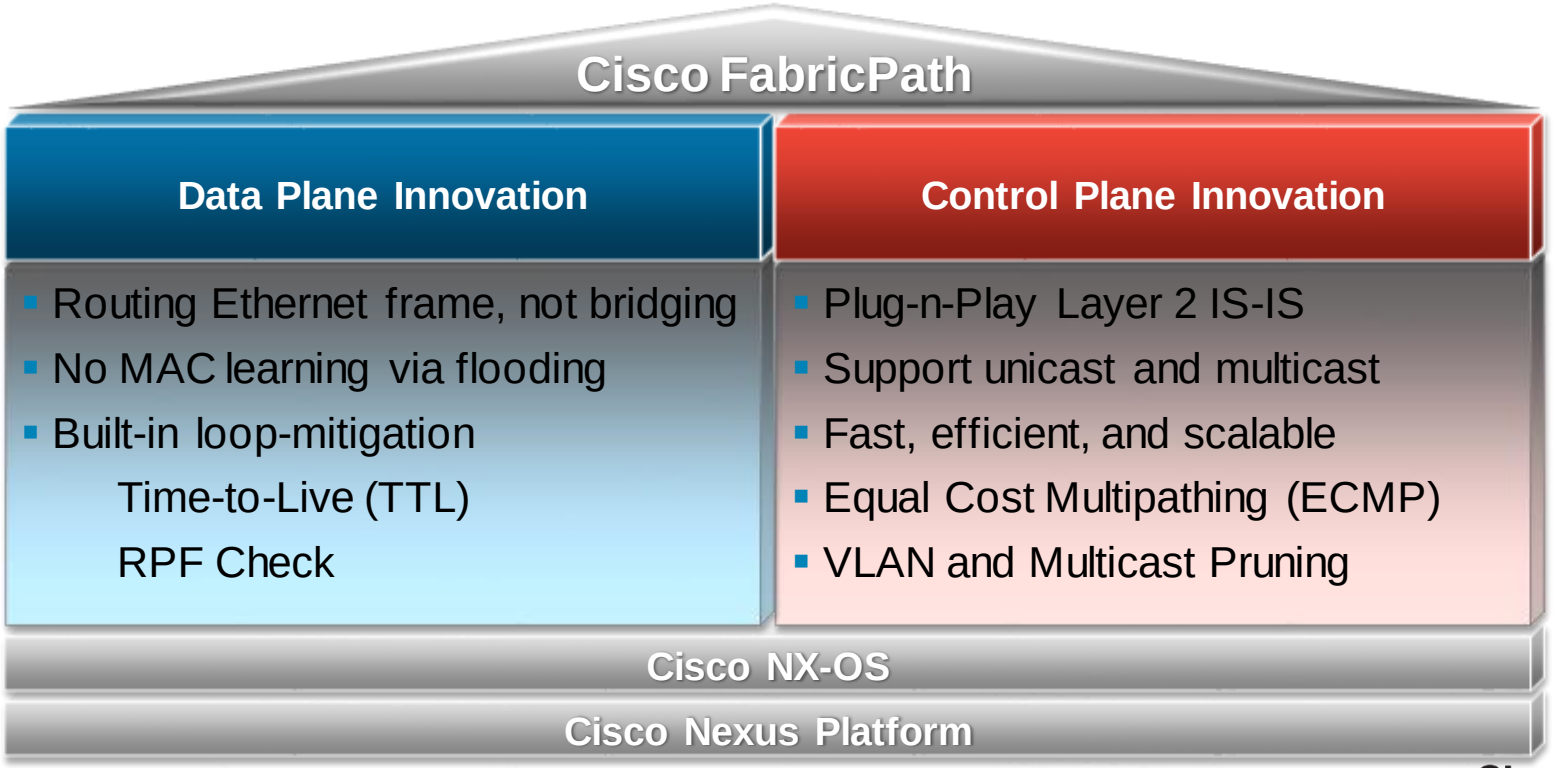
2/3 Tier Data Centre Building Blocks

Moving to Fabric Path



Introduction to Cisco Fabric Path

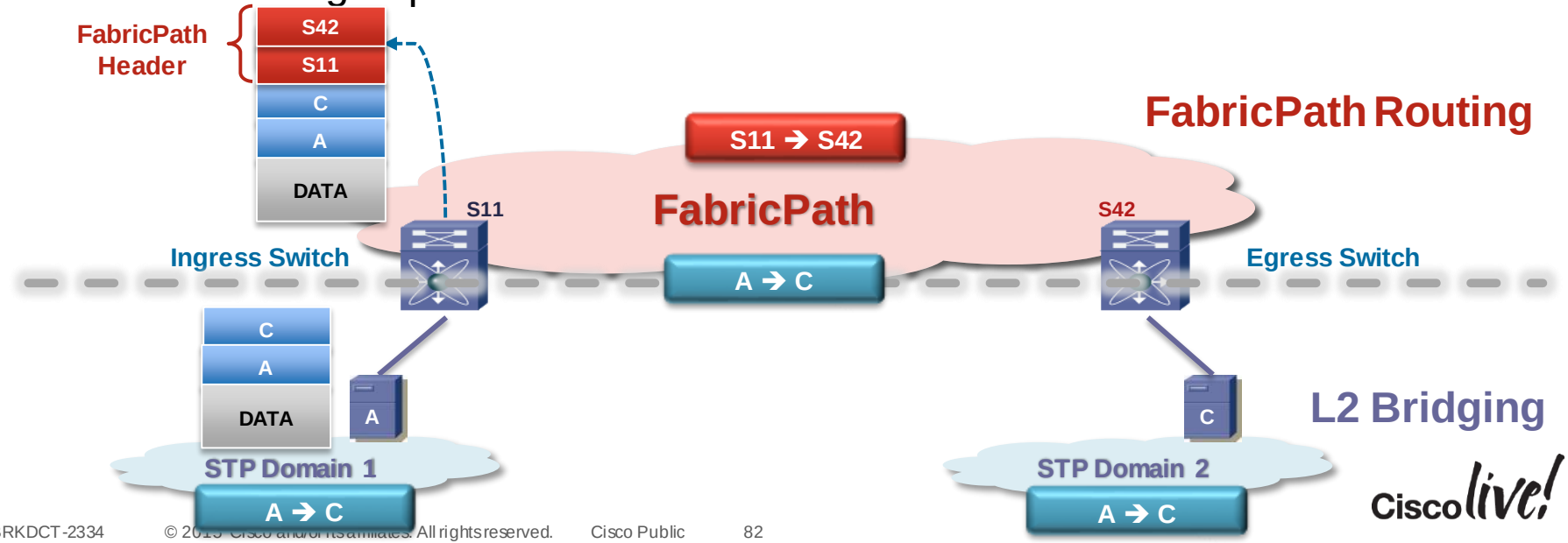
An NX-OS Innovation Enhancing L2 with L3



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

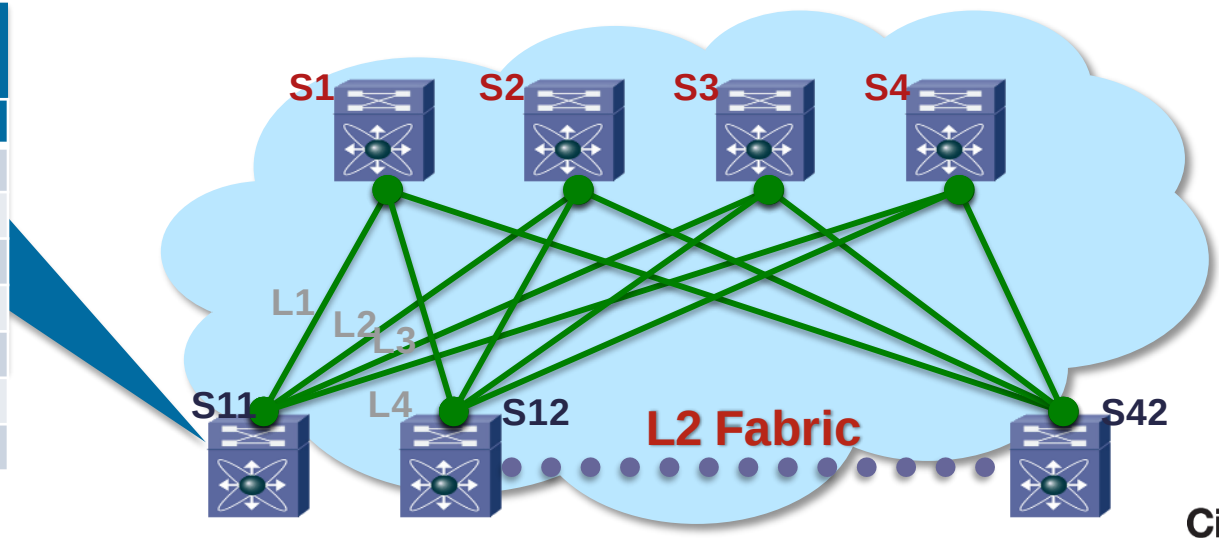


Control Plane Operation

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



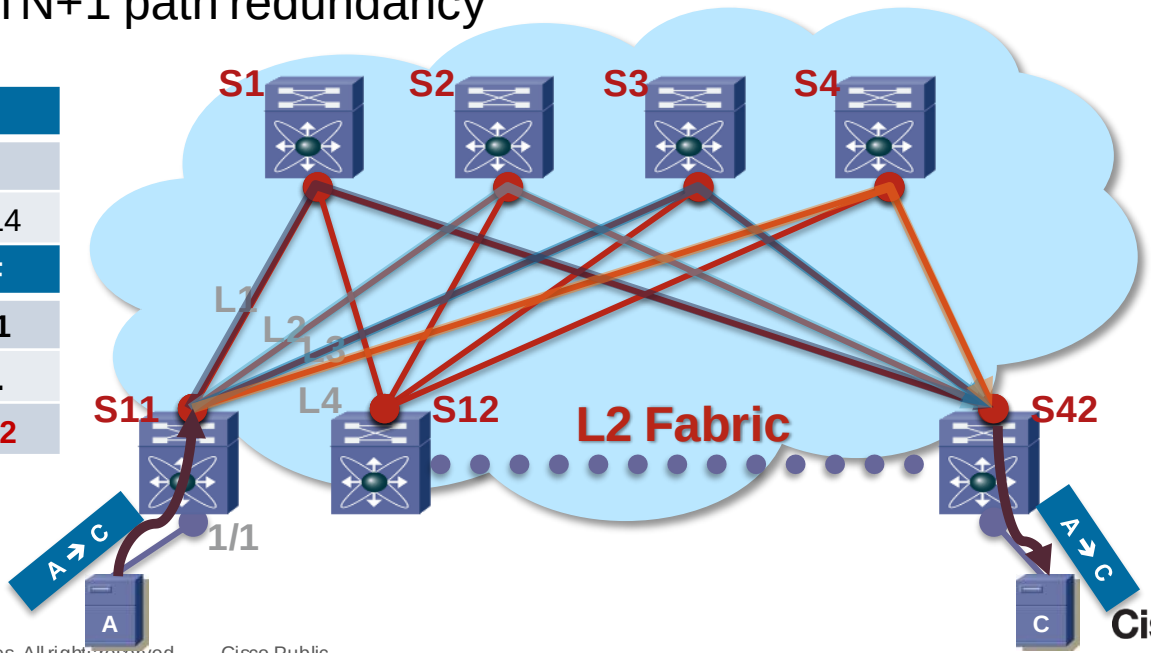
Unicast with FabricPath

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



Layer 3 Locations with Fabric Path

Layer 3 at Spine

- Overload Bit does not delay emulated switch id advertisement currently
- MAC scale is based off of the F2 or F3 modules being used
- Reduced points of configuration

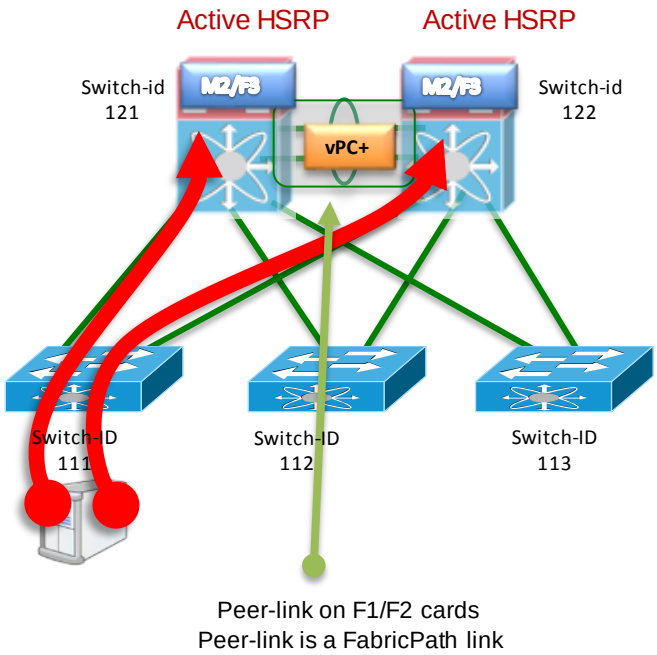
Layer 3 attached to Border Leaf

- Overload Bit provide fast failover on startup
- MAC scale can we scaled horizontally by adding in multiple GWs
- Common point of configuration for Layer 3

Distributed Layer 3 at each Leaf

- Overload Bit provides fast failover no startup
- MAC scale at edge
- Management application to synchronise configurations for Layer 3

