cisco The bridge to possible

ACI Multi-Site Architecture and Deployment

Max Ardica, Distinguished Engineer
@maxardica

Hall of Fame Speaker cisco Live!

cisco Live!

BRKDCN-2980

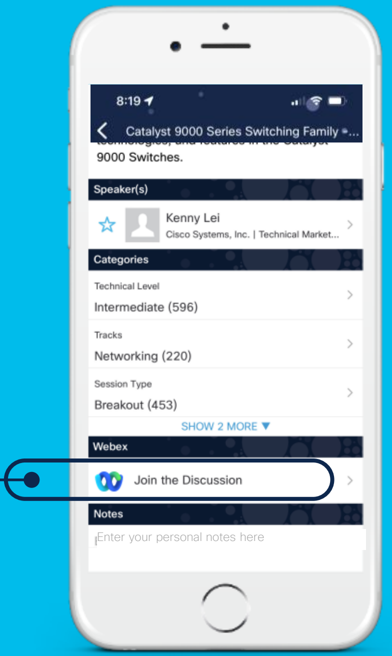
Cisco Webex App

Questions?

Use Cisco Webex App to chat
with the speaker after the session

How

1. Find this session in the Cisco Live 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



8:19

Catalyst 9000 Series Switching Family ...

9000 Switches.

Speaker(s)

Kenny Lei
Cisco Systems, Inc. | Technical Market...

Categories

Technical Level
Intermediate (596)

Tracks
Networking (220)

Session Type
Breakout (453)

SHOW 2 MORE ▼

Webex

Join the Discussion

Notes

Enter your personal notes here

Webex spaces will be moderated
until February 24, 2023.

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-Site architecture
- **Initial assumption:**
  - ✓ The audience already has a good knowledge of ACI main concepts (Tenant, BD, EPG, L2Out, L3Out, etc.)

Early Access. Yes, please.

Cisco U.

Tech learning, shaped to you.

CX Cisco **Customer Experience**

Accelerate your data center transformation with Cisco CX

Gain faster results by quickly transforming and boosting the agility and automation of your data center operations with CX expertise – so you can focus on the outcomes that matter most

Visit CX Data Center to learn more

cisco Live!

How we can help

Cisco Application Centric Infrastructure Services

Strategy and Solution Discovery
Discover and document business and technical requirements. Drive automation and infrastructure as code approach

Optimization and Monitoring Support
Unleash the full power of data center solutions through advanced network management

Assessments
Assess the readiness of your data center environment

Knowledge Transfer
Receive consulting support to drive your automation and DevOps priorities

Design, Implement, Migrate
Accelerate success throughout your network lifecycle with automation, micro-segmentation, and security

Learning | Certifications
Empower your workforce with efficiency and innovation

Optimize your network

Use and adopt

Protect your business

Accelerate multicloud

You don't have to do it alone.
For more insight, visit the Cisco CX stand in the World of Solutions for Lightning Talks and Demos

Data Center

ACI Technologies

Take a deep dive into ACI technologies, architecture and troubleshooting.

START

Feb 6 | 08:30
TECDCN-2840
Next Generation ACI Data Center Architecture, Deployment and Operations

Feb 7 | 08:30
BRKDCN-1601
Introduction to ACI

Feb 7 | 11:30
BRKDCN-2906
Introduction to Infrastructure as Code for ACI with Ansible and Terraform

Feb 7 | 14:00
BRKDCN-1688
How to operate your Nexus and ACI networks from the Cloud with Nexus Cloud

Feb 7 | 17:00
BRKDCN-2910
Why You Shouldn't Fear Upgrading Your ACI Fabric - The Handbook!

Feb 8 | 10:30
BRKDCN-2673
Nexus-as-Code - Kickstart your automation with ACI

Feb 8 | 12:00
BRKDCN-2949
Cisco ACI Multi-Pod Design and Deployment

Feb 8 | 14:30
BRKDCN-2980
ACI Multi-Site Architecture and Deployment

Feb 9 | 08:30
BRKDCN-2950
Nexus Cloud: How to manage your Nexus Data Center from the cloud

Feb 9 | 10:45
BRKDCN-3900
A Network Engineer's Blueprint for ACI Forwarding

Feb 9 | 13:45
BRKDCN-3982
ACI L4-L7 Policy-Based Redirect (PBR) Deep Dive and Tips

Feb 9 | 15:45
BRKDCN-3612
Secure Firewall in ACI

FINISH

Feb 10 | 11:00
BRKDCN-2969
Managing your data center network with ServiceNow

If you are unable to attend a live session, you can watch it On Demand after the event

cisco Live!

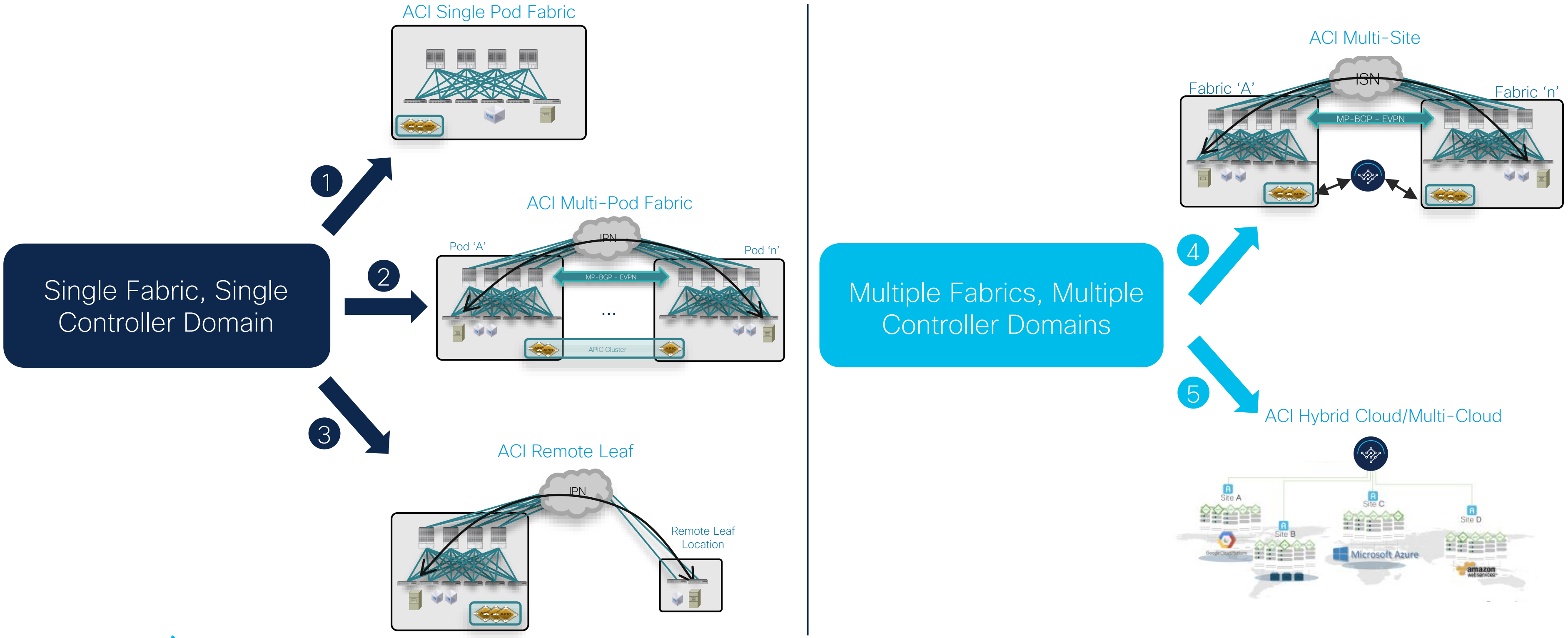
Agenda

- Introduction
- Nexus Dashboard Orchestrator (NDO) Architecture
- Provisioning Policies on NDO
- Inter-Site Connectivity Deployment Considerations
- ACI Multi-Site Control and Data Plane
- Connecting to the External L3 Domain
- Network Services Integration (Stretch Goal)

Introduction

ACI Architectural Options

Fabric and Policy Domain Evolution



Single Fabric, Single Controller Domain

1 ACI Single Pod Fabric

2 ACI Multi-Pod Fabric

3 ACI Remote Leaf

Multiple Fabrics, Multiple Controller Domains

4 ACI Multi-Site

5 ACI Hybrid Cloud/Multi-Cloud

Multi-Pod or Multi-Site?

That is the question…

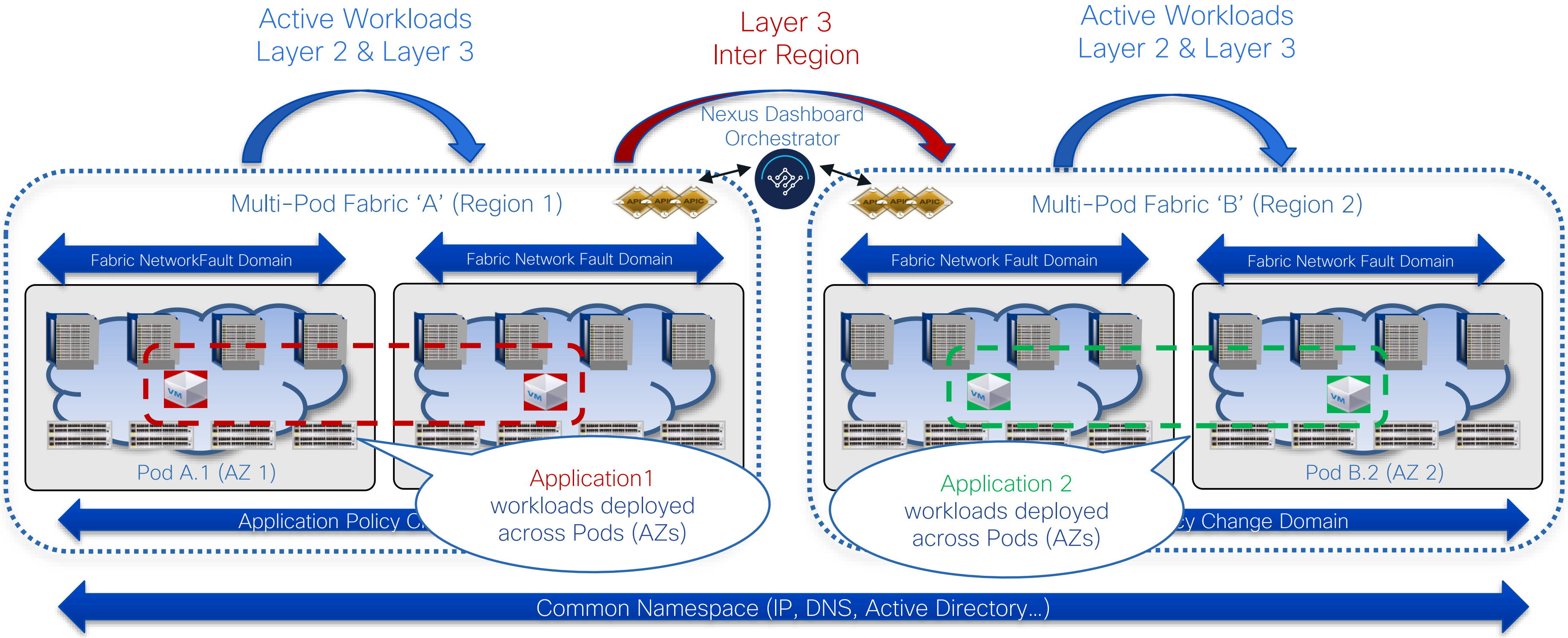
cisco Live!

And the answer is…

BOTH!

