What You Make Possible

TOMORROW
starts here.

Deployment Challenges with Interconnecting Data Centres

BRKDCT-3060

TOMORROW starts here.

Session: BRKDCT-3060 Abstract

Data Centre Networking: Deployment Challenges with Interconnecting Data Centres
This advanced session discusses the challenges and recommended solutions for extending LAN connectivity between geographically dispersed Data Centres. The Data Centre is now more and more spreading across multiple sites, and one very difficult point to solve is the extension of VLAN in a large scale with respect to the requirement for Spanning-Tree stability. The different requirements for providing a robust LAN extension solution will be discussed during this session, including end-to-end loop prevention, multi-homing considerations and optimal bandwidth utilisation. Detailed design guidance will be provided around the deployment of Ethernet based technologies, leveraging Multi Chassis EtherChannel functionalities like VSS and vPC, as well as MPLS based technologies (EoMPLS and VPLS) and an innovative IP based technology called Overlay Transport Virtualisation (OTV). Locator Identity Separation Protocol (LISP) will then be introduced as an emerging technology capable of providing both IP Mobility and Path Optimisation functionalities. This advanced session is intended for network design and operation engineers from Enterprises, Service Providers or Enterprise Hosting Service Providers that are willing to solve this difficult and controversial problem of Data-Centre Interconnect.

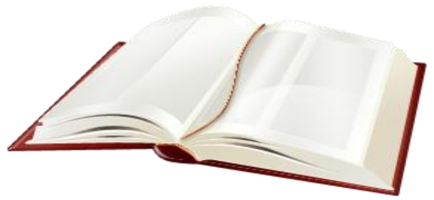
Goals of This Session…

- Highlighting the main business requirements driving Data Centre Interconnect (DCI) deployments
- Understand the functional components of the holistic Cisco DCI solutions
- Get a full knowledge of Cisco LAN extension technologies and associated deployment considerations
- Integrate routing aspect induced by the emerging application mobility offered by DCI
- This session does not include:
- Storage extension considerations associated to DCI deployments

Data Centre Interconnect

Agenda

- Mobility and Virtualisation in the Data Centre
- LAN Extension Deployment Scenarios
  - –Ethernet Based Solutions
  - –MPLS Based Solutions
    - EoMPLS
    - VPLS
    - A-VPLS
    - EVPN
  - –IP Based Solutions
- Encryption
- IP Mobility without LAN Extension
- Path optimisation
- VXLAN
- Summary and Conclusions
- Q&A



= For your Reference

Mobility and Virtualisation in the Data Centre

Distributed Data Centres

Building the Data Centre Cloud

Distributed Data Centre Goals

- Seamless workload mobility
- Distributed applications
- Pool and maximise global resources
- Business Continuity

Interconnect Challenges

- Complex operations
- Transport dependence
- IP subnets and mobility
- Failure containment

Geographically Disperse Data Centres

Connecting Virtualised Data Centres

Multi-tenancy/segmentation:

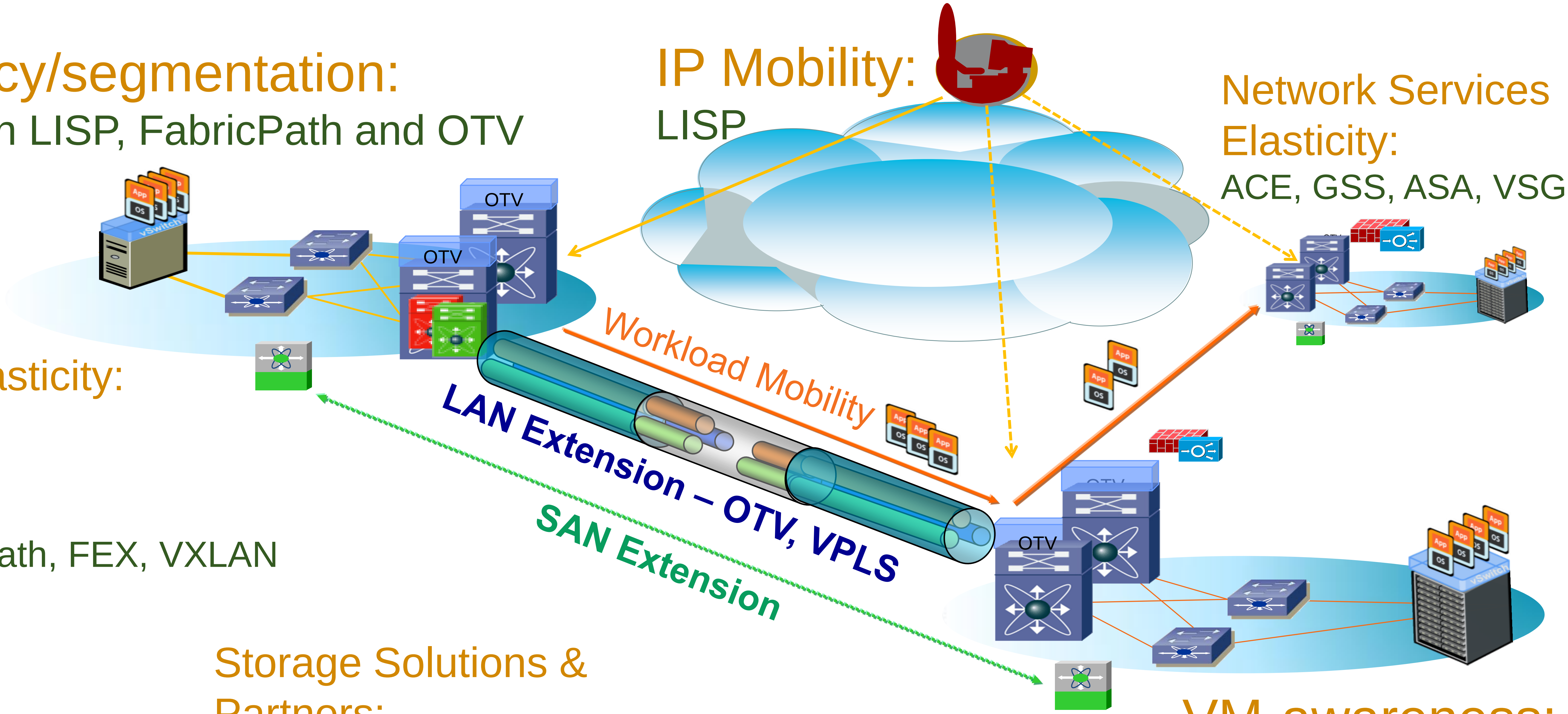
Segment-IDs in LISP, FabricPath and OTV

IP Mobility:

LISP

Network Services Elasticity:

ACE, GSS, ASA, VSG



L2 Domain Elasticity:

Inter-DC:

OTV/VPLS

Intra-DC:

vPC, FabricPath, FEX, VXLAN

Storage Solutions & Partners:

FCIP, Read/write Acceleration

EMC, NetApp

VM-awareness:

Port Profiles

Location of compute resources is transparent to the user

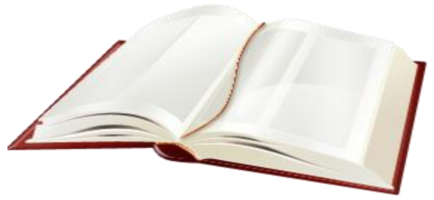
Layer 2 Use Cases

- Extending Operating System / File System clusters
- Extending Database clusters
- Virtual machine mobility
- Physical machine mobility
- Physical to Virtual (PtoV) Migrations
- Legacy devices/apps with embedded IP addressing
- Time to deployment and operational reasons
- Extend DC to solve power/heat/space limitations
- Data Centre co-location

Layer 2 Risks

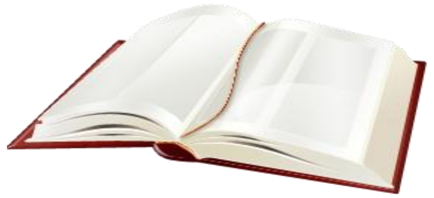
- Flooding of packets between Data Centres
- Spanning Tree (STP) is not easily scalable and risk grows as diameter grows
- STP has no domain isolation – issue in single DC can propagate
- First hop resolution and inbound service selection can cause verbose inter-Data Centre traffic
- In general Cisco recommends L3 routing for geographically diverse locations
- This session focuses on making limited L2 connectivity as stable as possible

MTU Requirements:



- EoMPLS Port Mode: 1522 Bytes
- EoMPLS VLAN Mode: 1526 Bytes
- VPLS: 1522 Bytes
- A-VPLS: 1530 Bytes
- OTV: 1542 Bytes
- LISP
  - IPv4 1536 Bytes
  - IPv6 1556 bytes
- FabricPath: 1516 Bytes
- VXLAN: 1550 Bytes
- GRE: 24 Bytes

MTU Requirements:

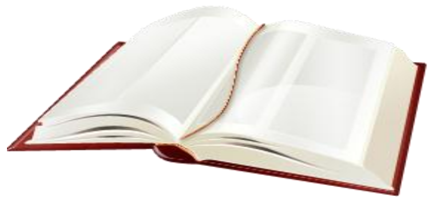


- 802.1ae: ~40 Bytes (16 for the 802.1AE header, 8 for the CMD(Cisco Metadata) and 16 for the ICV (integrity check value))
- IPSEC: 74 Bytes

Data Centre Interconnect

Agenda

- Mobility and Virtualisation in the Data Centre
- LAN Extension Deployment Scenarios
  - –Ethernet Based Solutions
  - –MPLS Based Solutions
    - EoMPLS
    - VPLS
    - A-VPLS
    - EVPN
- Overlay Transport Virtualisation (OTV)
- Encryption
- IP Mobility without LAN Extension
- Path optimisation
- VXLAN
- Summary and Conclusions
- Q&A



= For your Reference

Ethernet Based Solutions

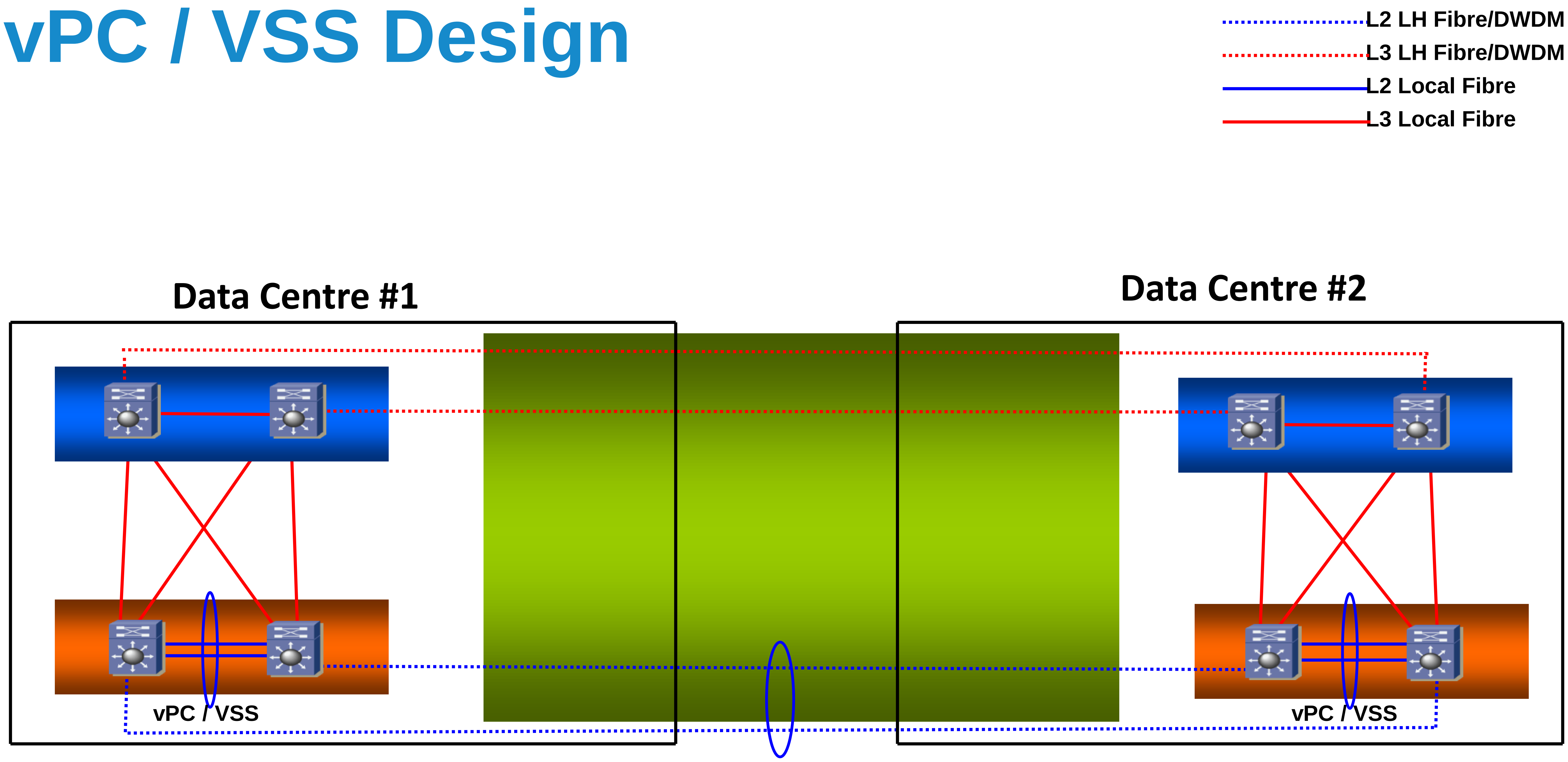
Layer 2 Prerequisites for All Options

- This session assumes a fairly detailed knowledge of Spanning Tree Protocol
- Items we leverage in this solution:
  - 802.1w
  - 802.1s
  - Port Fast
  - BPDU Filter
  - BPDU Guard
  - Root Guard
  - Loop Guard
  - Bridge Assurance (Catalyst 6500, Nexus 5000/5500 and 7000)

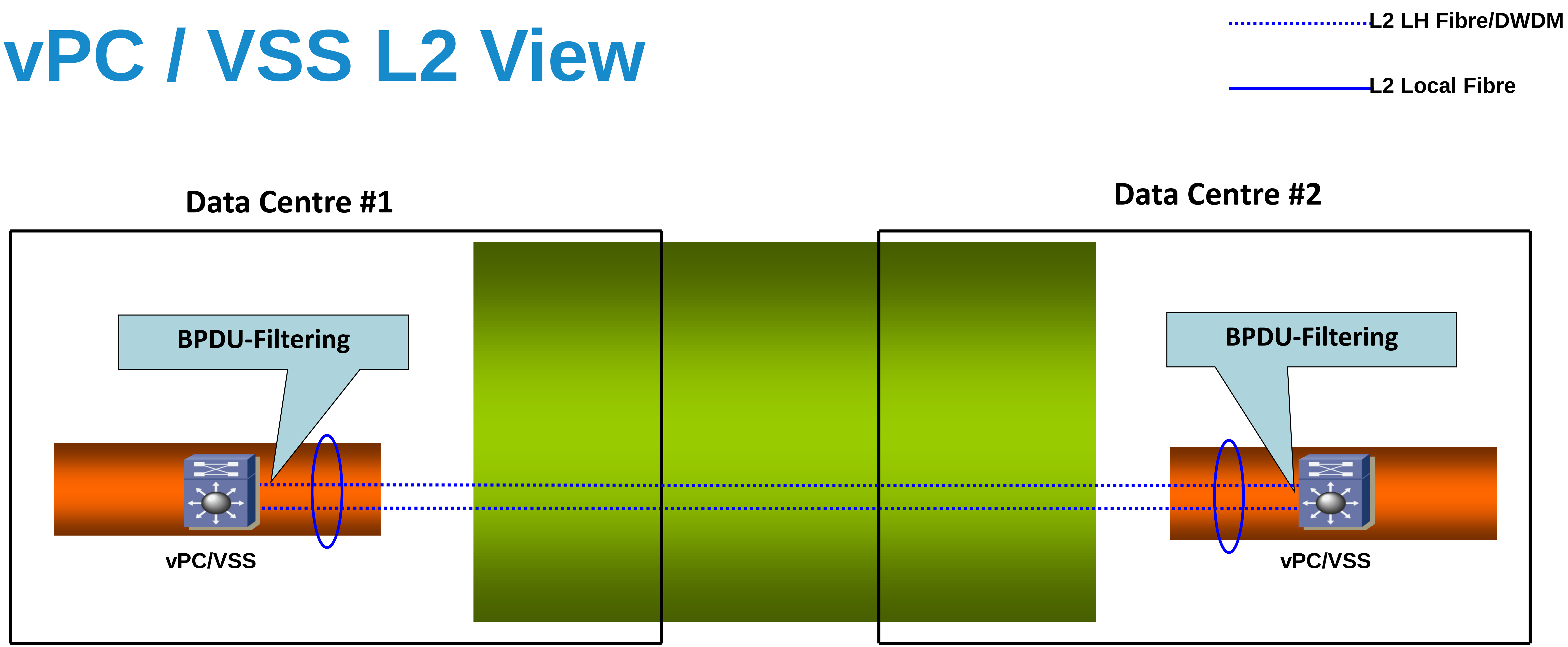
Layer 2 Extension Without Tunnels/Tags (vPC/VSS)

- 6500 with Virtual Switching System cluster (Supported distances at 80km (ZR) Dark Fibre)
- Nexus 7000 with Virtual Port-Channels (Supported distances at 80km (ZR-X2) Dark Fibre)
- All traffic flows to a vPC/VSS member node
- Hub-and-spoke topology from a layer 2 perspective
- Dedicated links to vPC/VSS members from each Data Centre aggregation switch
- Can consume lambda or fibre strands quickly
- Data plane rate limiting in L2 still needs protection
- STP domains are not isolated unless we BPDU-filter at all vPC/VSS aggregation switches

vPC / VSS Design



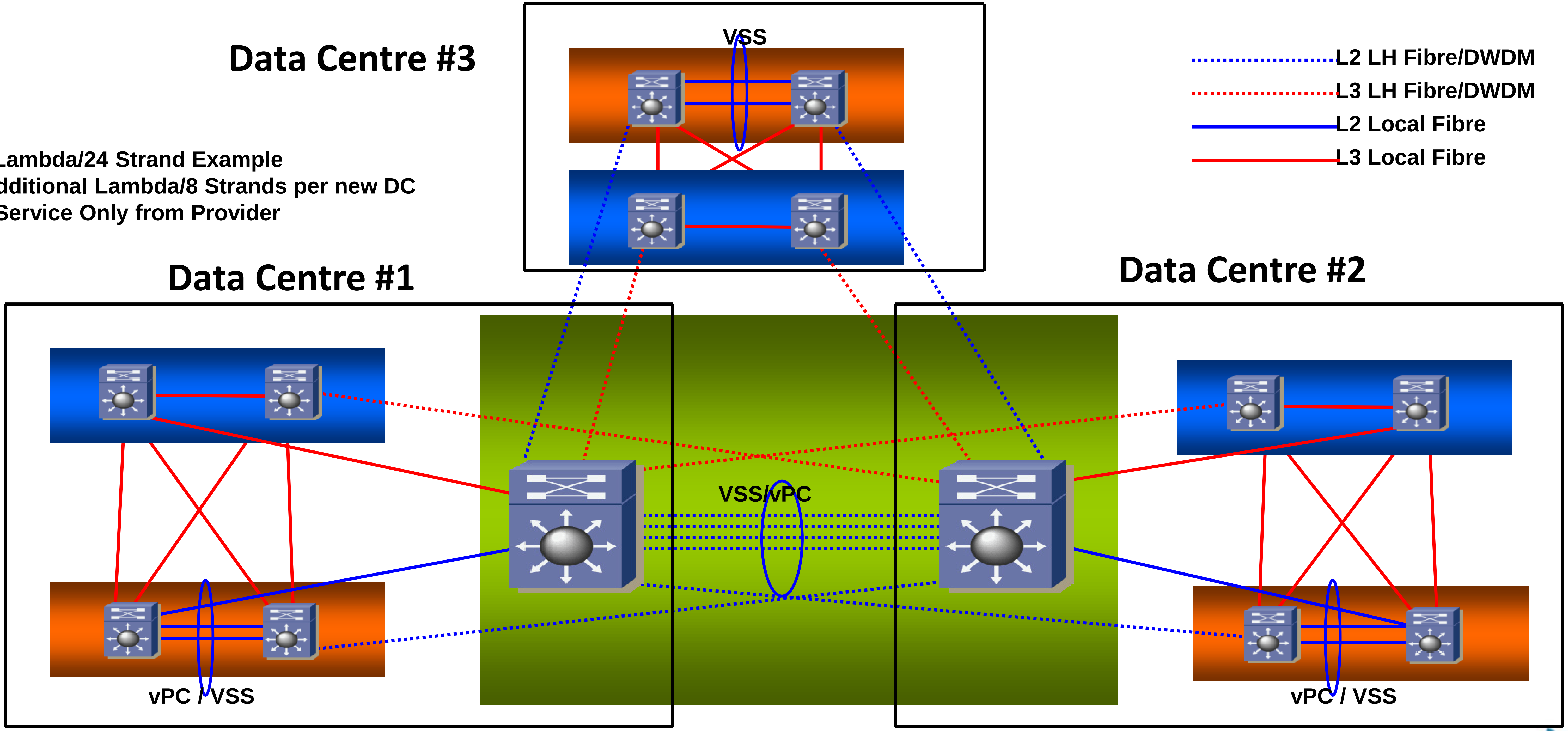
vPC / VSS L2 View



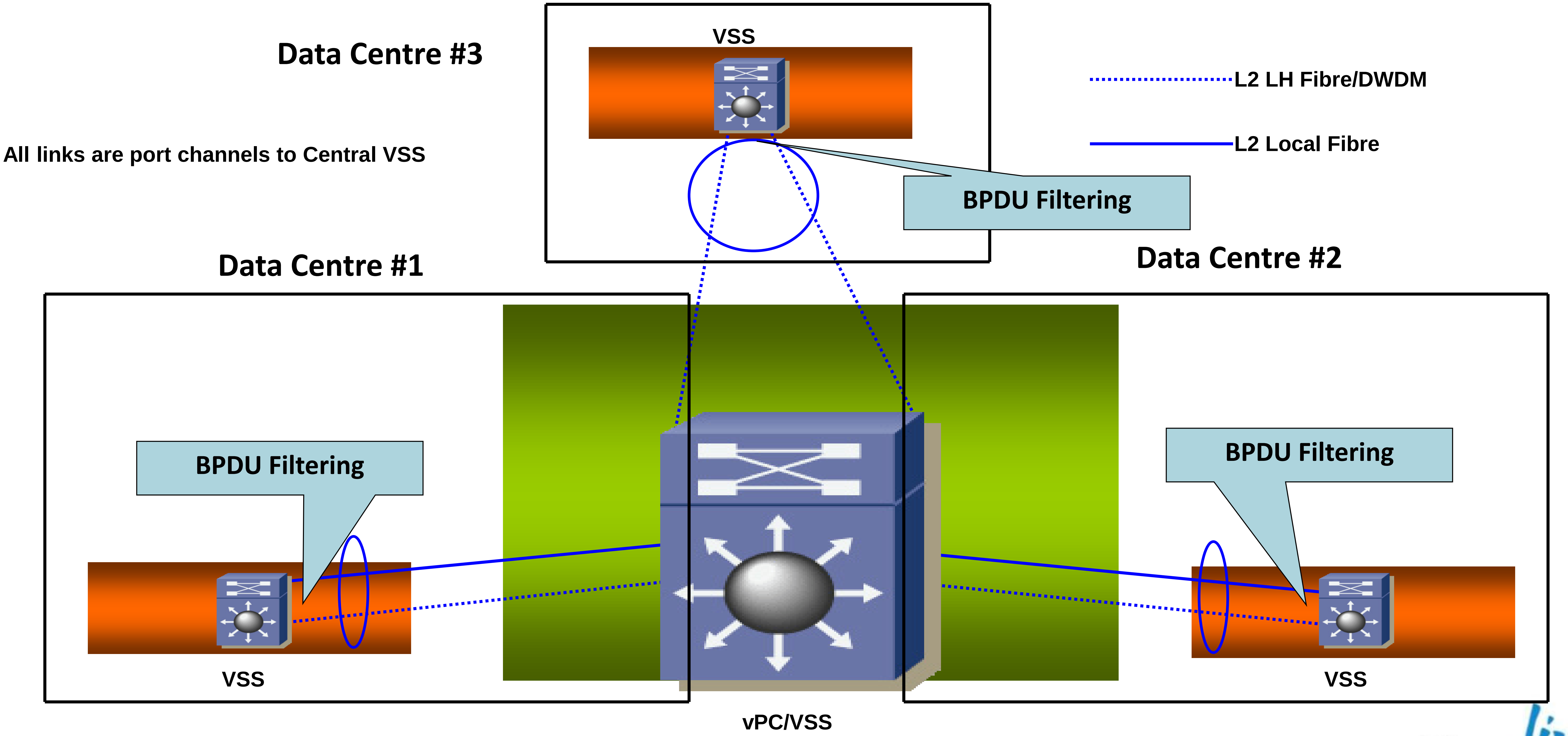
- vPC/VSS Domain ID for facing vPC/VSS layers should be different
- BPDU Filter on the edge devices to avoid BPDU propagation
- STP Edge Mode to provide fast failover times
- No Loop must exist outside the vPC/VSS domain
- No L3 peering between Nexus 7000 devices (i.e. pure layer 2)

vPC / VSS Design

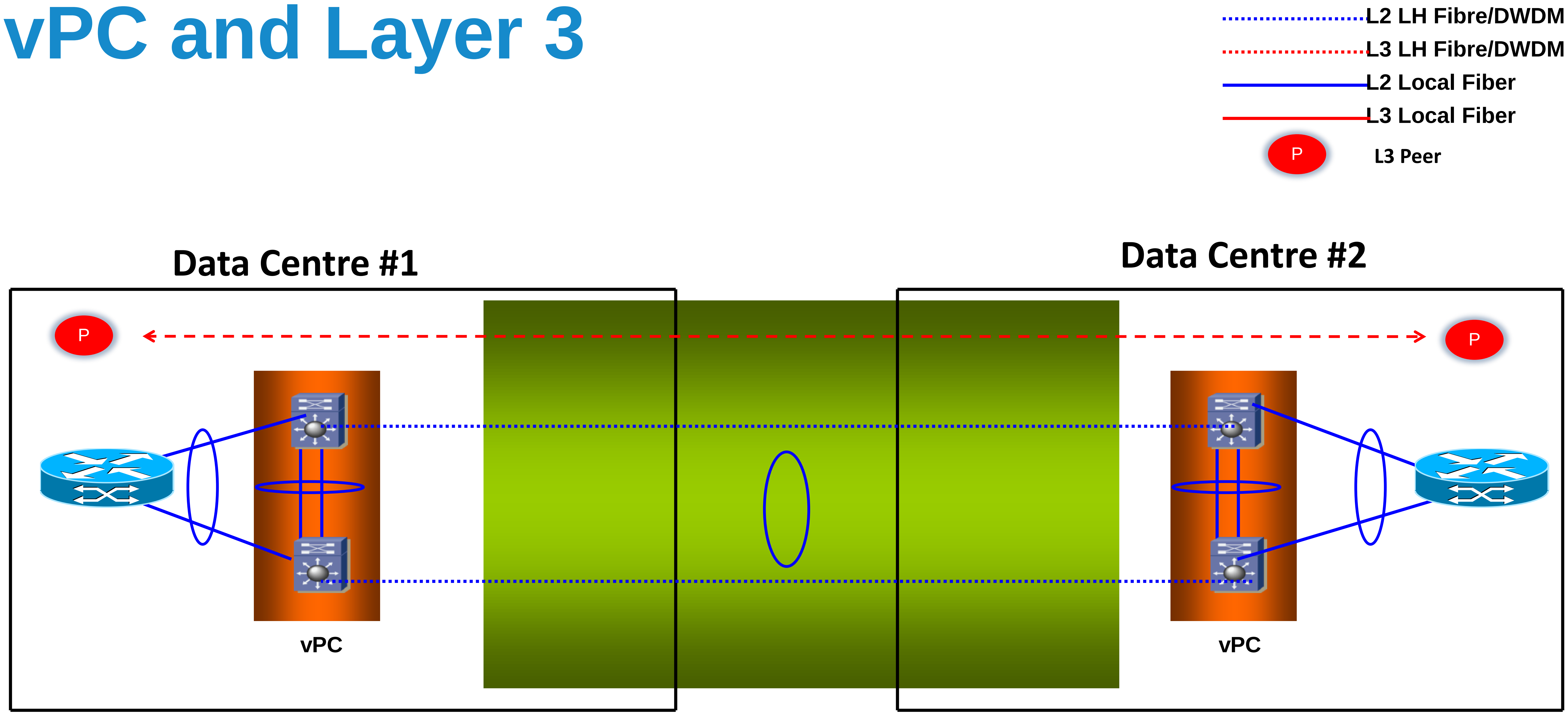
12 Lambda/24 Strand Example
4 Additional Lambda/8 Strands per new DC
L2 Service Only from Provider



vPC / VSS L2 View

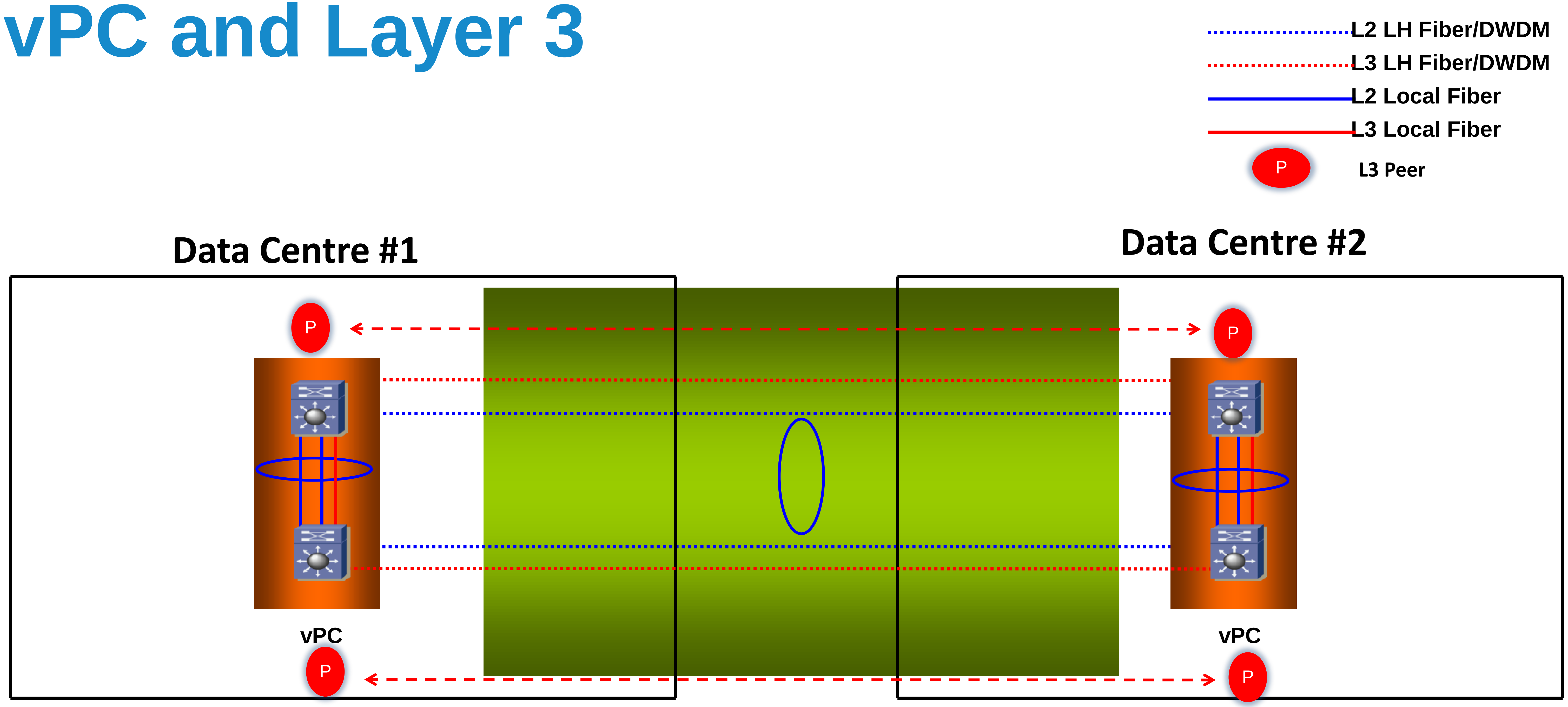


vPC and Layer 3



- Nexus 7000 configured for L2 Transport only
- SVI passive-interface (no IGP peering)