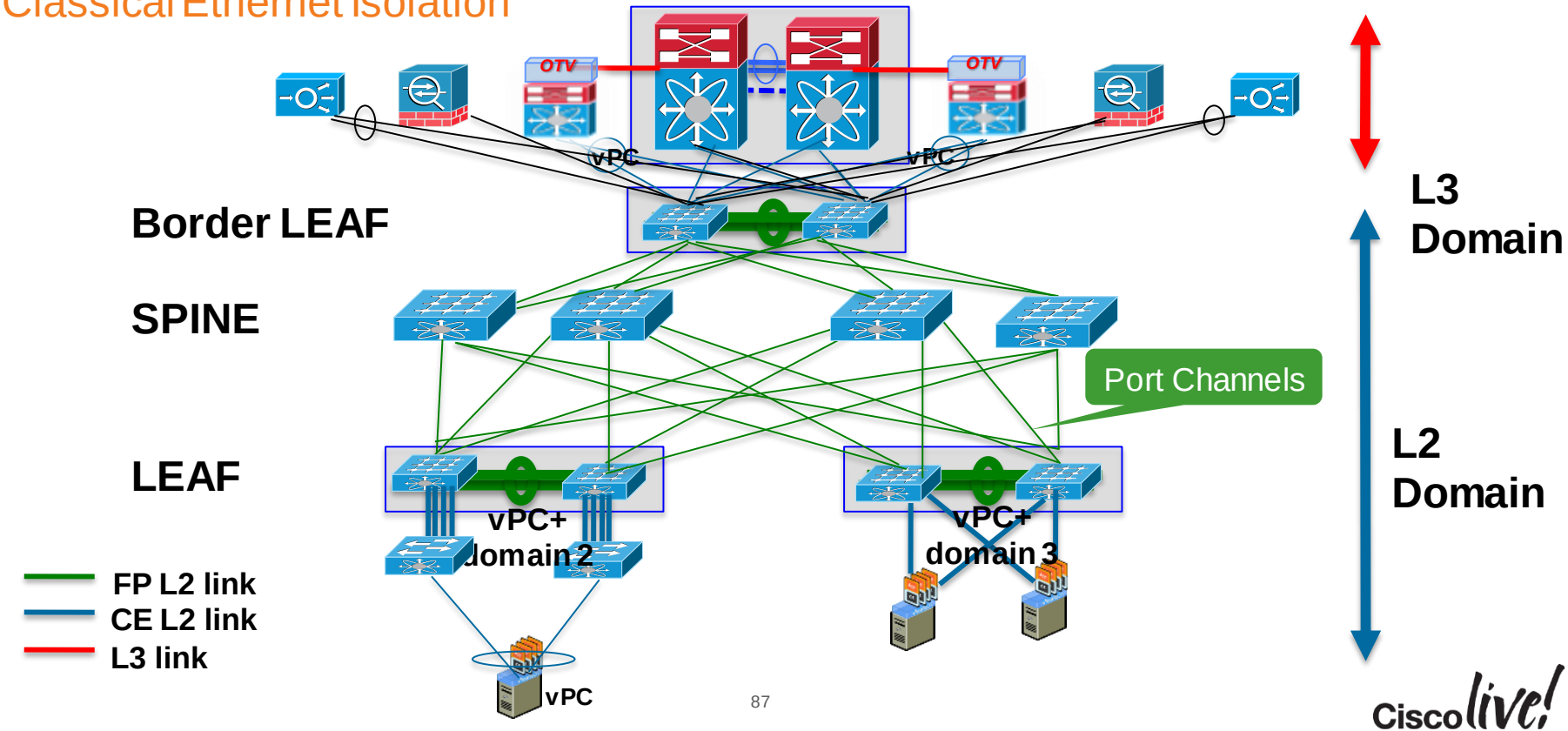
FabricPath - vPC+ at SPINE Layer



- 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

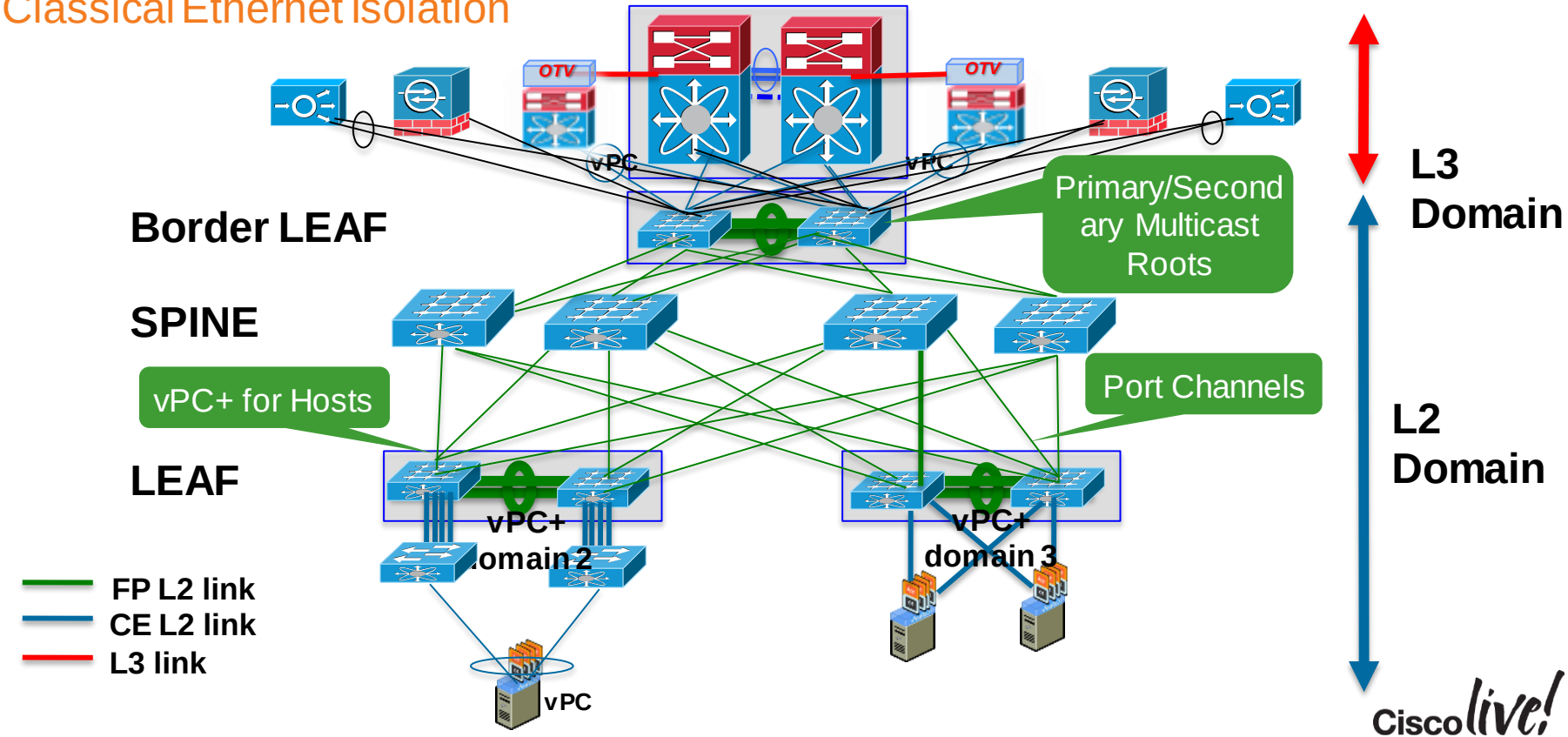
Fabric Path Based Data Centre

Classical Ethernet Isolation



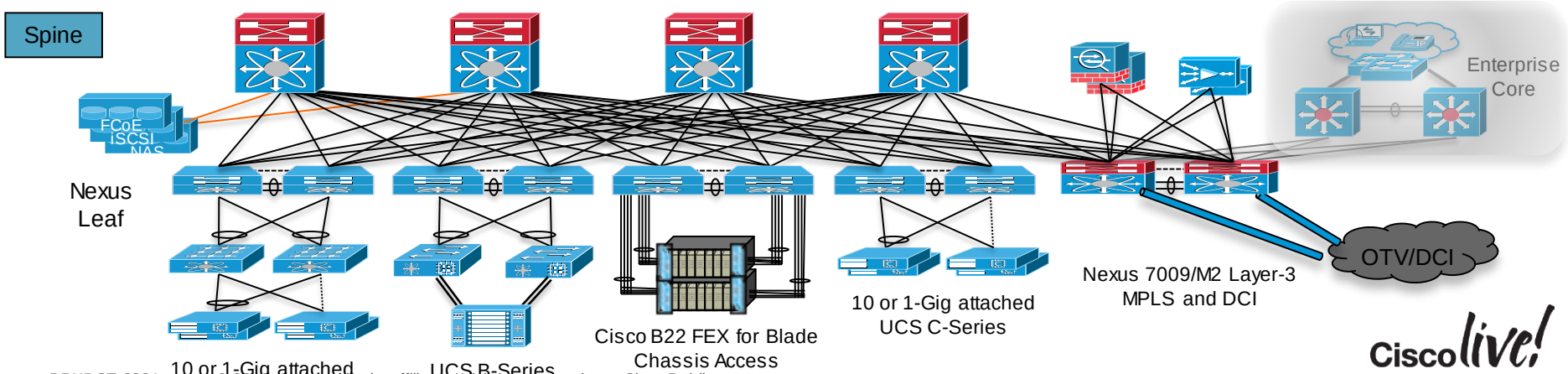
Fabric Path Based Data Centre

Classical Ethernet Isolation



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

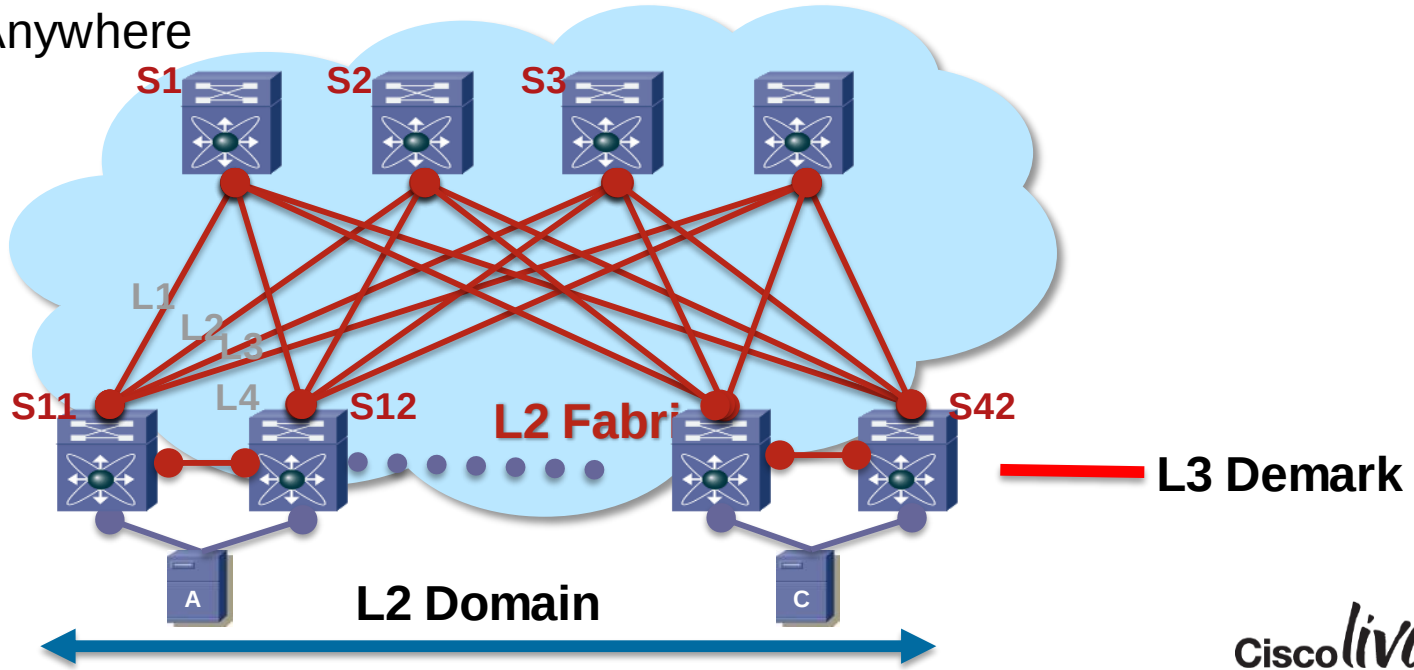
- Nexus 6004/7000 Spine layer creating Layer-2 FabricPath domain
- Nexus TOR switches deployed in vPC(+) pairs for edge link redundancy
- FEX, UCS, 3rd-party blade, and direct attach server models supported
- Nexus 7009-M2 Leaf switch pair acting as Layer-3 border for the FabricPath domain
- No MAC learning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



Layer 3 Distributed Gateway at TOR

See Session on Dynamic Fabric Automation

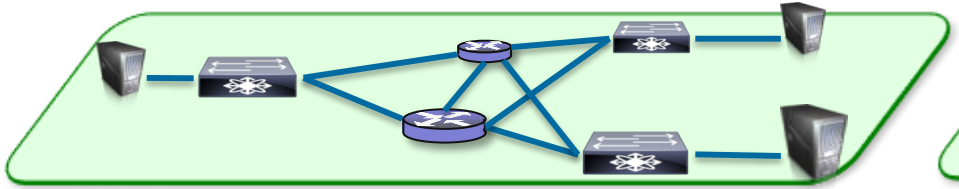
- Forwarding decision (Routing/Switching) as close as possible to the Workload
- Scale-Out Architecture
- Any Subnet, Anywhere

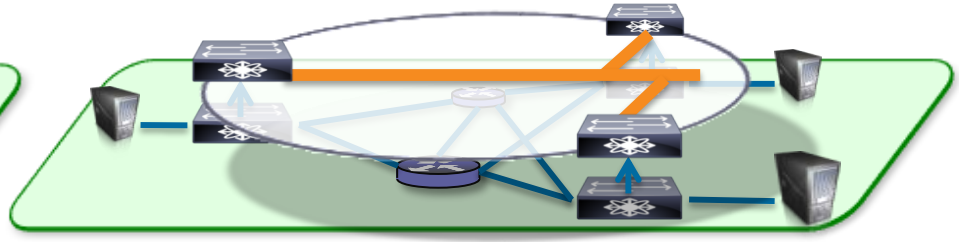


Overlays

Cisco live!

What about an Overlay?