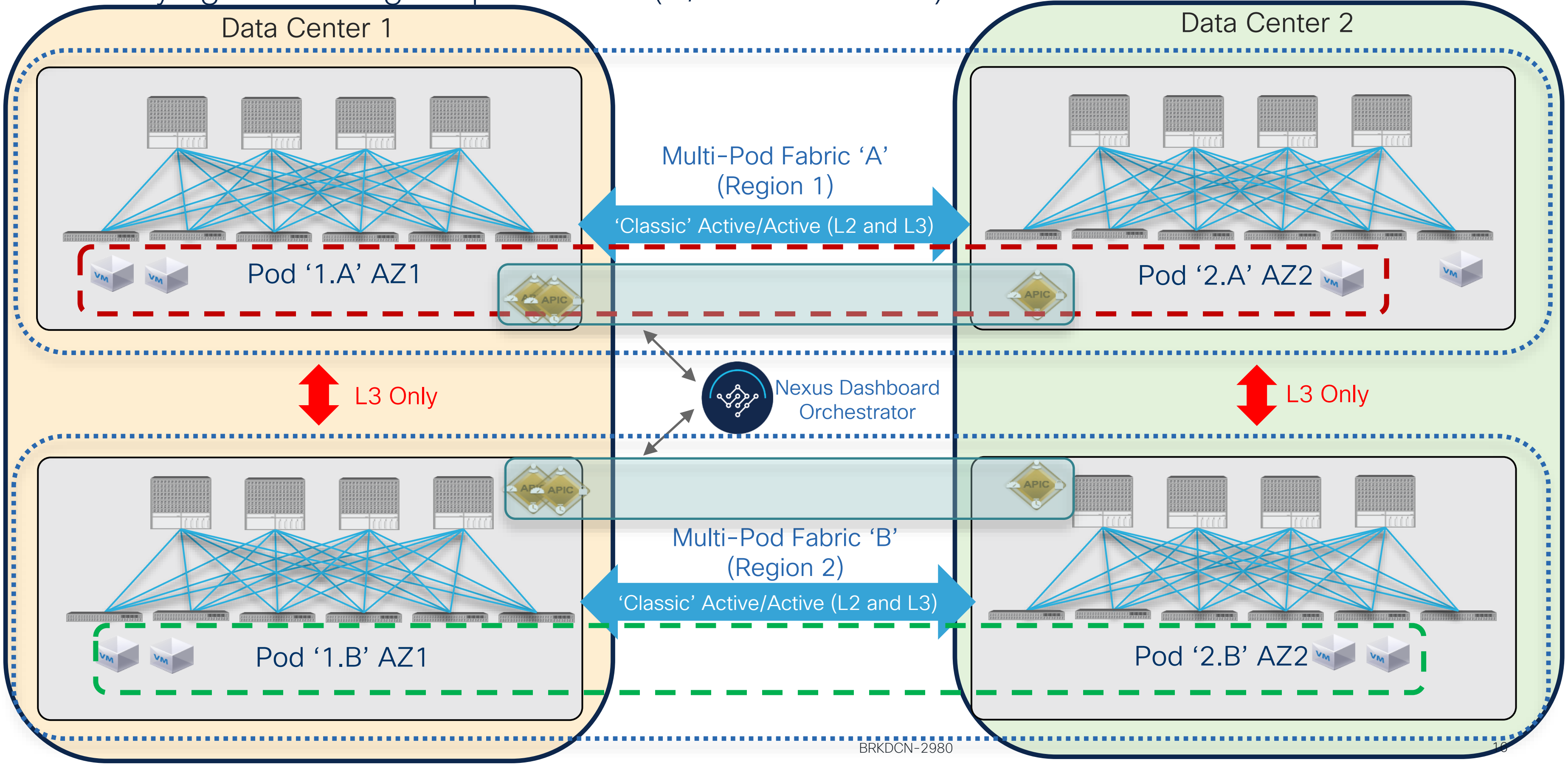
Systems View (How do these things relate)

Change and Network Fault Domain Isolation



Multi-Pod + Multi-Site

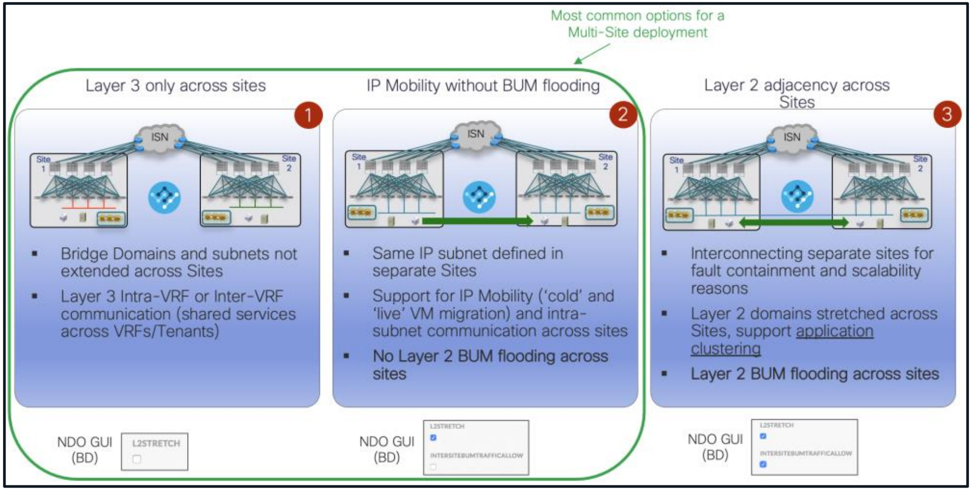
Satisfying Conflicting Requirements (A/A DCs and DR)



But wait! Couldn’t I deploy Multi-Site also to handle more typical Multi-Pod use cases?

Multi-Site for Active/Active Application Deployments?

Multi-Site for Active/Active Application Deployments



Most common options for a Multi-Site deployment

Layer 3 only across sites

- Bridge Domains and subnets not extended across Sites
- Layer 3 Intra-VRF or Inter-VRF communication (shared services across VRFs/Tenants)

NDO GUI (BD)

L2STRETCH □

IP Mobility without BUM flooding

- Same IP subnet defined in separate Sites
- Support for IP Mobility ('cold' and 'live' VM migration) and intra-subnet communication across sites
- No Layer 2 BUM flooding across sites

NDO GUI (BD)

L2STRETCH ☑

INTERSITEBUMTRAFFICALLOW □

Layer 2 adjacency across Sites

- Interconnecting separate sites for fault containment and scalability reasons
- Layer 2 domains stretched across Sites, support application clustering
- Layer 2 BUM flooding across sites

NDO GUI (BD)

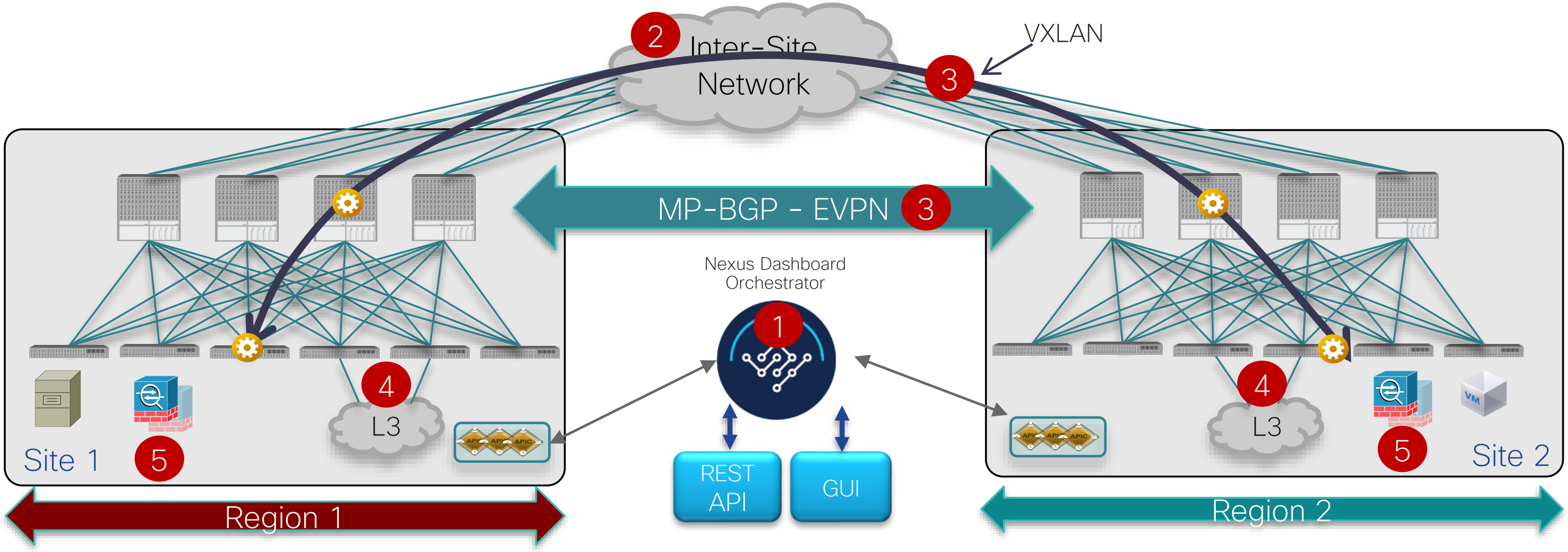
L2STRETCH ☑

INTERSITEBUMTRAFFICALLOW ☑

- ACI Multi-Site allows to extend connectivity and policies between separate APIC domains
  - Layer 3 only across sites
  - Layer 2 with and without BUM flooding
- Keep in mind some specific considerations before deploying Multi-Site for "classic" Active/Active application deployments (i.e. same application components deployed across sites)
  - Loss of change and network fault domain isolation across separate ACI domains
  - Creation of separate VMM domains by design (loss of intra-cluster functionalities like DRS, vSphere FT/HA, …)
  - Specific service node insertion deployment considerations (use of separate service nodes per fabric, limited support for service nodes clustering across sites, no support for vzAny + PBR, …)

ACI Multi-Site

The Ideal Architecture for "Loosely Coupled" DCs



- Separate ACI Fabrics with independent APIC clusters
- No latency limitation between Fabrics
- ACI Multi-Site Orchestrator pushes cross-fabric configuration to multiple APIC clusters providing scoping of all configuration changes

- MP-BGP EVPN control plane between sites
- Data Plane VXLAN encapsulation across sites
- End-to-end policy definition and enforcement

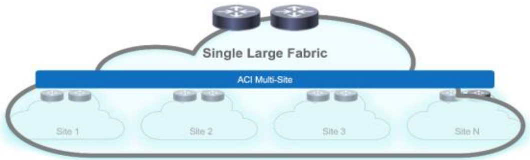
Want to know how to provision Multi-Pod and Multi-Site from scratch? Come to BRKDCN-2919 (Thu @ 8.30 am)

ACI Multi-Site Architecture

Most Common Use Cases

- Compartmentation/Scale

  Building Multiple Fabrics inside a single Data Center

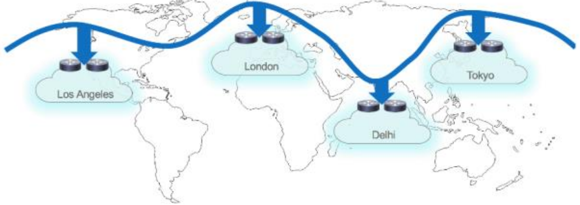


Optimized and controlled L2/L3 connectivity (including optimized/controlled BUM forwarding), scale out total number of leaf nodes (SP use case)

- Data Center Interconnect (DCI)

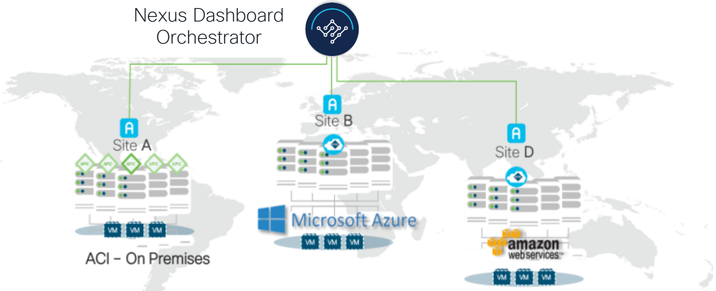
  Extend connectivity/policy between ‘loosely coupled’ DC sites

  Disaster Recovery and IP mobility use cases



- Hybrid-Cloud and Multi-Cloud

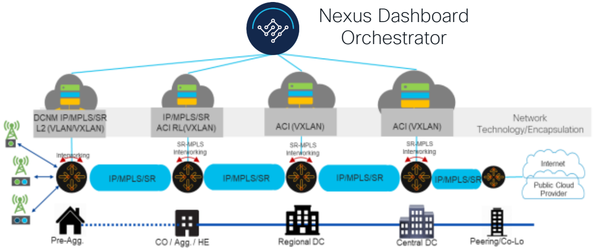
  Integration between on-prem and public clouds (AWS, Azure, GCP)



- SP 5G Telco DC/Cloud*

  Centralized DC Orchestration for “Autonomous Fabrics”

