What You Make Possible

TOMORROW
starts here.

Simplifying DCI with Overlay Transport Virtualisation

BRKDCT-2049

TOMORROW starts here.

Session Objectives

- The main goals of this session are:
- This session features a detailed analysis of the architectural aspects and deployment benefits behind OTV
- The attendees will learn how OTV is aimed at providing Layer 2 connectivity beyond the Layer 3 boundary while maintaining the failure containment and operational simplicity that the Layer 3 boundary provides
- The attendees will get a deep knowledge of how the OTV control-plane and data-plane work to provide the VLAN extension

Session Non-objectives

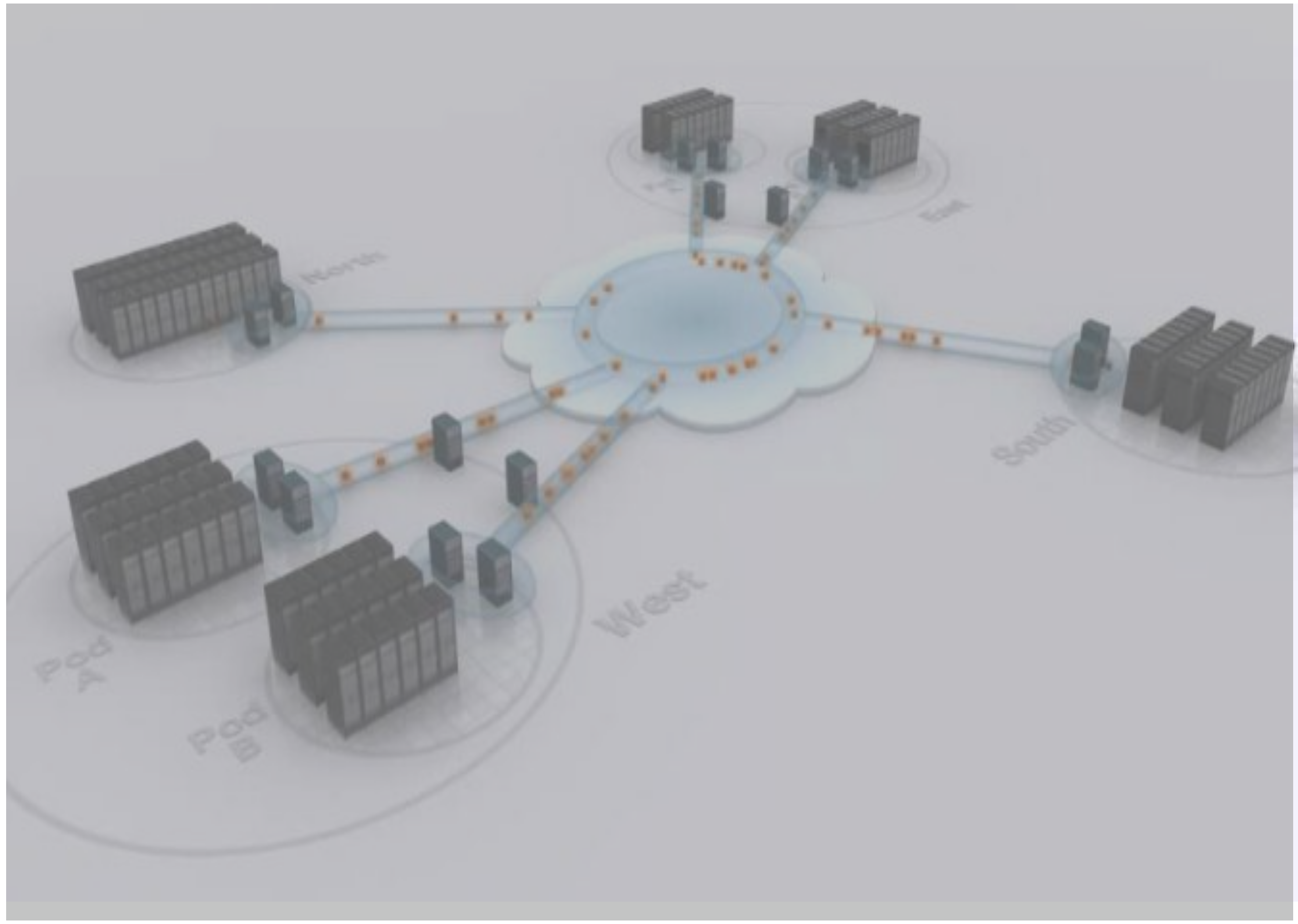
- This session does not include:
- In depth discussion of Path Optimisation technologies (ACE/GSS, LISP, etc.)
- Storage extension considerations associated to DCI deployments
- Workload mobility application specific deployment considerations

Related Cisco Live Events

Session-ID	Session Name
BRKDCT-2615	Active-Active Data Centre Strategies
BRKDCT-3060	Deployment Considerations with Interconnecting Data Centres
BRKDCT-2081	Cisco FabricPath Technology and Design
BRKDCT-3103	Advanced OTV – Configure, Verify and Troubleshoot OTV in Your Network

Agenda

- Distributed Data Centres: Goals and Challenges
- OTV Architecture Principles
- OTV at the Aggregation Layer

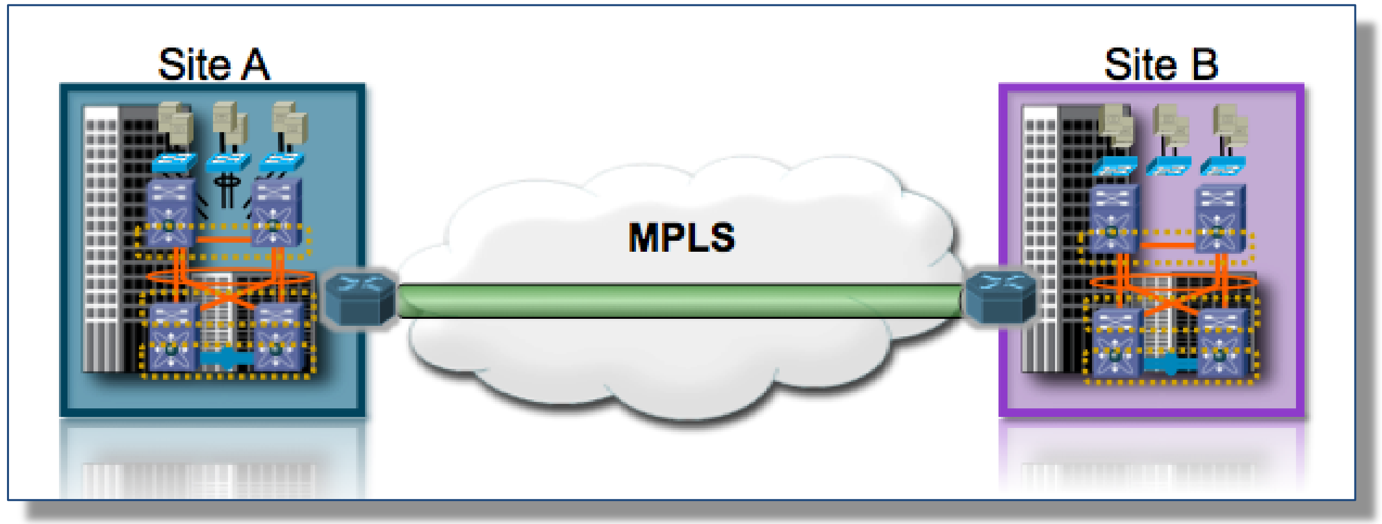


Distributed Data Centres Goals

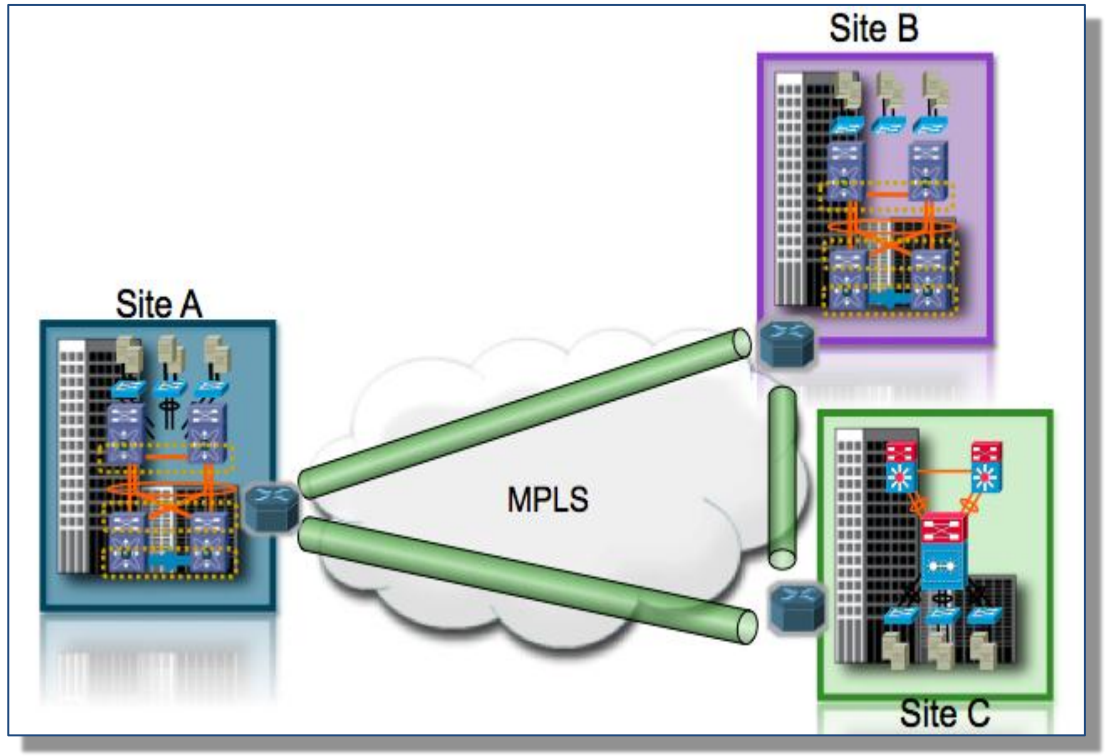
- Seamless workload mobility
- Distributed applications
- Maximise compute resources
- Ensure business continuity

Traditional Layer 2 Extension

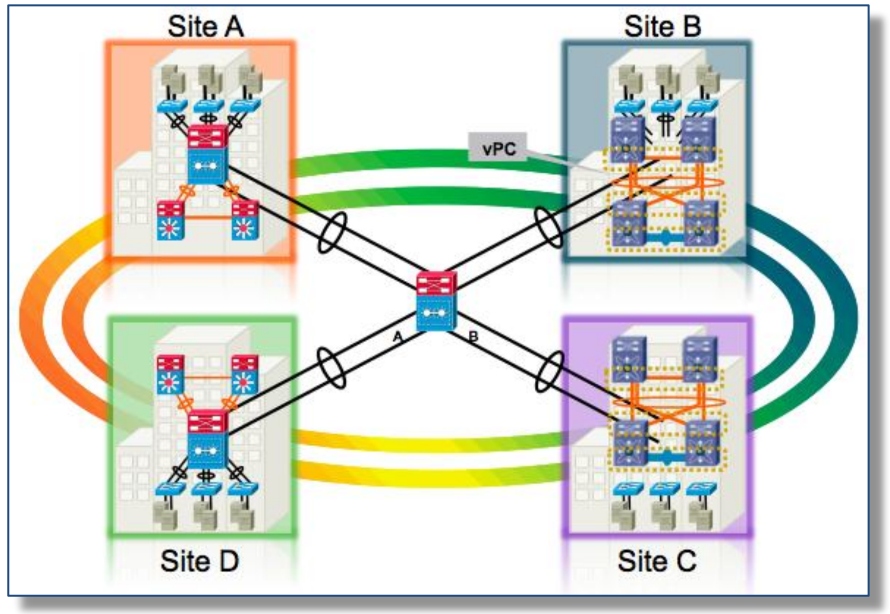
EoMPLS



VPLS



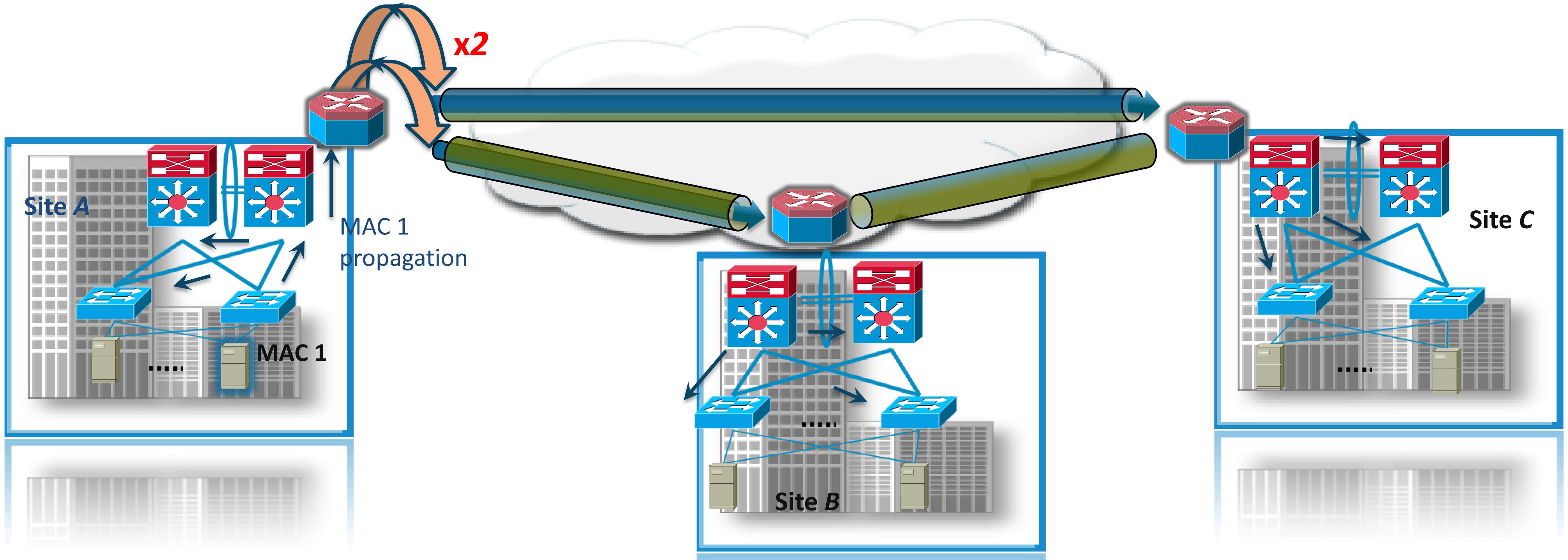
Dark Fibre



Traditional Layer 2 VPNs

Flooding Behaviour

- Unknown Unicast Flooding used to propagate MAC reachability
- Flooding domain extended to every site



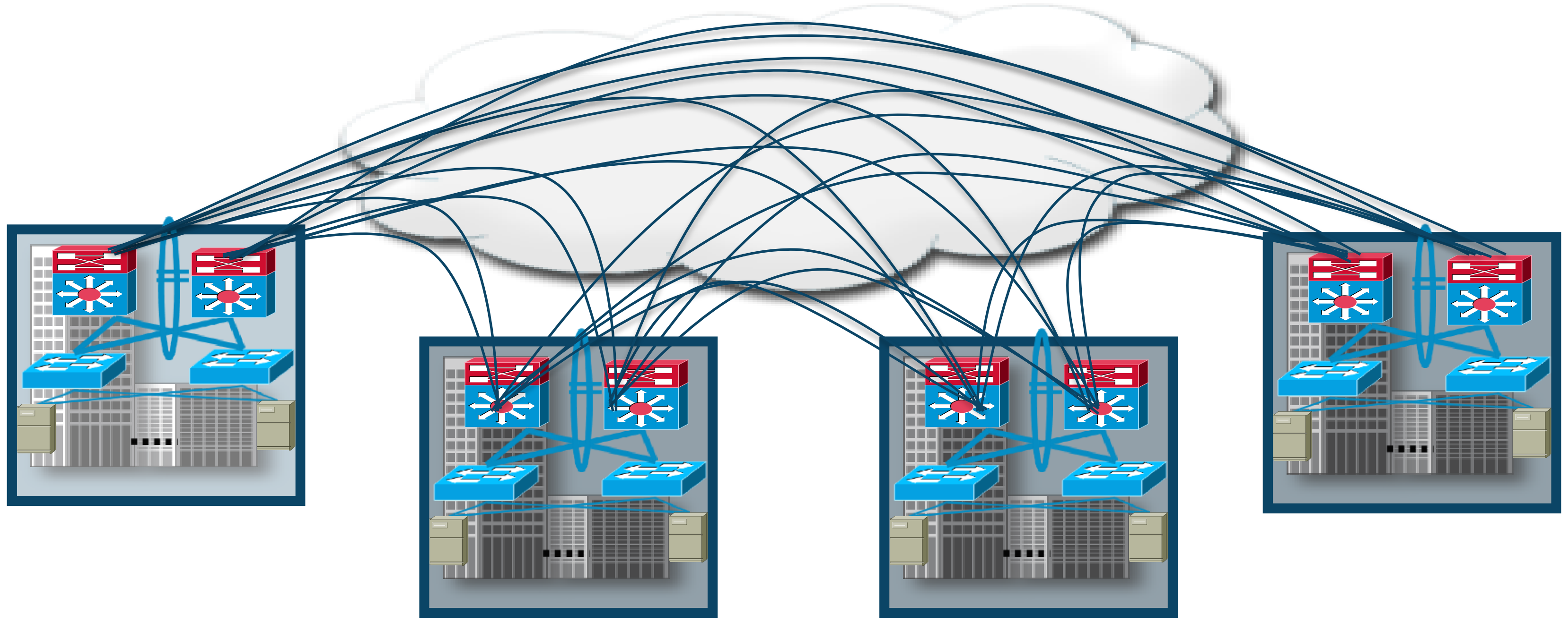
Our goal…

providing layer 2 connectivity, yet restrict the reach of the unknown unicast flooding domain in order to contain failures and preserve resiliency

Traditional Layer 2 VPNs

Pseudo-wire Maintenance

- Full mesh of pseudo-wires/tunnels must be in place
- Complex to add/remove sites. N sites = ***N*(N-1)/2*** pseudo-wires
- Head-end replication for mcast and bcast = Sub-optimal BW utilisation

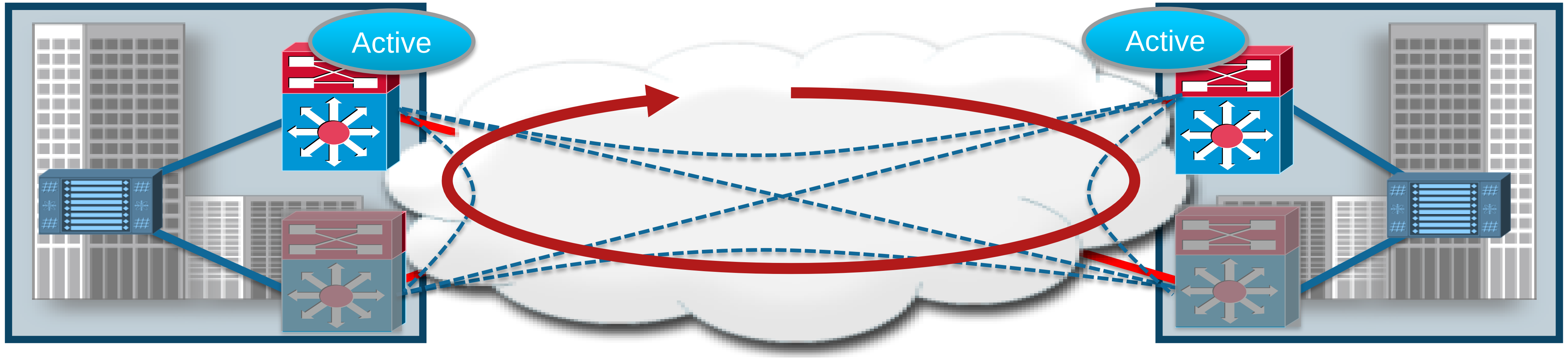


Our goal… providing point-to-cloud provisioning and optimal bandwidth utilisation in order to reduce cost

Traditional Layer 2 VPNs

Multi-Homing

- Requires additional protocols (BGP, ICC, EEM)
- STP often extended
- Malfunctions impact all sites



Our goal… natively providing automatic detection of multi-homing without the need of extending the STP domains, together with a more efficient load-balancing

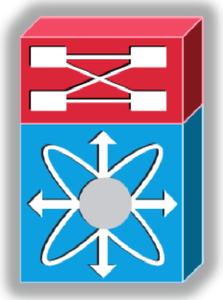
Overlay Transport Virtualisation

Technology Pillars

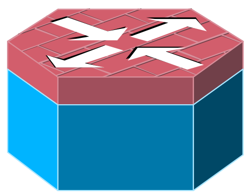
OTV is a “MAC in IP” technique to
extend Layer 2 domains
OVER ANY TRANSPORT

Dynamic Encapsulation

- No Pseudo-Wire State Maintenance
- Optimal Multicast Replication
- Multipoint Connectivity
- Point-to-Cloud Model



Nexus 7000
First platform to support OTV (since 5.0 NXOS Release)



ASR 1000
Now also supporting OTV (since 3.5 XE Release)

Protocol Learning

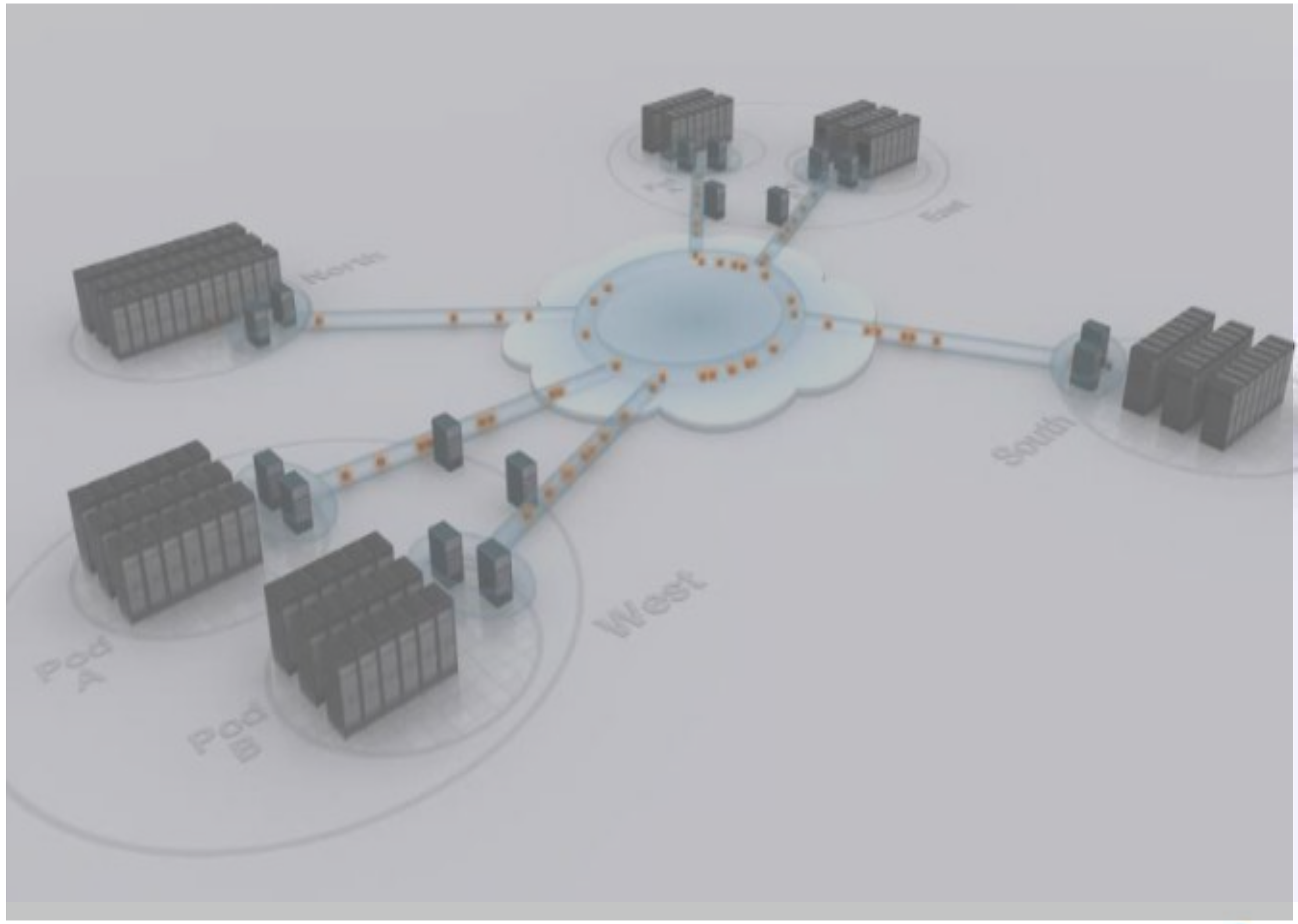
- Preserve Failure Boundary
- Built-in Loop Prevention
- Automated Multi-homing
- Site Independence

OTV Changes the Game…

- Flooding Based Learning → Control-Plane Based Learning
  - Control Plane Protocol for proactive advertisement of MAC addresses
- Pseudo-wires and Tunnels → Dynamic Encapsulation
  - Not static tunnel or pseudo-wire configuration
  - Optimal replication for a more efficient bandwidth utilisation
- Complex Dual-homing → Native Automated Multi-homing
  - Load balancing of flows
  - Site independence
  - STP isolation

Agenda

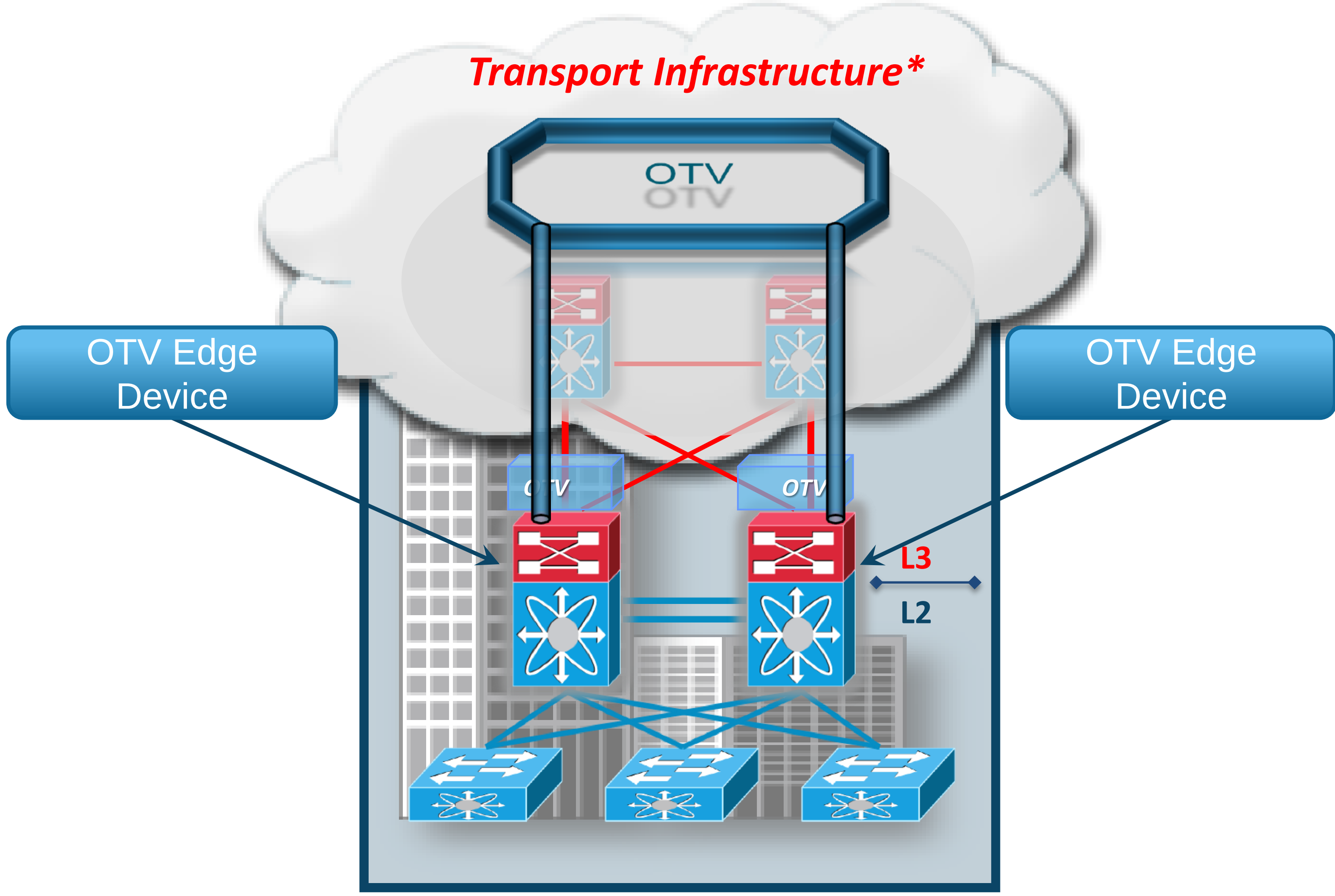
- Distributed Data Centres: Goals and Challenges
- OTV Architecture Principles
  - Control Plane and Data Plane
  - Failure Isolation
  - Multi-homing
  - L2 Multicast Forwarding
  - QoS and Scalability
  - Path Optimisation
- OTV at the Aggregation Layer