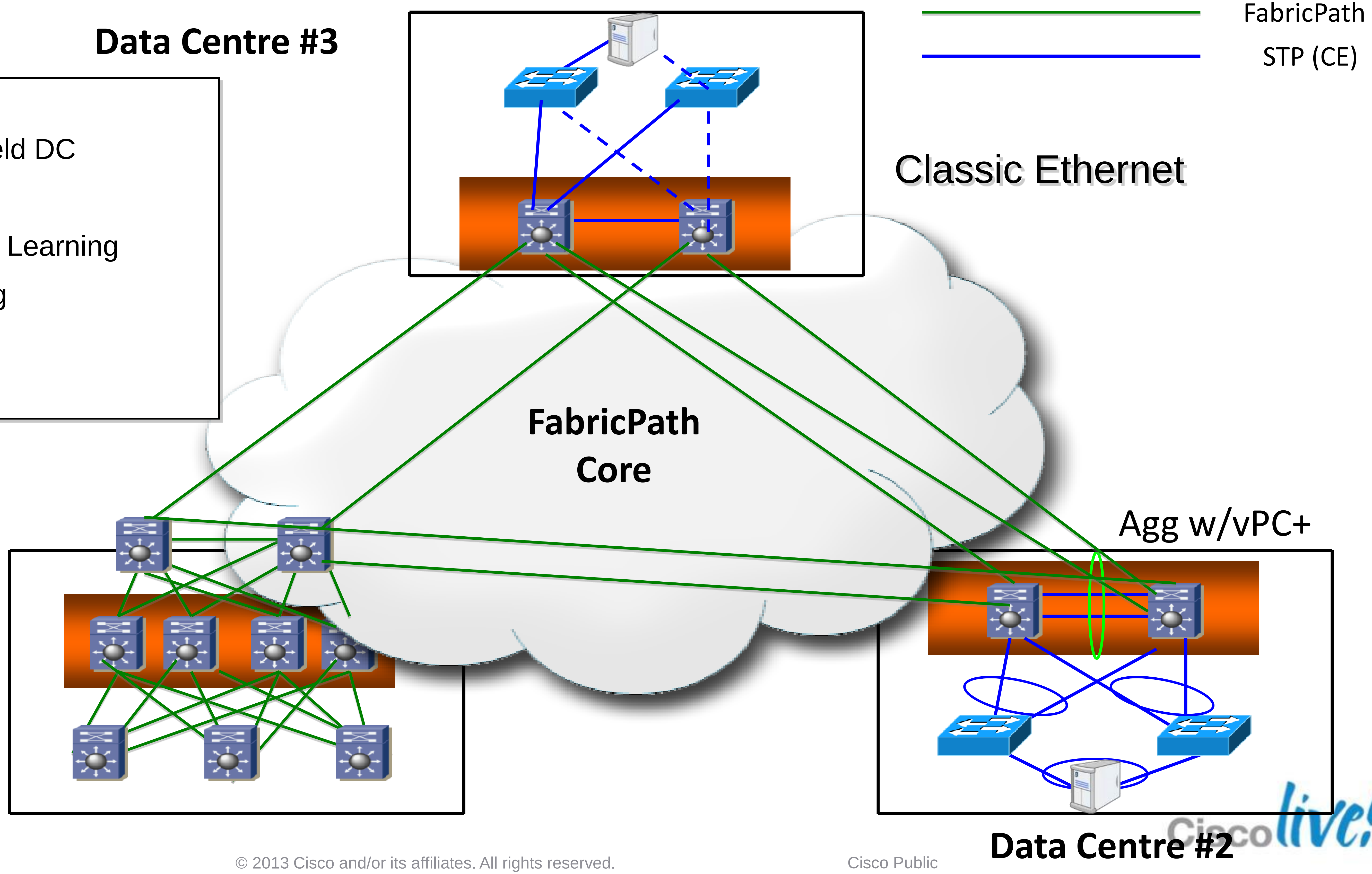
vPC and Layer 3



- Peering over a vPC inter-connection on parallel routed interfaces
- SVI passive-interface (no IGP peering)

FabricPath Design (Partial/Full/Ring Topology)

- Leverage vPC+
- Brownfield / Greenfield DC
- STP Integration
- Conversational MAC Learning
- Native VLAN Pruning
- TTL / RPF
- ECMP for L2



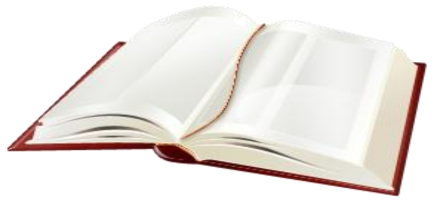
FabricPath Requirements

- FabricPath L2 ISIS adjacencies are Point to Point
  - Need for direct Point to Point L1 WAN Links
  - FabricPath over VPLS is not supported
  - L2 managed service : Dark Fiber, DWDM, EoMPLS
  - MTU requirements : 16 extra Bytes for FabricPath header
- BFD not supported
- Multi-desination Traffic: Multicast/ARP traffic across DCI can be non-optimal due to MDT (Multi-destination tree)
- FabricPath and HSRP Localisation does not work

Data Centre Interconnect

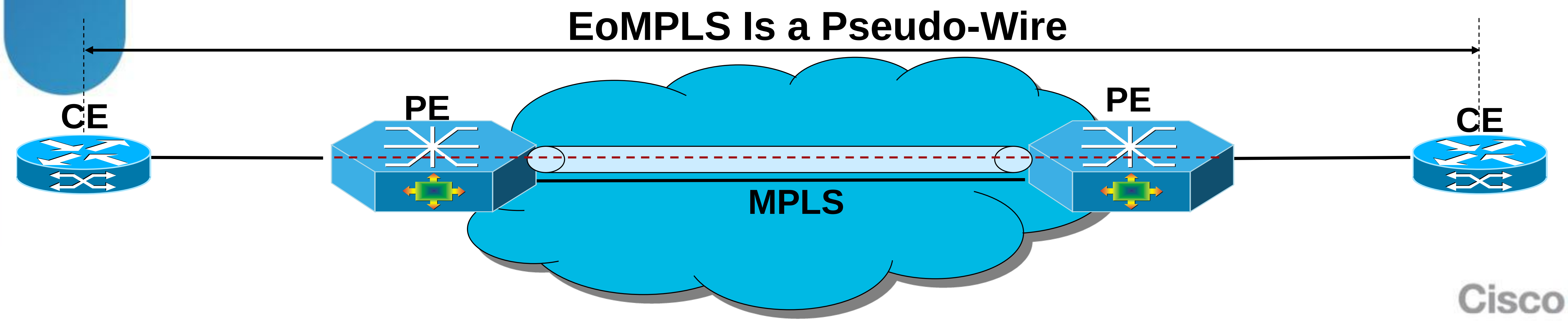
Agenda

- Mobility and Virtualisation in the Data Centre
- LAN Extension Deployment Scenarios
  - –Ethernet Based Solutions
  - –MPLS Based Solutions
    - EoMPLS
    - VPLS
    - A-VPLS
    - EVPN
- Overlay Transport Virtualisation (OTV)
- Encryption
- IP Mobility without LAN Extension
- Path optimisation
- VXLAN
- Summary and Conclusions
- Q&A


= For your Reference

EoMPLS (Ethernet Over MPLS)

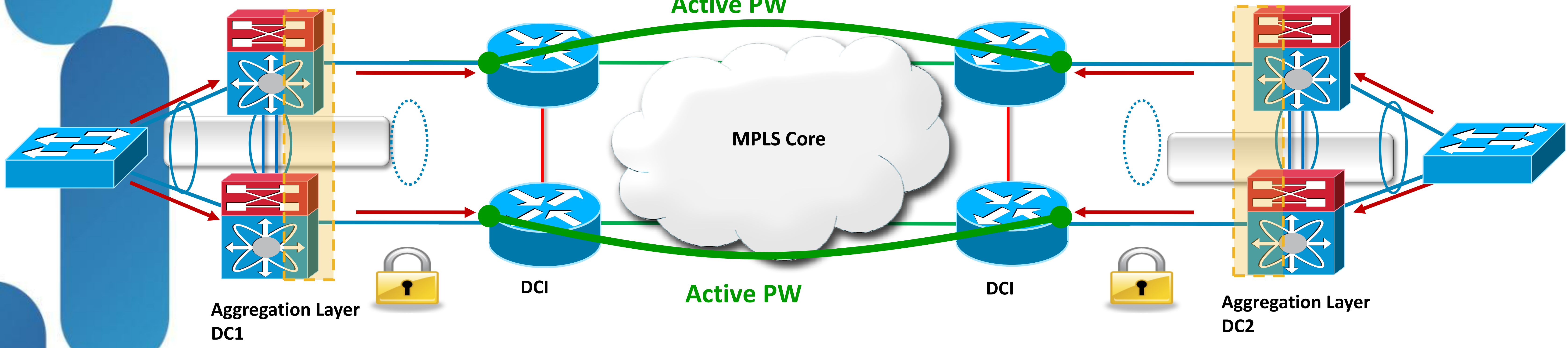
- Encapsulates Ethernet frames inside MPLS packets to pass layer 3 network
- EoMPLS has routing separation from metro core devices providing connectivity – CE flapping routes won't propagate inside MPLS
- Point to point links between locations
- Data plane rate limiting in L2 still needs protection



EoMPLS Usage with DCI

End-to-End Loop Avoidance using Edge to Edge LACP

On DCI Etherchannel:
- STP Isolation (BPDU Filtering)
- Broadcast Storm Control
- FHRP Isolation

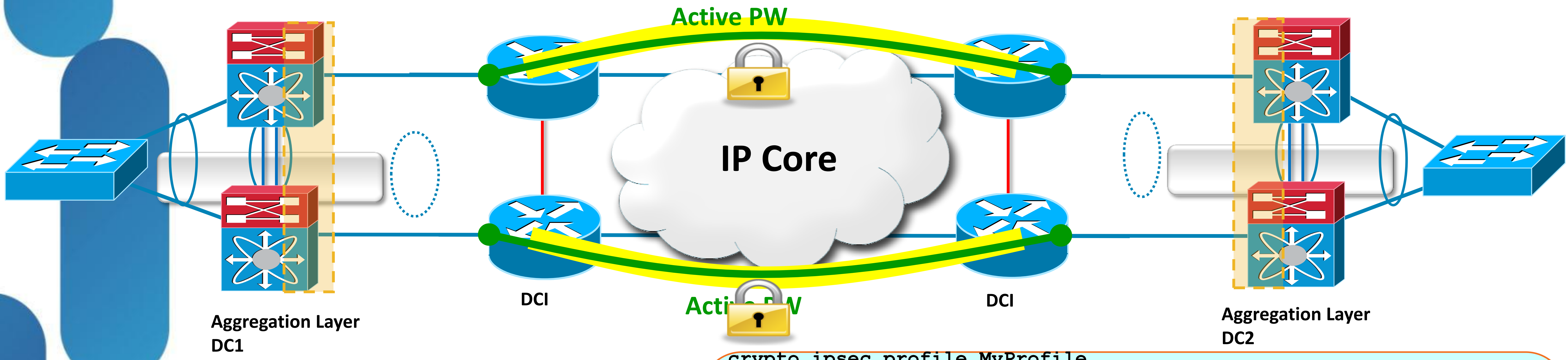


Encryption Services with 802.1AE

requires a full meshed vPC

➔ 4 PW

EoMPLS Usage with DCI
Over IP core




```
crypto ipsec profile MyProfile
 set transform-set MyTransSet

interface Tunnel100
 ip address 100.11.11.11 255.255.255.0
 ip mtu 9216
 mpls ip
 tunnel source Loopback100
 tunnel destination 12.11.11.21
 tunnel protection ipsec profile MyProfile
```

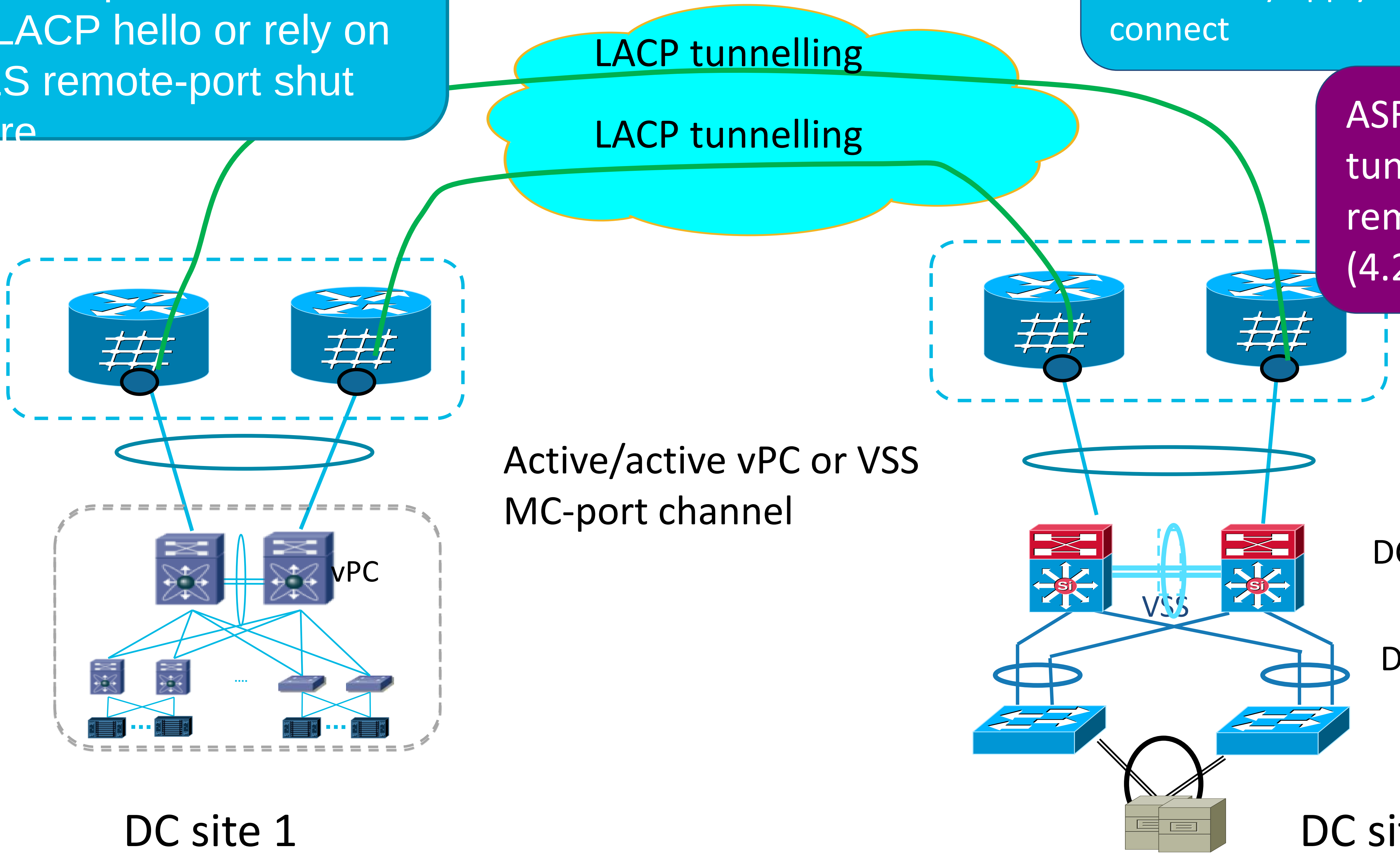
ASR 9000 DCI Solution – LACP Tunnelling

active-active redundancy with fast convergence

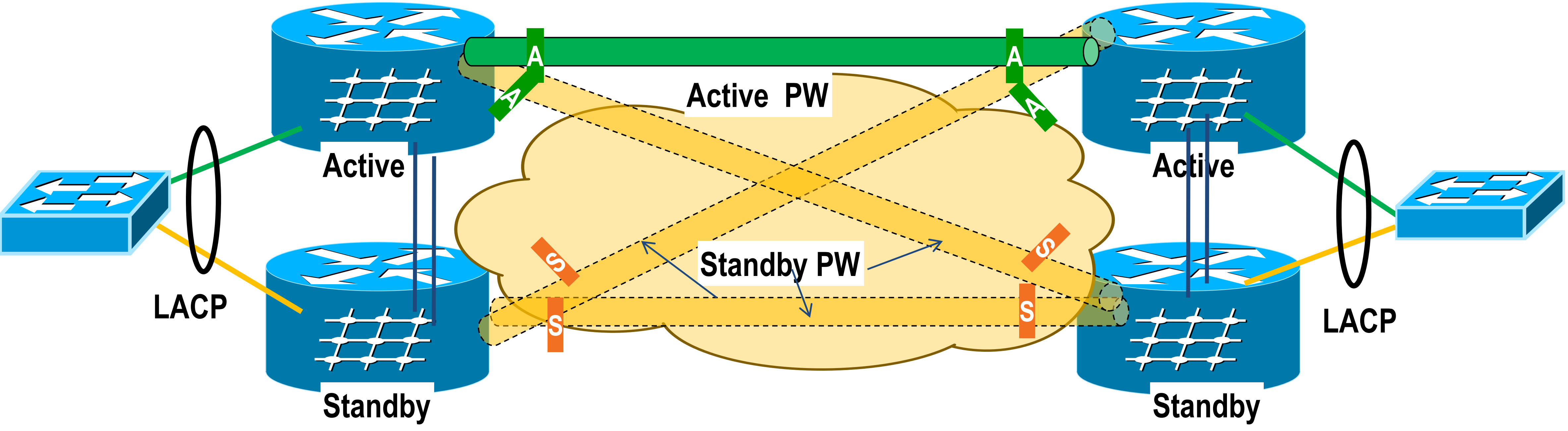
Port-mode EoMPLS, tunnel all packets, including LACP. Convergence depends on how fast of the LACP hello or rely on the EoMPLS remote-port shut down feature

Simple configuration, active/active load balancing. Transparent over PE and MPLS cloud. Only apply to two DC sites inter-connect

ASR 9000 support LACP tunnelling and EoMPLS remote-port shut down (4.2.1)

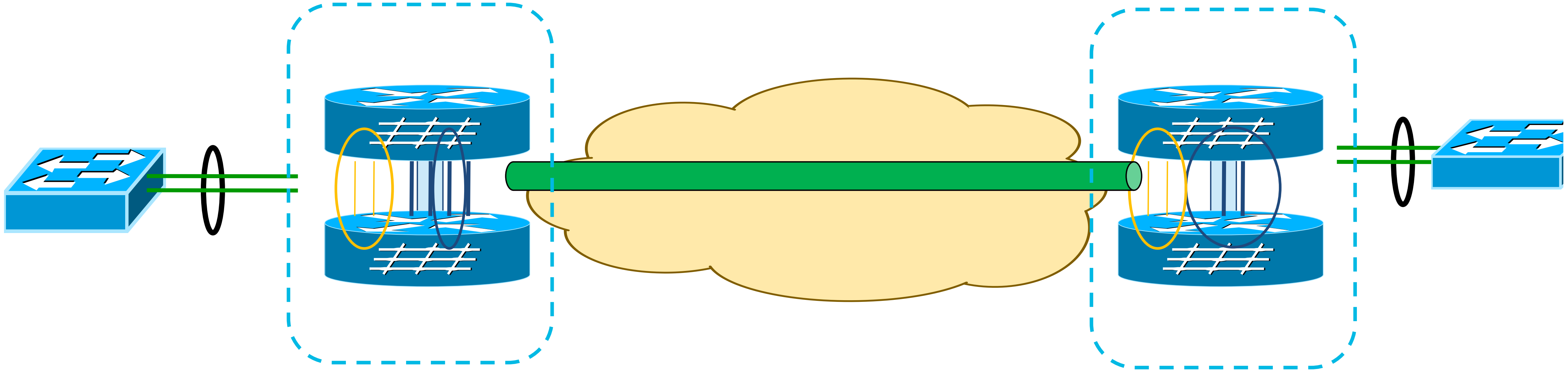


Deployment Example – L2VPN Service



Solution1: MC-LAG + 2-way PW redundancy

- Active/standby MC-LAG → bandwidth inefficiency
- 4 PWs with 3 standby → control plane overhead
- PW failover time depends on the number of PWs → slow convergence
- Require additional state sync (for example, IGMP Snooping table) to speed up service convergence → complex

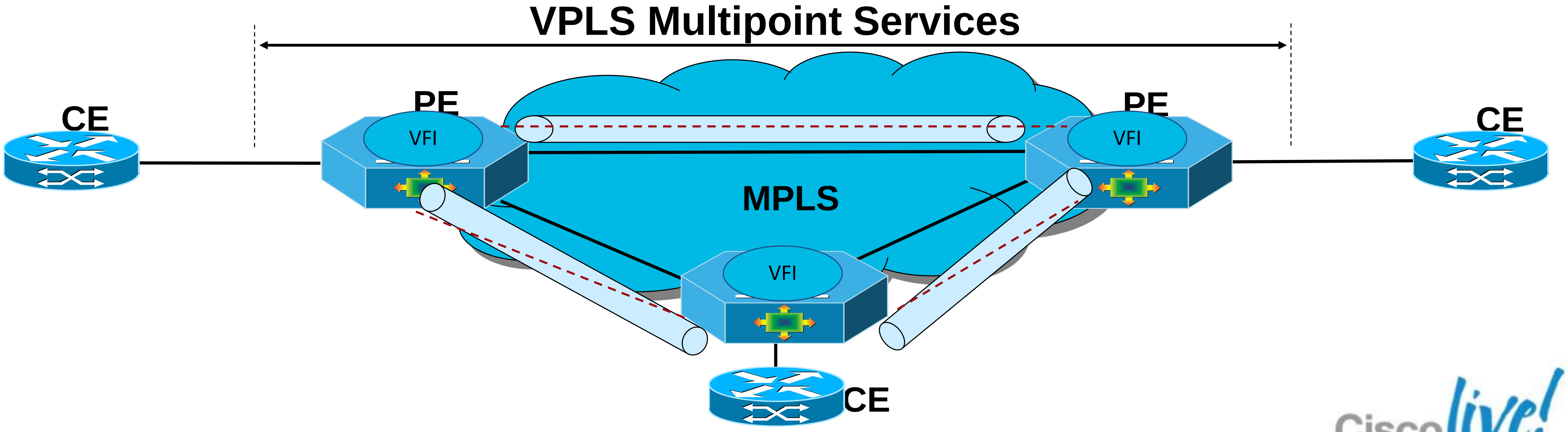


Solution 2: ASR 9000 Cluster

- Active/active regular LAG
- Single PW
- Link/Node failure is protected by LAG, PW is even not aware → super fast convergence
- State sync naturally
- Simple, fast solution

Virtual Private LAN Service (VPLS)

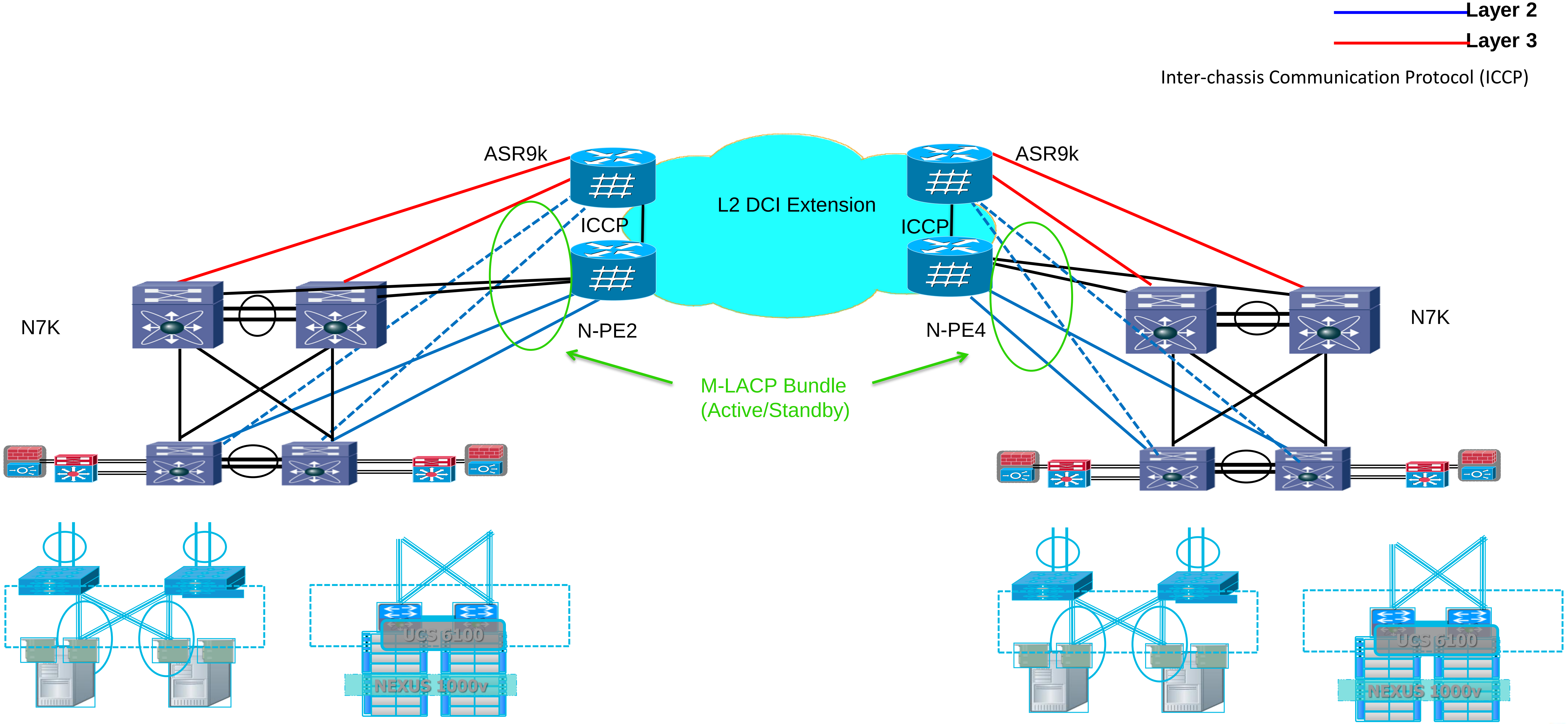
- VPLS defines an architecture that allows MPLS networks to offer Layer 2 multipoint Ethernet Services
- Metro Core emulates an IEEE Ethernet bridge (virtual)
- Virtual Bridges linked with EoMPLS Pseudo Wires
- Data plane rate limiting in L2 still needs protection



Virtual Forwarding Instance (VFI)

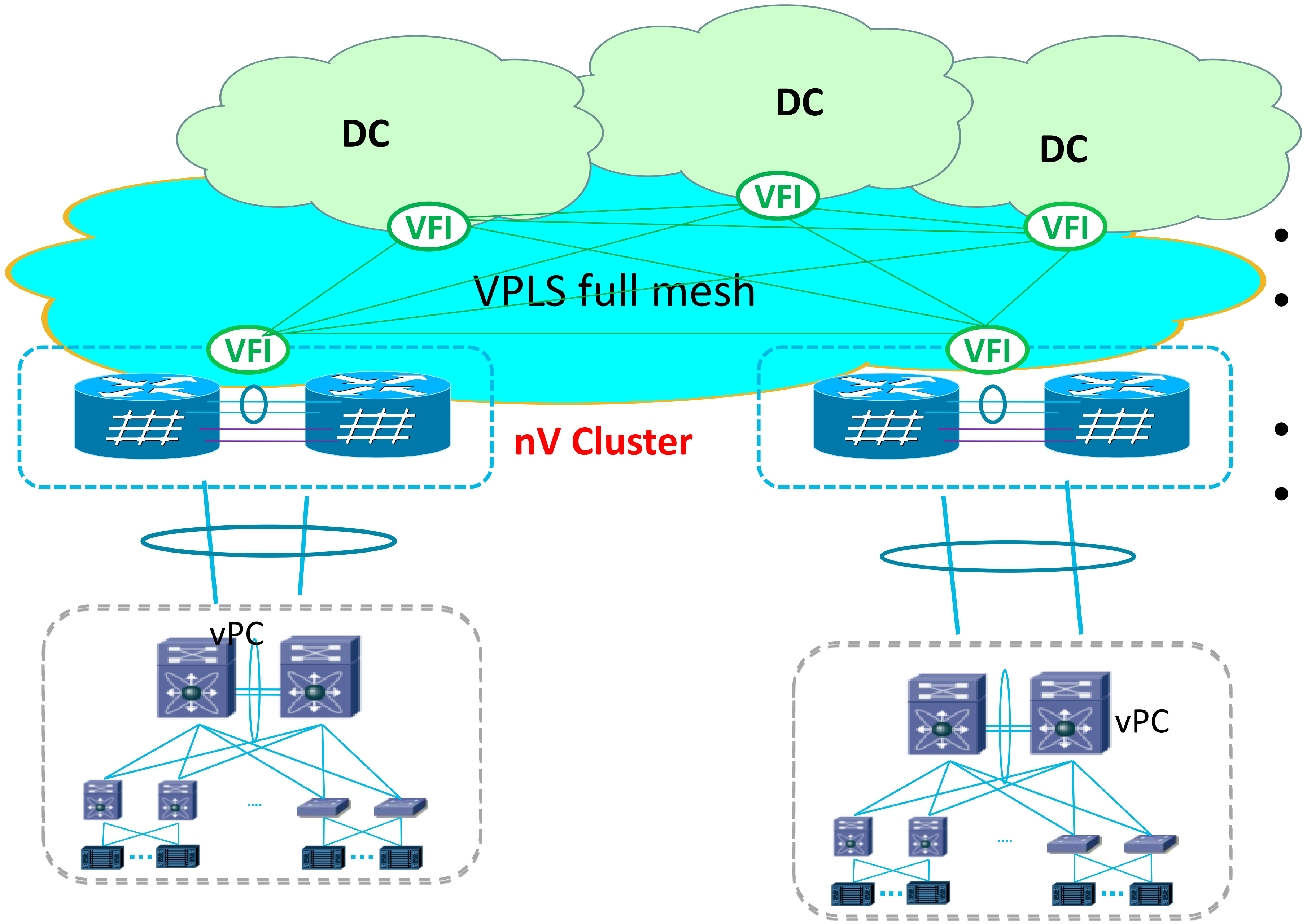
- IOS Representation of Virtual Switch Interface
- Flooding / Forwarding
  – MAC table instances per customer (port/VLAN) for each PE
  – VFI will participate in learning and forwarding process
  – Associate ports to MAC, flood unknowns to all other ports
- Address Learning / Aging
  – LDP enhanced with additional MAC List TLV (label withdrawal)
  – MAC timers refreshed with incoming frames
- Loop Prevention
  – Create full-mesh of Pseudo Wire VCs (EoMPLS)
  – Unidirectional LSP carries VCs between pair of N-PE Per
  – VPLS Uses “split horizon” concepts to prevent loops

MCLAG with ASR 9000 and Nexus 7000

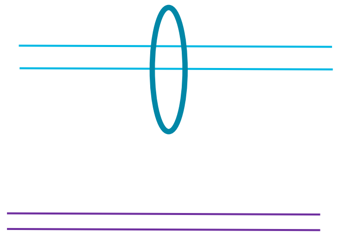


VPLS Multi-homing – ASR9K nV Cluster

Simple and faster network convergence



- Reduce the Number of PWs
- Simplify VPLS dual homing with active/active link bundle
- per-flow and per-VLAN load balancing
- Sub-second to 50msec fast convergence

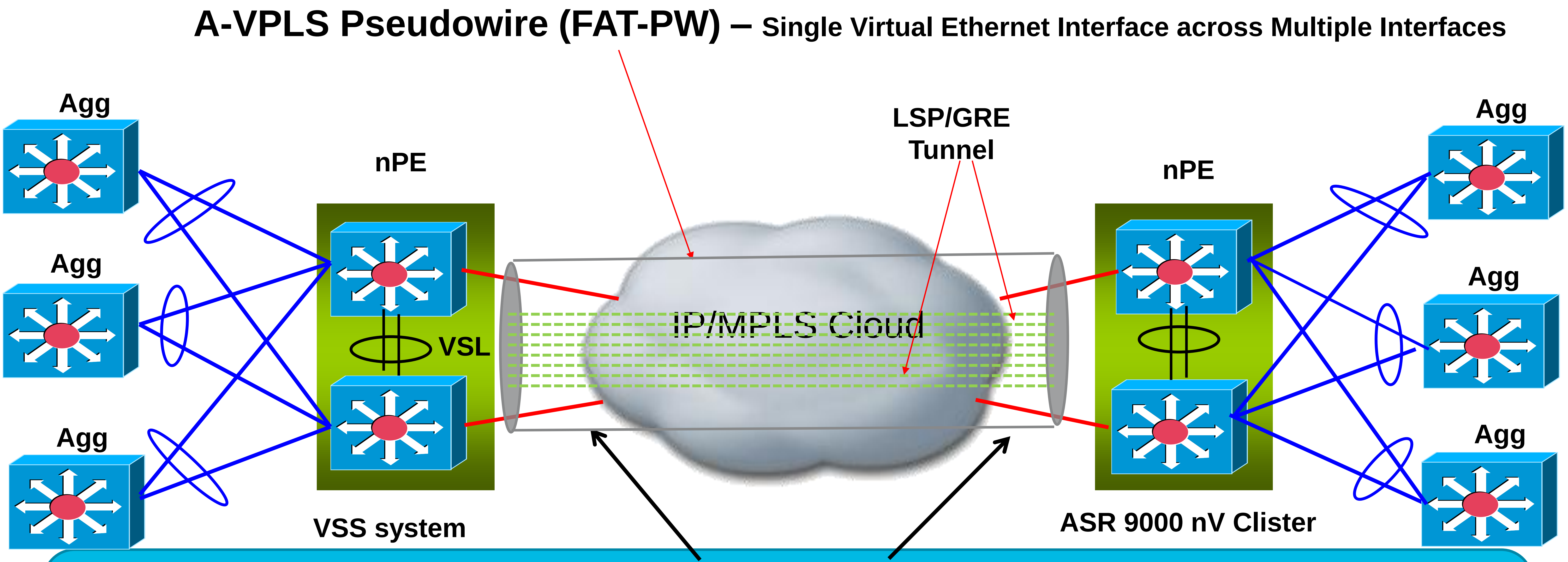


data-plane: port-channel used between the ASR9000 on any 10G or 100G Interfaces.

control-plane: One or two 10G/1G from each RSP this is a Special external EOBC 1G/10G ports on RSP.