  Optional SR-MPLS/MPLS Handoff on Border Leaf nodes

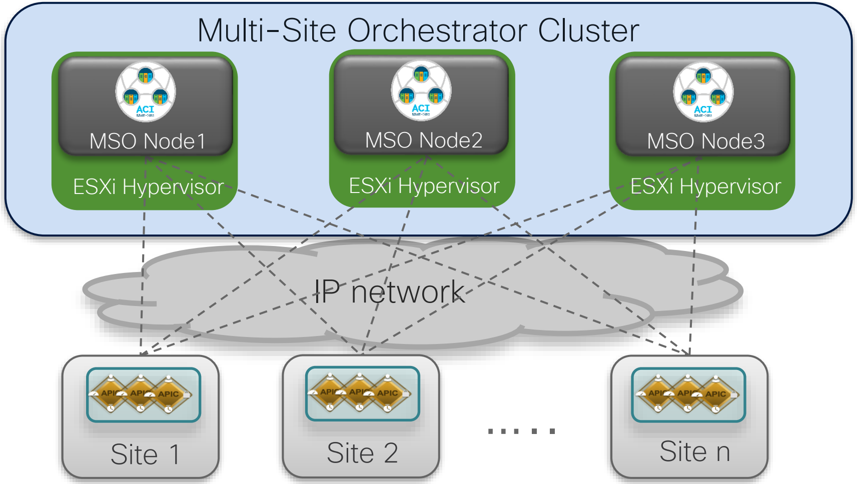


*May also apply to Enterprise deployments

Nexus Dashboard Orchestrator (NDO) Architecture

Original Multi-Site Orchestrator Option

VM Based MSO Cluster (OVA), Now EoL/EoS



- Supported from the beginning (MSO release 1.0(1))
- Each Cisco Multi-Site Orchestrator node is packaged in a VMware vSphere virtual appliance (OVA)
- For high availability, you should deploy each Cisco Multi-Site Orchestrator virtual machine on its own VMware ESXi host
- Requirements for MSO Release 1.2(x) and above:

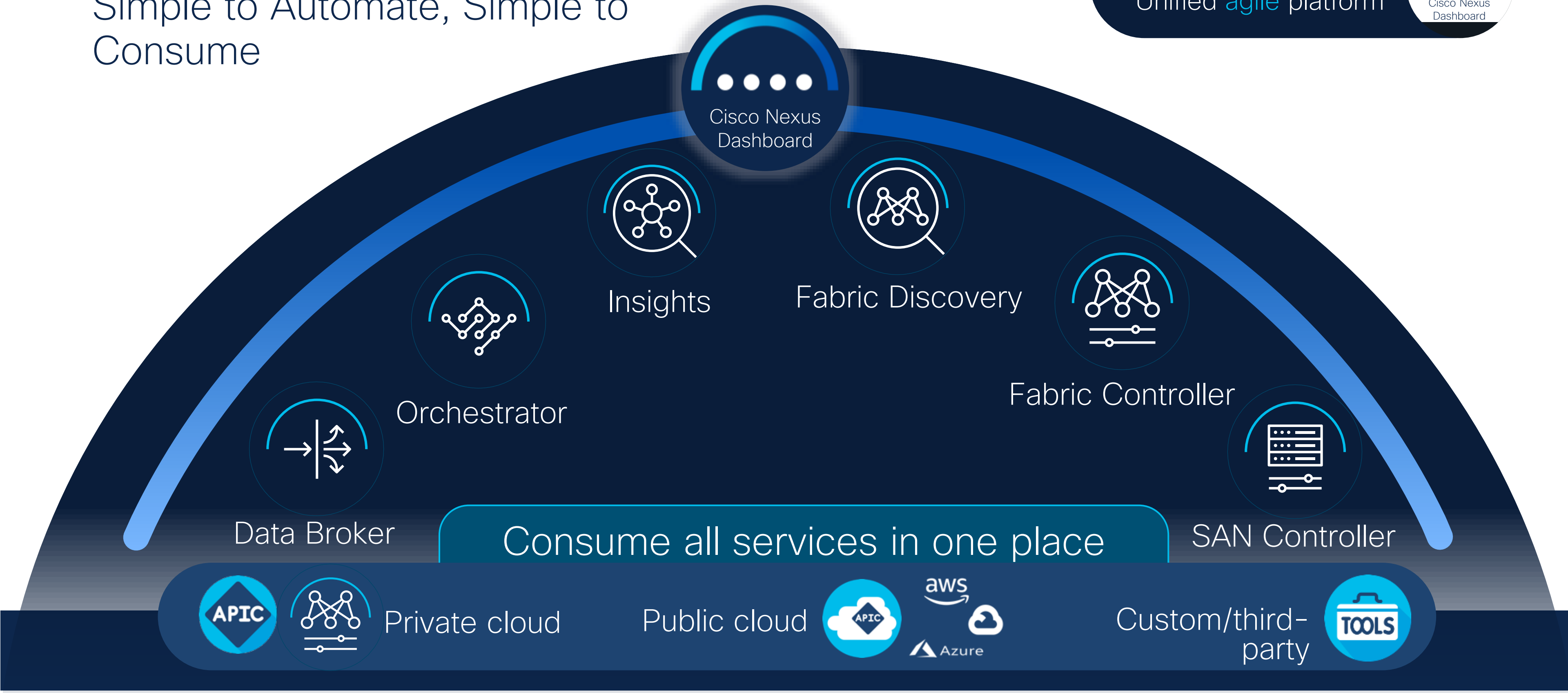
  VMware ESXi 6.0 or later

  Minimum of eight virtual CPUs (vCPUs), 48 Gbps of memory, and 100 GB of disk space
- MSO 3.1(1) last supported release with this form factor, now EoL/EoS

Cisco Nexus Dashboard

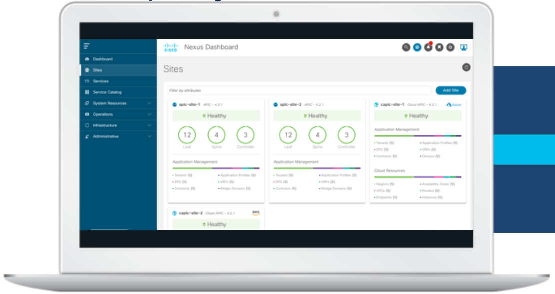
Simple to Automate, Simple to Consume

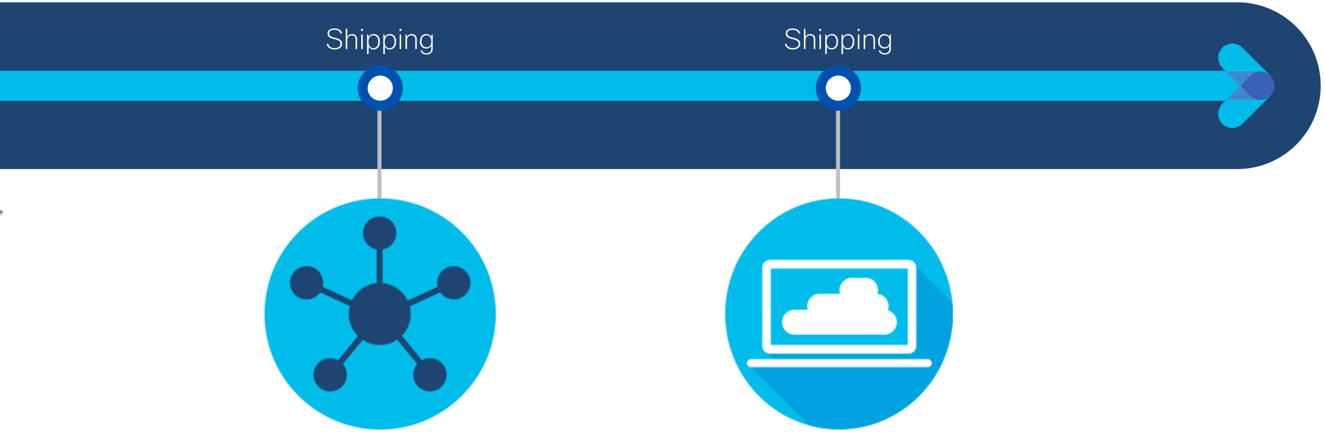
Powering automation
Unified agile platform



Cisco Nexus Dashboard

Deployment Evolution





Physical Cisco ND Platform Cluster

Virtual/Cloud Cisco ND Platform Cluster

ND virtual cluster supported on ESXi and KVM hypervisors

Spec: 16 vCPUs, 64Gb ram and 500Gb disk

ND cloud cluster supported for AWS and Azure

Cisco Multi-Site Orchestrator has become Cisco Nexus Dashboard Orchestrator

ACI
Multi-Site

Cisco Multi-Site Orchestrator

Cisco Nexus Dashboard Orchestrator

Up to release 3.1(1)

From release 3.2(1)

Nexus Dashboard Orchestrator

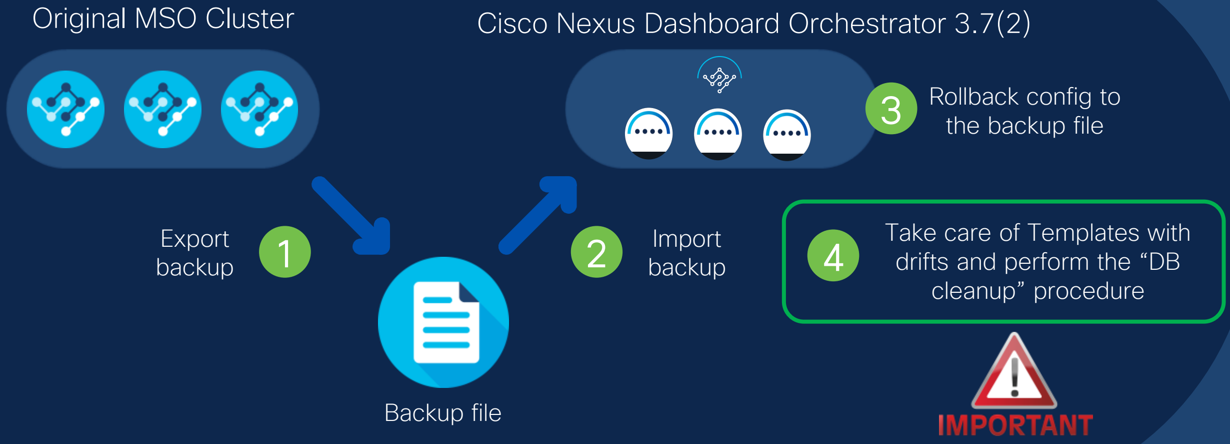
What NDO Release to Choose?

	Recommended Releases per Scenario	
	Current Release	Target Release
1	MSO/NDO 1.1.x to 3.7.1j	NDO 3.7(2) (Shipping)
2	NDO 3.7(2)	NDO 4.1(2) (Q2CY23)*
3	None - Greenfield	NDO 4.1(1) (Shipping)

*NDO 4.1(2) will support one-click GUI upgrade from NDO 3.x releases

Migrating the MSO Cluster to Cisco NDO

- All MSO releases are officially End-of-Life (EOL)
- Customer should (and must) migrate from MSO to NDO
- NDO 3.7(2) release is the recommended target release for this migration


Original MSO Cluster
Cisco Nexus Dashboard Orchestrator 3.7(2)
1
Export backup
Backup file
2
Import backup
3
Rollback config to the backup file
4
Take care of Templates with drifts and perform the “DB cleanup” procedure
IMPORTANT

Note

This procedure is only supported for SW upgrades but not for SW downgrades

https://www.cisco.com/c/en/us/td/docs/dcn/ndo/3x/deployment/cisco-nexus-dashboard-orchestrator-deployment-guide-371/ndo-deploy-migrate-37x.html

