



Cisco ACI Multi-Pod Design and Deployment

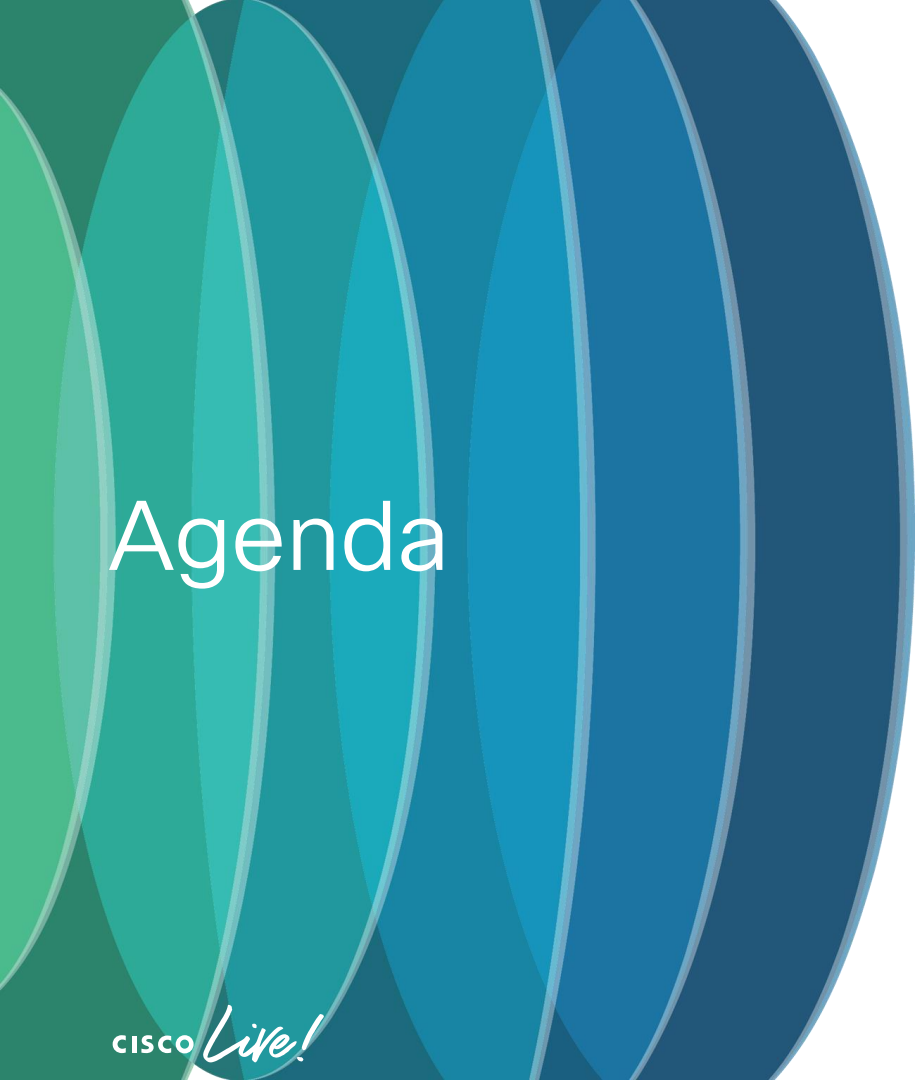
John Weston - Technical Marketing Engineer
Data Center Networking
BRKDCN-2949



Session Objectives



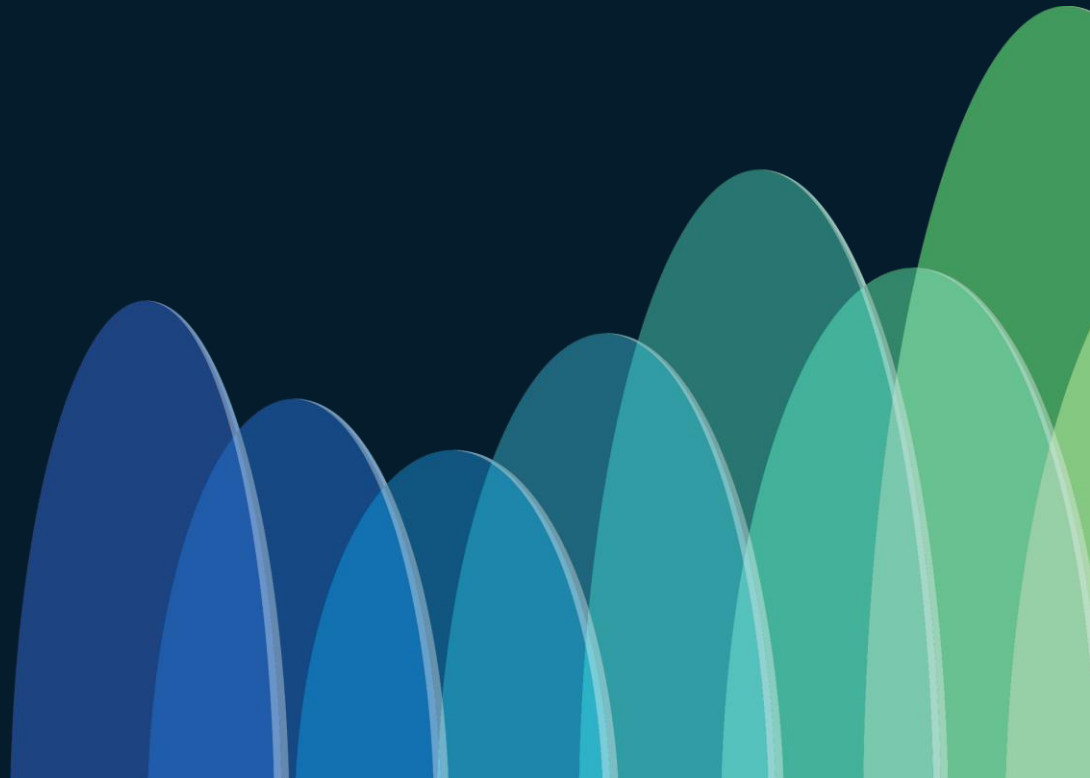
- At the end of the session, the participants should be able to:
 - Articulate the different deployment options to interconnect Cisco ACI networks (Multi-Pod and Multi-Site) and when to choose one vs. the other
 - Understand the functionalities and specific design considerations associated to the ACI Multi-Pod architecture
- Initial assumption:
 - The audience already has a good knowledge of ACI main concepts (Tenant, BD, EPG, L3Out, etc.)



Agenda

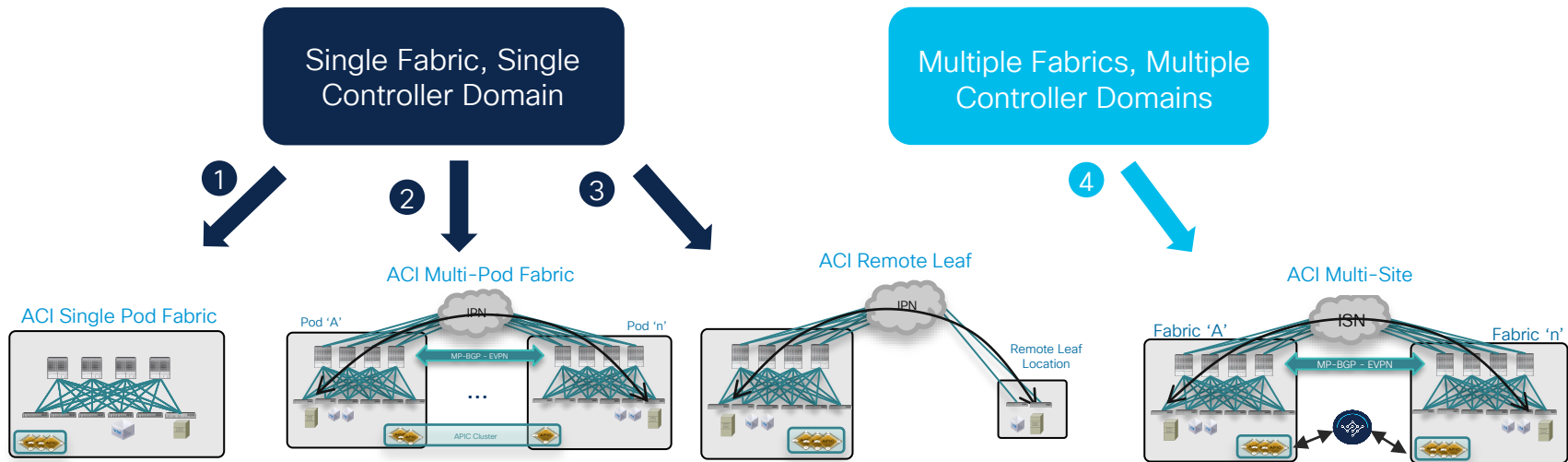
- Introduction
- Supported Topologies
- APIC Cluster Deployment
- Inter-Pod Connectivity
- Control and Data Planes
- Connecting to External Networks
- Network Services Integration
- Multi-Pod and Remote Leaf

Introduction



ACI Architectural Options

Fabric and Policy Domain Evolution



Multi-Pod or Multi-Site?

Where to Get More Information

- ACI Multi-Pod White Paper

<https://www.cisco.com/c/en/us/solutions/collateral/data-center-virtualization/application-centric-infrastructure/white-paper-c11-737855.html>

- ACI Multi-Site White Paper

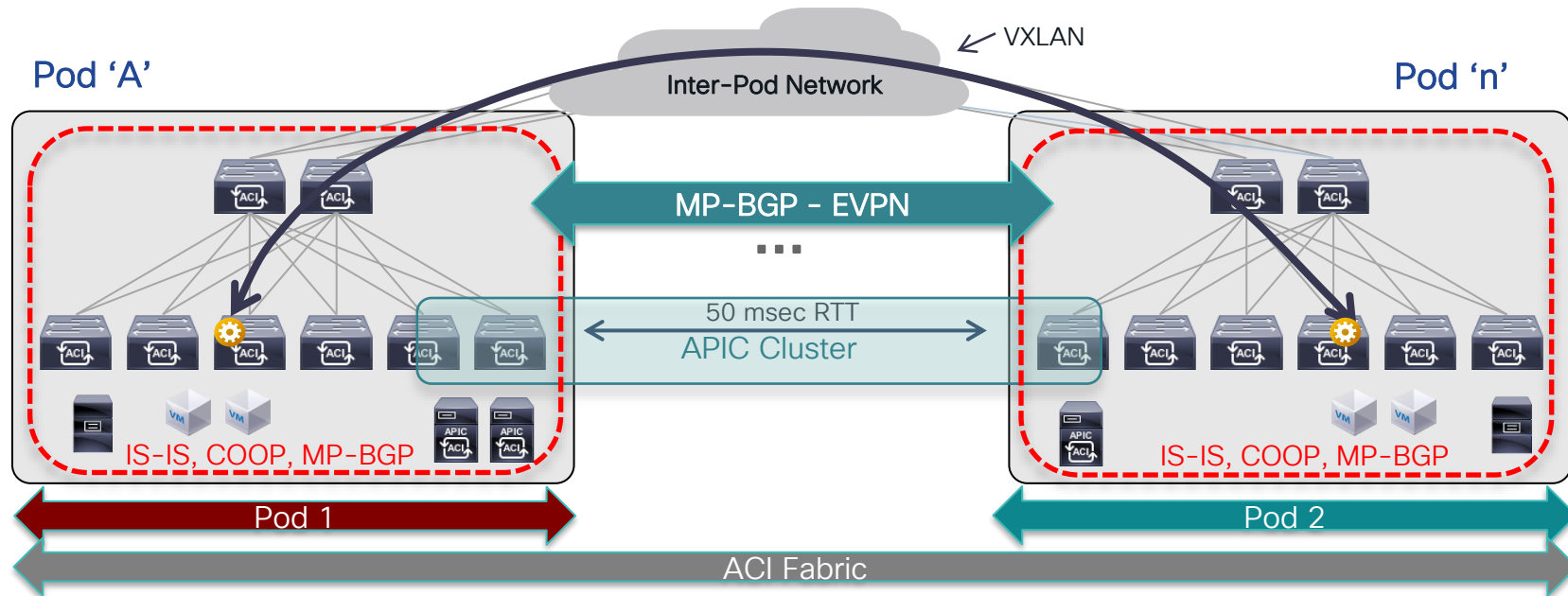
<https://www.cisco.com/c/en/us/solutions/collateral/data-center-virtualization/application-centric-infrastructure/white-paper-c11-739609.html>

- ACI Multi-Site Cisco Live 2020 Digital Breakout Session

<https://www.ciscolive.com/on-demand/on-demand-library.html?search=ardica&search=ardica#/video/1636411349156002rlx8>

ACI Multi-Pod

The Ideal Architecture for Active/Active DC Deployments

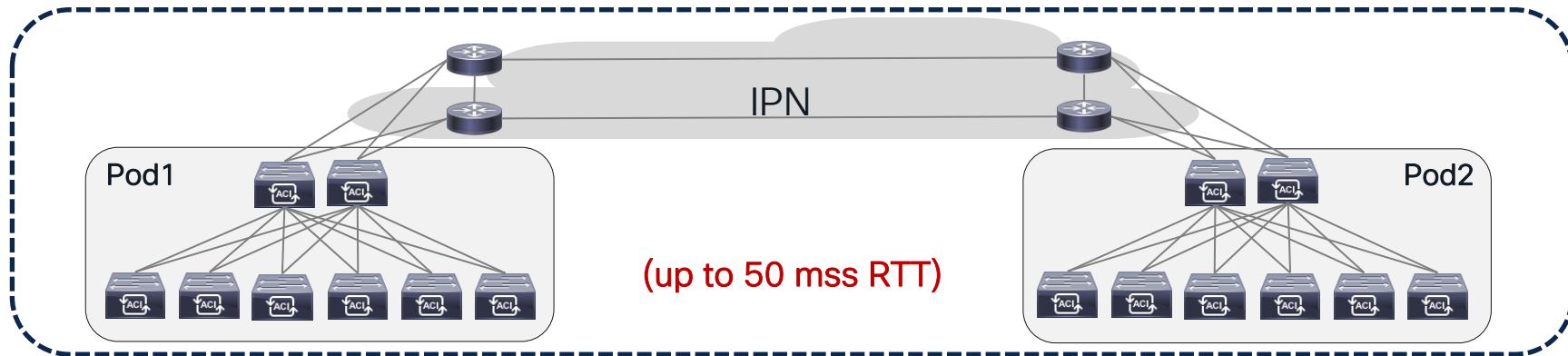


- Multiple ACI Pods connected by an IP Inter-Pod L3 network, each Pod consists of leaf and spine nodes
- Managed by a single APIC Cluster
- Single Management and Policy Domain
- Forwarding control plane (IS-IS, COOP) fault isolation
- Data Plane VXLAN encapsulation between Pods
- End-to-end policy enforcement

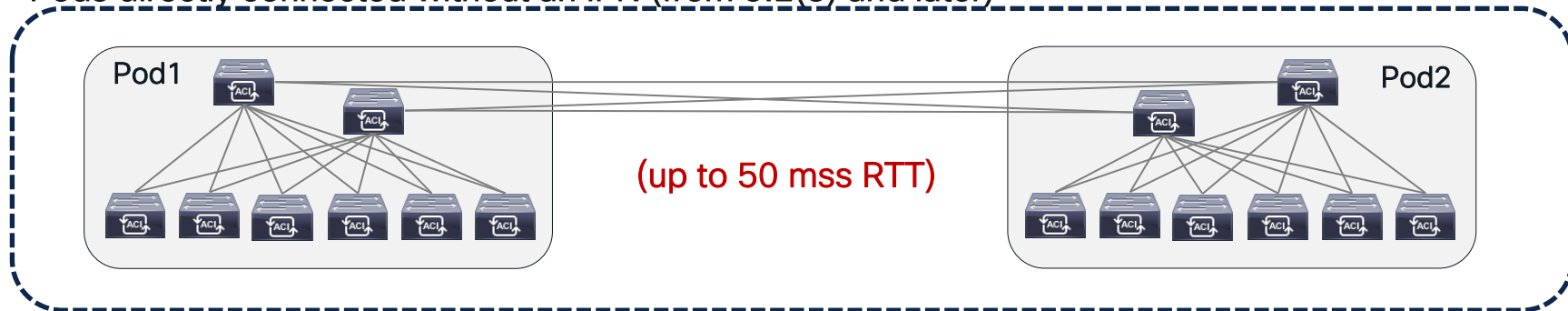
Supported Topologies

Multi-Pod Supported Topologies

Pods connected via an Inter-Pod Network (IPN)

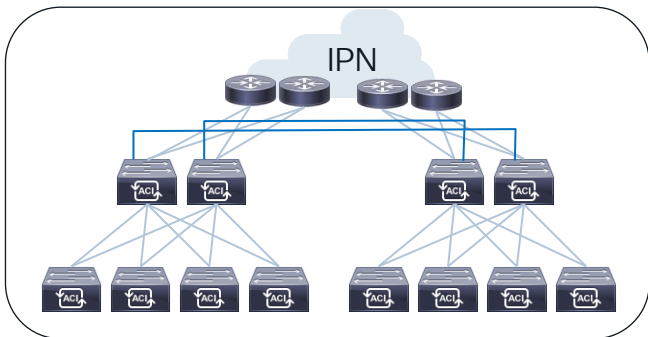
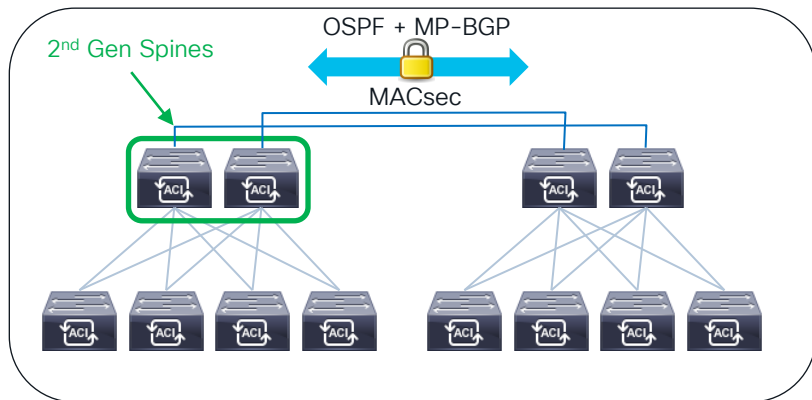


Pods directly connected without an IPN (from 5.2(3) and later)



Multi-Pod Spines Back-to-Back

Guidelines and Restrictions



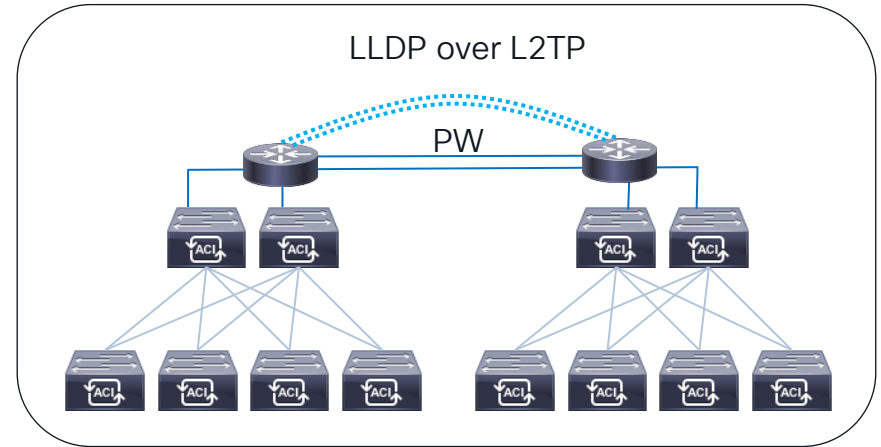
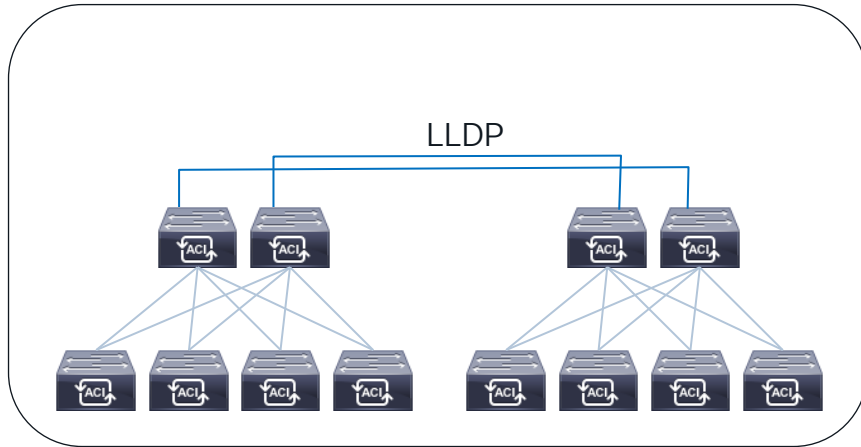
Back-to-Back + IPN (Migration Only)

- Support is limited to a topology with 2 Pods
- Supported on 2nd generation spines or later
- OSPF underlay peering, MP-BGP overlay peering between the spines in separate Pods
 - No need for PIM-Bidir (spines do not run PIM)
- MACsec encryption supported across Pods
- Not compatible functions (any feature requiring an IPN connection)
 - ACI Multi-Site
 - Remote Leaf
 - GOLF
 - APIC connectivity via L3 network
- Back-to-Back + IPN only supported for migration purposes (migration is disruptive)

Multi-Pod Spines Back-to-Back

Supported Topologies

- Back-to-back spine connectivity must be point-to-point (physical or logical)
- Spines discover back-to-back connections via LLDP
- Links can be directly connected or must support tunneling of LLDP packets



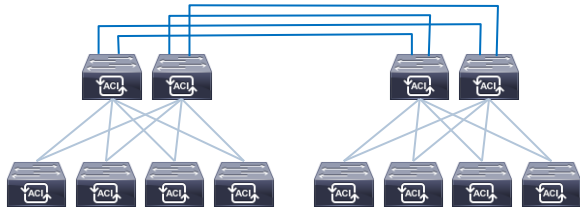
Multi-Pod Spines Back-to-Back

Supported Topologies

It is not mandatory for all spines in a Pod to connect to all the spines in the other Pod, the design decision must be made based on resiliency/bandwidth considerations

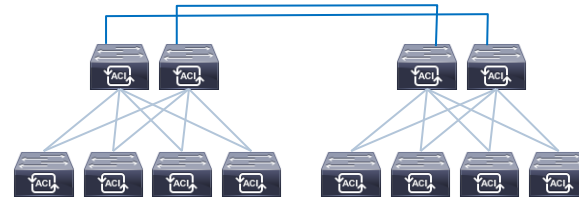
Recommended

Full mesh between spines

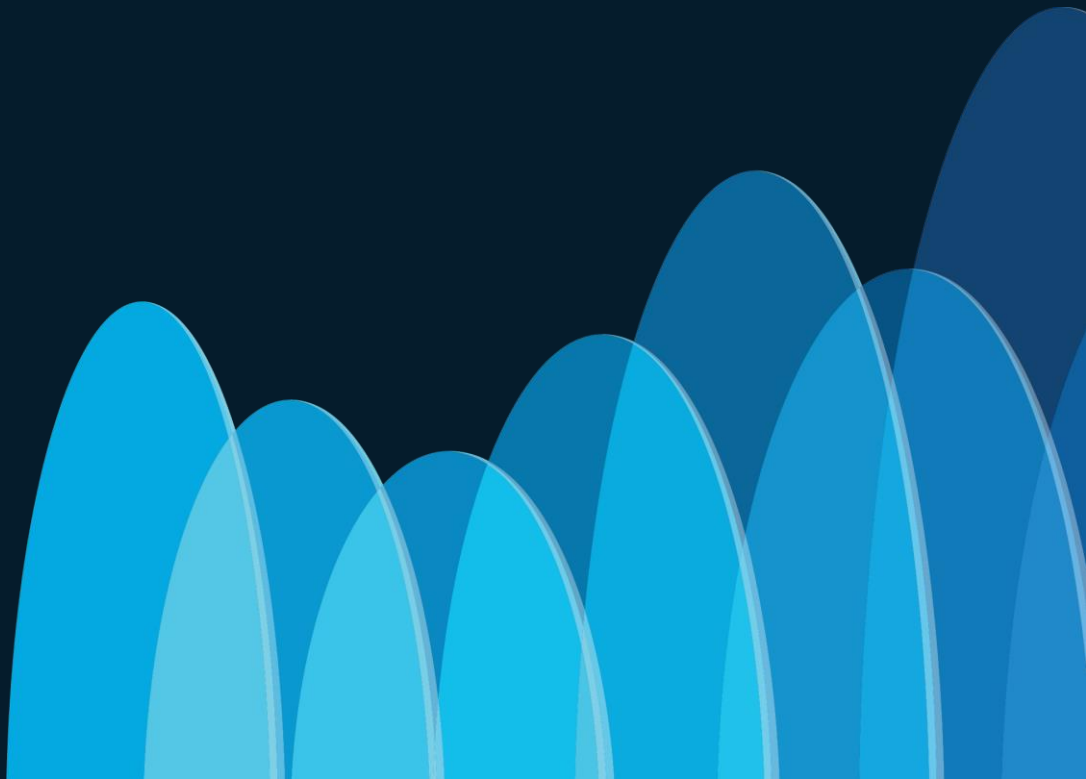


Supported

Partial mesh between spines



APIC Cluster Deployment Considerations



APIC – Distributed Multi-Active Data Base



Each shard is replicated across three APICs

Sharding:

The data of the ACI fabric is divided into smaller, separate parts called shards and distributed across active APIC nodes.