Note: Split-brain: keepalive over any L2 cloud Management port or any regular data port or interface or sub-interface.

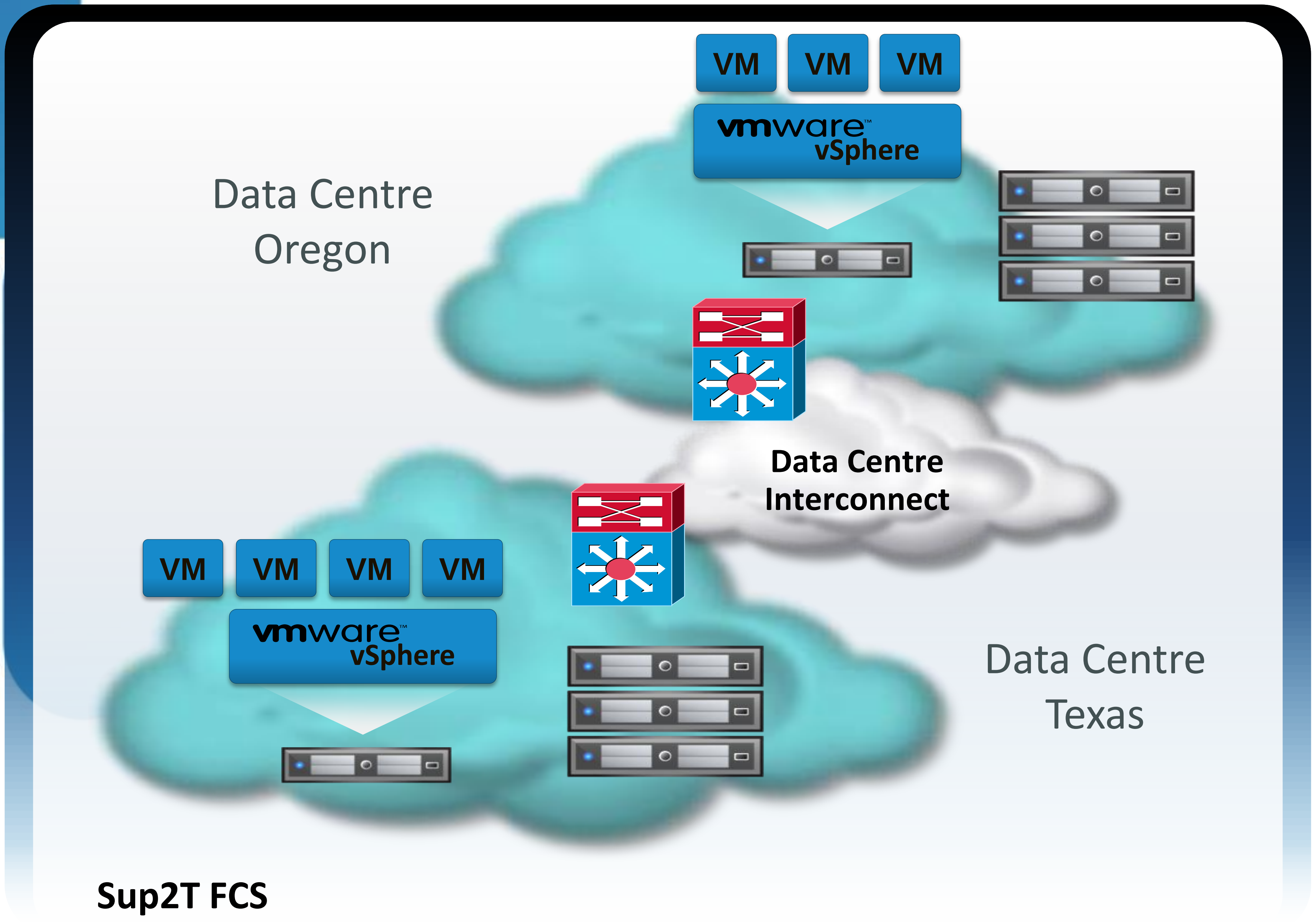
Multi-Pathing with A-VPLS (6500 and ASR9000)



Up to 8 equal cost paths between any two sites
A label is assigned to each equal cost path based on routing reachability of neighbor
Simplified CLI: Virtual Ethernet interface
Loadbalancing at L2/L3/L4

Supervisor 2T VPLS on Any Port

Data Centre Interconnect using VSS



- Native VPLS support
- EoMPLS – port-mode and sub-interface mode
- QUAD Supervisor VSS NSF/SSO
- CapEx Savings: No need for SIP Based linecards
- Application VM mobility
  - Redistribute compute workloads
  - No Service Disruption
  - Capacity management
  - Disaster avoidance
  - Data Centre upgrades

EVPN – The Principle

From PE1

iBGP L3-NLRI:

- **next-hop: PE1**
- **<C-IP1, L1>**

iBGP L2-NLRI

- **next-hop: PE1**
- **<C-MAC1, L2>**

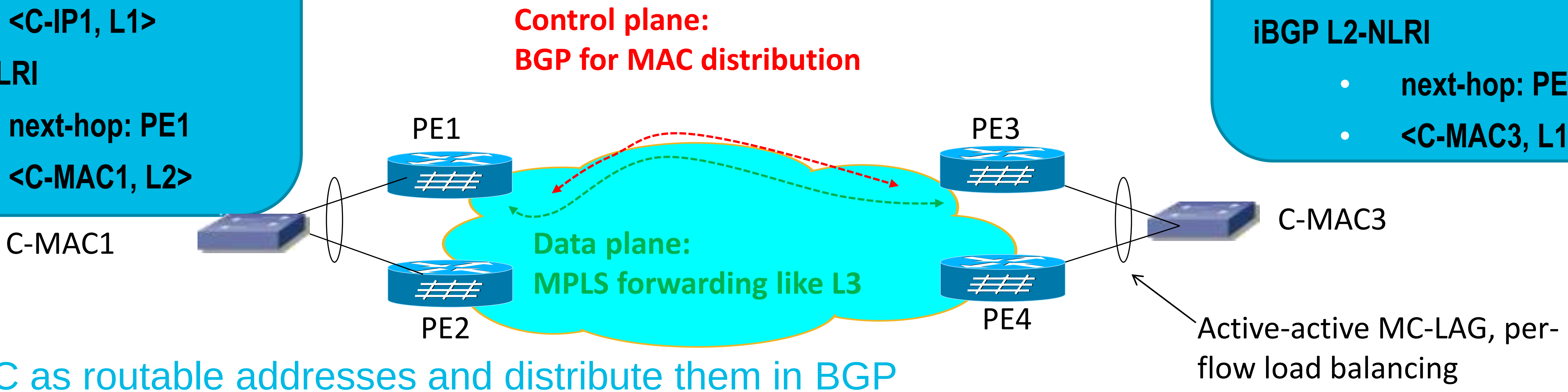
From PE3

iBGP L3-NLRI:

- **next-hop: PE3**
- **<C-IP5, L1>**

iBGP L2-NLRI

- **next-hop: PE3**
- **<C-MAC3, L1>**



- Treat MAC as routable addresses and distribute them in BGP
- Receiving PE injects these MAC addresses into forwarding table along with its associated adjacency like IP prefix
- When multiple PE nodes advertise the same MAC, then multiple adjacency is created for that MAC address in the forwarding table: multi-paths
- When forwarding traffic for a given unicast MAC DA, a hashing algorithm based on L2/L3/L4 header is used to pick one of the adjacencies for forwarding: per-flow load balancing
- PW is not required

Note: Network Layer Reachability Information (NLRI)

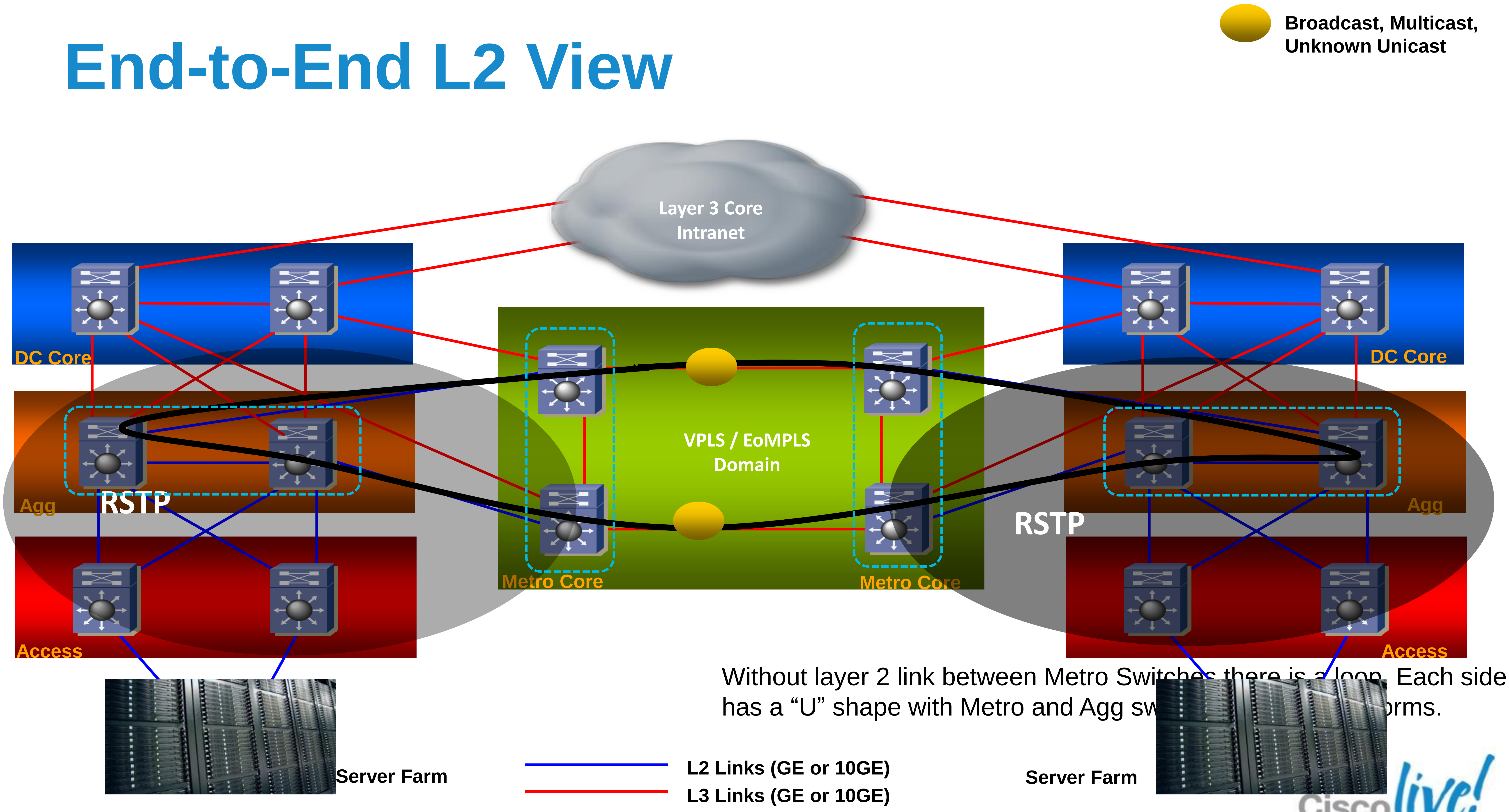
Why Evolve to EVPN?

- Optimised forwarding for both unicast and multicast, per-flow based load balancing like L3 ECMPs
- Simple access multi-homing, active-active per-flow load balancing
- Fast convergence as L3 network
- Highly scale as L3 network
- Flexible policy control for E-tree, extranet, etc
- Same inter-AS solution like L3VPN
- Consistent operation as L3VPN service → truly converged network
  - Same BGP control plane for both L2 and L3VPN
  - Same MPLS based forwarding plane for both L2 and L3VPN
  - No EoMPLS/VPLS PW required, no control plane signalling overhead

Spanning Tree

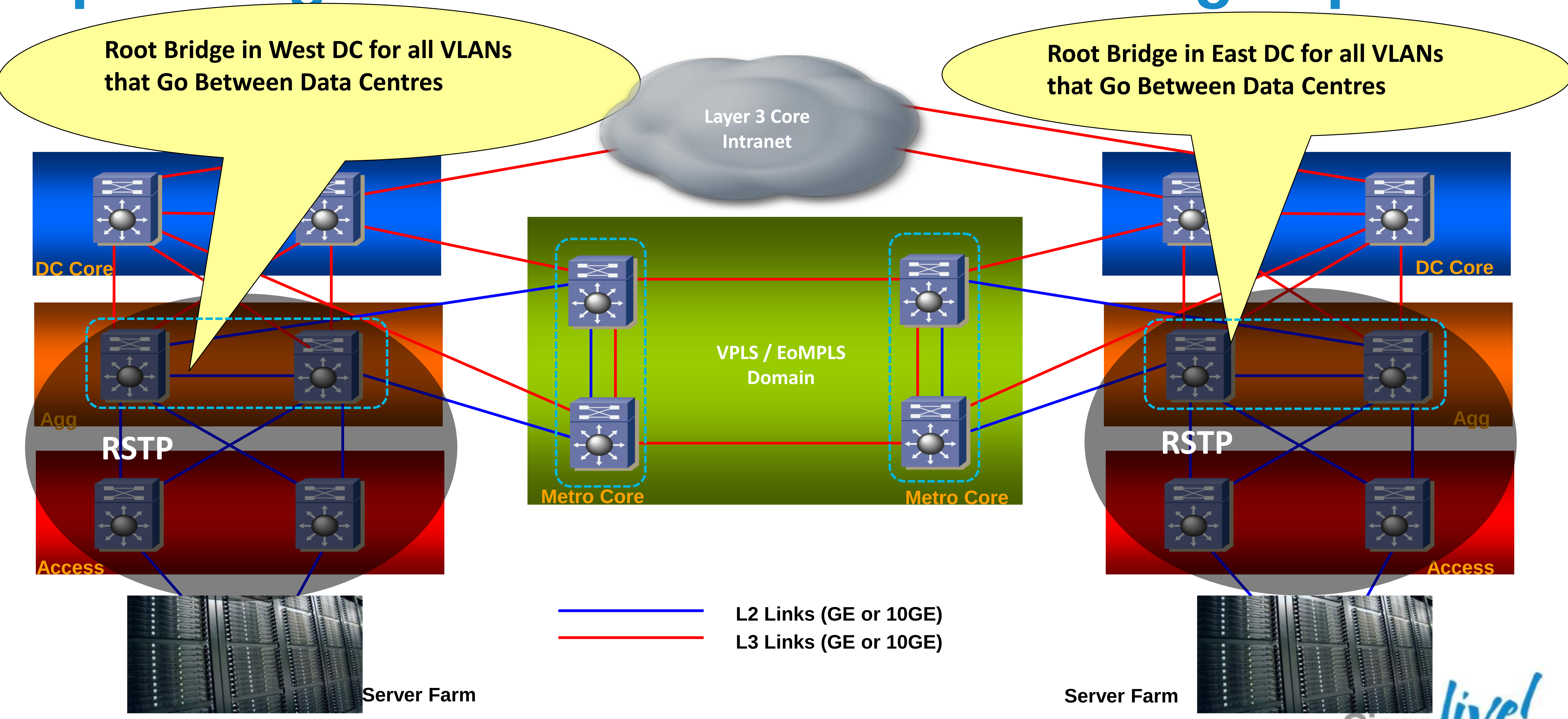
- Spanning-Tree BPDUs will NOT traverse between the Data Centres – It isn’t needed (and blocked) with VPLS
- We still need to control data plane layer 2 events (i.e., limit the traffic)
- Since enterprises want dual N-PE devices, and VPLS blocks BPDUs, we require method to block within a local DC

End-to-End L2 View



Without layer 2 link between Metro Switches there is a loop. Each side has a “U” shape with Metro and Agg sw … rms.

Spanning Tree – Local STP Root Bridges per DC



Storm Control

- Traffic storms when packets flood the LAN
- Traffic storm control feature prevents LAN ports from being disrupted by broadcast or multicast flooding
- Rate limiting for unknown unicast (UU) must be handled at Data Centre aggregation; unknown unicast flood rate-limiting (UUFRL):
  - mls rate-limit layer2 unknown rate-in-pps [burst-size]
- Storm Control is configured as a percentage of the link that storm traffic is allowed to use.
  - storm-control broadcast level 1.00 (% of b/w may vary – need to baseline)
  - storm-control multicast level 1.00 (% of b/w may vary – need to baseline)

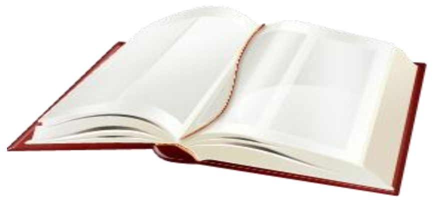
Summary of Tagging Section

- EoMPLS well suited for Router-Router links
- VPLS well suited for Switch-Switch links
- Straightforward to scale to multiple Data Centre locations
- MST and MC-LAG both work well
  - One tradeoff is QinQ support against number of VLANs to pass
  - Another is the root of the spanning tree for inter-DC VLANs
- A-VPLS
  - Backwards Compatible
  - Load Balancing Enhancements
  - Simplified Configuration
  - Single virtual nPE

Data Centre Interconnect

Agenda

- Mobility and Virtualisation in the Data Centre
- LAN Extension Deployment Scenarios
  - –Ethernet Based Solutions
  - –MPLS Based Solutions
    - EoMPLS
    - VPLS
    - A-VPLS
    - EVPN
- Overlay Transport Virtualisation (OTV)
- Encryption
- IP Mobility without LAN Extension
- Path optimisation
- VXLAN
- Summary and Conclusions
- Q&A


= For your Reference

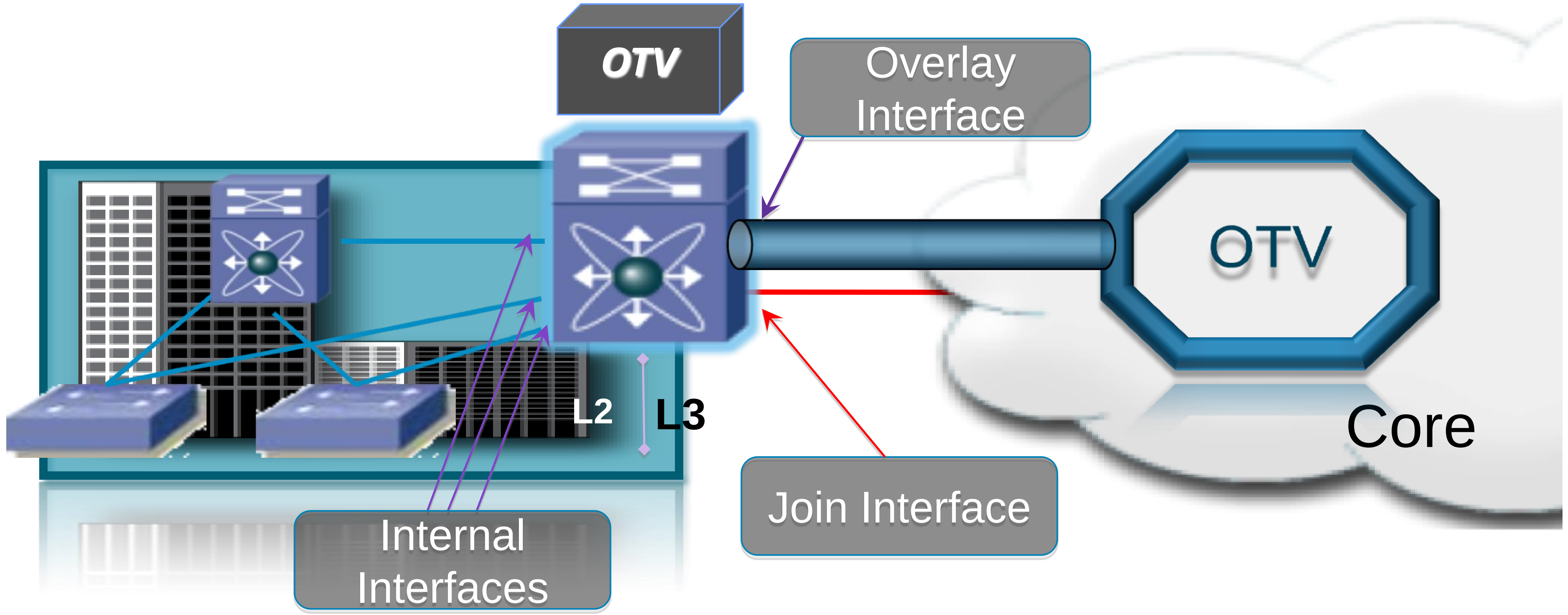
Overlay Transport Virtualisation (OTV)

Overlay Transport Virtualisation (OTV)

- OTV is a MAC-in-IP method that extends Layer 2 connectivity
- Ethernet LAN Extension over any Network
- Ethernet in IP “MAC routing”
- Multi-dataCentre scalability
- Simplified Configuration & Operation
- Seamless overlay - no network re-design
- Single touch site configuration
- High Resiliency
- Failure domain isolation
- Seamless Multi-homing
- Maximises available bandwidth
- Automated multi-pathing
- Optimal multicast replication

OTV Interface Types

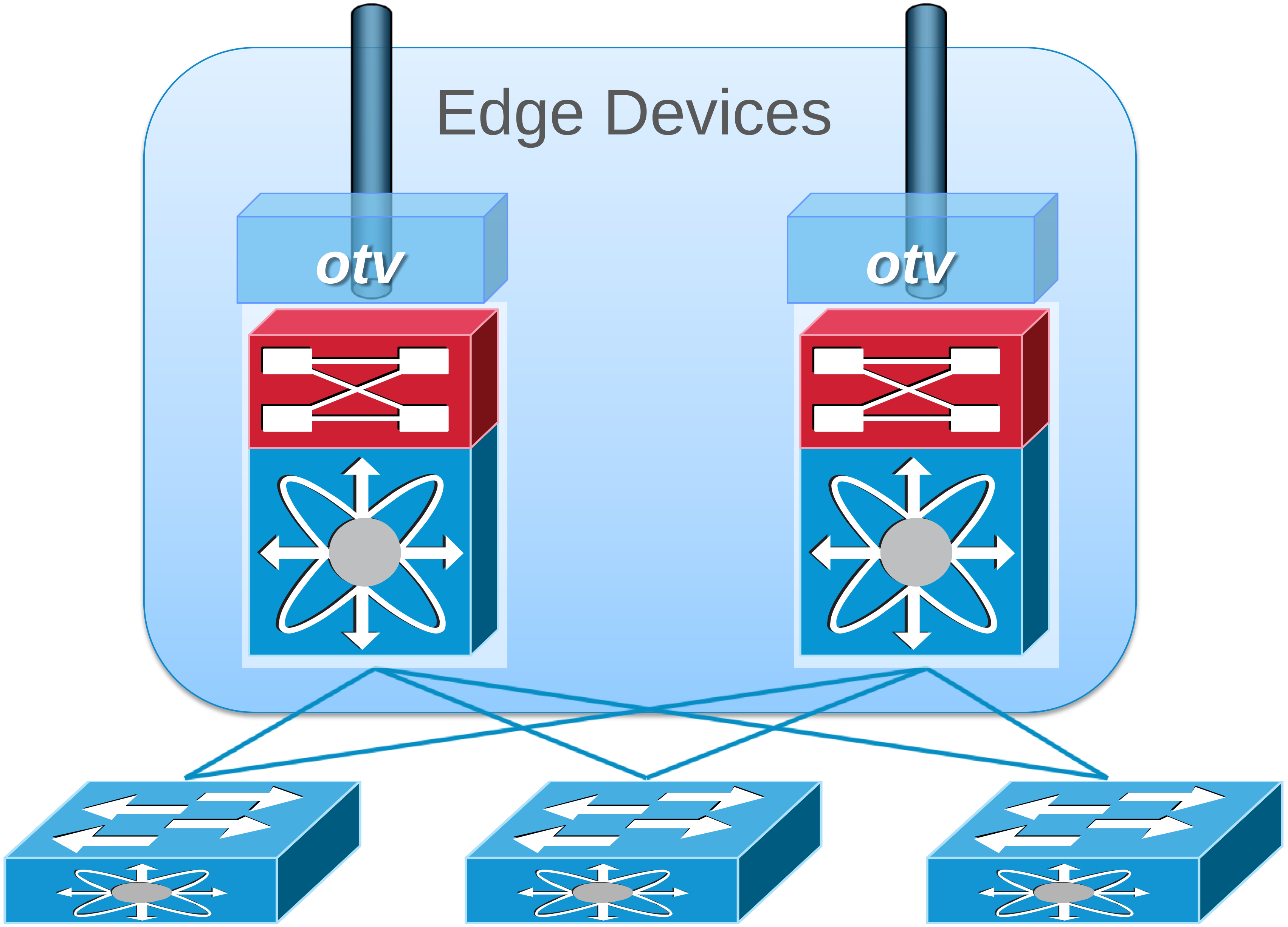
- Edge Device
- Internal Interfaces
- External Interface
- Join Interface
- Overlay Interface



Introduction

Terminology: Edge Device

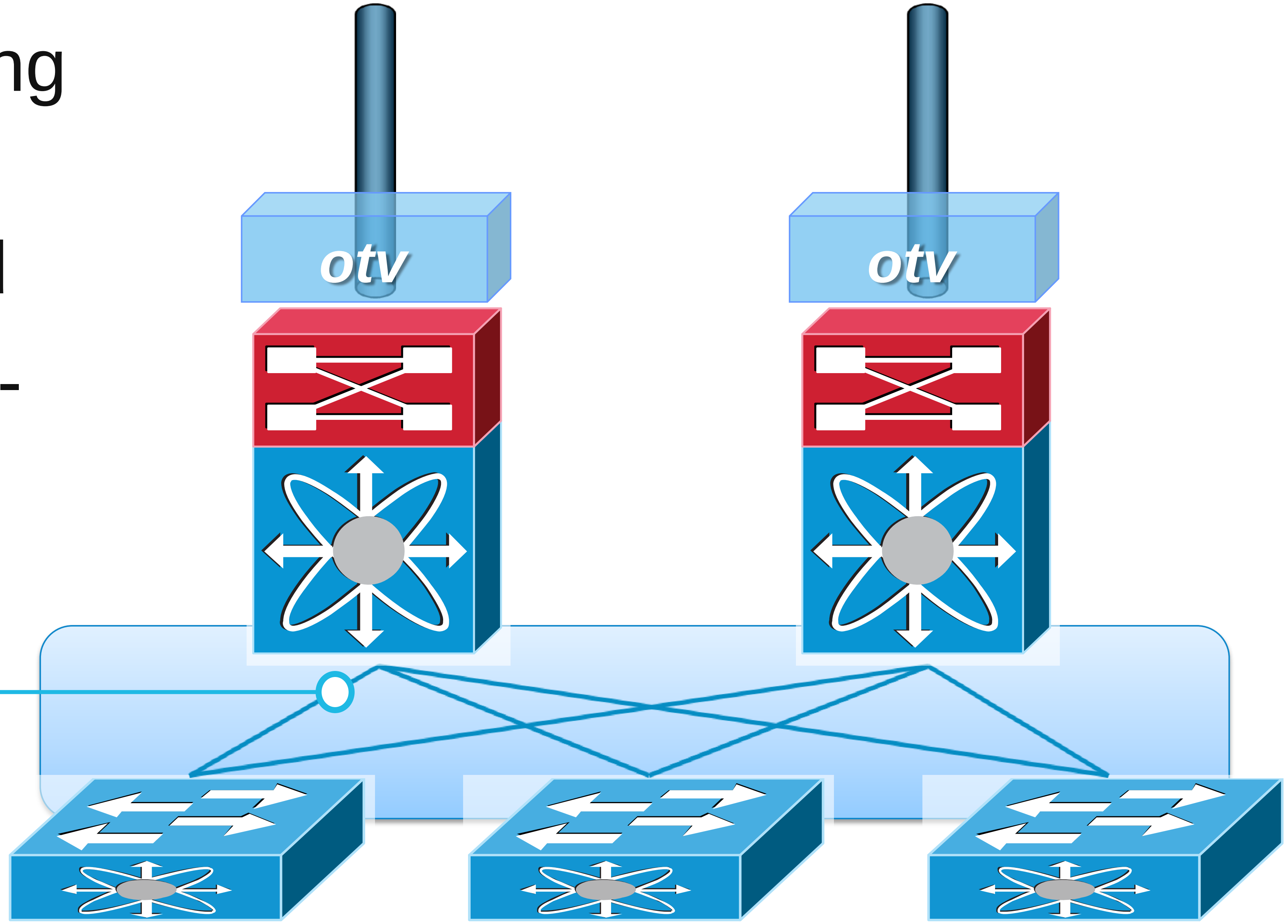
- Performs OTV functions
- Support multiple OTV devices per site
- OTV requires the Transport Services (TRS) license
- Creating non default VDC's requires Advanced Services license



Introduction

Terminology: Internal Interfaces

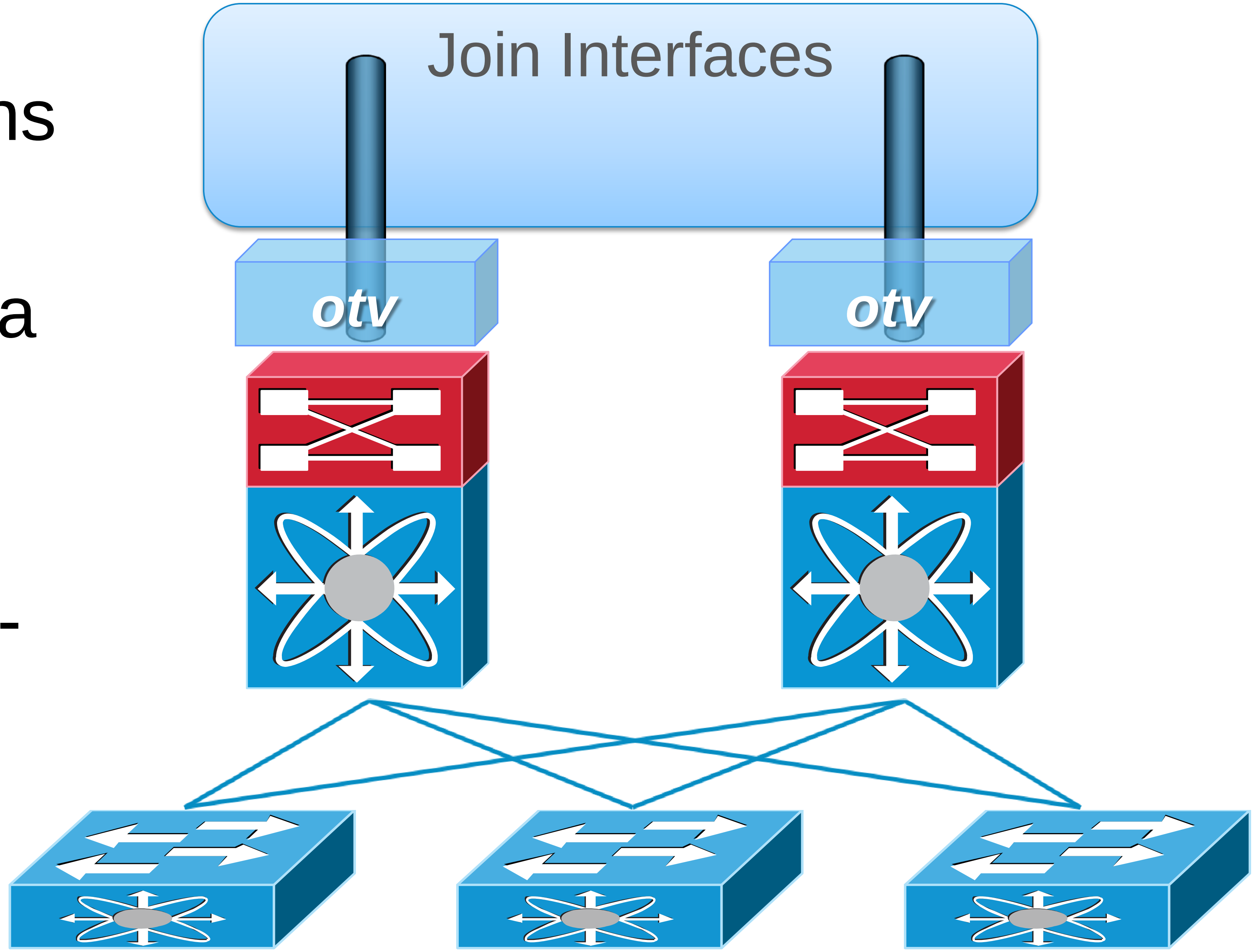
- Regular layer 2 interfaces facing the site
- No OTV configuration required
- Currently supported only on M-series modules