Provisioning Policies on NDO

Supporting Different Types of Policies

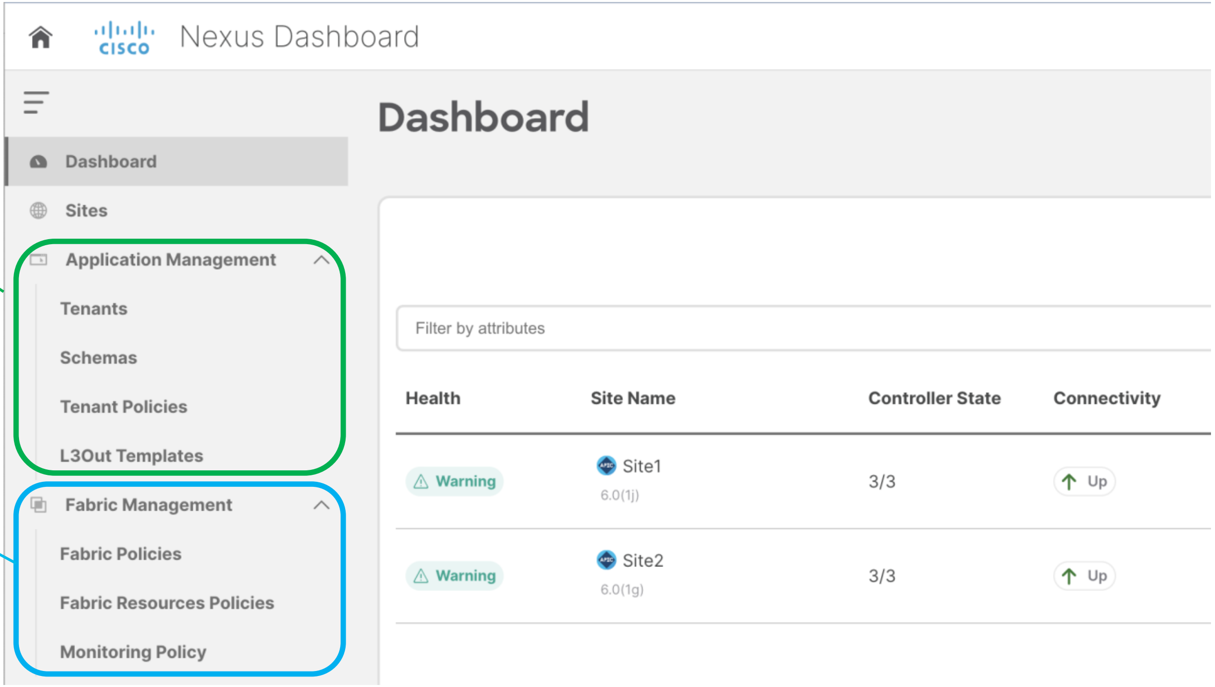
Cisco Nexus Dashboard Orchestrator 4.0(1)

Application Management Policies

Used to define tenant policies (Application Network Profiles, EPGs, BDs, VRFs, etc.)

Fabric Management Policies

Used to define fabric access policies, interface and monitoring policies



Dashboard

Health	Site Name	Controller State	Connectivity
Warning	Site1 6.0(1j)	3/3	Up
Warning	Site2 6.0(1g)	3/3	Up

Provisioning Policies on NDO

Multiple Template Types

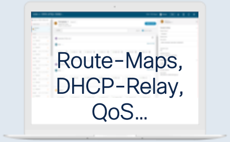
Cisco Nexus Dashboard Orchestrator 4.0(1)

Optimized in 4.0(1)

EPGs, BDs, VRFs, Contracts, etc.

Application Template

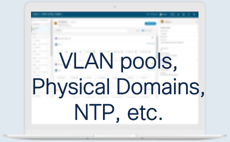
New in 4.0(1)



Route-Maps, DHCP-Relay, QoS…

Tenant Policy Template

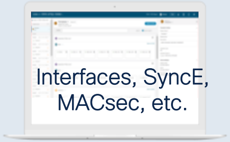
New in 4.0(1)



VLAN pools, Physical Domains, NTP, etc.

Fabric Policies Template

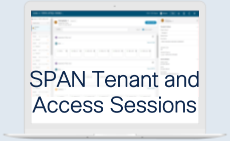
New in 4.0(1)



Interfaces, SyncE, MACsec, etc.

Fabric Resources Policies Template

New in 4.0(1)



SPAN Tenant and Access Sessions

Monitoring Policies Template

New in 4.1(1)

Full L3Out Configuration

L3Out Template

Benefits

Simplify | Single Pane of Glass

Provisioning Policies on NDO

Why do we “Templatize” the Configuration?

Features

NDO 3.4(1) Template versioning and rollback	NDO 3.4(1) Template deployment plan visibility	NDO 3.4(1) Change control workflow	NDO 3.4(1) Detach templates from Sites	NDO 3.6(1) Configuration drift reconciliation workflow
Support rollback of template from newer to older version-id Label a template as Golden	Shows preview of what NDO is going to provisioning to each site	New personas for management and provisioning of configuration	Configuration is not removed from the APIC/NDFC domains	NDO workflow that synchronizes and merges any config changes made in APIC or NDFC domains
Granular roll back of templates specific configuration	Better visibility to reduce errors and seize the impact of a template’s deployment	More structured deployments which enables increased flexibility	Ease of use for migration	Simplify the understanding and reconciliation of config drifts between NDO and APIC/NDFC

Benefits

Provisioning Policies on NDO

Template-Level Operational Enhancements

For more information and demonstrations of all those NDO template-level operational enhancements:

- Template Versioning

  https://video.cisco.com/video/6277140235001

- Template Deployment Plan Visibility

  https://video.cisco.com/video/6277137504001

- Change Control Workflow

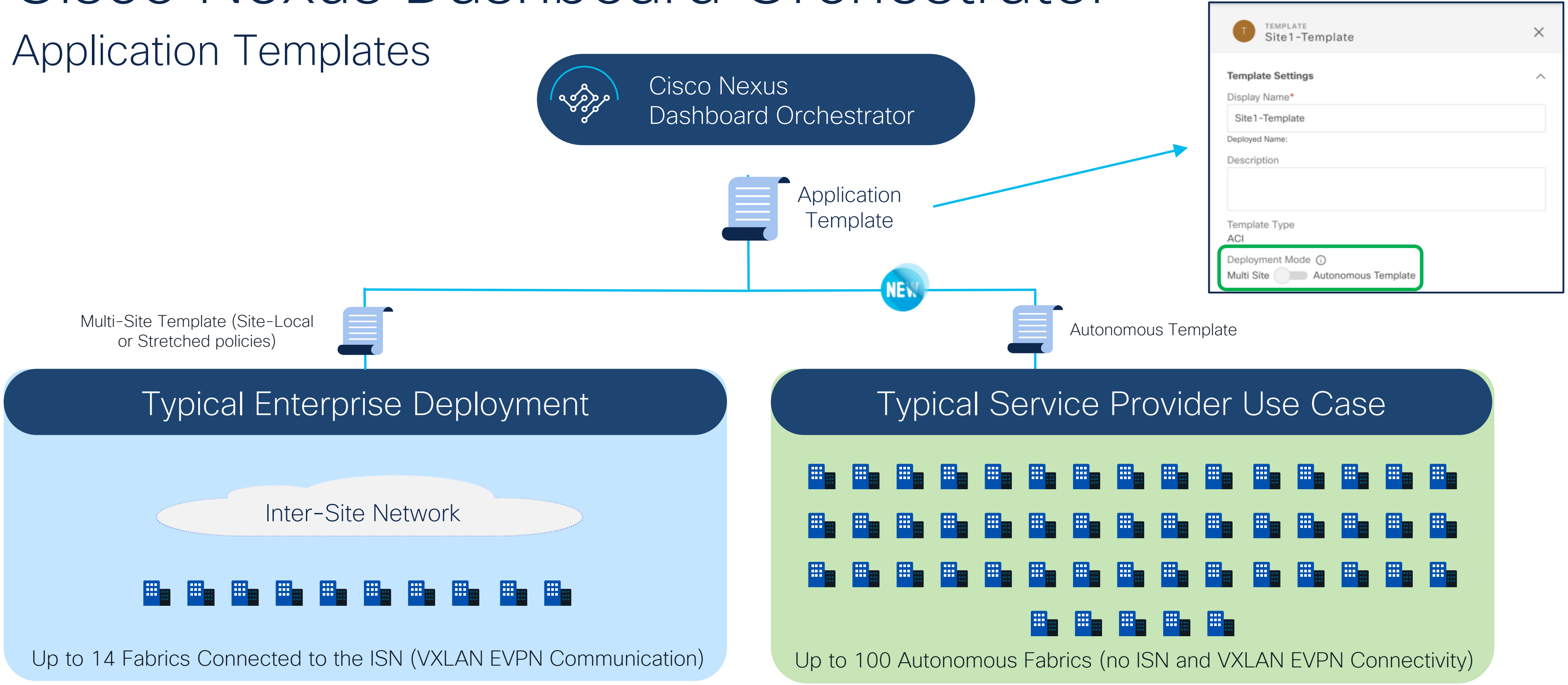
  https://video.cisco.com/video/6277140011001

Provisioning Application Policies on NDO

Cisco Nexus Dashboard Orchestrator

Application Templates

Cisco Nexus Dashboard Orchestrator 4.0(1)



Cisco Nexus Dashboard Orchestrator

Application Template

Multi-Site Template (Site-Local or Stretched policies)

Autonomous Template

Typical Enterprise Deployment

Inter-Site Network

Up to 14 Fabrics Connected to the ISN (VXLAN EVPN Communication)

Typical Service Provider Use Case

Up to 100 Autonomous Fabrics (no ISN and VXLAN EVPN Connectivity)

TEMPLATE
Site1-Template

Template Settings

Display Name*: Site1-Template

Deployed Name:

Description

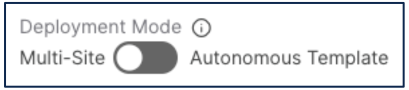
Template Type: ACI

Deployment Mode: Multi Site ○ Autonomous Template

Different Template Deployment Modes Can Apply to the Same Set of Fabrics

Application Templates

Multi-Site Templates



Deployment Mode ⓘ

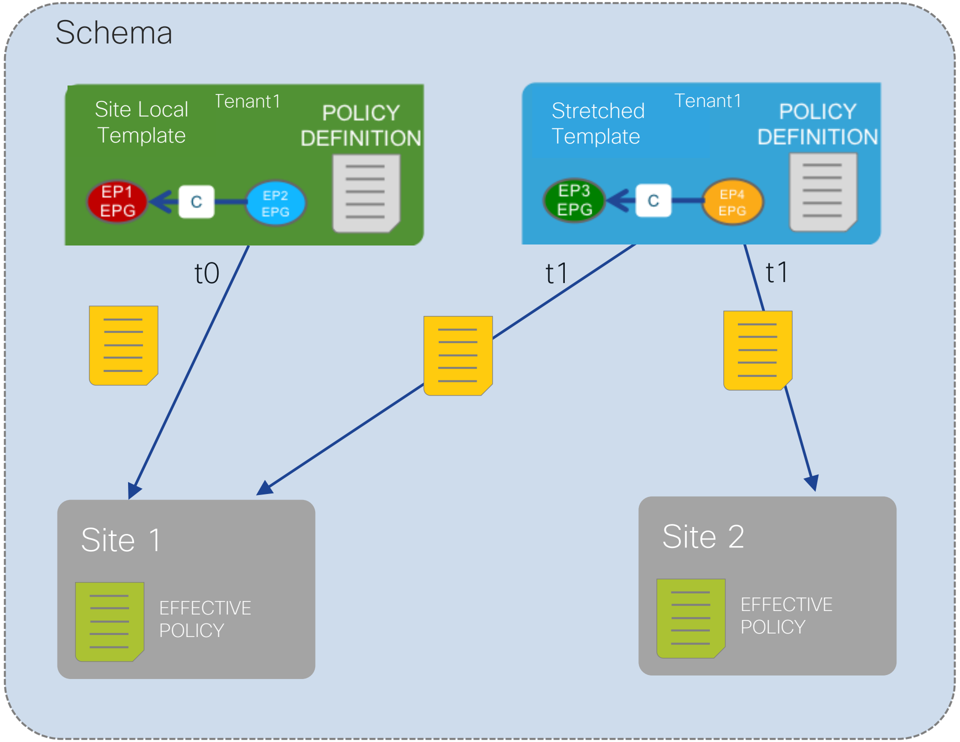
Multi-Site ○ Autonomous Template

- Application Template = ACI policy definition (ANP, EPGs, BDs, VRFs, etc.)
- Schema = container of Application Templates sharing a common use-case

  As a typical use case, a schema can (and should) be dedicated to a Tenant
- The template is the <u>atomic unit of change for policies</u>