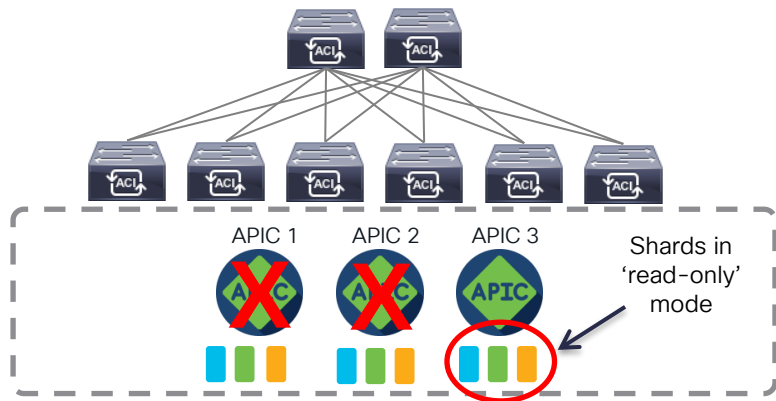
 ← Data shards

Replica:

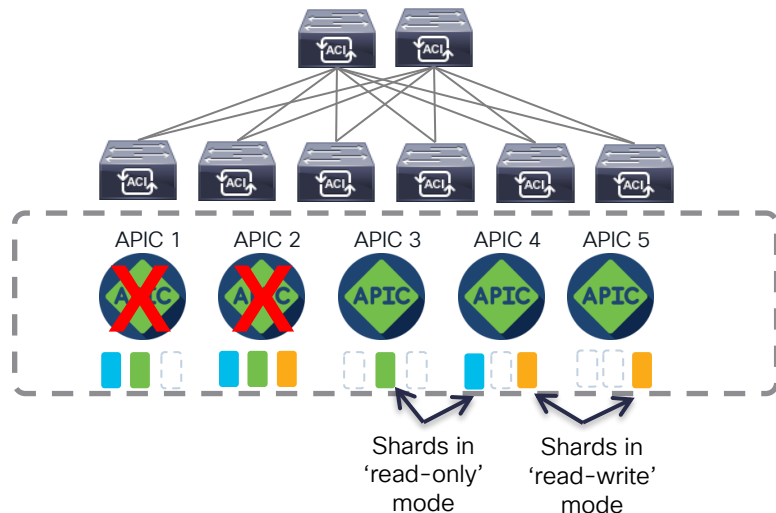
Each shard is replicated up to three replicas regardless of the size of your cluster.

APIC Cluster Deployment Considerations

Single Pod Scenario



- APIC will allow read-only access to the DB when only one node remains active (standard DB quorum)
- Hard failure of two nodes cause all shards to be in 'read-only' mode (of course reboot etc. heals the cluster after APIC nodes are up)



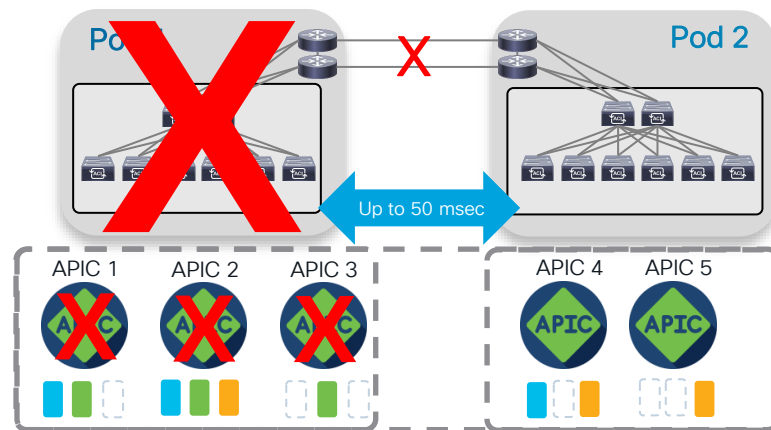
- Additional APIC will increase the system scale (up to 7* nodes supported) but does not add more redundancy
- Hard failure of two nodes would cause inconsistent behavior across shards (some will be in 'read-only' mode, some in 'read-write' mode)

APIC Cluster Deployment Considerations

Multi-Pod – 2 Pods Scenario



- **Pod isolation scenario:** changes still possible on APIC nodes in Pod1 but not in Pod2
- **Pod hard failure scenario:** recommendation is to activate a standby node to make the cluster fully functional again

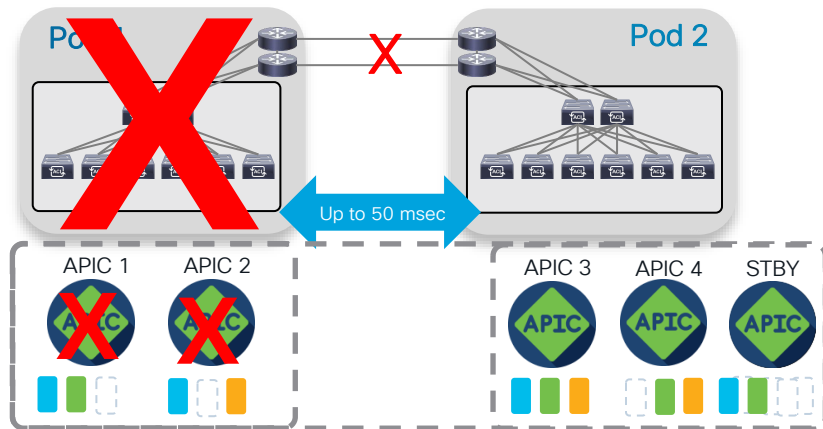


- **Pod isolation scenario:** same considerations as with single Pod (different behavior across shards)
- **Pod hard failure scenario:** may cause the loss of information for the shards replicated across APIC nodes in the failed Pod

Possible to restore the whole fabric state to the latest taken configuration snapshot ('ID Recovery' procedure – **needs BU and TAC involvement or Warm Standby from 6.1(2))**)

APIC Cluster Deployment Considerations

Two pods? 200 or fewer leaf nodes? Deploy a 4-APIC cluster



- Avoids Fabric-ID recovery scenario
- **Pod isolation scenario**: same considerations as with 5 nodes (different behavior across shards)
- **Pod hard failure scenario**
 - No chance of total loss of information for any shard
 - Can bring up a standby node in the second site to regain full majority for all the shards



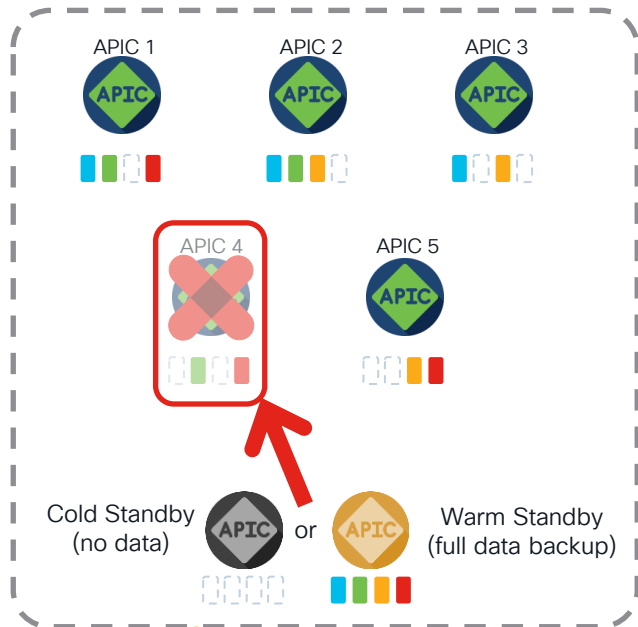
Warm Standby APIC

ACI Release 6.1(2)

- ACI release 6.1(2) introduces Warm Standby APIC
- Warm standby solves disaster recovery scenarios when fabric is in unrecoverable state (3 or more active APICs lost). **Avoids BU and TAC involved fabric-id recovery**
- Warm Standby can replace single APIC failures like Cold Standby (falls back to Cold Standby behavior)
- Warm standby is a fabric level setting. When configured, all Cold Standby APICs will be converted to Warm Standby
- Max 3 Warm Standby APICs supported

Standby APIC Use Cases

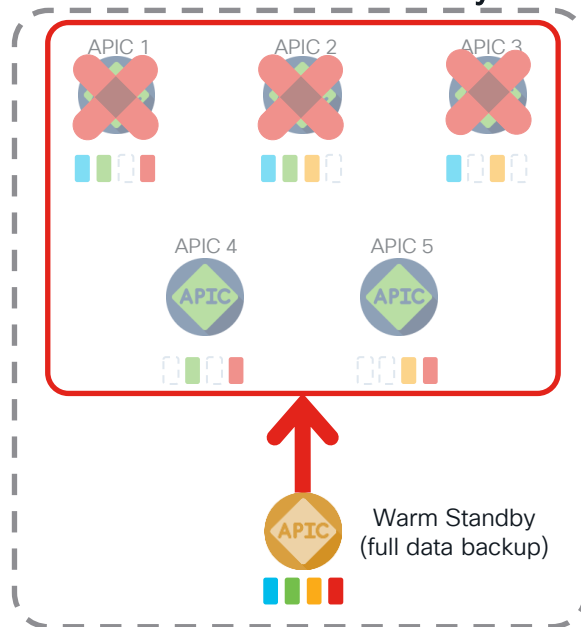
Single APIC Replacement



Replace a broken APIC as soon as possible. Data is synchronized from remaining APICs after replacement.



Rebuild the cluster for disaster recovery

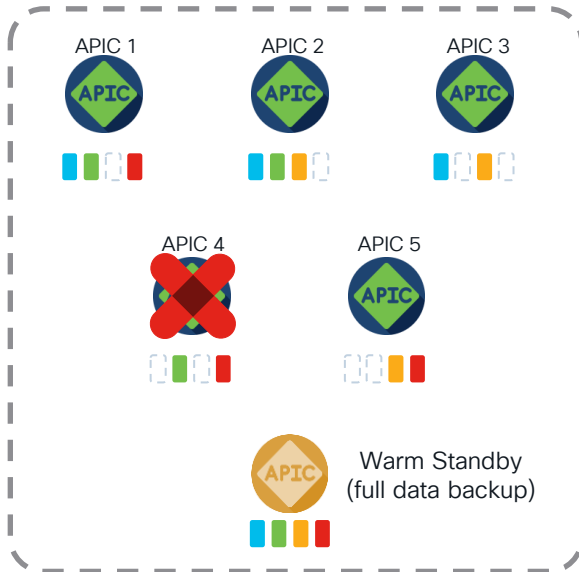


Data loss due to three or more APICs are lost. Rebuild the entire APIC cluster with the standby as new APIC1 via the data stored in the standby.

Single APIC Replacement with Warm Standby APIC

The procedure is the same as Cold Standby

1. Hardware failure on APIC4.
Need a replacement.



2. Promote the Cold Standby APIC as APIC 4



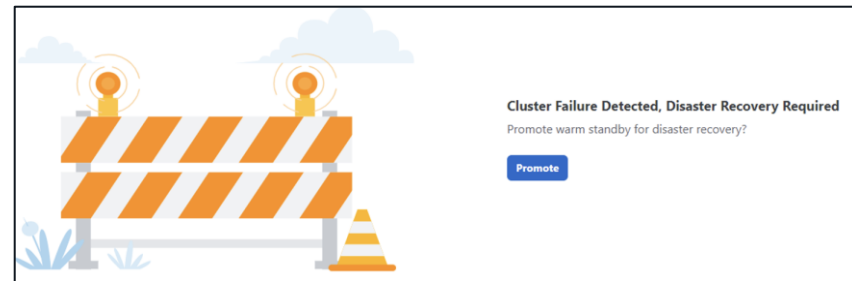
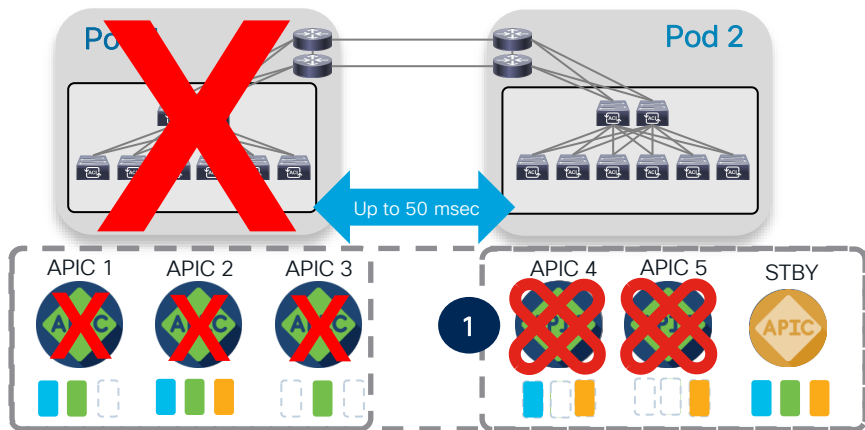
3. Data is synchronized from the remaining APICs.



(under the hood)
Data stored in the warm standby node is cleared

APIC Cluster Deployment Considerations

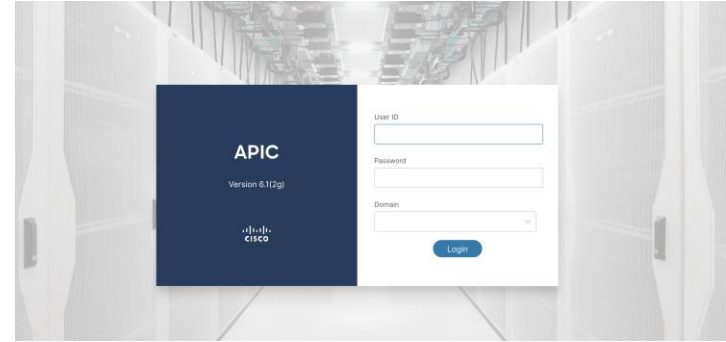
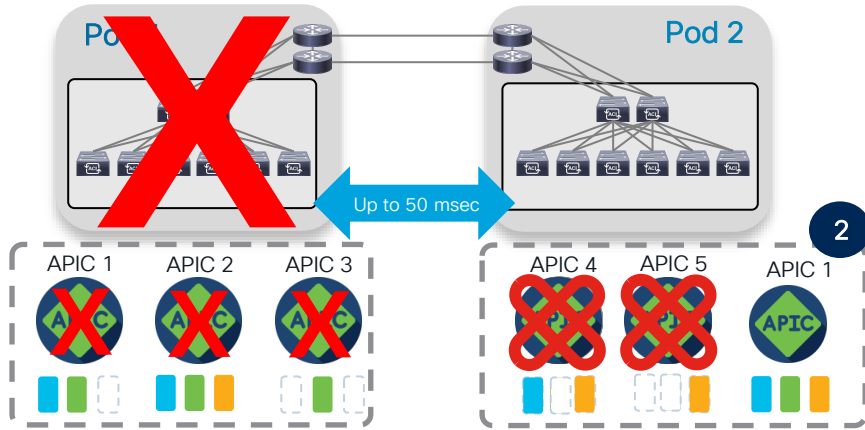
Multi-Pod – 5 APIC DR Scenario with Warm Standby APIC



1. Warm standby APIC disconnects APIC 4 and 5 fabric infra links before being promoted to APIC 1

APIC Cluster Deployment Considerations

Multi-Pod – 5 APIC DR Scenario with Warm Standby APIC

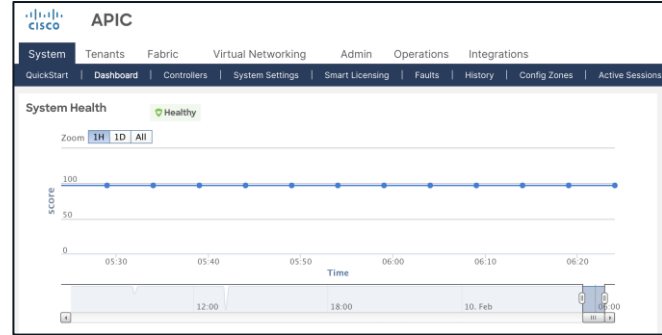
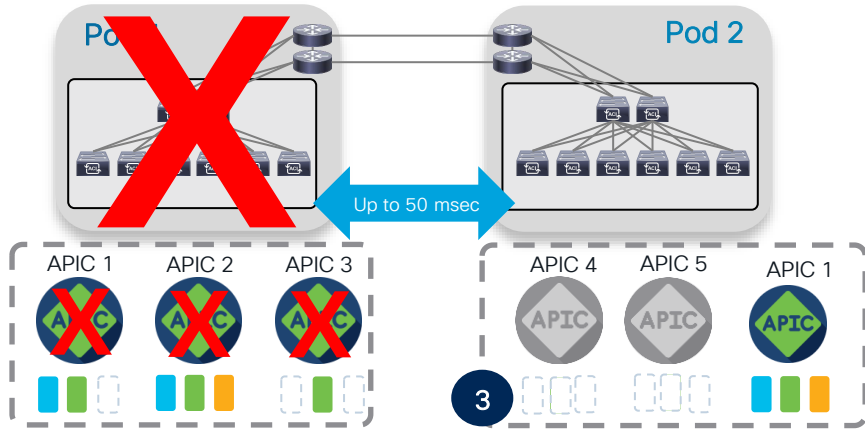


2. Warm Standby promoted to APIC 1

When warm standby comes up as APIC 1, it is read/write and can be used to manage the fabric

APIC Cluster Deployment Considerations

Multi-Pod – 5 APIC DR Scenario with Warm Standby APIC

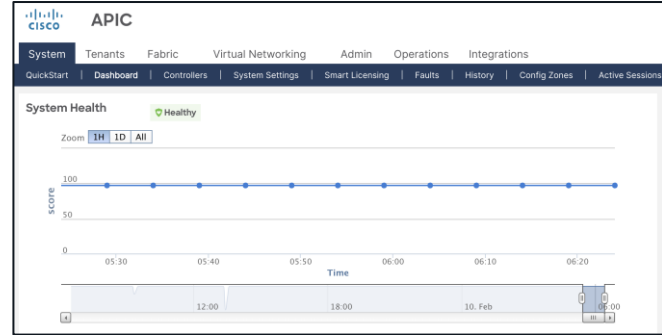
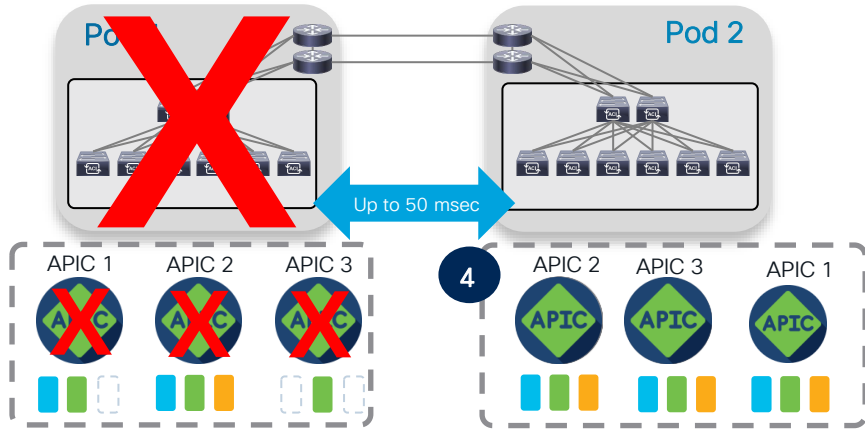


3. User clean reboots APIC 4 and 5

- **Pod hard failure scenario:** Complete loss of data for some shards. Disaster recovery scenario
- User initiates Disaster Recovery procedure from UI of Warm Standby APIC

APIC Cluster Deployment Considerations

Multi-Pod – 5 APIC DR Scenario with Warm Standby APIC



4. User rebuilds the cluster by adding back APICs 4 and 5 as APIC 2 and 3

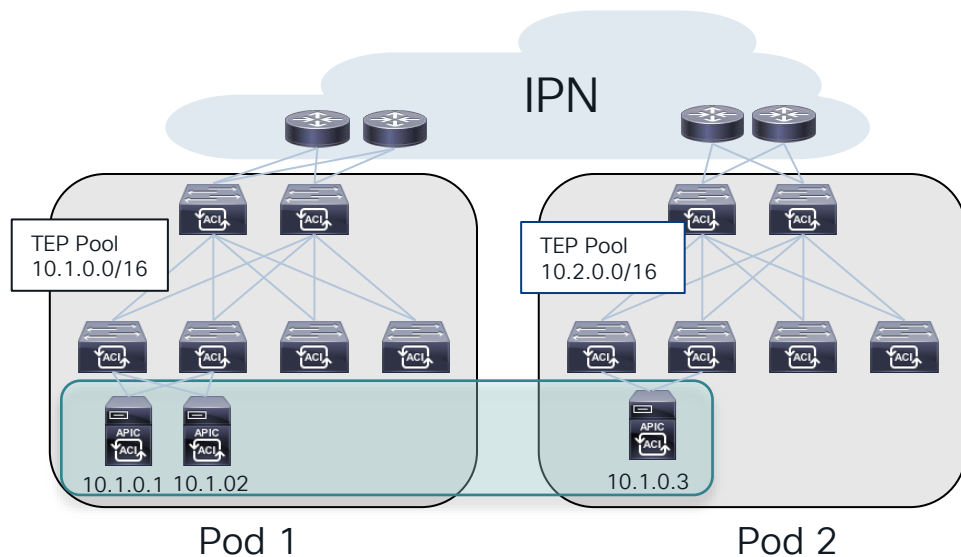
- Pod hard failure scenario: Complete loss of data for some shards. Disaster recovery scenario
- User initiates Disaster Recovery procedure from UI of Warm Standby APIC