Robust Underlay/Fabric

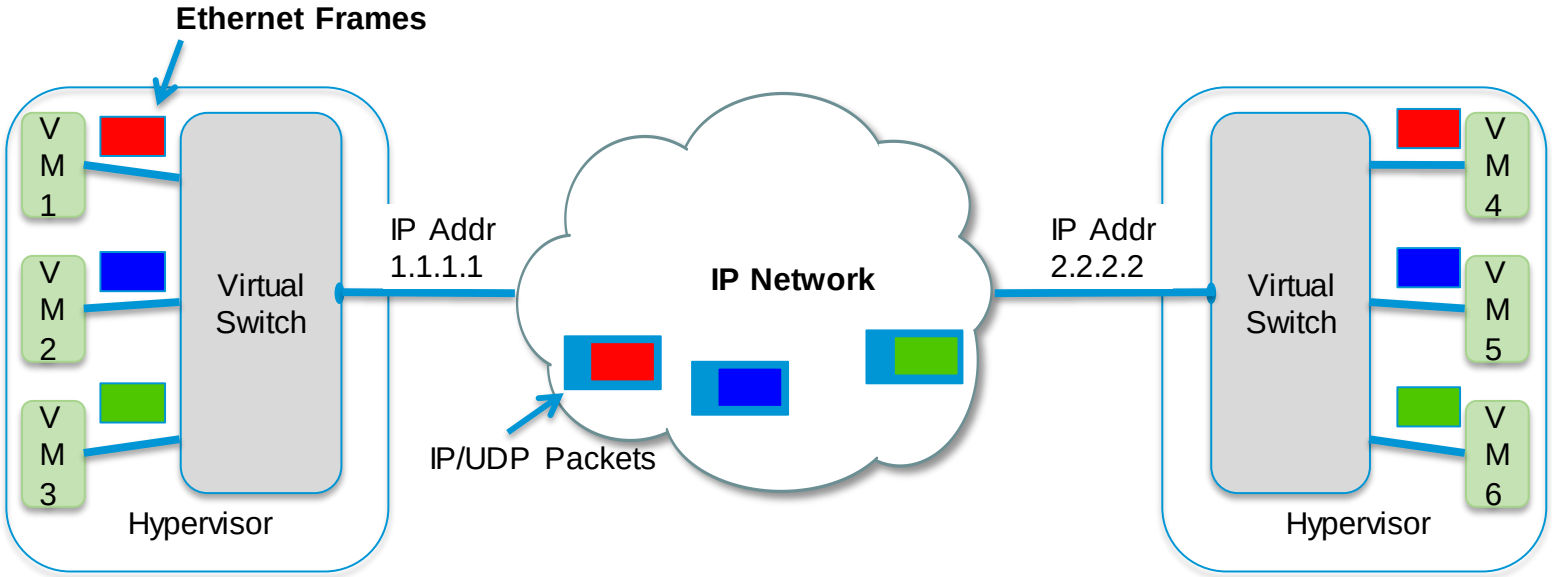
- High Capacity Resilient Fabric
- Intelligent Packet Handling
- Programmable & Manageable

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



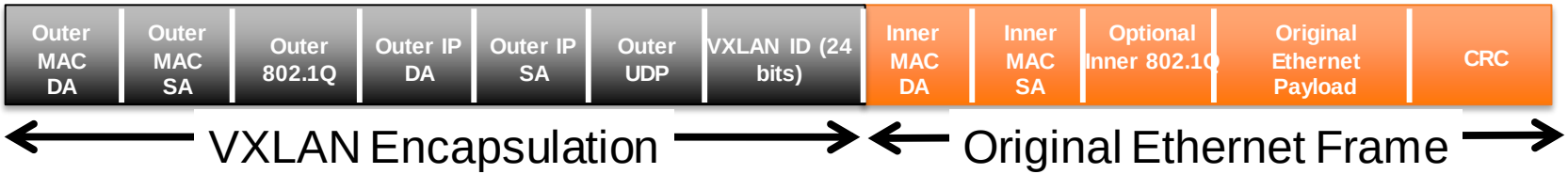
Virtual Overlay Encapsulations and Forwarding

- 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)
- Two main Hypervisor based Overlays
  - VXLAN Virtual Extensible Local Area Newtork
  - NVGRE, Network Virtualisation Generic Router Encapsulation
- Network Based Overlays
  - OTV, Overlay Transport Virtualisation
  - VPLS, EVPN
  - 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

For Your Reference

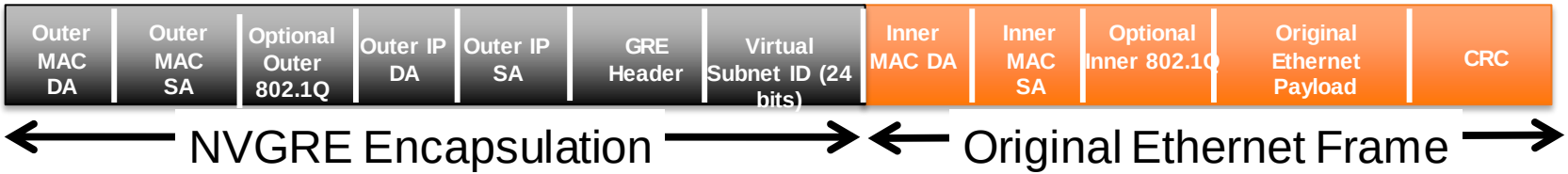
- https://datatracker.ietf.org/doc/draft-sridharan-virtualization-nvgre/
- Generic Routing Encapsulation (GRE) header for Network Virtualisation (NVGRE) in multi-tenant data centres
- 24 Bit Segment ID
- NVGRE Encapsulation 42 bytes
- Port Channel Load Distribution will be polarised
  - Most current 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

For Your Reference

- 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

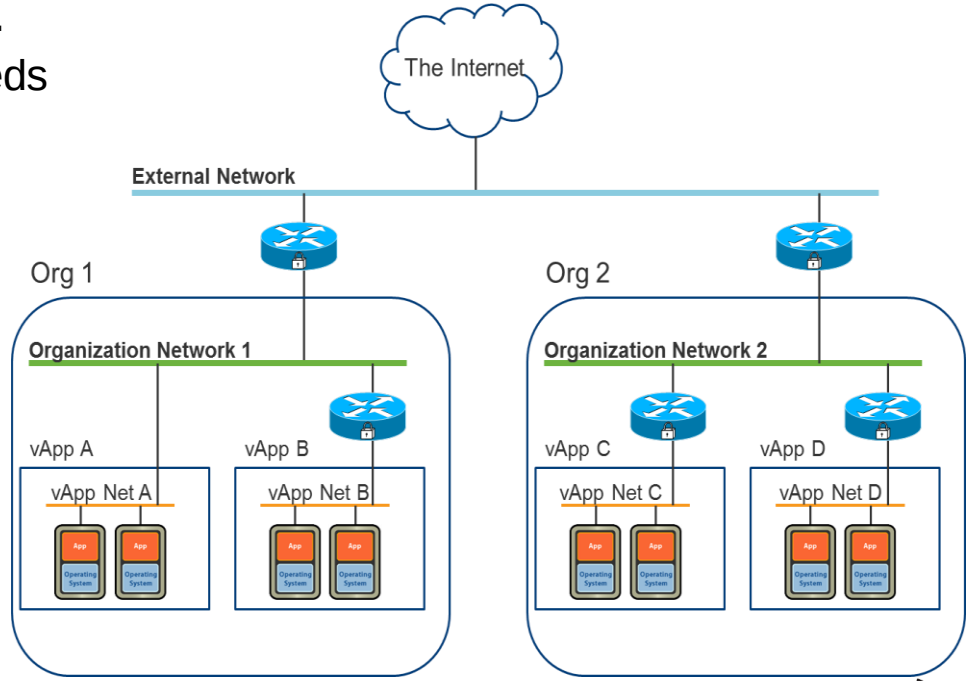
- Tunnel between End Points
  - VMs do NOT see NVGRE Encapsulation Hypervisor removes.
- IP multicast used for L2 broadcast/multicast, unknown unicast
- Technology submitted to IETF for standardisation
  - With Microsoft, Intel, Broadcom and Others



Multi-Tenancy and vApps Drive Layer 2 Segments

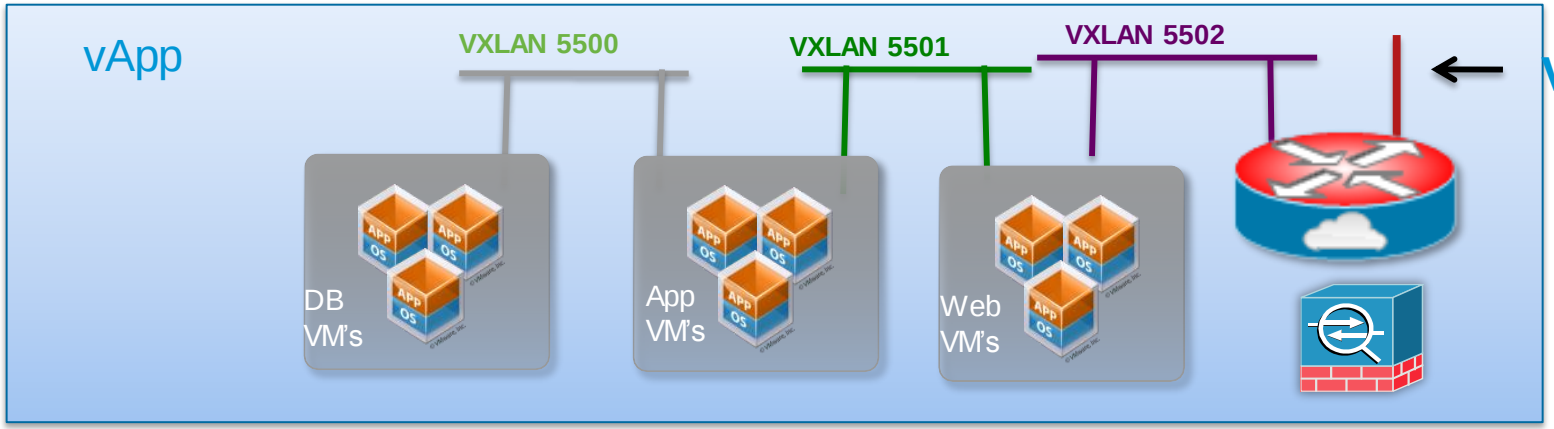
- Both MAC and IP addresses could overlap between two tenants, or even within the same tenant in different vApps.
  - Each overlapping address space needs
  - a separate segment

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



What is a vApp?

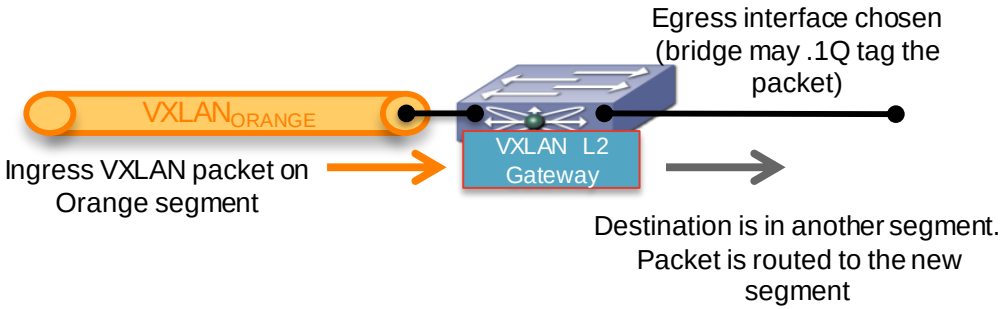
- A Cloud Provider using vCloud Director offers catalogs of vApps to their Users
- When cloned, new vApps retain the same MAC and IP addresses
- Duplicate MACs within different vApps requires L2 isolation
- Duplicate IP addresses requires L2/L3 isolation (NAT of externally facing IP addresses)
- Usage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