Terminology

OTV Edge Device

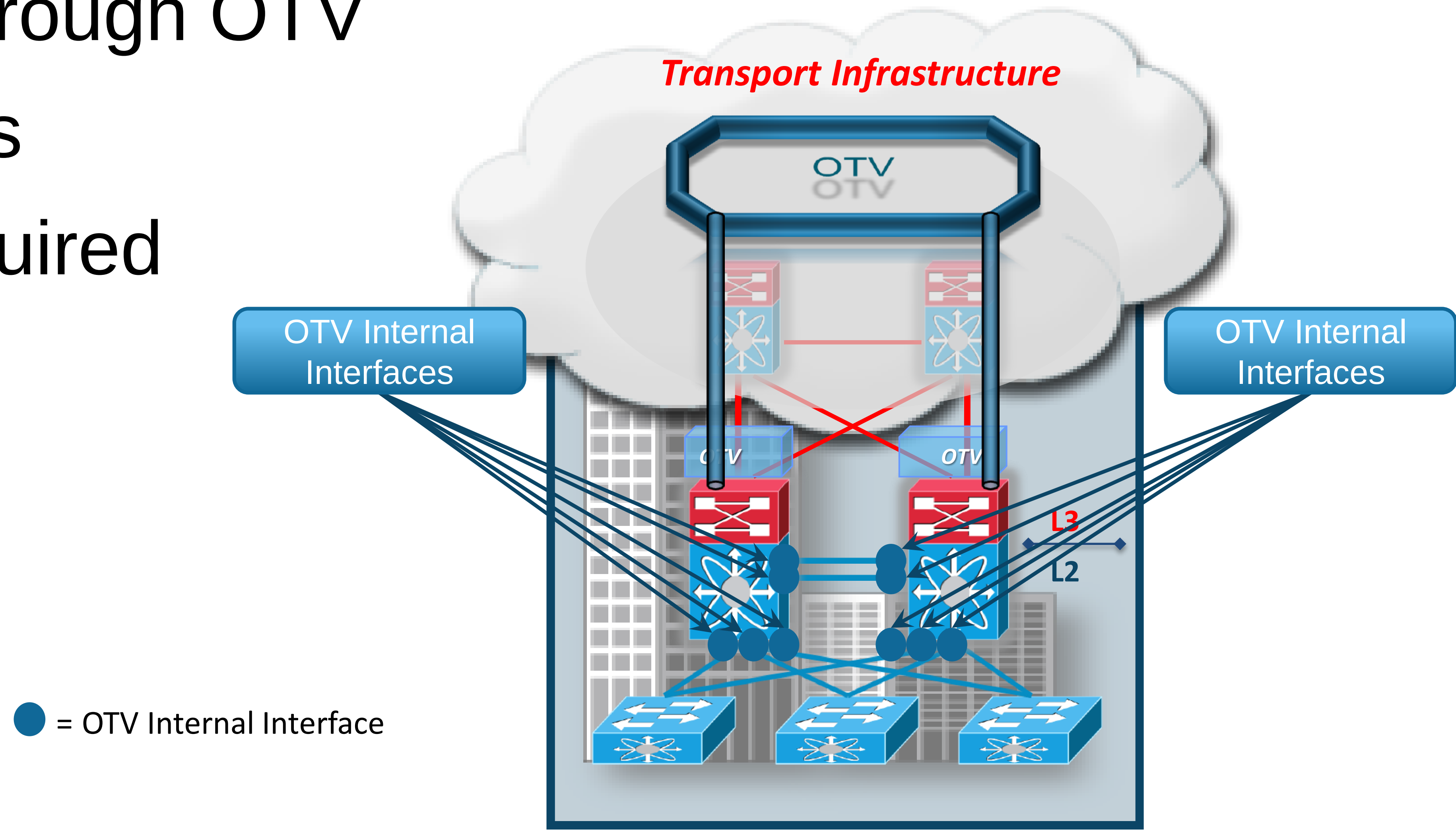
- Performs all OTV functionality
- Usually located at the Aggregation Layer or at the Core Layer
- Support for multiple OTV Edge Devices (multi-homing) in the same site



*   It can be owned by the Enterprise or by the Service Provider

Terminology
Internal Interfaces

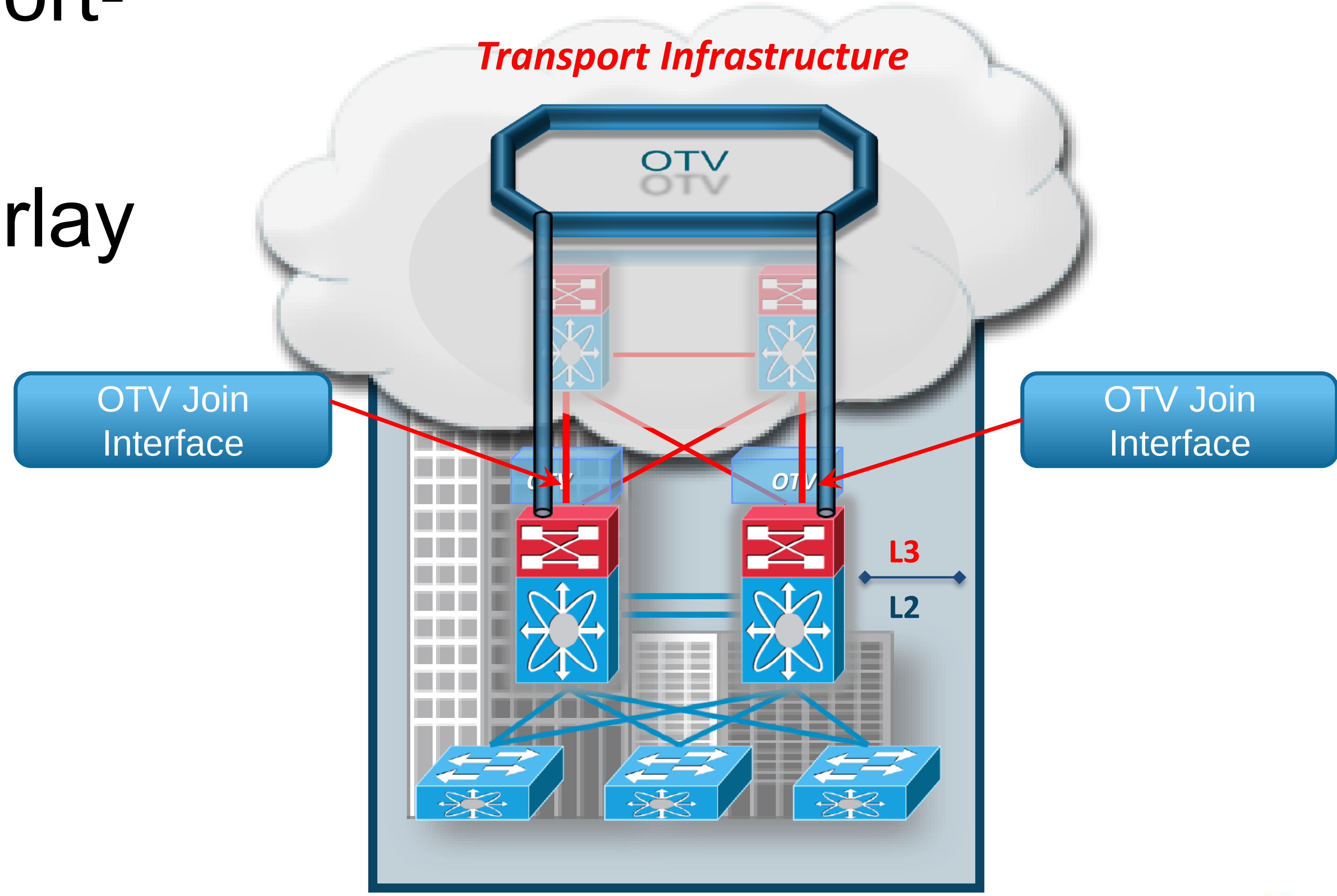
- Site facing Interfaces of the Edge Devices
- Carry VLANs extended through OTV
- Regular Layer 2 interfaces
- No OTV configuration required



Terminology

Join Interfaces

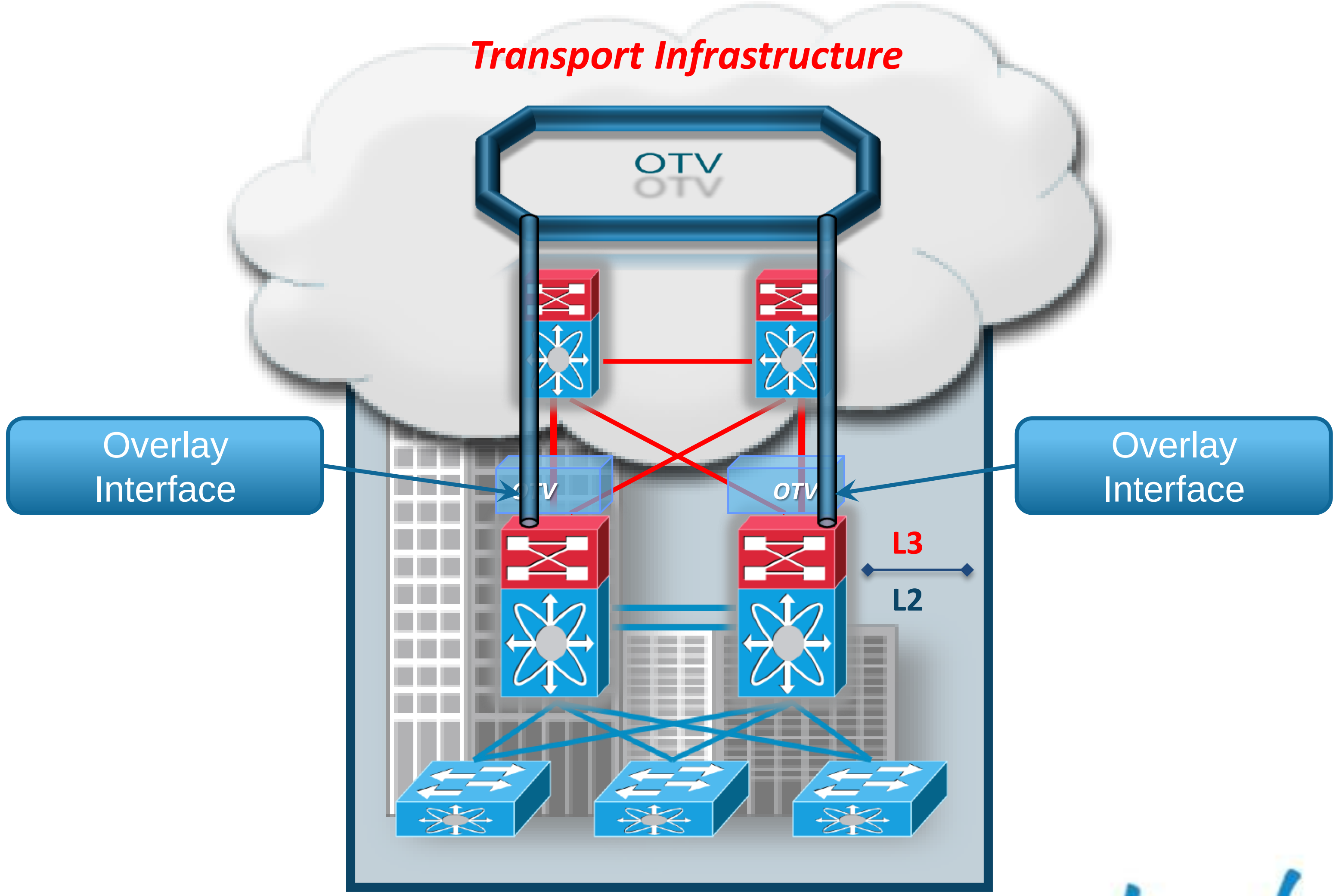
- One of the uplink of the Edge Device
- Point-to-point routed interface (port-channel supported)
- Used to physically “join” the Overlay network
- No OTV specific configuration required



Terminology

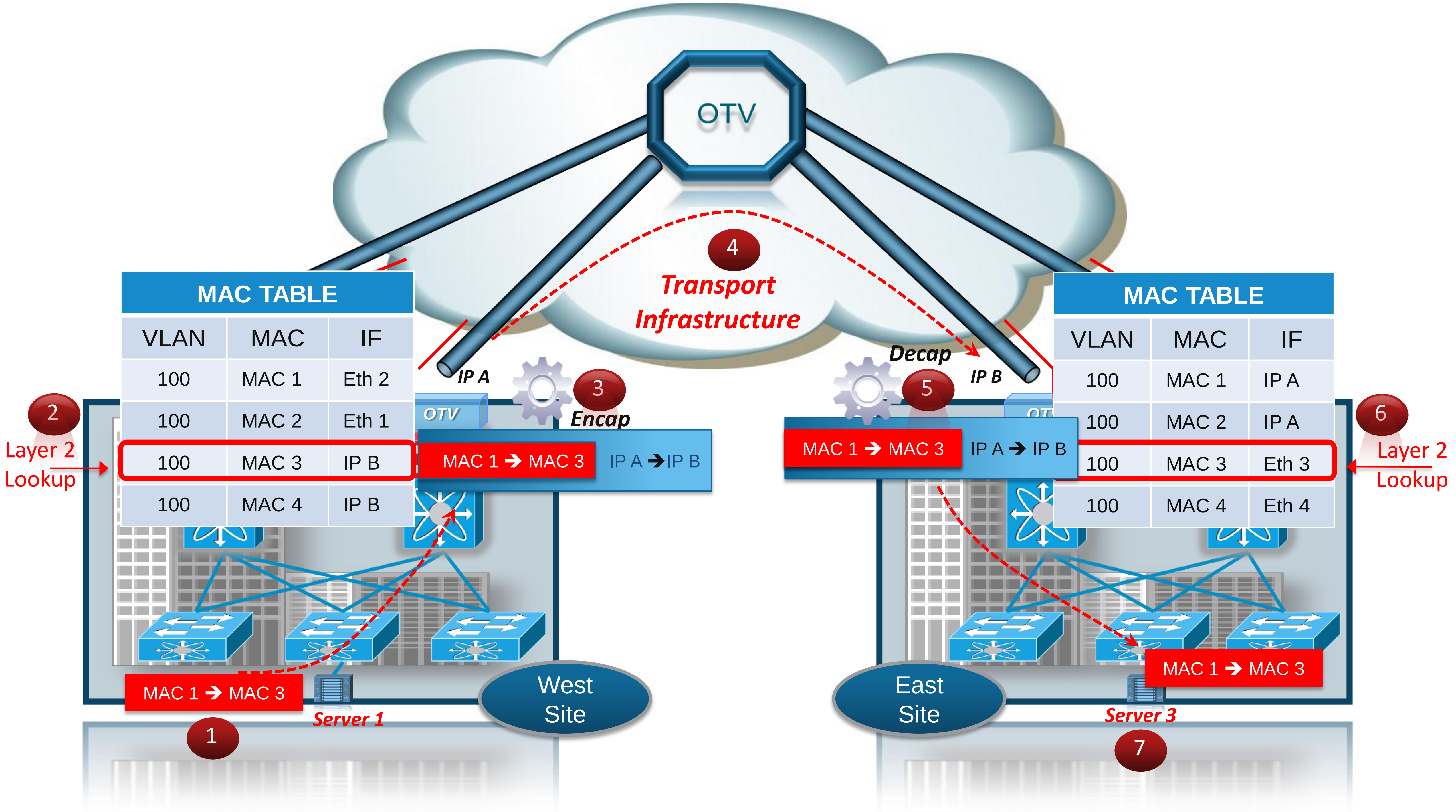
Overlay Interface

- **Virtual** interface with most of the OTV configuration
- Logical multi-access multicast-capable interface
- Encapsulates Layer 2 frames in IP unicast or multicast



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

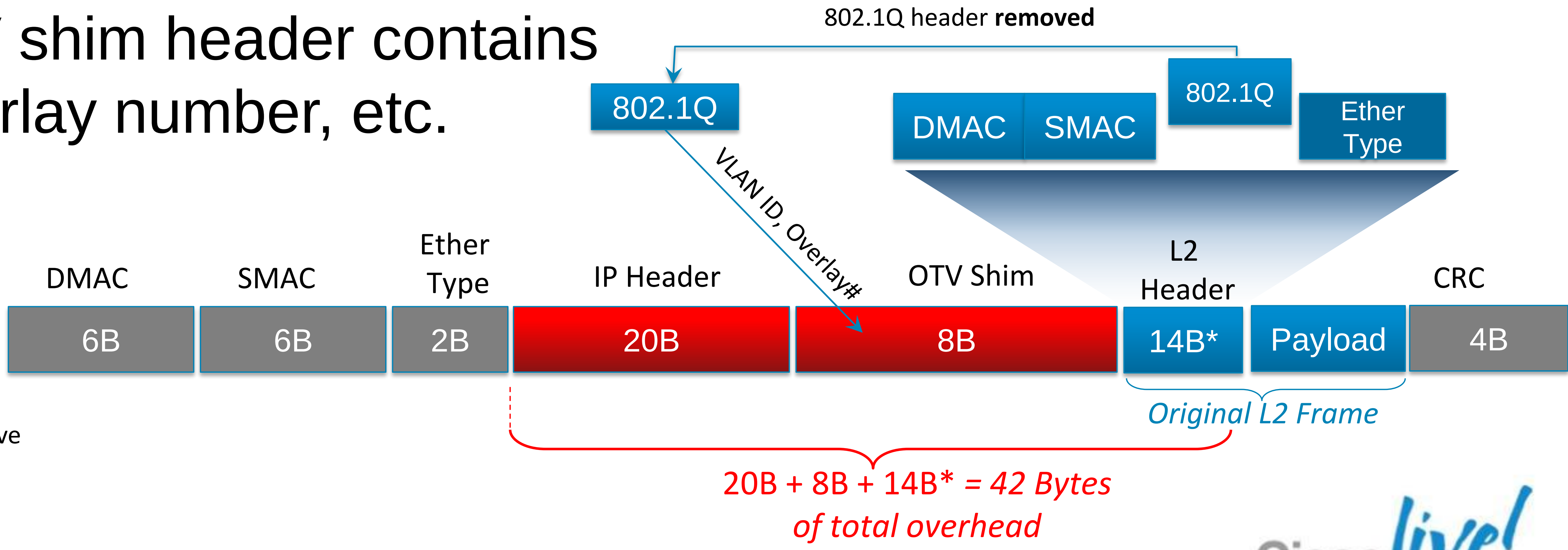
MAC TABLE (East Site)

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

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.

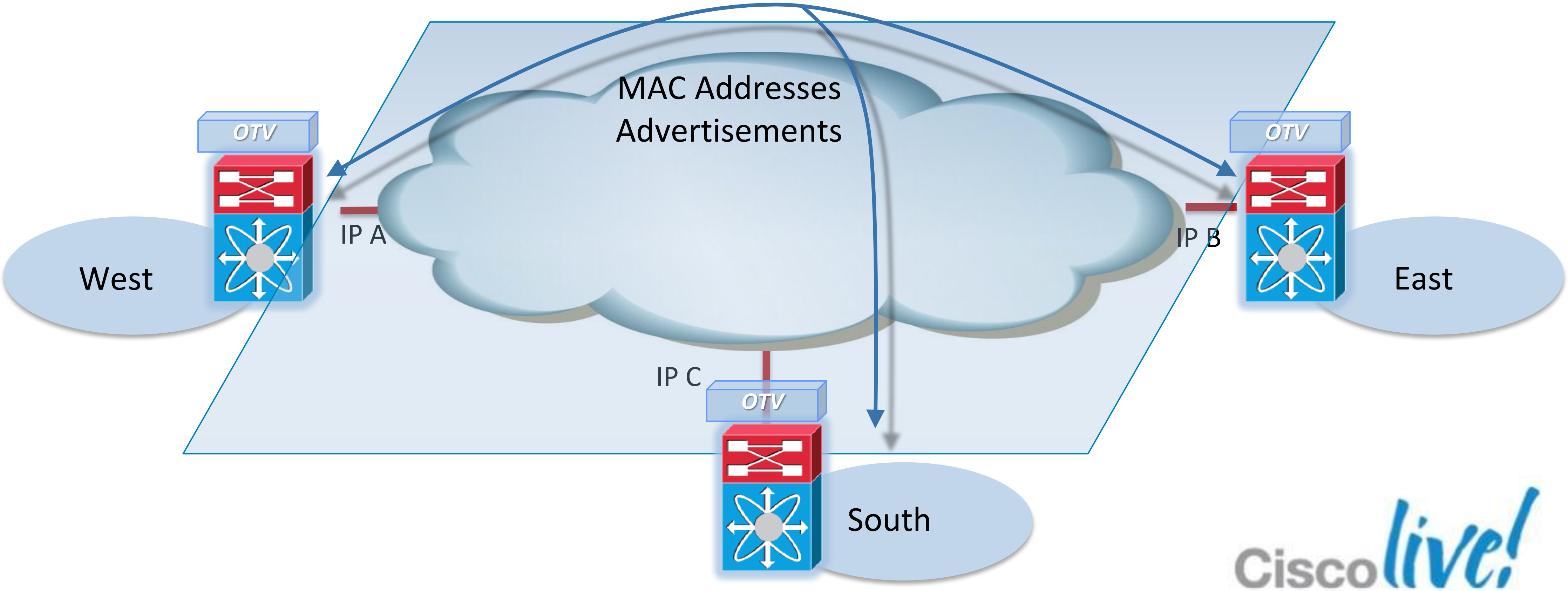


* The 4 Bytes of .1Q header have already been removed

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



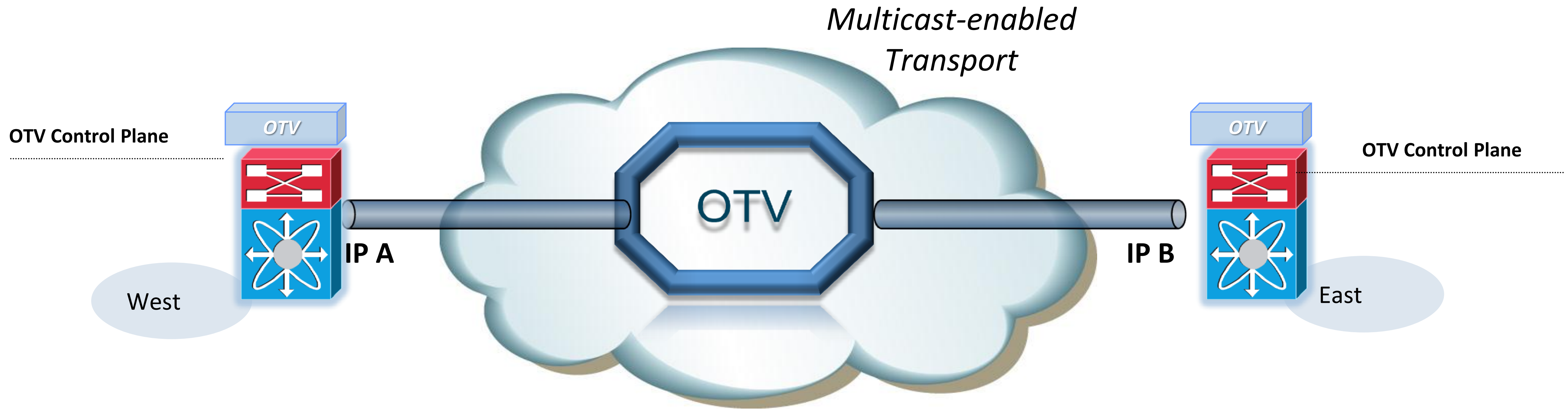
OTV Control Plane

Neighbour Discovery and Adjacency Formation

- Before any MAC address can be advertised the OTV Edge Devices must:
  - –Discover each other
  - –Build a neighbour relationship with each other
- Neighbour Relationship built over a transport infrastructure:
  - –Multicast-enabled (all shipping releases)
  - –Unicast-only (from release 5.2)

OTV Control Plane

Neighbour Discovery (over Multicast Transport)



Mechanism

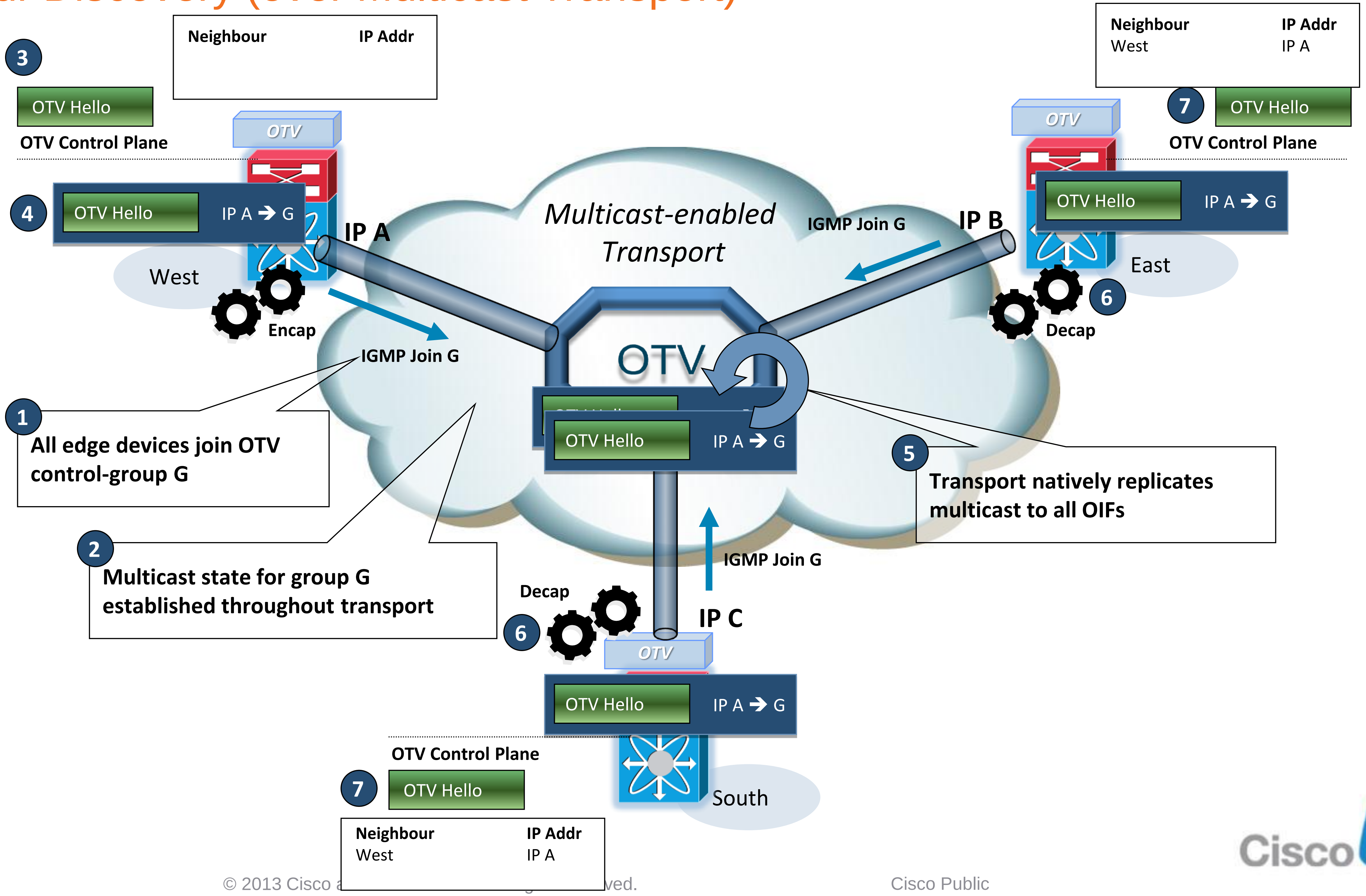
- Edge Devices (EDs) join an multicast group in the transport, as they were hosts (no PIM on EDs)
- OTV hellos and updates are encapsulated in the multicast group

End Result

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

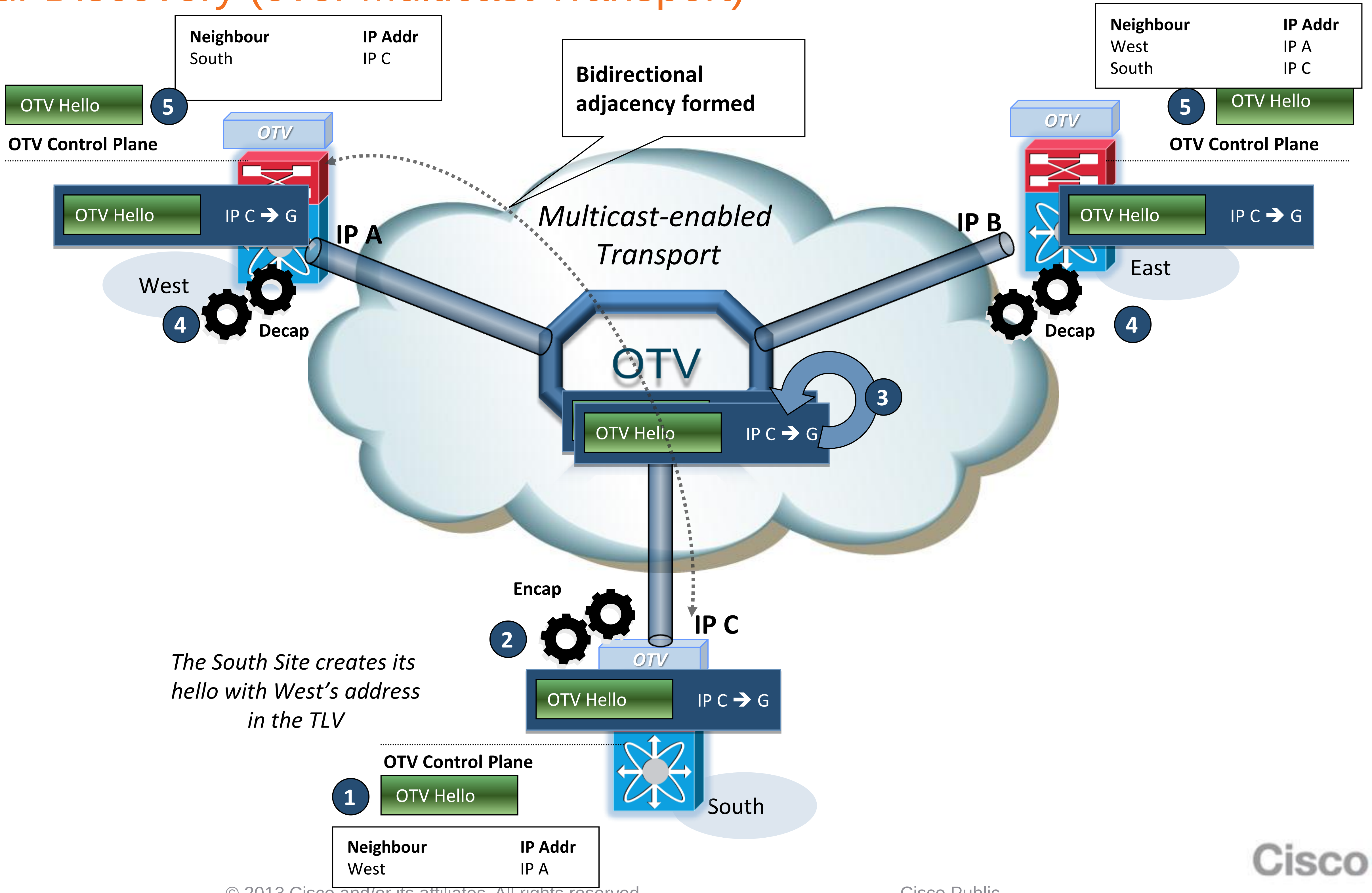
OTV Control Plane

Neighbour Discovery (over Multicast Transport)