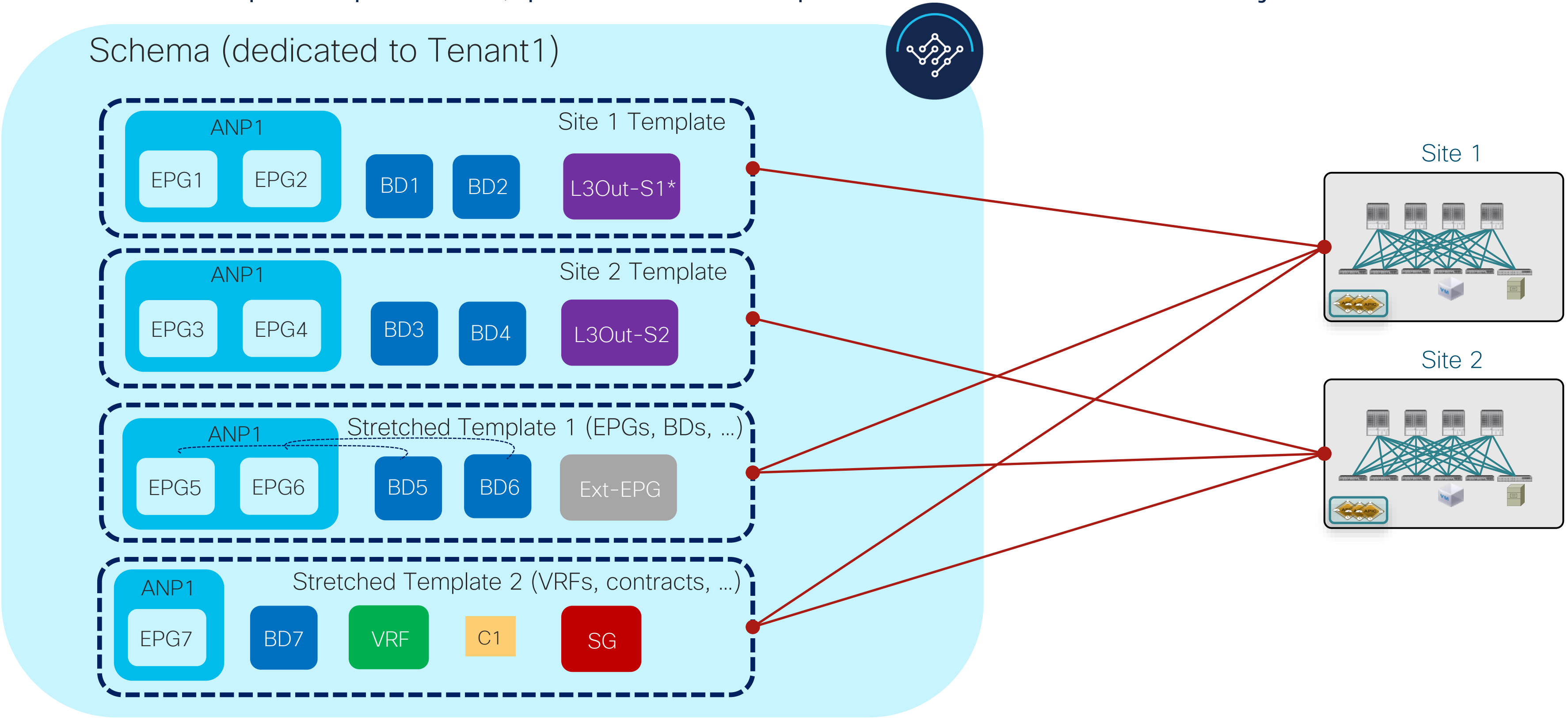
  A Multi-Site template associated to a single site can be pushed only to that site

  A Multi-Site template associated to multiple sites is concurrently pushed to all those sites



Best Practices for Multi-Site Templates

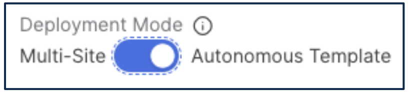
One Template per Site, plus Two Templates for “Stretched Objects”



*L3Out defined in a separate “L3Out Template” from NDO 4.1(1)

Application Templates

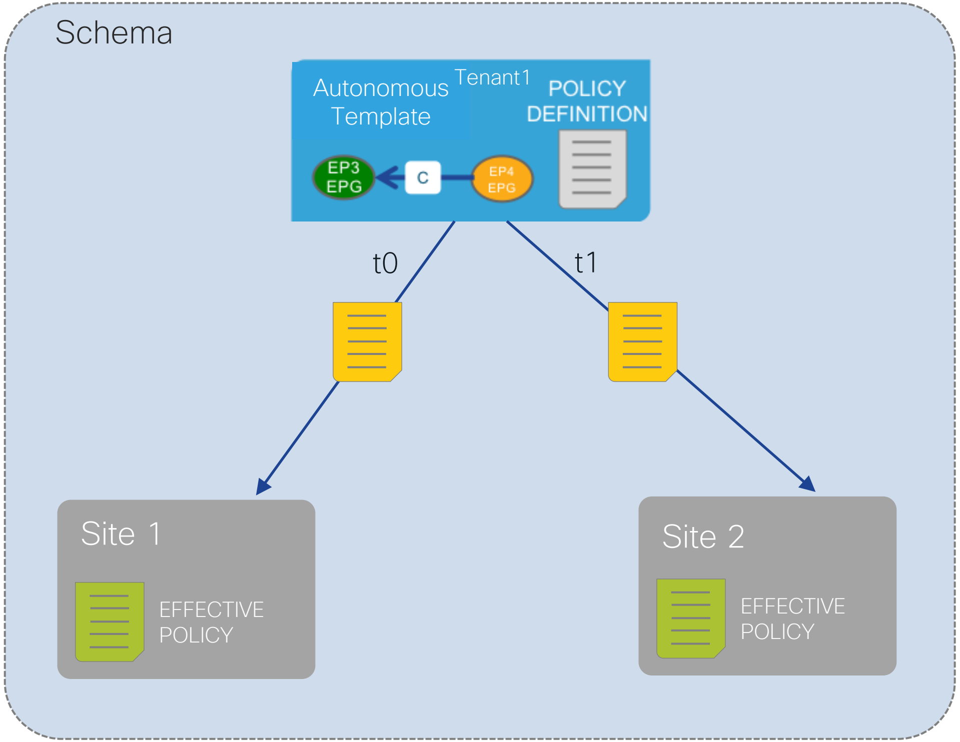
Autonomous Templates



Deployment Mode ⓘ

Multi-Site ● Autonomous Template

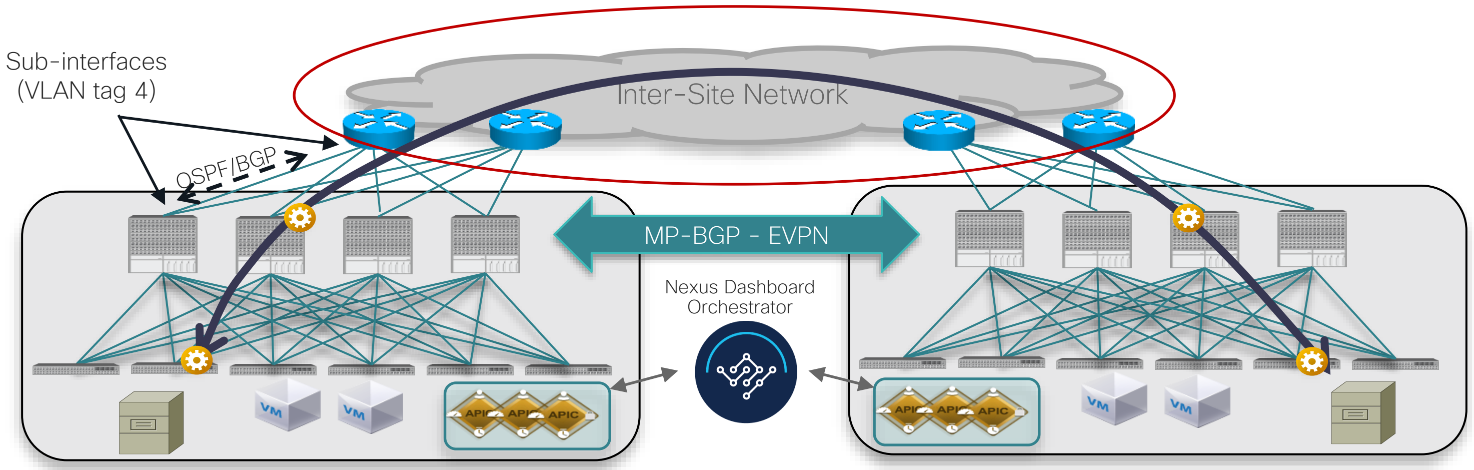
- Autonomous templates can also be associated to one or more fabrics
- Differently than for Multi-Site templates, the deployment of an Autonomous template to different sites won't cause the "stretching" of configuration objects (VRFs, BDs, EPGs,…)
- NDO performs a "configuration replication" function to multiple sites
- Autonomous Templates can be deployed to different fabrics at different points in time*
- Other template types behave as Application Autonomous templates



*Roadmap feature planned for CY23

Inter-Site Connectivity Deployment Considerations

Inter-Site Network (ISN) Functional Requirements



- Not managed by APIC or NDO, must be independently configured (day-0 configuration)
- IP topology can be arbitrary, not mandatory to connect all the spine nodes to the ISN
- ISN main functional requirements:
  - ✓ OSPF/BGP* to peer with the spine nodes and exchange TEP address reachability

    Must use sub-interfaces (with VLAN tag 4) toward the spines
  - ✓ No multicast requirement for BUM traffic forwarding across sites
  - ✓ Increased end-to-end MTU support (at least 50/54 extra Bytes)

*Requires ACI 5.2(1) and NDO 3.5(1)

Inter-Site Connectivity

Frequently Asked Questions

1 What platforms can or should I deploy in the ISN? →

- Any network device capable of routing traffic and supporting packets with increased MTU size can be deployed in the ISN
- Need sub-interfaces support for the ISN devices directly connected to the spines

2 Do I need to run L3 multicast inside the ISN? →

- No, ingress replications is performed by the ACI spine nodes to forward BUM traffic across sites
- This function is only required for the BDs that are stretched across sites with BUM flooding enabled

3 Can I use a Layer 2 only infrastructure as ISN? →

- No, the only officially supported configuration consists in deploying the ISN nodes as L3 network devices (particularly the ISN devices connected to the spines)

Inter-Site Connectivity

Frequently Asked Questions (2)

4 Do I need to deploy a dedicated infrastructure as ISN?

- No, the network providing ISN services for Multi-Site could also be used for other functions
- It is recommended (but not mandatory) to use a dedicated VRF for providing ISN connectivity

5 Is there a minimum bandwidth I should deploy between sites?

- No, the bandwidth required between sites mostly depends on the amount of east-west connectivity expected between sites

6 Is OSPF the only protocol supported to peer with the ISN network?

- No, from ACI release 5.2(1) and NDO release 3.5(1) we introduced support also for BGP peering between the spines and the first L3 hop ISN devices