APIC Connectivity over L3 Network



APIC Connectivity Options

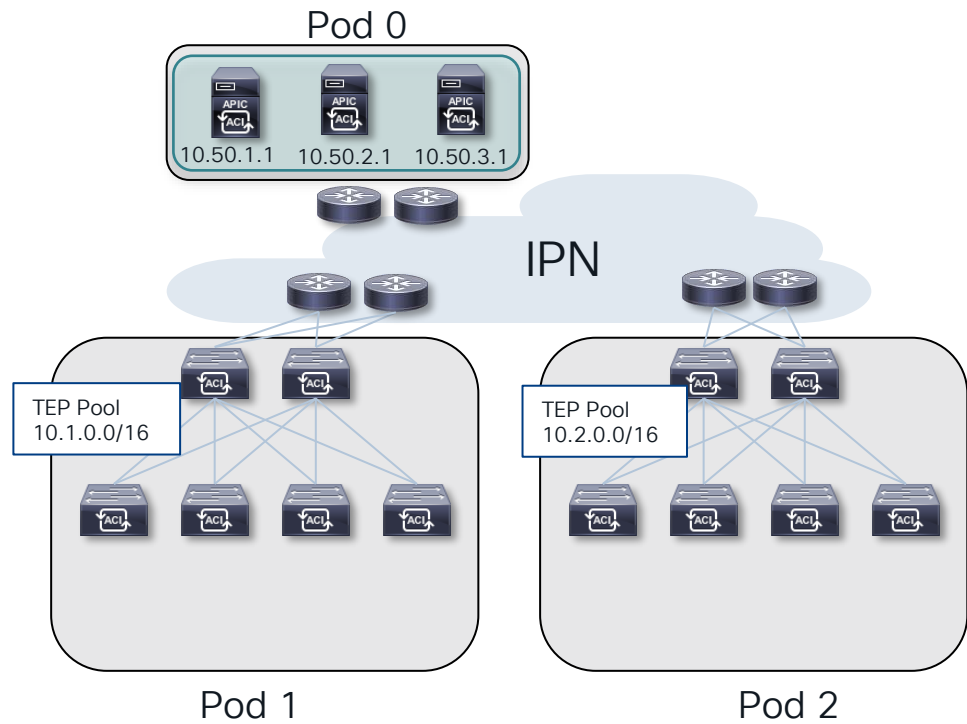
APIC Cluster directly connected to fabric



- APICs can be placed in any pod
- APIC fabric IP addresses are always assigned from pod 1 TEP pool
- Recommended to distribute APICs across pods so loss of a pod does not bring down the entire cluster

APIC Connectivity Options

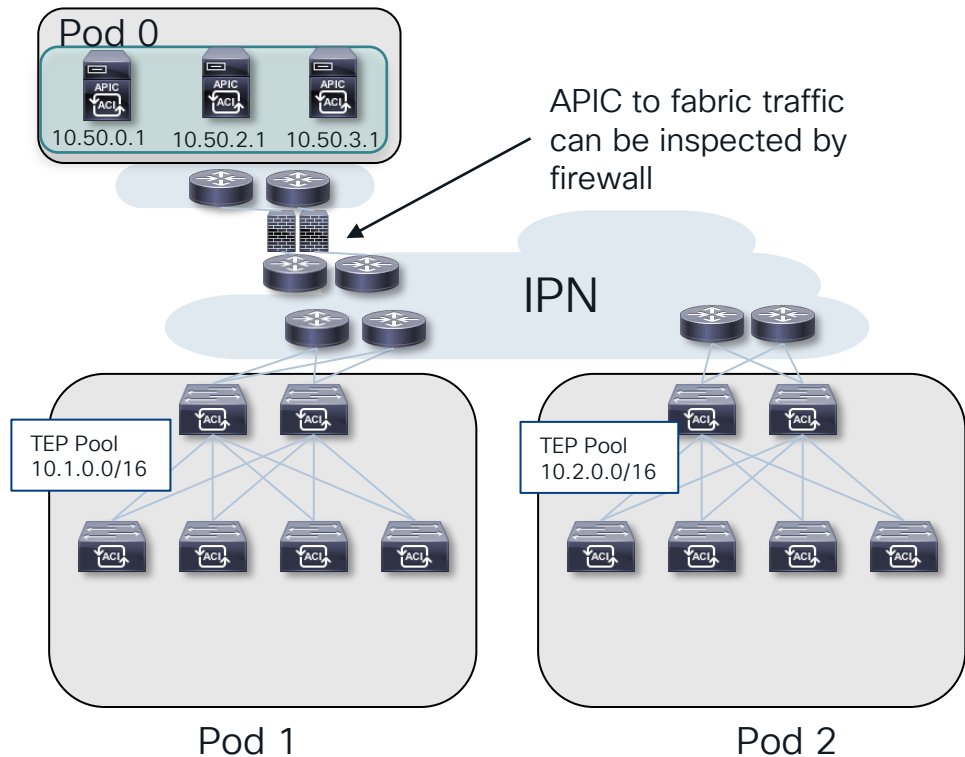
APIC cluster connected over L3 Network



- APICs are connected over IPN, not directly to leaf switches
- APICs connected to IPN are considered part of Pod 0
- APIC fabric IP addresses are user configurable. Not assigned from pod TEP range
- Each APIC fabric IP can be in a different subnet
- APICs can be geographically distributed within the Multi-Pod 50 msec RTT limit

APIC Connectivity Options

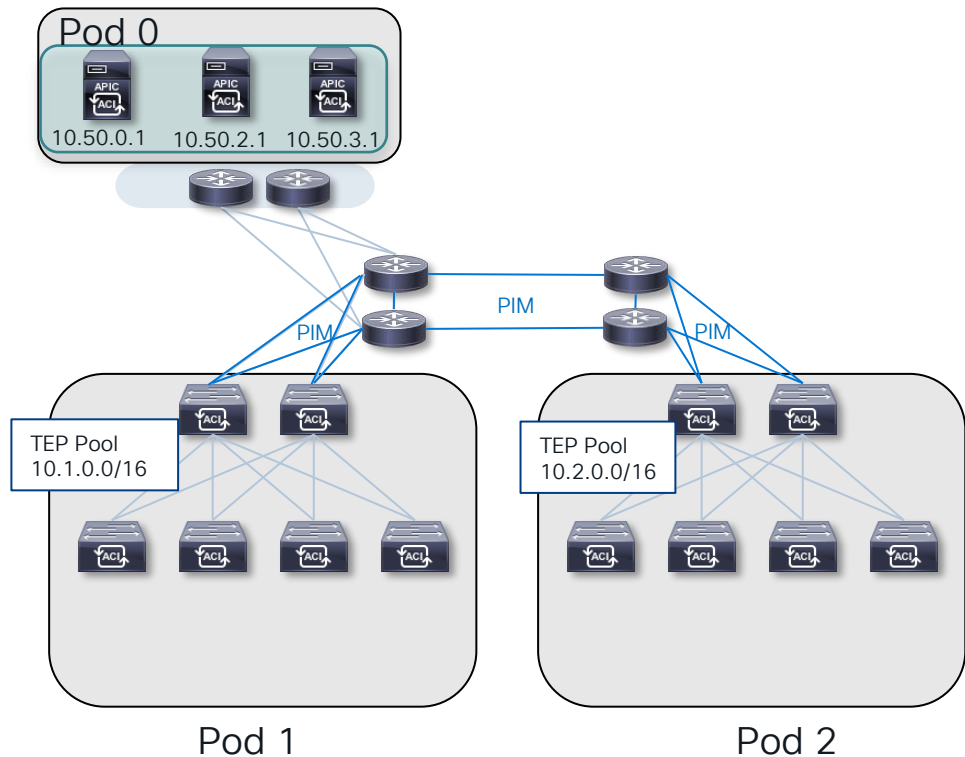
APIC cluster connected over L3 Network, Secure Zone Use Case



- APIC Cluster over L3 Network supports use case where all traffic between APIC and switches must be inspected by a firewall
- APIC cluster can be placed in a secure zone where all traffic into and out of the zone is inspected by a firewall

APIC Connectivity Options

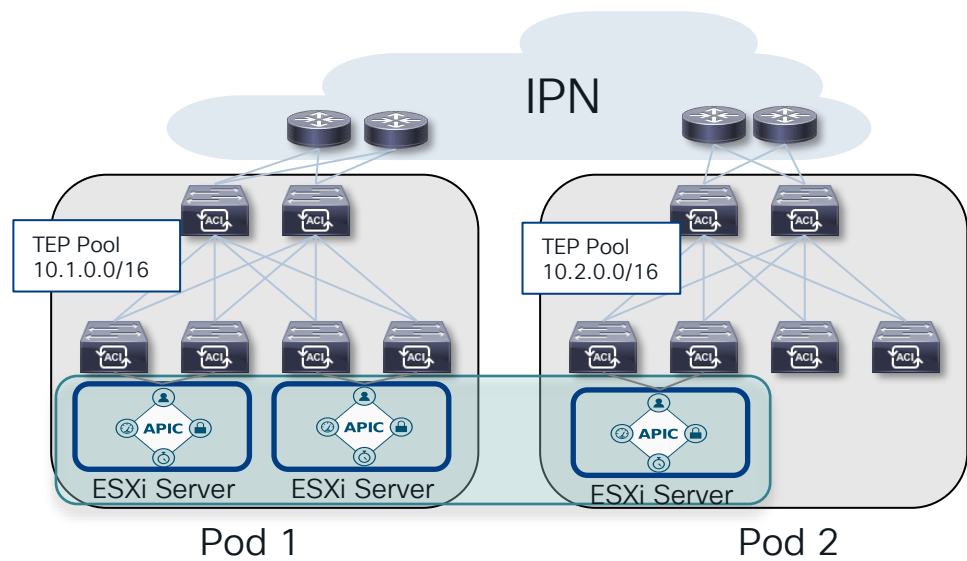
APIC cluster connected over L3 Network, IPN Multicast Requirement



- Multicast (PIM Bidir) is only required for inter-pod BUM traffic
- If APIC cluster over L3 network is managing only one pod, multicast is not required in the IPN
- If it is a Multi-Pod fabric, multicast is only required on the links interconnecting the pods

APIC Connectivity Options

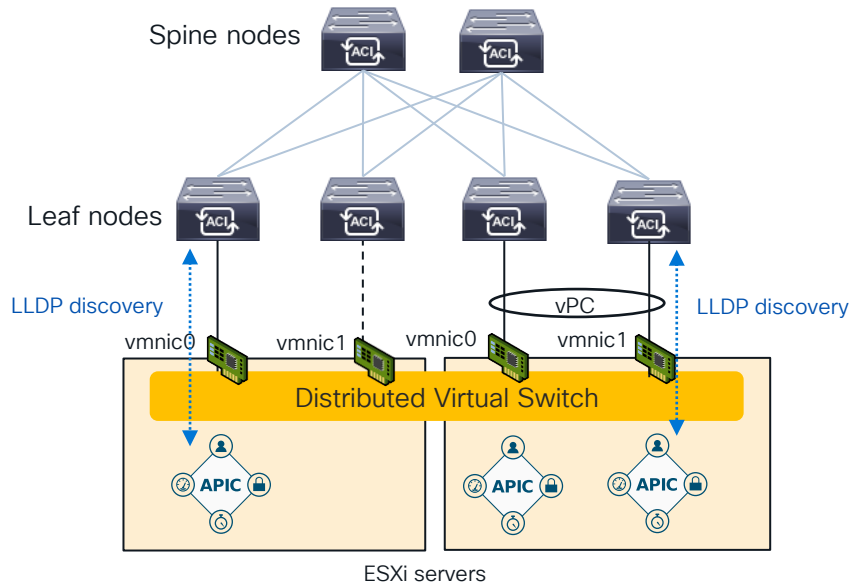
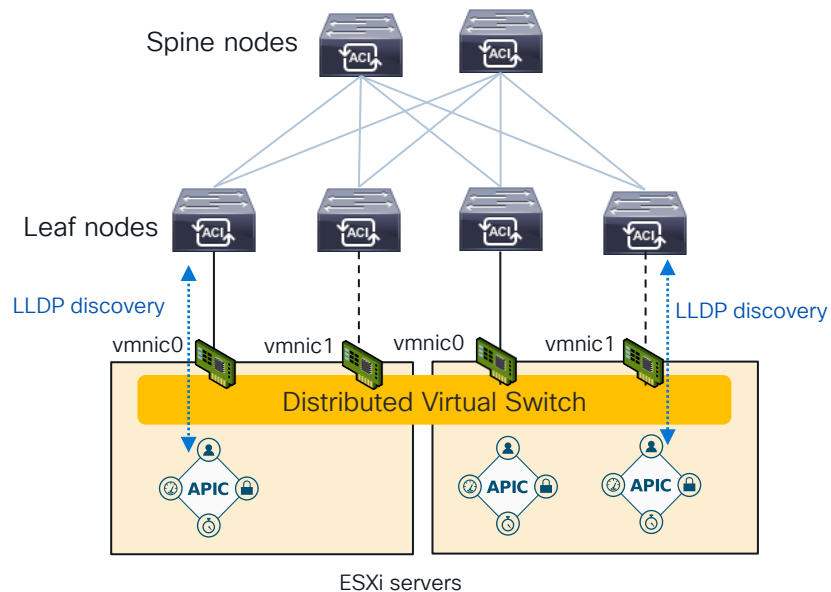
Virtual APIC cluster



- Virtual APIC cluster (all virtual APICs)
- Runs as a VM on an ESXi hypervisor
- ESXi server directly connected to fabric
- No mixed cluster support. Must be all virtual or all physical
- Supports all types of deployments, Remote Leaf, Multi-Pod, Multi-Site.

Topology considerations for virtual APIC on ESXi Directly Attached

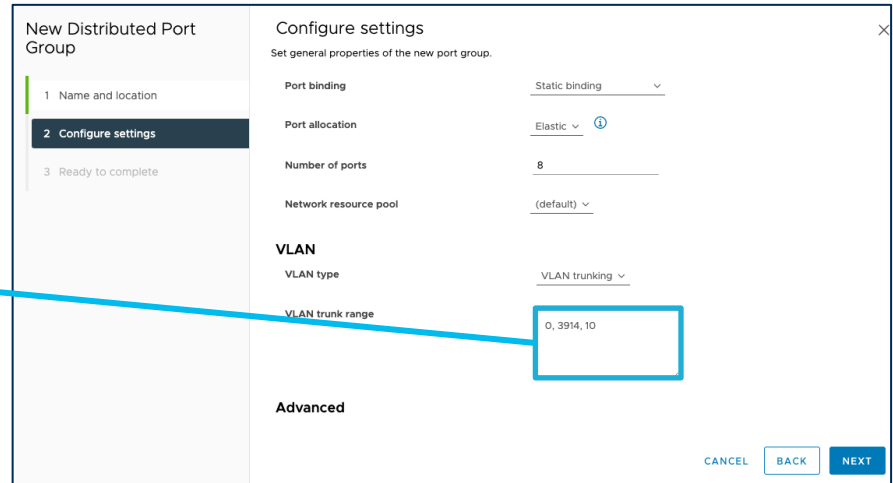
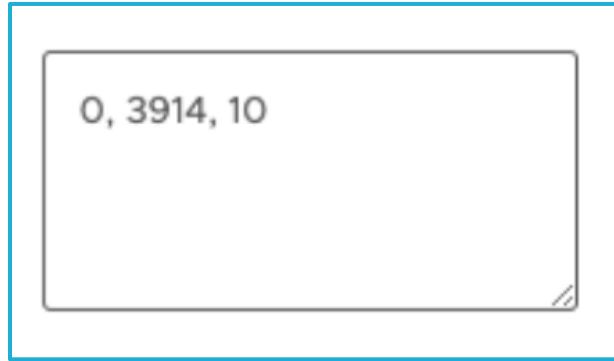
- ESXi servers need to be directly connected to ACI leaf nodes via individual links or vPC. (APIC1 must use Active-Standby instead of Active-Active with vPC)
- LLDP must be disabled on the virtual switch for LLDP discovery between leaf nodes and vAPICs.



Distributed Switch configuration

Port Group VLAN configuration

- VLAN type: VLAN Trunking
- VLAN trunk range:
 - VLAN 0 (VLAN 0 is required for APIC LLDP discovery)
 - ACI Infra VLAN (for example, 3914 is used as the default value during APIC initial setup)
 - Inband VLAN(s) (VLAN 10 in the example)



New Distributed Port Group

Configure settings

Set general properties of the new port group.

1 Name and location

2 Configure settings

3 Ready to complete

Port binding: Static binding

Port allocation: Elastic

Number of ports: 8

Network resource pool: (default)

VLAN

VLAN type: VLAN trunking

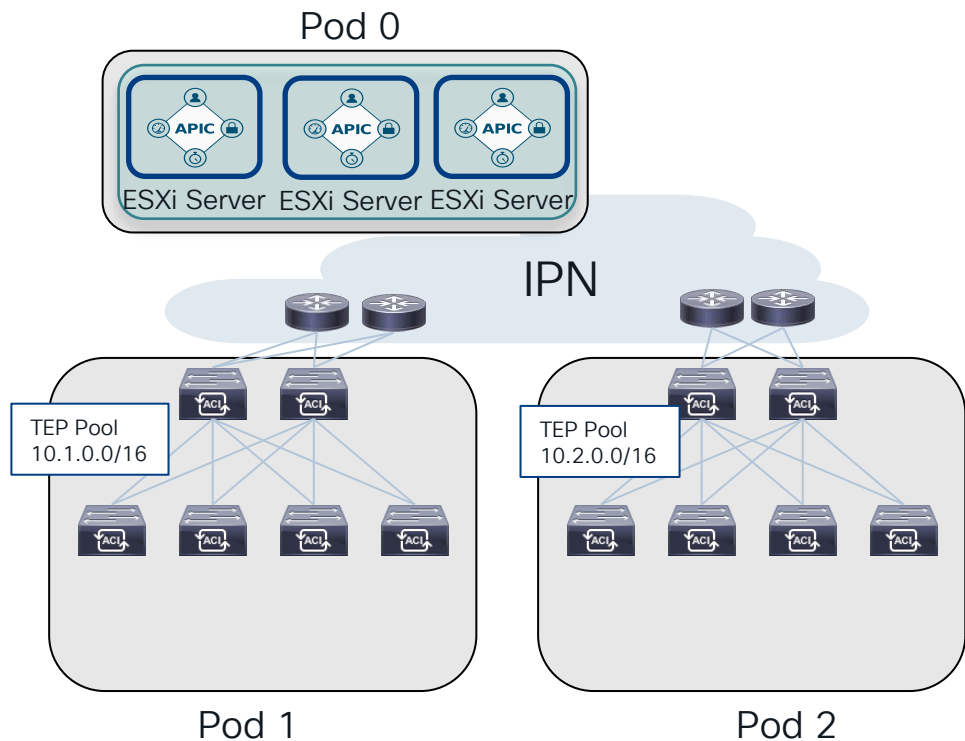
VLAN trunk range: 0, 3914, 10

Advanced

CANCEL BACK NEXT

APIC Connectivity Options

Virtual APIC cluster over L3 Network



- Virtual APIC over L3 Network
- Same or different IP addresses per APIC same as physical APIC over L3 network
- Cannot mix virtual APIC over L3 Network with directly connected virtual APIC



Physical cluster to virtual cluster migration

Support for cluster migration added in ACI 6.1(1) release

Cluster As Seen By Node Refresh

Cluster Overview Migrate Migration History Edit

Fabric Name, Infra VLAN, TEP Pool and GiPo Pool can't be edited unless the fabric is reset to its factory settings

Cluster Member's Health 3 Fully Fit 3

Fabric Attachment Directly Attached	Fabric Name S2-Fabric	Fabric ID 1	Infra VLAN 3914
TEP Pool 10.2.0.0/16	Current Cluster Size 3	GiPo Pool Address 225.0.0.0/15	Target Cluster Size 3
Discovery Mode Permissive	Shrink In Progress No	Standby Type Warm	

Controllers

Filter by attributes Actions

ID	Controller Name	IP Address	Pod ID	Admin State	Operation State	Health	Failover Status	Mode	Serial Number	Type
1	S2-API...	172.31.184.1 View All	1	In S...	Avail...	Full...	No Ap...	Acti...	FCH19...	Physical
2	S2-API...	172.31.184.1 View All	1	In S...	Avail...	Full...	No Ap...	Acti...	FCH19...	Physical
3	S2-API...	172.31.184.1 View All	1	In S...	Avail...	Full...	No Ap...	Acti...	FCH19...	Physical

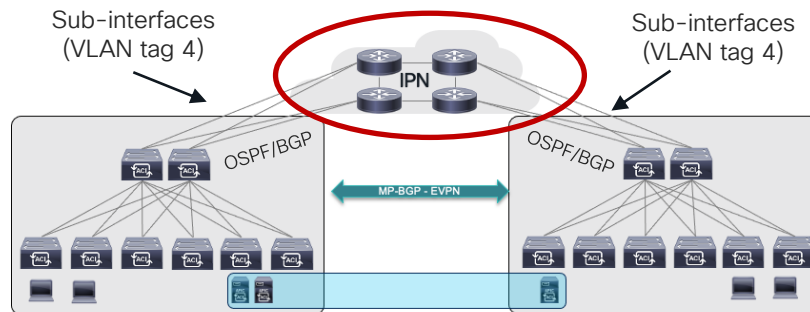
- From 6.1(1) you can now from a physical APIC cluster to a virtual APIC cluster (and vice versa)
- Migrate workflow handles the migration (decommission, commission) for all nodes automatically
- Mix of physical and virtual cluster not supported