VXLAN L2 and L3 Gateways

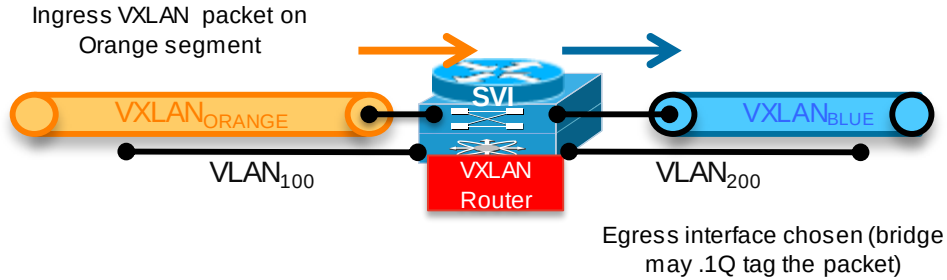
Connecting VXLAN to the broader network

L2 Gateway: VXLAN to VLAN Bridging



L3 Gateway: VXLAN to X Routing

- VXLAN
- VLAN



VXLAN Gateway Functionality

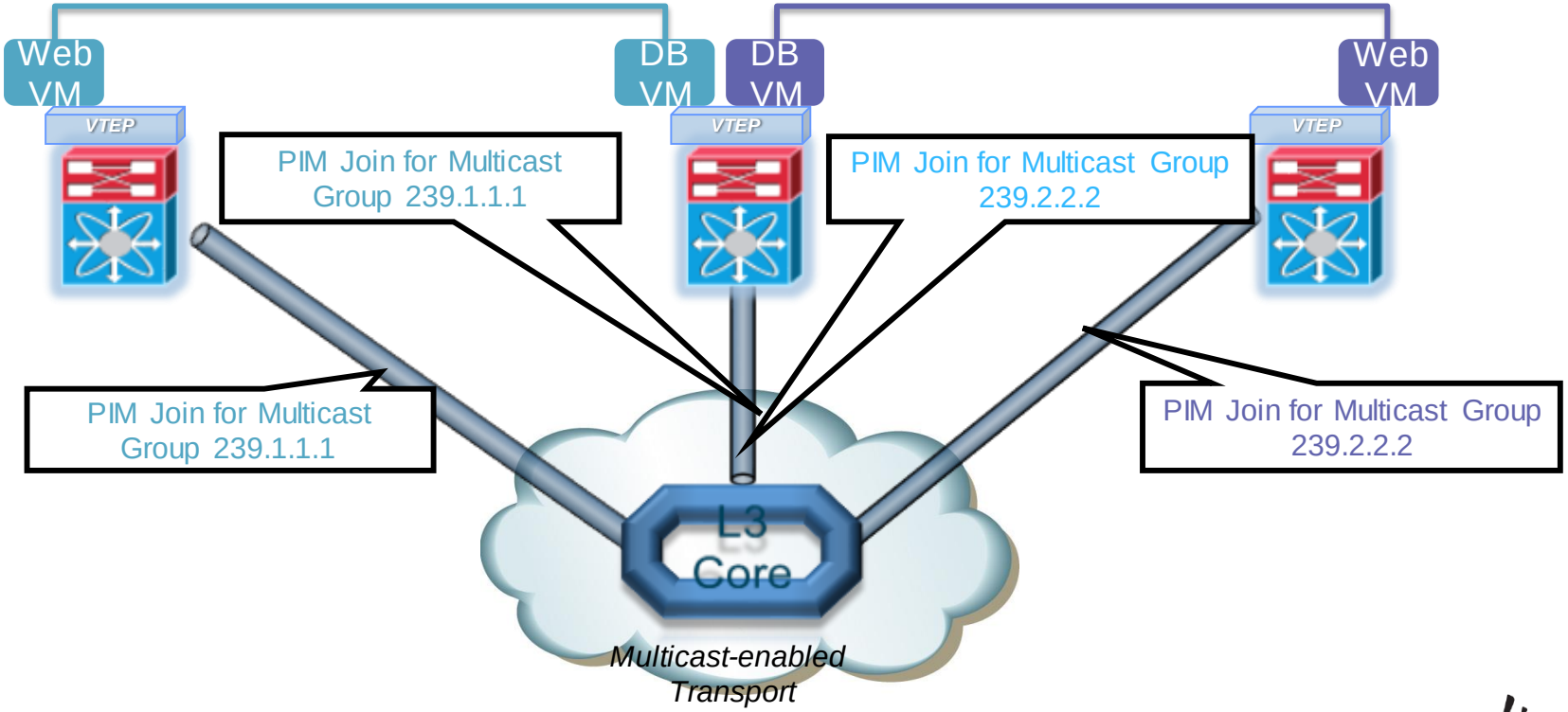
For Your Reference

PLATFORM	VXLAN Bridging and/or VXLAN Routing)	Starting Release	PLATFORM	VXLAN Bridging and/or VXLAN Routing)	Starting Release
DATA CENTRE			**ENTERPRISE NETWORKING**		
Nexus 1000v	Yes: Bridging and Routing	4.2(1)SV1(5.1) (MCast 5.2(1)SV3 (BGP CP)	**ASR 1K**	Yes **Bridging only**	IOS XE 3.13S (Bridging)
Nexus 3100	Yes **Bridging Only**	NX-OS 6.0(2)U3(2)	**ASR 9K**	Yes **(Routing and Bridging)**	IOS XR 5.2.0 (Bridging and Routing)
Nexus 5600	Yes (Bridging and Routing)	NX-OS 7.1(0)N1(1a)			
Nexus 9300 (Standalone)	VXLAN Bridging VXLAN Routing	6.1.(2)I2 7.0(3)I1(1)			

Cisco live!

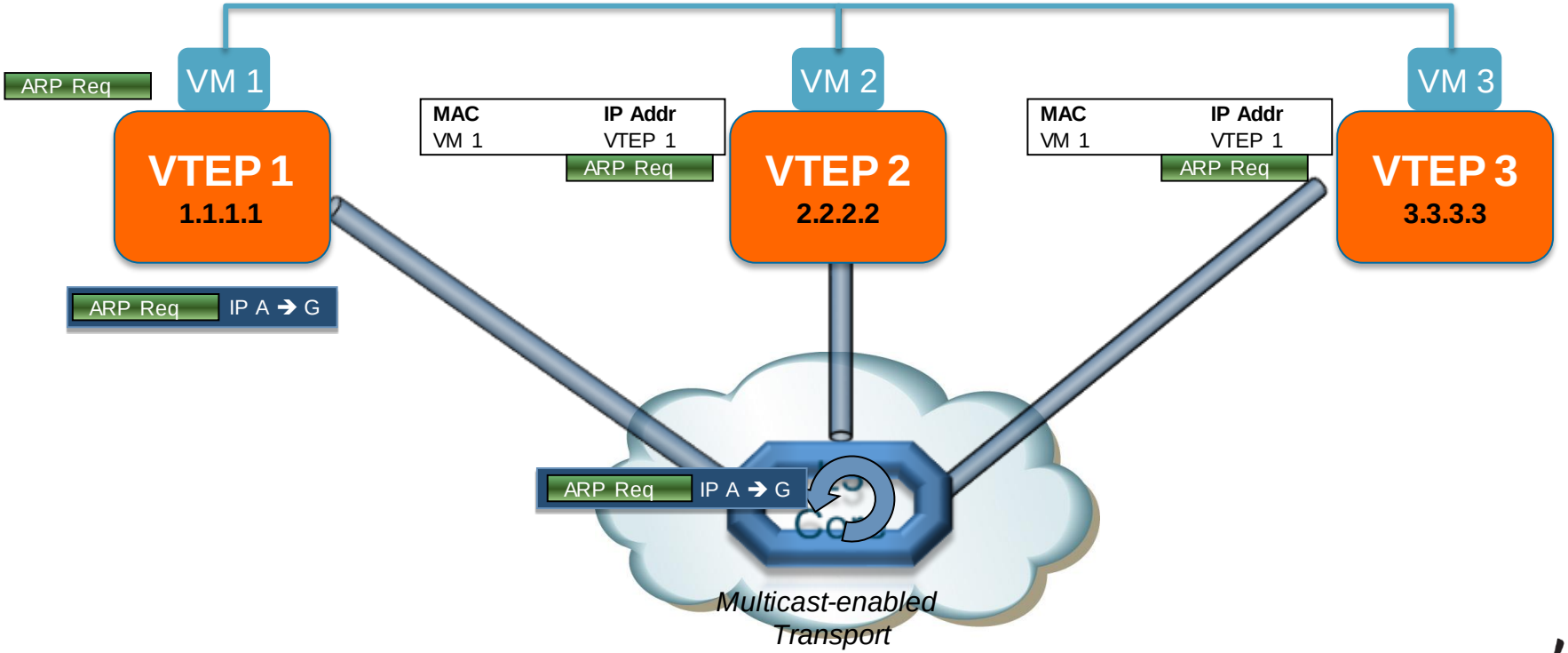
Data Plane Learning

Dedicated Multicast Distribution Tree per VNI



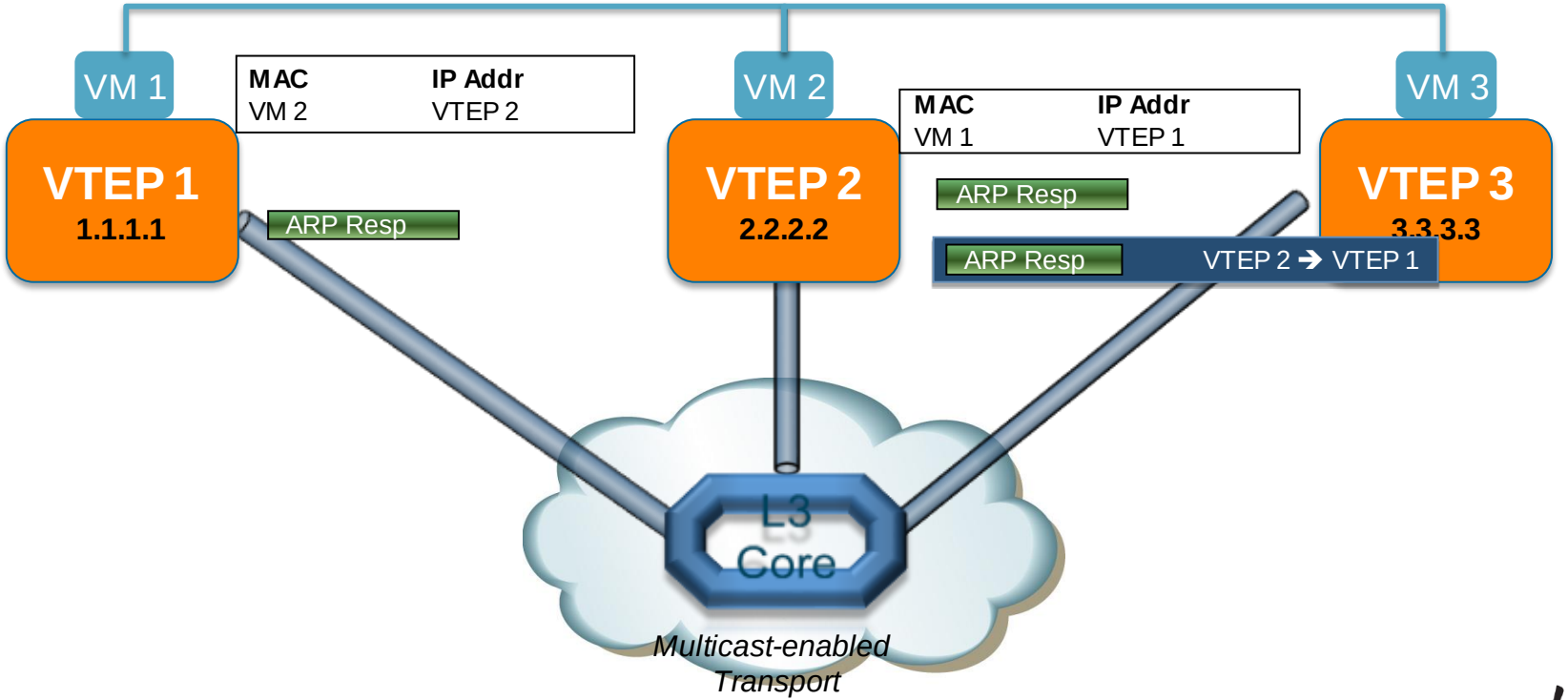
Data Plane Learning

Learning on Broadcast Source - ARP Request Example



Data Plane Learning

Learning on Unicast Source - ARP Response Example



VXLAN Configuration – Mapping VLANs to VNIs

Layer 2 Gateway on Multicast Enabled Fabric

```
feature vn-segment-vlan
feature nv overlay

Vlan 102
  vn-segment 10102

interface nve1
  no shutdown
  source-interface loopback1
  member vni 10102 mcast-group 239.1.1.102

interface <phy if>
   switchport mode access
   switch port access vlan 102
```

VXLAN Identifier

Tunnel Interface

Used for the VTEP

IP Multicast Group for Multi-Destination Traffic

Locally Significant VLAN

Cisco live!

VXLAN Configuration – Mapping VLANs to VNIs

Layer 3 Gateway

```
feature vn-segment-vlan
feature nv overlay
feature interface-vlan
Feature pim

Vlan 102
  vn-segment 10102

interface loopback 1
  ip addres 10.1.1.1/32
  ip router ospf 1 area 0.0.0.0
  ip pim sparse-mode

interface nve1
  no shutdown
  source-interface loopback1
  member vni 10102 mcast-group 239.1.1.102
```

VXLAN Identifier

Tunnel Interface

Used for the VTEP

IP Multicast Group for Multi-Destination Traffic

VXLAN Config – Mapping VNIs to VLAN Interfaces

Layer 3 Gateway Continued…

```
feature pim
route-map SPINE permit 10
  match ip multicast group 239.1.1.0/24

ip pim rp-address 10.10.10.50 route-map SPINE
ip pim ssm range 232.0.0.0/8
```

Enabling PIM

Refer to Slide 57

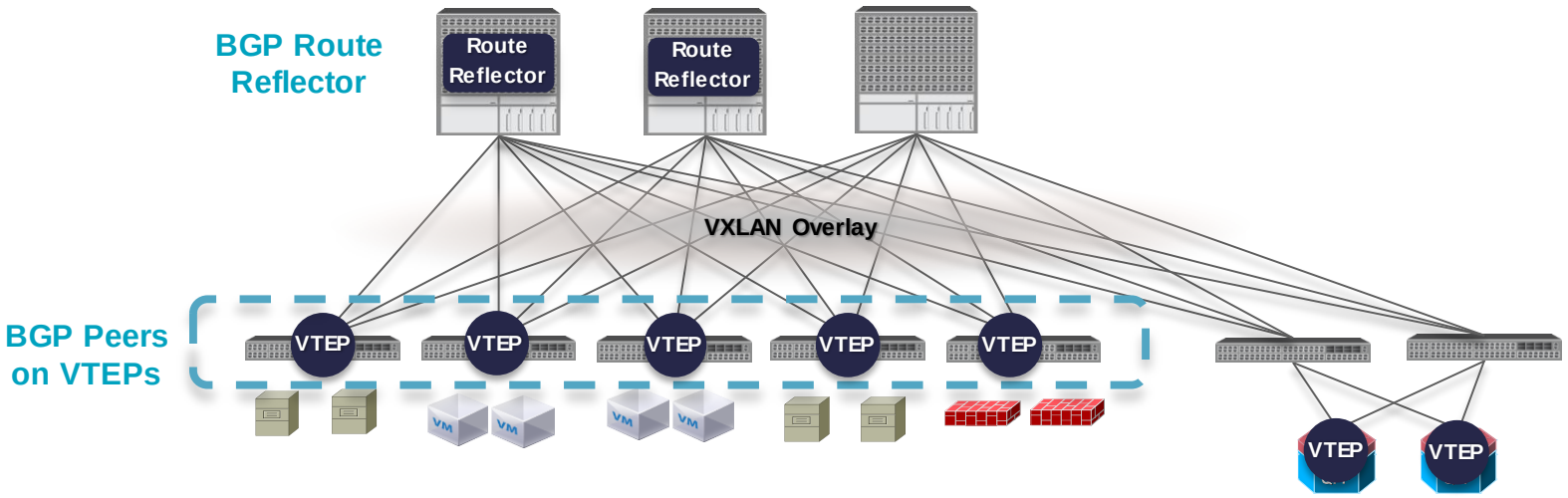
```
interface Ethernet2/1
  description Uplink to Core Eth1/1
  no switchport
  ip address 192.168.1.1/30
  ip ospf network point-to-point
  ip router ospf 1 area 0.0.0.0
  ip pim sparse-mode
  no shutdown
```