ACI Multi-Site Control and Data Plane

Namespace Normalization and Shadow Objects

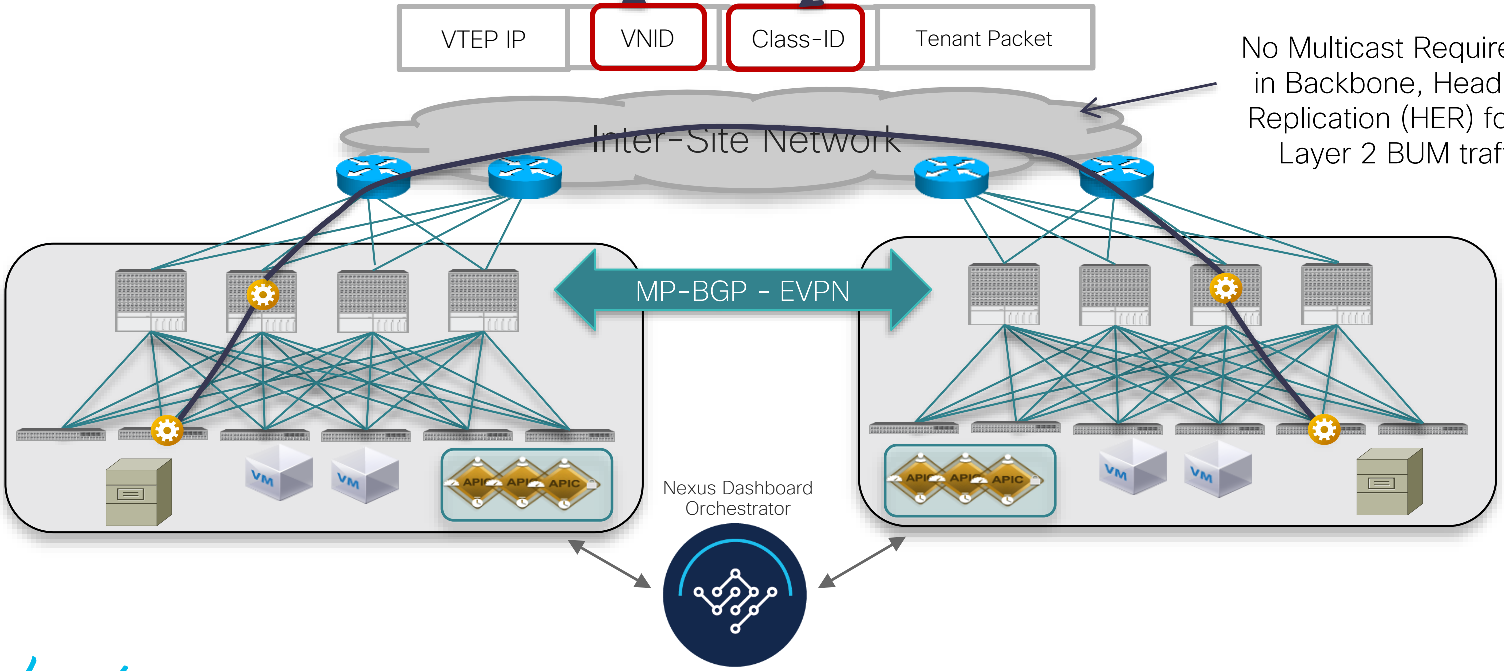
ACI Multi-Site

Network and Identity Extended between Fabrics

Deployment Mode ⓘ
Multi-Site ○ Autonomous Template

Network information carried across Fabrics (Availability Zones)

Identity information carried across Fabrics (Availability Zones)



No Multicast Requirement in Backbone, Head-End Replication (HER) for any Layer 2 BUM traffic)

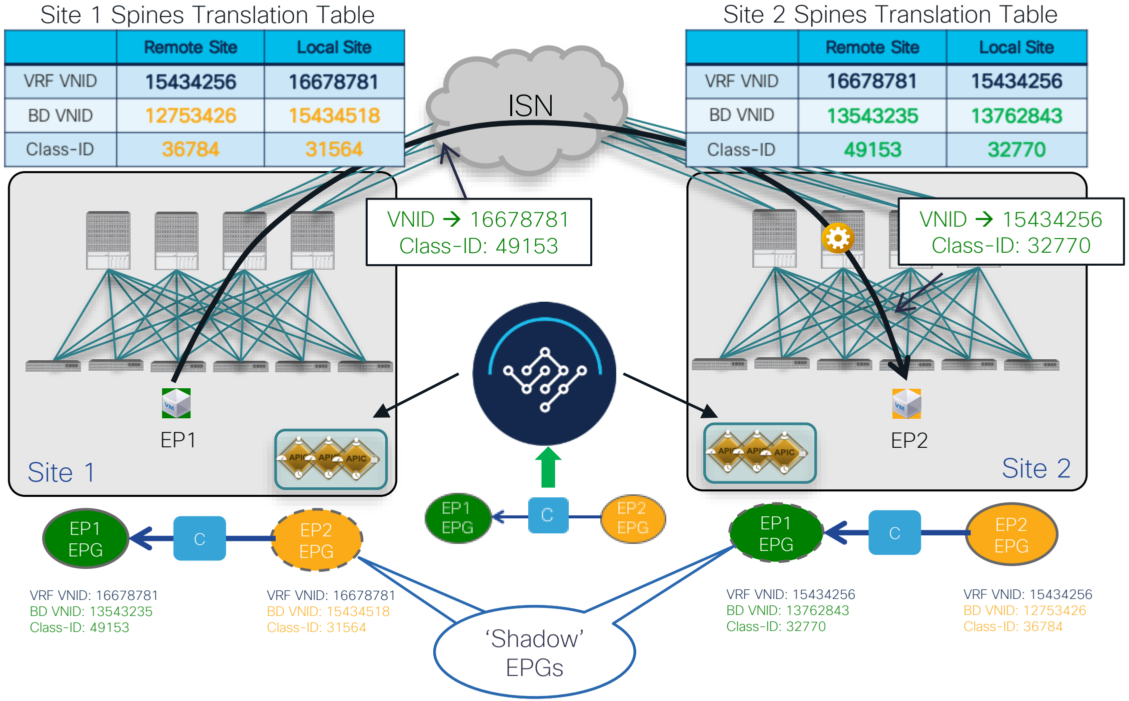
ACI Multi-Site

Inter-Site Policies and Spines' Translation Tables

Deployment Mode ⓘ

Multi-Site ○ Autonomous Template

- Inter-Site policies defined on the ACI Nexus Dashboard Orchestrator are pushed to the respective APIC domains
  - End-to-end policy consistency
  - Creation of 'Shadow' EPGs to locally represent the policies
- Inter-site communication requires the installation of translation table entries on the spines (namespace normalization)
- Translation entries are populated in different cases:
  - Stretched EPGs/BDs
  - Creation of a contract between not stretched EPGs
  - Preferred Group or vzAny deployments



Site 1 Spines Translation Table

	Remote Site	Local Site
VRF VNID	15434256	16678781
BD VNID	12753426	15434518
Class-ID	36784	31564

Site 2 Spines Translation Table

	Remote Site	Local Site
VRF VNID	16678781	15434256
BD VNID	13543235	13762843
Class-ID	49153	32770

VNID → 16678781
Class-ID: 49153

VNID → 15434256
Class-ID: 32770

VRF VNID: 16678781
BD VNID: 13543235
Class-ID: 49153

VRF VNID: 16678781
BD VNID: 15434518
Class-ID: 31564

VRF VNID: 15434256
BD VNID: 13762843
Class-ID: 32770

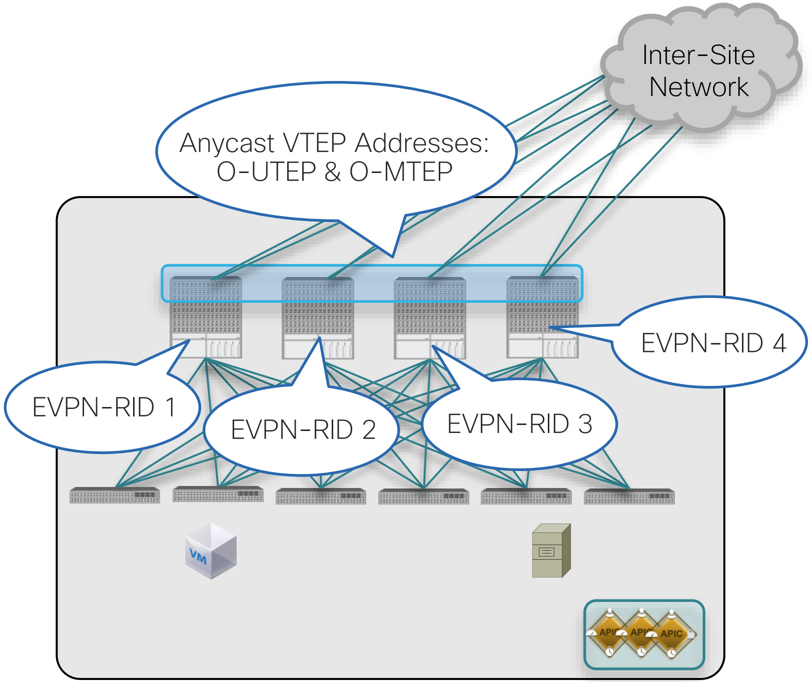
VRF VNID: 15434256
BD VNID: 12753426
Class-ID: 36784

'Shadow' EPGs

Underlay and Overlay Control Plane Considerations

ACI Multi-Site

BGP Inter-Site Peers



- Spines connected to the Inter-Site Network perform two main functions:
  1. Establishment of MP-BGP EVPN peerings with spines in remote sites
     - One dedicated Control Plane address (EVPN-RID) is assigned to each spine running MP-BGP EVPN
  2. Forwarding of inter-sites data-plane traffic
     - Anycast Overlay Unicast TEP (O-UTEP): assigned to all the spines connected to the ISN and used to source and receive L2/L3 unicast traffic
     - Anycast Overlay Multicast TEP (O-MTEP): assigned to all the spines connected to the ISN and used to receive L2 BUM traffic
- EVPN-RID, O-UTEP and O-MTEP addresses are assigned from the Nexus Dashboard Orchestrator and must be routable across the ISN

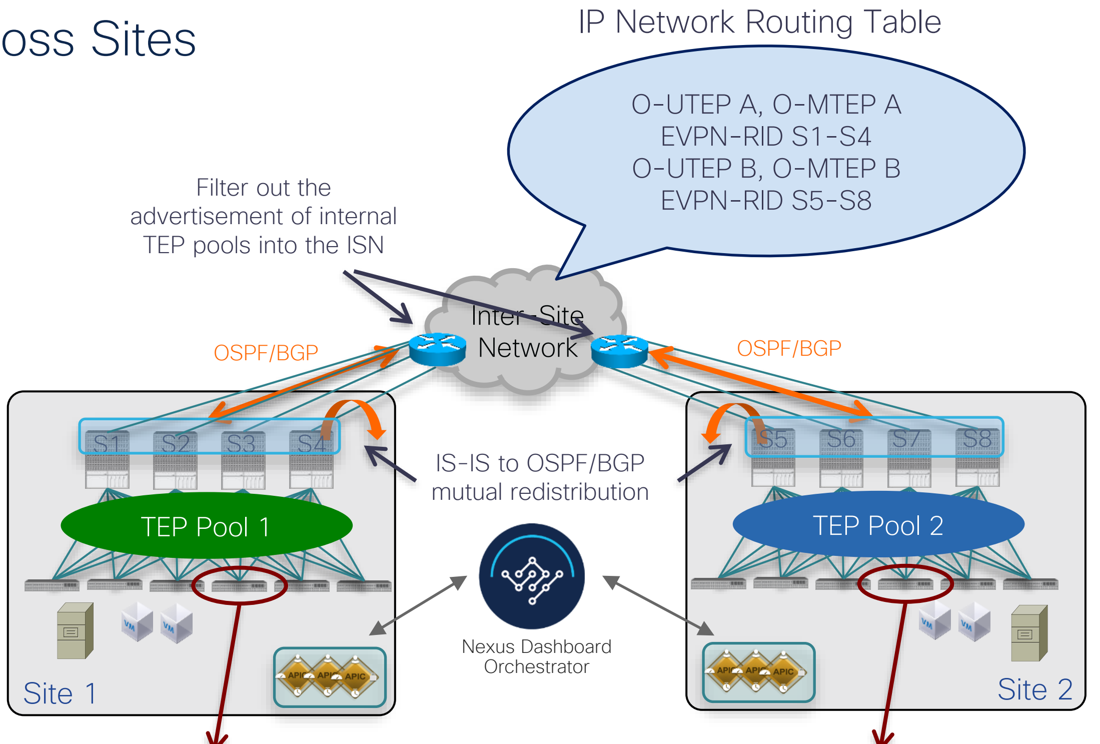
ACI Multi-Site

Exchanging TEP Information across Sites

- OSPF or BGP peering between spines and Inter-Site network
    - Mandates the use of L3 sub-interfaces (with VLAN 4 tag) between the spines and the ISN
- Exchange of External Spine TEP addresses (EVPN-RID, O-UTEP and O-MTEP) across sites

  Internal TEP Pool information not needed to establish inter-site communication (should be filtered out on the first-hop ISN router)

  Use of overlapping internal TEP Pools across sites is fully supported



Leaf Routing Table

IP Prefix	Next-Hop
O-UTEP B, O-MTEP B, EVPN-RID S5-S8	Site1-S1, Site1-S2, Site1-S3, Site1-S4

Leaf Routing Table

IP Prefix	Next-Hop
O-UTEP A, O-MTEP A, EVPN-RID S1-S4	Site2-S5, Site2-S6, Site2-S7, Site2-S8