OTV Control Plane

Neighbour Discovery (over Multicast Transport)



OTV Control Plane

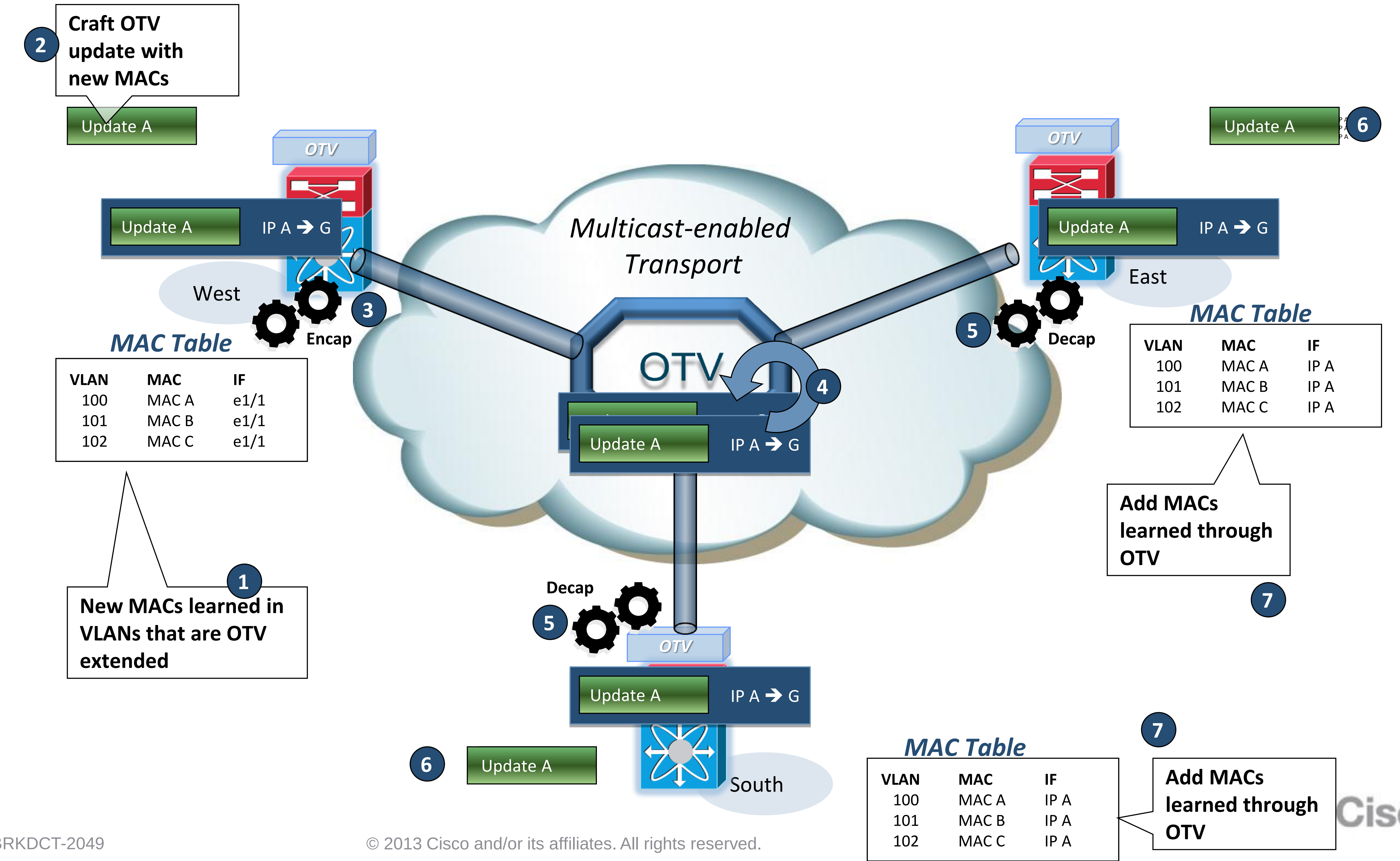
MAC Advertisements (over Multicast Transport)

- MAC addresses are advertised with their VLAN IDs, IP next hop and Site-ID
- IP next hops are the addresses of Edge Devices' Join interfaces*
- Each OTV update can contain multiple MAC addresses for different VLANs
- When the MAC address ages out from the OTV Device MAC Table, an update is created and sent to the remote OTV Edge Devices (MAC Withdraw)

** In a future release (6.2) the "IP next hops" can be loopback addresses*

OTV Control Plane

MAC Advertisements (over Multicast Transport)



OTV Configuration

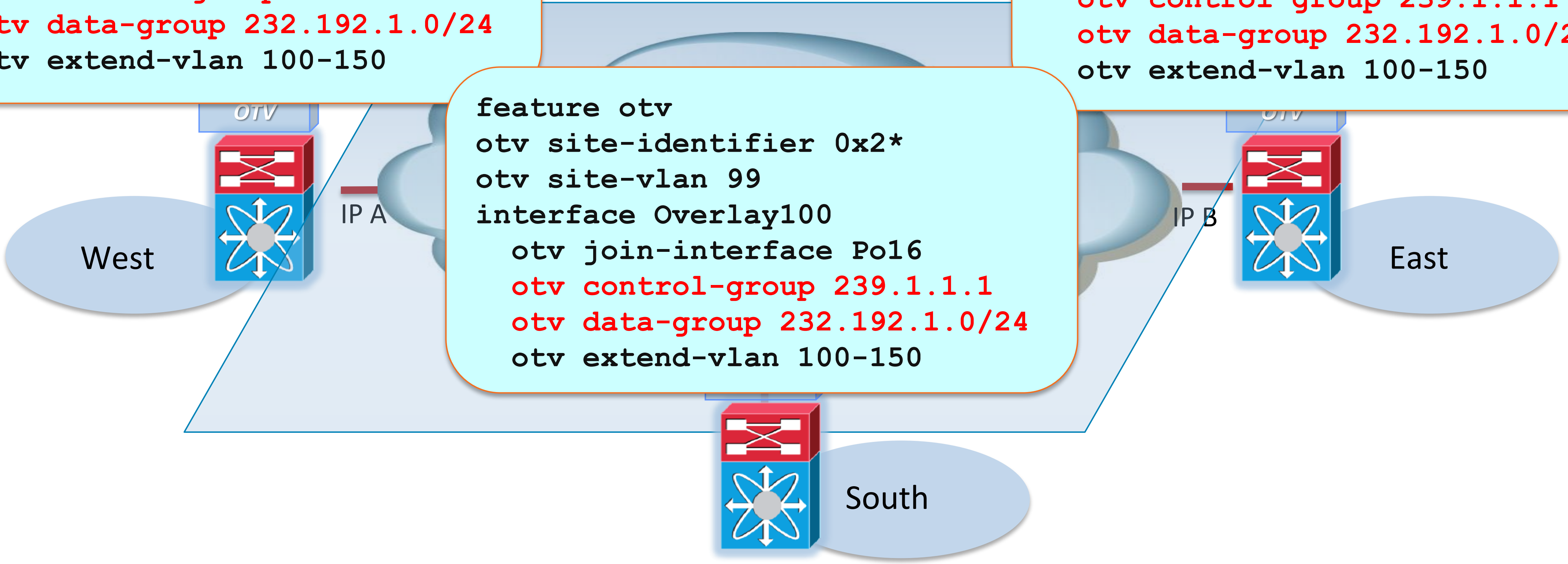
OTV over Multicast Transport

- Minimal configuration required to get OTV up and running

```
feature otv
otv site-identifier 0x1*
otv site-vlan 99
interface Overlay100
  otv join-interface e1/1
  otv control-group 239.1.1.1
  otv data-group 232.192.1.0/24
  otv extend-vlan 100-150
```

```
feature otv
otv site-identifier 0x3*
otv site-vlan 99
interface Overlay100
  otv join-interface e1/1.10
  otv control-group 239.1.1.1
  otv data-group 232.192.1.0/24
  otv extend-vlan 100-150
```

```
feature otv
otv site-identifier 0x2*
otv site-vlan 99
interface Overlay100
  otv join-interface Po16
  otv control-group 239.1.1.1
  otv data-group 232.192.1.0/24
  otv extend-vlan 100-150
```



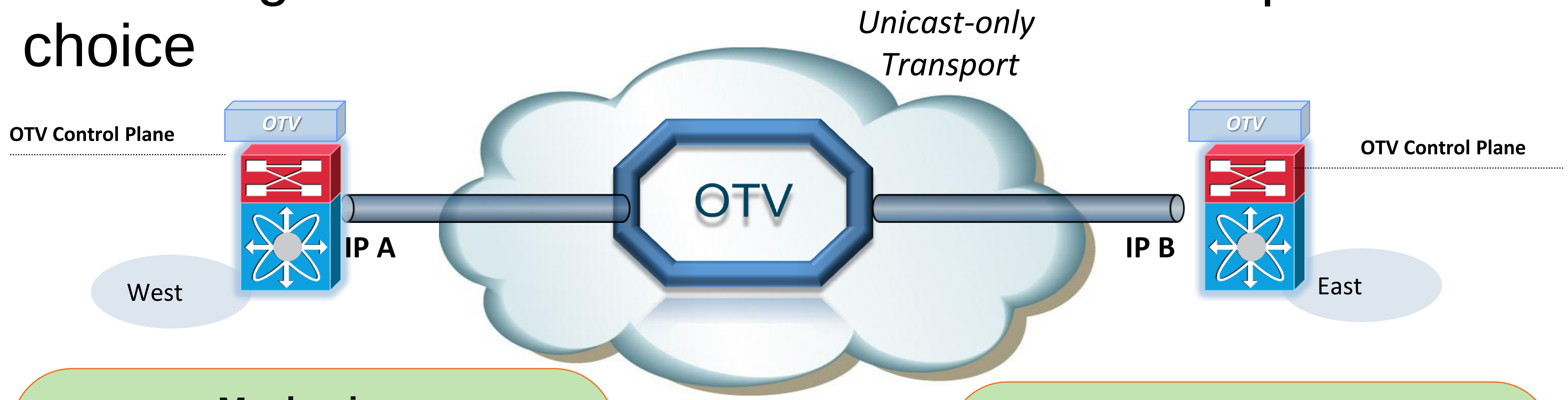
*Introduced from release 5.2

OTV Control Plane

Neighbour Discovery (Unicast-only Transport)

Release 5.2 and above

- 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"*

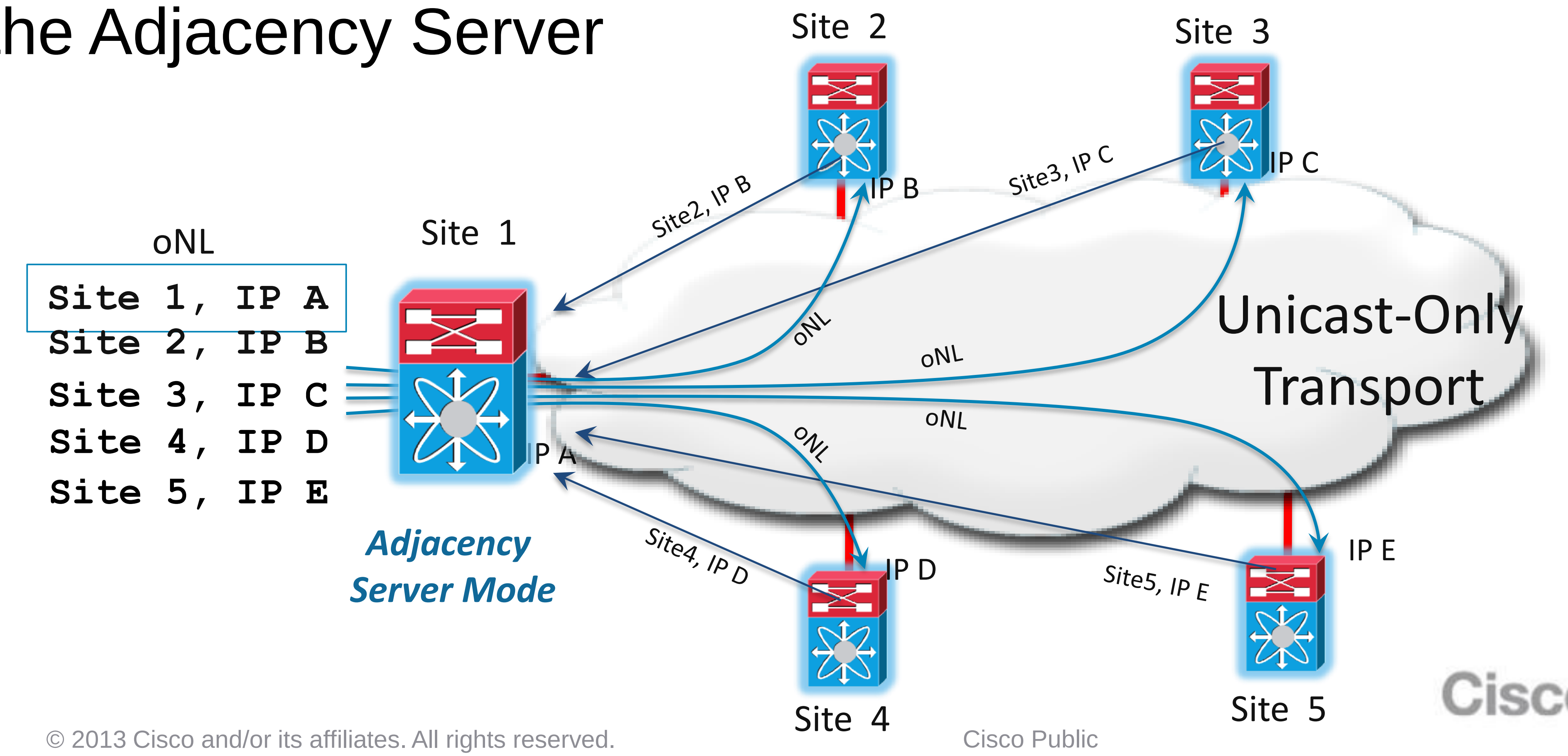
All signalling must be replicated for each neighbour

Data traffic must also be replicated at the head-end

OTV Control Plane

Neighbour Discovery (Unicast-only Transport)

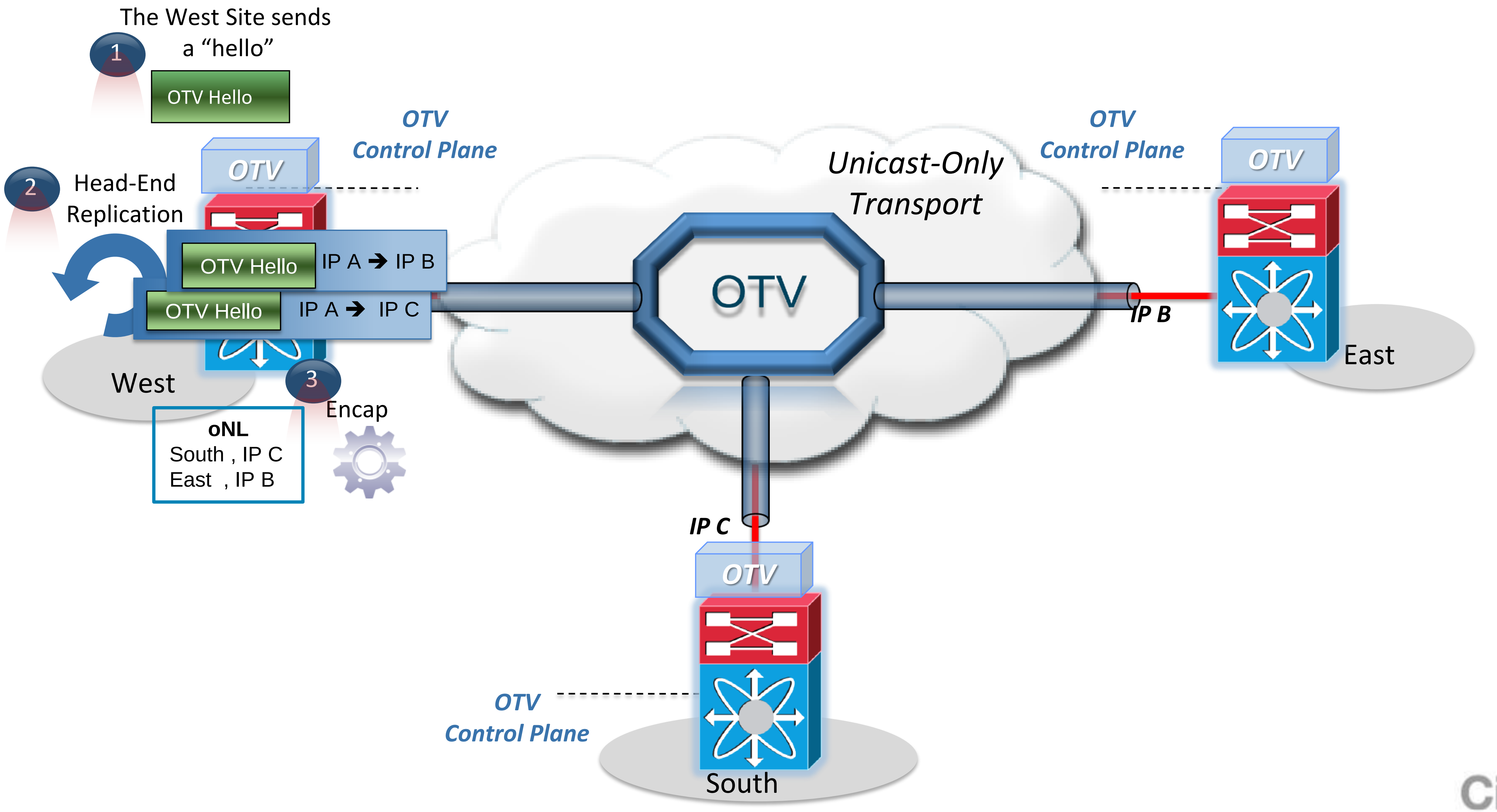
- Adjacency server is a process that can run on any OTV edge device
- Advertises IP of each Edge Device (ED) to all other EDs (OTV neighbour list – oNL)
- All subsequent communications happen directly between EDs without going through the Adjacency Server



OTV Control Plane

Neighbour Discovery (Unicast-Only Transport)

Release 5.2 and above

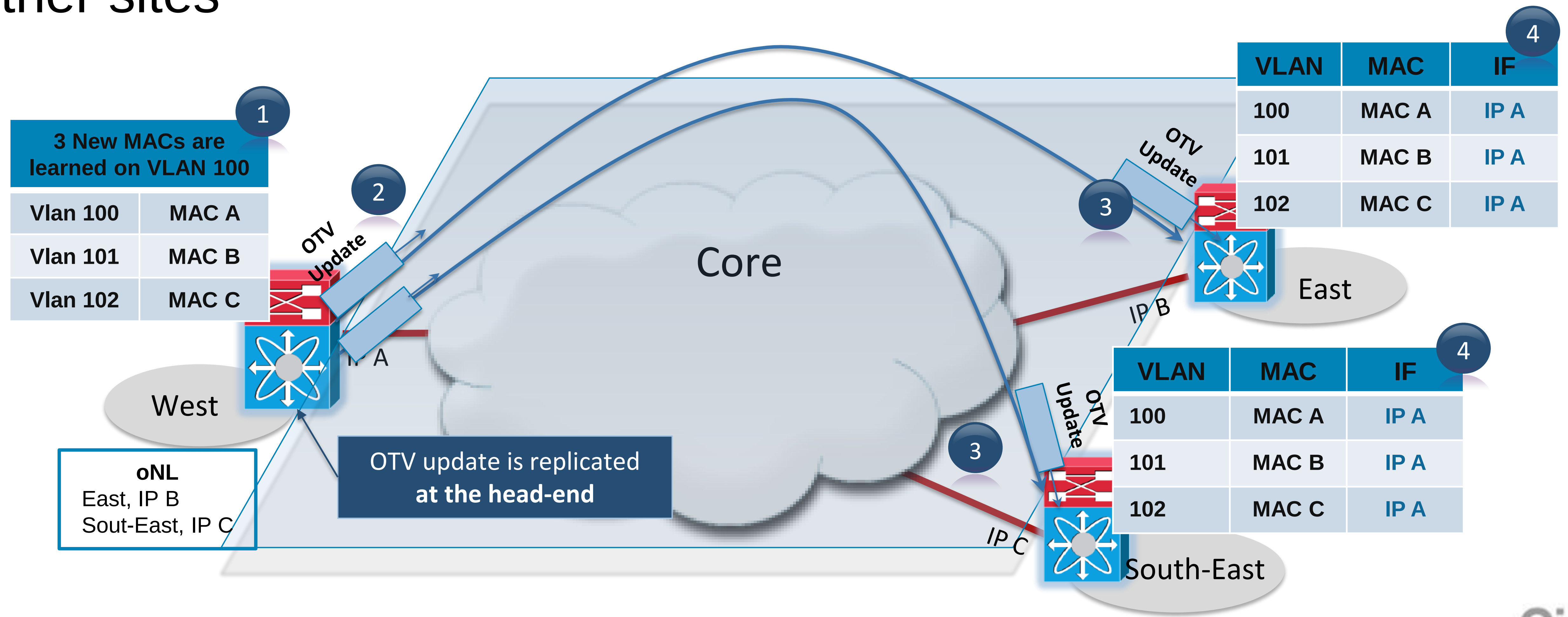


Release 5.2 and above

OTV Control Plane

MAC Advertisements (Unicast-Only Transport)

- A single update needs to be created for each destination Edge Device present on the Overlay
- Same for the sites' multicast and broadcast packets to be sent to the other sites



Release 5.2 and above

OTV Configuration

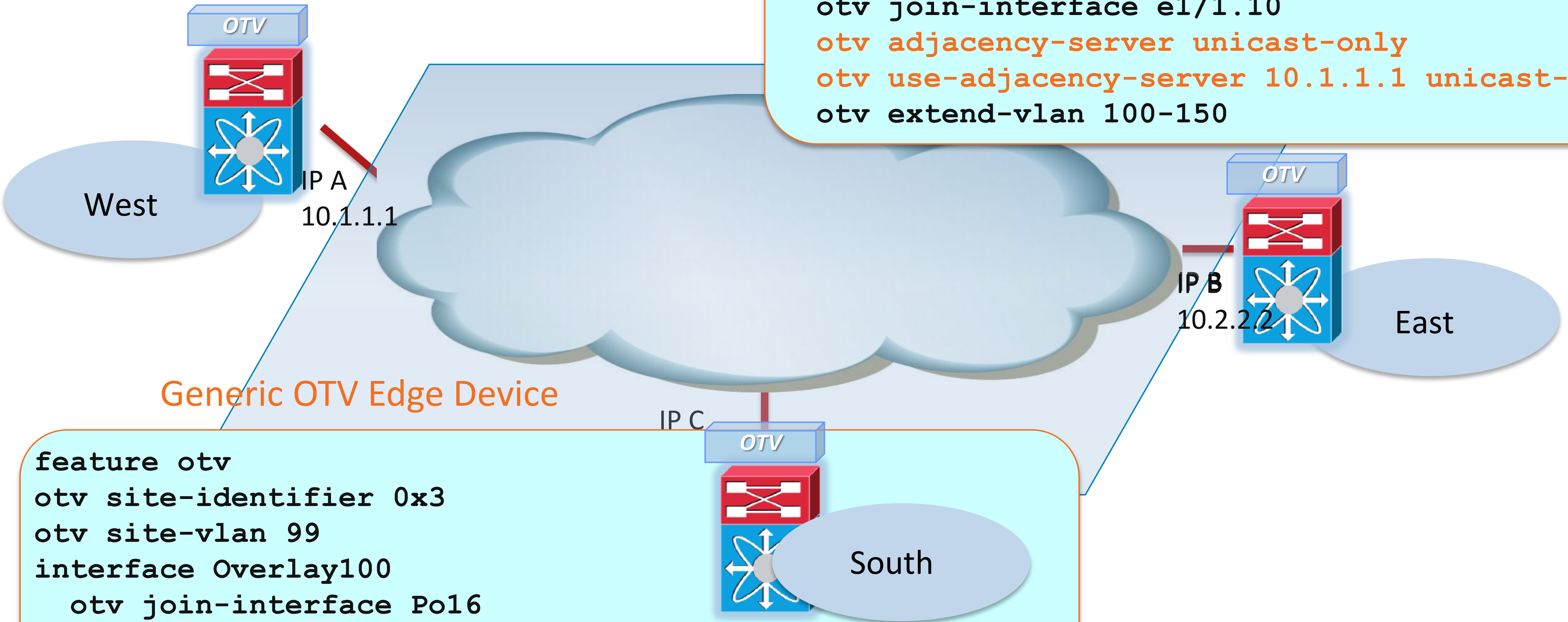
OTV over a Unicast only Transport

Primary Adjacency Server

```
feature otv
otv site-identifier 0x1
otv site-vlan 99
interface Overlay100
  otv join-interface e1/1
  otv adjacency-server unicast-only
  otv extend-vlan 100-150
```

Secondary Adjacency Server

```
feature otv
otv site-identifier 0x2
otv site-vlan 99
interface Overlay100
  otv join-interface e1/1.10
  otv adjacency-server unicast-only
  otv use-adjacency-server 10.1.1.1 unicast-only
  otv extend-vlan 100-150
```



Generic OTV Edge Device

```
feature otv
otv site-identifier 0x3
otv site-vlan 99
interface Overlay100
  otv join-interface Po16
  otv use-adjacency-server 10.1.1.1 10.2.2.2 unicast-only
  otv extend-vlan 100-150
```

OTV Control Plane

CLI Verification

- Establishment of control plane adjacencies between OTV Edge Devices (multicast or unicast transport):

```
dc1-agg-7k1# show otv adjacency

Overlay Adjacency database
Overlay-Interface Overlay100  :
Hostname       System-ID        Dest Addr     Up Time     Adj-State
dc2-agg-7k1    001b.54c2.efc2   20.11.23.2    15:08:53    UP
dc1-agg-7k2    001b.54c2.e1c3   20.12.23.2    15:43:27    UP
dc2-agg-7k2    001b.54c2.e142   20.22.23.2    14:49:11    UP
```

- Unicast MAC reachability information:

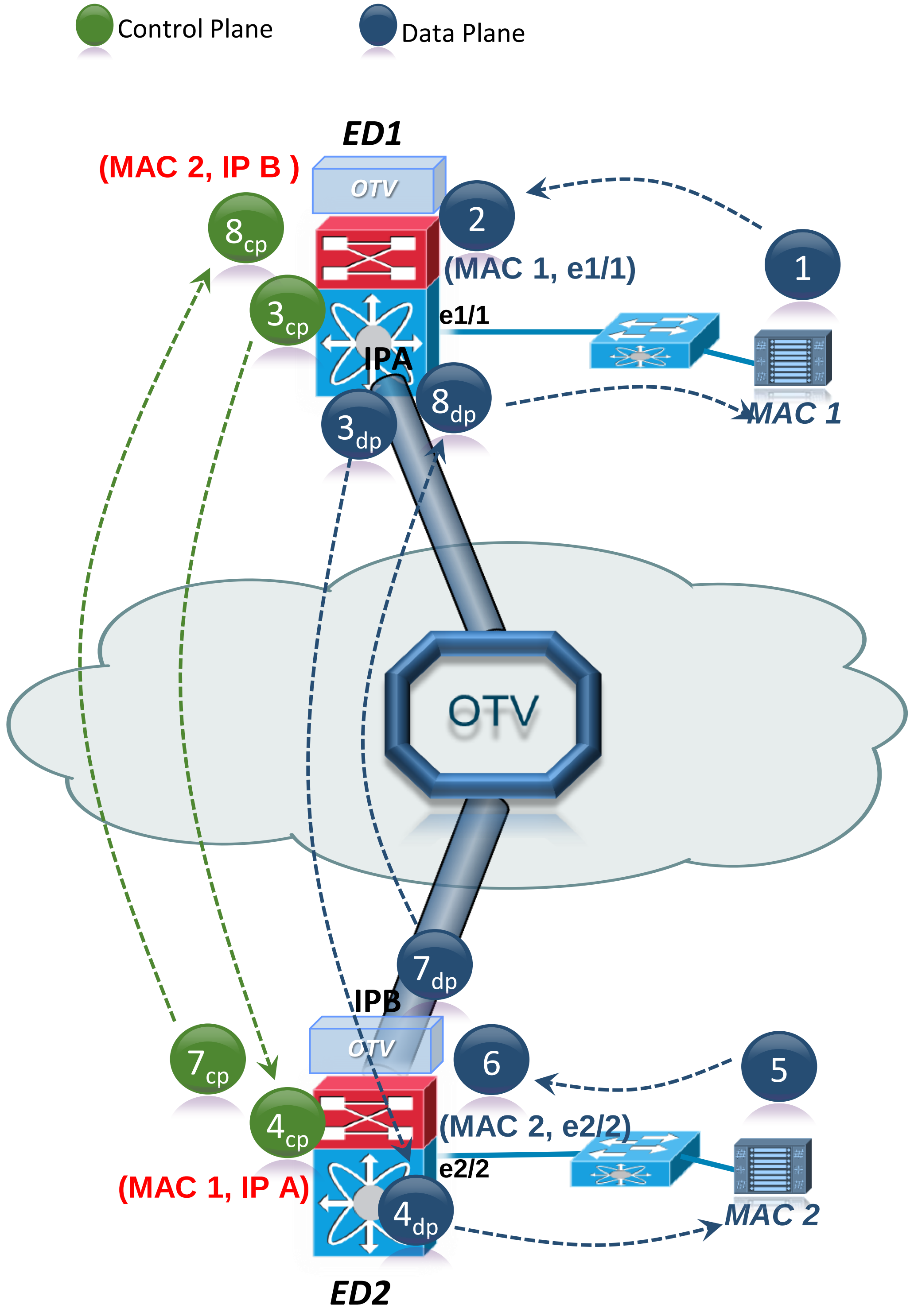
```
dc1-agg-7k1# show otv route
OTV Unicast MAC Routing Table For Overlay100
VLAN MAC-Address     Metric Uptime    Owner      Next-hop(s)
---- -------------- ------ --------  ---------  -----------
2001 0000.0c07.ac01  1      3d15h     site       Ethernet1/1
2001 0000.1641.d70e  1      3d15h     site       Ethernet1/2
2001 0000.49f3.88ff  42     2d22h     overlay    dc2-agg-7k1
2001 0000.49f3.8900  42     2d22h     overlay    dc2-agg-7k2
```