Inter-Pod Connectivity Deployment Considerations



ACI Multi-Pod

Inter-Pod Network (IPN) Requirements



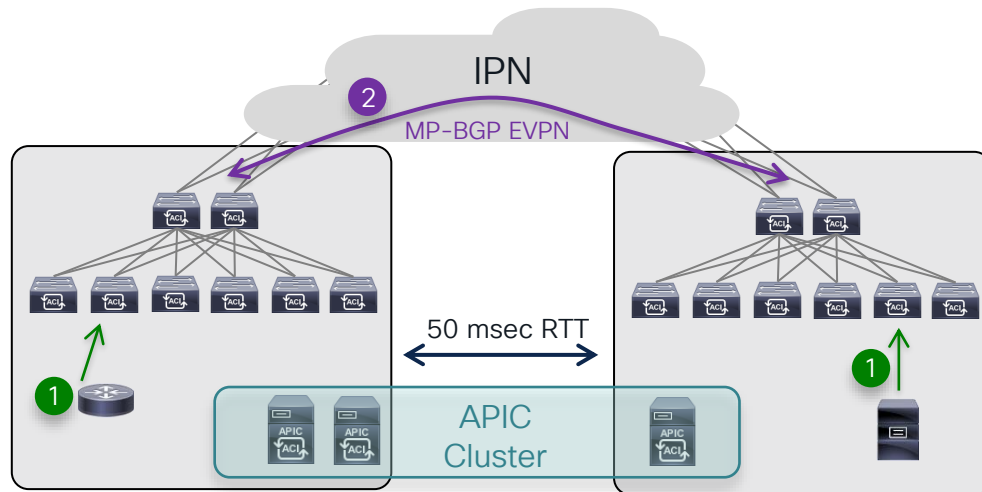
- Not managed by APIC, must be separately configured (day-0 configuration)
- IPN topology can be arbitrary, not mandatory to connect to all spine nodes
- Main requirements:
 - Multicast BiDir PIM → needed to handle Layer 2 BUM* traffic
 - OSPF or BGP to peer with the spine nodes and learn VTEP reachability
 - Increase MTU support to handle VXLAN encapsulated traffic
 - DHCP-Relay

* Broadcast, Unknown unicast, Multicast

ACI Multi-Pod and MTU

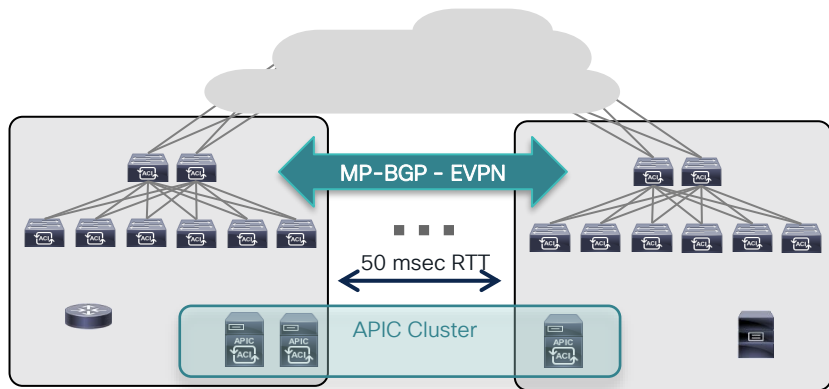
Different MTU Meanings

1. **Data Plane MTU:** MTU of the traffic generate by endpoints (servers, routers, service nodes, etc.) connected to ACI leaf nodes
 - Need to account for 50B of overhead (VXLAN encapsulation) for inter-Pod communication
2. **Control Plane MTU:** for CPU generated traffic like EVPN across sites
 - The default value is **9000B**, can be tuned to the maximum MTU value supported in the IPN

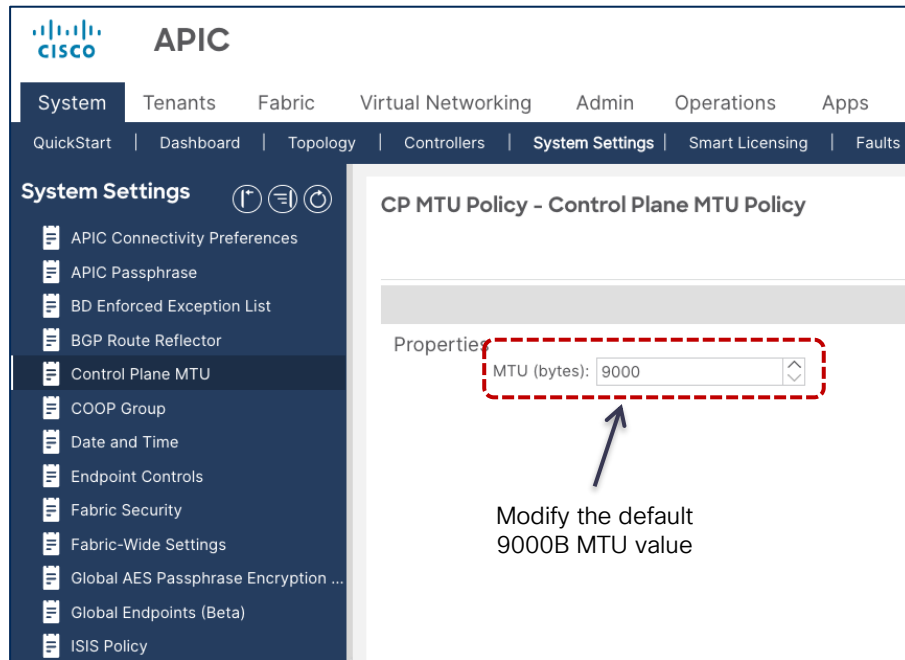


ACI Multi-Pod and MTU

Tuning CP MTU for EVPN Traffic across Pods

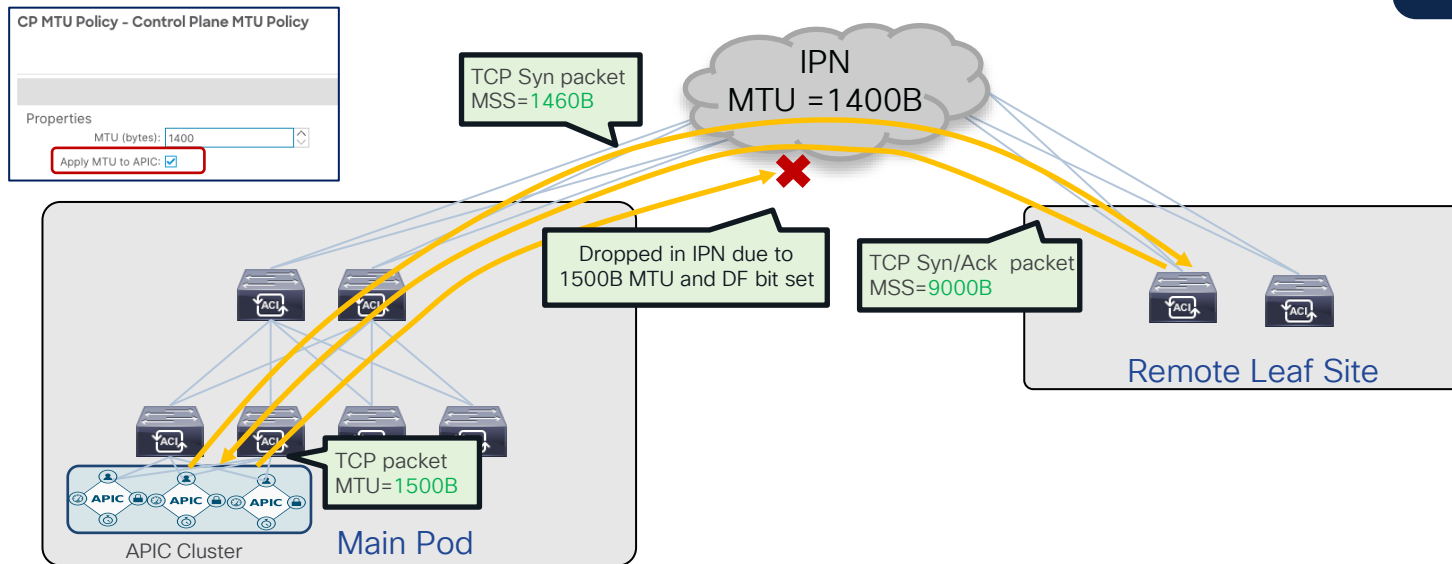


- Control Plane MTU can be set leveraging the “CP MTU Policy” on APIC
- The required MTU in the IPN would then depend on this setting and on the Data Plane MTU configuration
 - Always need to consider the VXLAN encapsulation overhead for data plane traffic (50/54 bytes)



Apply Control Plane MTU to APIC

ACI Release 6.0(4)F

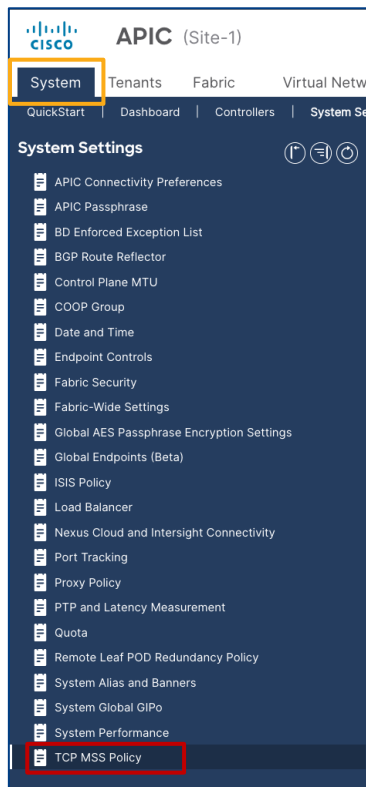


- Leaf discovery may fail for Remote Leaf and Multi-Pod switches if the IPN network supports less than 1500-byte MTU
- From ACI release 6.0(4) you can set the APIC fabric interface to the configured CP MTU value (if less than 1500 bytes)

ACI Multi-Site and MTU Size

Introducing the TCP-MSS Adjust Functionality

ACI Release 6.0(3)F



The screenshot shows the 'TCP MSS Policy' configuration page. The 'Type' is set to 'Global'. The 'IPv4' and 'IPv6' fields are both set to 8888. A red box highlights the 'IPv4' and 'IPv6' fields, and a blue arrow points from the text 'Supported values are 688-9104 bytes' to the 'IPv4' field.

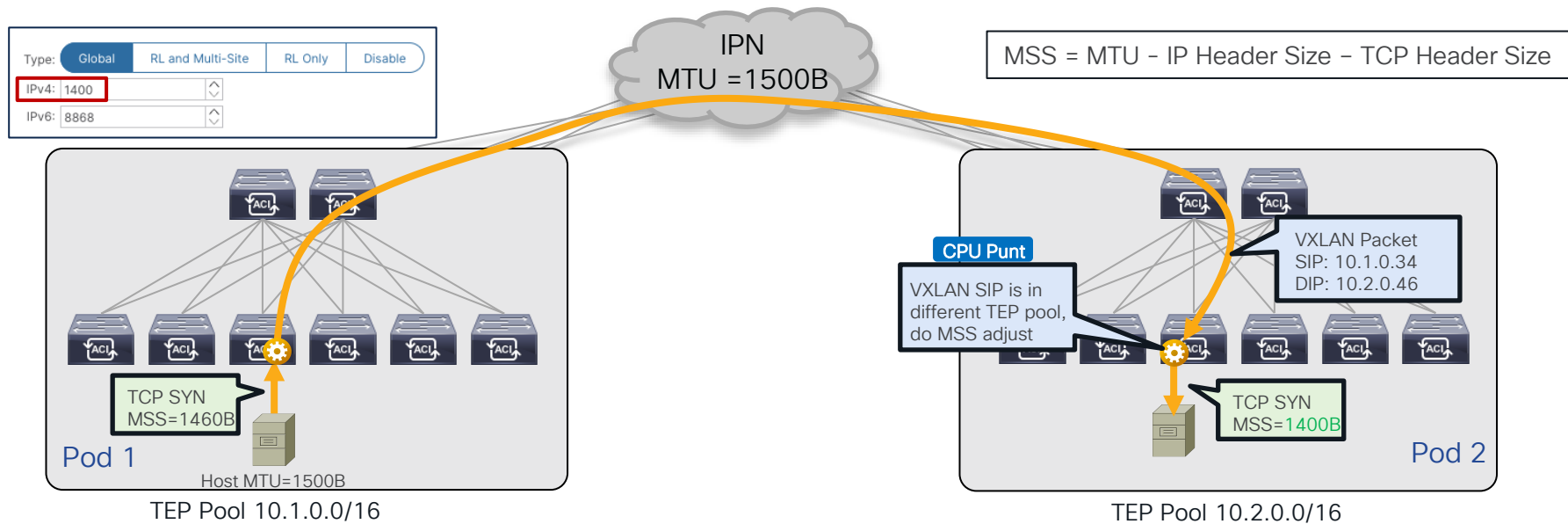
Supported values are
688-9104 bytes

- TCP MSS adjust policy is enabled at System Settings level
- Supports different TCP MSS adjust setting for IPv4 and IPv6
- Supports three different options:
 1. **Global**: applies to all flows (Multi-Pod, Multi-Site, RLs to LLs/RLs to RLs)
 2. **RL and Msite**: applies to Multi-Site and RLs to LLs/RLs to RLs flows
 3. **RL Only**: applies only to RLs to LLs/RLs to RLs flows

TCP-MSS Adjust Functionality

SYN Packet

ACI Release 6.0(3)F

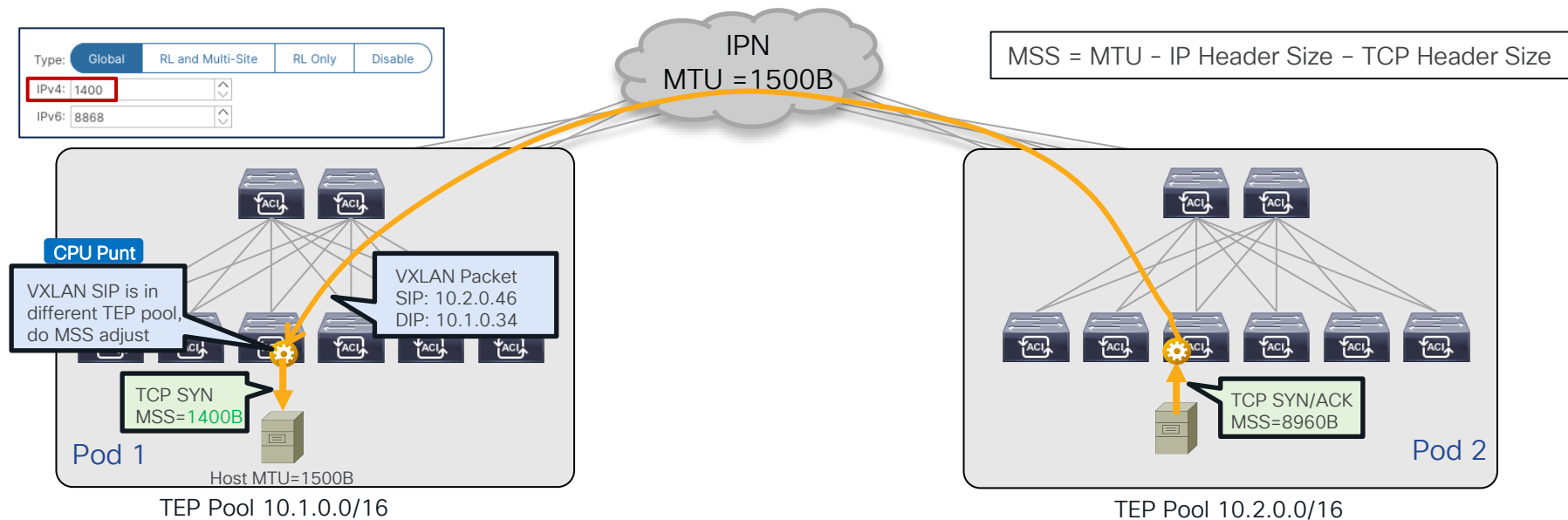


- TCP MSS adjust is always performed on the egress leaf node
- Adjusts TCP MSS value on SYN and SYN/ACK packets
- Checks for Source IP in the VXLAN header → TCP-MSS adjusts performed if the source IP is not part of the pods local internal TEP pool

TCP-MSS Adjust Functionality

SYN/ACK Packet

ACI Release 6.0(3)F

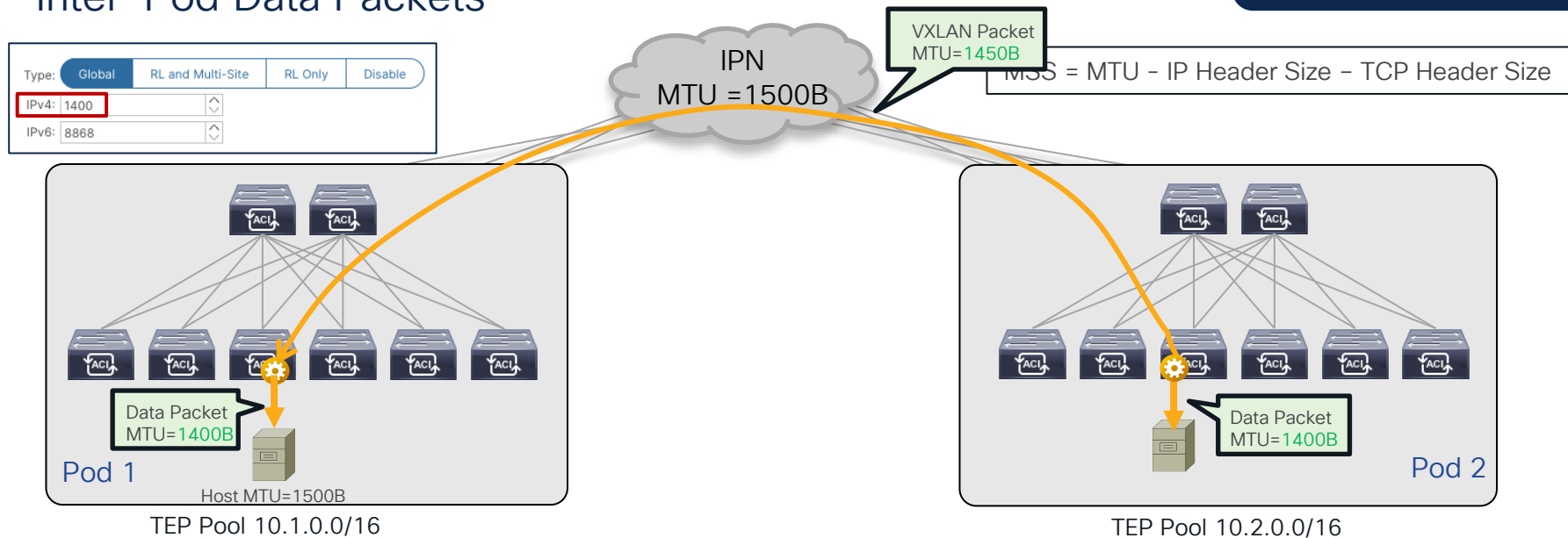


- TCP MSS adjust is always performed on the egress leaf node
- Adjusts TCP MSS value on SYN and SYN/ACK packets
- Checks for Source IP in the VXLAN header → TCP-MSS adjusts performed if the source IP is not part of the pods local internal TEP pool

TCP-MSS Adjust Functionality

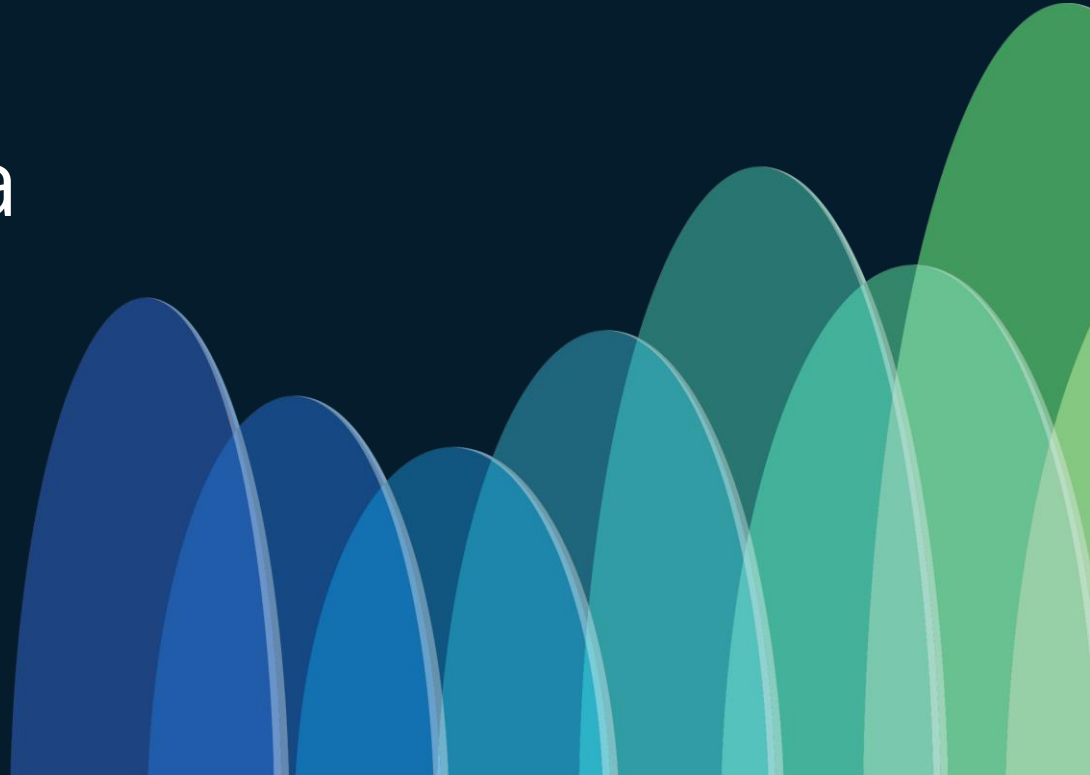
Inter-Pod Data Packets

ACI Release 6.0(3)F



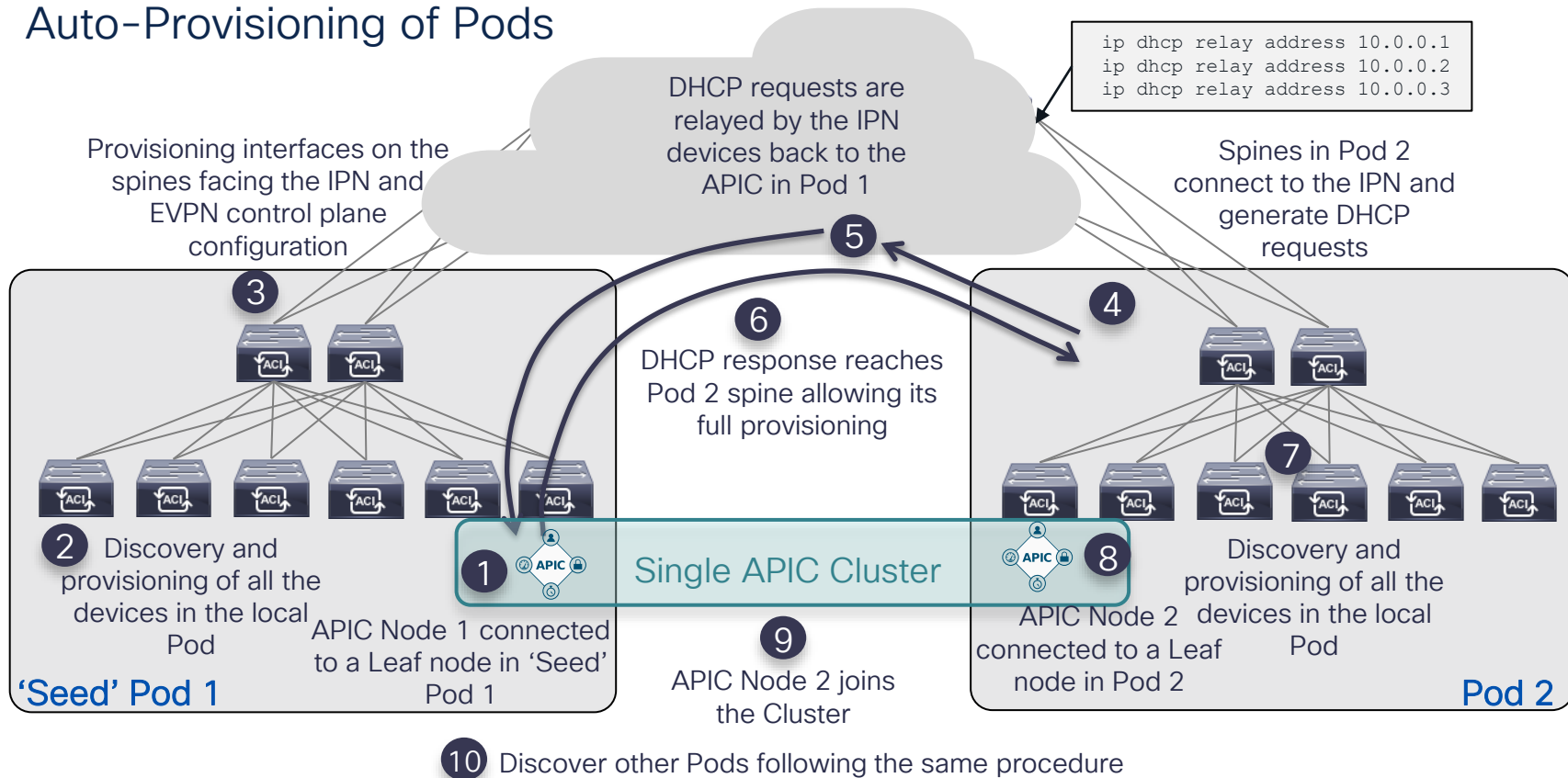
- As a result of the MSS negotiation, the endpoints generate packets for that TCP communication with MTU 1400B (irrespectively of the local Host MTU)
- The VXLAN encapsulated traffic can be successfully forwarded across the IPN