Introduction

Terminology: Join Interface

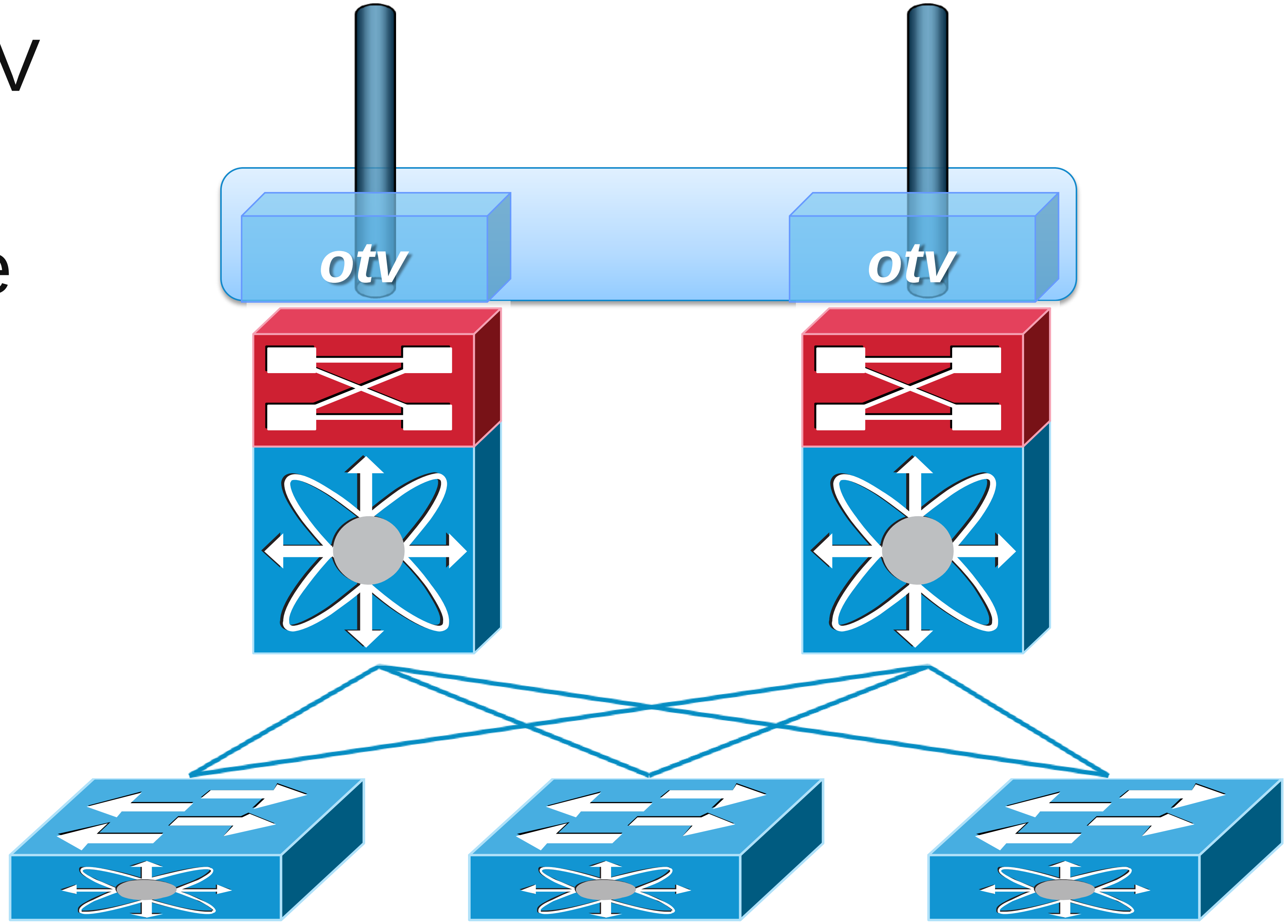
- Uplink on Edge device that joins the Overlay
- Forwards OTV control and data traffic
- Layer 3 interface
- Currently supported only on M-series modules



Introduction

Terminology: Overlay Interface

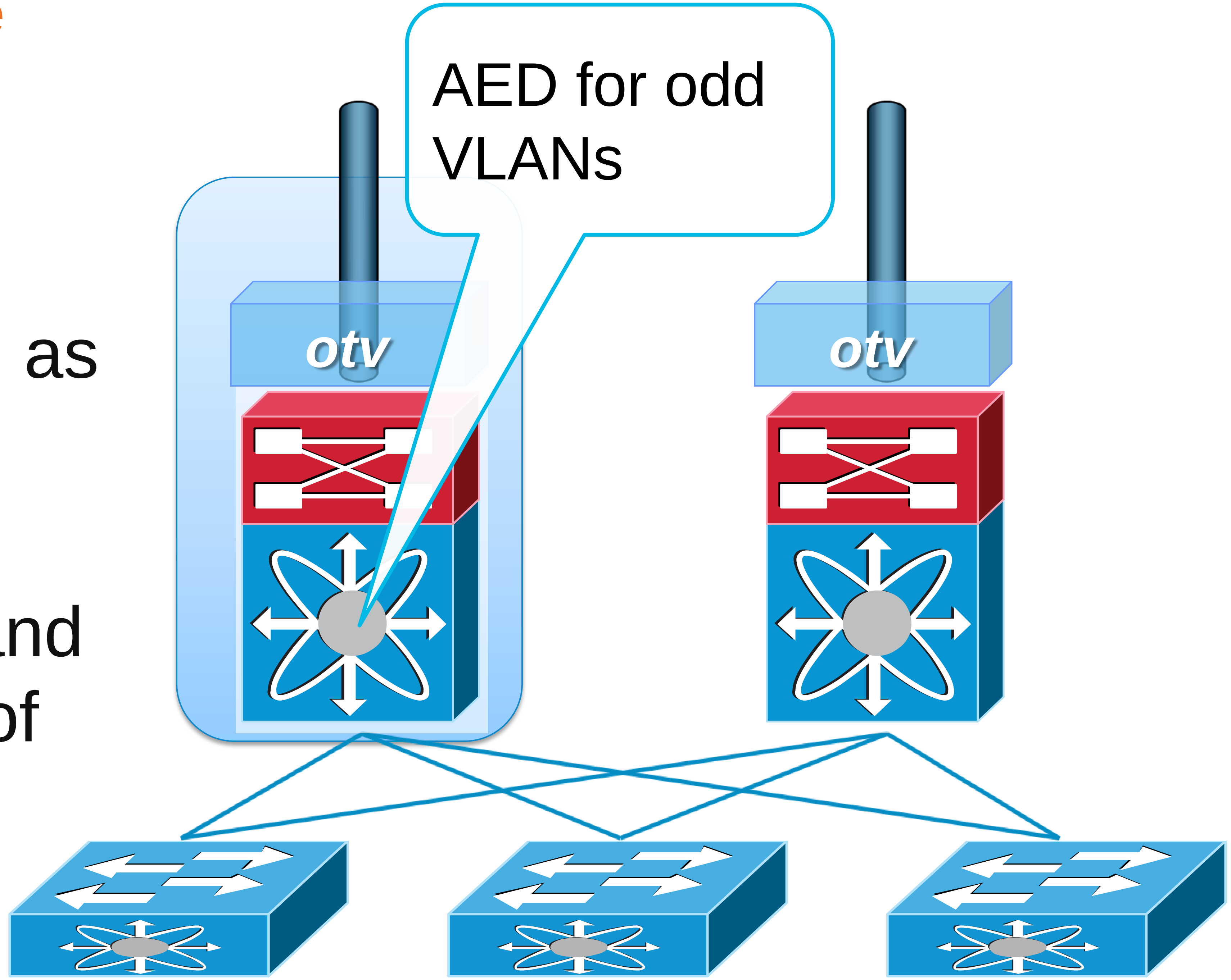
- Virtual Interface where the OTV configurations are applied
- Multi-access multicast-capable interface
- Encapsulates Layer 2 frames



Introduction

Terminology: Authoritative Edge Device

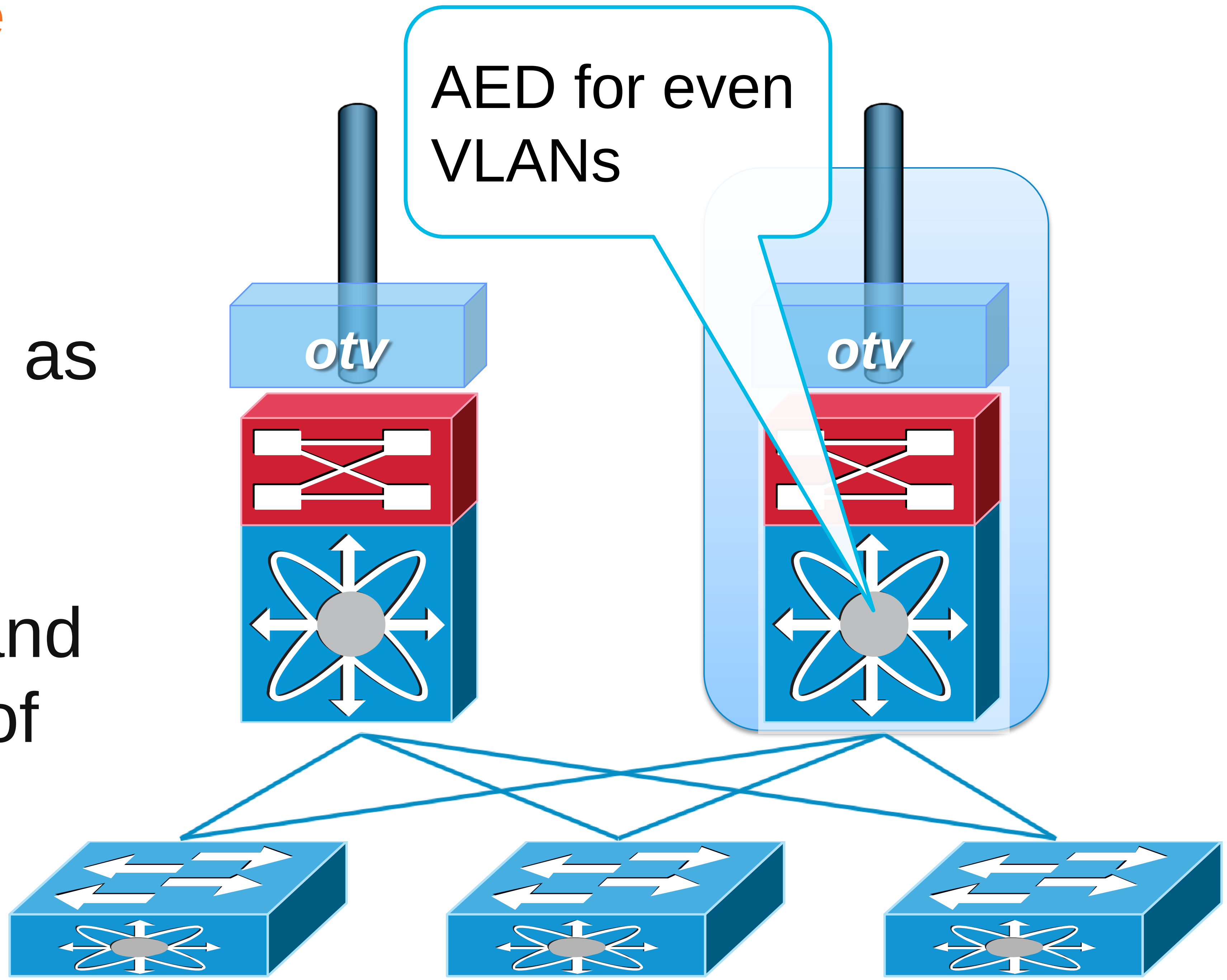
- OTV supports multiple edge devices per site
- A single OTV device is elected as AED on a per-vlan basis
- The AED is responsible for advertising MAC reachability and forwarding traffic into and out of the site for its VLANs



Introduction

Terminology: Authoritative Edge Device

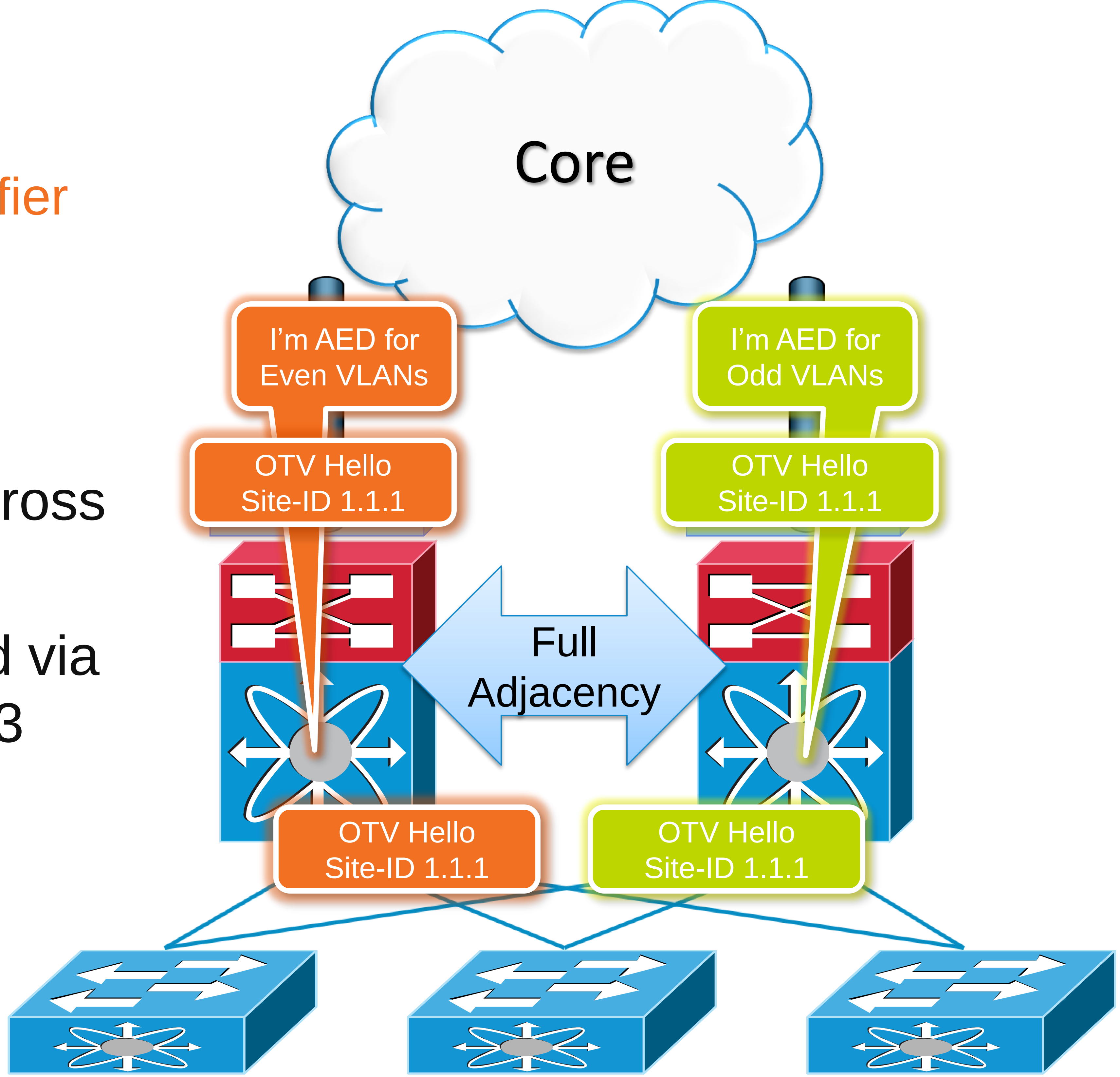
- OTV supports multiple edge devices per site
- A single OTV device is elected as AED on a per-vlan basis
- The AED is responsible for advertising MAC reachability and forwarding traffic into and out of the site for its VLANs



Introduction

Terminology: Site VLAN and Site Identifier

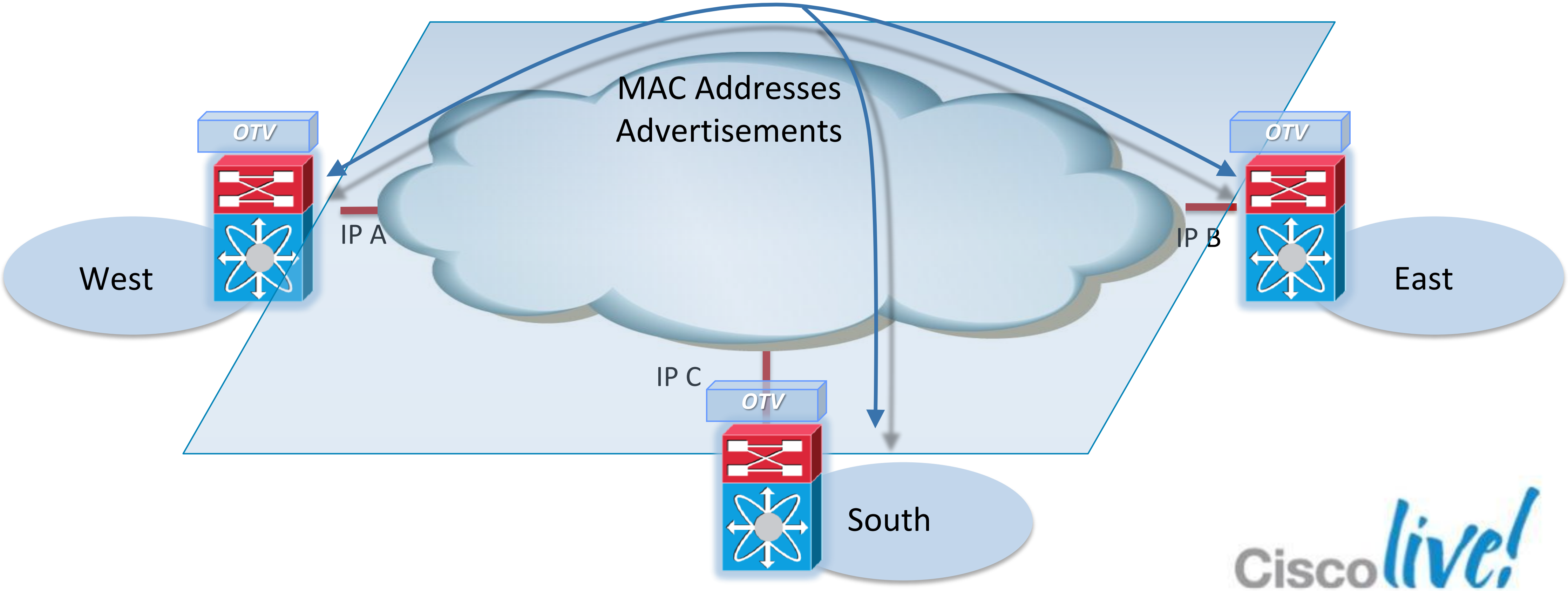
- 5.2(1) added **Dual Site Adjacency**
  1. **Site Adjacency** established across the site vlan
  2. **Overlay Adjacency** established via the Join interface across Layer 3 network



OTV Control Plane

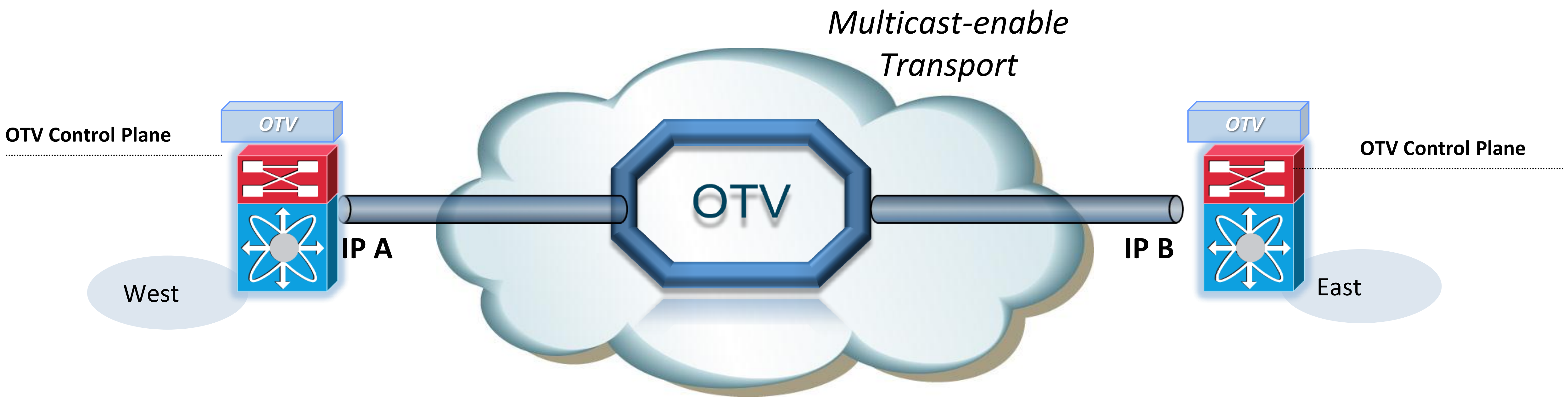
Building the MAC Tables

- **No unknown unicast flooding**
- **Control Plane Learning with proactive MAC advertisement**
- Background process with no specific configuration
- IS-IS used between OTV Edge Devices



OTV Control Plane

Neighbour Discovery (over Multicast Transport)



Mechanism

- Edge Devices (EDs) join an multicast group in the transport, as they were hosts (no PIM on EDs)
- OTV hellos and updates are encapsulated in the multicast group

End Result

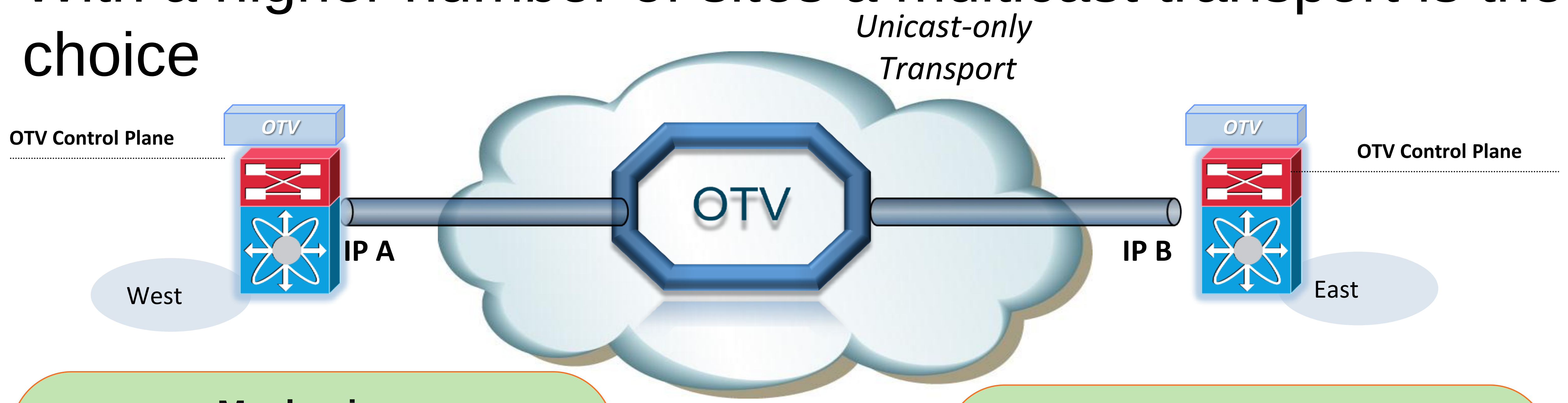
- Adjacencies are maintained over the multicast group
- A single update reaches all neighbours

Release 5.2 and above

OTV Control Plane

Neighbour Discovery (Unicast-only Transport)

- Ideal for connecting a small number of sites
- With a higher number of sites a multicast transport is the best choice



Mechanism

- Edge Devices (EDs) register with an *"Adjacency Server"* ED
- EDs receive a full list of Neighbours (oNL) from the *AS*
- OTV hellos and updates are encapsulated in IP and ***unicast*** to each neighbour

End Result

Neighbour Discovery is automated by the *"Adjacency Server"*

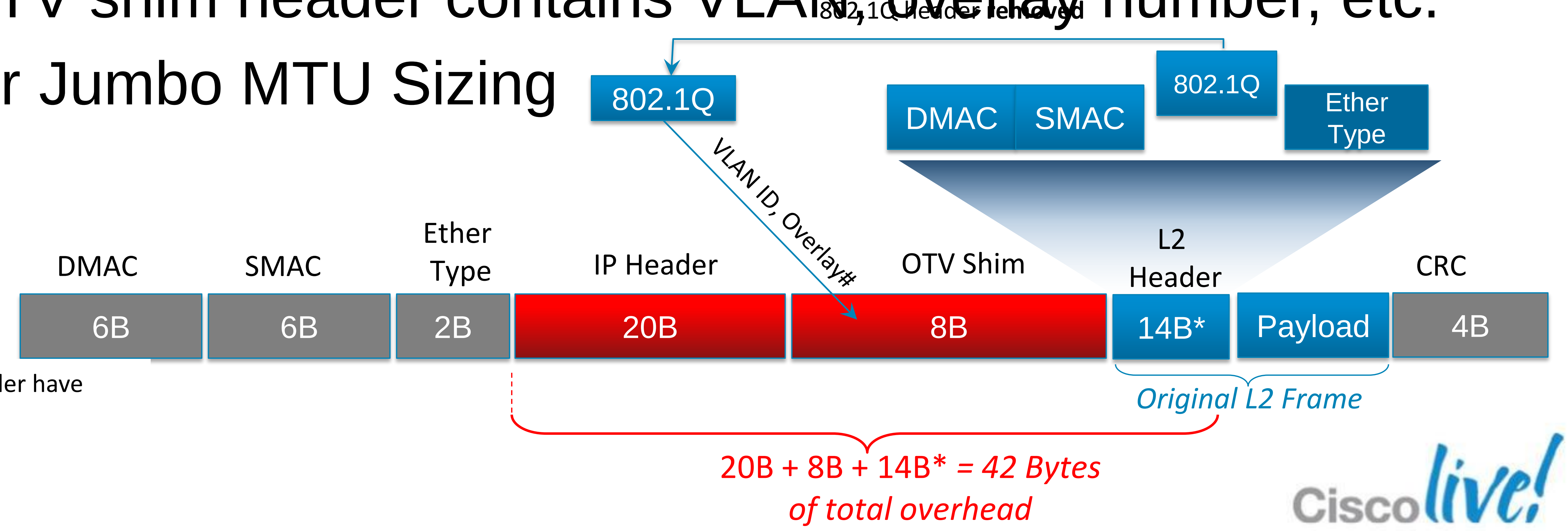
All signaling must be replicated for each neighbour

Data traffic must also be replicated at the head-end

OTV Data Plane

Encapsulation

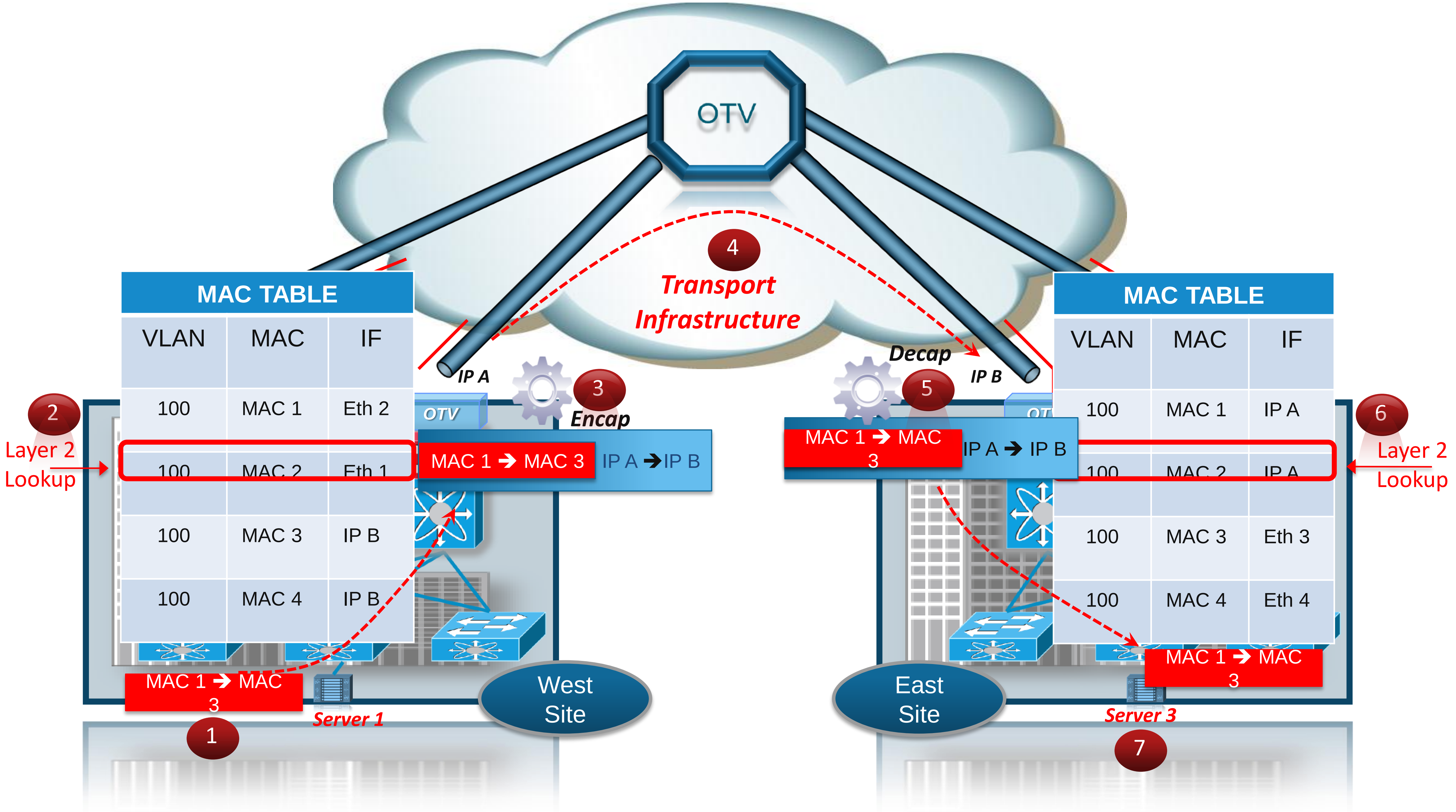
- **42 Bytes** overhead to the packet IP MTU size
  - Outer IP + OTV Shim - Original L2 Header (w/out the .1Q header)
- 802.1Q header is **removed** and the VLAN field copied over to the OTV shim header
- Outer OTV shim header contains VLAN, overlay number, etc.
- Consider Jumbo MTU Sizing



* The 4 Bytes of .1Q header have already been removed

OTV Data Plane

Inter-Site Packet Flow



MAC TABLE (West Site)

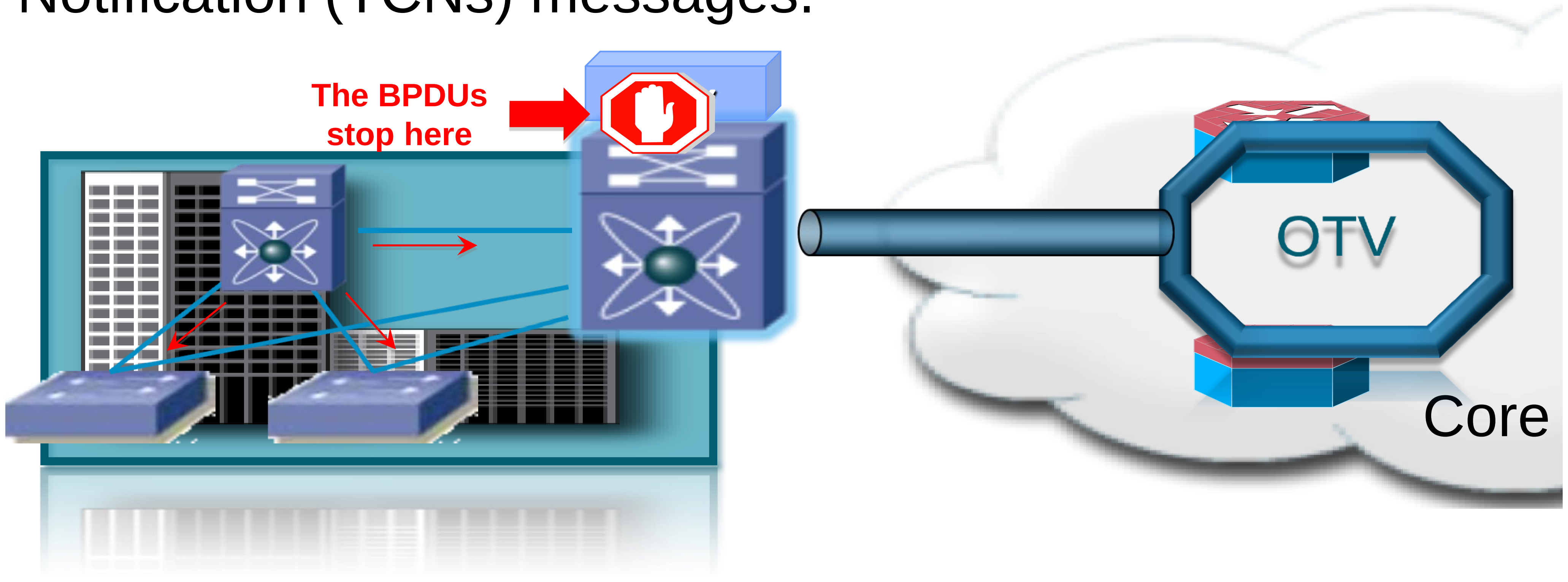
VLAN	MAC	IF
100	MAC 1	Eth 2
100	MAC 2	Eth 1
100	MAC 3	IP B
100	MAC 4	IP B

MAC TABLE (East Site)

VLAN	MAC	IF
100	MAC 1	IP A
100	MAC 2	IP A
100	MAC 3	Eth 3
100	MAC 4	Eth 4

STP BPDU Handling

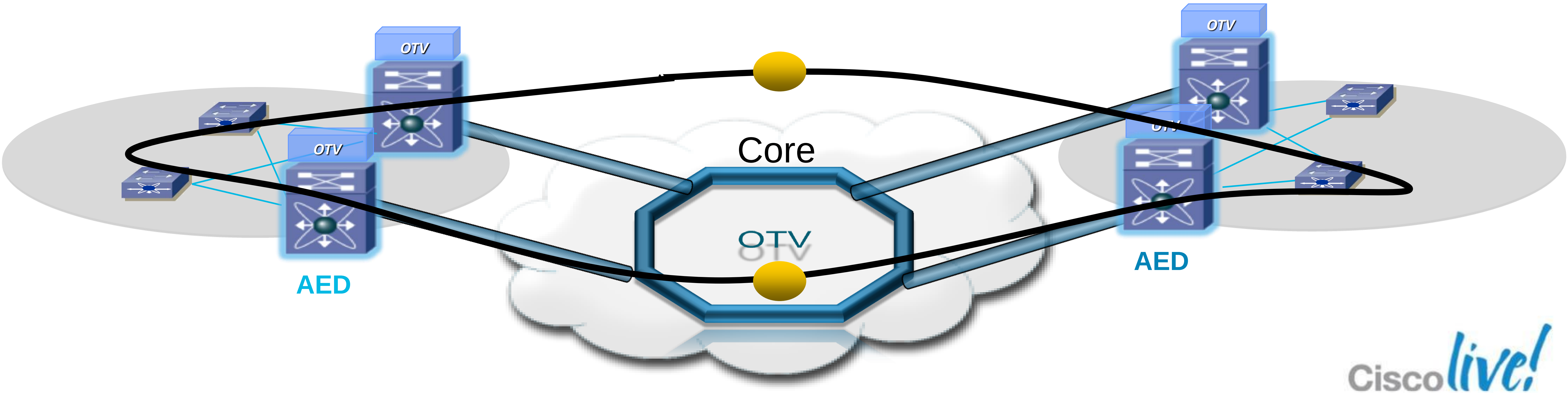
- When STP is configured at a site, an Edge Device will send and receive BPDUs on the **internal interfaces**.
- An OTV Edge Device will not originate or forward BPDUs on the overlay network.
- An OTV Edge Device can become (but it is not required to) a root of one or more spanning trees within the site.
- An OTV Edge Device will take the typical action when receiving Topology Change Notification (TCNs) messages.