Local Site MAC

Remote Site MAC

OTV Packet Walk

Establishing Inter-Site Unicast Communication



1 – Broadcast ARP for MAC 2

2 – Broadcast ARP received by ED1. MAC 1 is learnt on its internal interface

3_{cp} – ED1 advertises MAC 1 in an OTV Update sent to the other EDs part of the Overlay

4_{cp} – ED2 receives the update and stores MAC1 in MAC table, next-hop is ED1

3_{dp} – ED1 forward the broadcast frame to the Overlay. All EDs receive it

4_{dp} – ED2 decapsulates the frame and forwards the ARP broadcast request into the site

5 – Server 2 receives the ARP and replies with a unicast ARP to MAC 1

6 – ED2 learns MAC 2 on its internal interface

7_{cp} – ED2 advertises MAC 2 with an update sent to the other EDs

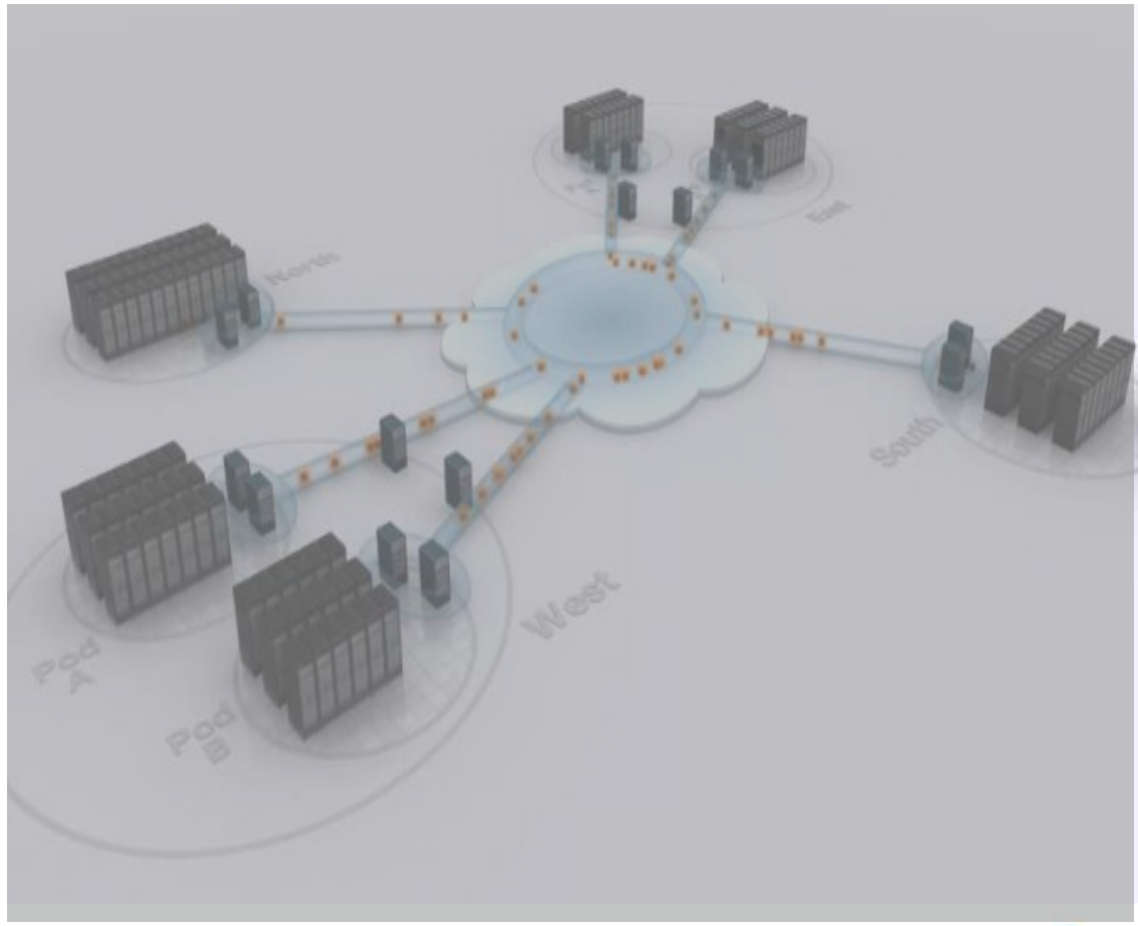
8_{cp} – ED1 receives the update and stores MAC2 in MAC table, next-hop is ED2

7_{dp} – ED2 knows that MAC 1 is reachable via IP A. It encapsulates the packet (IP A is dest IP) and sends it unicast to ED1

8_{dp} – Core delivers packet to ED1. ED1 decapsulates and forwards it into the site to MAC 1

Agenda

- Distributed Data Centres: Goals and Challenges
- OTV Architecture Principles
  - Control Plane and Data Plane
  - Failure Isolation
  - Multi-homing
  - L2 Multicast Forwarding
  - QoS and Scalability
  - Path Optimisation
- OTV at the Aggregation Layer

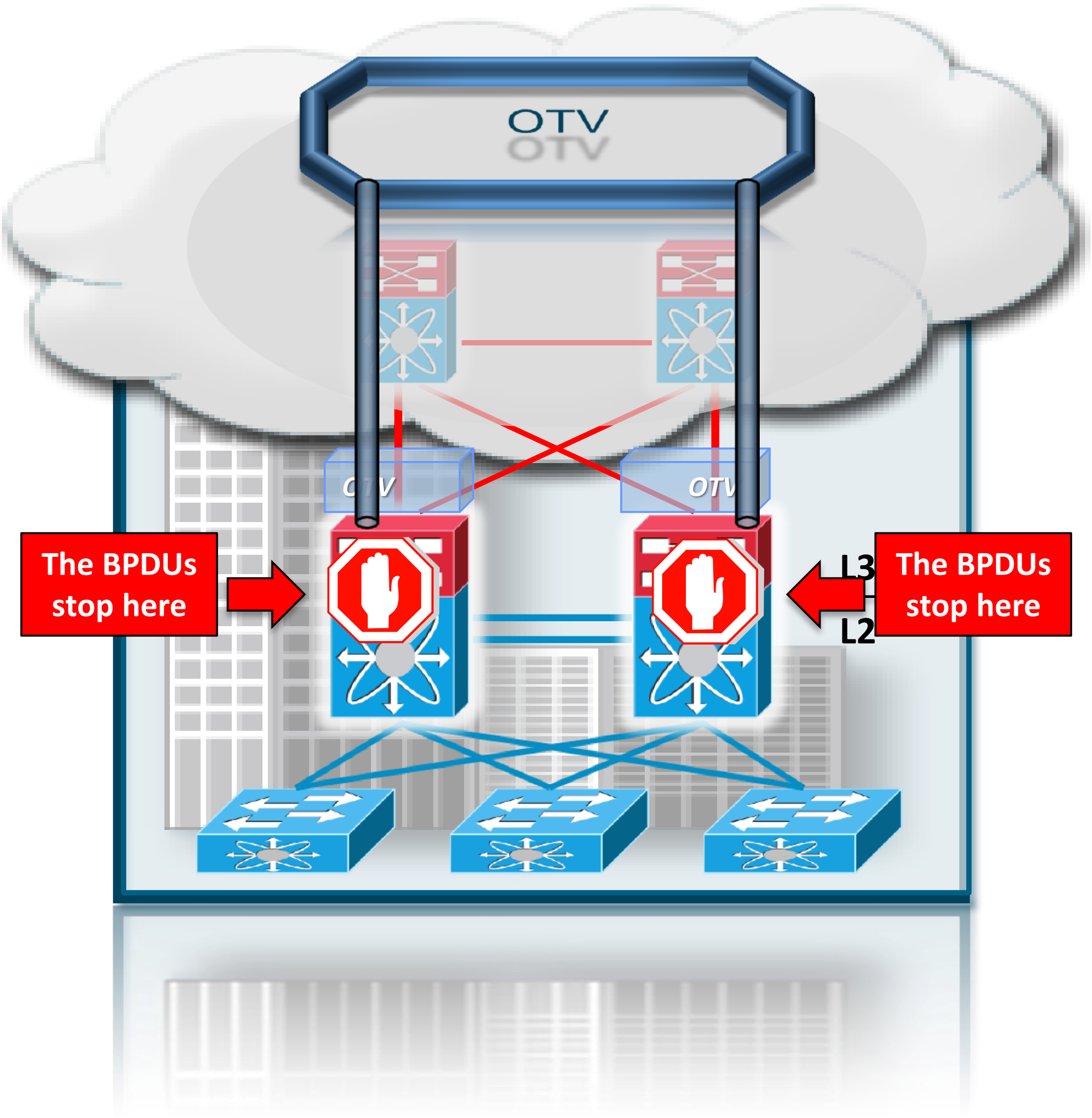


Cisco live!

Spanning Tree and OTV

Site Independence

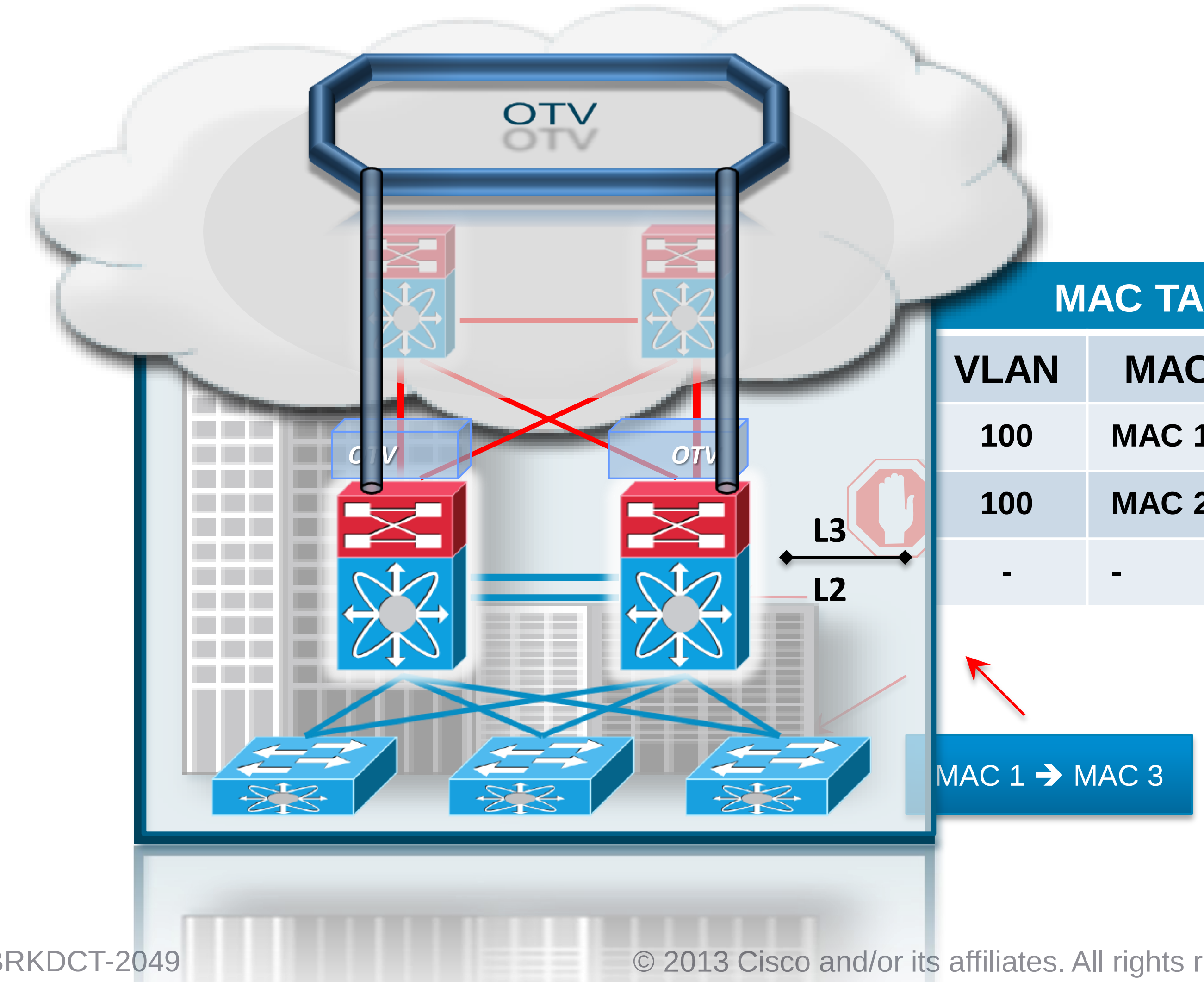
- Site transparency: no changes to the STP topology
- Total isolation of the STP domain
- **Default behaviour: no configuration is required**
- BPDUs sent and received ONLY on Internal Interfaces



Unknown Unicast and OTV

No Longer Unknown Unicast Storms Across the DCI

- No requirements to forward unknown unicast frames
- Assumption: end-host are not silent or uni-directional
- **Default behaviour: no configuration is required**



MAC TABLE

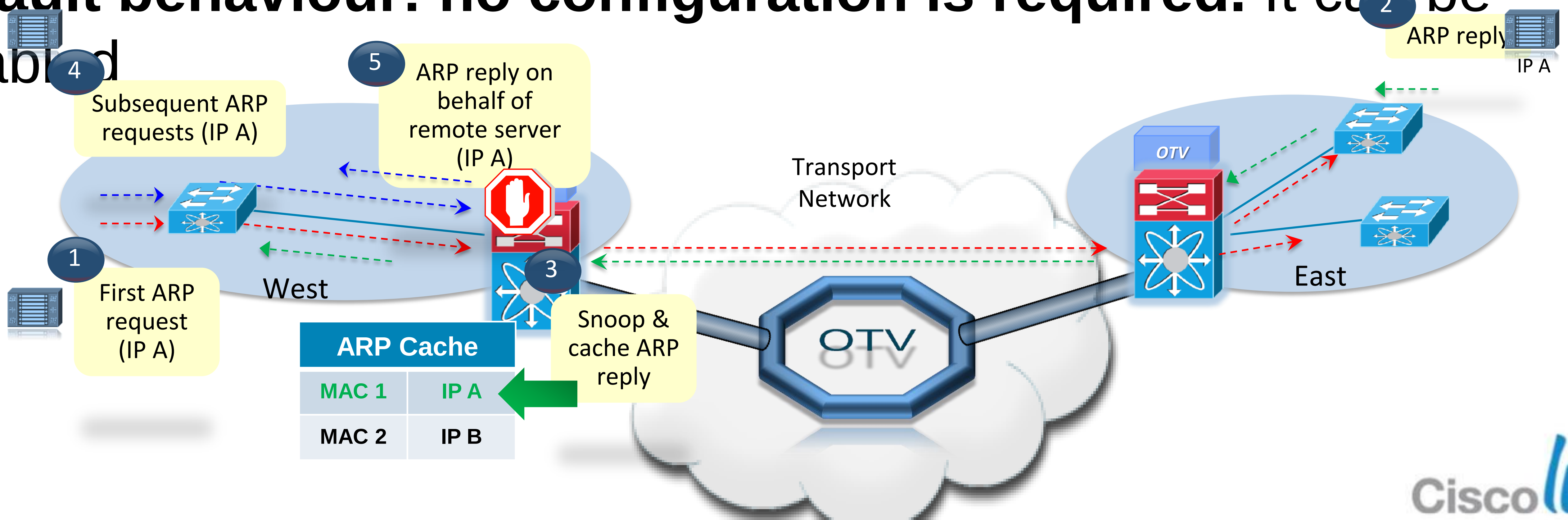
VLAN	MAC	IF
100	MAC 1	Eth1
100	MAC 2	IP B
-	-	

No MAC 3 in the MAC Table

Controlling ARP Traffic

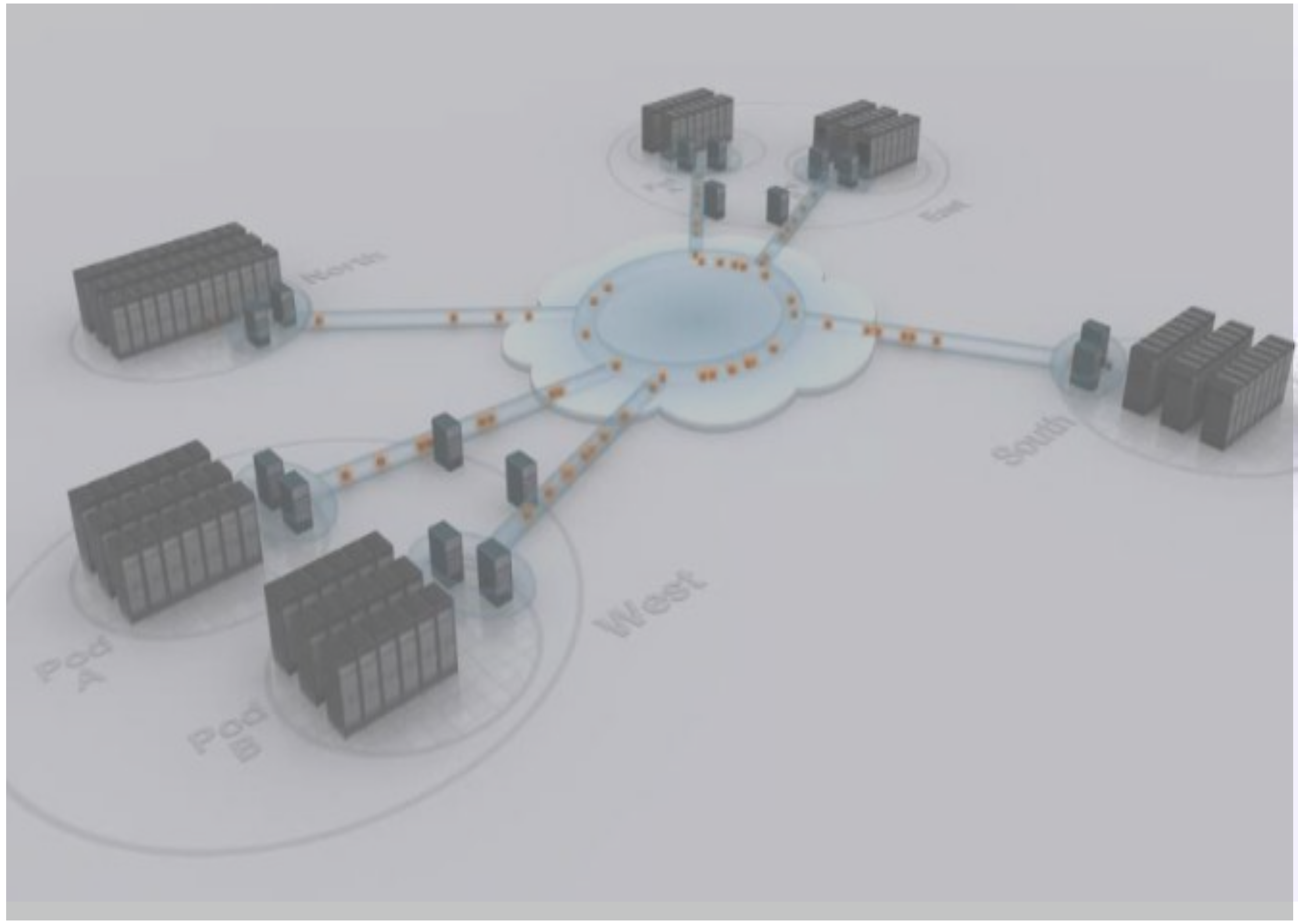
ARP Neighbour-Discovery (ND) Cache

- ARP cache maintained in Edge Device by snooping ARP replies
- First ARP request is broadcasted to all sites. Subsequent ARP requests are replied by local Edge Device
- **Drastic reduction of ARP traffic on DCI**
- **Default behaviour: no configuration is required.** It can be disabled



Agenda

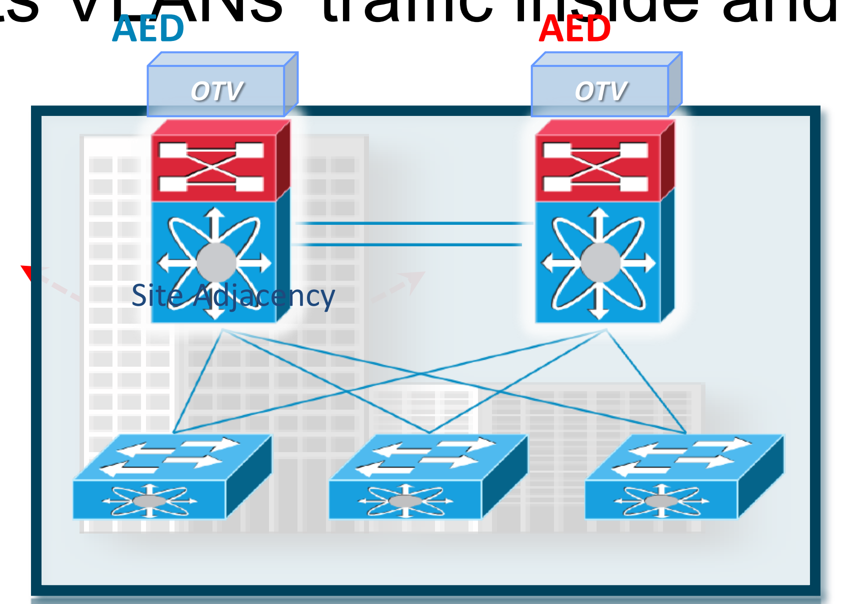
- Distributed Data Centres: Goals and Challenges
- OTV Architecture Principles
  - Control Plane and Data Plane
  - Failure Isolation
  - Multi-homing
  - L2 Multicast Forwarding
  - QoS and Scalability
  - Path Optimisation
- OTV at the Aggregation Layer



OTV Multi-homing

Fully Automated Multi-homing

- **No additional protocols required (i.e. BGP)**
- **OTV site-vlan** used to discover OTV neighbour in the same site
- **Authoritative Edge Device (AED)** Election takes place
- Extended VLANs are split across the AEDs
- The AED is responsible for:
  - MAC addresses advertisement for its VLANs
  - Forwarding its VLANs' traffic inside and outside the site



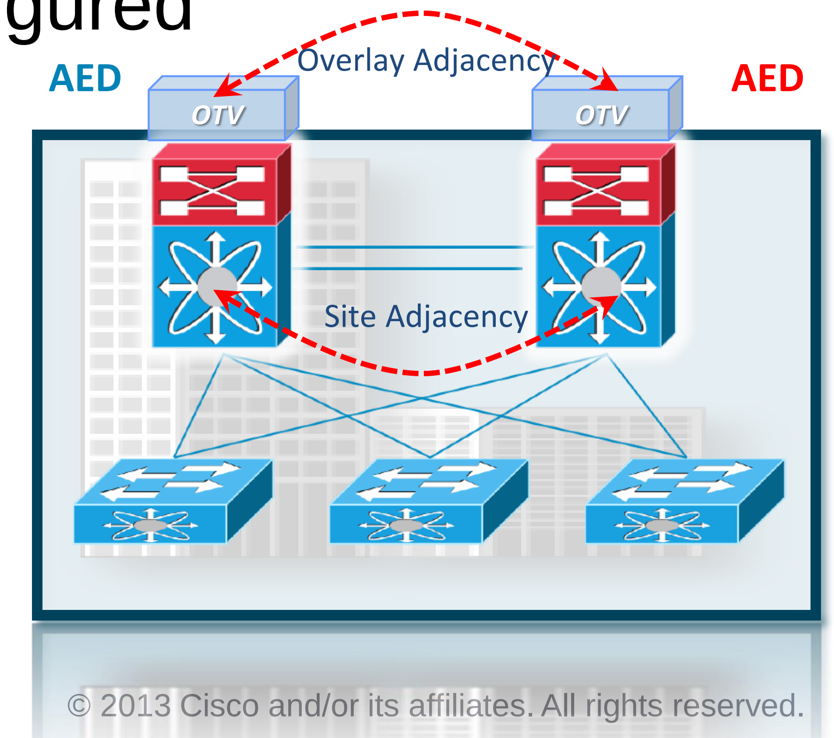
Site Adjacency used for AED election

Release 5.2 and above

Hardened Multi-homing

Introducing OTV Site-identifier

- All devices on a site must be configured with a common site-identifier
- The site-id information is included in the control plane
- Makes OTV multi-homing more robust and resilient to user configuration errors
  - Site Adjacency and Overlay Adjacency are now **both** leveraged for AED election
- An overlay will not come up until a site-id is configured
  - Warning: ISSU to 5.2 results in an overlay down condition until site-ids are manually configured

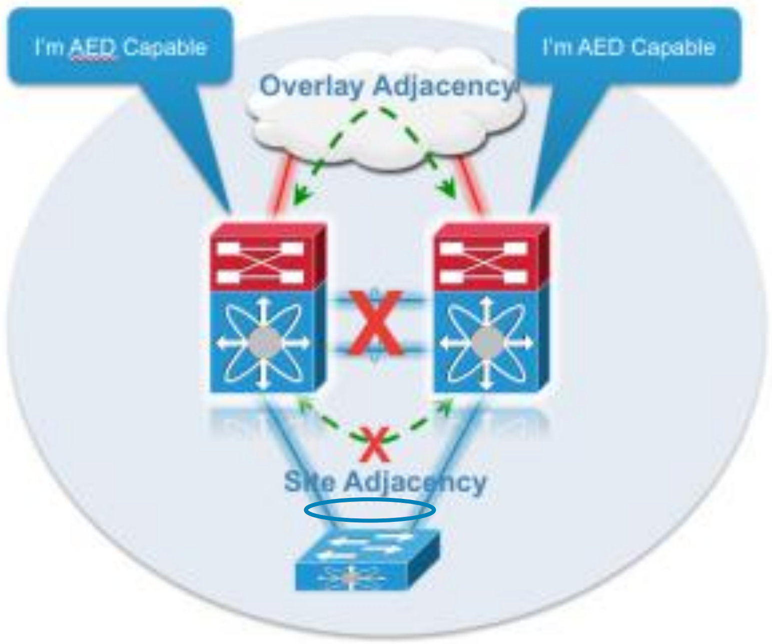



```
feature otv
otv site-identifier 0x1
otv site-vlan 99
```

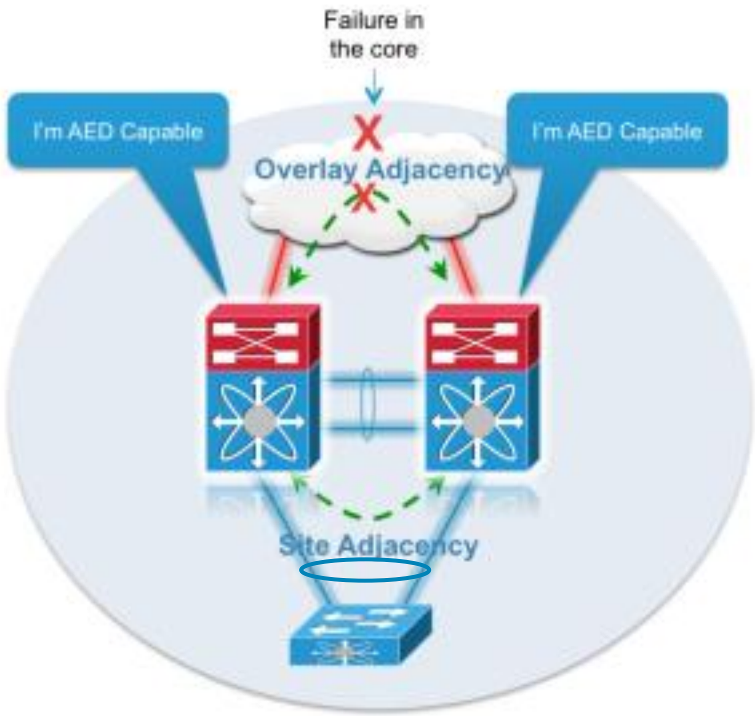
Hardened Multi-homing

Failure Scenarios

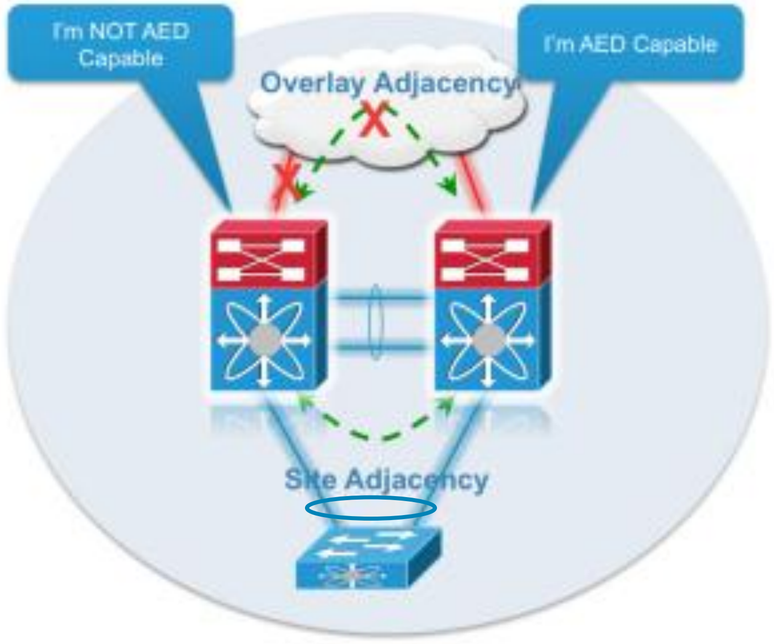
Release 5.2 and above



- Local DC failures may cause the Site Adjacency to fail
- OTV Edge Devices still communicate via the Overlay Adjacency and retain their own AED role