Control and Data Planes



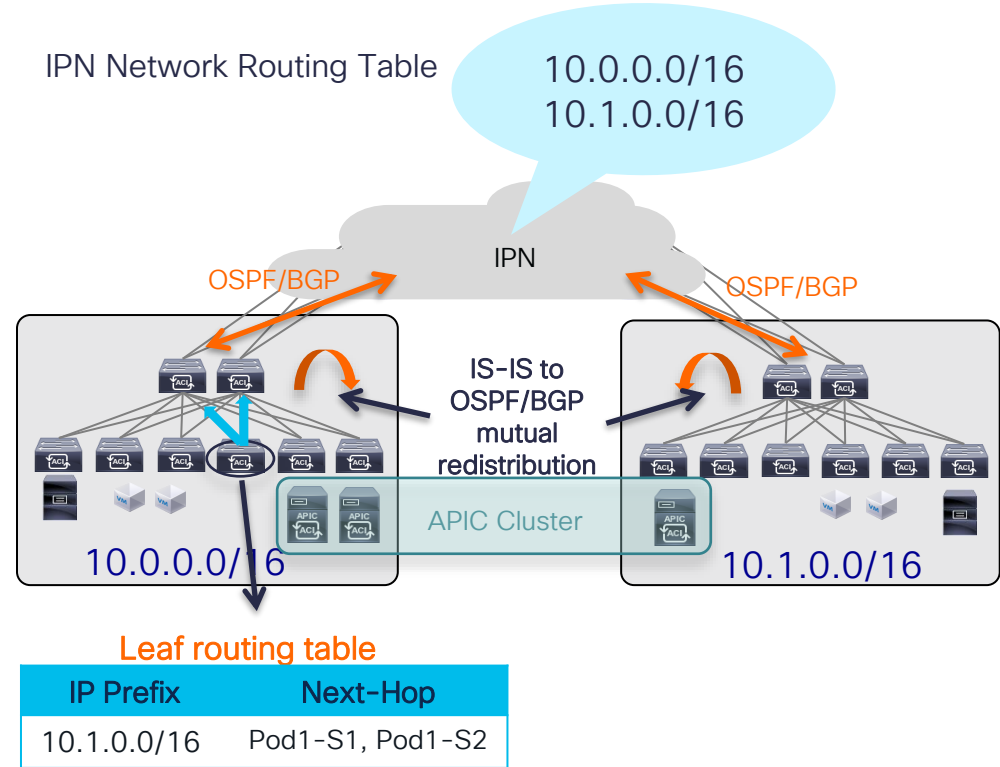
ACI Multi-Pod

Auto-Provisioning of Pods



Exchanging TEP information across pods

- Separate IP address pools for VTEPs assigned by APIC to each Pod
 - Summary routes advertised toward the IPN via OSPF or BGP routing
 - IS-IS convergence events local to a Pod not propagated to remote Pods
- Spine nodes redistribute other Pods summary routes into the local IS-IS process
 - Needed for local VTEPs to communicate with remote VTEPs



Exchanging TEP information across pods

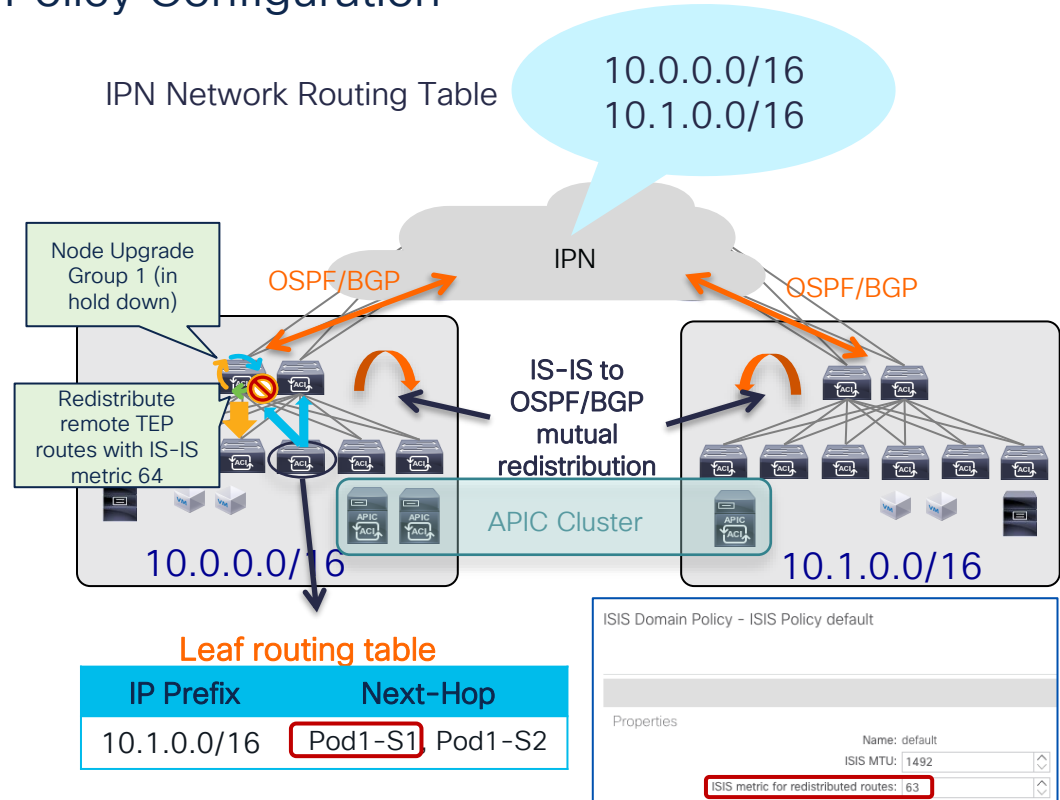
Issue with Default IS-IS Metric Policy Configuration

Default fabric wide IS-IS metric is set at 63 (max value)

During upgrade, spines set the overload mode while policy is being downloaded

If fabric-wide value is already using the max value for routes redistributed into IS-IS, the overload functionality is ineffective

This can create unexpected traffic interruption if leaf sends traffic to a spine which is not fully upgraded (and ready to forward traffic)

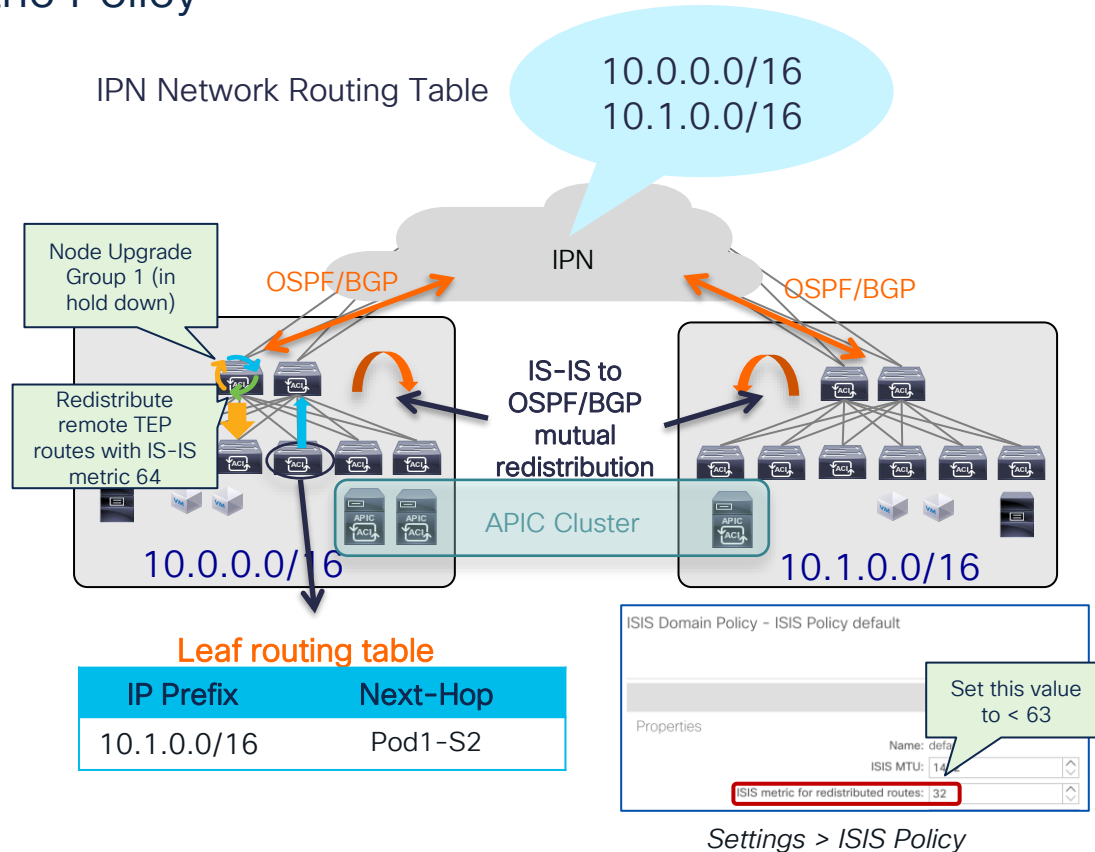


Settings > ISIS Policy (Default Config)

Exchanging TEP information across pods

Lowering the Default IS-IS Metric Policy

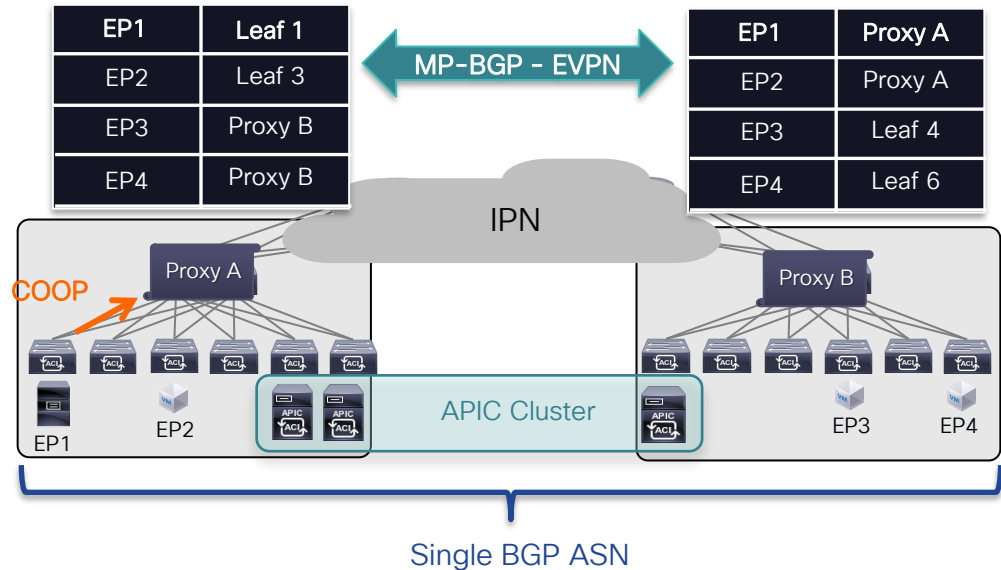
- By lowering the default ISIS metric value, connectivity to TEP prefixes received from the remote site will be preferred through the remaining spines
- This behavior gives time to the spine for completing the upgrade



ACI Multi-Pod

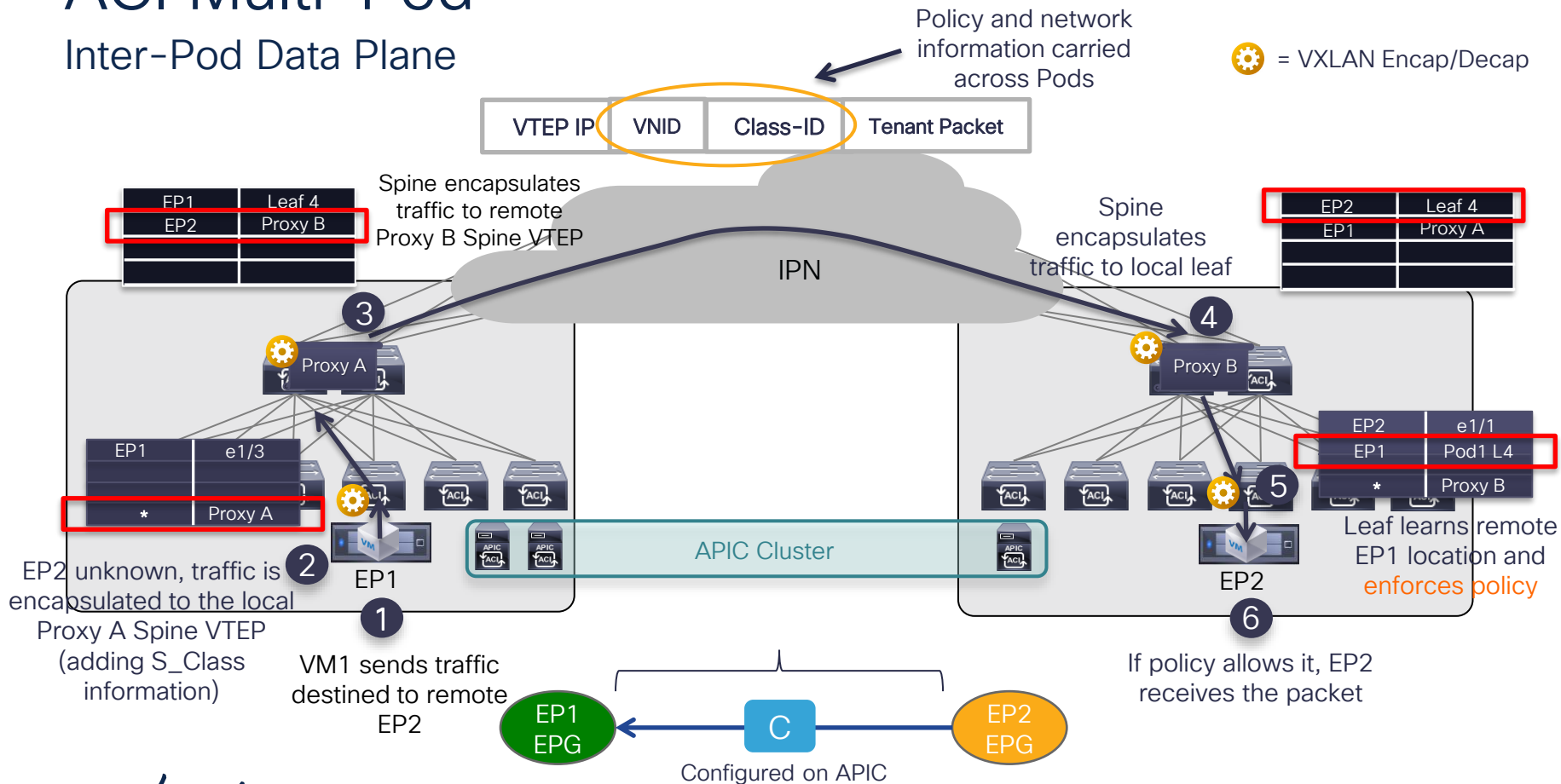
Inter-Pod MP-BGP EVPN Control Plane

- MP-BGP EVPN to sync Endpoint (EP) and Multicast Group information
 - All remote Pod entries associated to a Proxy VTEP next-hop address (not part of local TEP Pool)
 - Same BGP AS across all the Pods
- iBGP EVPN sessions between spines in separate Pods
 - Full mesh MP-iBGP EVPN sessions between local and remote spines (default behavior)
 - Optional RR deployment (recommended one RR in each Pod for resiliency)



ACI Multi-Pod

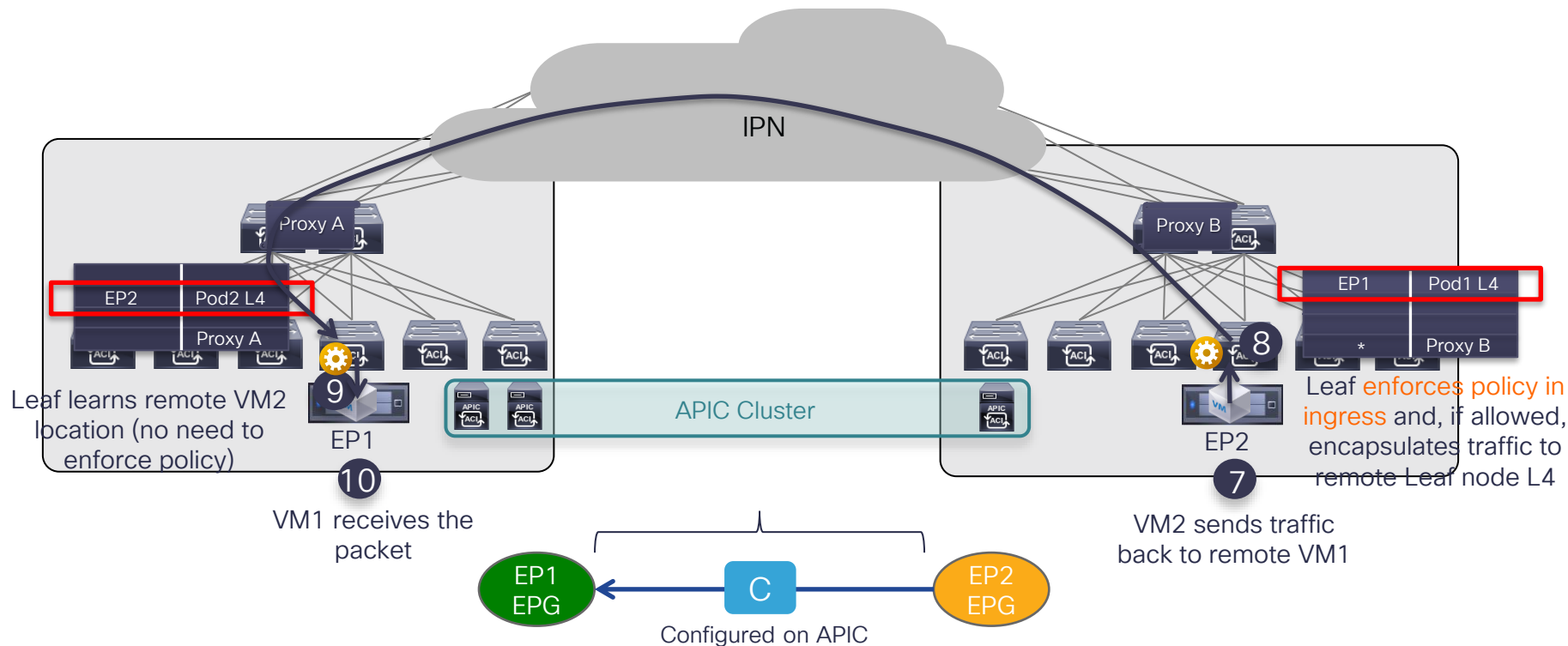
Inter-Pod Data Plane



ACI Multi-Pod

Inter-Pod Data Plane (2)

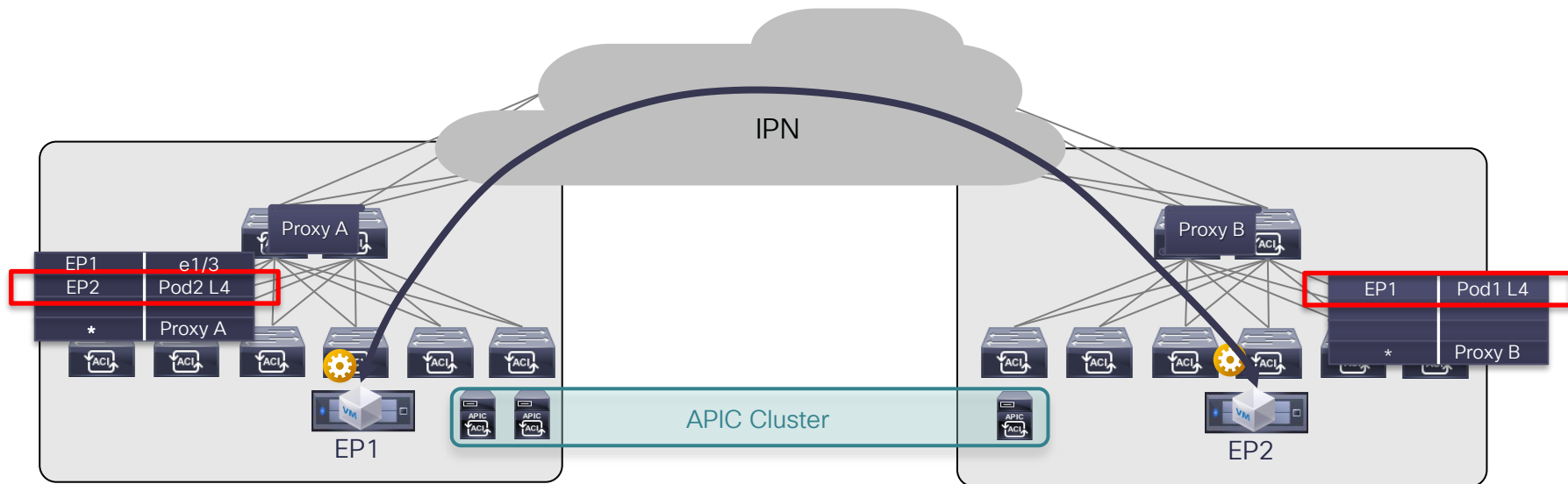
⚙️ = VXLAN Encap/Decap



ACI Multi-Pod

Inter-Pod Data Plane (3)