Handling Data-plane Loop Prevention

Broadcast/Multicast Handling

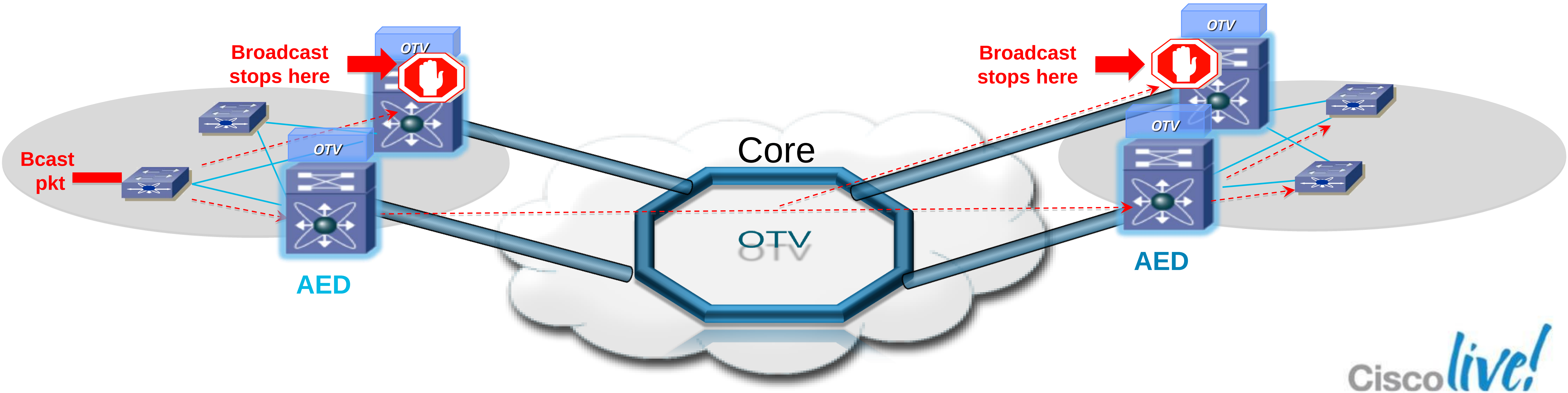
- Brodcast/M-cast packets reach all Edge Devices within a site.
- **The AED for the VLAN is the only Edge Device that forwards b-cast/ m-cast packets onto the overlay network**
- The b-cast/m-cast packet is replicated to all the Edge Devices on the overlay.
- Only the AED at each remote site will forward the packet from the overlay onto the site.
- Once sent into the site, the b-cast/m-cast packet is replicated per regular switching



Multi-homing

AED and Broadcast/Multicast Handling

- Broadcast/M-cast packets reach all Edge Devices within a site.
- **The AED for the VLAN is the only Edge Device that forwards b-cast/ m-cast packets onto the overlay network**
- The b-cast/m-cast packet is replicated to all the Edge Devices on the overlay.
- Only the AED at each remote site will forward the packet from the overlay onto the site.
- Once sent into the site, the b-cast/m-cast packet is replicated per regular switching

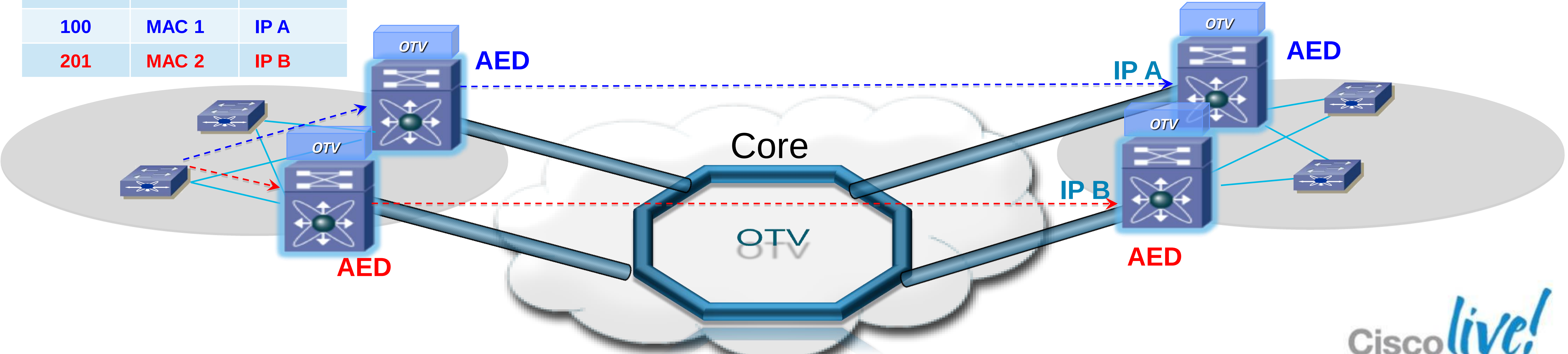


Multi-homing

AED and Unicast Forwarding

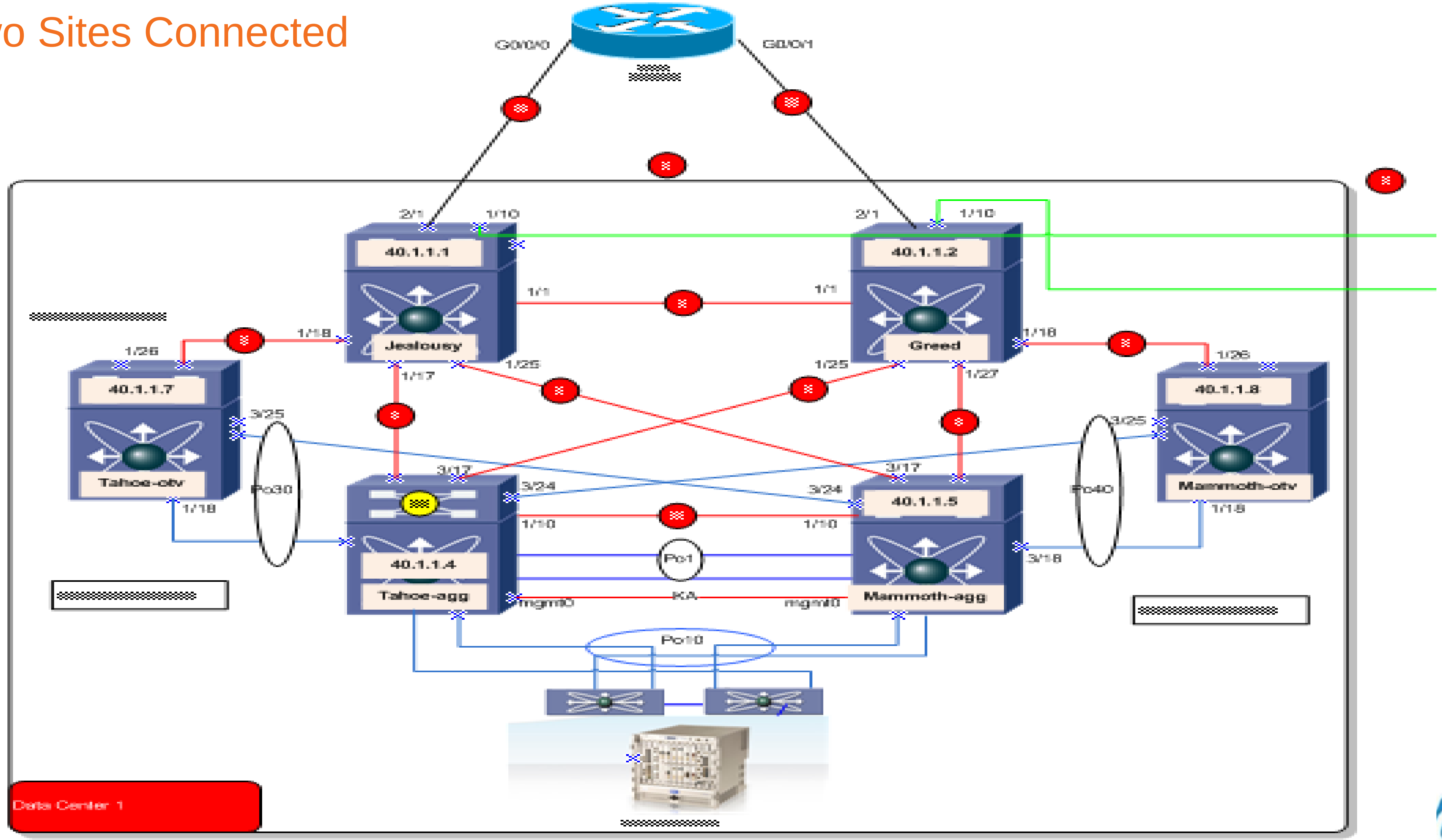
- One AED is elected for each VLAN on each site
- Different AEDs can be elected for each VLAN to balance traffic load
- Only the AED forwards unicast traffic to and from the overlay
- Only the AED advertises MAC addresses for any given site/VLAN
- Unicast routes will point to the AED on the corresponding remote site/VLAN

MAC TABLE		
VLAN	MAC	IF
100	MAC 1	IP A
201	MAC 2	IP B



OTV Use Case

Two Sites Connected



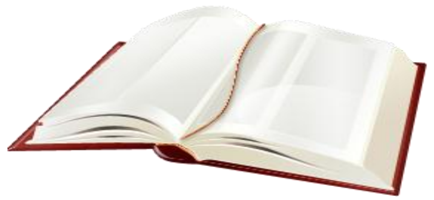
OTV Summary

- STP Isolation: BPDUs are not forwarded over the overlay
- Automated Multi-homing support
- Optimal Multicast Replication
- Control-plane MAC based learning and forwarding
- Simplified Configuration
- Operational Simplicity
- IP Based / Transport Agnostic (IP/MPLS)
- End-to-End loop prevention

Data Centre Interconnect

Agenda

- Mobility and Virtualisation in the Data Centre
- LAN Extension Deployment Scenarios
  - –Ethernet Based Solutions
  - –MPLS Based Solutions
    - EoMPLS
    - VPLS
    - A-VPLS
    - EVPN
- Overlay Transport Virtualisation (OTV)
- Encryption
- IP Mobility without LAN Extension
- Path optimisation
- VXLAN
- Summary and Conclusions
- Q&A



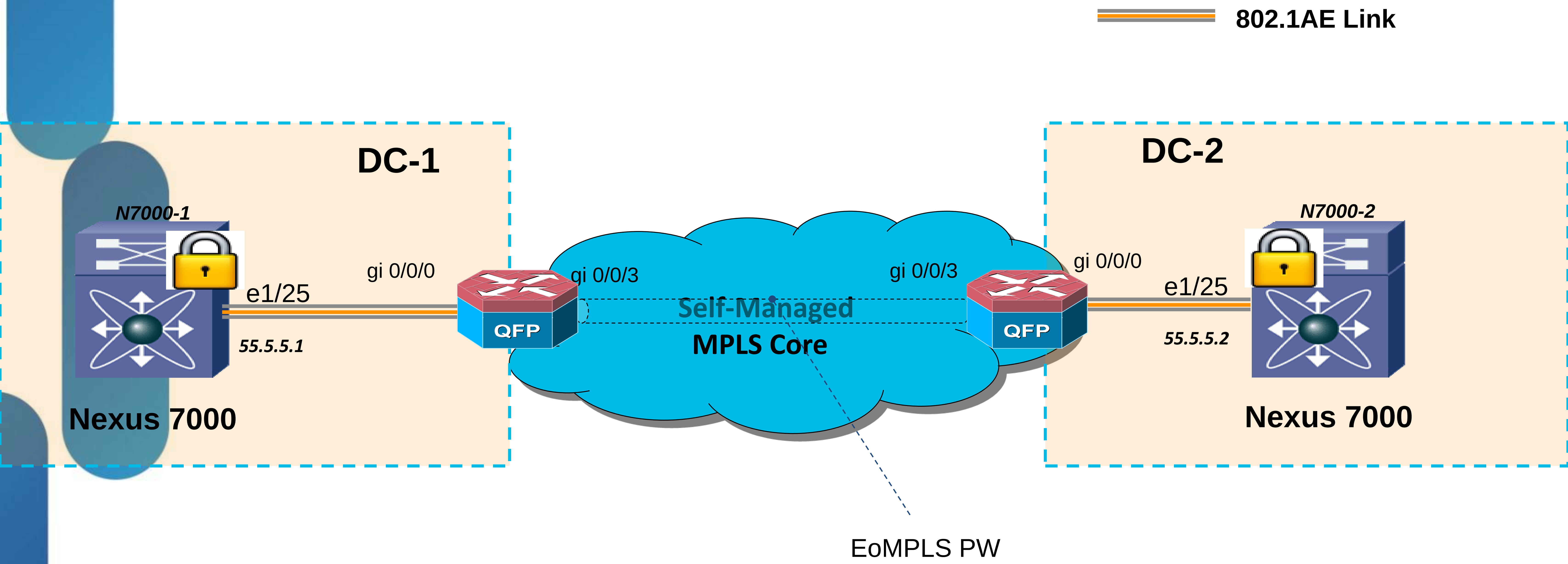
= For your Reference

Encryption

Point-to-Point Encryption Solution

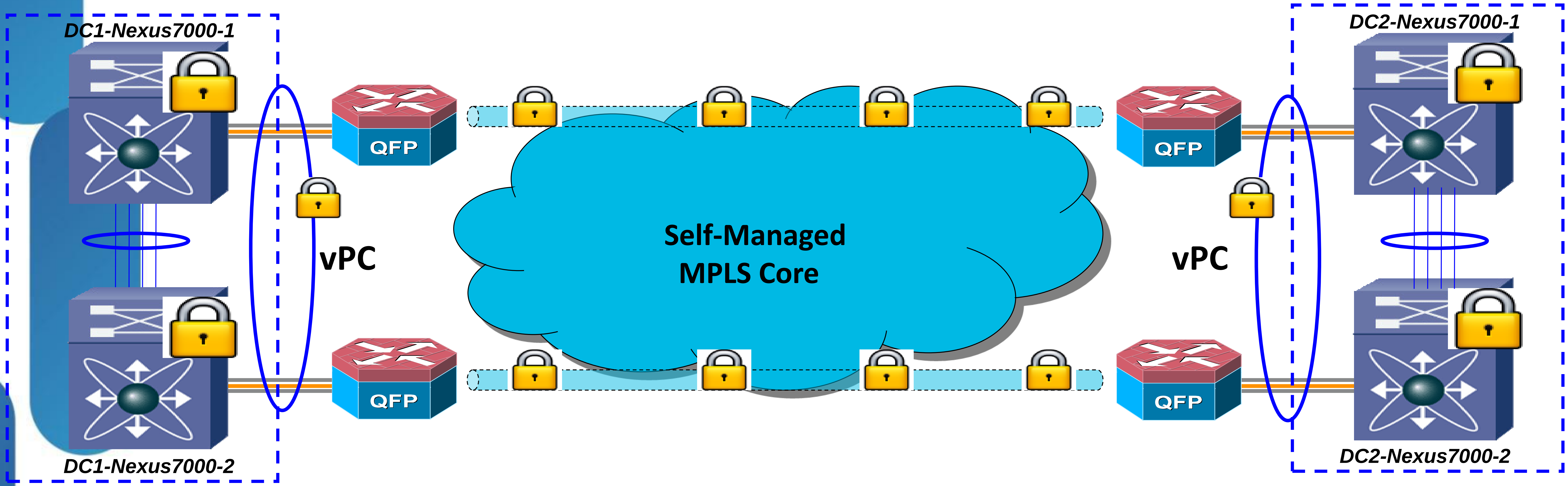
Nexus 7000 Trustsec can be used to secure data across remote data-Centre if Layer 2 and BPDU transparency is ensured (e.g. dark fibre or DWDM transport).

Encryption Solution



* Remote port shutdown (ASR Only)

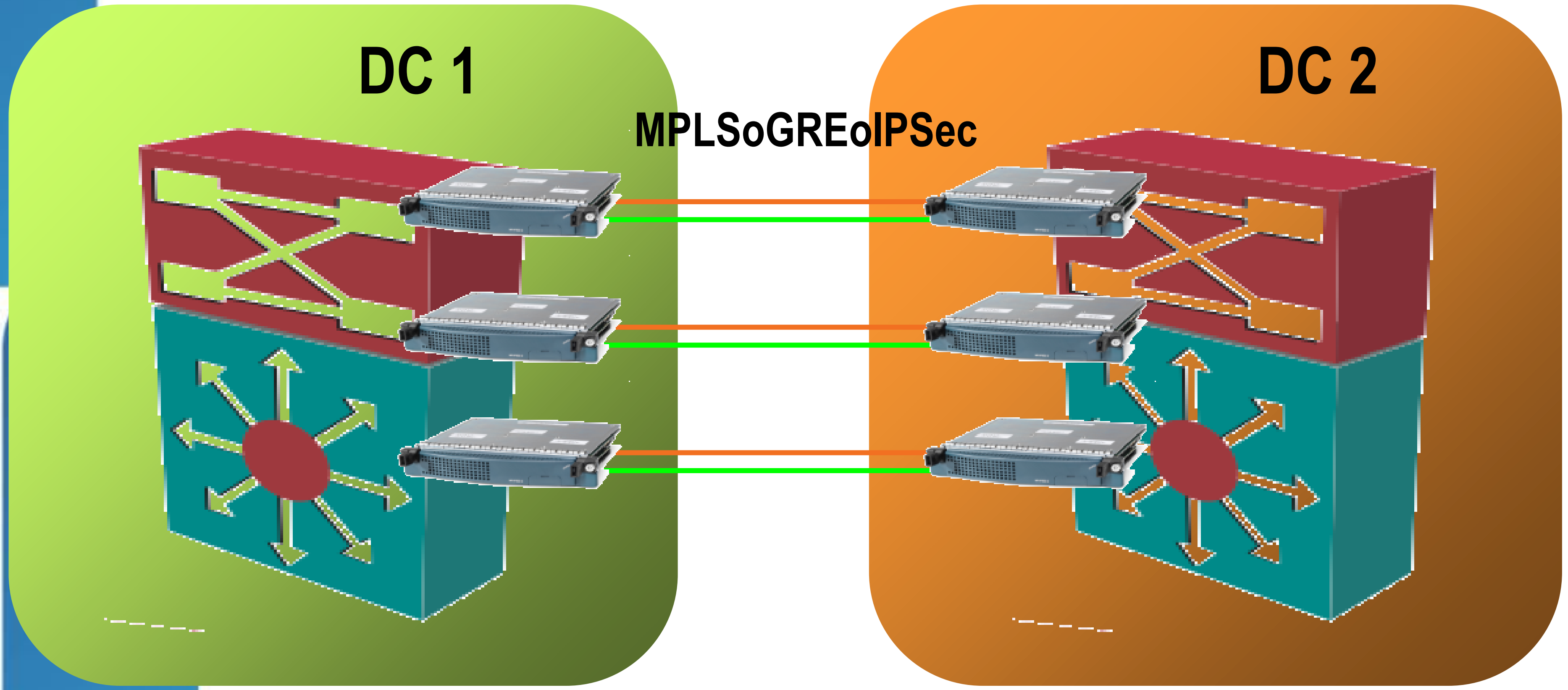
Nexus 7000 vPC Encryption Solution



* Remote port shutdown (ASR)

VSPA/ASR1000/ASA Solution Overview

DataCentre Interconnect with MPLSoGREoIPSec



- Leverage ECMP to load balance flows over multiple GRE/IPSec
- Duplicate tunnels per VSPA allow redundant 10GE links to be provisioned
- Inherent crypto engine HA: Traffic will rebalance in the event of a VSPA outage

Solution Objective

- Provide a high speed Layer 2 connection between two or more DCs.. Two or more redundant links are used between the DCs.

VSPA Performance

- Three VSPAs can drive a 10 GE link with IMIX traffic. Single chassis can encrypt three 10 GE links at IMIX rates.

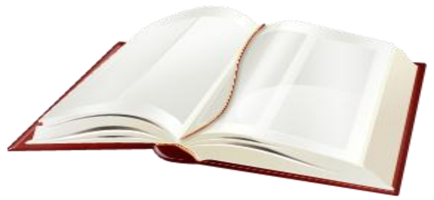
ASR-1000 Performance

- ASR1000-ESP5-1.8Gbps IPSec
- ASR1000-ESP10-4Gbps IPSec
- ASR1000-ESP20-8Gbps IPSec
- ASR1006-2/ESP20-16Gbps IPSec
- ASR1006-2/ESP40 – 25.8Gbps IPSec
- ASA-5585-X Performance
- IPSec 5Gbps

Data Centre Interconnect

Agenda

- Mobility and Virtualisation in the Data Centre
- LAN Extension Deployment Scenarios
  - –Ethernet Based Solutions
  - –MPLS Based Solutions
    - EoMPLS
    - VPLS
    - A-VPLS
    - EVPN
- Overlay Transport Virtualisation (OTV)
- Encryption
- IP Mobility without LAN Extension
- Path optimisation
- VXLAN
- Summary and Conclusions
- Q&A

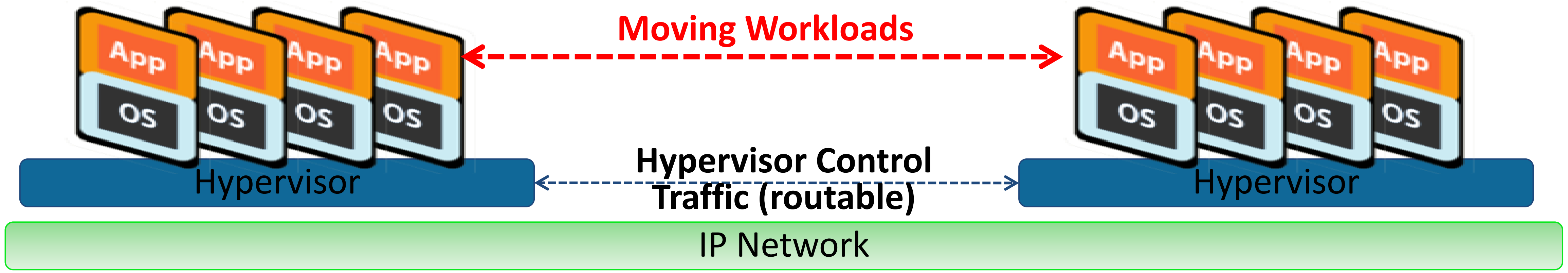


= For your Reference

IP Mobility without LAN Extension

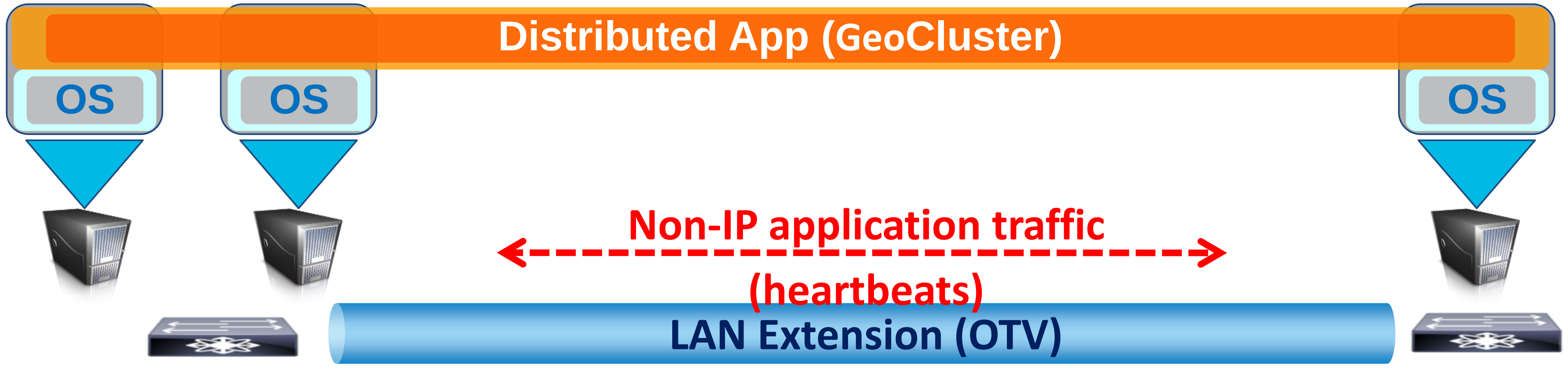
Moving vs. Distributing Workloads

Why do we really need LAN Extensions?



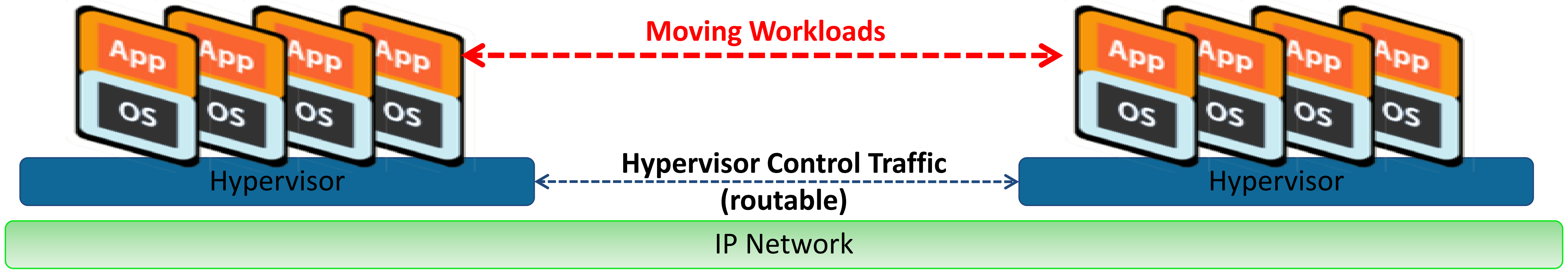
- **<u>Move workloads</u>** with IP mobility solutions: LISP Host Mobility
  – IP preservation is the real requirement (LAN extensions not mandatory)

- **<u>Distribute workloads</u>** with LAN extensions
  – Application High Availability with Distributed Clusters



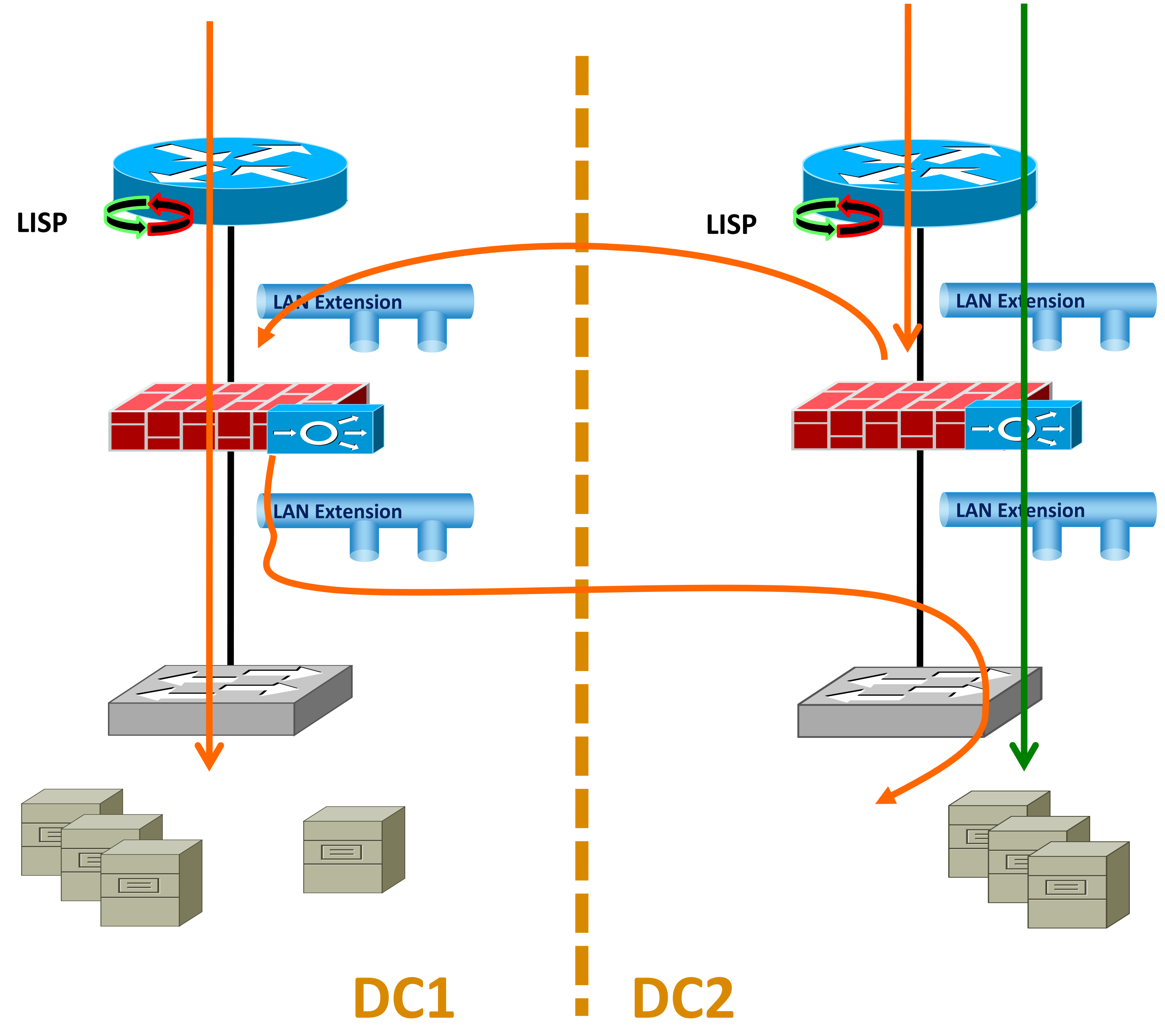
Live Moves or Cold Moves

- **<u>Live (hot) Moves</u>** preserve existing connections and state
  - e.g. vMotion, Cluster failover
  - Requires synchronous storage and network policy replication ➔ Distance limitations
- **<u>Cold Moves</u>** bring machines down and back up elsewhere
  - e.g. Site Recovery Manager
  - No state preservation: less constrained by distances or services capabilities