- Remote core failure may cause the Overlay Adjacency to fail
- OTV Edge Devices still communicate via the Site Adjacency and retain their own AED role



- Local failure of directly connected Join Interface may cause the Overlay Adjacency to fail
- OTV Edge Devices still communicate via the Site Adjacency
- The OTV Device experiencing the direct failure advertises itself as not AED capable → responsibility for LAN extension for al the VLANs is moved to the other Edge Device

OTV Multi-homing

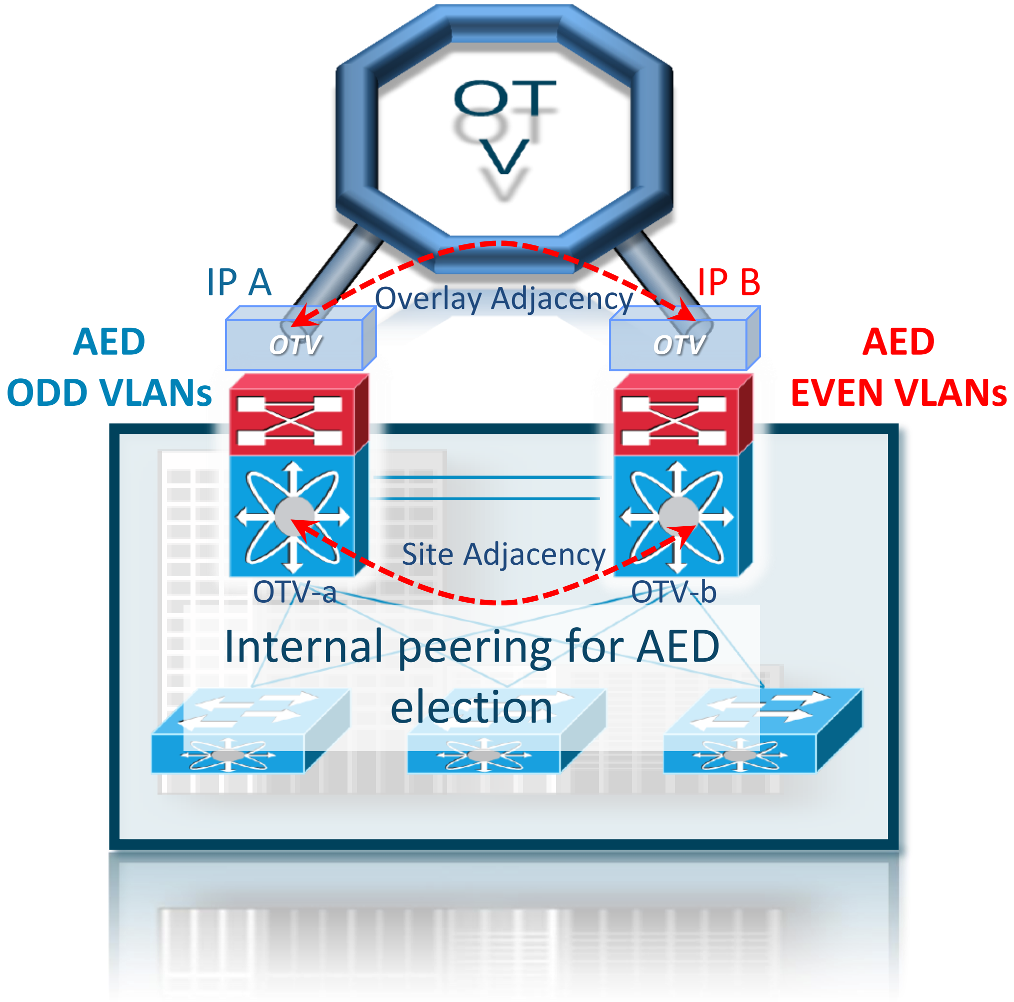
VLANs Split across AEDs

- Automated and deterministic algorithm (not configurable)
- In a dual-homed site:
  - Lower IS-IS System-ID (Ordinal 0) = EVEN VLANs
  - Higher IS-IS System-ID (Ordinal 1) = ODD VLANs
- Future functionality will allow to tune the behaviour

```
OTV-a# show otv vlan

OTV Extended VLANs and Edge Device State Information (* - AED)

VLAN   Auth. Edge Device           Vlan State         Overlay
----   ------------------          ----------         -------
 100   East-b                      inactive(Non AED)  Overlay100
 101*  East-a                      active             Overlay100
 102   East-b                      inactive(Non AED)  Overlay100
```

```
OTV-b# show otv vlan

OTV Extended VLANs and Edge Device State Information (* - AED)

VLAN   Auth. Edge Device           Vlan State         Overlay
----   ------------------          ----------         -------
 100*  East-b                      active             Overlay100
 101   East-a                      inactive(Non AED)  Overlay100
 102*  East-b                      active             Overlay100
```

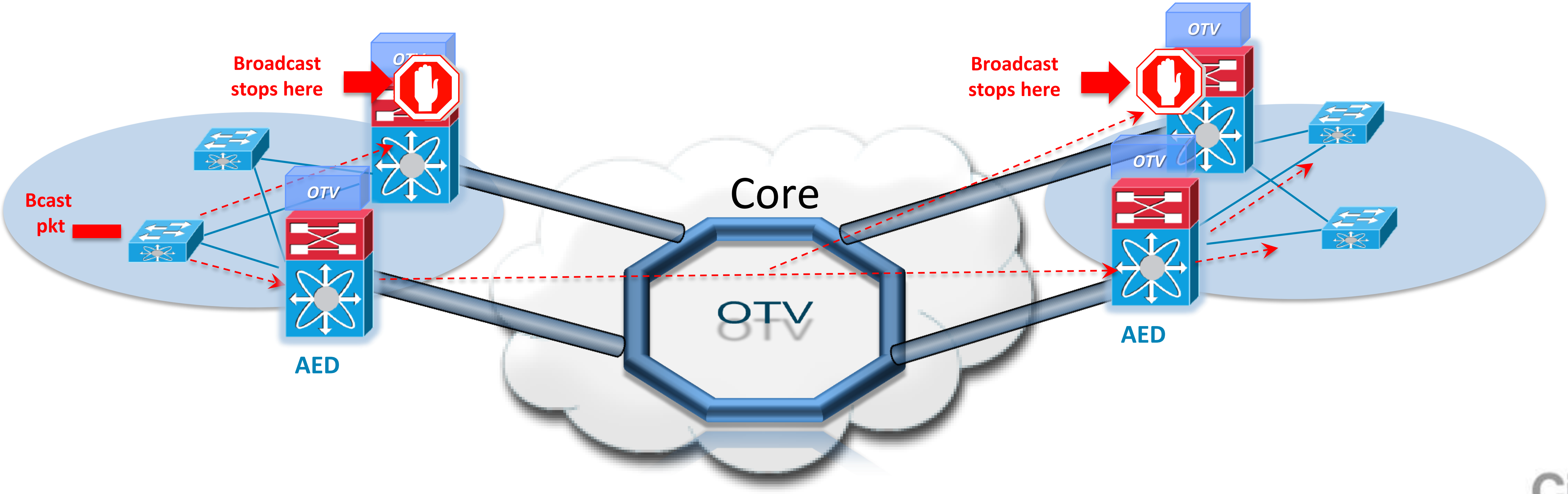
Remote OTV Device MAC Table		
VLAN	MAC	IF
100	MAC 1	IP A
101	MAC 2	IP B



OTV Multi-homing

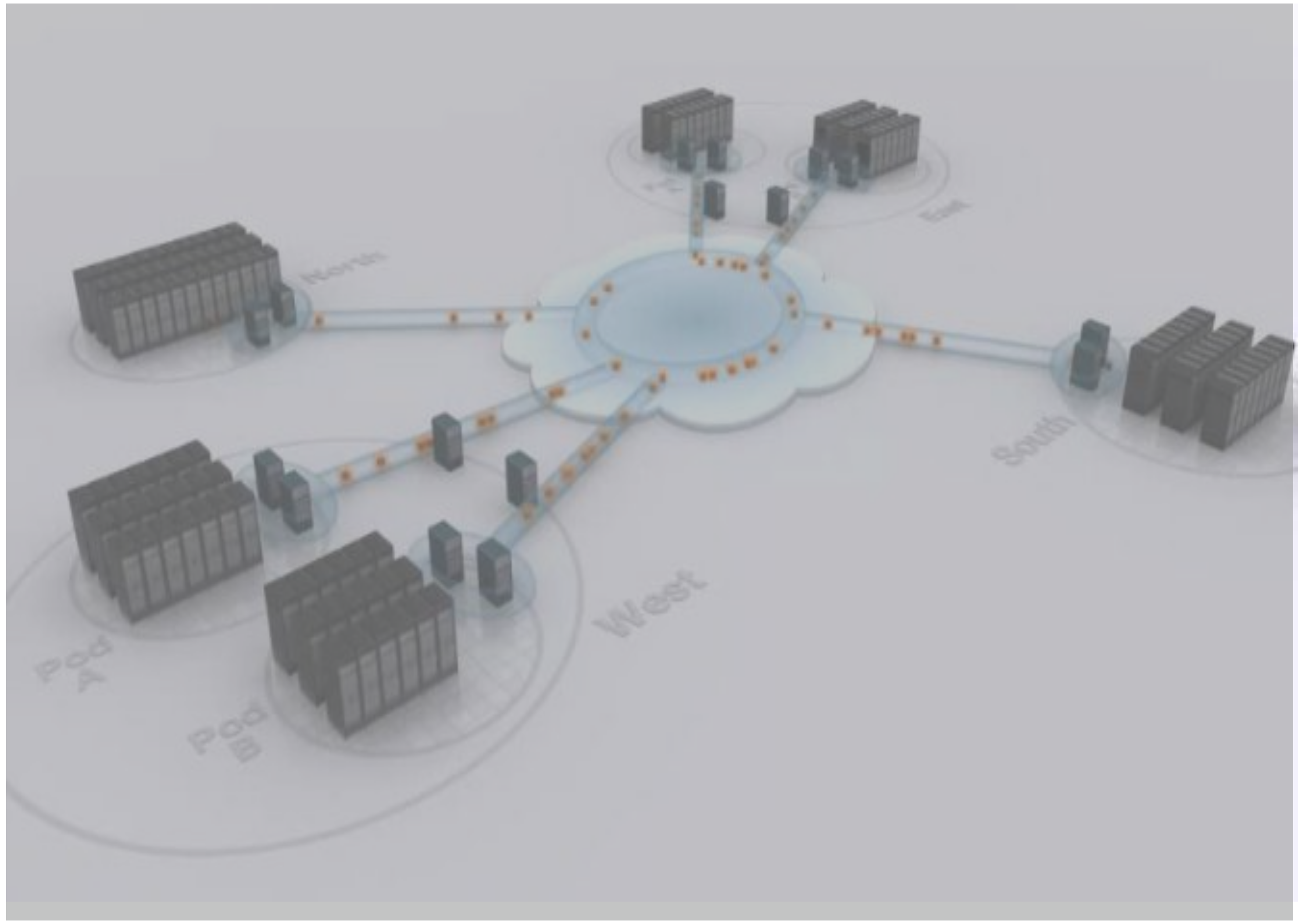
AED and Broadcast Handling

1. Broadcast reaches all the Edge Devices within the site
2. Only the AED forwards the traffic to the Overlay
3. All the Edge Devices at the other sites receive the broadcast
4. At the remote sites only the AEDs forward it into the site



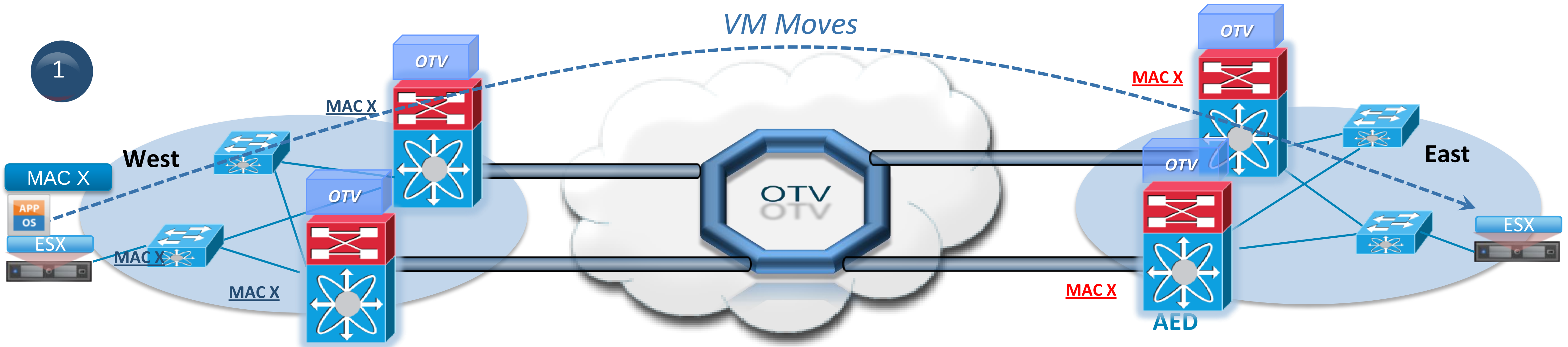
Agenda

- Distributed Data Centres: Goals and Challenges
- OTV Architecture Principles
  - Control Plane and Data Plane
  - Failure Isolation
  - Multi-homing
  - Mobility
  - L2 Multicast Forwarding
  - QoS and Scalability
  - Path Optimisation
- OTV at the Aggregation Layer



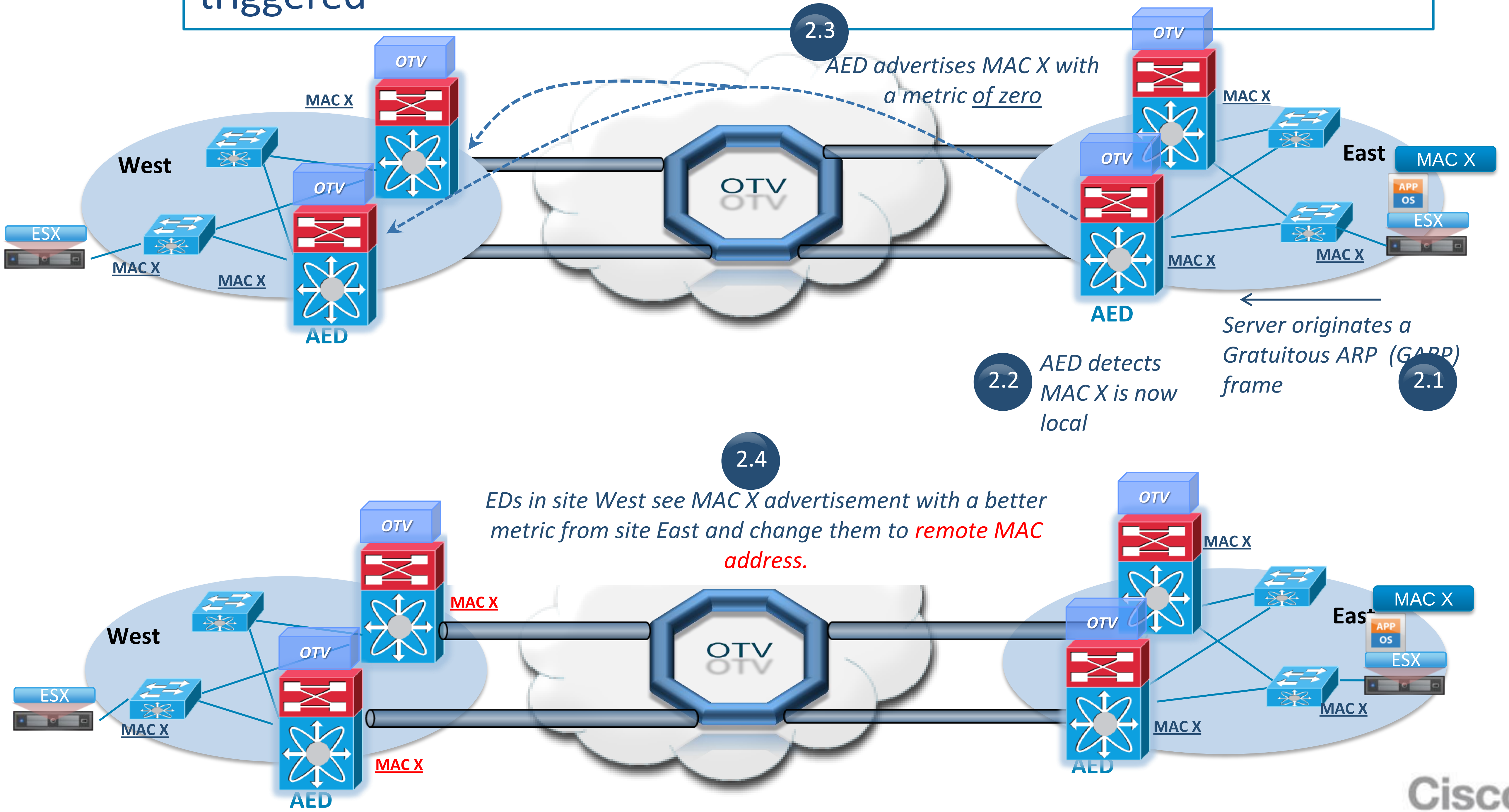
OTV and MAC Mobility (1)

Step 1 – Workload is moved between Data Centre sites



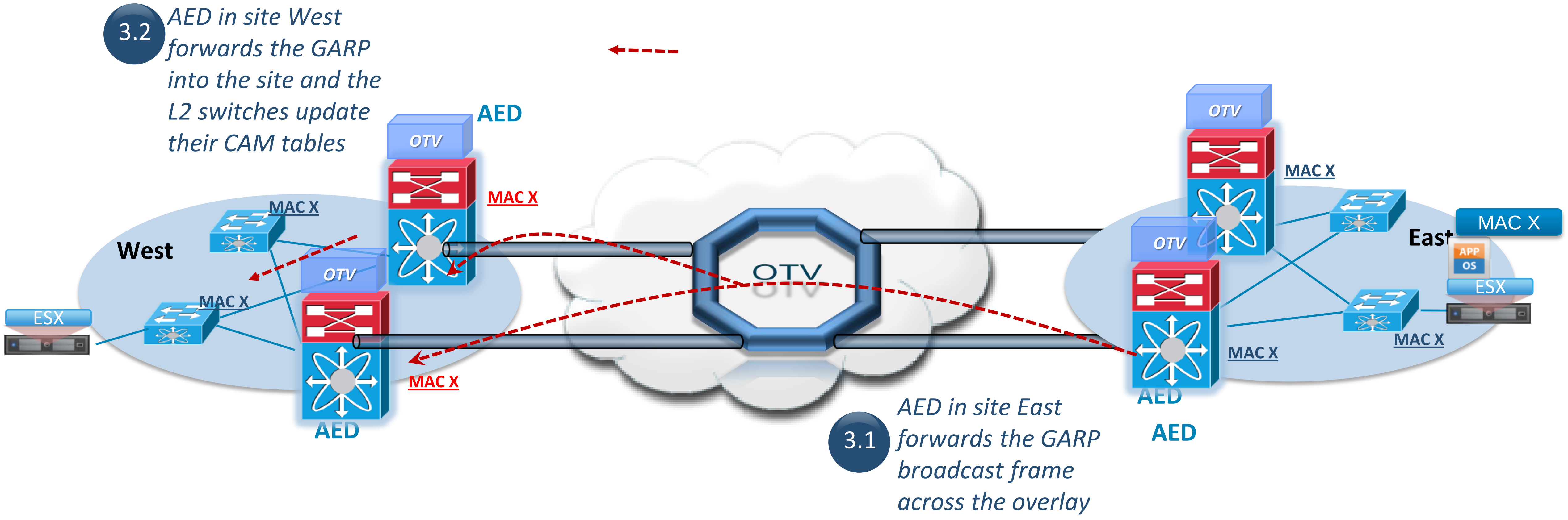
OTV and MAC Mobility (2)

Step 2 – Workload is detected in East DC and OTV control plane is triggered



OTV and MAC Mobility (3)

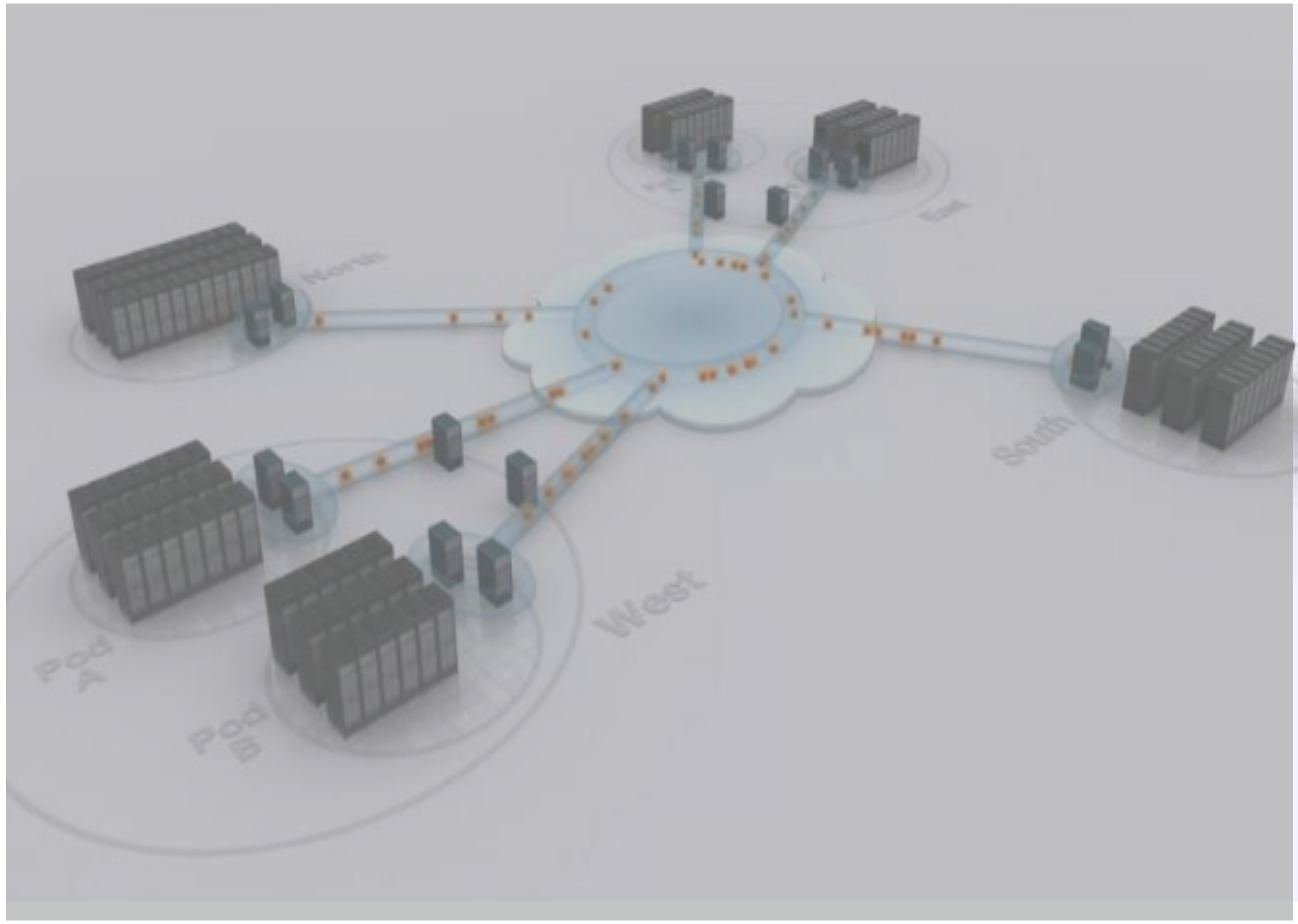
Step 3 – East to West OTV data plane traffic allows to update the MAC tables of the L2 devices in West Site



NOTE: GARP is used as example of traffic, same behaviour is achieved with any other L2 broadcast frames exchanged between East and West sites

Agenda

- Distributed Data Centres: Goals and Challenges
- OTV Architecture Principles
  - Control Plane and Data Plane
  - Failure Isolation
  - Multi-homing
  - L2 Multicast Forwarding
  - QoS and Scalability
  - Path Optimisation
- OTV at the Aggregation Layer



L2 Multicast Traffic between Sites

Multicast Enabled Transport

OTV can leverage the multicast support available in the transport network to **optimise** the delivery of the multicast traffic for the VLANs stretched across sites

Three steps:

1. **Automated mapping** of the sites' multicast groups to a **range of multicast groups** in the transport network
2. Creation of the Multicast state information at the OTV Edge Devices
3. Sites' Multicast traffic delivered over the Overlay

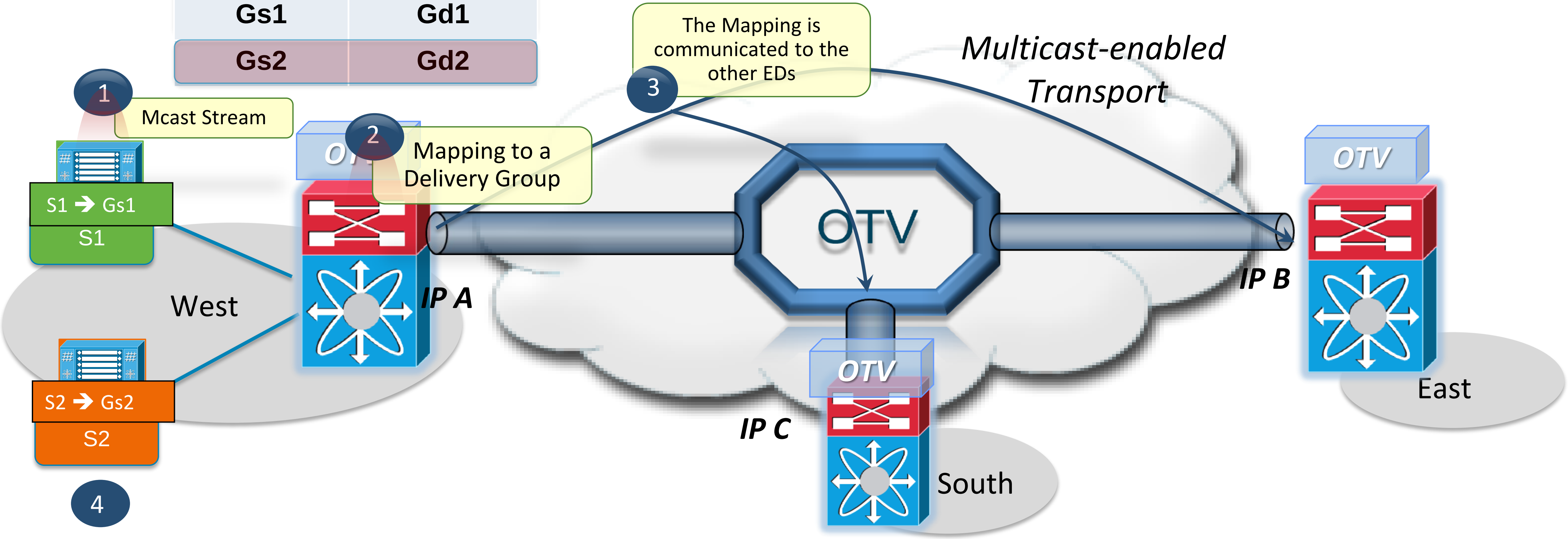
L2 Multicast with Multicast Transport

Step 1—Mapping of the Site Multicast Groups

- The site multicast groups are mapped to a ***SSM group range*** in the core
- Each (Si,Gsi) maps to a different SSM group in **round-robin fashion**

Mcast Group Mapping

Site Group	Core Group
Gs1	Gd1
Gs2	Gd2



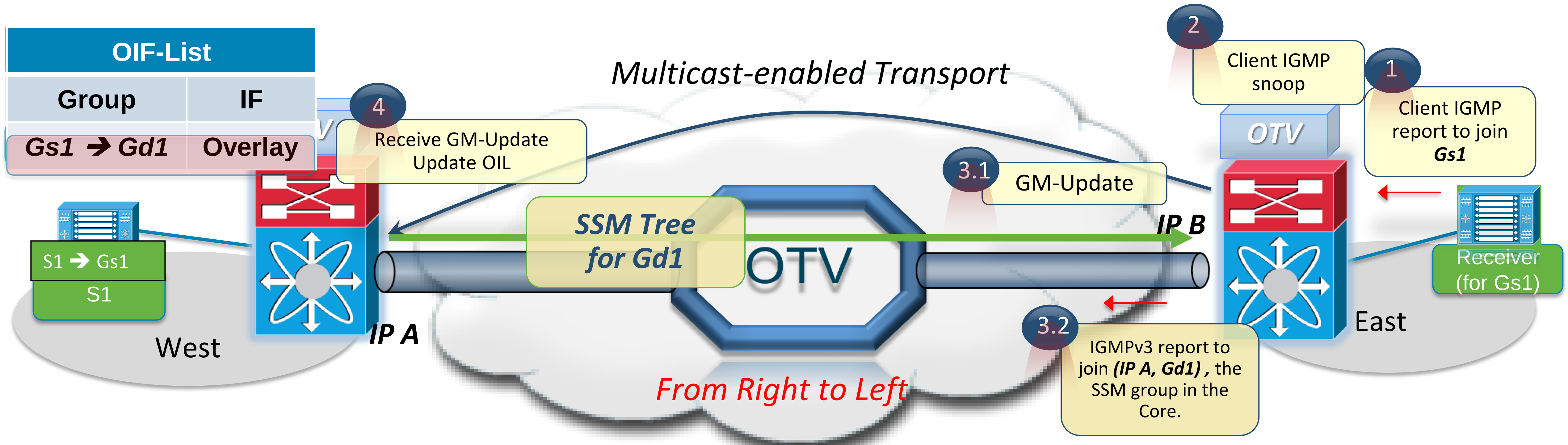
1. The Mcast source starts sending traffic to the group ***Gs1***
2. The West ED maps (S1,Gs1) to a ***delivery group Gd1***
3. The West ED communicates the mapping information (including the source VLAN) to the other EDs
4. Same process happens once source S2 is enabled (sending to a different group **Gs2**)