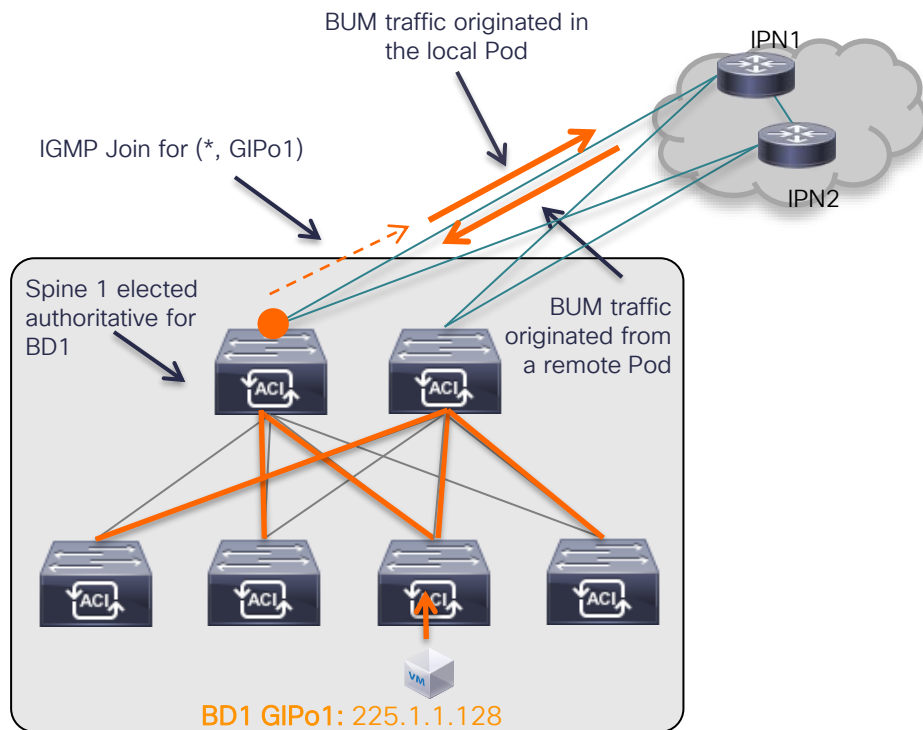
⚙️ = VXLAN Encap/Decap



- 11 From this point EP1 to EP2 communication is encapsulated Leaf to Leaf (VTEP to VTEP) and policy always applied at the ingress leaf (applies to both L2 and L3 communication)

ACI Multi-Pod

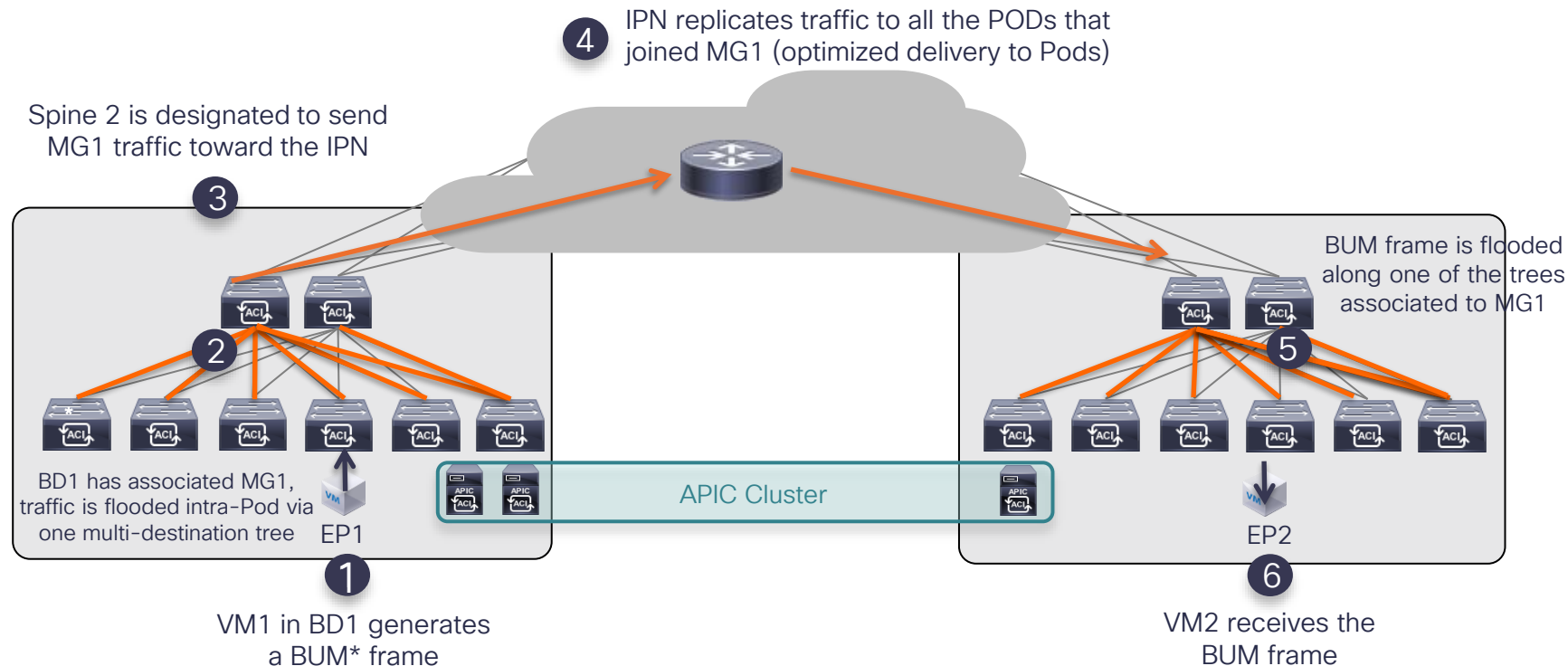
Use of Multicast for Inter-Pod Layer 2 BUM Traffic



- Ingress replication for BUM* traffic not supported with Multi-Pod
- PIM Bidir is the only validated and supported option
 - Scalable: only a single (*,G) entry is created in the IPN for each BD
 - Fast-convergent: no requirement for data-driven multicast state creation
- A spine is elected authoritative for each Bridge Domain:
 - Generates an IGMP Join on a specific link toward the IPN
 - Always sends/receives BUM traffic on that link

ACI Multi-Pod

Use of Multicast for Inter-Pod BUM Traffic

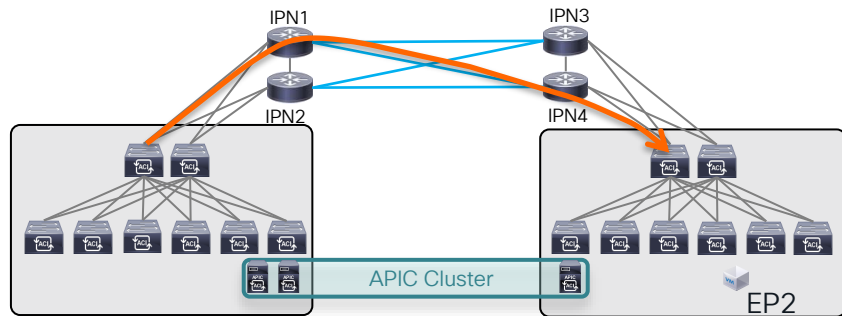


BUM: Layer 2 Broadcast, Unknown Unicast, Multicast

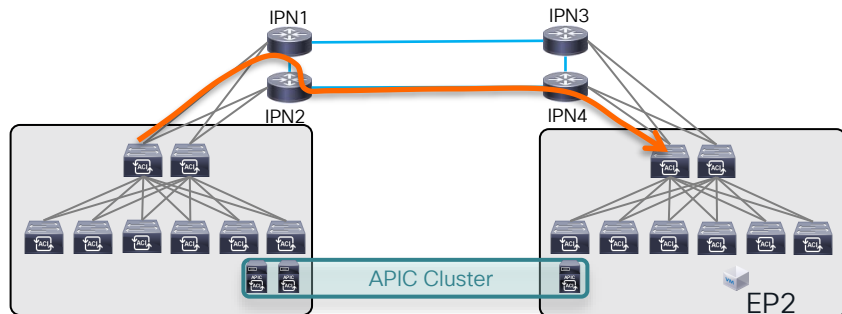
ACI Multi-Pod

PIM Bidir for BUM – Supported Topologies

Full Mesh between remote IPN devices

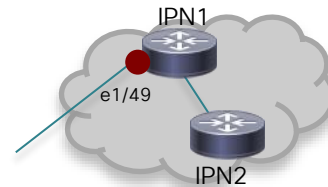


Directly connect local IPN devices



- Create full-mesh connections between IPN devices
- More costly for geo-dispersed Pods, as it requires more links between sites
- Alternatively, connect local IPN devices with a port-channel interface (for resiliency)
- In both cases, it is **critical** to ensure that the preferred path toward the RP from any IPN devices is not via a spine
- Recommendation is to increase the OSPF cost of the interfaces between IPN and spines

```
interface Ethernet1/49.4
  description L3 Link to Pod1-Spine1
  mtu 9150
  encapsulation dot1q 4
  ip address 192.168.1.1/31
  ip ospf cost 100
  ip ospf network point-to-point
  ip router ospf IPN area 0.0.0.0
  ip pim sparse-mode
  ip dhcp relay address 10.1.0.2
  ip dhcp relay address 10.1.0.3
```



ACI Multi-Pod

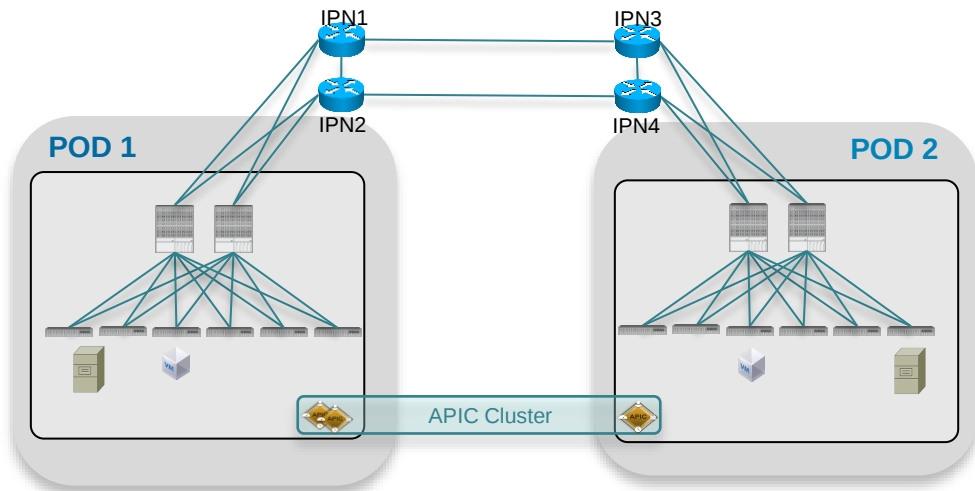
RP Redundancy with PIM Bidir

Active RP (IPN1)

```
ip pim rp-address 192.168.100.2 group-list 225.0.0.0/15 bidir
ip pim rp-address 192.168.100.2 group-list 239.255.255.240/28 bidir
interface loopback1
  description Phantom RP
  ip address 192.168.100.1/30
  ip ospf network point-to-point
  ip router ospf IPN area 0.0.0.0
  ip pim sparse-mode
```

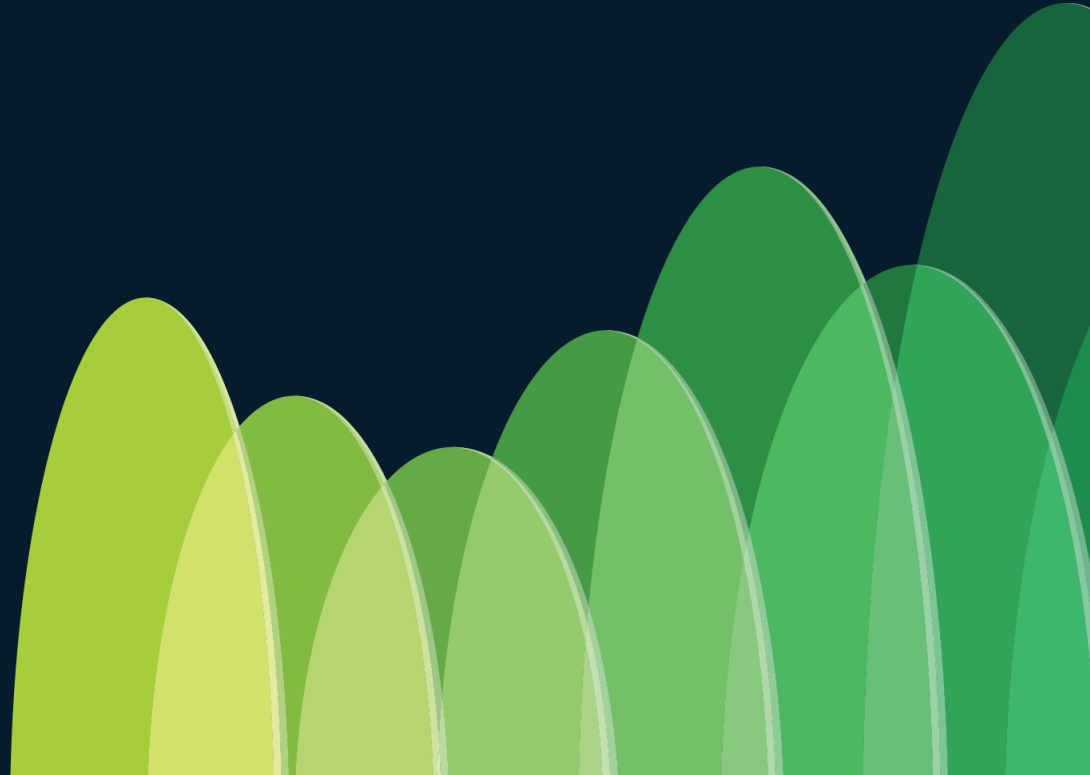
Standby RP (IPN3)

```
ip pim rp-address 192.168.100.2 group-list 225.0.0.0/15 bidir
ip pim rp-address 192.168.100.2 group-list 239.255.255.240/28 bidir
interface loopback1
  description Phantom RP
  ip address 192.168.100.1/29
  ip ospf network point-to-point
  ip router ospf IPN area 0.0.0.0
  ip pim sparse-mode
```



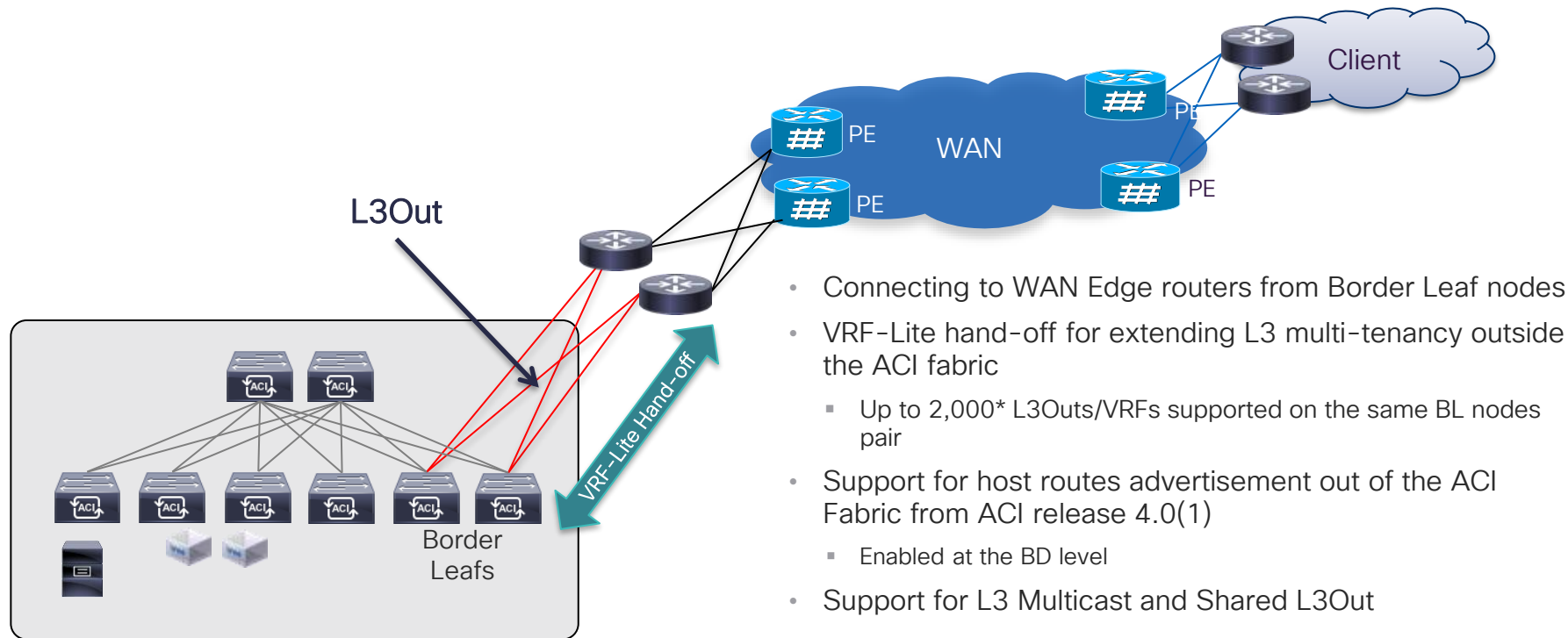
- In PIM Bidir, only one device functions as active RP for a given group at the same time
- Can leverage the ‘Phantom RP’ configuration for providing resiliency
- The RP role is ‘active’ on the device announcing the most specific route for the RP’s address
- In case of failure of the ‘active’ RP device, the shared tree is immediately rebuilt toward the ‘standby’ RP based on routing convergence event

Connecting to the External Layer 3 Domain



Connecting ACI to Layer 3 Domain

'Traditional' L3Out on the BL Nodes

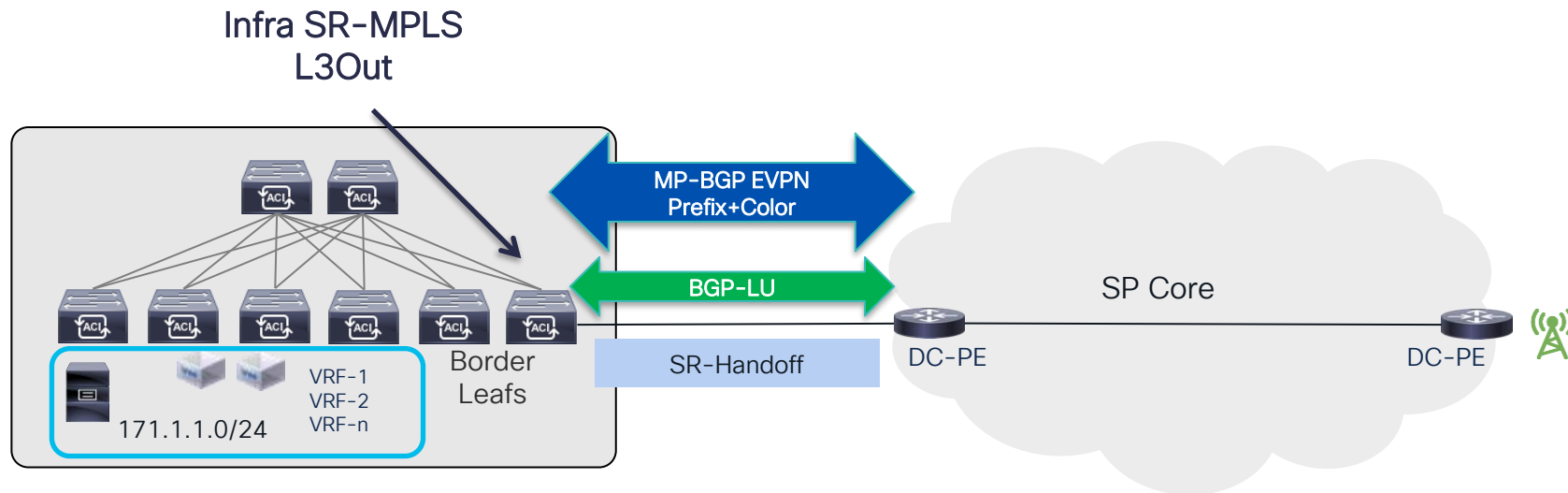


* From release 6.0(3), 800 L3Outs 6.0(2) and earlier

Connecting ACI to Layer 3 Domain

'SR-MPLS Handoff'

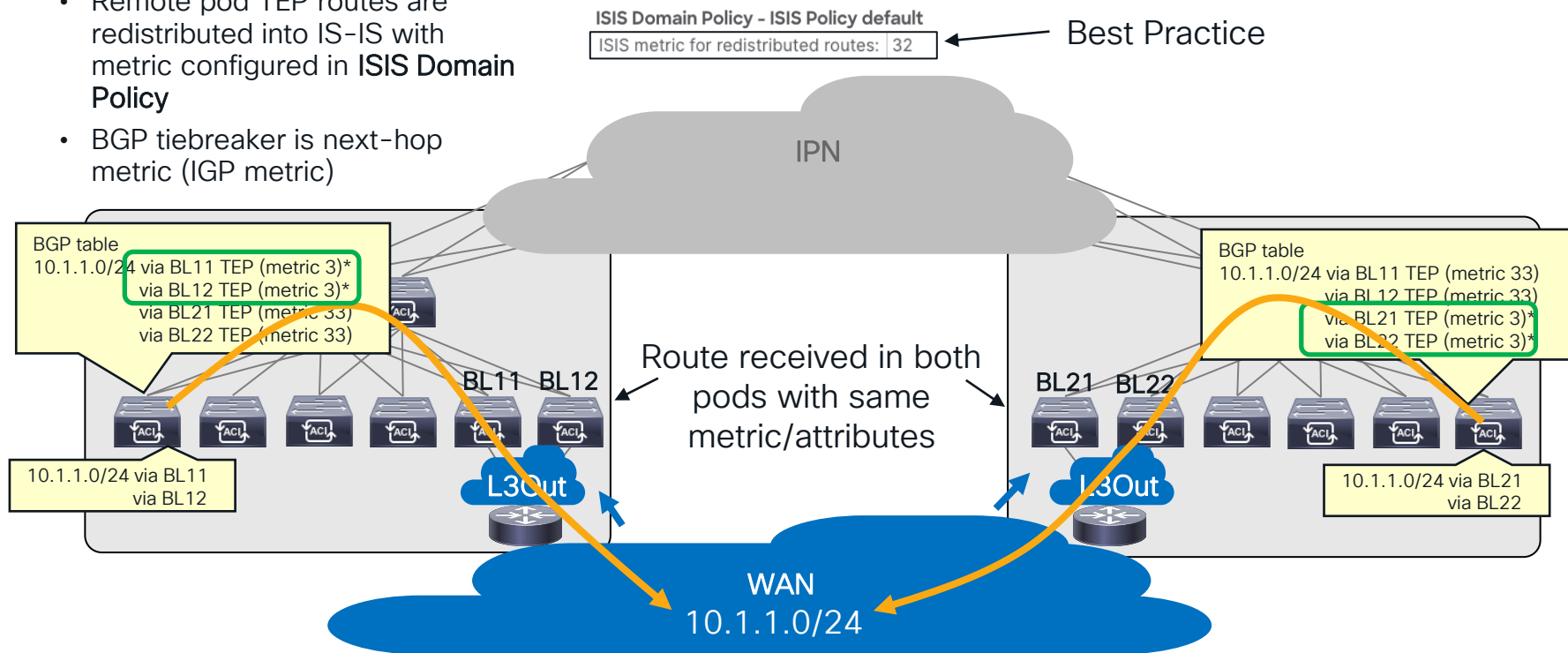
- Border Leafs connect to PE router in SP core
- Single BGP EVPN session for all VRFs
- ACI BL is advertising EVPN type-5 routes with BGP color community