Routed Uplink to Spine

```
Interface Vlan 102
  ip address 10.2.2.1/24
  ip router ospf 1 area 0.0.0.0
  ip pim sparse-mode
```

VXLAN L3 GW Network

VXLAN Evolution - BGP EVPN Control Plane

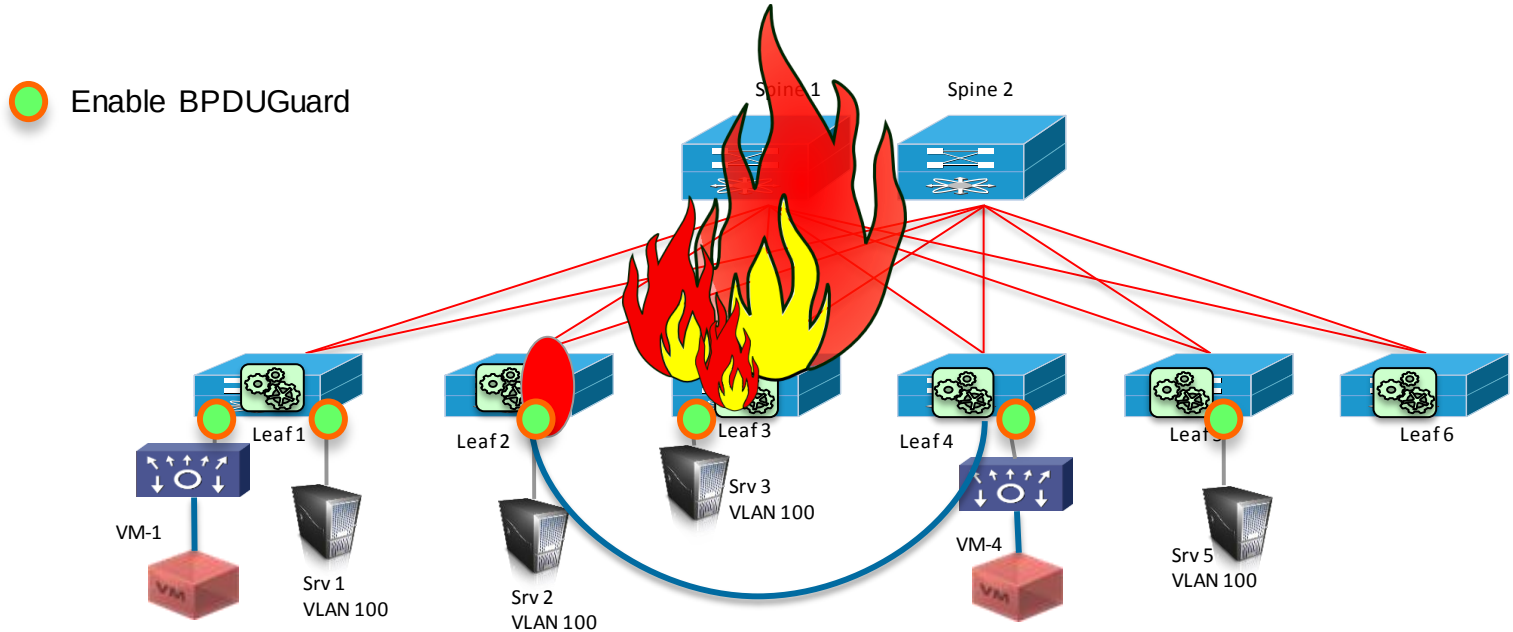


Uses Multi-Protocol BGP w EVPN Address Family for Dynamic Tunnel Discovery and Host reachability

Supported across the product line: Nexus and ASR

VXLAN and Layer 2 Loop Avoidance

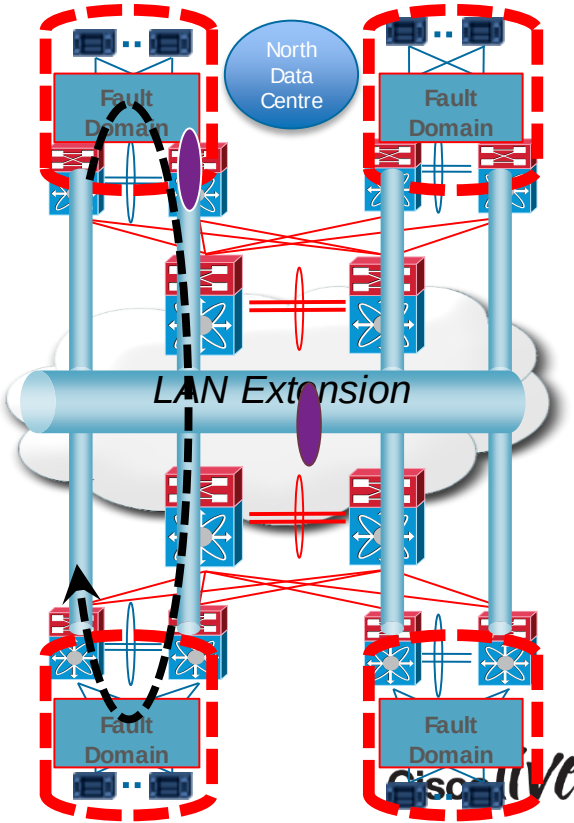
- VXLAN doesn't implement a native L2 loop detection and protection
- BPDU's are not forwarded across the VXLAN domain
- A backdoor link can be established between two or more TORs



L2 Network Overlays for Data Centre Interconnect

OTV/VPLS

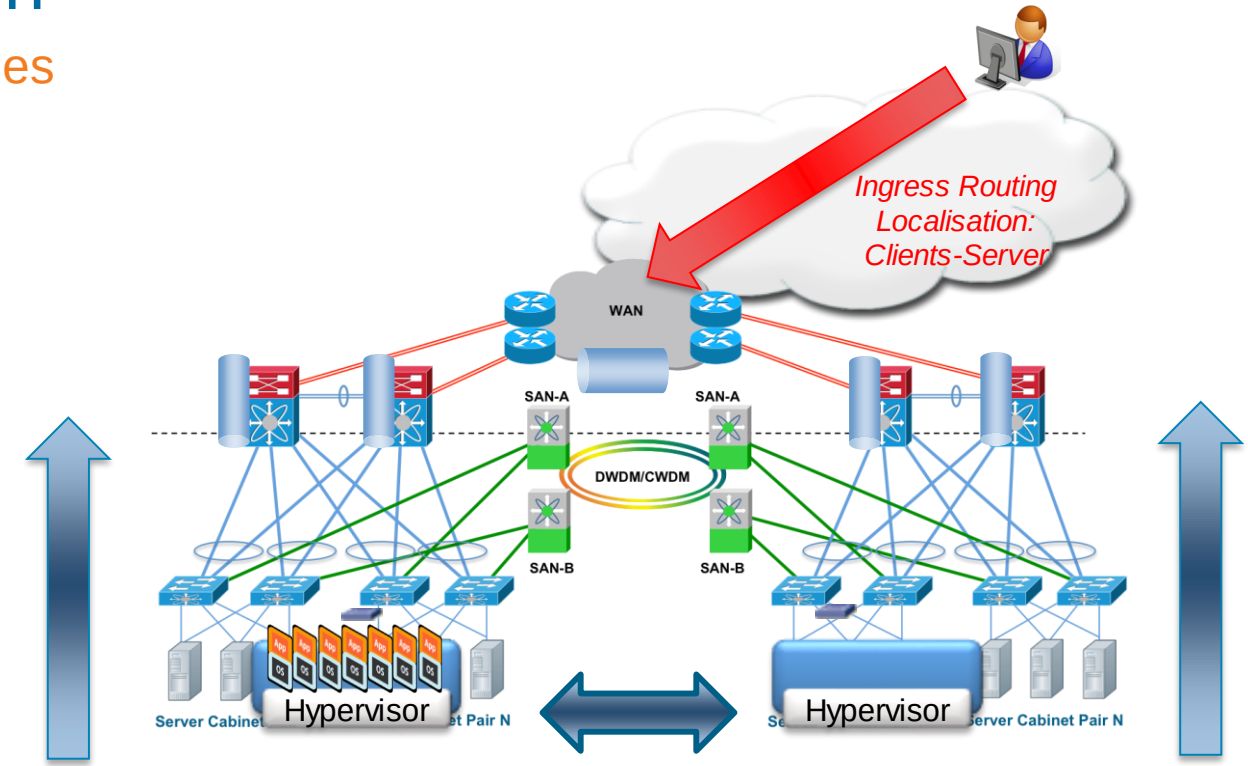
- OTV/VPLS resilient geo-extension of segments
- Preserve failure isolation between locations
- Network resiliency and multi-pathing
- Built in loop handling
- Optimal traffic handling
- Streamlined operations
- Egress routing optimisation
- HW Accelerated high performance connectivity



Path Optimisation

Optimal Routing Challenges

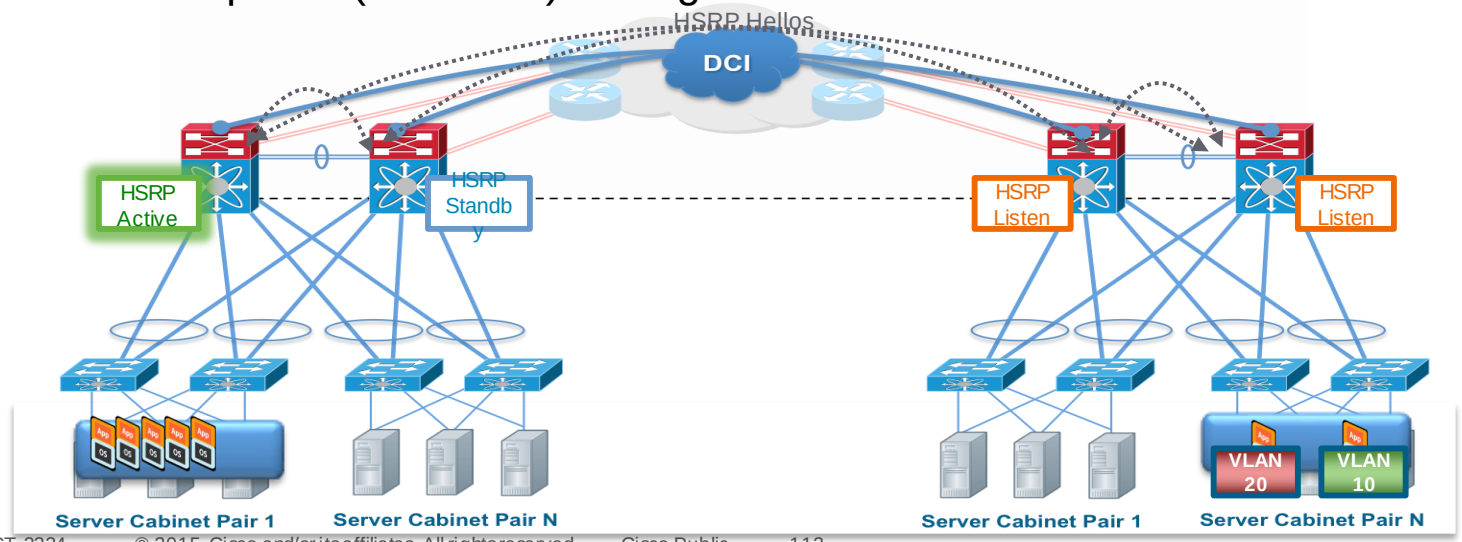
- Layer 2 extensions represent a challenge for optimal routing
- Challenging placement of gateway and advertisement of routing prefix/subnet



Path Optimisation

Egress Routing with LAN Extension

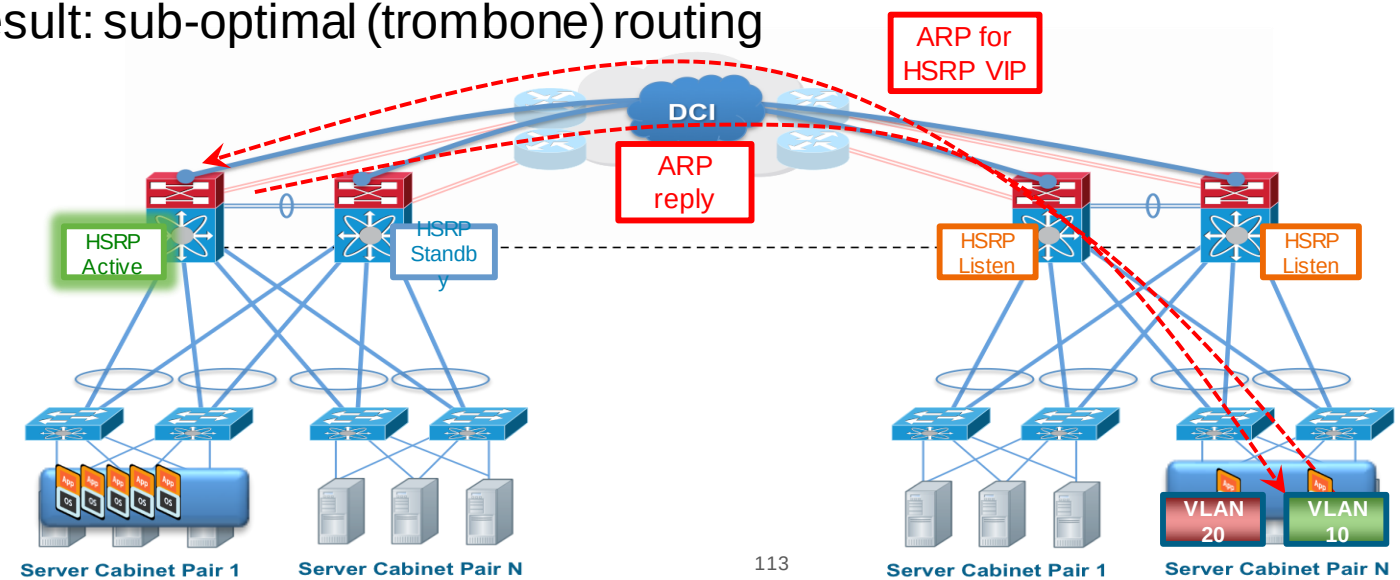
- Extended VLANs typically have associated HSRP groups
- By default, only one HSRP router elected active, with all servers pointing to HSRP VIP as default gateway
- Result: sub-optimal (trombone) routing