L2 Multicast with Multicast Transport

Step 2—Multicast State Creation

1. A receiver in the East site sends an IGMP join for Gs1
2. The OTV ED snoops the IGMP join (without forwarding it)
3. ED performs two actions:
   - Announces the receivers in a Group-Membership Update (GM-Update) to all other EDs
   - Sends an IGMPv3 report to join the ***(IP A, Gd1) SSM* group** in the core
4. The source ED adds the Overlay interface to the *Outbound Interfaces* (OIF)
5. The SSM tree for Gd1 (rooted at the source ED) is built in the core

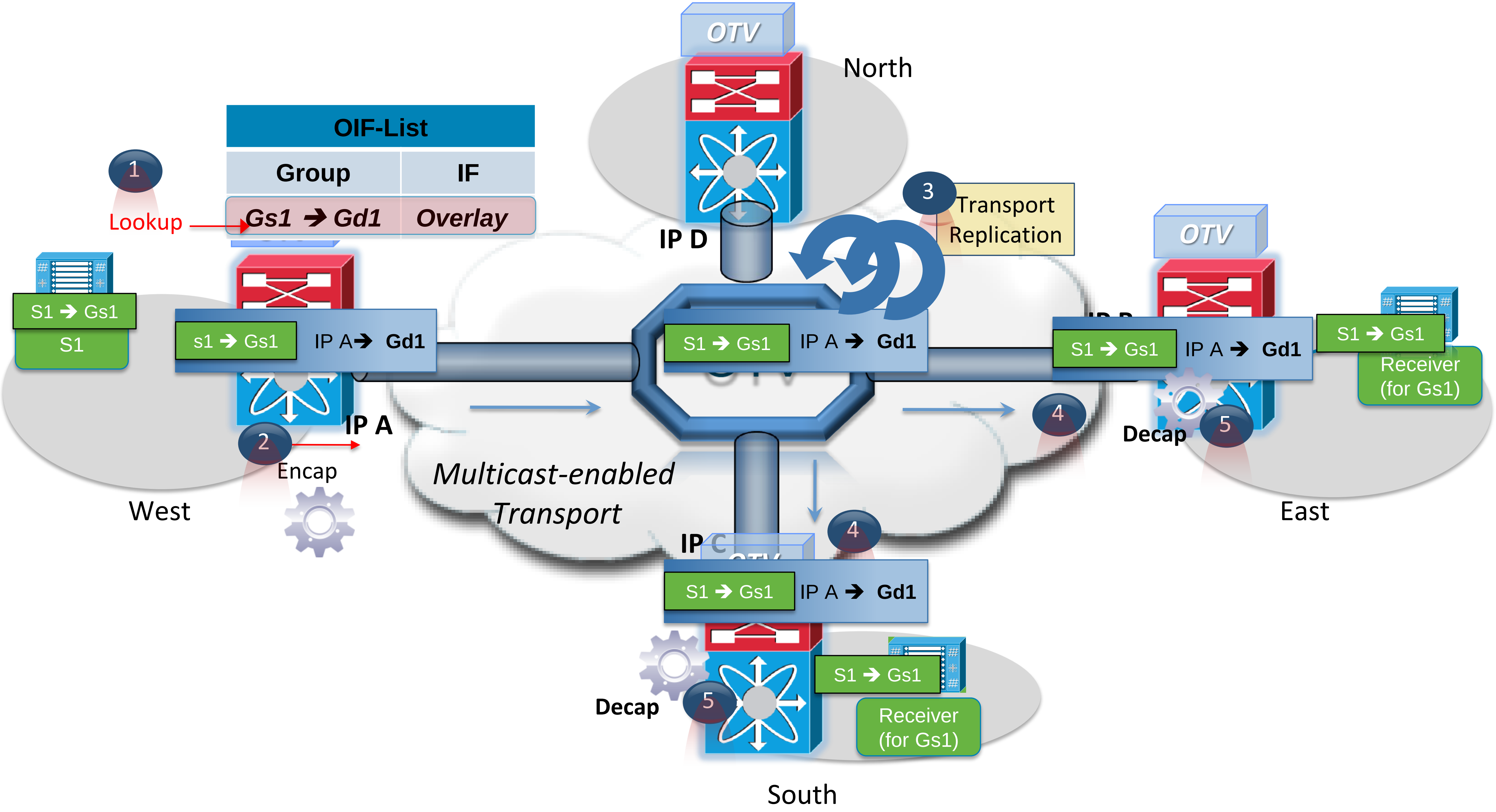
OIF-List	
Group	IF
Gs1 ➔ Gd1	Overlay



It is important to clarify that the edge devices join the core multicast groups *as hosts, not as routers!*

L2 Multicast with Multicast Transport

Step 3—Multicast Packet Flow



L2 Multicast with Multicast Transport

Multicast Groups in the Core

OTV can leverage the benefits of a multicast-enabled transport for both control and data planes. The following summarises the requirements for a multicast transport:

- **Control group** – Single PIM-SM or PIM-Bidir group used to form adjacencies and exchange MAC reachability information
- **Data groups** – Range of SSM groups used to carry multicast data traffic generated by the sites

  The right number of SSM groups to be used depends on a tradeoff between the amount of multicast state to be maintained in the core and the optimisation of Layer 2 multicast traffic delivery

```
interface Overlay100
  otv join-interface e1/1
  otv control-group 239.1.1.1
  otv data-group 232.192.1.0/24
  otv extend-vlan 100-150
```

L2 Multicast Traffic between Sites

Unicast Only Transport

Release 5.2 and above

OTV can leverage the Adjacency Server mode of operation to allow for the delivery of L2 multicast traffic for the VLANs stretched across sites

Two steps:

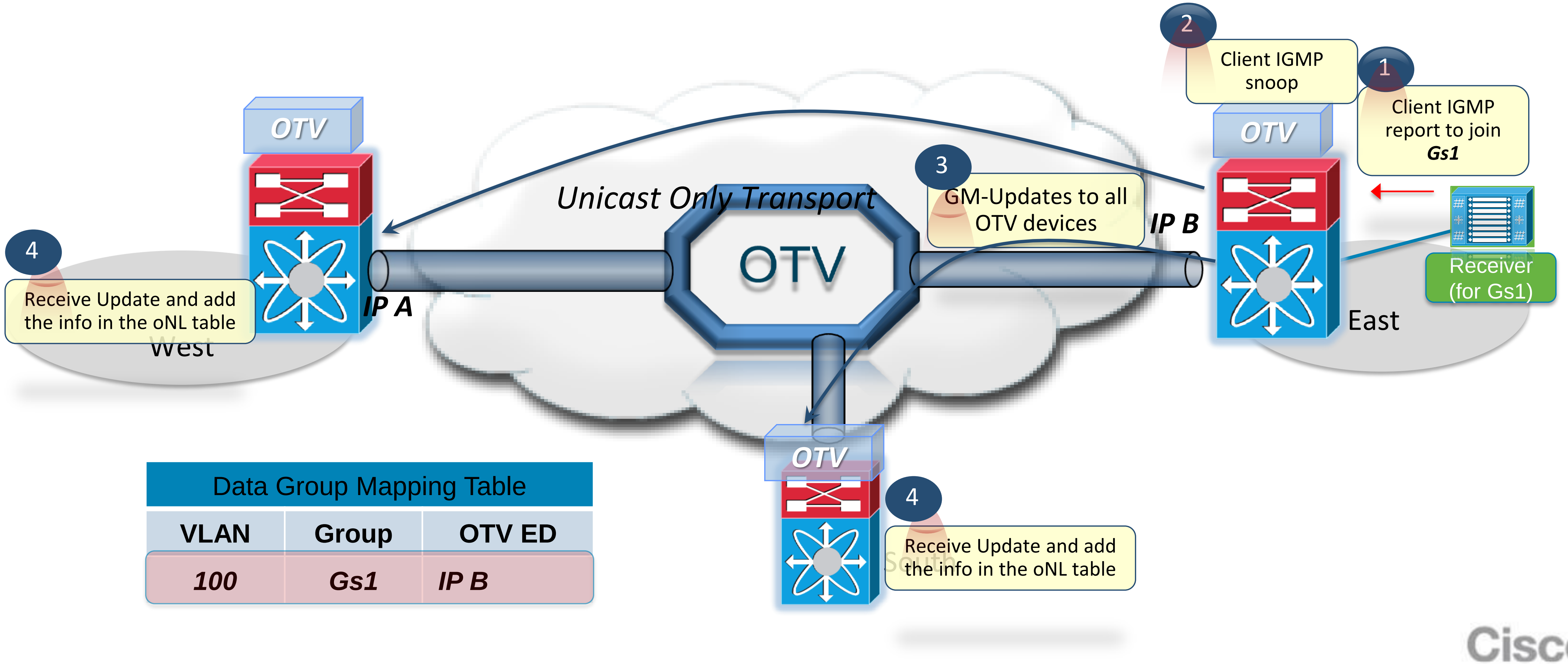
1. Communication of the existence of a receiver to all the OTV devices via Control Plane messages (GM-Updates)
2. The OTV Device connected to the source is then able to deliver the L2 multicast stream only to the sites where receivers interested to that specific group are located

Release 5.2 and above

L2 Multicast with Unicast Transport

Control Plane Updates for Interested Receivers

1. A receiver in the East site sends an IGMP join for Gs1
2. The OTV ED snoops the IGMP join (without forwarding it)
3. The East ED announces the receiver via a GM-Update sent to all OTV
4. Remote OTV devices store the information in the Data Group Mapping Table



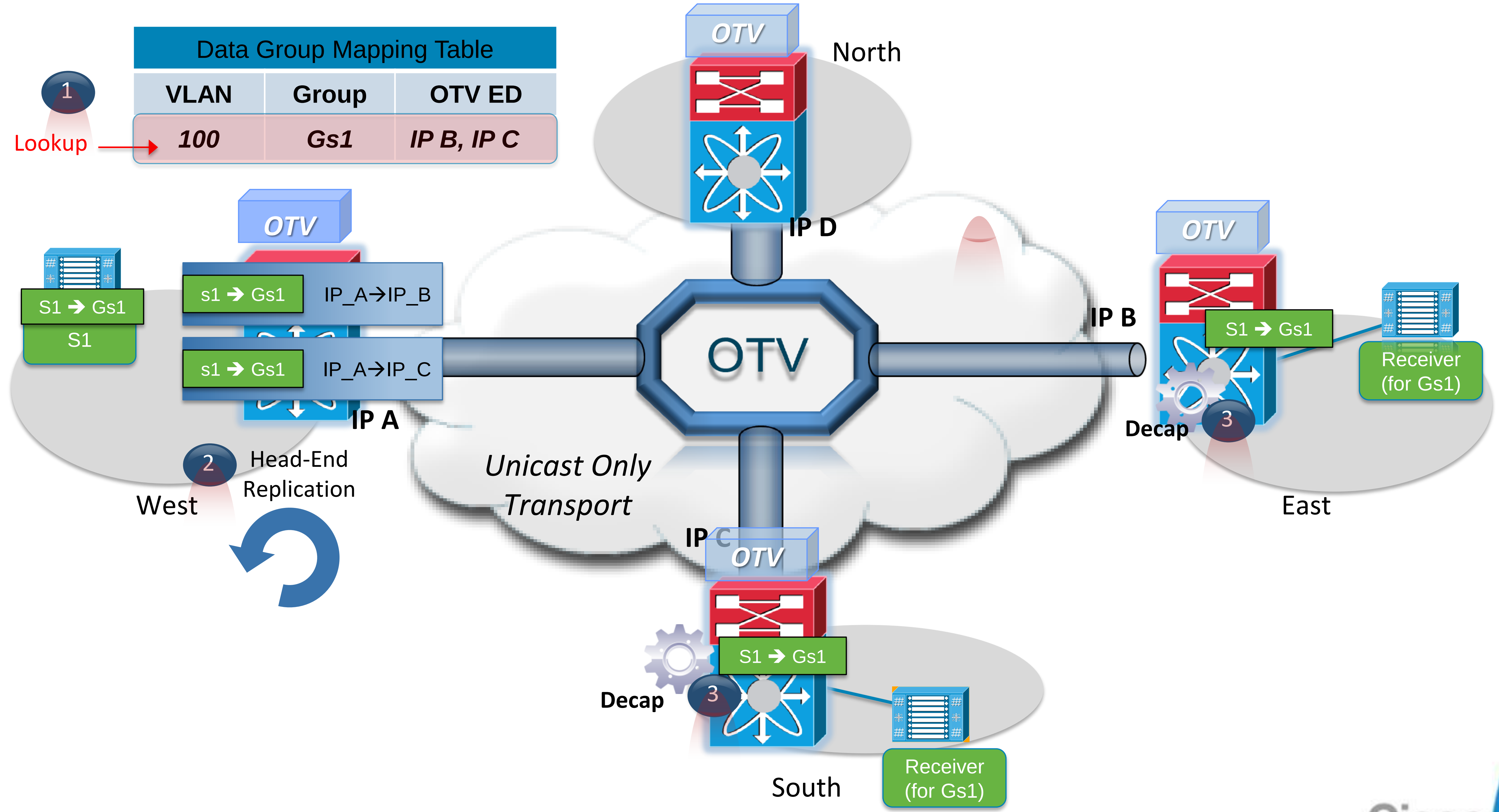
Data Group Mapping Table		
VLAN	**Group**	**OTV ED**
100	*Gs1*	*IP B*

L2 Multicast with Unicast Transport

L2 Multicast Packet Flow

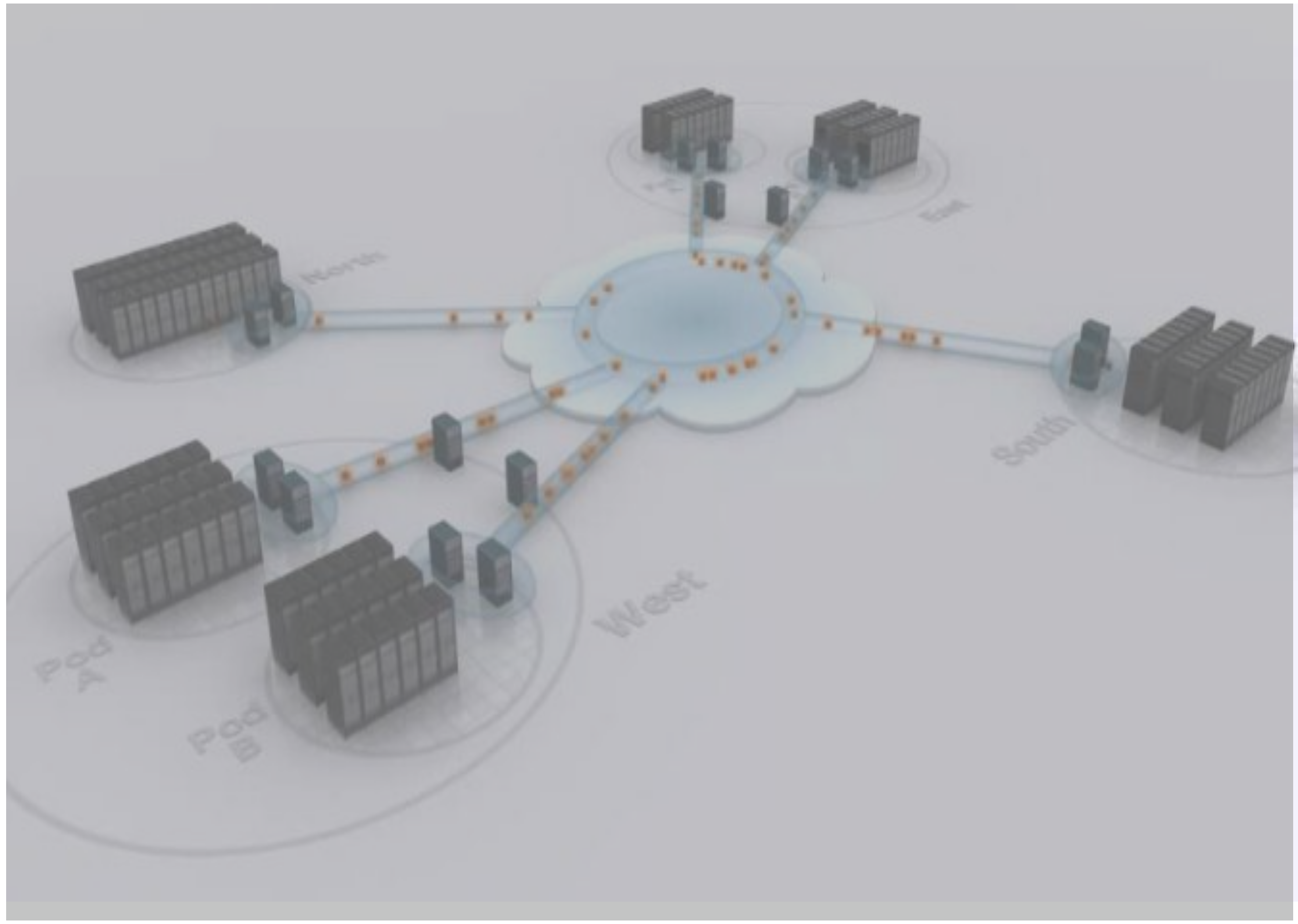
Release 5.2 and above

Data Group Mapping Table		
VLAN	Group	OTV ED
100	*Gs1*	*IP B, IP C*



Agenda

- Distributed Data Centres: Goals and Challenges
- OTV Architecture Principles
  - Control Plane and Data Plane
  - Failure Isolation
  - Multi-homing
  - L2 Multicast Forwarding
  - QoS and Scalability
  - Path Optimisation
- OTV at the Aggregation Layer

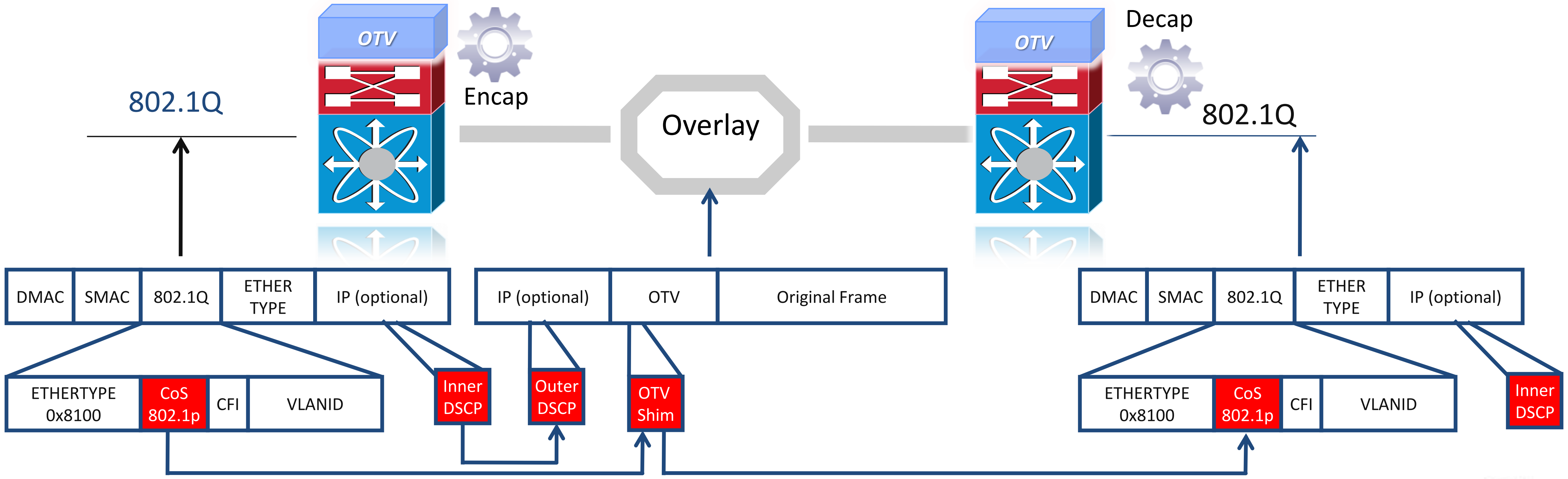


QoS and OTV

Traffic Marking

Release 5.2 and above

- On encapsulation:
  - CoS bits (802.1p) copied to the OTV shim header
  - If IP traffic: The original (inner) DSCP value is also copied to “outer” DSCP
- On de-capsulation:
  - CoS value is recovered from the OTV shim and added to the 802.1Q header
- Original CoS and DSCP are both preserved
- OTV Control Traffic is statically marked at CoS = 6/DSCP = 48



OTV Scalability

Current and Future Supported Values

The following values are supported in 5.2, 6.0, 6.1 releases:

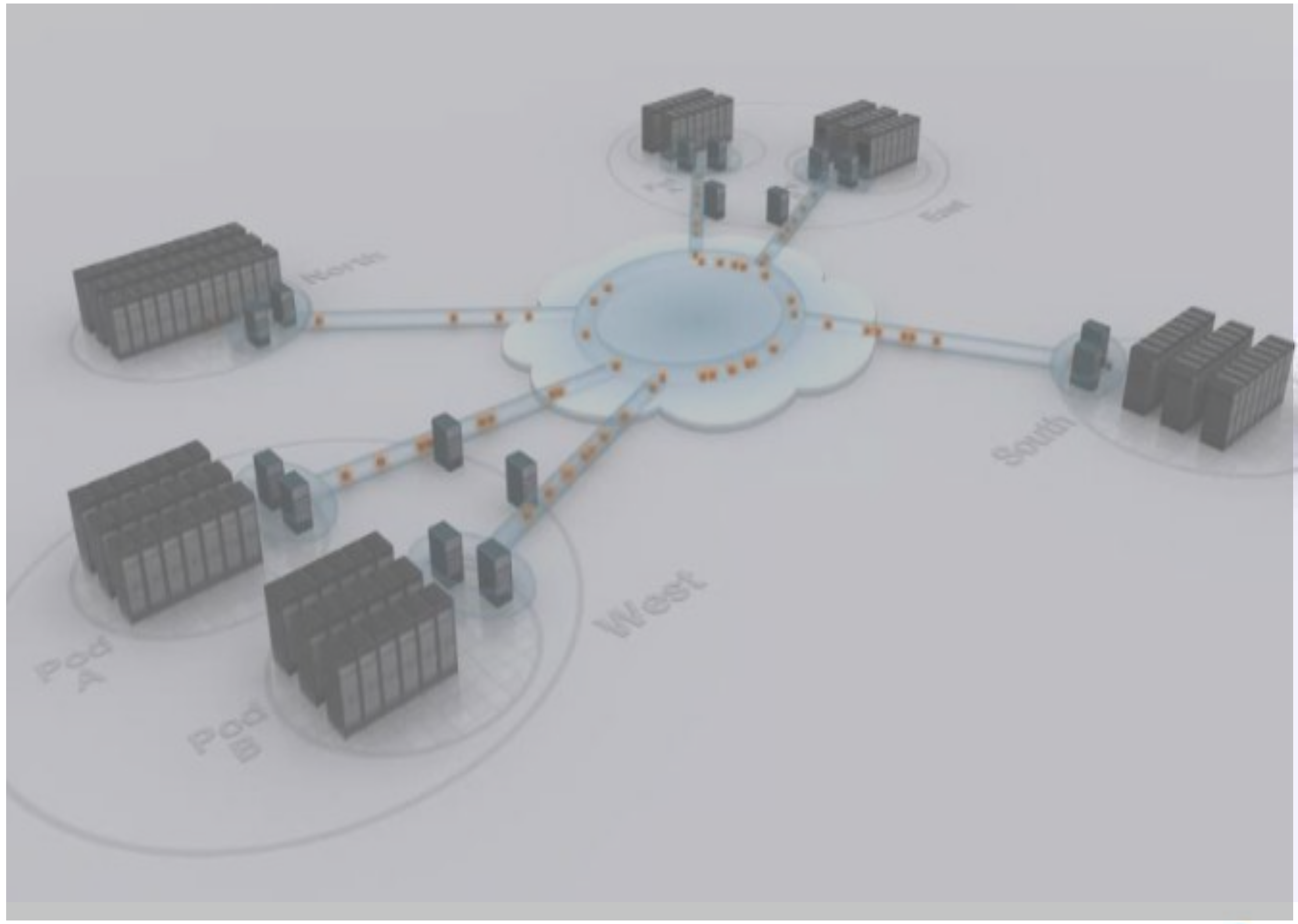
- 6 Sites
- 256 OTV extended VLANs*
- 16K MAC Addresses across all the extended VLANs
- 2000 Sites' Multicast Data Groups

The values below will be supported in 6.2 release (2Q13):

- 10 Sites
- 2000 OTV extended VLANs
- 32K MAC Addresses across all the extended VLANs
- 4000 Sites' Multicast Data Groups

Agenda

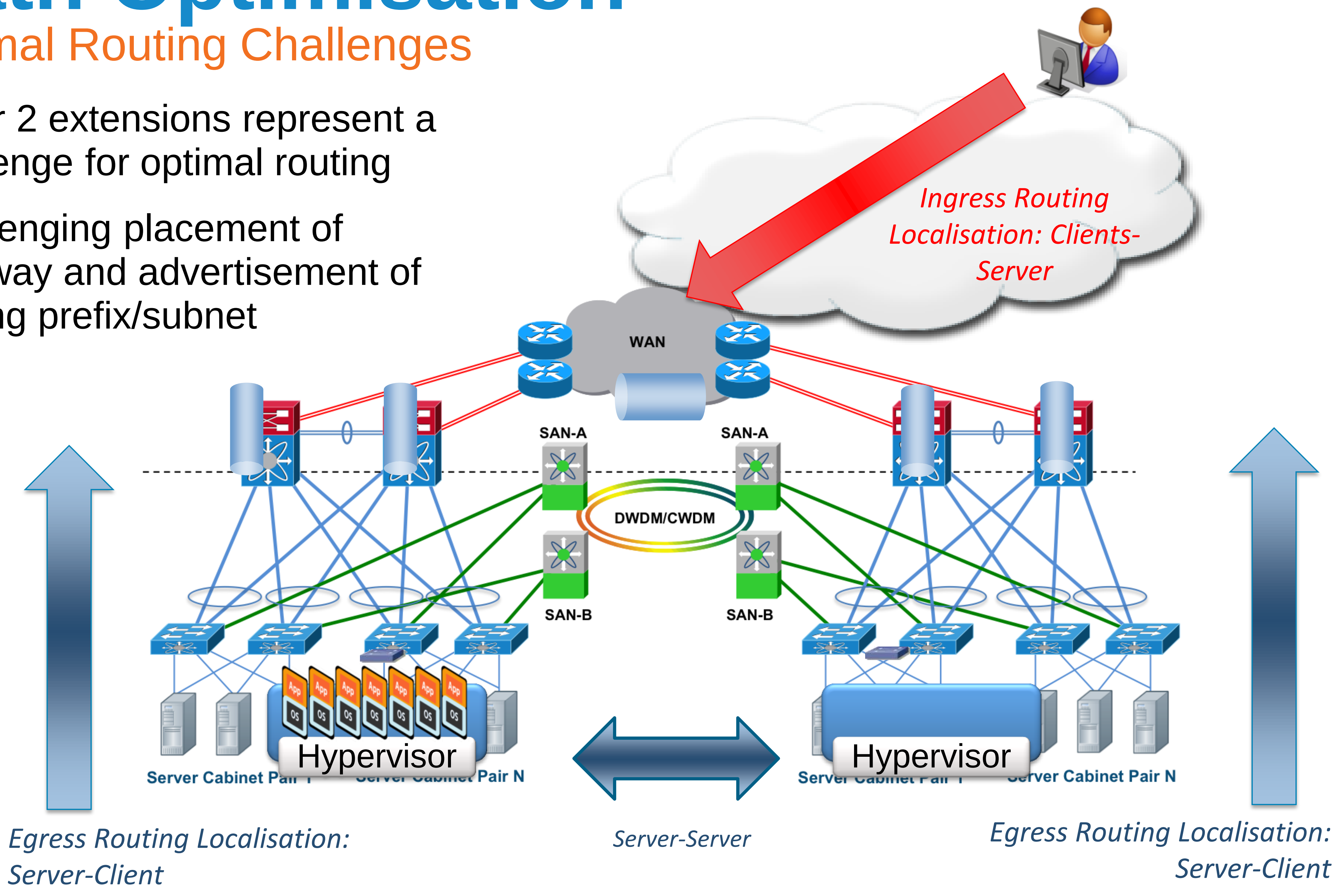
- Distributed Data Centres: Goals and Challenges
- OTV Architecture Principles
  - Control Plane and Data Plane
  - Failure Isolation
  - Multi-homing
  - L2 Multicast Forwarding
  - QoS and Scalability
  - Path Optimisation
- OTV at the Aggregation Layer