Connecting to the External L3 Domain

Local L3Outs preferred over L3Outs in remote pods

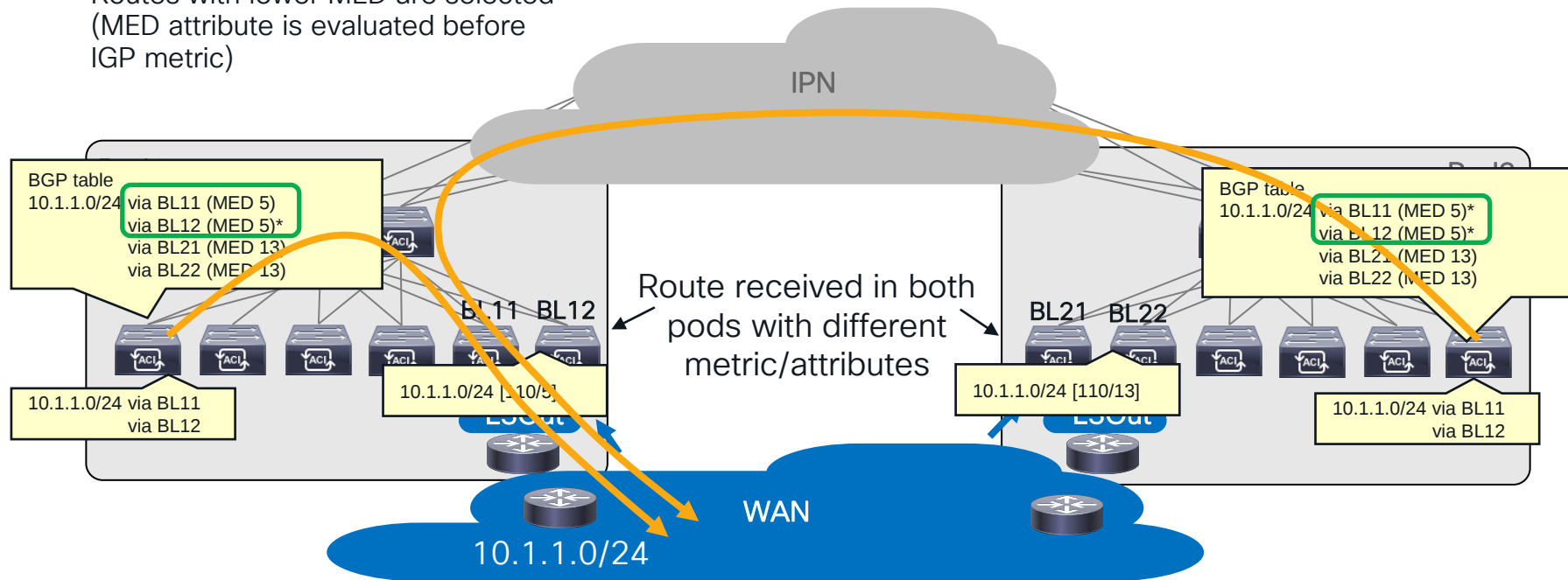
- Remote pod TEP routes are redistributed into IS-IS with metric configured in **ISIS Domain Policy**
- BGP tiebreaker is next-hop metric (IGP metric)



Connecting to the External L3 Domain

Remote pod L3Out may be used if it has a better external metric

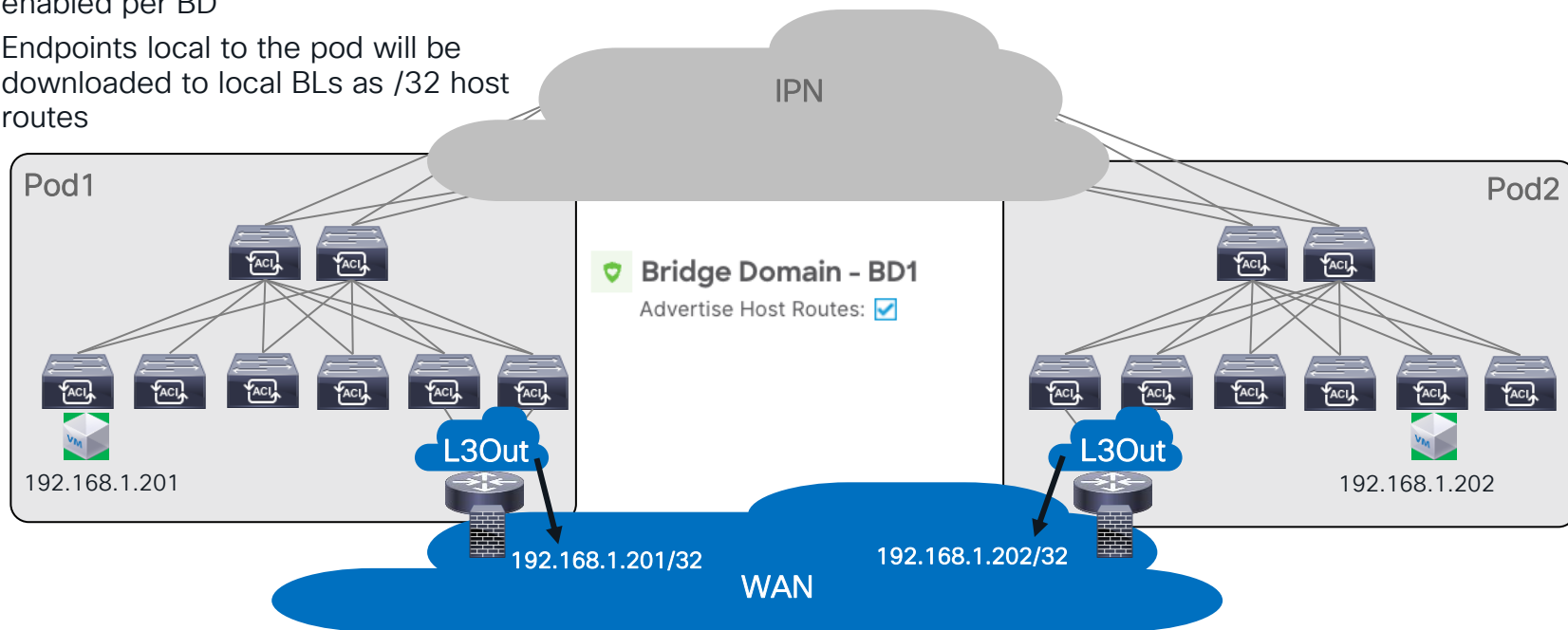
- BGP MED is set to OSPF metric when redistributed into fabric
- Routes with lower MED are selected (MED attribute is evaluated before IGP metric)



Connecting to the External L3 Domain

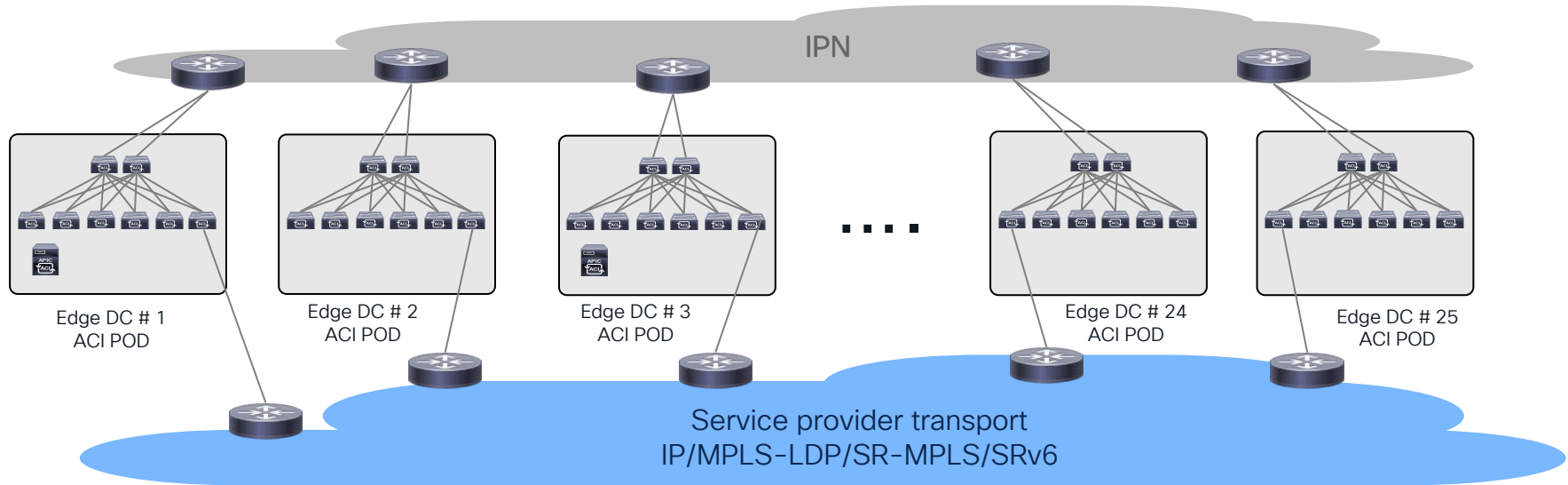
Influencing inbound path: Host route advertisement

- Host route advertisement can be enabled per BD
- Endpoints local to the pod will be downloaded to local BLs as /32 host routes

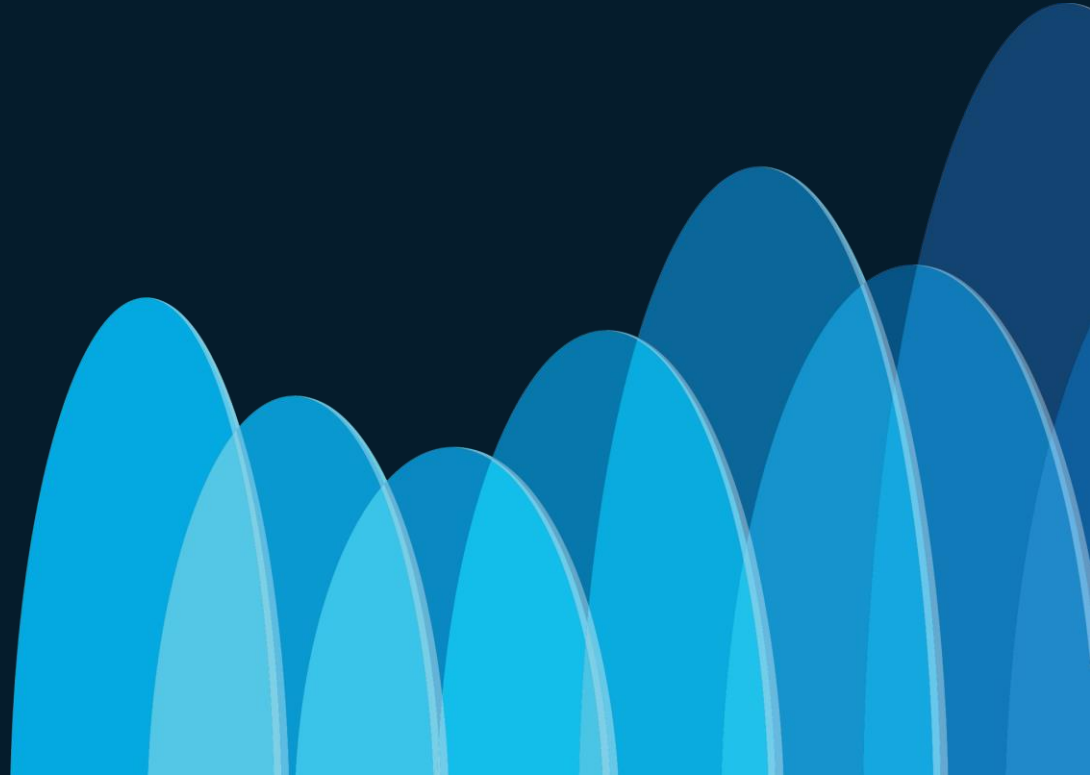


Edge DCs with Multi-Pod architecture

- APIC controllers are needed only in some Pods
- Communication across Pods is typically through SR-MPLS L3out
- 25 Pods per fabric is supported starting 6.0(1) release
- Leaf scale per fabric remains same. 2 Spines per Pod is supported
- Latency requirement remains same - 50 msec RTT requirement across APIC clusters and between switches and APIC
- No need to enable PIM-Bidir in IPN if L2 extension across Pod is not required



Network Services Integration



ACI Multi-Pod

Design options

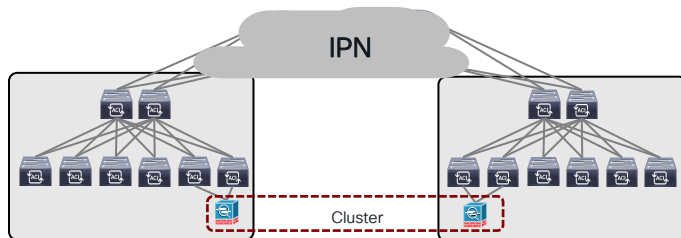


For Your
Reference

Typical options for an
Active/Active DC use case



- Active and Standby pair deployed across Pods
- No issues with asymmetric flows



- Active/Active FW cluster nodes stretched across Sites (single logical FW)
- Requires the ability of discovering the same MAC/IP info in separate sites at the same time
- Supported from ACI release 3.2(4d) with the use of Service-Graph with PBR



- Independent Active/Standby pairs deployed in separate Pods
- Use of Symmetric PBR to avoid the creation of asymmetric paths crossing different active FW nodes

ACI Multi-Pod: Active/Active cluster across pods

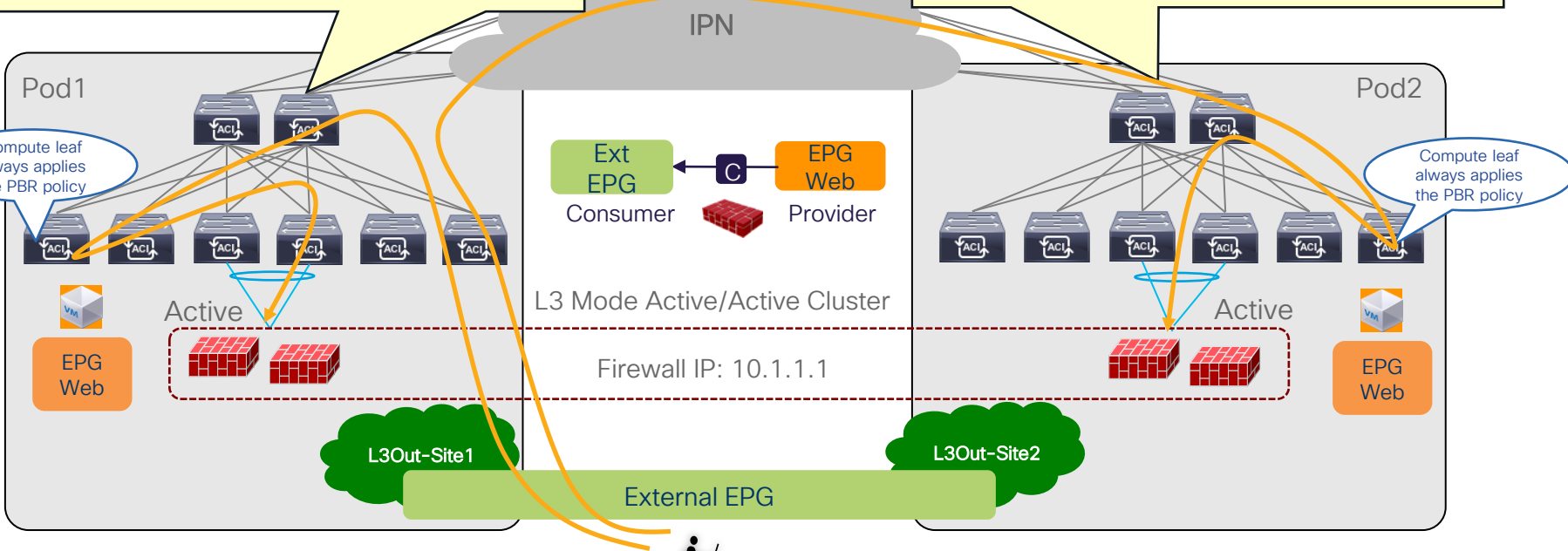
North-South Traffic Flow

Spines in Pod1

- 10.1.1.1 via Service Leaf in Pod1 (preferred)
- 10.1.1.1 via Pod2

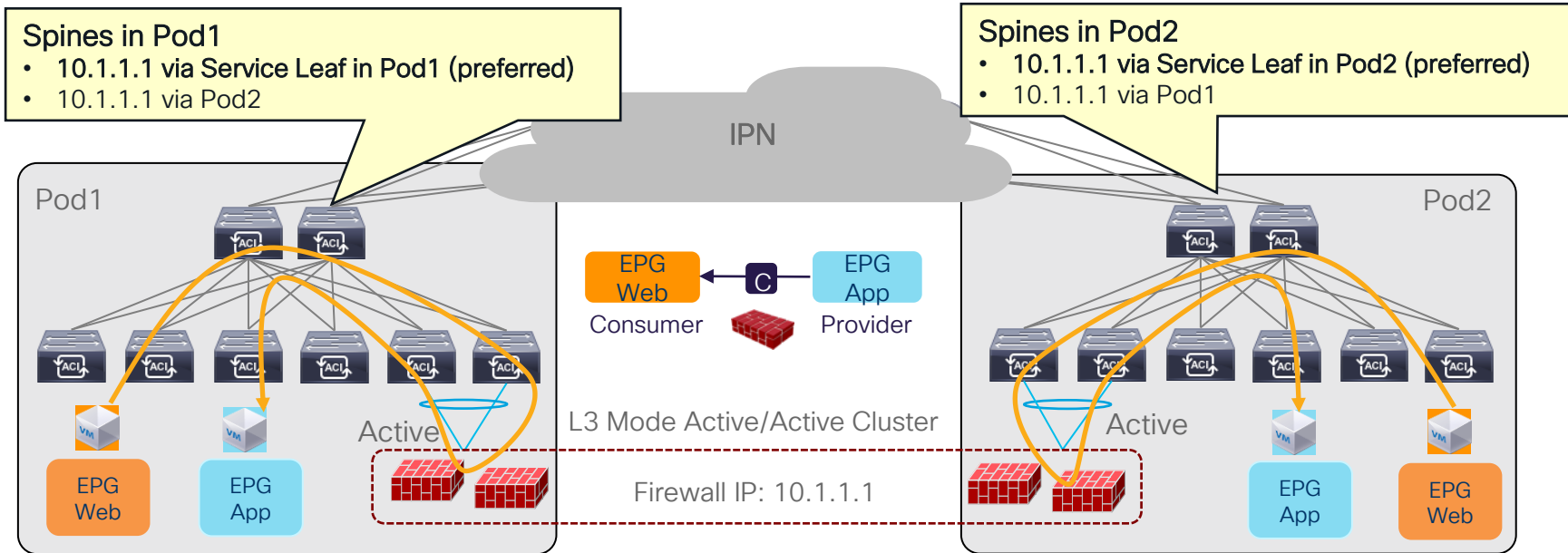
Spines in Pod2

- 10.1.1.1 via Service Leaf in Pod2 (preferred)
- 10.1.1.1 via Pod1



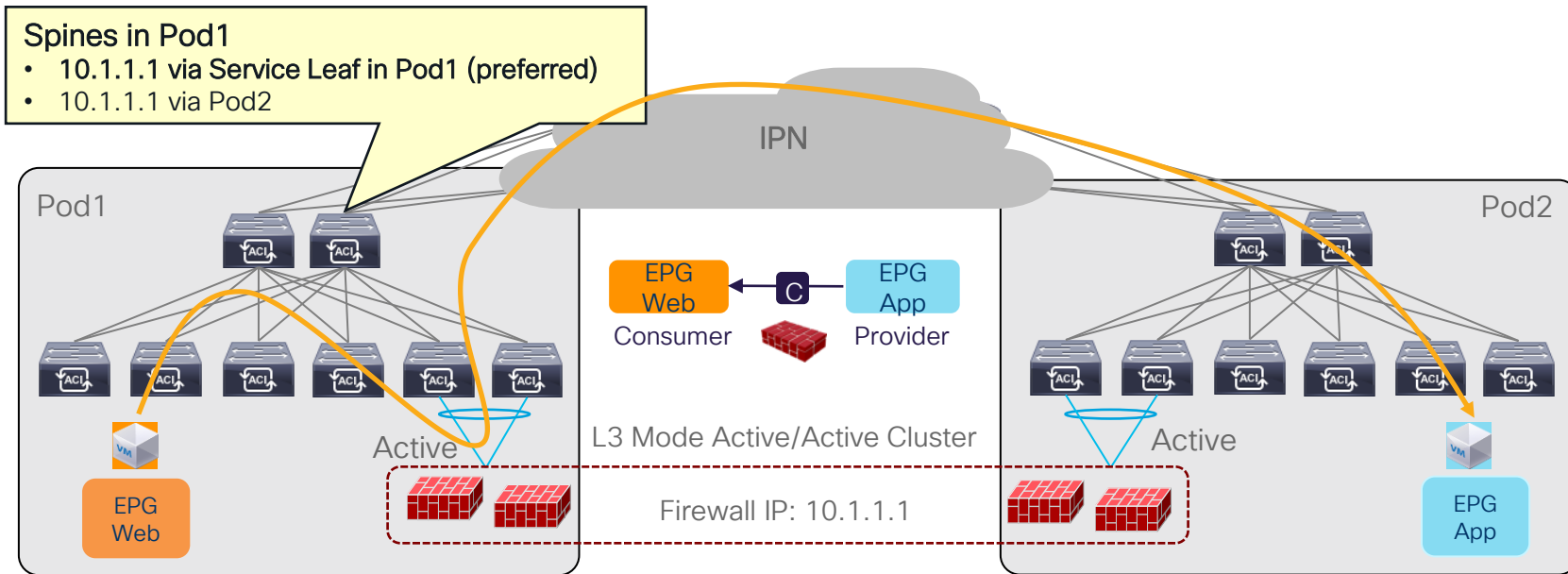
ACI Multi-Pod: Active/Active cluster across pods

East-West Traffic Flow (Intra-Pod)