ACI Multi-Site

Inter-Site MP-BGP EVPN Control Plane

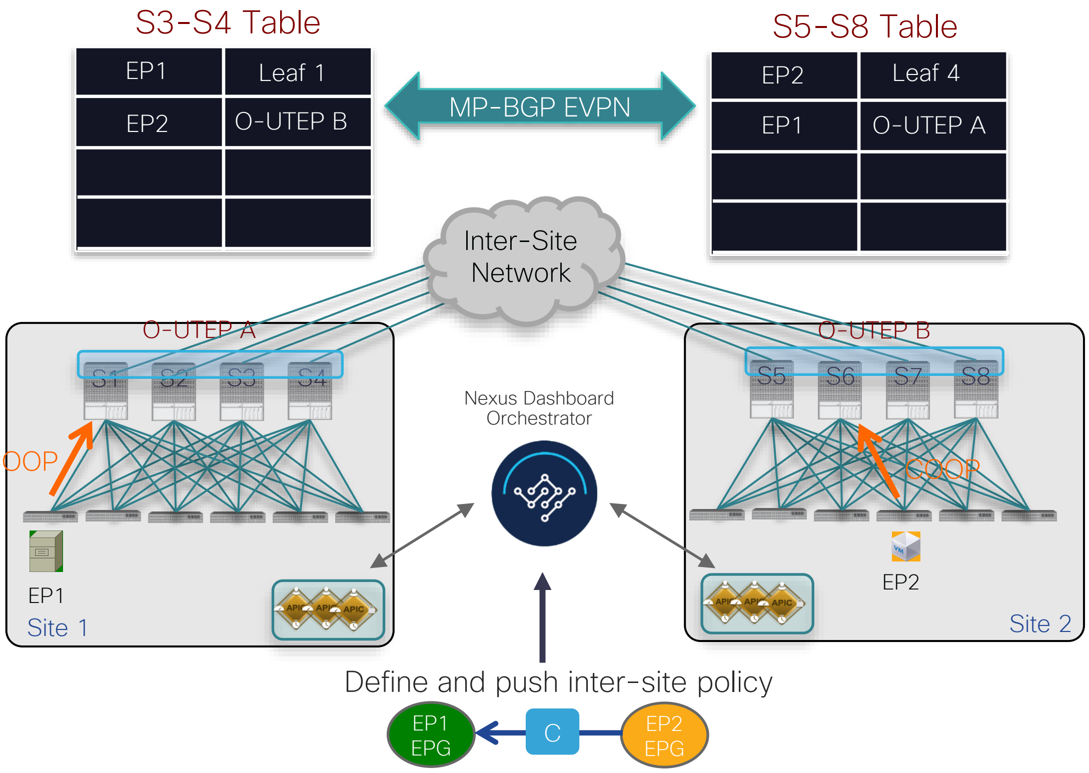
- MP-BGP EVPN used to communicate Endpoint (EP) information across Sites

  MP-iBGP or MP-EBGP peering options supported

  Remote host route entries (EVPN Type-2) are associated to the remote site Anycast O-UTEP address

- Automatic filtering of endpoint information across Sites

  Host routes are exchanged across sites **only** if there is a cross-site contract requiring communication between endpoints



S3-S4 Table

EP1	Leaf 1
EP2	O-UTEP B

S5-S8 Table

EP2	Leaf 4
EP1	O-UTEP A

Data Plane Communication across Sites

ACI Multi-Site

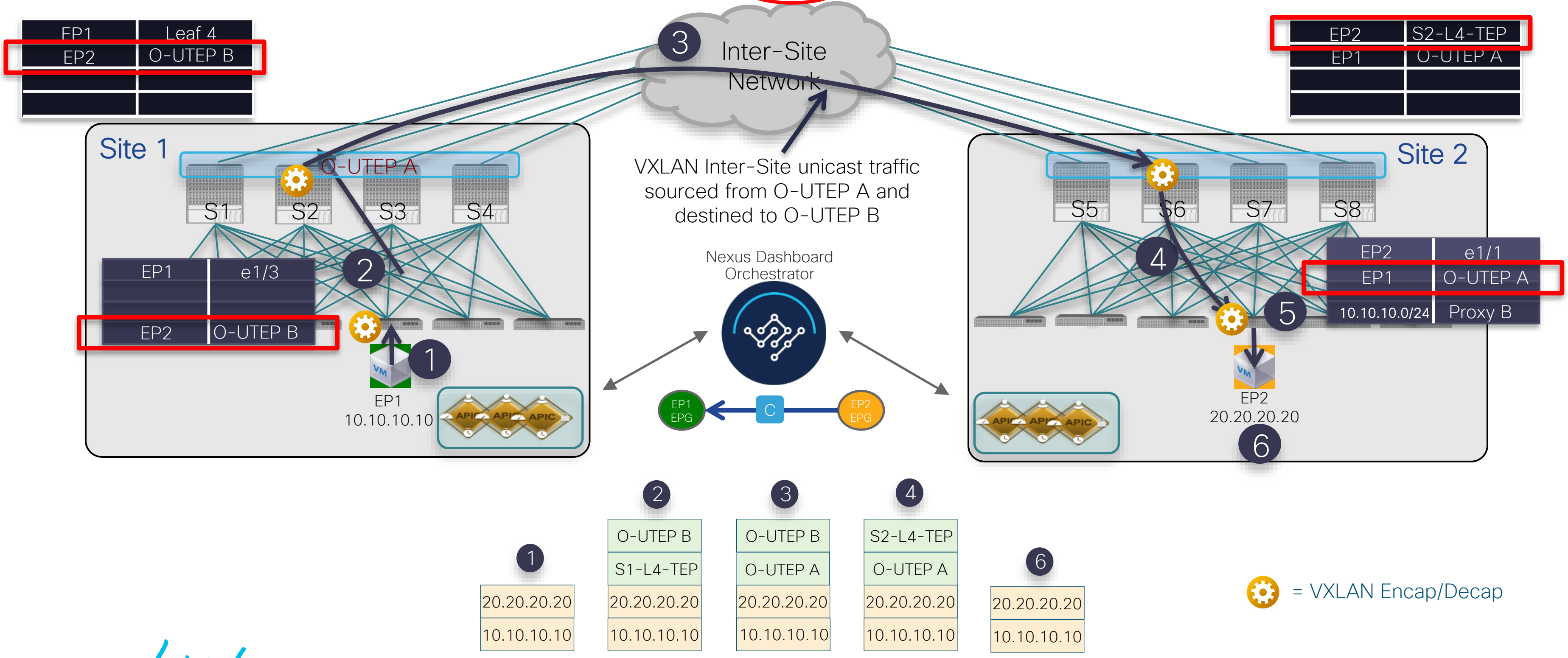
Inter-Sites Unicast Data Plane

Deployment Mode ⓘ

Multi-Site ○ Autonomous Template

VTEP IP	VNID	Class-ID	Tenant Packet

Policy information carried across Pods



VXLAN Inter-Site unicast traffic sourced from O-UTEP A and destined to O-UTEP B

EP1	Leaf 4
EP2	O-UTEP B

EP1	e1/3
EP2	O-UTEP B

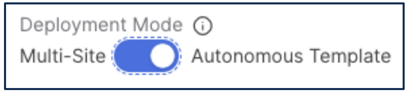
EP2	S2-L4-TEP
EP1	O-UTEP A

EP2	e1/1
EP1	O-UTEP A
10.10.10.0/24	Proxy B

1	2	3	4	6
	O-UTEP B	O-UTEP B	S2-L4-TEP	
	S1-L4-TEP	O-UTEP A	O-UTEP A	
20.20.20.20	20.20.20.20	20.20.20.20	20.20.20.20	20.20.20.20
10.10.10.10	10.10.10.10	10.10.10.10	10.10.10.10	10.10.10.10

ACI Multi-Site

L3 Only across Sites ("Autonomous Sites")

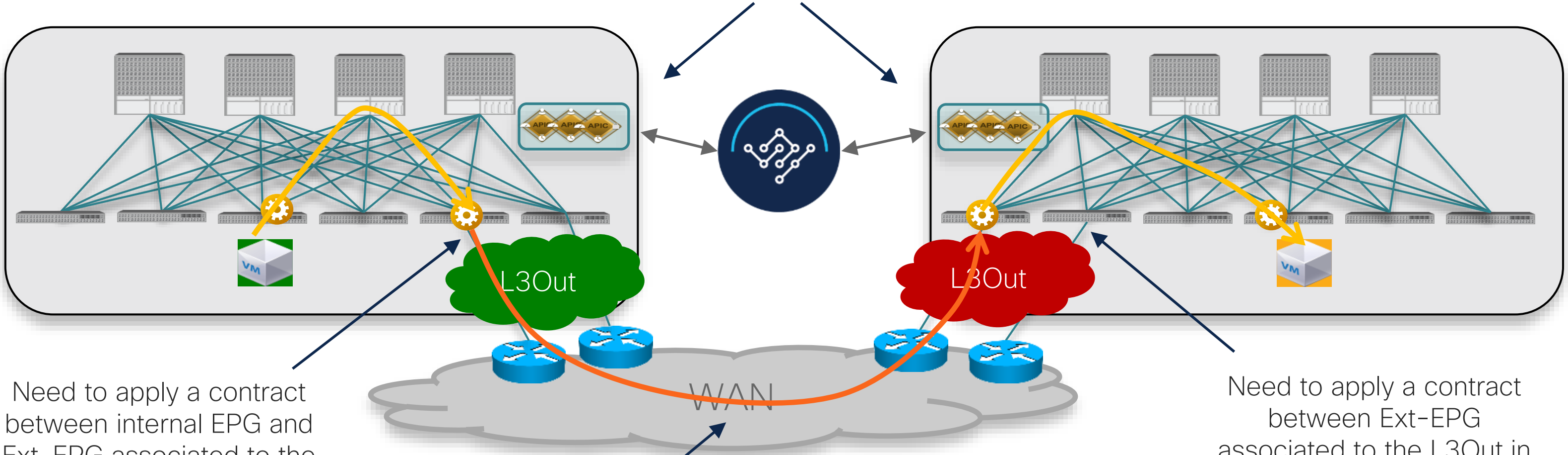


Deployment Mode ⓘ

Multi-Site ● Autonomous Template

Routing across sites via the WAN backbone

"Autonomous ACI Fabrics" (no ISN)



Need to apply a contract between internal EPG and Ext-EPG associated to the L3Out in Fabric 1

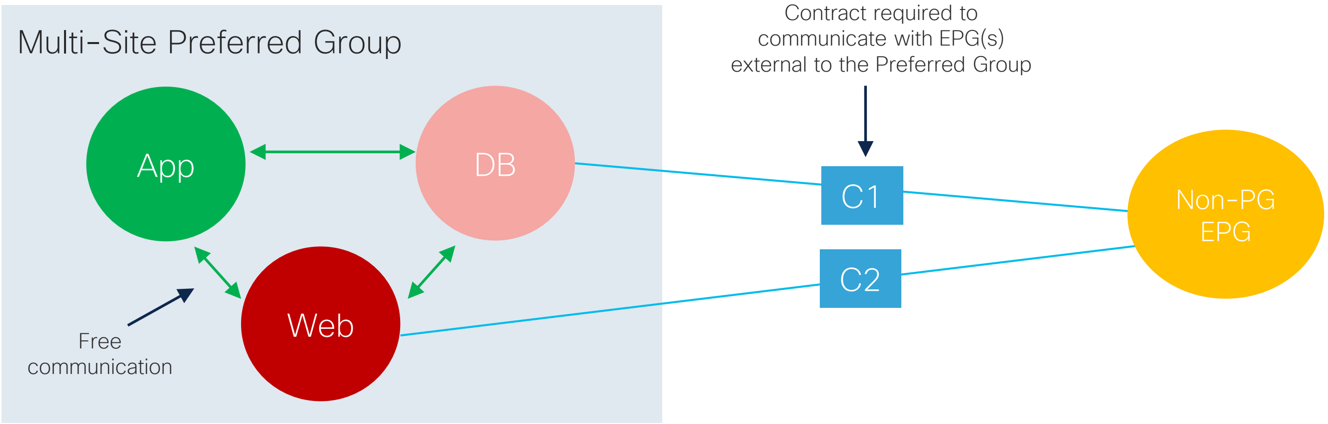
Mandates the use of a multi-VRF capable backbone network (VRF-Lite, MPLS-VPN, etc.) to extend multiple VRFs across fabrics

Need to apply a contract between Ext-EPG associated to the L3Out in Fabric 2 and internal EPG