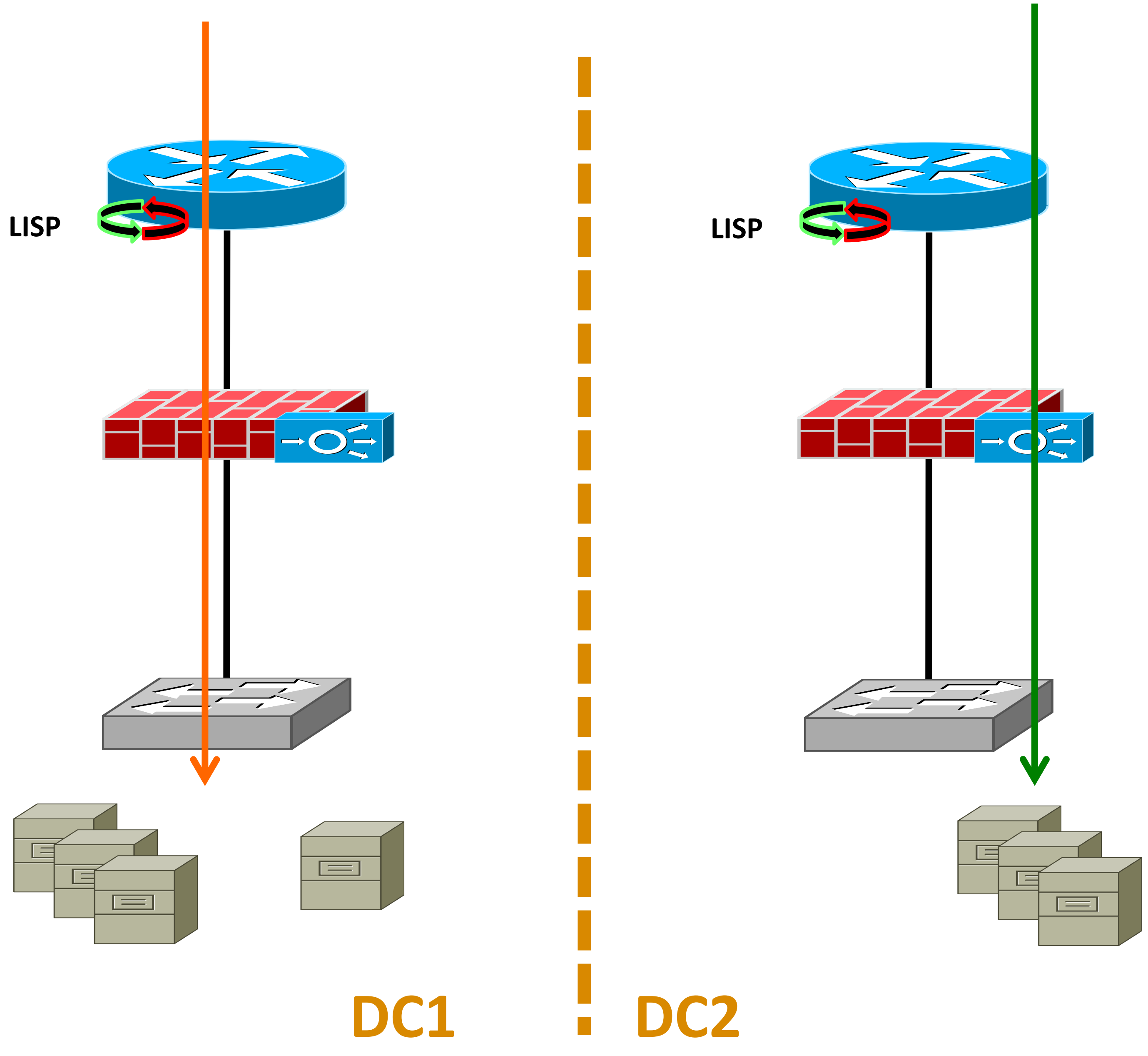
Mobility across PODs within a site or across different locations

Services - Live Moves



- Redirection of established flows:
  - Extended Clusters
  - Cluster or LISP based re-direction

Services – Cold Moves

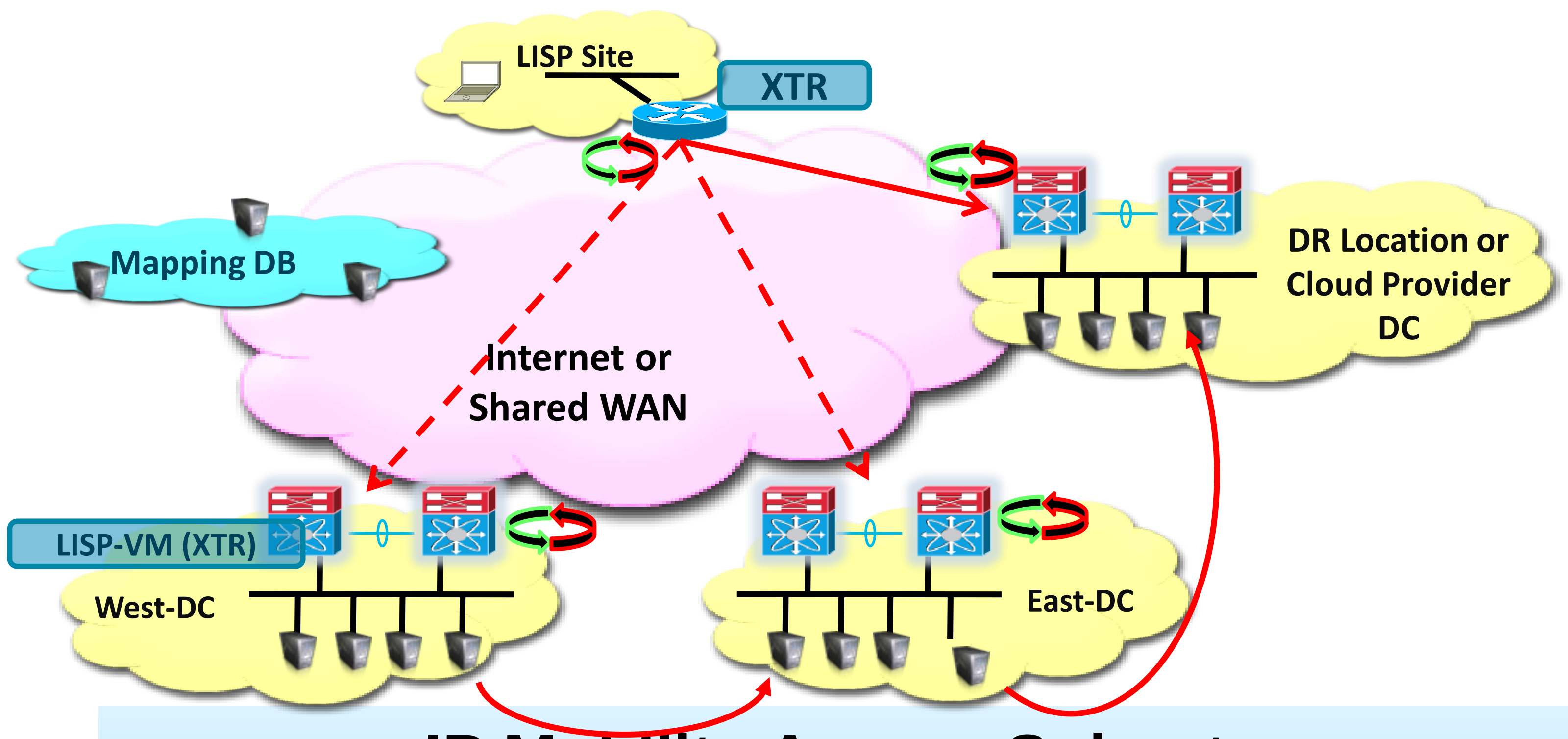


- IP preservation ➔ Uniform Policies

Established before the move
Established after the move

Host-Mobility Scenarios

Moves Without LAN Extension



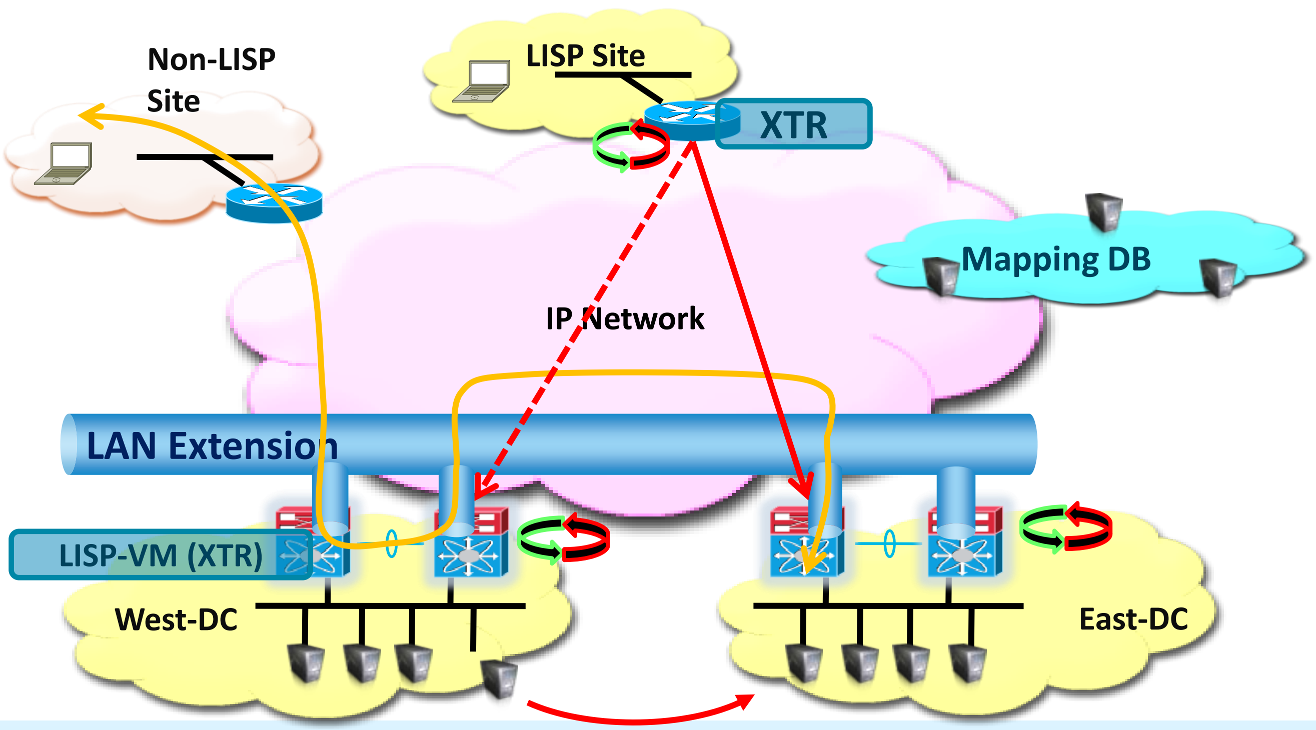
IP Mobility Across Subnets

Disaster Recovery

Cloud Bursting

Application Members in One Location

Moves With LAN Extension



Routing for Extended Subnets

Active-Active Data Centres

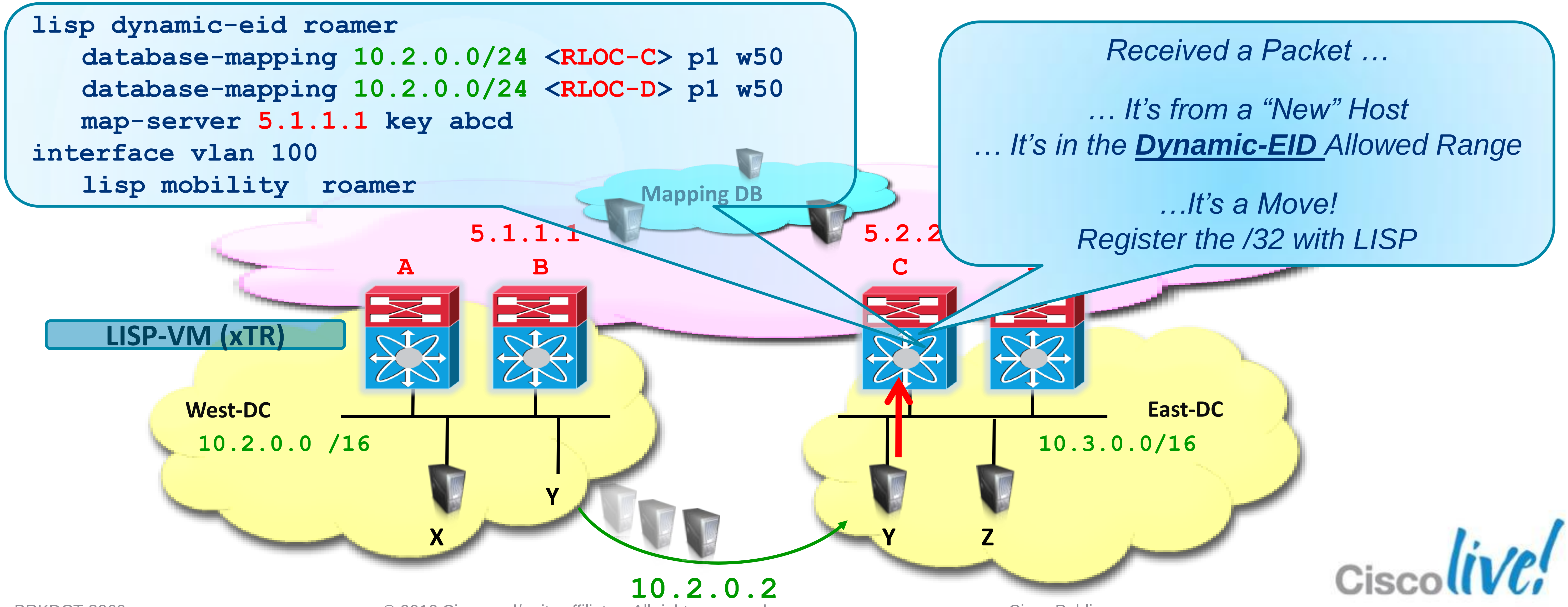
Distributed Clusters

Application Members Distributed (Broadcasts across sites)

LISP Host-Mobility – Move Detection

Monitor the source of Received Traffic

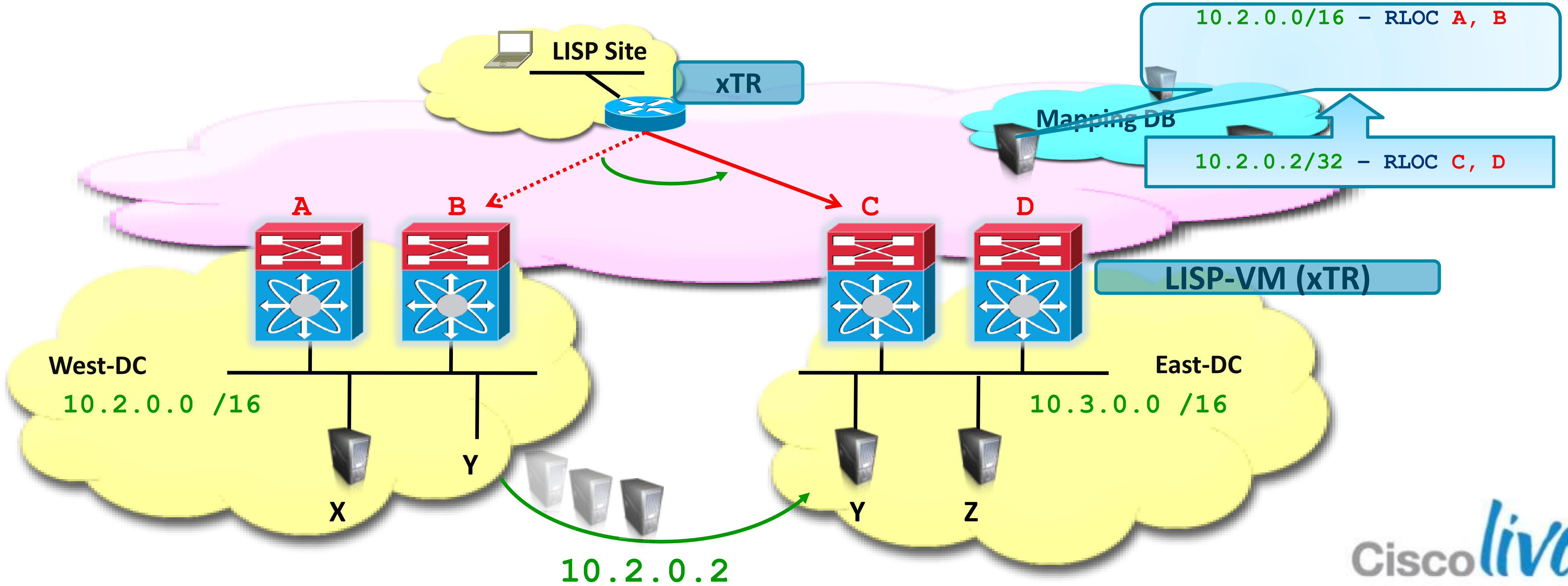
- The new xTR checks the source of received traffic
- Configured dynamic-EIDs define which prefixes may roam



LISP Host-Mobility – Traffic Redirection

Update Location Mappings for the Host System Wide

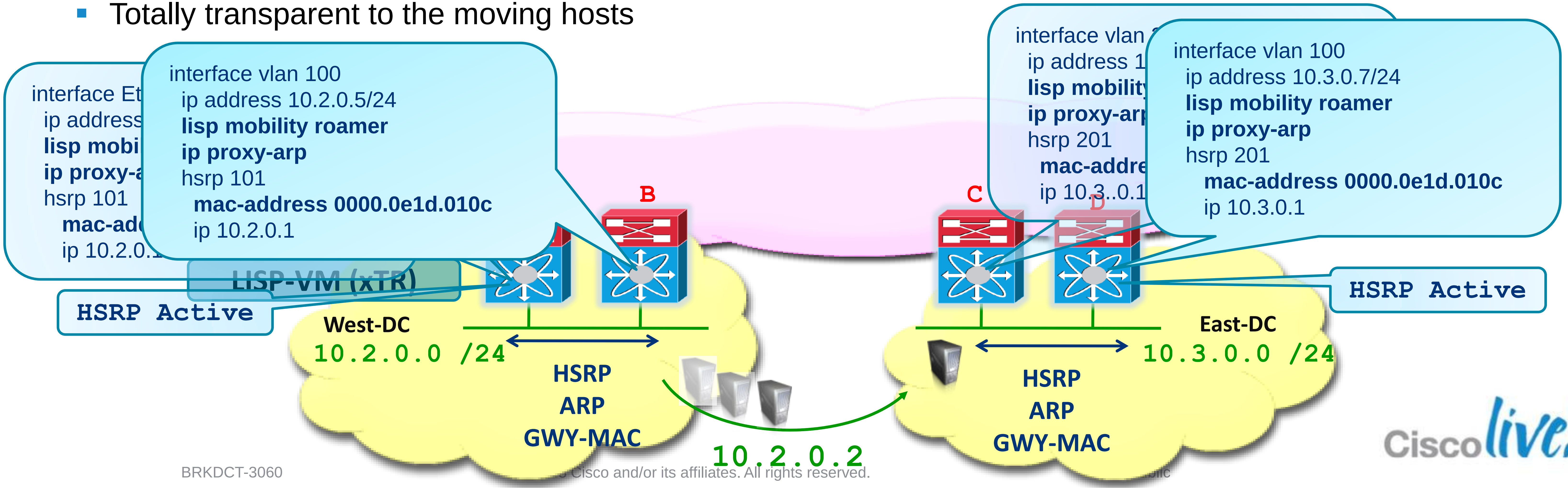
- When a host move is detected, updates are triggered:
  - The host-to-location mapping in the Database is updated to reflect the new location
  - The old ETR is notified of the move
  - ITRs are notified to update their Map-caches
- Ingress routers (ITRs or PITRs) now send traffic to the new location
- Transparent to the underlying routing and to the host



LISP Host-Mobility – First Hop Routing

No LAN Extension

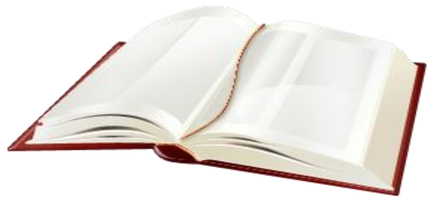
- SVI (Interface VLAN x) and HSRP configured as usual
  - Consistent GWY-MAC configured across all dynamic subnets
- The lisp mobility <dyn-eid-map> command enables proxy-arp functionality on the SVI
  - The LISP-VM router services first hop routing requests for both local and roaming subnets
- Hosts can move anywhere and always talk to a local gateway with the same MAC
- Totally transparent to the moving hosts



Data Centre Interconnect

Agenda

- Mobility and Virtualisation in the Data Centre
- LAN Extension Deployment Scenarios
  - –Ethernet Based Solutions
  - –MPLS Based Solutions
    - EoMPLS
    - VPLS
    - A-VPLS
    - EVPN
- Overlay Transport Virtualisation (OTV)
- Encryption
- IP Mobility without LAN Extension
- Path optimisation
- VXLAN
- Summary and Conclusions
- Q&A



= For your Reference

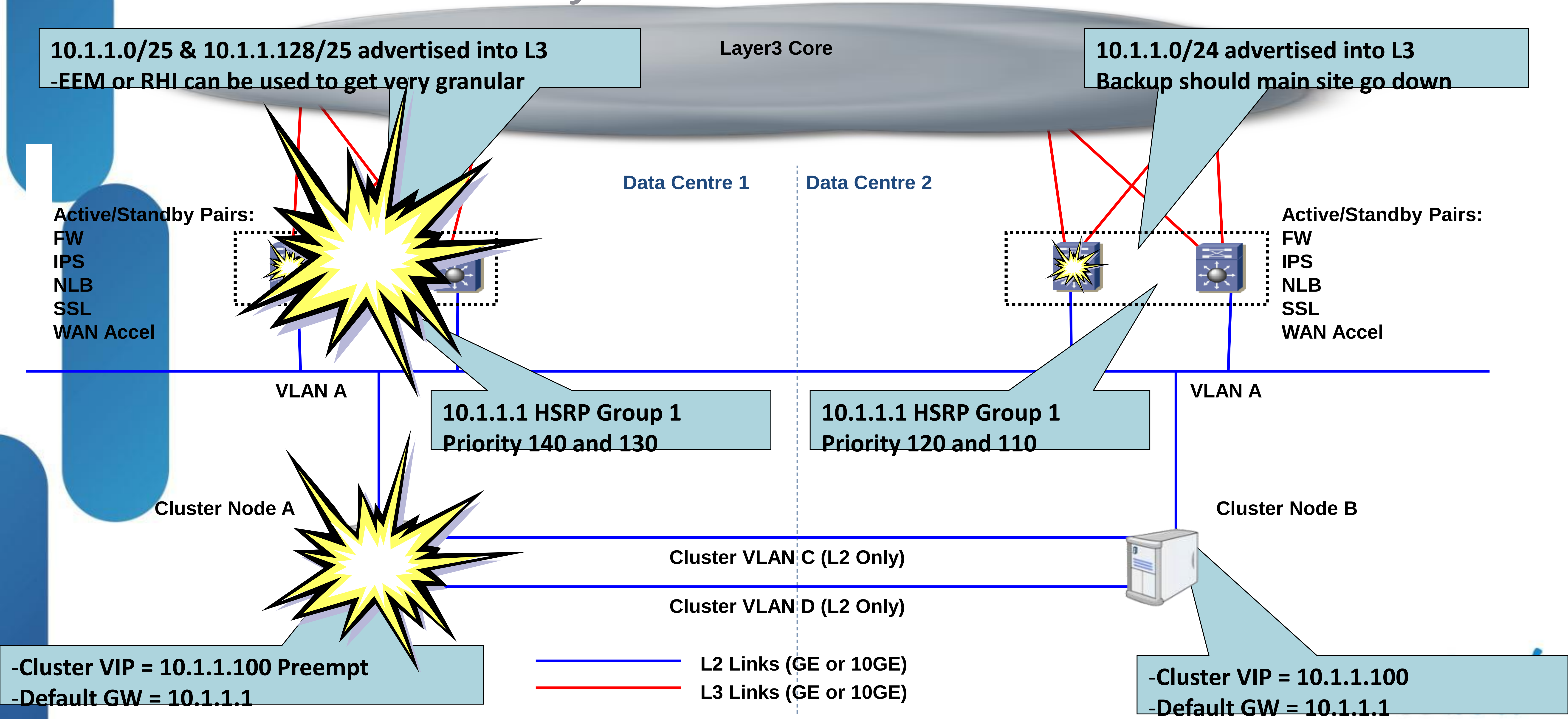
Flow Optimization and Symmetry
Site Selection and Inbound Flows
First Hop Outbound

Optimising Traffic Patterns and HA Design

- Many tradeoffs in understanding flows in multi-DC design
- Slides that follow are a specific recommendation that meets the following requirements:
  - Minimise inter-DC traffic to maintenance/failure scenario's
  - Ability to extend clusters between locations (OS, FS, DB, VMware DRS, etc.)
  - Desire to keep flows symmetric in/out of a location for DC services (FW, LB, IPS, WAAS, etc.)
  - Site failure will allow failover, with IP mobility to resolve caching issues
  - Single points of failure in gear won't cause site failover
  - Indicate a location preference for a service to the Layer 3 network
  - If broadcast storm in DC, limit impacts to other DCs
  - If DCI Layer 2 adjacency fails
  - Ability to connect to services in both DC locations (active/active per application)
  - DNS to round-robin clients to DC
  - Allow backup server farms with same service VIP (for backup connections on site fail)
  - Localised HSRP (egress)
  - Inbound traffic draw via LISP (ingress)
- This is a solution in production at some customers

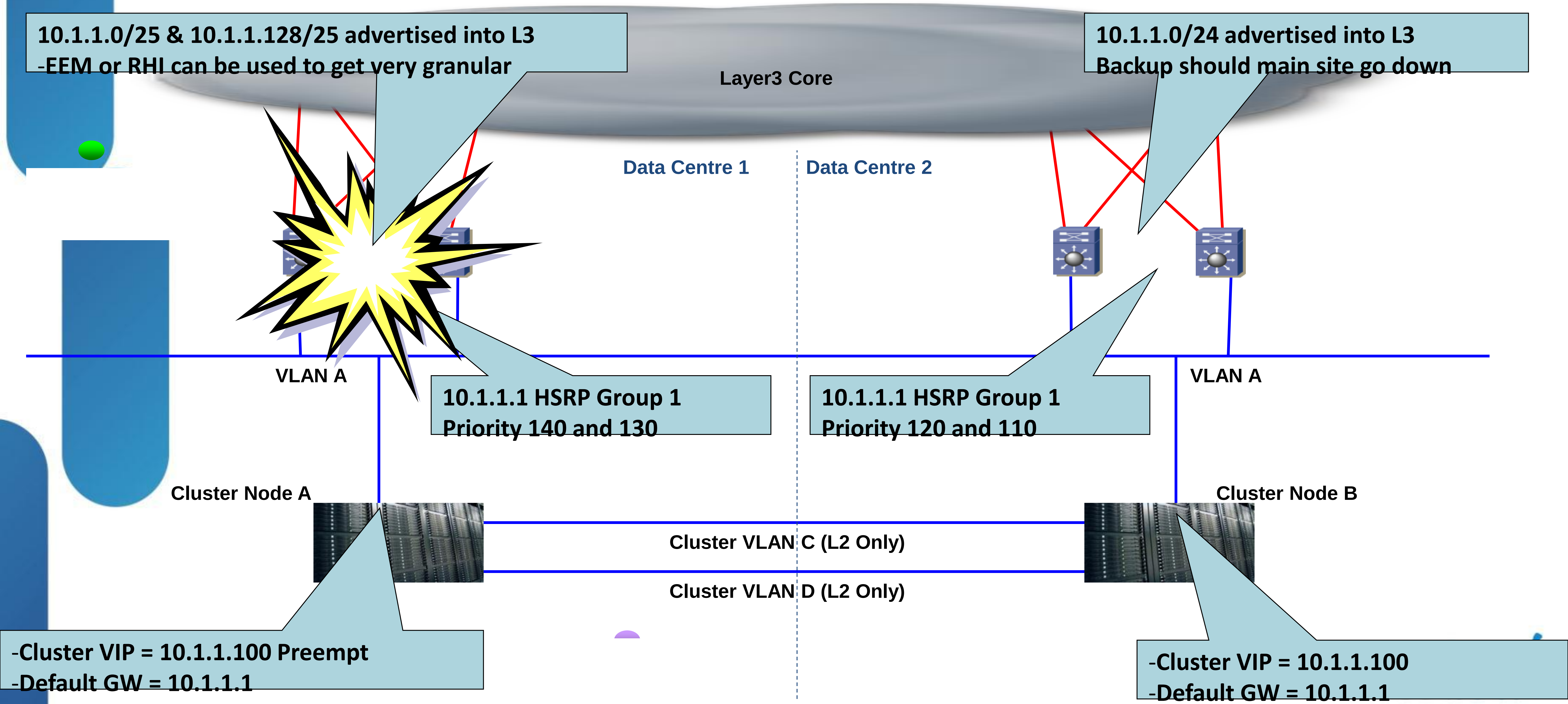
Sample Cluster – Service Normally in Left DC

Default Gateway Shared Between Sites



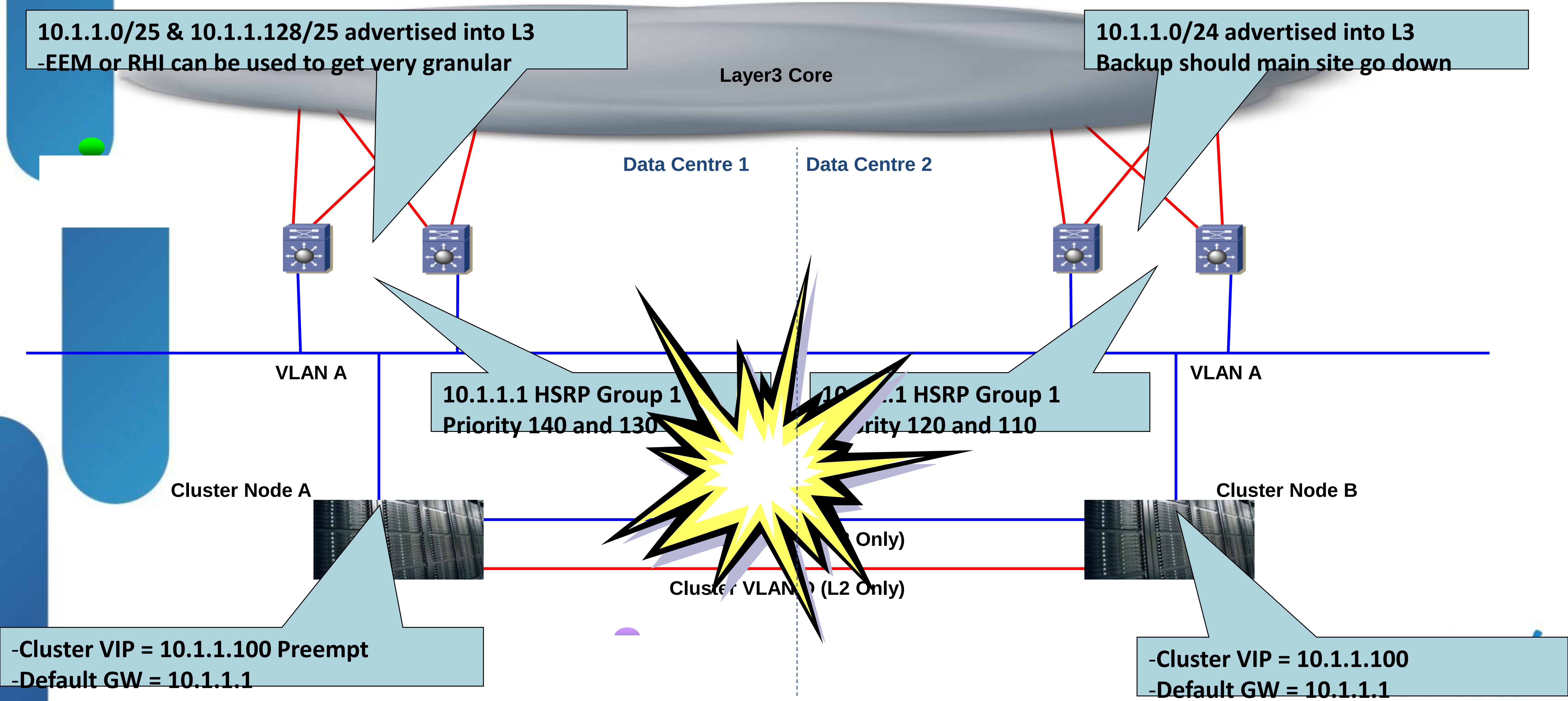
Sample Cluster – Broadcast Storm in Left DC