Path Optimisation

Egress Routing with LAN Extension

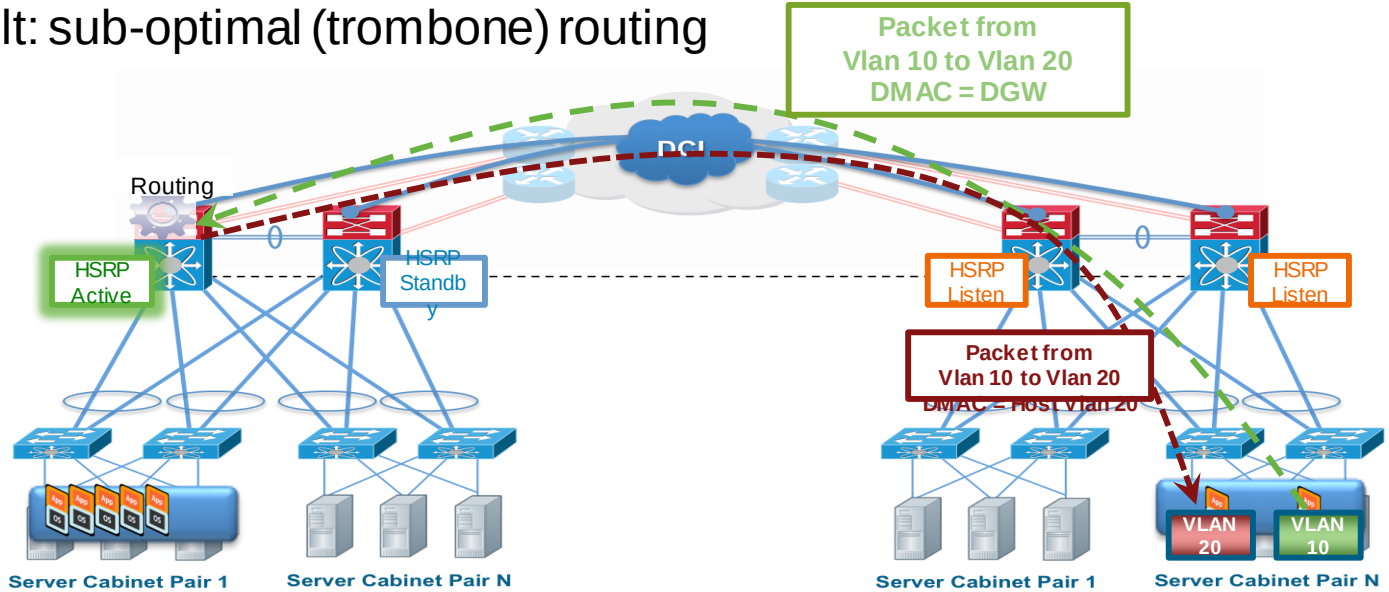
- Extended VLANs typically have associated HSRP groups
- By default, only one HSRP router elected active, with all servers pointing to HSRP VIP as default gateway
- Result: sub-optimal (trombone) routing



Path Optimisation

Egress Routing with LAN Extension

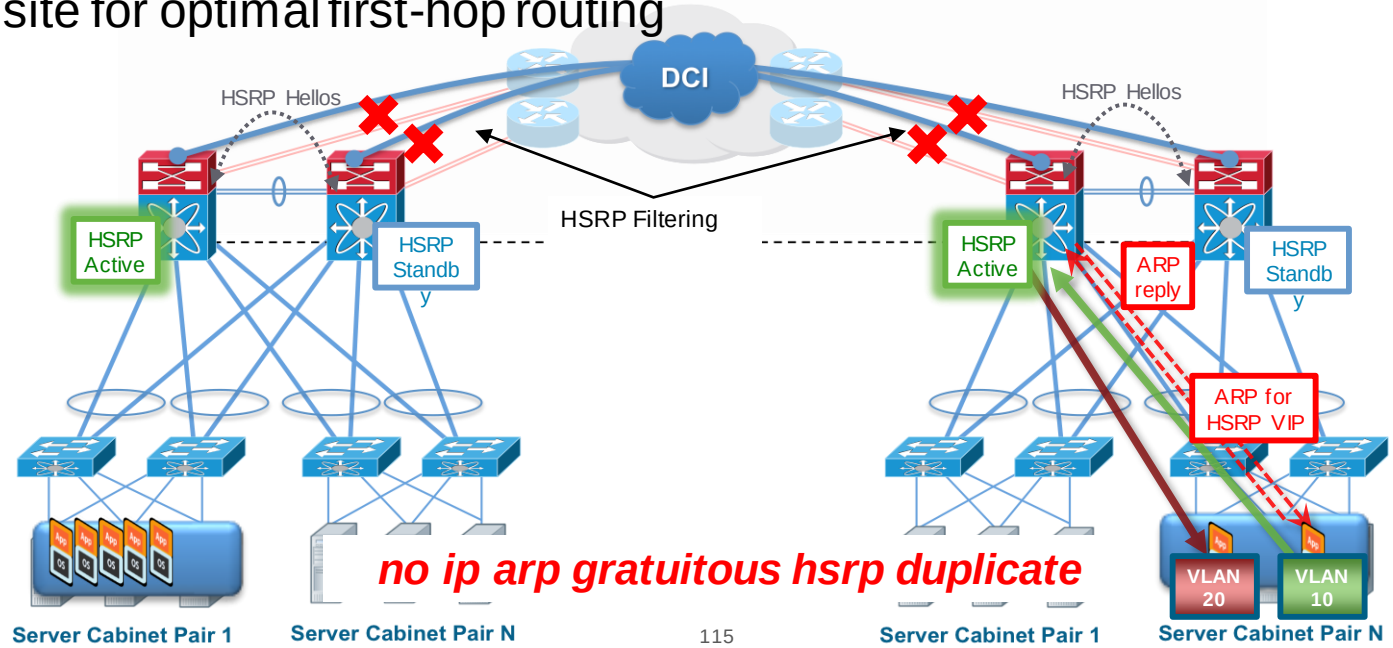
- Extended VLANs typically have associated HSRP groups
- By default, only one HSRP router elected active, with all servers pointing to HSRP VIP as default gateway
- Result: sub-optimal (trombone) routing



Egress Routing Localisation

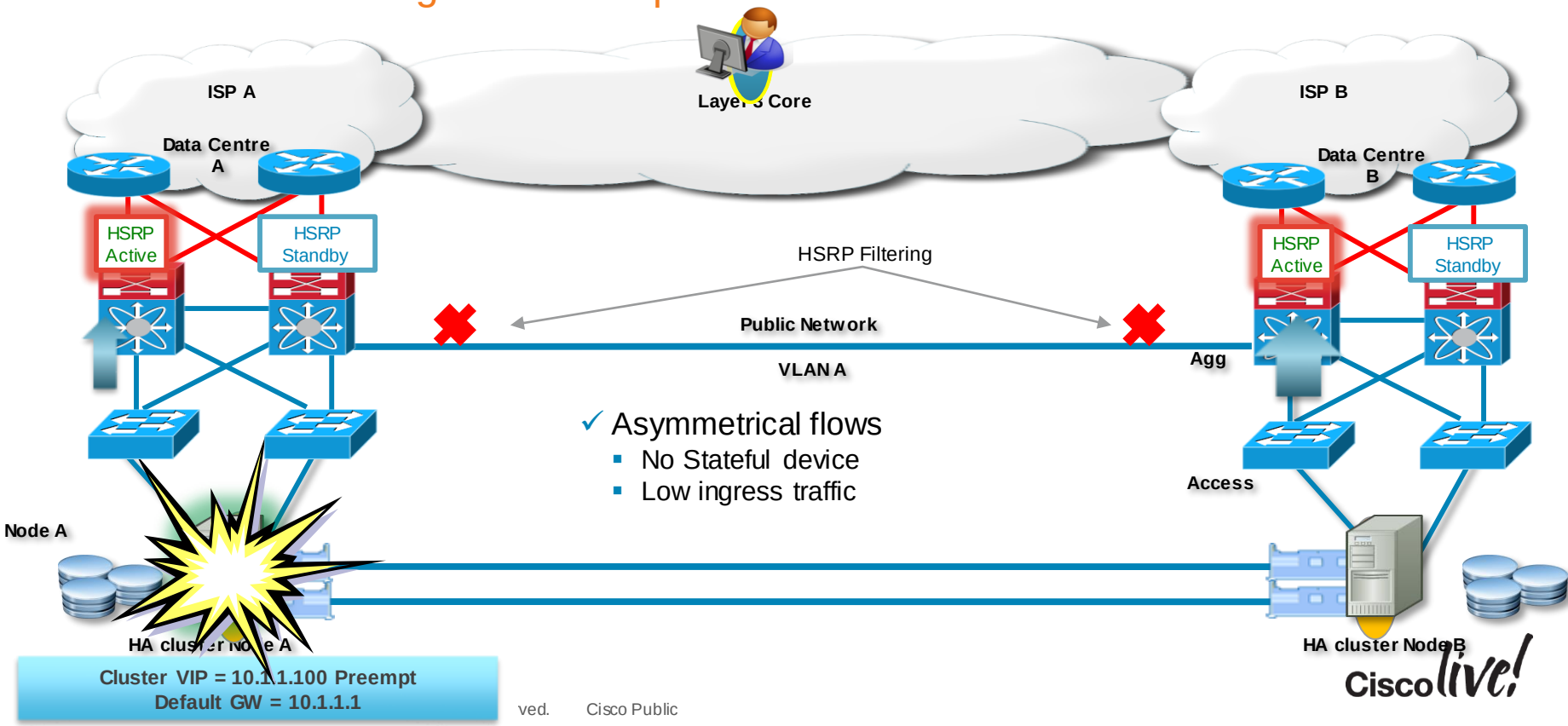
FHRP Filtering Solution

- 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



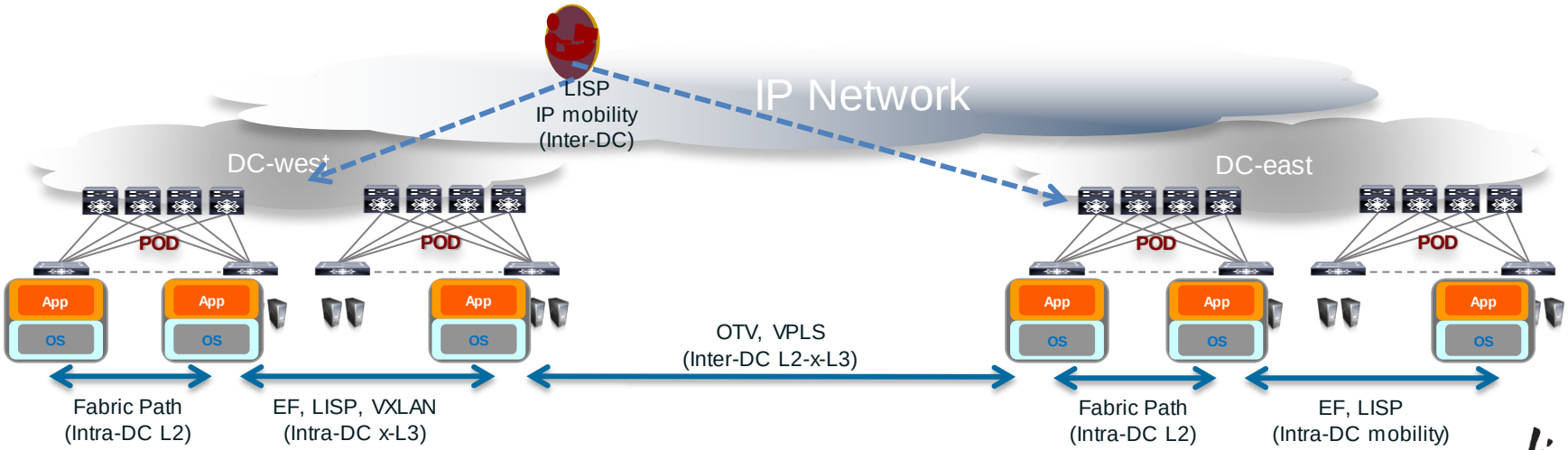
Sample Cluster - Primary Service in Left DC