Path Optimisation

Optimal Routing Challenges

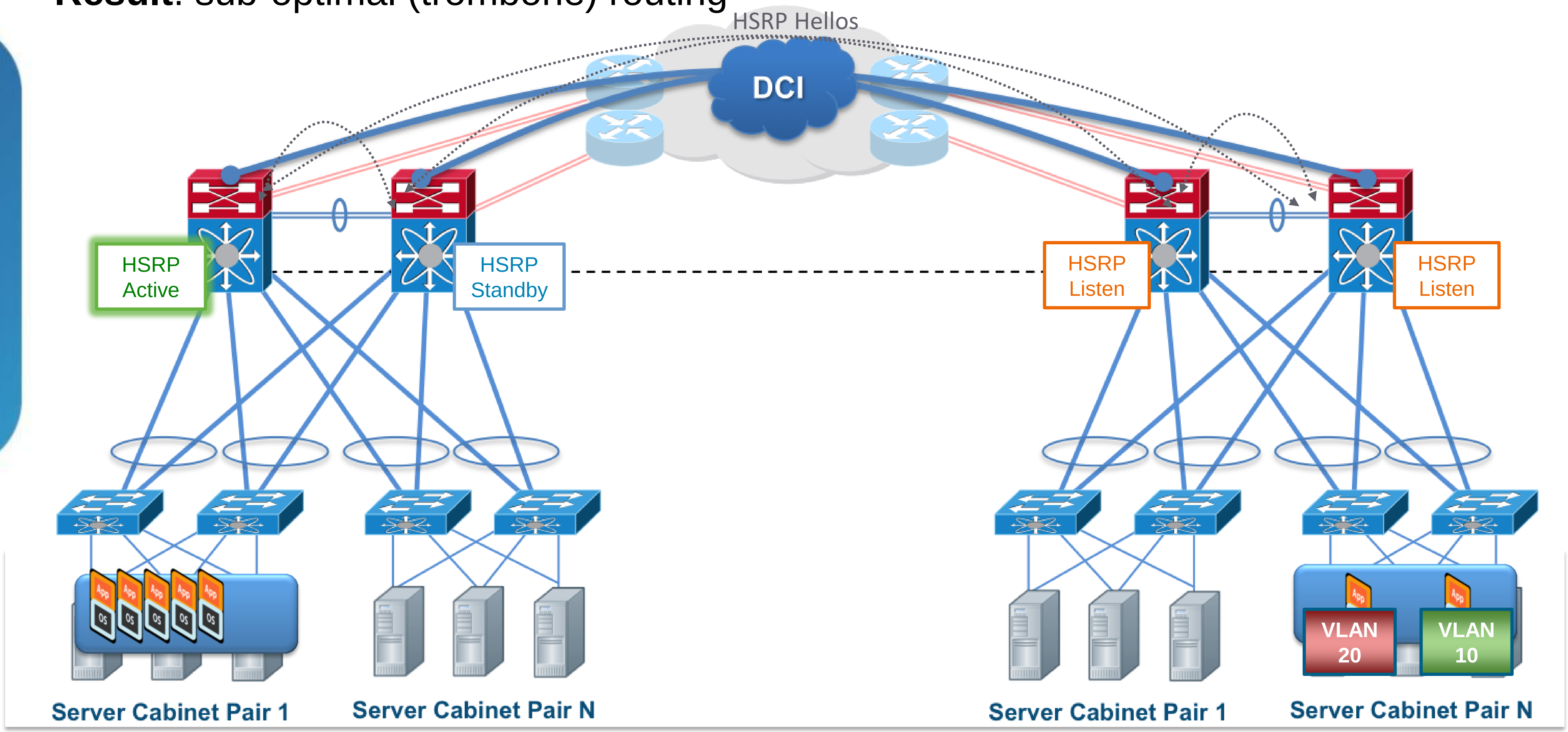
- Layer 2 extensions represent a challenge for optimal routing
- Challenging placement of gateway and advertisement of routing prefix/subnet



Path Optimisation

Egress Routing with LAN Extension

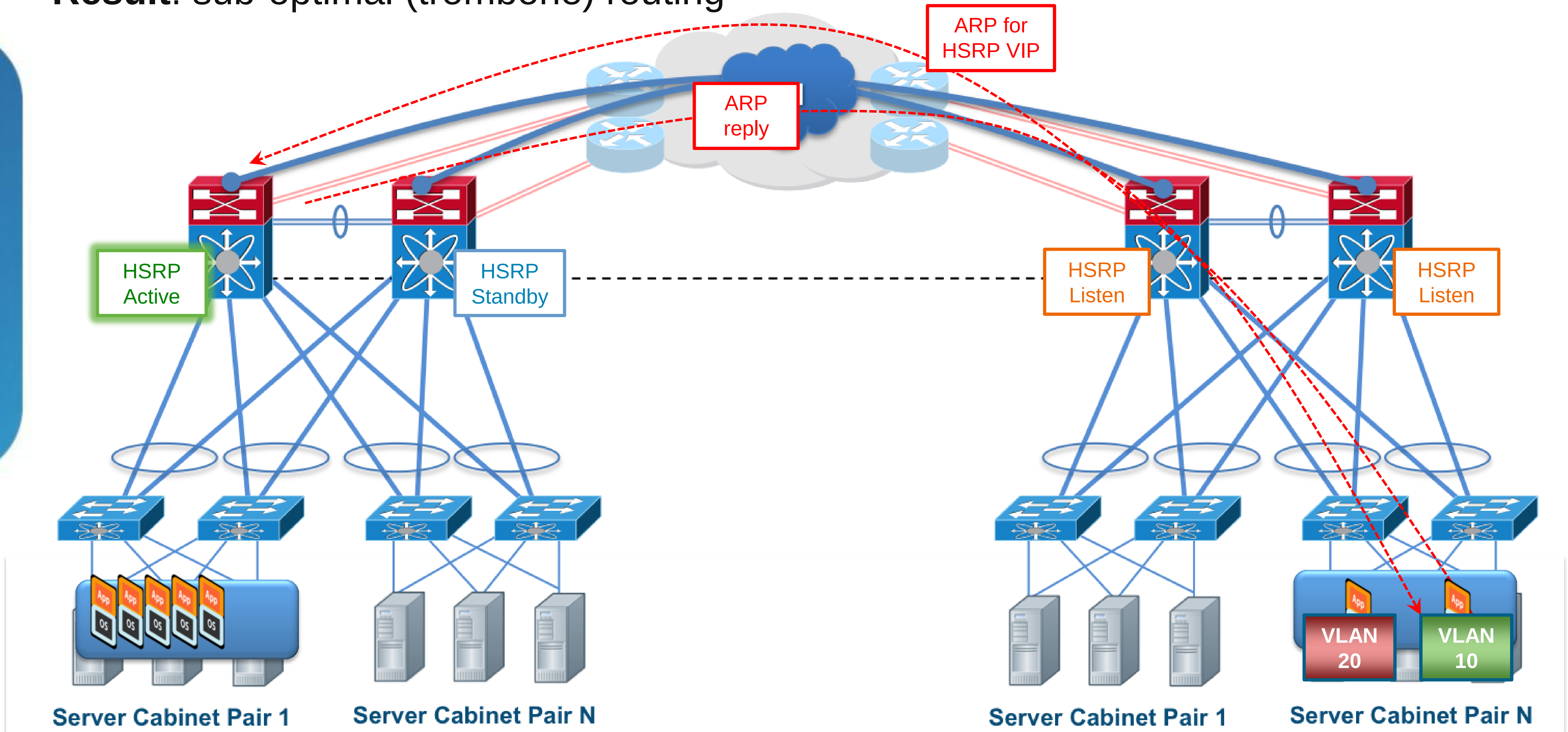
- Extended VLANs typically have associated HSRP groups
- By default, only one HSRP router elected active, with all servers pointing to HSRP VIP as default gateway
- **Result**: sub-optimal (trombone) routing



Path Optimisation

Egress Routing with LAN Extension

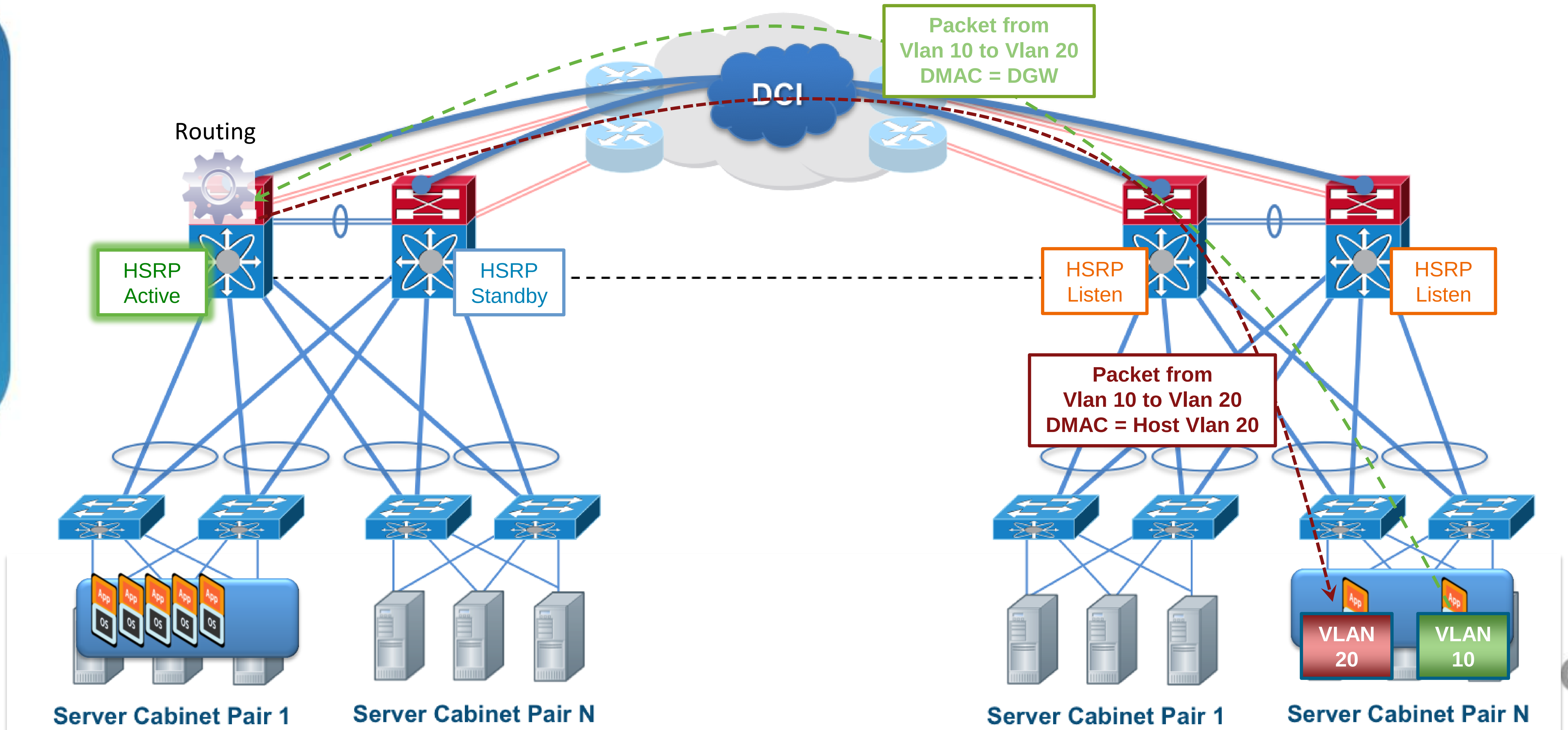
- Extended VLANs typically have associated HSRP groups
- By default, only one HSRP router elected active, with all servers pointing to HSRP VIP as default gateway
- **Result**: sub-optimal (trombone) routing



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

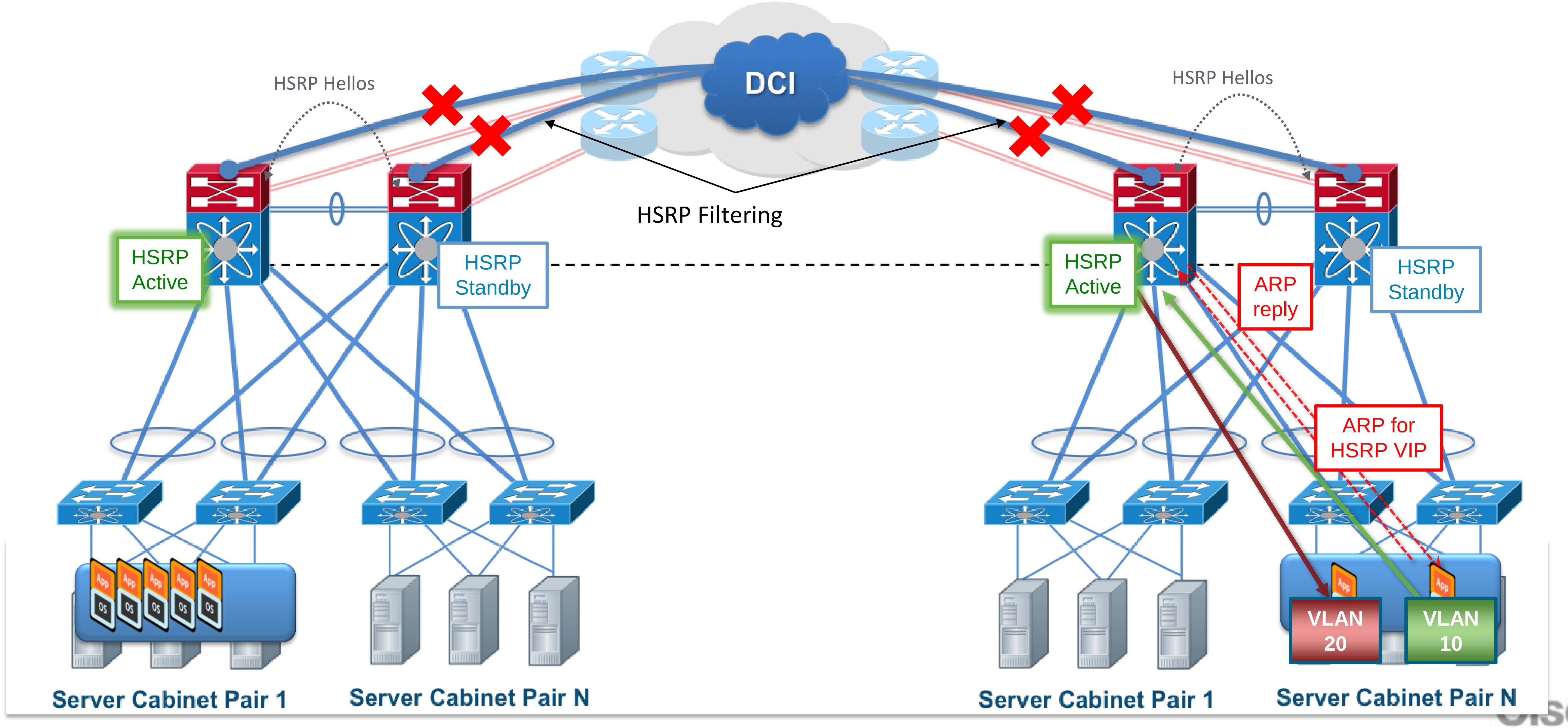


Cisco live!

Egress Routing Localisation

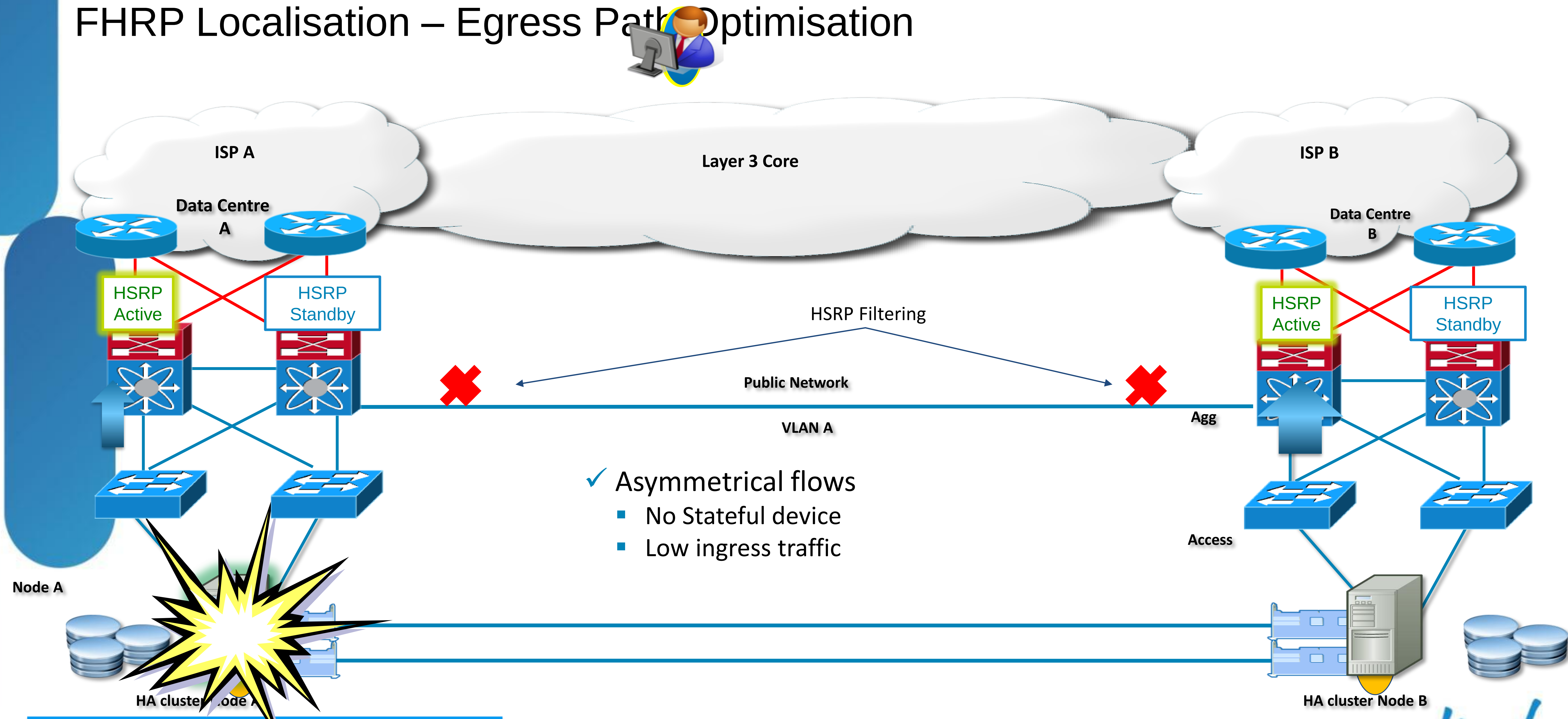
FHRP Filtering Solution

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



Sample Cluster - Primary Service in Left DC

FHRP Localisation – Egress Path Optimisation

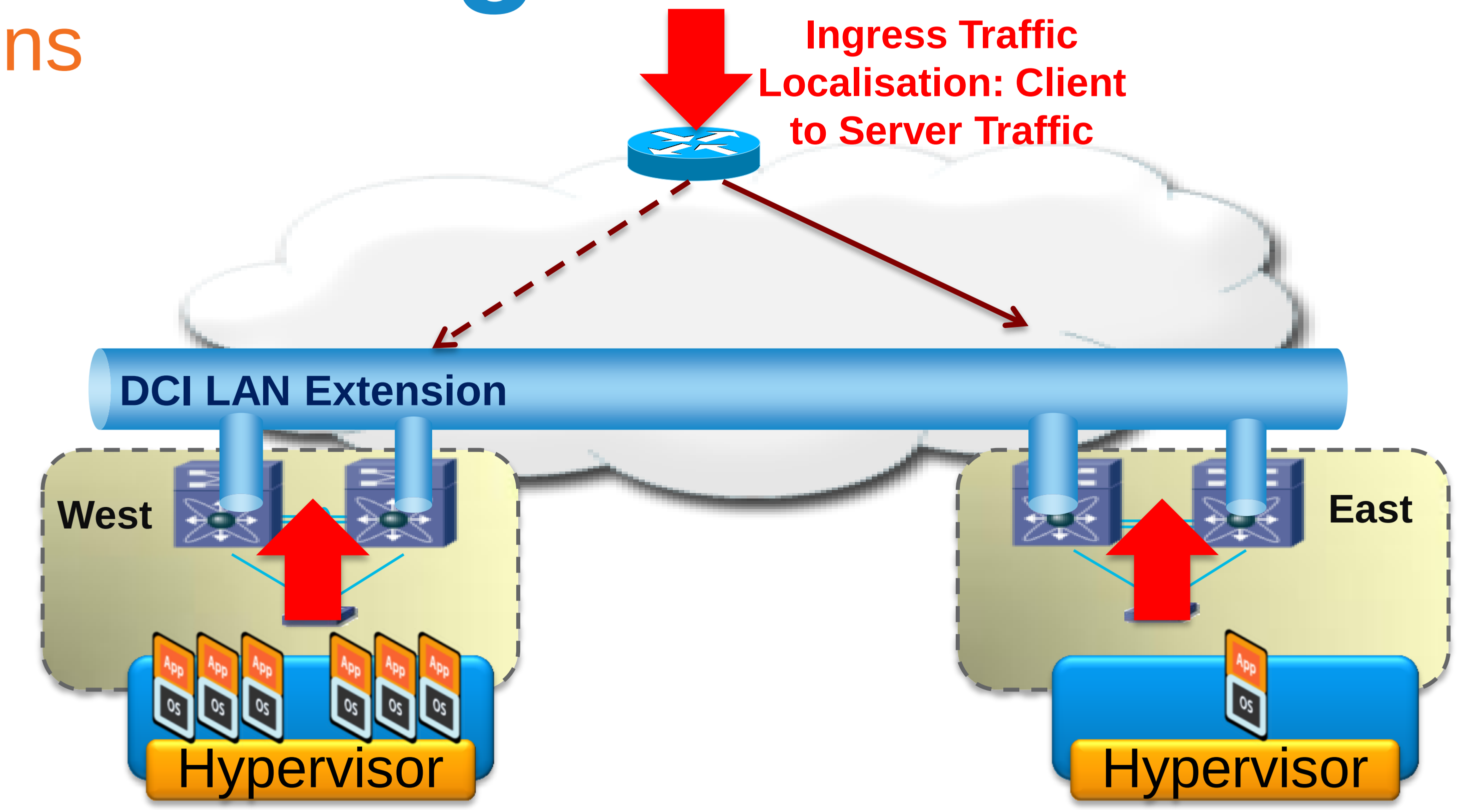


- ✓ Asymmetrical flows
  - No Stateful device
  - Low ingress traffic

Cluster VIP = 10.1.1.100 Preempt
Default GW = 10.1.1.1

Ingress Routing Localisation

Possible Solutions



Challenge

- Subnets are spread across locations
- Subnet information in the routing tables is not specific enough
- Routing doesn't know if a server has moved between locations
- Traffic may be sent to the location where the application is not available

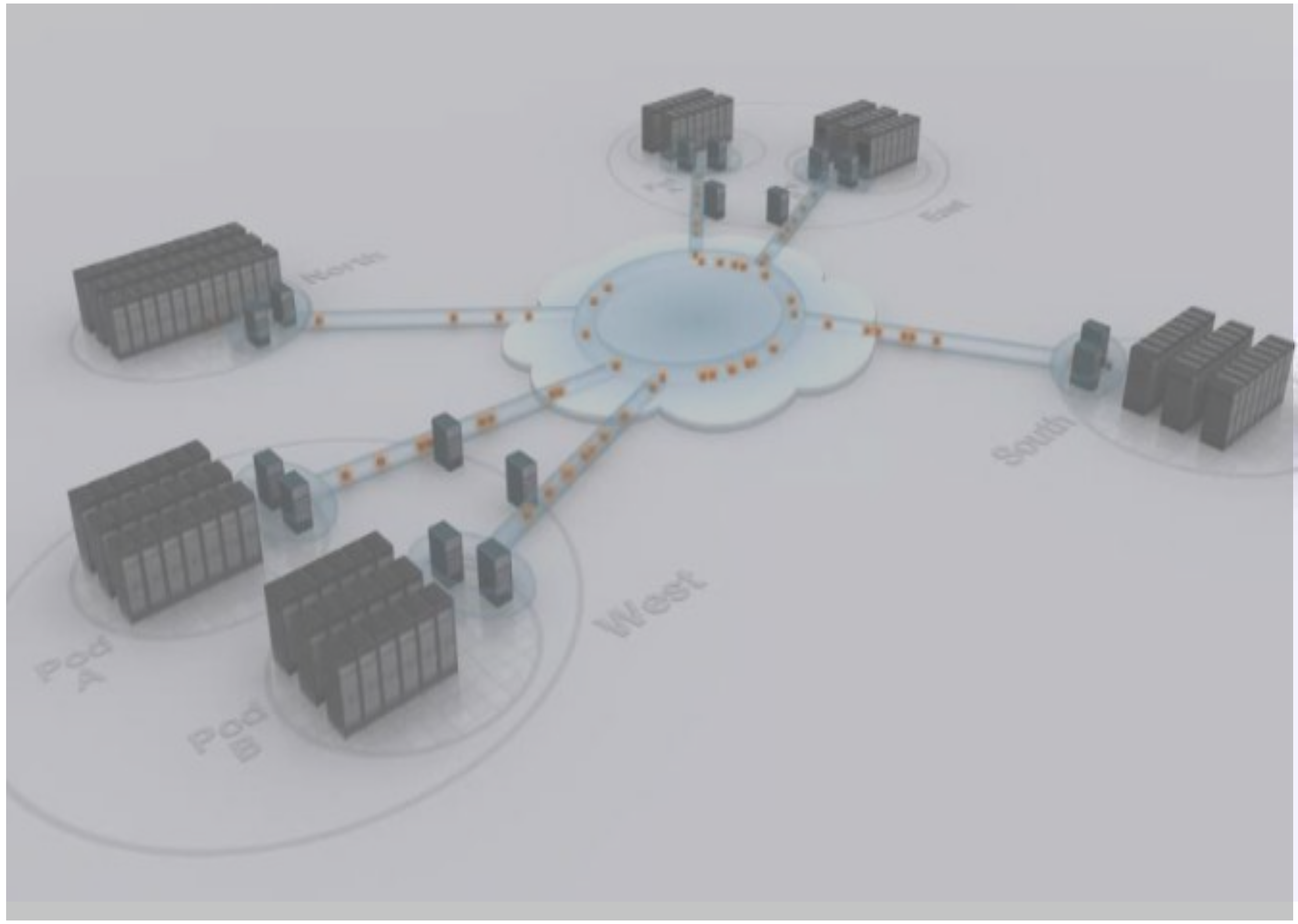
Options

- DNS Based
  1. DNS redirection with ACE/GSS
- Routing Based
  2. Route Injection
  3. LISP

For more details on LISP and OTV Deployment see: BRKDCT-3060

Agenda

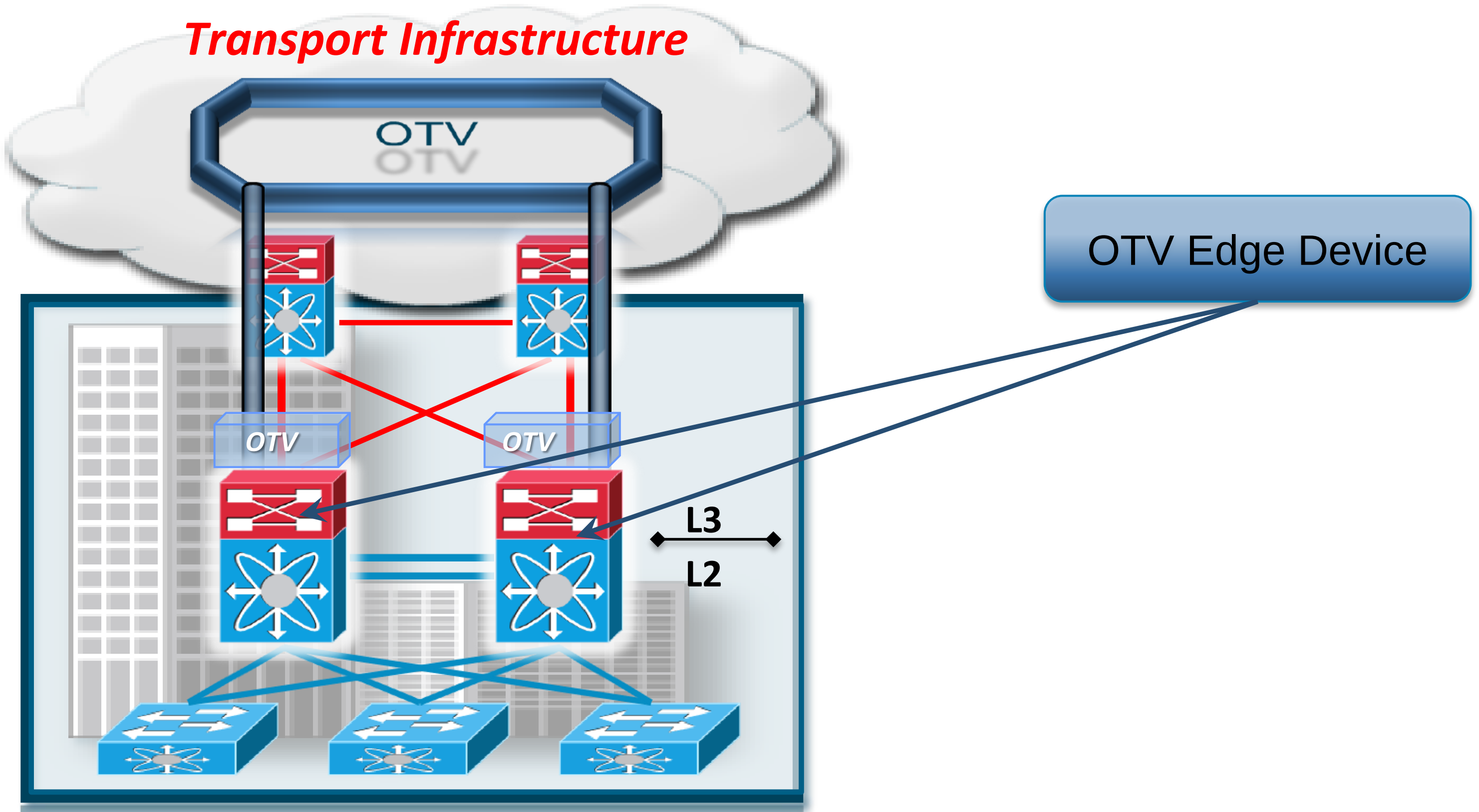
- Distributed Data Centres: Goals and Challenges
- OTV Architecture Principles
- OTV at the Aggregation Layer