Broadcast, Multicast, Unknown Unicast



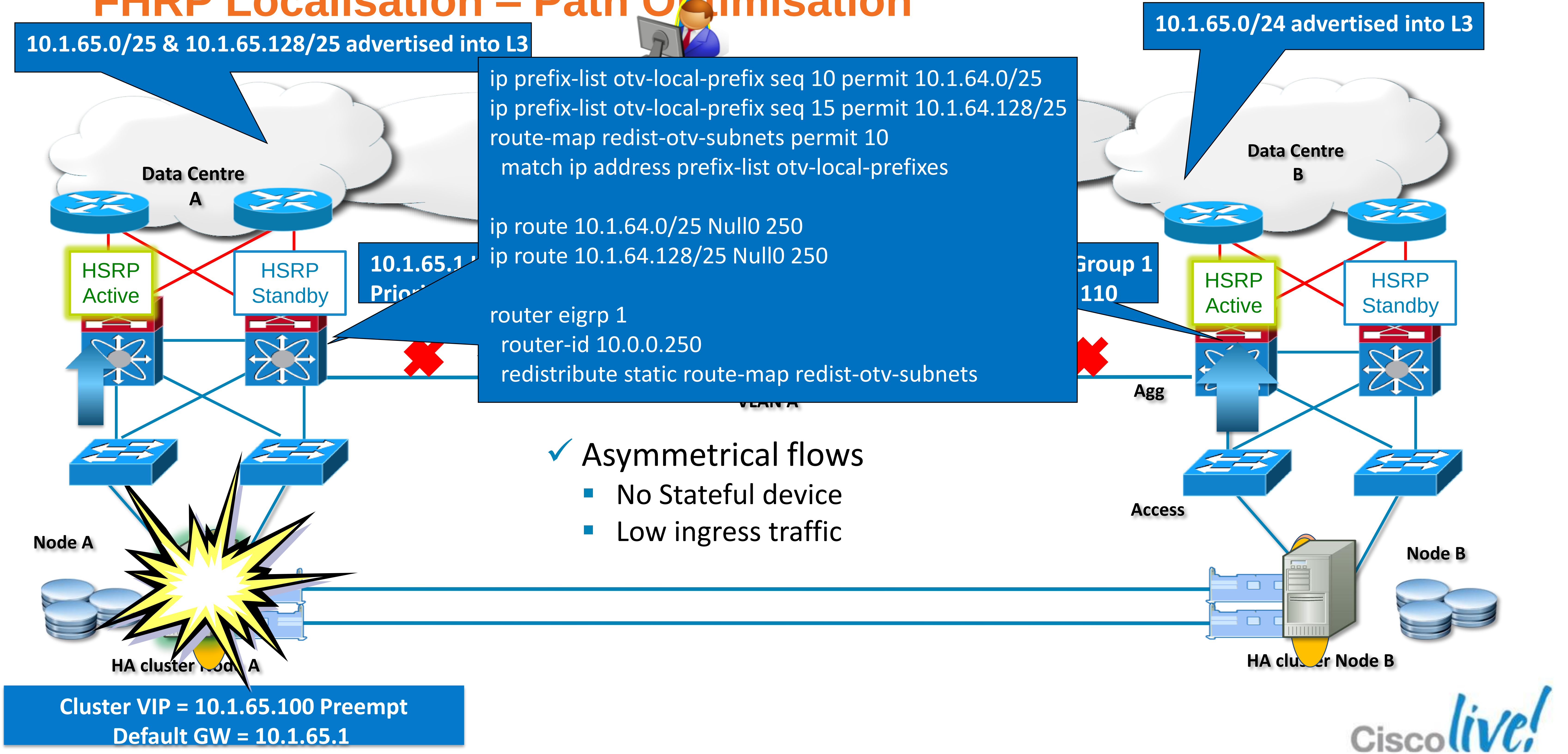
Sample Cluster – L2 Interconnect Failure

Broadcast, Multicast, Unknown Unicast



Sample Cluster - Primary Service in Left DC

FHRP Localisation – Path Optimisation

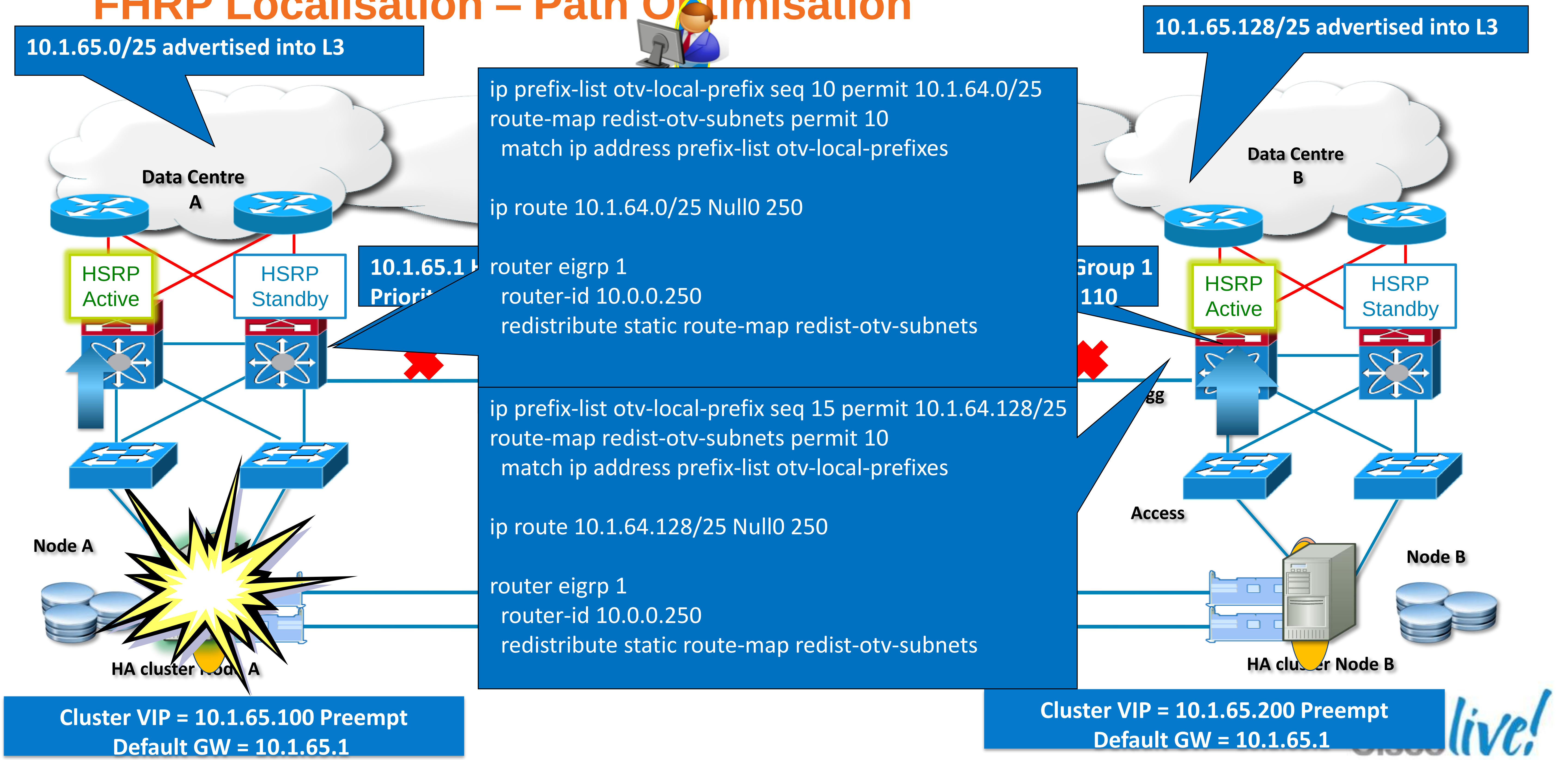



```
ip prefix-list otv-local-prefix seq 10 permit 10.1.64.0/25
ip prefix-list otv-local-prefix seq 15 permit 10.1.64.128/25
route-map redist-otv-subnets permit 10
  match ip address prefix-list otv-local-prefixes

ip route 10.1.64.0/25 Null0 250
ip route 10.1.64.128/25 Null0 250

router eigrp 1
  router-id 10.0.0.250
  redistribute static route-map redist-otv-subnets
```

✓ Asymmetrical flows
- No Stateful device
- Low ingress traffic

Sample Cluster – Active / Active DC

FHRP Localisation – Path Optimisation

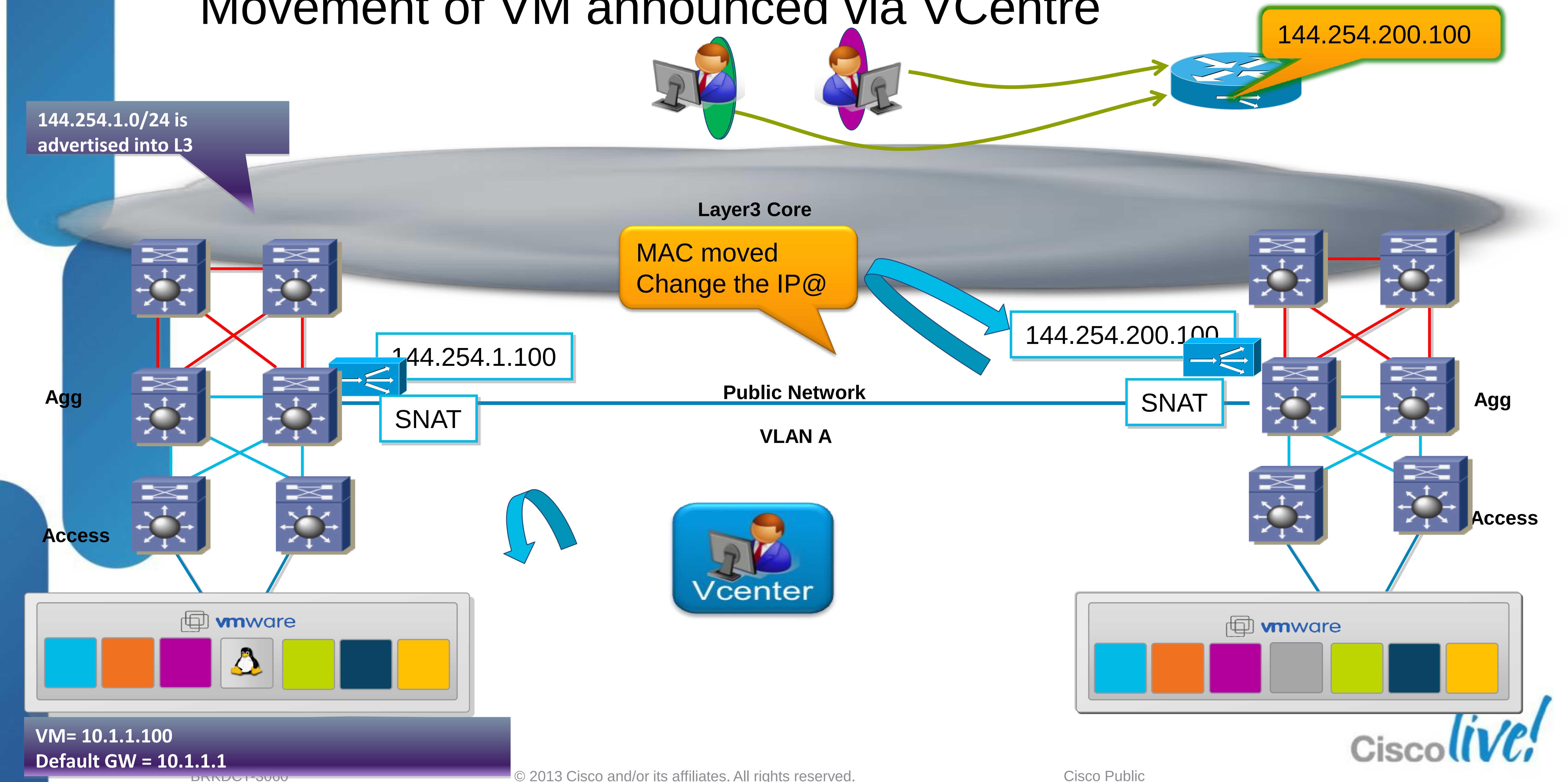



```
ip prefix-list otv-local-prefix seq 10 permit 10.1.64.0/25
route-map redist-otv-subnets permit 10
  match ip address prefix-list otv-local-prefixes

ip route 10.1.64.0/25 Null0 250

router eigrp 1
  router-id 10.0.0.250
  redistribute static route-map redist-otv-subnets
```

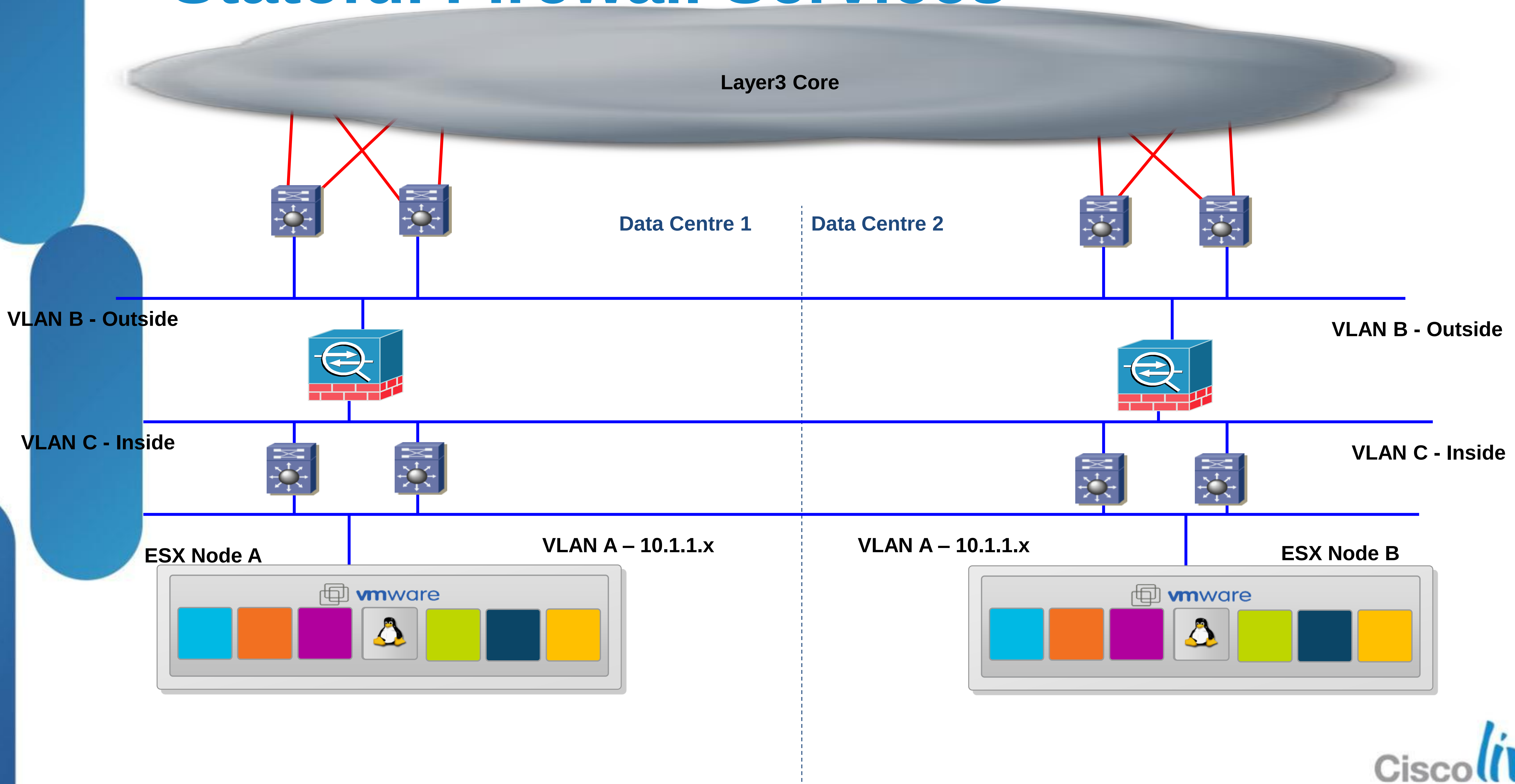
```
ip prefix-list otv-local-prefix seq 15 permit 10.1.64.128/25
route-map redist-otv-subnets permit 10
  match ip address prefix-list otv-local-prefixes

ip route 10.1.64.128/25 Null0 250

router eigrp 1
  router-id 10.0.0.250
  redistribute static route-map redist-otv-subnets
```

Primary Service in Left DC – DR/SRM

Movement of VM announced via VCentre



Stateful Firewall Services



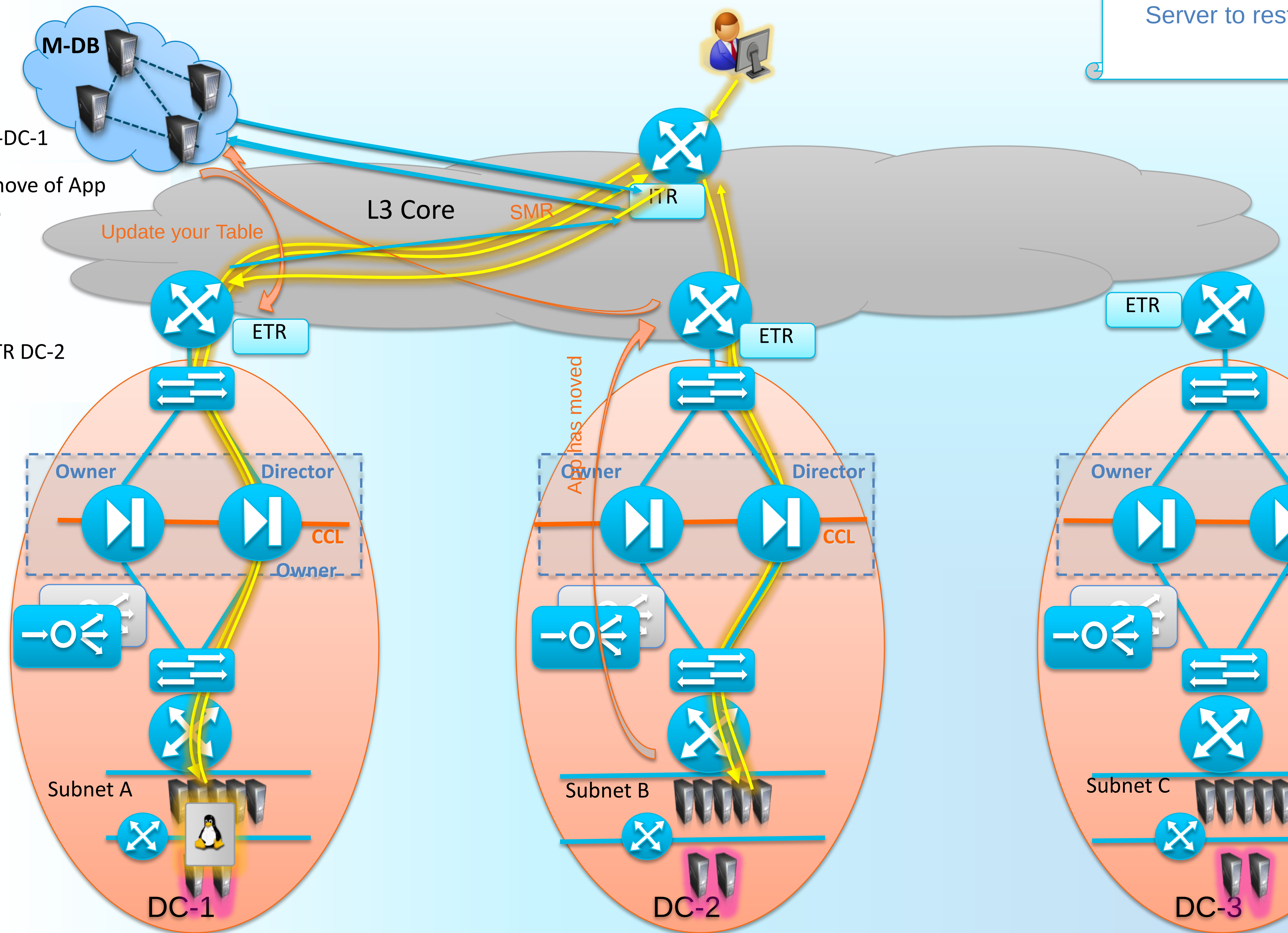
ASA Clustering per DC across Multiple sites

LISP Across Subnet Mode with ASA Clustering (Cold migration)

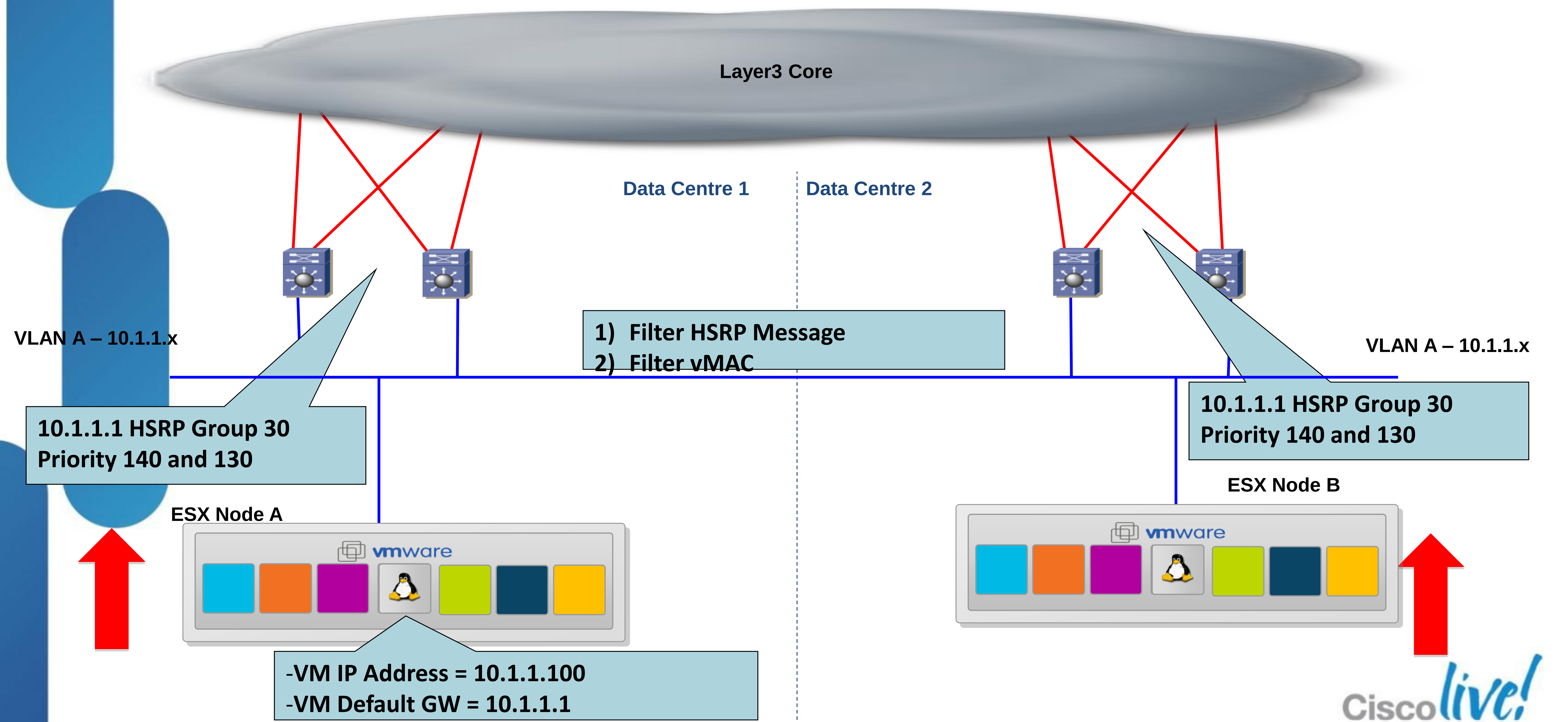
- Business continuity assumes the user re-establish a new session
- TCP session is re-initiated
- Cold Migration implies the Server to restart

1 - End-user sends Request to App
2 - ITR intercepts the Req and check the localisation
3 - MS replies location for Subnet A being ETR DC-1
3" - ITR encaps the packet and sends it to RLOC ETR-DC-1

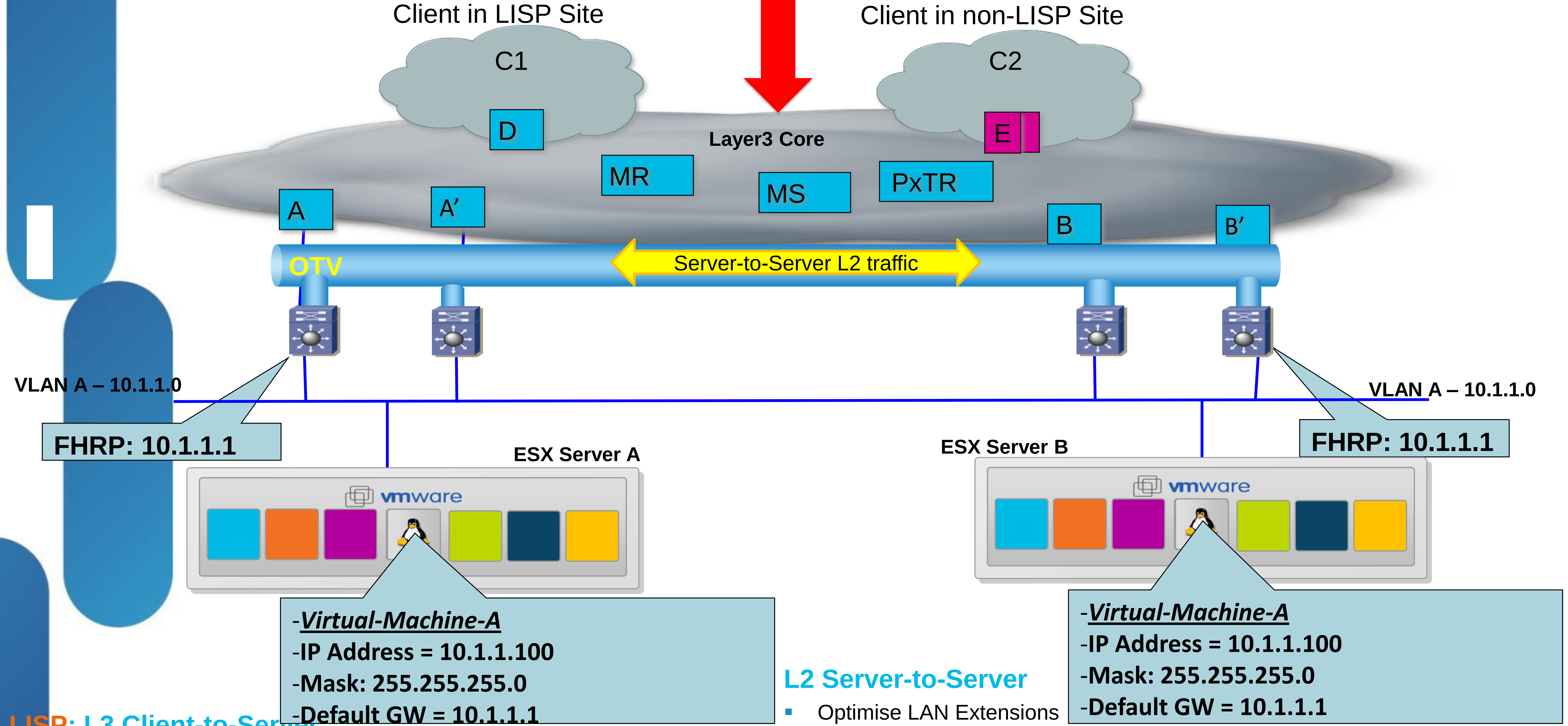
4 – LISP Multi-hop informs ETR on DC-2 about the move of App
5 – ETR DC-2 informs MS about new location of App
6 – MR updates ETR DC-1
7 – ETR DC-1 updates its table (App:Null0)
8 – ITR sends traffic to ETR DC-1
9 – ETR DC-1 replies Solicit Map Req
8 – ITR sends a Map Req and redirects the Req to ETR DC-2



Localised First Hop



Locator/ID Separation Protocol (LISP) and L2 Extension Workload Mobility



LISP: L3 Client-to-Server

- Optimise L3 Routing providing granular location information
- Optimised mobility within or across subnets
- Scale the network so host routes are in mapping database

L2 Server-to-Server

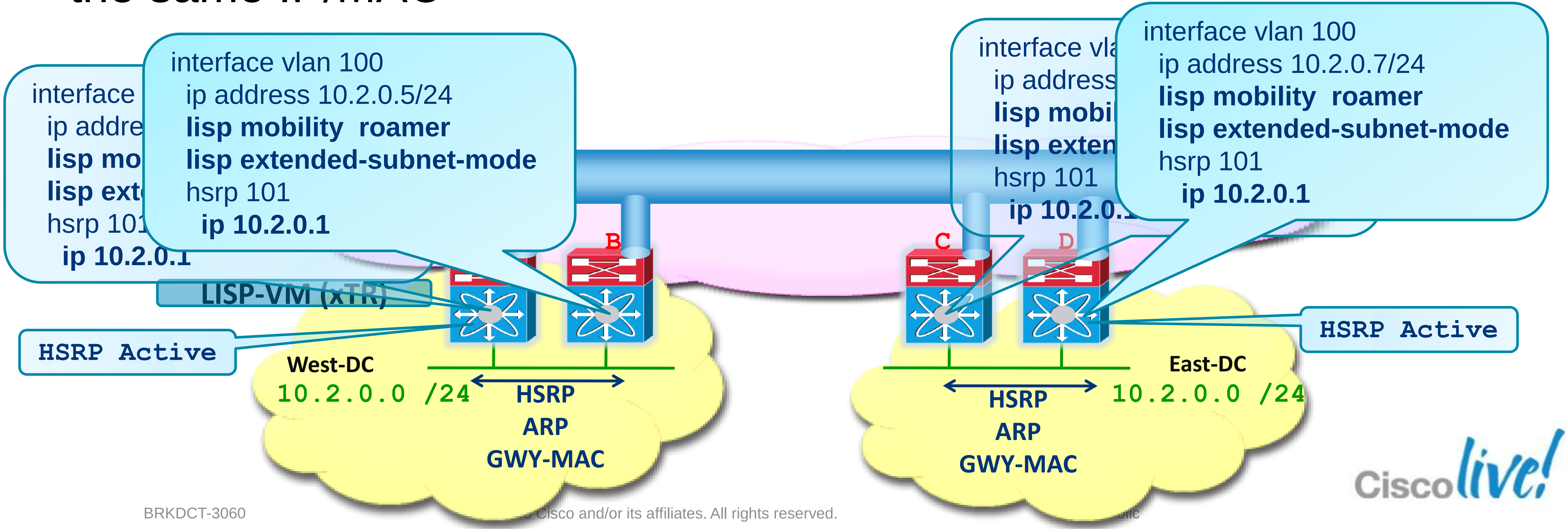
- Optimise LAN Extensions
- Enable dispersion of app clusters
- App discovery based on MAC level broadcast and link-local multicast
- General application communication may require L2 connectivity