ACI Multi-Pod: Active/Active cluster across pods

East-West Traffic Flow (Inter-Pod) incoming traffic



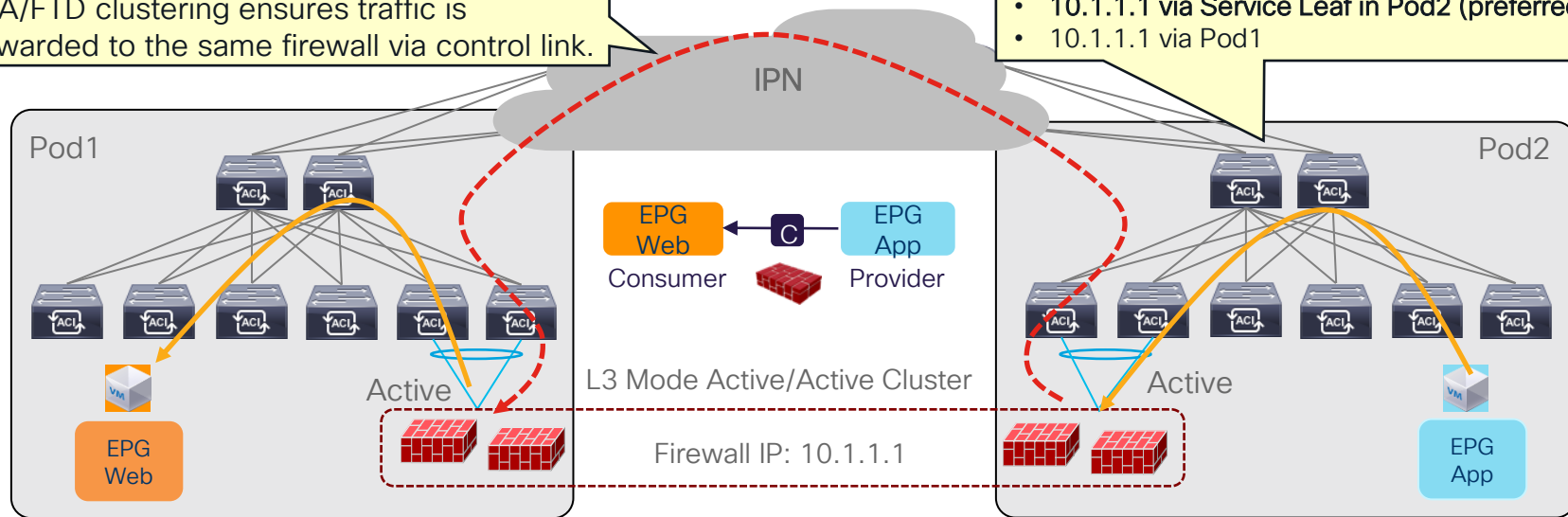
ACI Multi-Pod: Active/Active cluster across pods

East-West Traffic Flow (Inter-Pod) return traffic

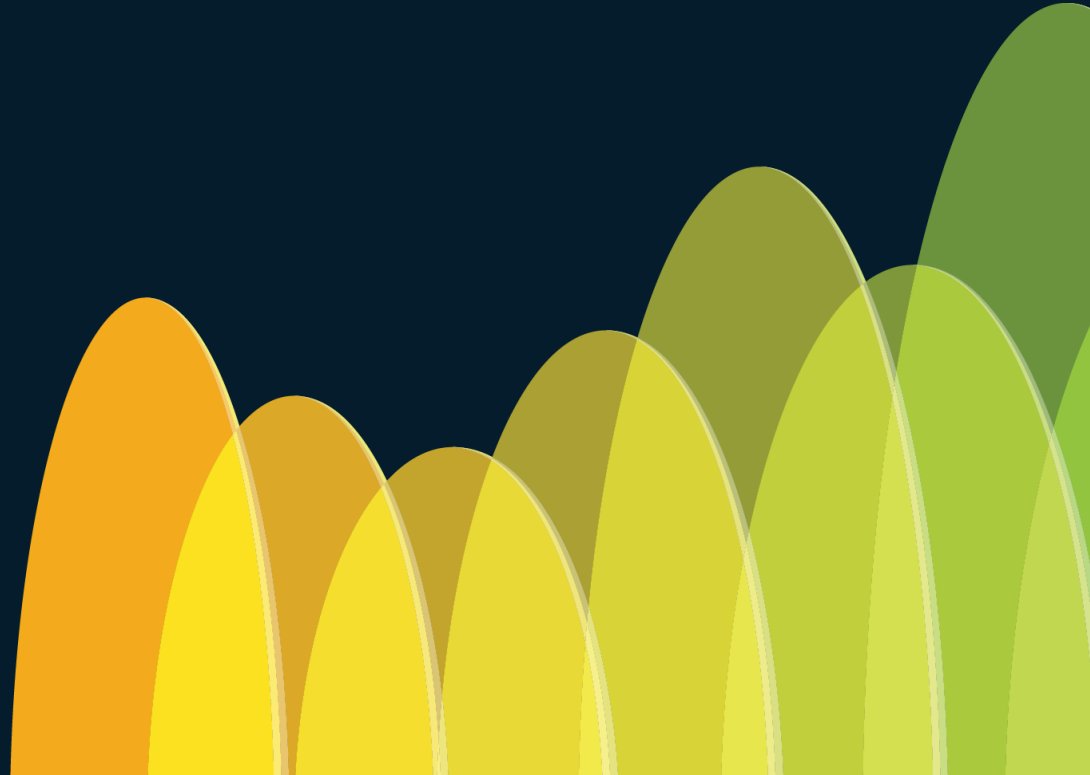
Even if asymmetric redirection happens, ASA/FTD clustering ensures traffic is forwarded to the same firewall via control link.

Spines in Pod2

- 10.1.1.1 via Service Leaf in Pod2 (preferred)
- 10.1.1.1 via Pod1

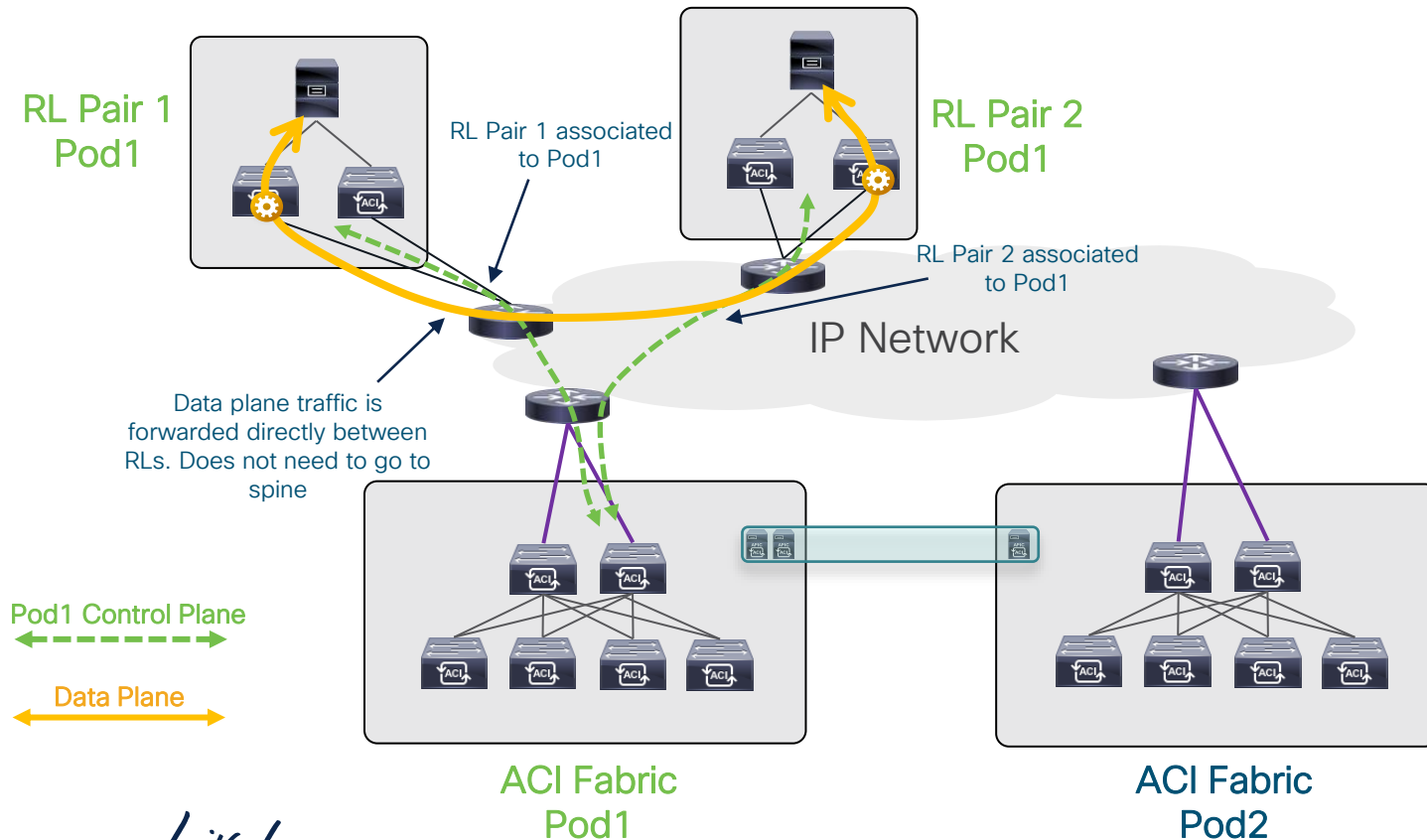


Multi-Pod with Remote Leaf



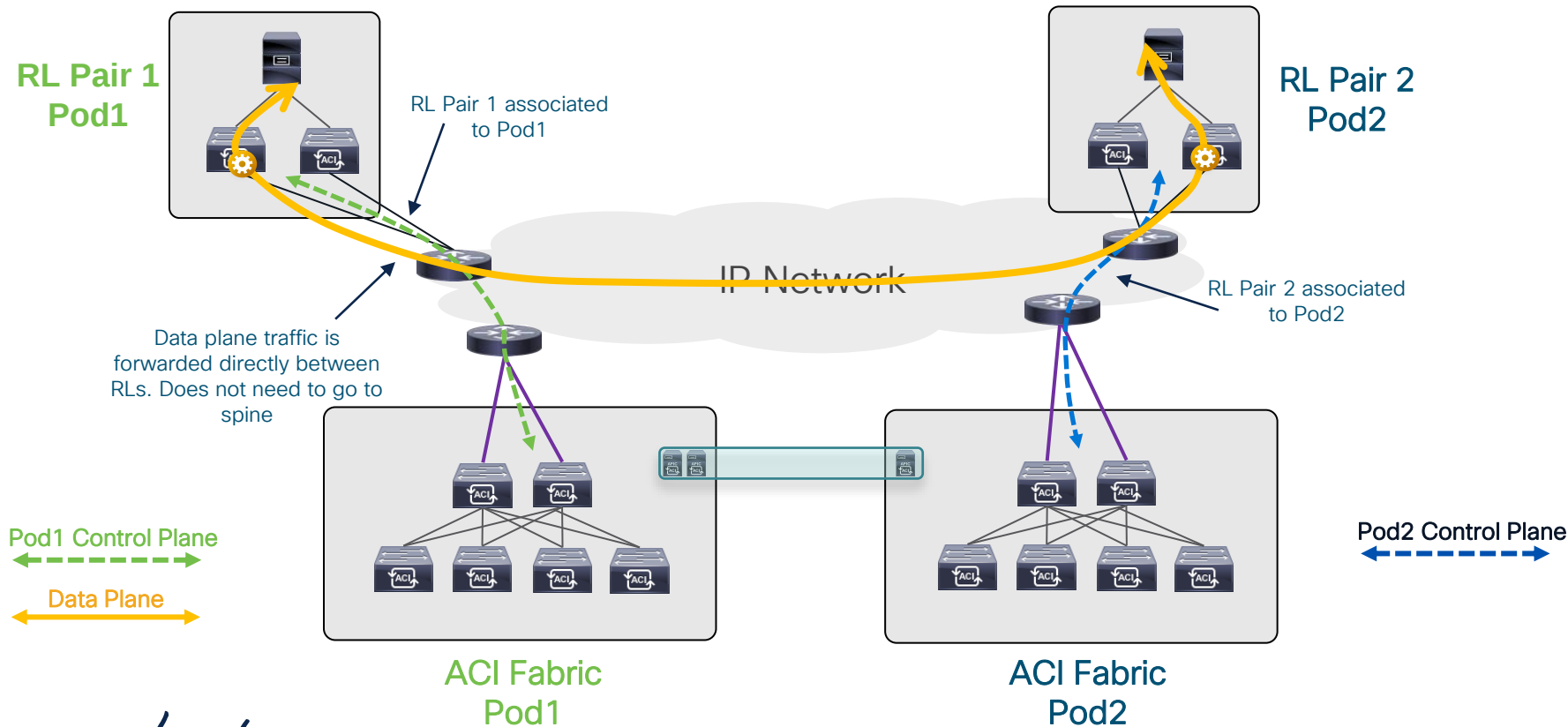
ACI Remote Leaf with Multi-Pod

Direct Forwarding between RL Pairs Part of the Same Pod



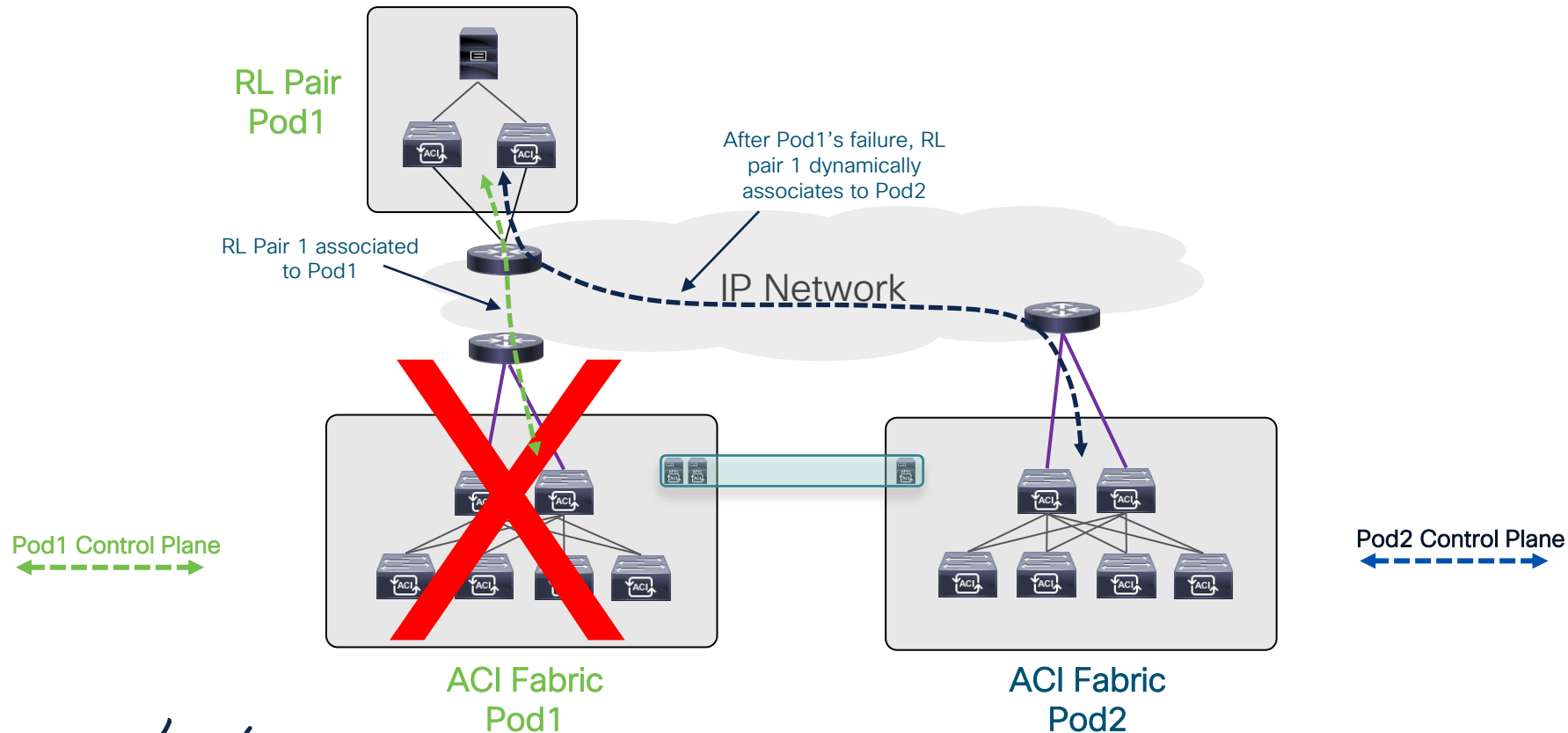
ACI Remote Leaf with Multi-Pod

Direct Forwarding between RL Pairs Part of Different Pods



ACI Remote Physical Leaf

RL Pair Resiliency in a Pod Failure Scenario



Useful Links

- ACI Multi-Pod White Paper
 - <https://www.cisco.com/c/en/us/solutions/collateral/data-center-virtualization/application-centric-infrastructure/white-paper-c11-737855.html>
- ACI Multi-Pod Configuration Paper
 - <https://www.cisco.com/c/en/us/solutions/collateral/data-center-virtualization/application-centric-infrastructure/white-paper-c11-739714.html>
- ACI Multi-Pod and Service Node Integration White Paper
 - <https://www.cisco.com/c/en/us/solutions/collateral/data-center-virtualization/application-centric-infrastructure/white-paper-c11-739571.html>
- ACI Remote Leaf Architecture White Paper
 - <https://www.cisco.com/c/en/us/solutions/collateral/data-center-virtualization/application-centric-infrastructure/white-paper-c11-740861.html>

Webex App

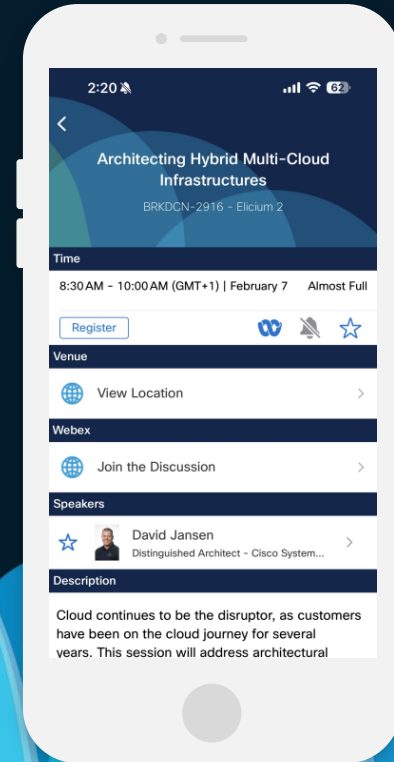
Questions?

Use the Webex app to chat with the speaker after the session

How

- 1 Find this session in the Cisco Events mobile app
- 2 Click “Join the Discussion”
- 3 Install the Webex app or go directly to the Webex space
- 4 Enter messages/questions in the Webex space

Webex spaces will be moderated by the speaker until February 28, 2025.



Fill Out Your Session Surveys



Participants who fill out a minimum of 4 session surveys and the overall event survey will get a unique Cisco Live t-shirt.

(from 11:30 on Thursday, while supplies last)



All surveys can be taken in the Cisco Events mobile app or by logging in to the Session Catalog and clicking the 'Participant Dashboard'



Content Catalog

Continue your education



- Visit the Cisco Showcase for related demos
- Book your one-on-one Meet the Engineer meeting
- Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs
- Visit the On-Demand Library for more sessions at ciscolive.com/on-demand. Sessions from this event will be available from March 3.



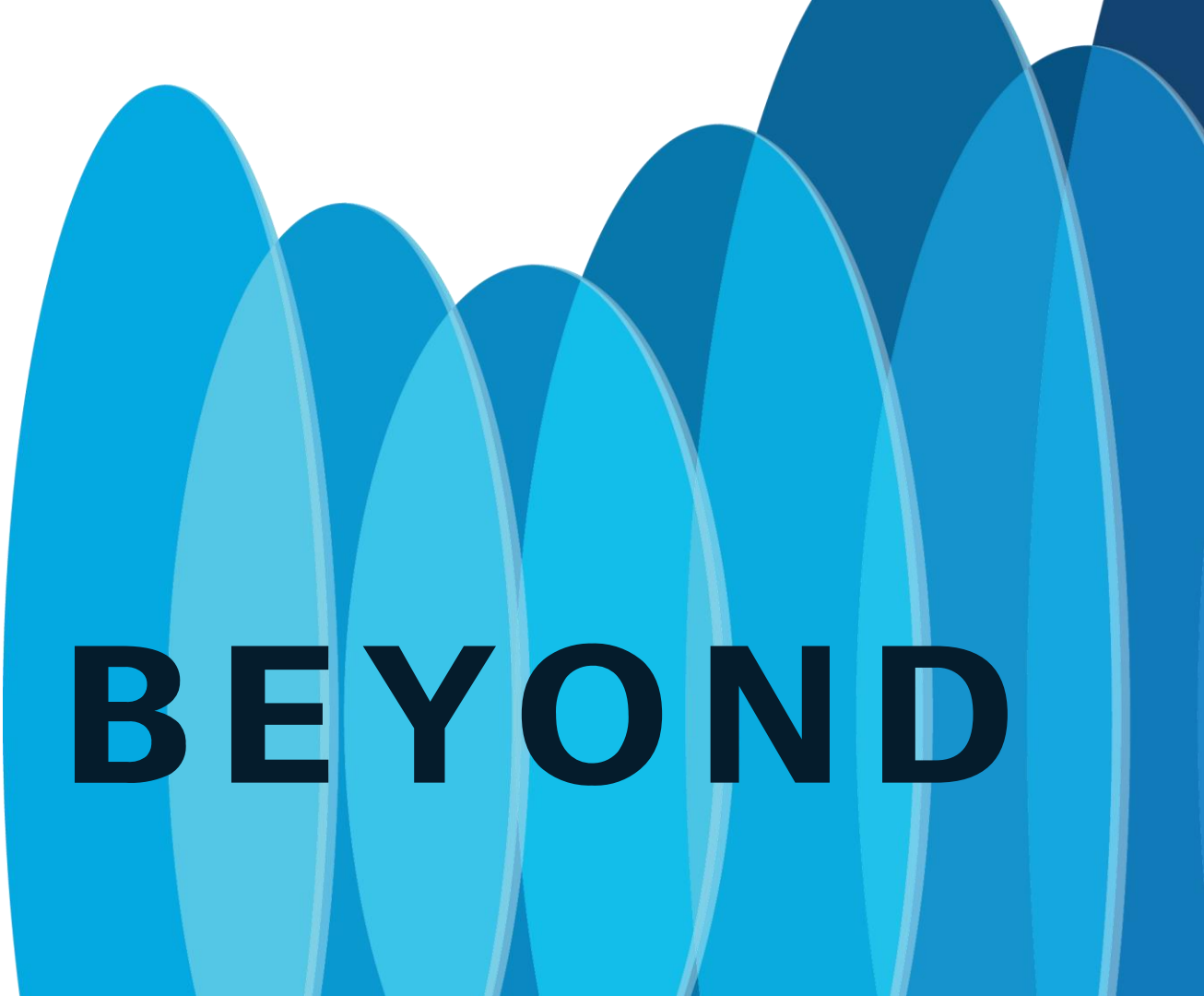
Thank you

CISCO *Live!*



CISCO *Live!*

GO BEYOND

The background of the slide features a series of overlapping, teardrop-shaped elements in various shades of blue, ranging from light sky blue to deep navy blue. These shapes are arranged in a way that creates a sense of depth and movement, resembling a stylized mountain range or a series of waves. The overall aesthetic is clean and modern.