OTV Deployment Options

OTV at the Aggregation Layer

- No universal response where to place the OTV Edge Device
- Main Options:
  - OTV at the Core Layer
  - OTV at the Aggregation Layer (most common – discussed in this presentation)

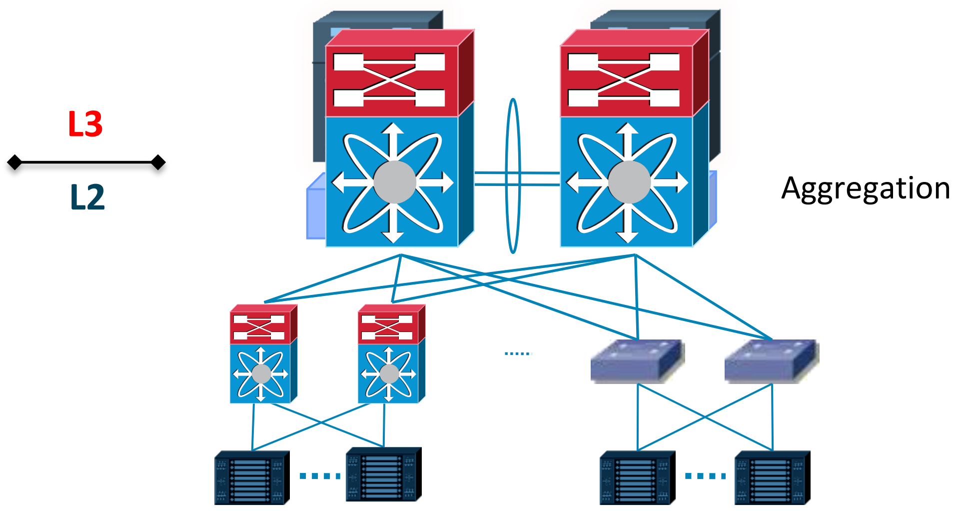


For more details on OTV Deployment Models see: BRKDCT-3060

OTV and SVI Routing

Functional Separation

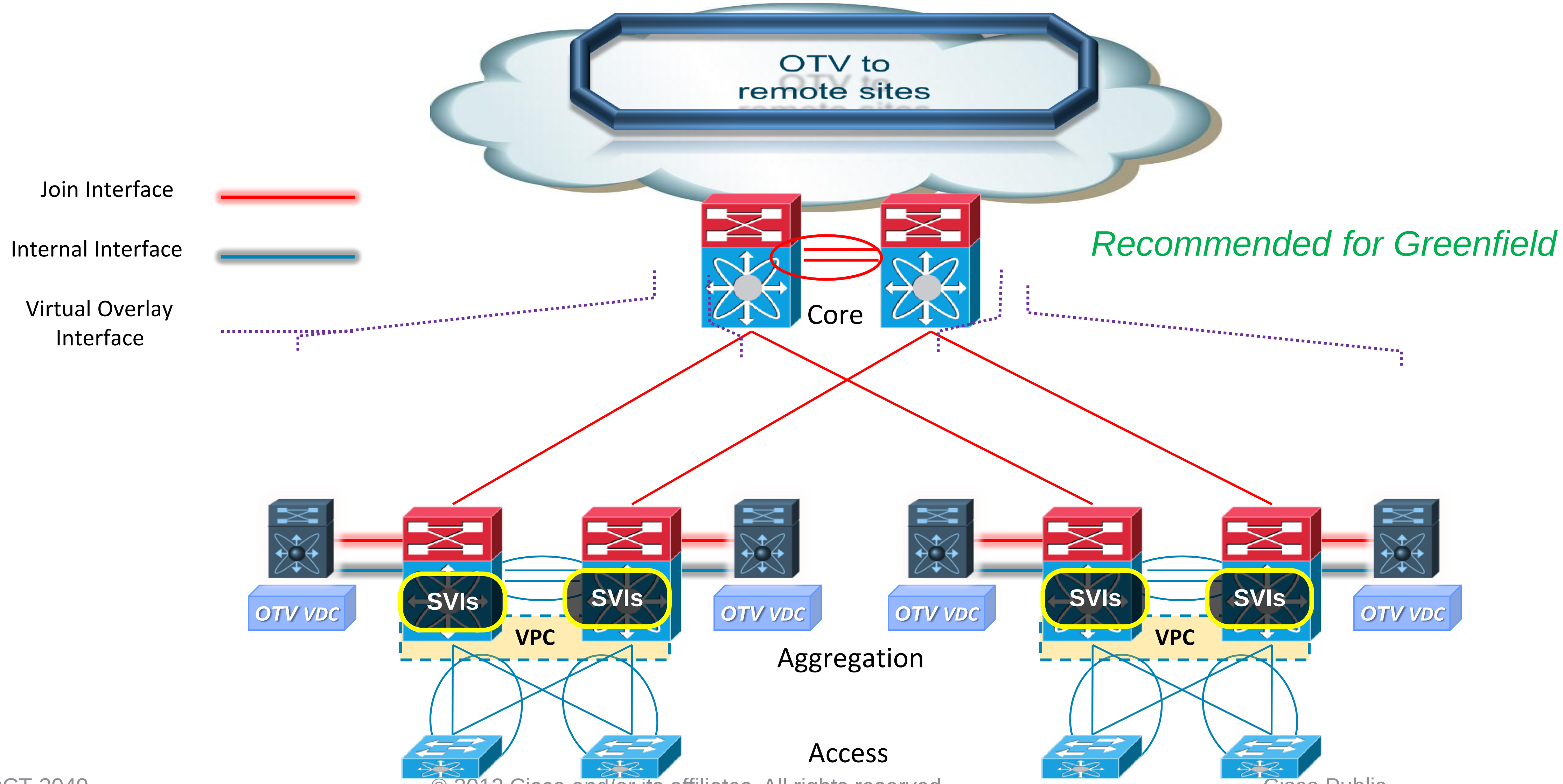
- **Guideline:** The current OTV implementation on the Nexus 7000 enforces the separation between SVI routing and OTV encapsulation for any extended VLAN
- This separation can be achieved with having two separate devices to perform these two functions
- An alternative **cleaner and less intrusive** solution is the use of **Virtual Device Contexts (VDCs)** available with Nexus 7000 platform:
    - A dedicated OTV VDC to perform the OTV functionalities
    - The Aggregation-VDC used to provide SVI routing support



OTV at the Aggregation Layer

OTV at the Aggregation with L3 Boundary

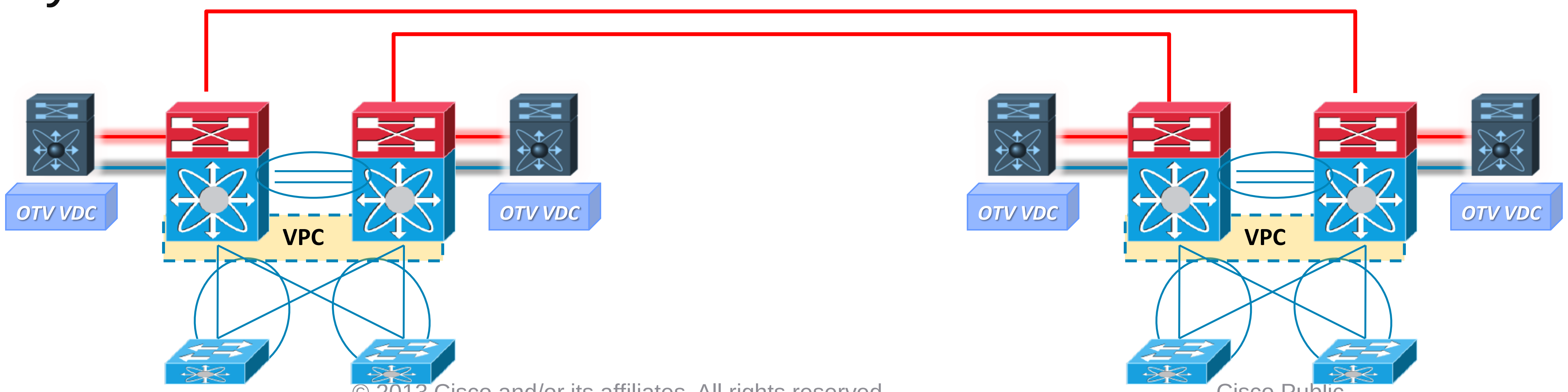
- DC Core performs only Layer 3 role
- STP and unknown unicast domains isolated between PODs
- Intra-DC and inter-DC LAN extension provided by OTV
- Ideal for single aggregation block topology



OTV at the Aggregation Layer

OTV in Point-to-Point Deployments

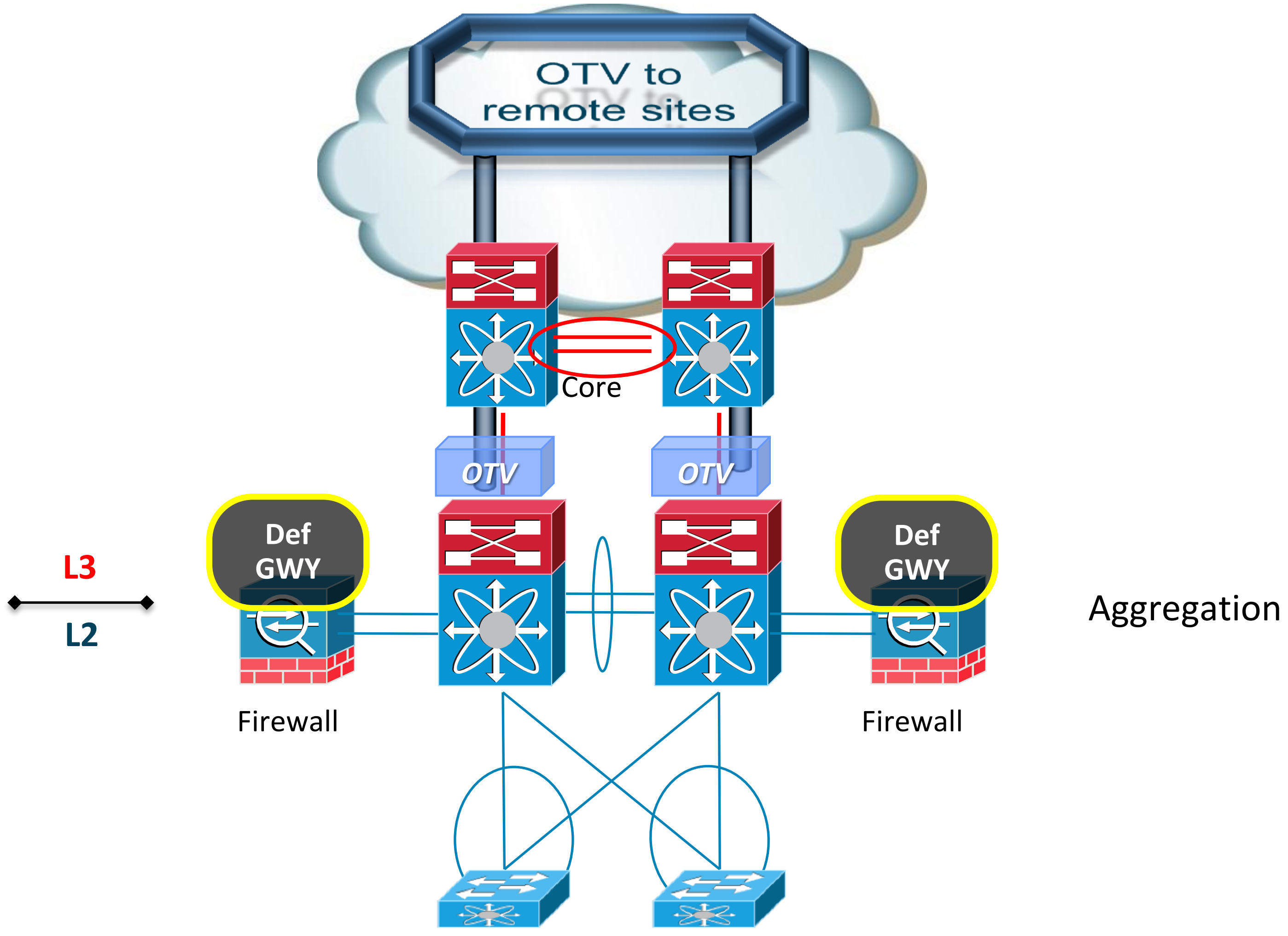
- Data Centres directly connected at the Aggregation
- OTV (virtual) Appliance on a Stick
- Advantages over VSS-vPC:
  - ✓ Provision of L2 and L3 connectivity leveraging the same dark fibre connections
  - ✓ Native STP isolation: no need to explicitly configure BPDU filtering
  - ✓ ARP Optimisation with the OTV ARP Cache
  - ✓ Simplified provisioning of FHRP isolation (roadmap)
  - ✓ Easy addition of sites



OTV at the Aggregation

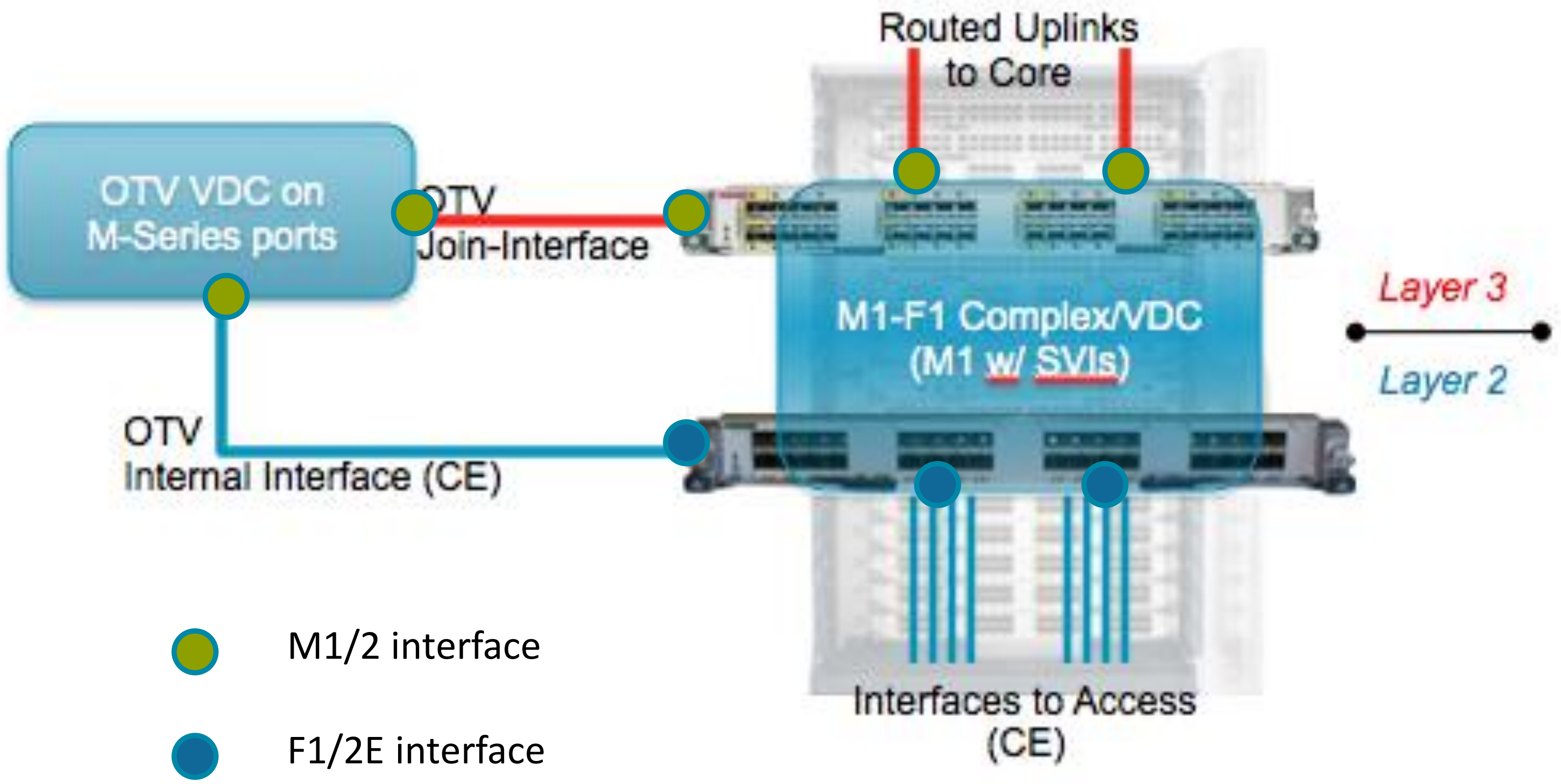
OTV at the Aggregation with L3 boundary on the Firewalls

- The Firewalls host the Default Gateway
- No SVIs at the Aggregation Layer
- No Need for the OTV VDC



OTV in the DC Aggregation

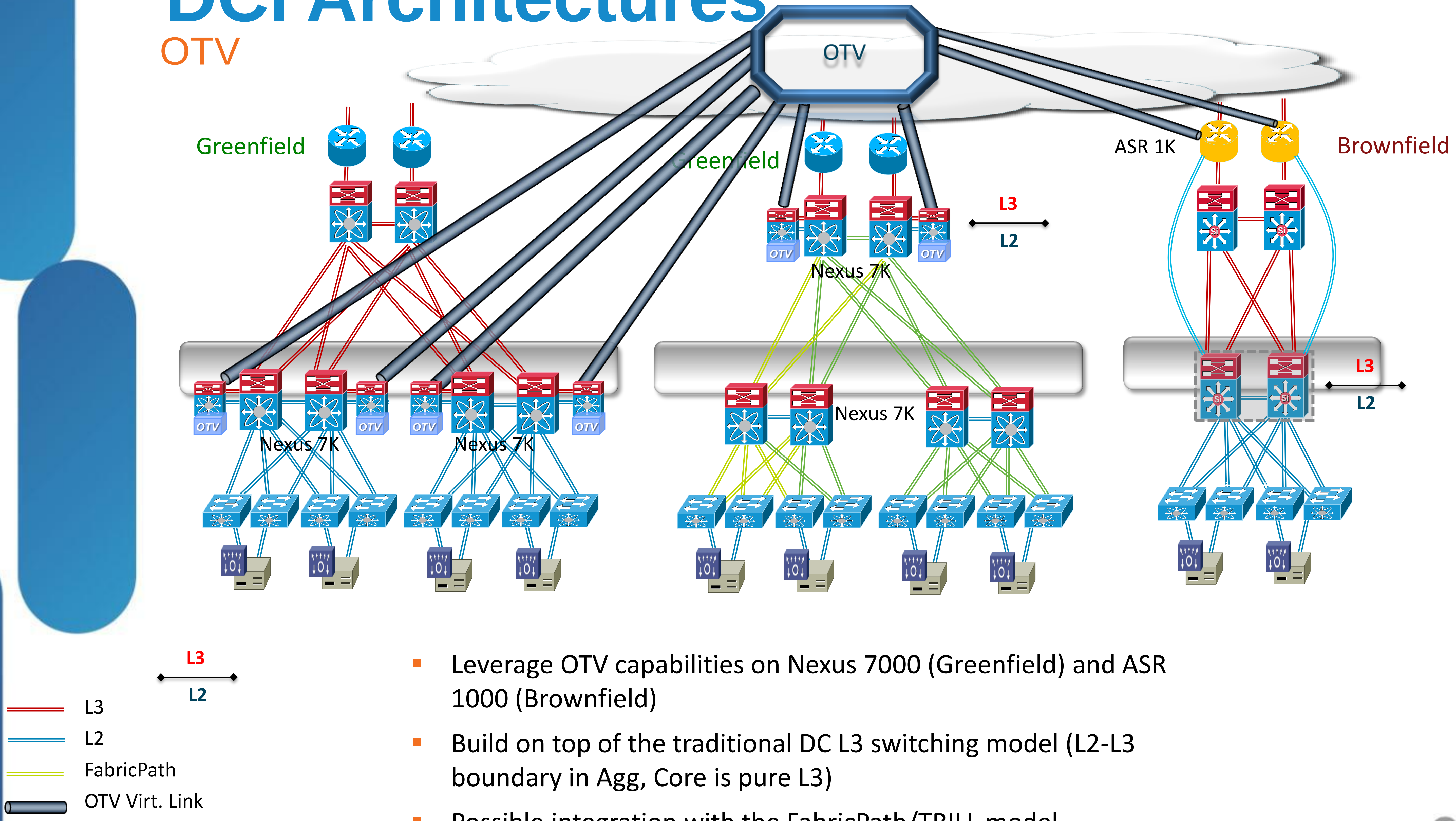
Nexus 7000 HW Support



- F1 and F2 linecards do not support OTV natively
- As of today, the OTV VDC must use only M-series ports for both internal and join interfaces
  – Recommendation is to allocate M only interfaces to the OTV VDC
  – All M series modules supported (M1-48, M1-32, M1-08, M2 series)
- Native OTV support on F-series is targeted for a future HW release

DCI Architectures

OTV



- Leverage OTV capabilities on Nexus 7000 (Greenfield) and ASR 1000 (Brownfield)
- Build on top of the traditional DC L3 switching model (L2-L3 boundary in Agg, Core is pure L3)
- Possible integration with the FabricPath/TRILL model

OTV Support on ASR1000

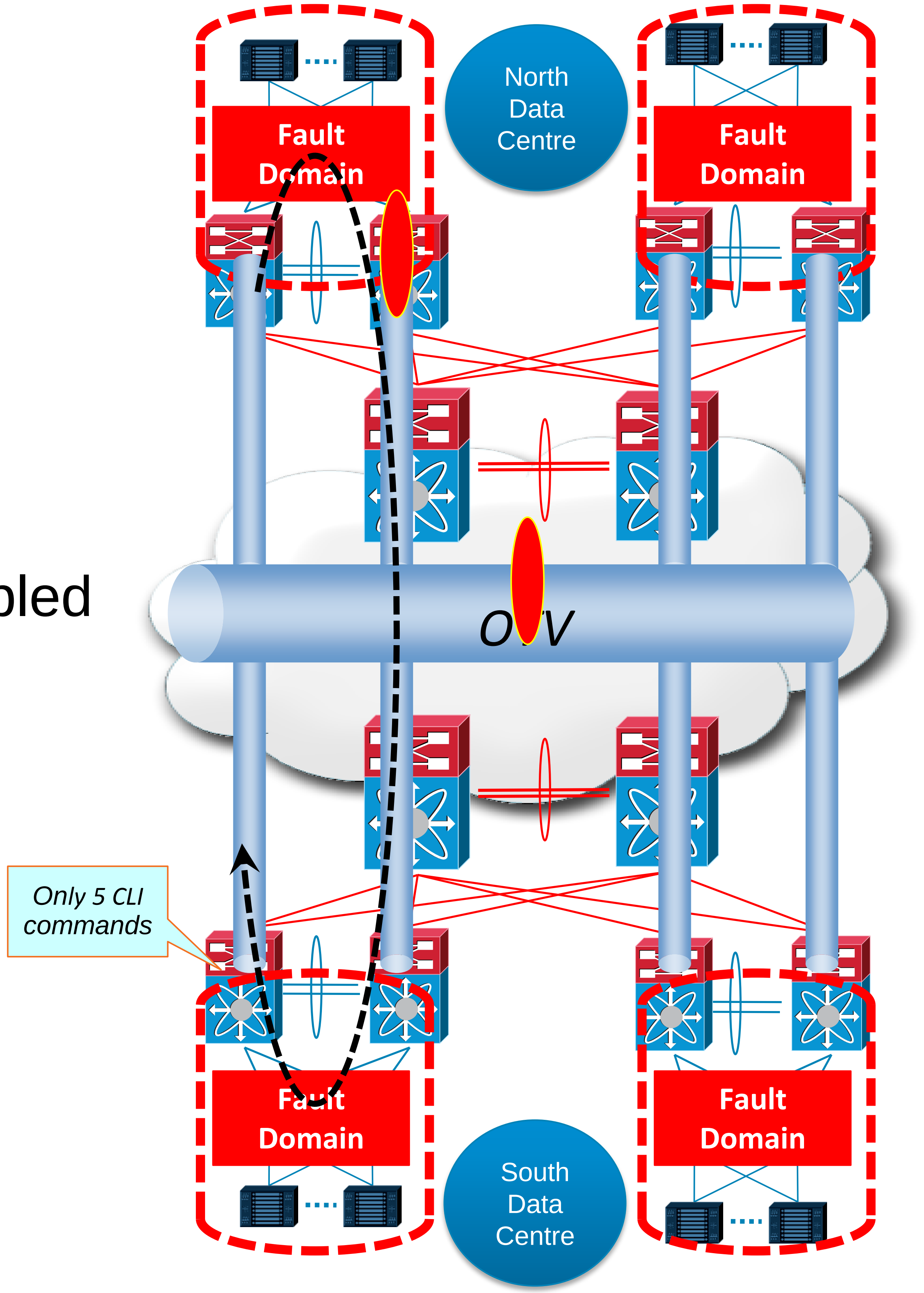
- OTV phase 1 introduced in IOS XE 3.5 (Nov 2011)
  - Multicast enabled core required for OTV control and data planes
- OTV phase 2 introduced in IOS XE 3.8 (Feb 2013)
  - Can use unicast or multicast enabled core
- IOS XE 3.6 and XE 3.7 provided OTV scale and convergence improvements
- To use OTV on ASR1000, you require:
  - Advance Enterprise Image or Advance IP Service is (AES or AIS) to have the CLI enabled
  - A purchase of OTV feature license. The OTV feature License is an honour based license (not enforced) right to use license (RTU)

OTV
Summary

- Extensions over any transport (IP, MPLS)
- Failure boundary preservation
- Site independence
- Optimal BW utilisation with multicast enabled transport infrastructure
(no head-end replication)
- Automated Built-in Multi-homing
- End-to-End loop prevention
- Scalability

    Sites, VLANs, MACs

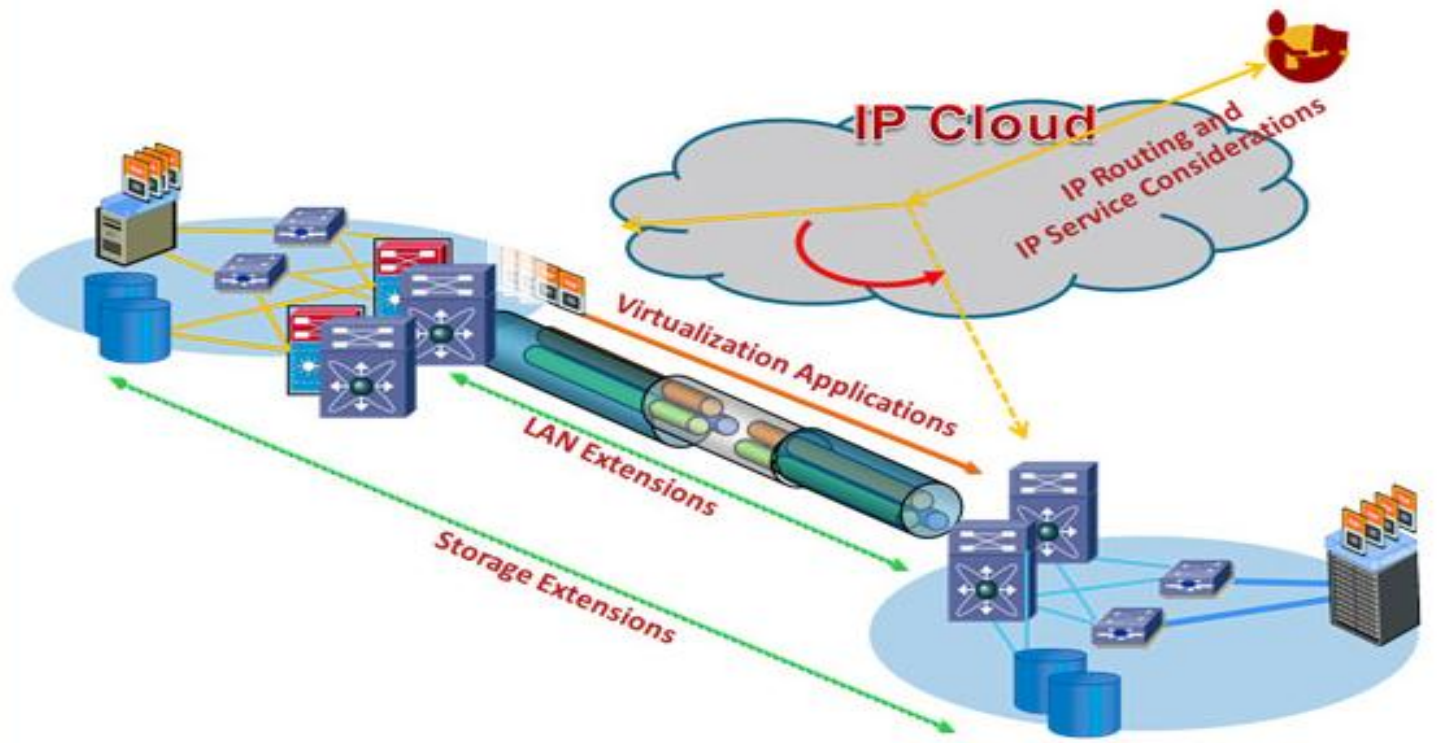
- Operational simplicity



Data Centre Interconnect

Where to Go for More Information

Enabling Long Distance Workload Mobility Among Data Centers



What's New | Overview | Architecture and System Validation | Selling Resources

In January 2010, Cisco announced Overlay Transport Virtualization (OTV) on Cisco Nexus 7000 series switches. The top 3 benefit are:

- Non-disruptive overlay;
- Isolate failure domains;
- Optimal BW utilization

In June 2010, Advanced VPLS (A-VPLS) will be available as software download from Catalyst 6500 on Cisco.com. A-VPLS represents evolution in Cisco's VPLS portfolio with Cisco Catalyst 6500 series switches. The top 3 benefit are:

http://www.cisco.com/go/dci
http://www.cisco.com/en/US/netsol/ns749/networking_solutions_sub_program_home.html

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

Cisco live! 365

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