L3 Router   LISP Router or infrastructure device

LISP Host-Mobility – First Hop Routing

With Extended Subnets

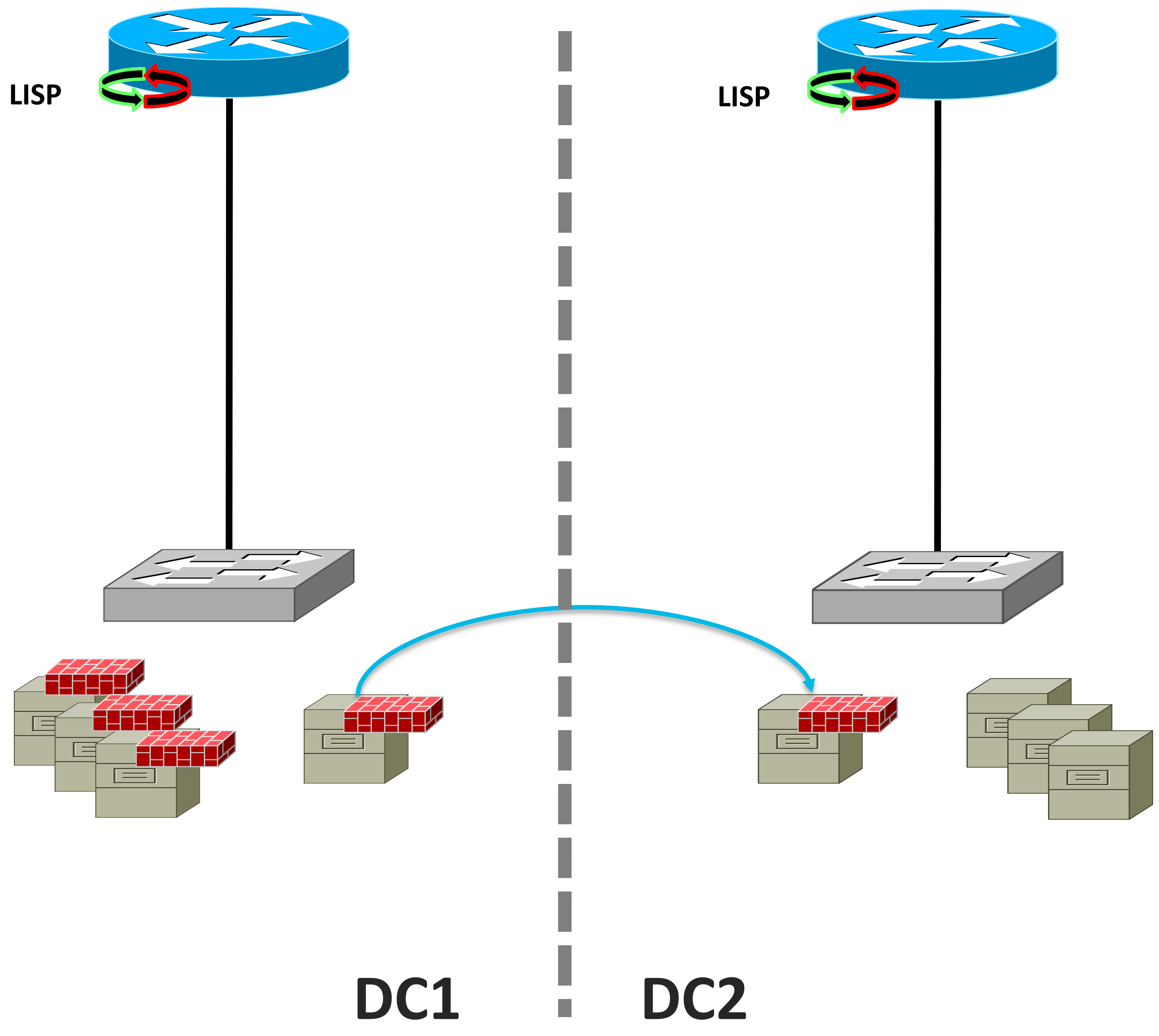
- Consistent GWY-IP and GWY-MAC configured across all sites
  - Consistent HSRP group number across sites ➔ consistent GWY-MAC
- Servers can move anywhere and always talk to a local gateway with the same IP/MAC



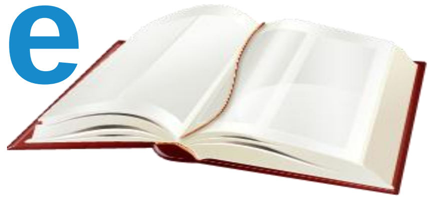
Service State Mobility

vPath and the Virtual Services Gateway (VSG)

- VSG uses the vPath model
- FW policies are maintained centrally
- FW state/enforcement is distributed to the hypervisor switch
- FW state moves granularly with each VM



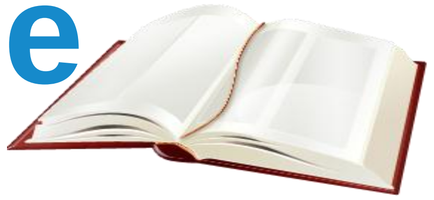
OTV - HSRP Localisation – OTV Edge Device



1) Define HSRPv1 and HSRPv2 to block HSRP Hello Messages

```
ip access-list ALL_IPs
  10 permit ip any any
!
mac access-list ALL_MACs
  10 permit any any
!
ip access-list HSRP_IP
  10 permit udp any 224.0.0.2/32 eq 1985
  20 permit udp any 224.0.0.102/32 eq 1985

vlan access-map HSRP_Local 10
        match ip address HSRP_IP
        action drop
vlan access-map HSRP_Local 20
        match ip address ALL
        action forward
```

OTV - HSRP Localisation – OTV Edge Device

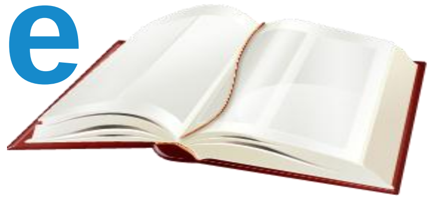


2) Prevent Duplicate HSRP Gratuitous ARP from HSRP VIP

```
arp access-list HSRP_VMAC_ARP
  10 deny ip any mac 0000.0c07.ac00 ffff.ffff.ff00
  20 deny ip any mac 0000.0c9f.f000 ffff.ffff.f000
  30 permit ip any mac any

feature dhcp
ip arp inspection filter HSRP_VMAC_ARP 10,11,600, 601, 700, 701


interface Vlan10
  no shutdown
  no ip redirects
  ip address 192.168.10.3/24
  no ip arp gratuitous hsrp duplicate
  hsrp 10
    priority 110
    ip 192.168.10.1
```

Message without: %ARP-3-DUP_VADDR_SRC_IP: arp [3849] Source address of packet received from 0000.0c07.ac1f on Vlan10(port-channel10) is duplicate of local virtual ip, 192.168.10.1

OTV - HSRP Localisation – OTV Edge Device

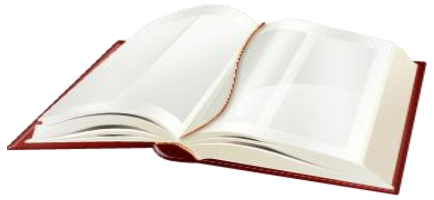


3) Filter learning HSRP Virtual MAC address across OTV

```
mac access-list HSRP_VMAC
  10 permit 0000.0c07.ac00 0000.0000.00ff any
  20 permit 0000.0c9f.f000 0000.0000.0fff any
!
vlan access-map HSRP_Localization 10
        match mac address HSRP_VMAC
        match ip address HSRP_IP
        action drop
!
vlan access-map HSRP_Localization 20
        match mac address ALL_MACs
        match ip address ALL_IPs
        action forward
!
vlan filter HSRP_Local vlan-list 10,11,600, 601, 700, 701
```

```
mac-list HSRP_VMAC_Deny seq 5 deny 0000.0c07.ac00
ffff.ffff.ff00
mac-list HSRP_VMAC_Deny seq 10 deny 0000.0c9f.f000
0000.0000.0fff
mac-list HSRP_VMAC_Deny seq 15 permit 0000.0000.0000
0000.0000.0000
!
route-map stop-HSRP permit 10
match mac-list HSRP_VMAC_Deny
!
otv-isis default
vpn Overlay0
redistribute filter route-map stop-HSRP
```

VPLS Localisation

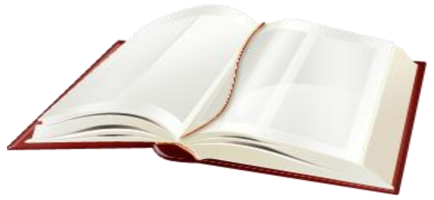


1) Configure virtual port-channel (vPC) on BOTH Nexus 7000 aggregation switches and filter HSRP

```
interface Ethernet2/1
lacp rate fast
switchport
switchport mode trunk
switchport trunk allowed vlan 1,76-80,100-349
channel-group 31 mode active
no shutdown

interface Ethernet2/2
lacp rate fast
switchport
switchport mode trunk
switchport trunk allowed vlan 1200-1449
channel-group 32 mode active
no shutdown
```

```
interface Ethernet2/6
lacp rate fast
switchport
switchport mode trunk
switchport trunk allowed vlan 1,76-80,100-349
channel-group 31 mode active
no shutdown

interface Ethernet2/3
lacp rate fast
switchport
switchport mode trunk
switchport trunk allowed vlan 1200-1449
channel-group 32 mode active
no shutdown
```

VPLS Localisation



2) Access list to filter HSRP hellos configured on both aggregation switches

```
ip access-list HSRP_Deny
statistics per-entry
10 deny udp any 224.0.0.102/32 eq 1985
20 permit ip any any
```

VPLS Localisation

3)Configure port-channel interface on BOTH Nexus 7000 aggregation switches

```
interface port-channel31
switchport
switchport mode trunk
ip port access-group HSRP_Deny in
switchport trunk allowed vlan 1,76-80,100-349
spanning-tree port type edge trunk
spanning-tree bpdufilter enable
vpc 31
```

```
interface port-channel32
switchport
switchport mode trunk
ip port access-group HSRP_Deny in
switchport trunk allowed vlan 1200-1449
spanning-tree port type edge trunk
spanning-tree bpdufilter enable
lacp max-bundle 1
vpc 32
```

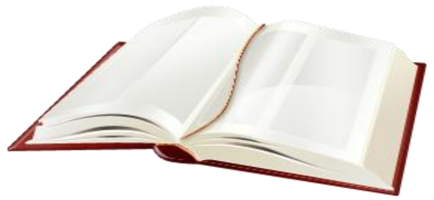
Summary State-full devices placement with DCI

- Ping-Pong effect might have a bad impact in term of perf with long distances:
  - Greedy bandwidth
  - Latency
- It is commonly accepted to distribute traditional A/S state-full devices between 2 Twin DC for short Metro Distances (+/- 10km max)
  - Keep transparency and easy to operate
  - limited to 2 Active DC
- As of today the preferred method is to deploy Stretch ASA clustering across distributed DC (Metro)
  - All ASA active
  - Not limited to 2 Active DC
- For Geographical Distributed DC
  - if Hot migration is required (i.e. Geo VPLEX), use ASA cluster stretched over multiple sites with LAN extension
  - for Cold migration use ASA cluster distributed per site in conjunction with LISP
- Ingress Path Optimisation
  - LISP Mobility is the preferred choice – It requires LISP Multi-hop
  - GSLB (DNS and KAP-AP) can help to redirect the traffic accordingly, but may face some caveats with proxy DNS and client caching
  - RHI can help but offers App based granularity only for Intranet core (Enterprise owns the L3 core)
- The recommended choice is ASA clustering in conjunction with the traditional DNS and LISP Mobility.
  - Stretched across multiple DC with LAN extension for Hot Migration
  - Confined inside each DC without LAN extension for Cold Migration

Data Centre Interconnect

Agenda

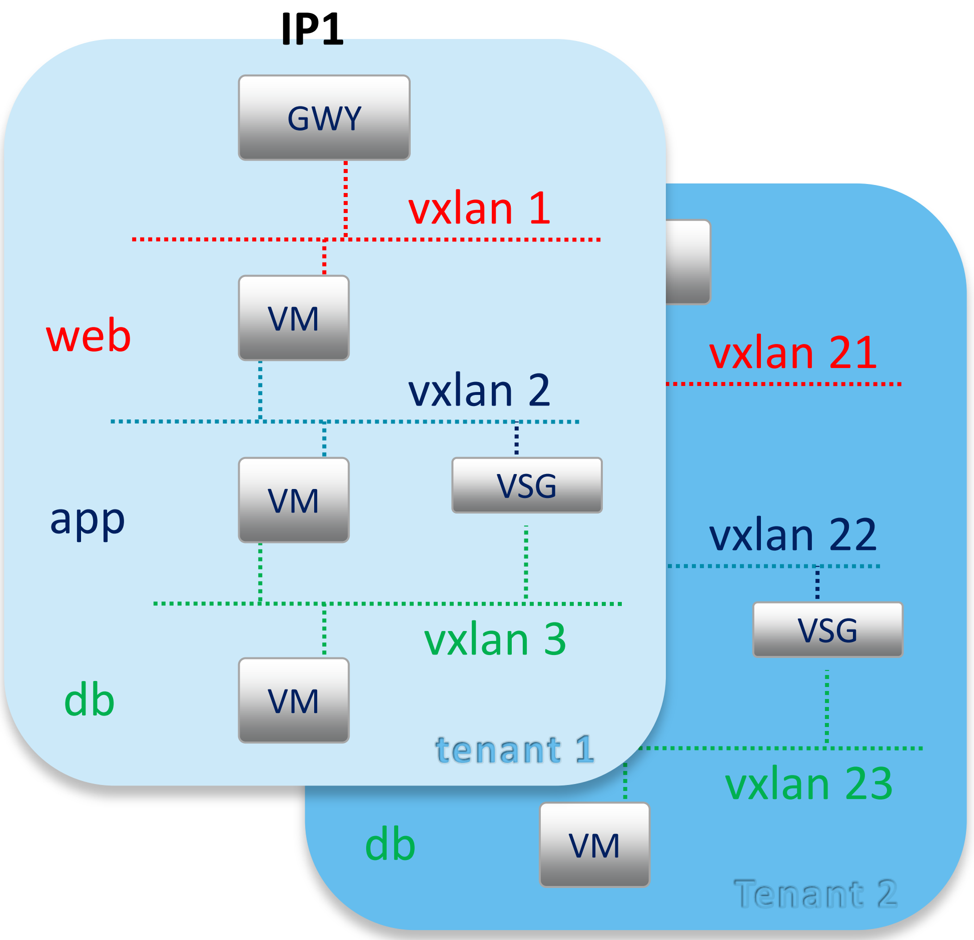
- Mobility and Virtualisation in the Data Centre
- LAN Extension Deployment Scenarios
  - –Ethernet Based Solutions
  - –MPLS Based Solutions
    - EoMPLS
    - VPLS
    - A-VPLS
    - EVPN
- Overlay Transport Virtualisation (OTV)
- Encryption
- IP Mobility without LAN Extension
- Path optimisation
- VXLAN
- Summary and Conclusions
- Q&A



= For your Reference

VXLAN

L2 Host Overlays and Virtualisation – VXLAN
Creating virtual segments



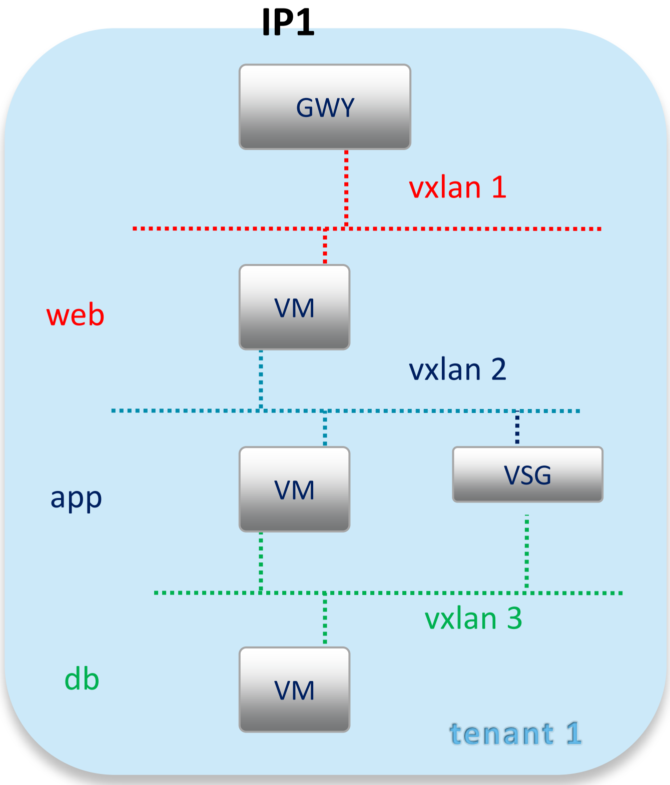
Multi-tier Virtual App = VMs + Segments + Gateway

Application: Cloud Services

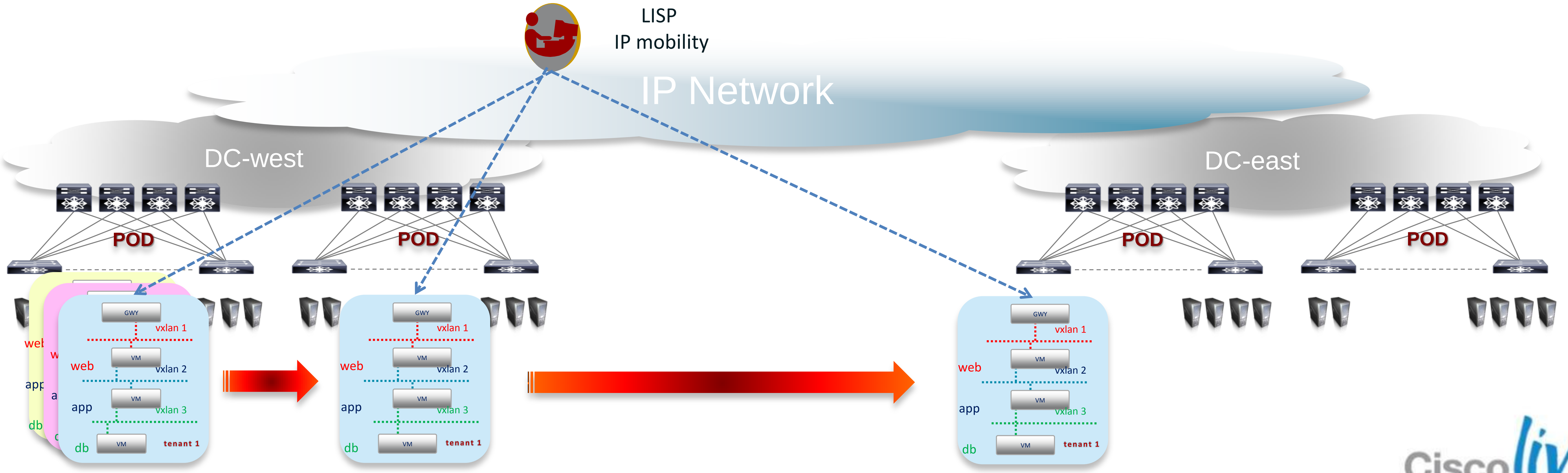
VXLAN elastic creation of **virtual Segments**

- Small Segments
  - Usually don't stretch outside of a POD
- Mobile: Can be instantiated anywhere
  - Move along with VMs as necessary
- Very large number of segments
  - Do not consume resources in the network core
- Host overlays are initiated at the hypervisor virtual switch ➔ Virtual hosts only
- Gateway to connect to the non-virtualised world
- VXLAN shipping since 2011 on Cisco Nexus 1000v, other variants: NVGRE, STT

LISP enables VXLAN to deliver vApp mobility



- Move virtual Applications (vApps) among private cloud PODs
  – Move VMs and virtual Segments (VXLANs)
- LISP host mobility allows the vApp to roam
  – Maintain optimal path for Client-Server connectivity
  – Maintain GWY IP address, segmentation and optimal reachability



VXLAN for DC Geo-Dispersion?

There are better suited tools

North-south VXLAN limitations

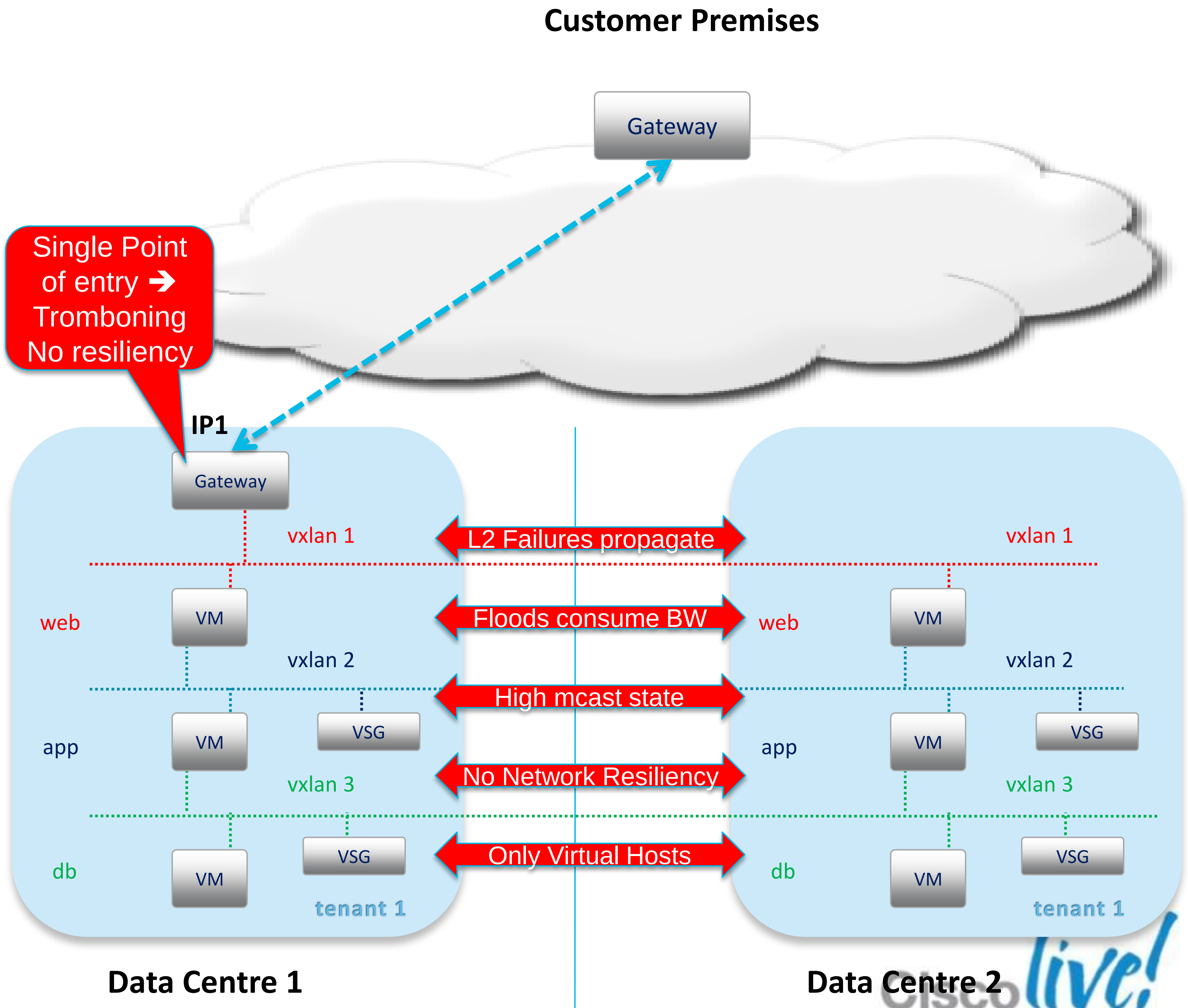
- Only one gateway per segment
  - More than one Gateway will lead to loops
  - Traffic is tromboned to the Gateway
  - Defeats the purpose of the geographic dispersion

East-west VXLAN limitations

- No isolation of L2 failures
- Excessive flood traffic and BW exhaustion
- Large amounts of IP multicast between DCs
- Not HW accelerated, virtual elements only
- No network resiliency or multi-pathing of the L2 overlay

The DCI toolkit solves all these issues in LISP, OTV and EVPN

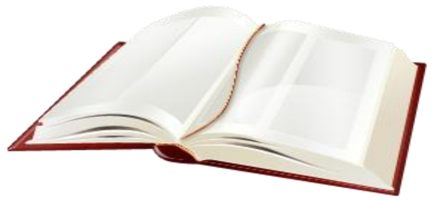
VXLAN is designed for small mobile segments, not extended segments



Data Centre Interconnect

Agenda

- Mobility and Virtualisation in the Data Centre
- LAN Extension Deployment Scenarios
  - –Ethernet Based Solutions
  - –MPLS Based Solutions
    - EoMPLS
    - VPLS
    - A-VPLS
    - EVPN
- Overlay Transport Virtualisation (OTV)
- Encryption
- IP Mobility without LAN Extension
- Path optimisation
- VXLAN
- Summary and Conclusions
- Q&A



= For your Reference

Summary

- Discussed different deployment options and transport options
- Tightly coupled Data Centre with FabricPath
- Spanning-tree isolation
- Traffic Optimisation Egress and Ingress Symmetry
- Encryption Solutions

References

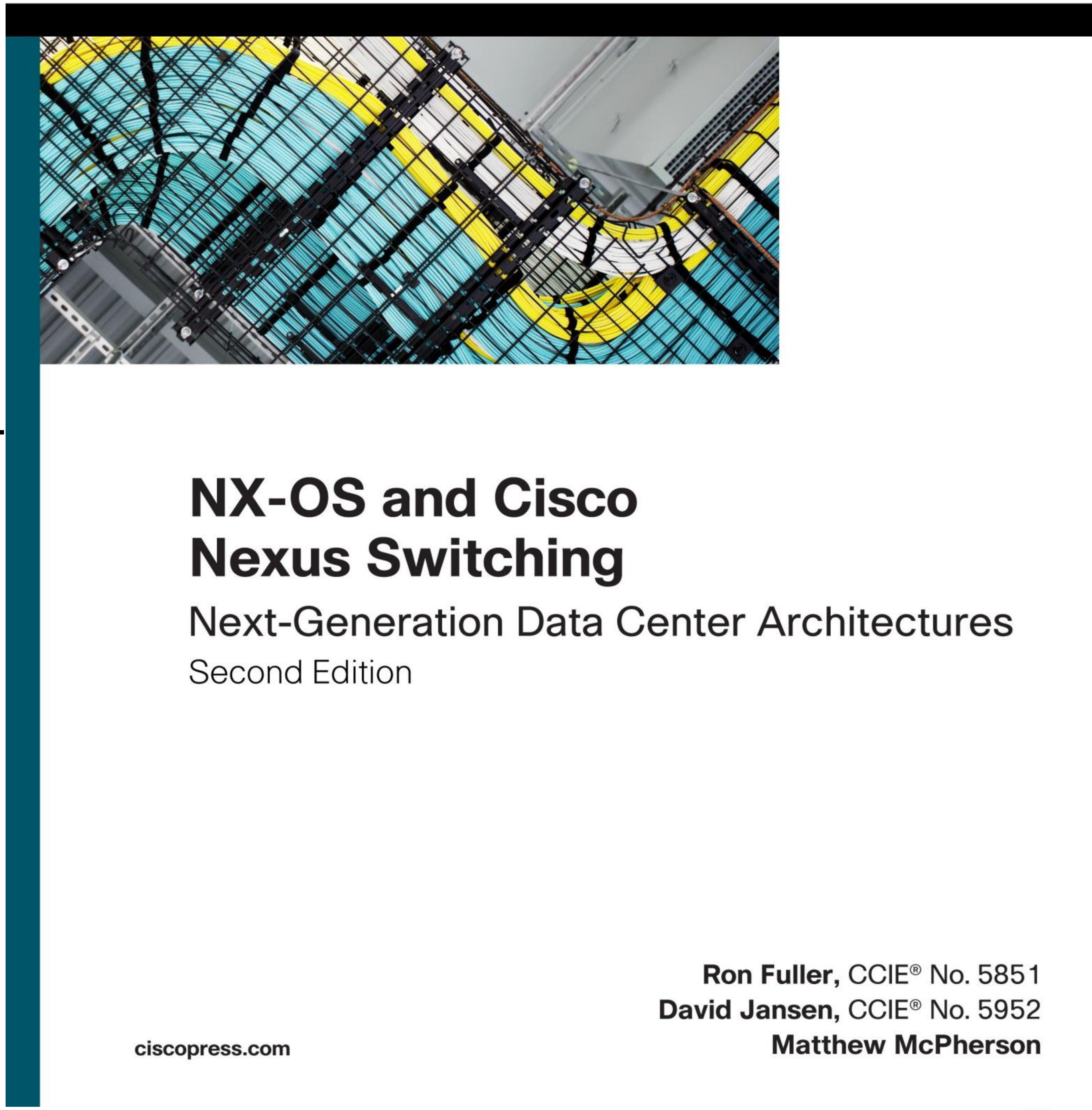
- Cisco Validated Design – DCI Solutions

http://www.cisco.com/en/US/solutions/ns340/ns414/ns742/ns743/ns749/landing_dci_mpls.html

- Discussed different deployment options and transport options
- Tightly coupled Data Centre with FabricPath
- Spanning-tree isolation
- Traffic Optimisation Egress and Ingress Symmetry
- Encryption Solutions

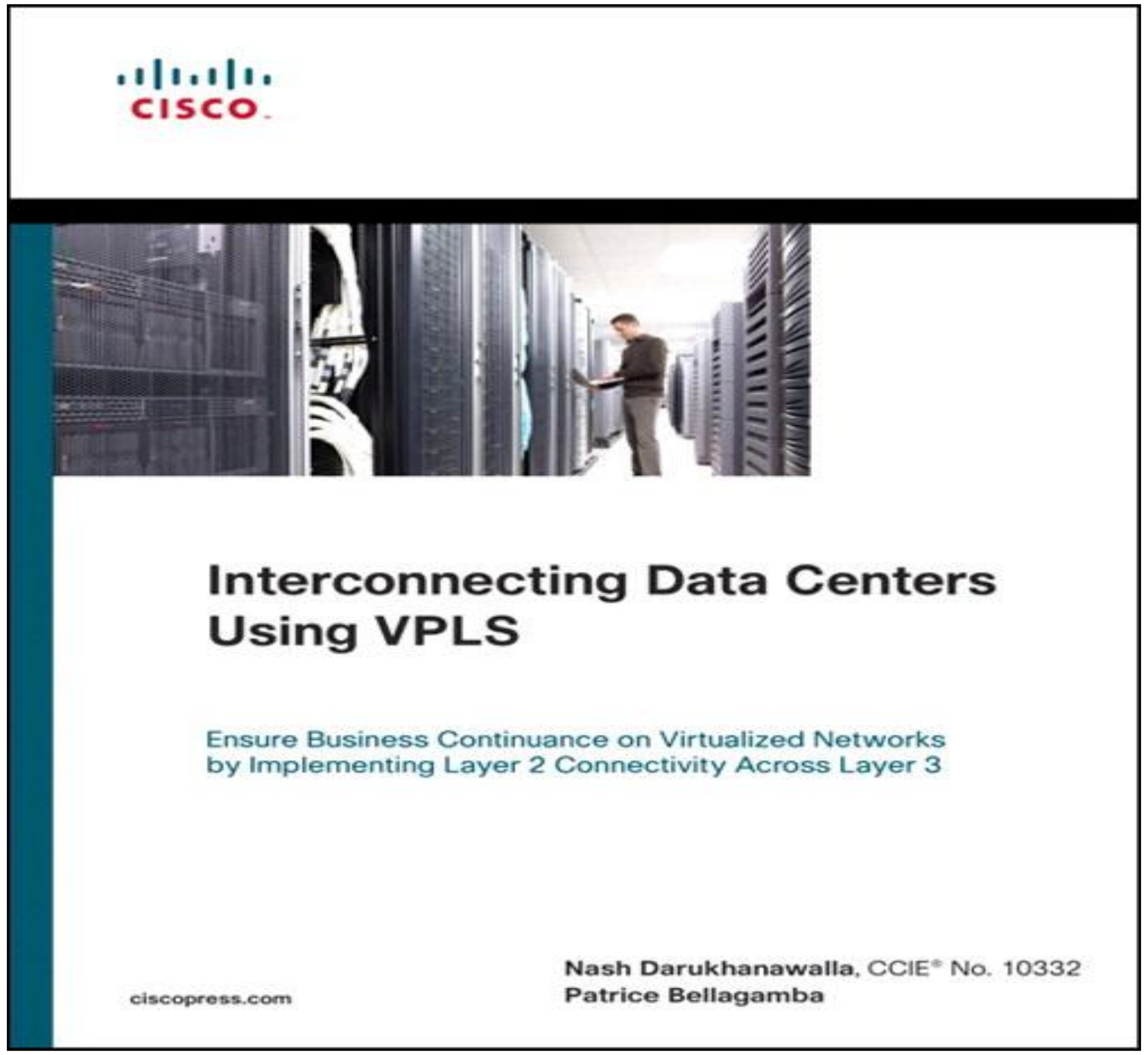
Recommended Reading

- NX-OS and Cisco Nexus Switching 2nd Edition (ISBN: 1587143046), by David Jansen, Ron Fuller, Matthew McPherson. Cisco Press 2013.
- NX-OS and Cisco Nexus Switching (ISBN: 1587058928), by David Jansen, Ron Fuller, Kevin Corbin. Cisco Press 2010.
- Interconnecting Data Centres Using VPLS (ISBN-10: 1-58705-992-4; ISBN-13: 978-1-58705-992-6), by Nash Darukhanawalla, Patrice Bellagamba . Cisco Press. 2009.
- Layer 2 VPN Architectures (ISBN: 1-58705-848-0), by Wei Luo, Carlos Pignataro, Anthony Chan, Dmitry Bokotey. Cisco Press. 2005.
- Cisco LAN Switching Configuration Handbook (2nd Edition) (ISBN-1587056100; ISBN-13: 978-1587056109), by Steve McQuerry, David Jansen, David Hucaby, Cisco Press. 2009.



Recommendations

- Check the Recommended Reading flyer for suggested books
- Additional Information on LISP:
  - http://www.lisp4.net
  - http://lisp4.cisco.com
  - http://www.cisco.com/go/lisp



Q & A

Complete Your Online Session Evaluation

Give us your feedback and receive a Cisco Live 2013 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 8 March 12:00pm-2:00pm

Don't forget to activate your Cisco Live 365 account for access to all session material, communities, and on-demand and live activities throughout the year. Log into your Cisco Live portal and click the "Enter Cisco Live 365" button.
www.ciscoliveaustralia.com/portal/login.ww