FHRP Localisation – Egress Path Optimisation

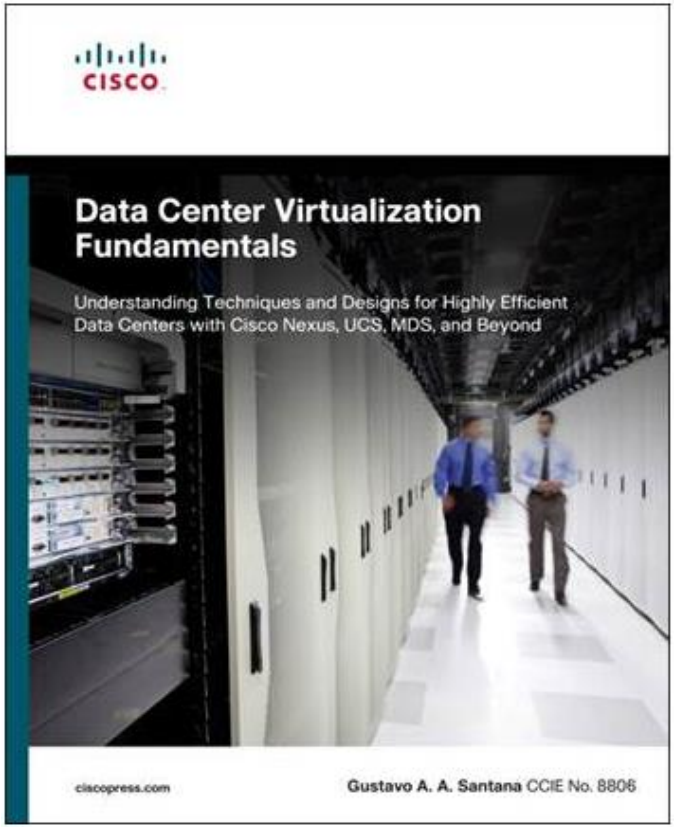


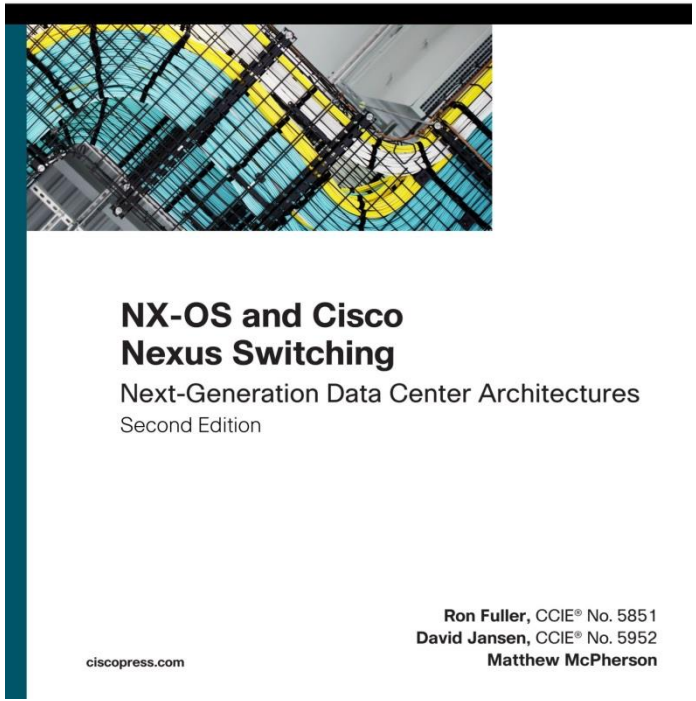
Technologies Intra-DC and Inter-DC

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





Continue Your Education

- Data Centre Breakout Sessions
- Demos in the World of Solutions
- Walk-in Self-Paced Labs
- Meet the Expert 1:1 meetings
- DevNet Programme


Q & A
Cisco live!

Complete Your Online Session Evaluation

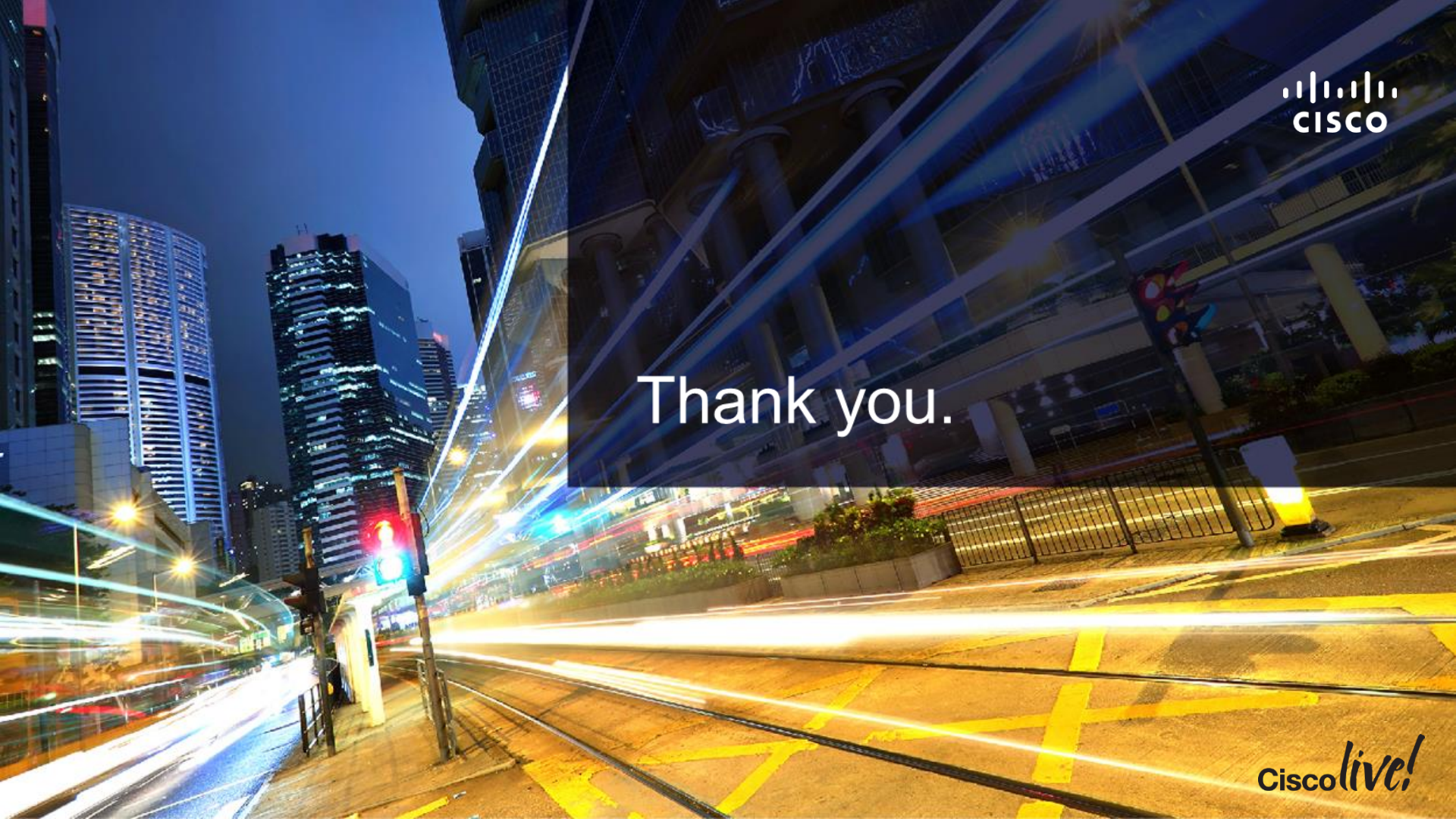
Give us your feedback and receive a Cisco Live 2015 T-Shirt!

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 http://showcase.genie-connect.com/clmelbourne2015
- Visit any Cisco Live Internet Station located throughout the venue

T-Shirts can be collected in the World of Solutions on Friday 20 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


Thank you.

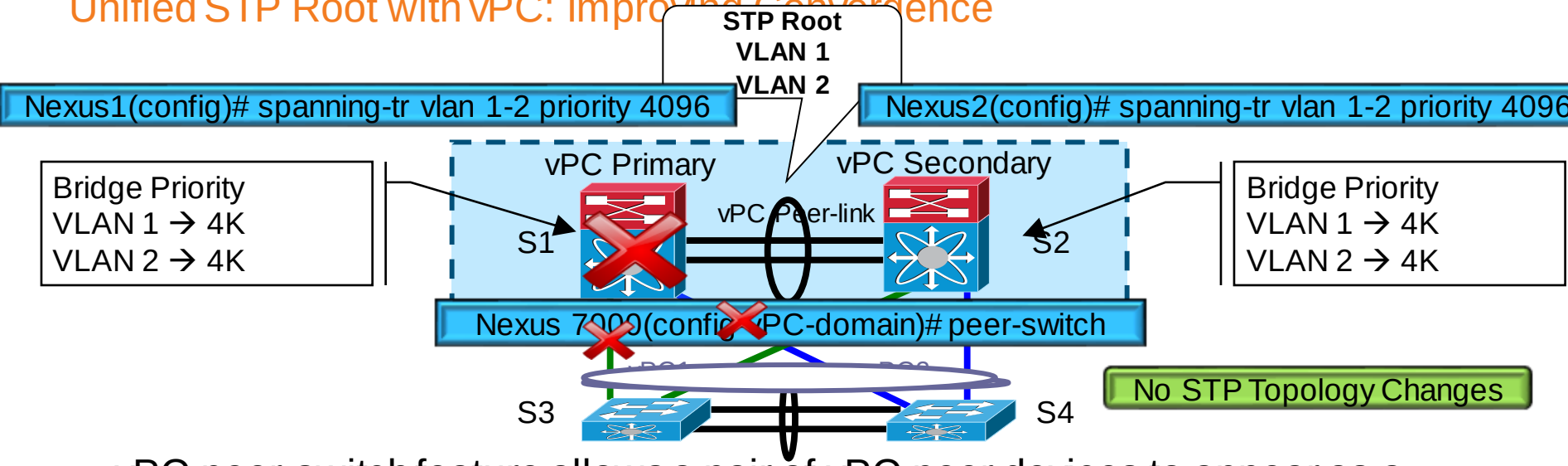

CISCO

Backup Slides

Cisco live!

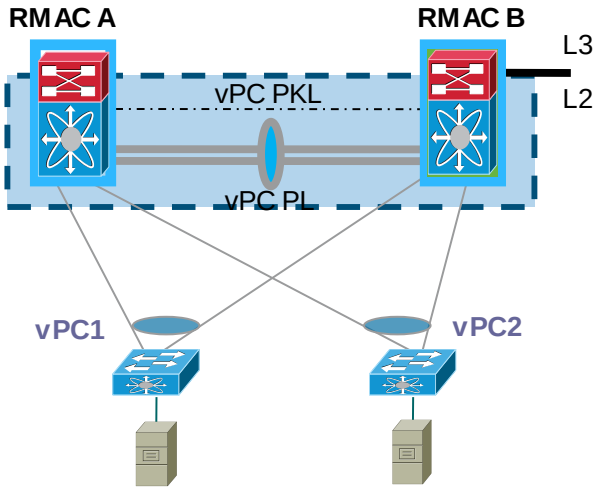
vPC Peer-switch

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



- 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

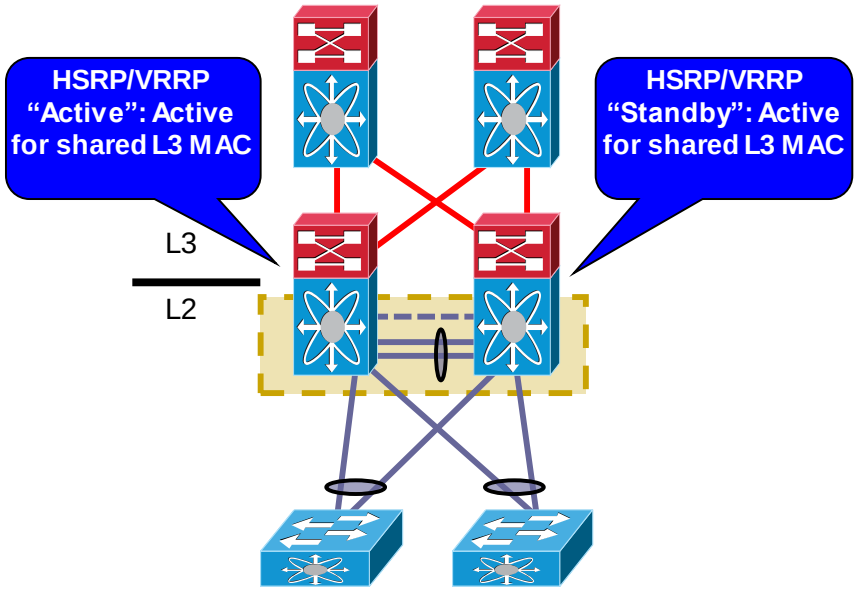
  Disable IP redirects on all SVIs associated with vPC VLANs (Default with NX-OS 5.1)

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

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

See notes section for more information

HSRP with vPC

FHRP Active Active



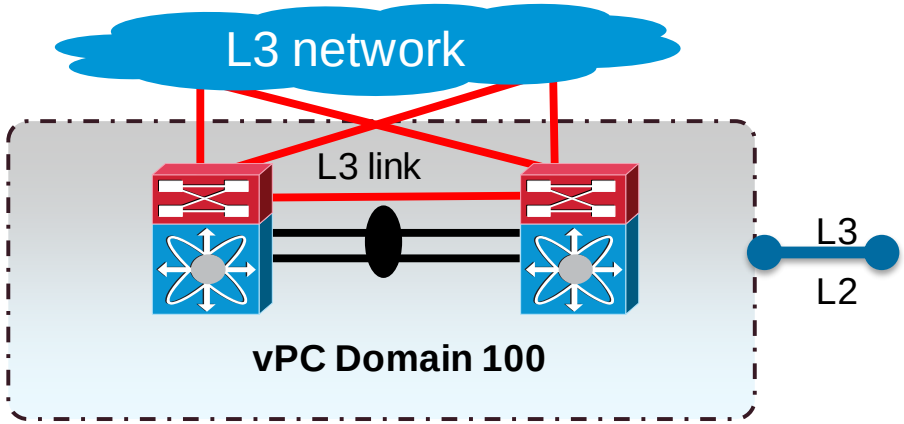
- Support for HSRP/VRRP protocols in Active/Active mode with vPC
  - HSRP or VRRP operate in Active/Active mode from data plane standpoint
  - HSRP or VRRP operate in Active/Standby mode from control plane standpoint (Active instance responds to ARP requests)
- Recommendations:
  - Do not tune HSRP timers (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

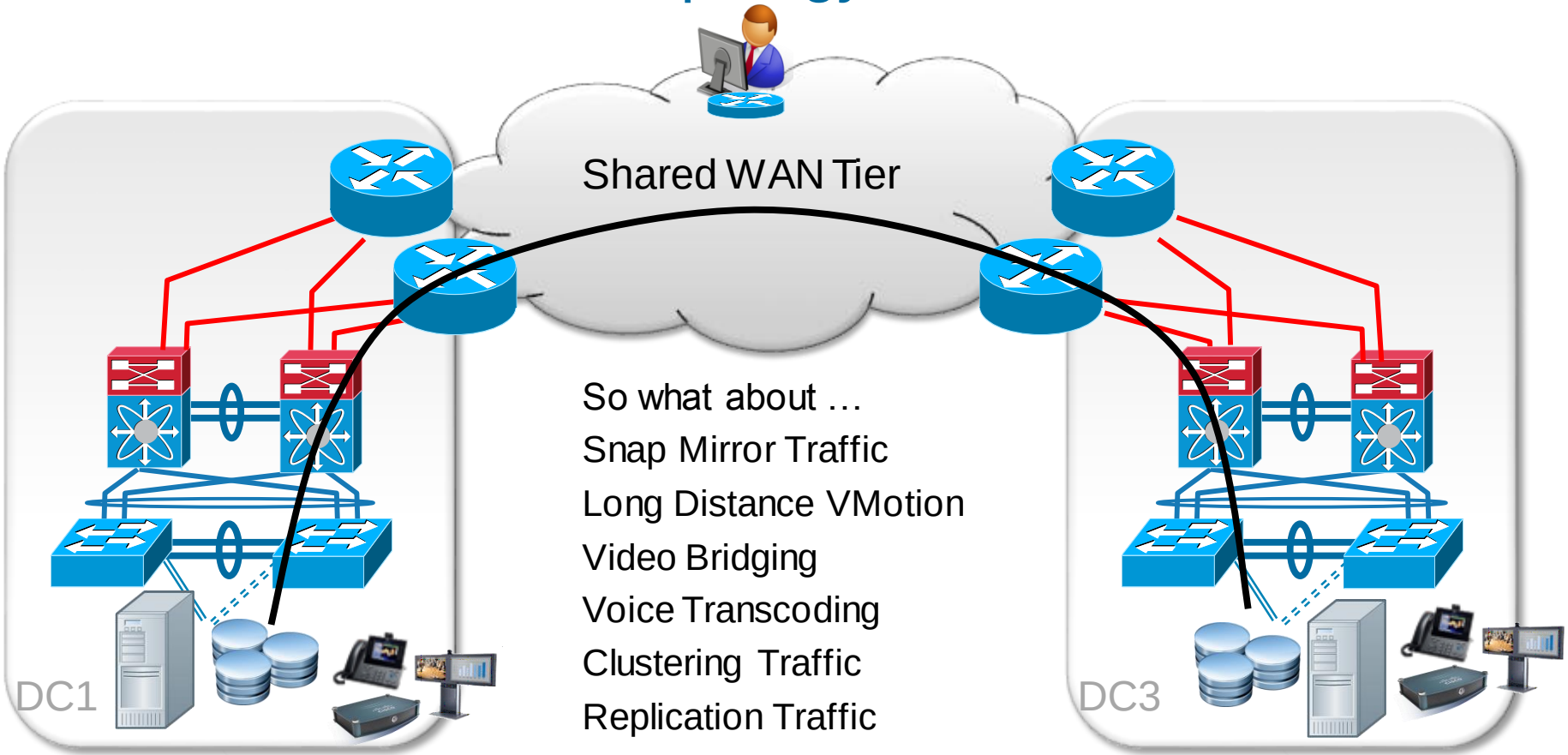
- Peering between two N7k for alternative path in case uplinks fail
- Recommend to have dedicated L3 interface and run routing protocol over L3 interconnect
- Alternately can use SVI over L2 link or vPC as alternate secondary option.
- Unique vPC Domain ID per pair.
  - vPC Domain ID is used at the vPC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



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!

Dual DC Reference Topology



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

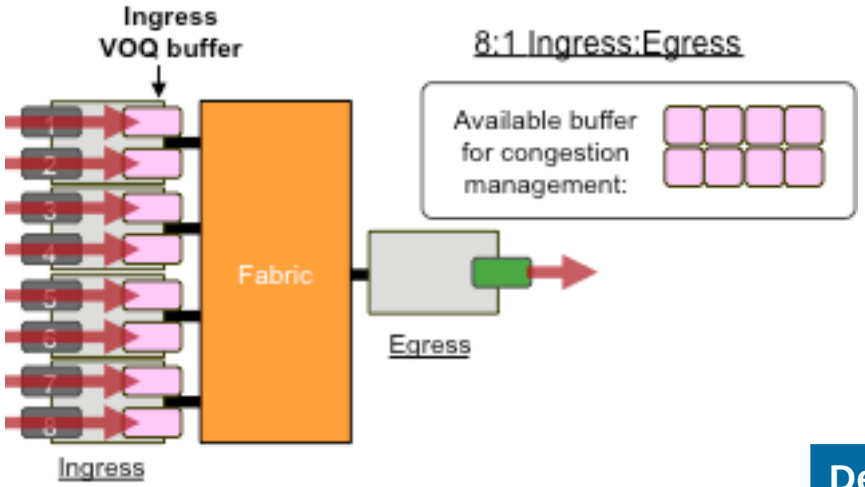
Users can notice a single packet lost in 10,000—Making HD Video ***One Hundred Times More Sensitive to Packet Loss than VoIP!***

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

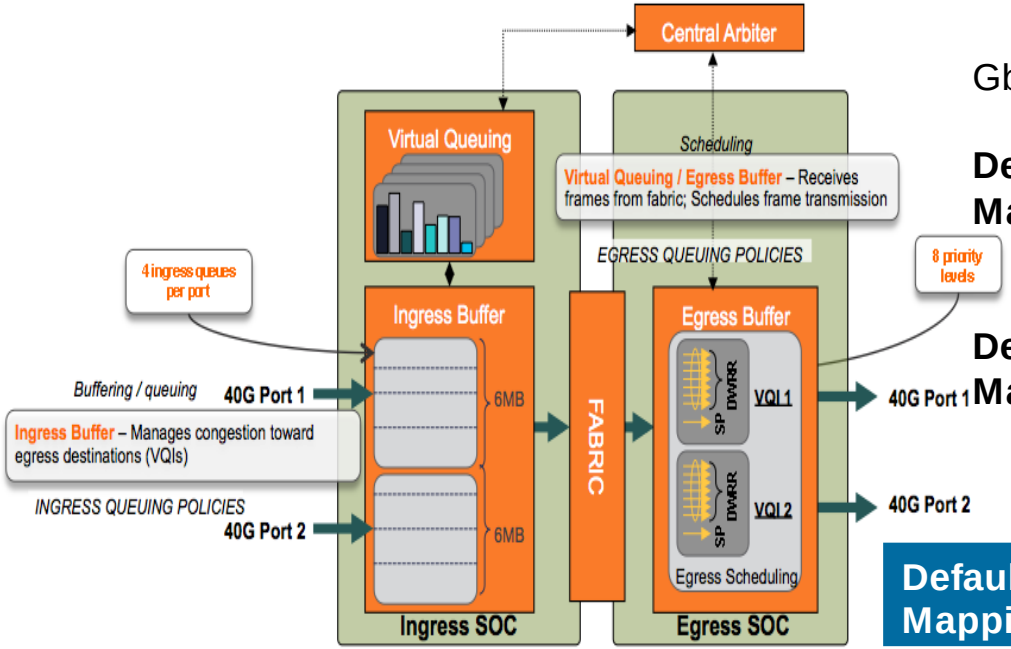
Total Per-Port Buffer (1:1): 1.5 MB
Total Per-Port Buffer (8:1): 12MB

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_Centre/MSDC/1.0/MSDC1_C.html#wp1120251

NEXUS F3 Module Buffer Structure 40G Port



Gbps Line Rate: 10 Gbps = 5,000 MB/s
or 5,000 KB/ms

Default Per-Port Buffer : 6 MB
Max Per-Port Buffer : 12MB

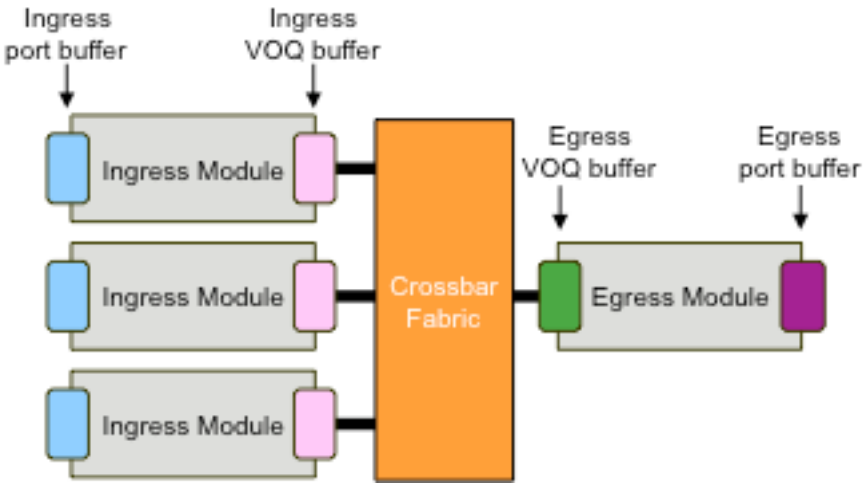
Default Per Port Buffering Capacity (6 MB): ~1.2 ms
Max Per Port Buffering Capacity (12 MB) : ~2.4 ms

Default Queue Mapping	COS Values	Buffer Allocated
Queue 0	COS 0 to COS 4	88% 1.35 MB
Queue 1	COS 5 to COS 7	10% 0.15 MB

Nexus 7000 F3 12 Port Module 72 MB VOQ Buffer

Nexus 7700 F3 24 Port Module 144 MB VOQ Buffer

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

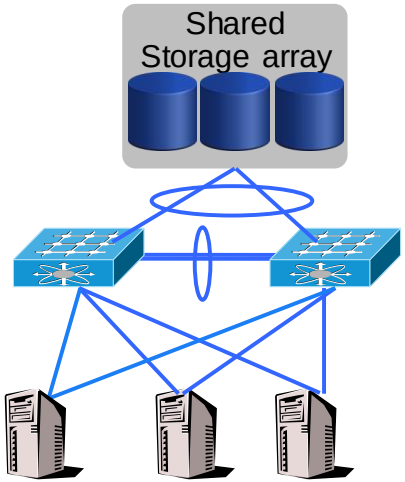
Per Port Ingress Buffer: 5.2 MB
Queue 0 default Buffer 2.6 MB
Queue 1 default Buffer 2.6 MB
Per Port Ingress VoQ Buffer: 4.5 MB
Total Ingress Per-Port Buffer: 9.7 MB

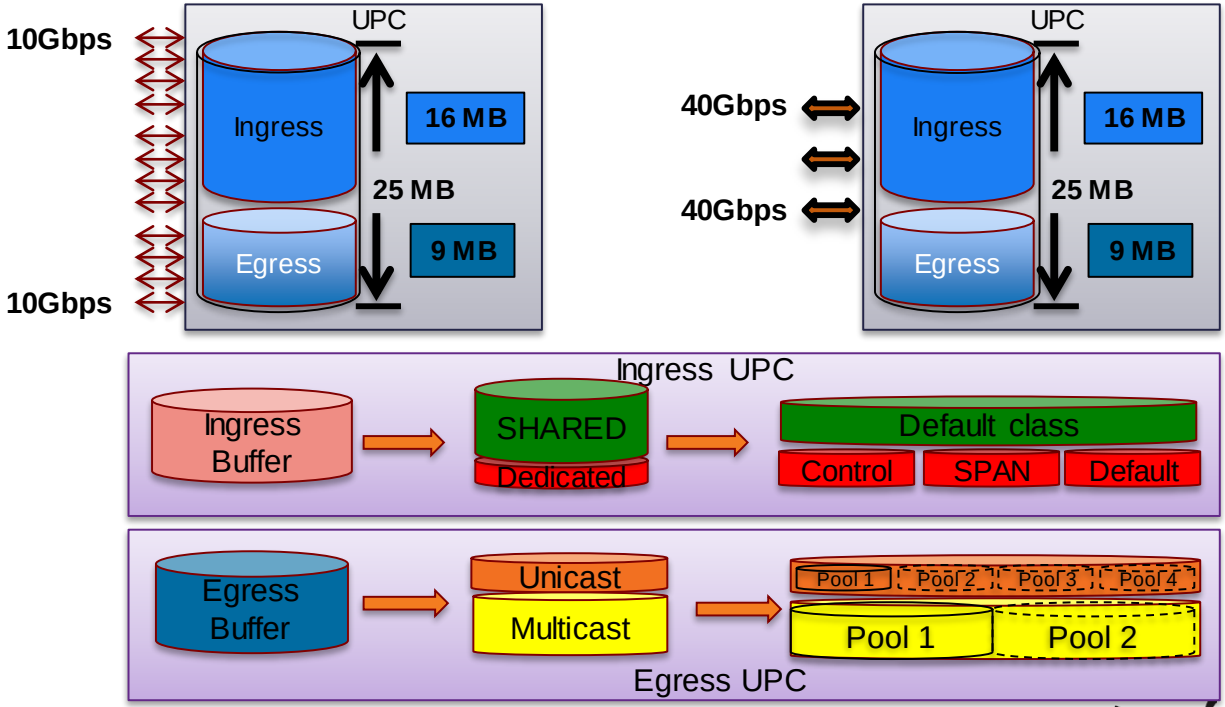
Per Port Egress Buffer: 5 MB
Per Port Egress VoQ Buffer: 380 Kb
Total Egress Per-Port Buffer: 5.3MB
Total Ingress+Egress Per-Port Buffer: 15MB

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 Queueing Model

Nexus 5600 Example





All Tenants use COS = 0

Notes on Changing Default QoS Configuration

- Queuing:
  - COS/DSCP-to-queue mappings (type queuing class-maps) have system-wide scope
  - If you change default COS/DSCP-to-queue mappings, make sure all interfaces in all VDCs have queuing policy applied that defines behaviour for all active queues
  - Queuing policies must include all available queues for a given direction (regardless of whether COS/DSCP values are mapped)
- QoS:
  - If you apply non-default QoS policy, default behaviour of implicit DSCP→COS mapping no longer applies for that interface


CISCO