Simplify Policy Application Preferred Group and vzAny

ACI Multi-Site

Simplify Policy Enforcement: Preferred Groups

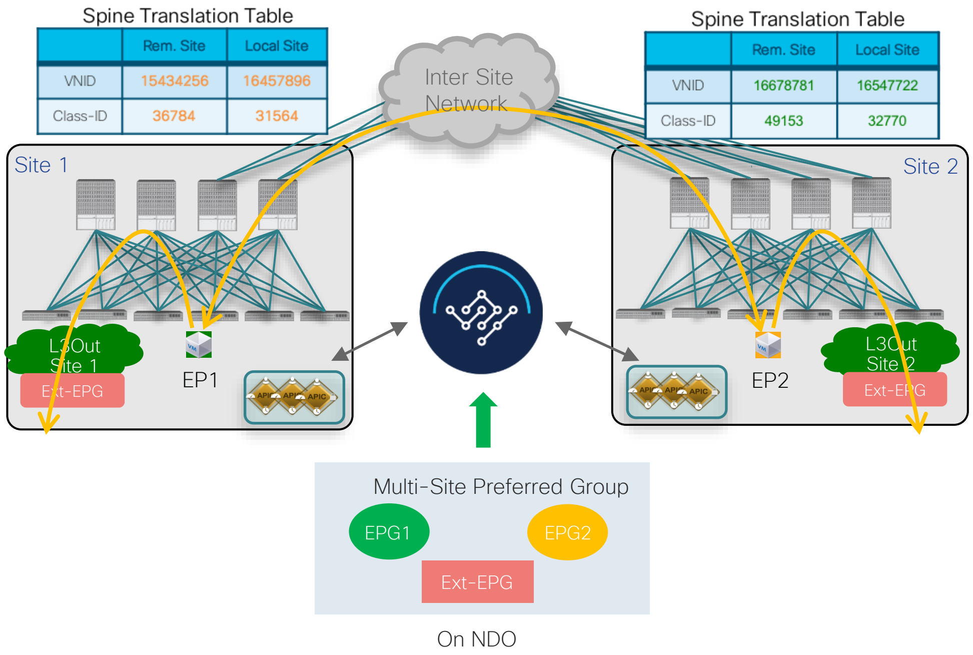


Multi-Site Preferred Group
App
DB
Web
Free communication
Contract required to communicate with EPG(s) external to the Preferred Group
C1
C2
Non-PG EPG

- ”VRF unenforced” not supported with Multi-Site
- Multi-Site Preferred Group configuration from the Multi-Site Orchestrator is supported from MSO 2.0(2) release
  - Creates ‘shadow’ EPGs and translation table entries ‘under the hood’ to allow ‘free’ inter-site communication
  - 250 Preferred Groups supported as MSO release 2.2(3), 1000 from MSO release 2.2(4)
- Typically desired in legacy to ACI migration scenarios

Simplify Policy Enforcement

Preferred Groups for E-W and N-S Flows



Spine Translation Table

	Rem. Site	Local Site
VNID	15434256	16457896
Class-ID	36784	31564

Spine Translation Table

	Rem. Site	Local Site
VNID	16678781	16547722
Class-ID	49153	32770

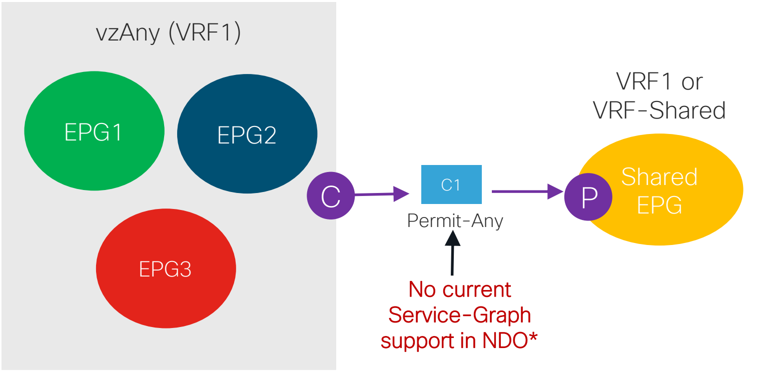
- Adding internal EPGs and External EPGs (associated to L3Outs) to the Preferred Group allows to enable free east-west and north-south connectivity
- When adding the Ext-EPG to the Preferred Group:
  - Can't use 0.0.0.0/0 for classification, needs more specific prefixes
  - As workaround it is possible to use 0.0.0.0/1 and 128.0.0.0/1 to achieve the same result
  - Must ensure Ext-EPG is a stretched object
- Intersite L3Out not supported if the Ext-EPG is part of a Preferred Group (as of NDO 4.1(1))

Simplify Policy Enforcement

vzAny Support

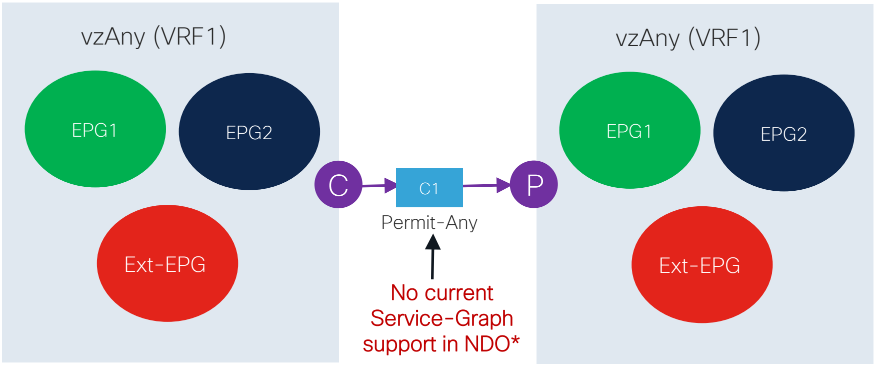
What is vzAny? Logical object representing all the EPGs in a VRF

Use case 1: Many-to-One communication (Shared Services)



- Multiple EPGs part of a specific VRF1 consume the services provided by a shared EPG (part of VRF1 or of a VRF-shared)
- VRF-shared can be part of the same tenant or of a different tenant

Use case 2: Enable free communication inside a VRF

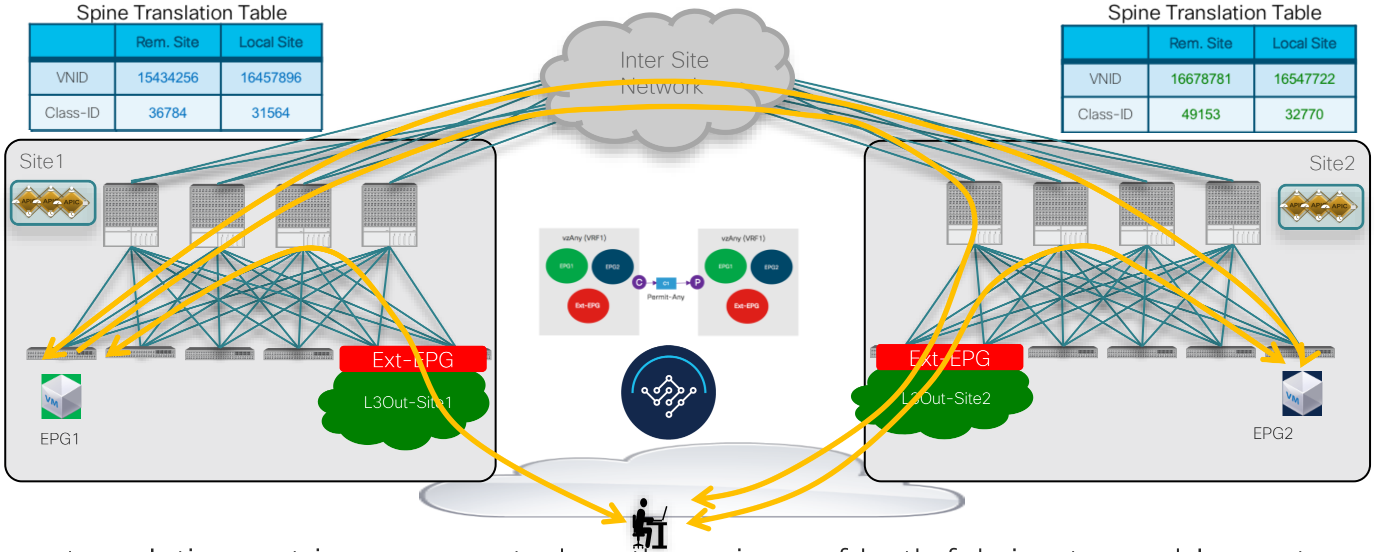


- vzAny provides and consumes a contract with an associated “Permit-any” filter
- Use ACI fabric only for network connectivity without policy enforcement
- Equivalent to “VRF unenforced”

*Planned for NDO 4.2(1) release

ACI Multi-Site and vzAny

Enable Inter-Site Free Communication Inside a VRF



Spine Translation Table

	Rem. Site	Local Site
VNID	15434256	16457896
Class-ID	36784	31564

Spine Translation Table

	Rem. Site	Local Site
VNID	16678781	16547722
Class-ID	49153	32770

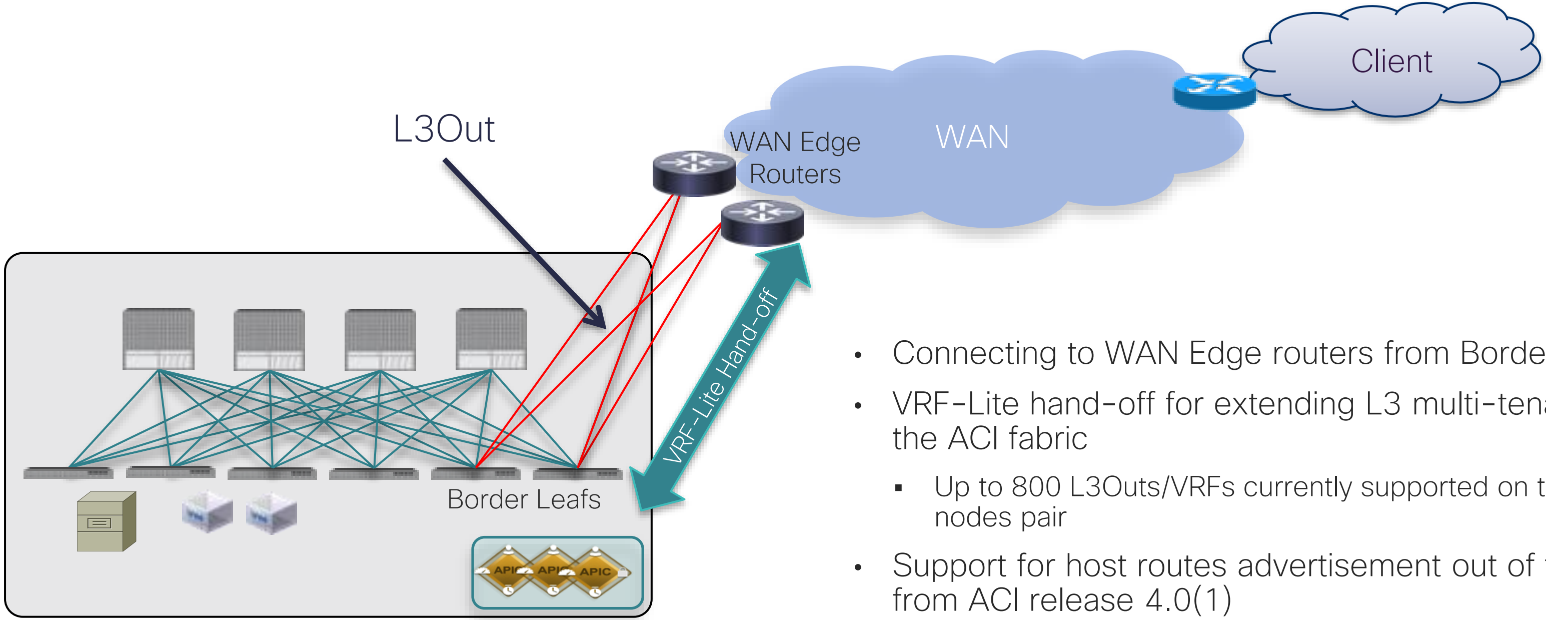
- Proper translation entries are created on the spines of both fabrics to enable east-west communication
- Supported also for connecting to the external Layer 3 domain
- vzAny + PBR support for any-to-any communication planned for a future NDO release

Connecting to the External L3 Domain

Different Types of L3Outs

Connecting to the External Layer 3 Domain

‘Traditional’ IP-Based L3Outs (Recommended Option)

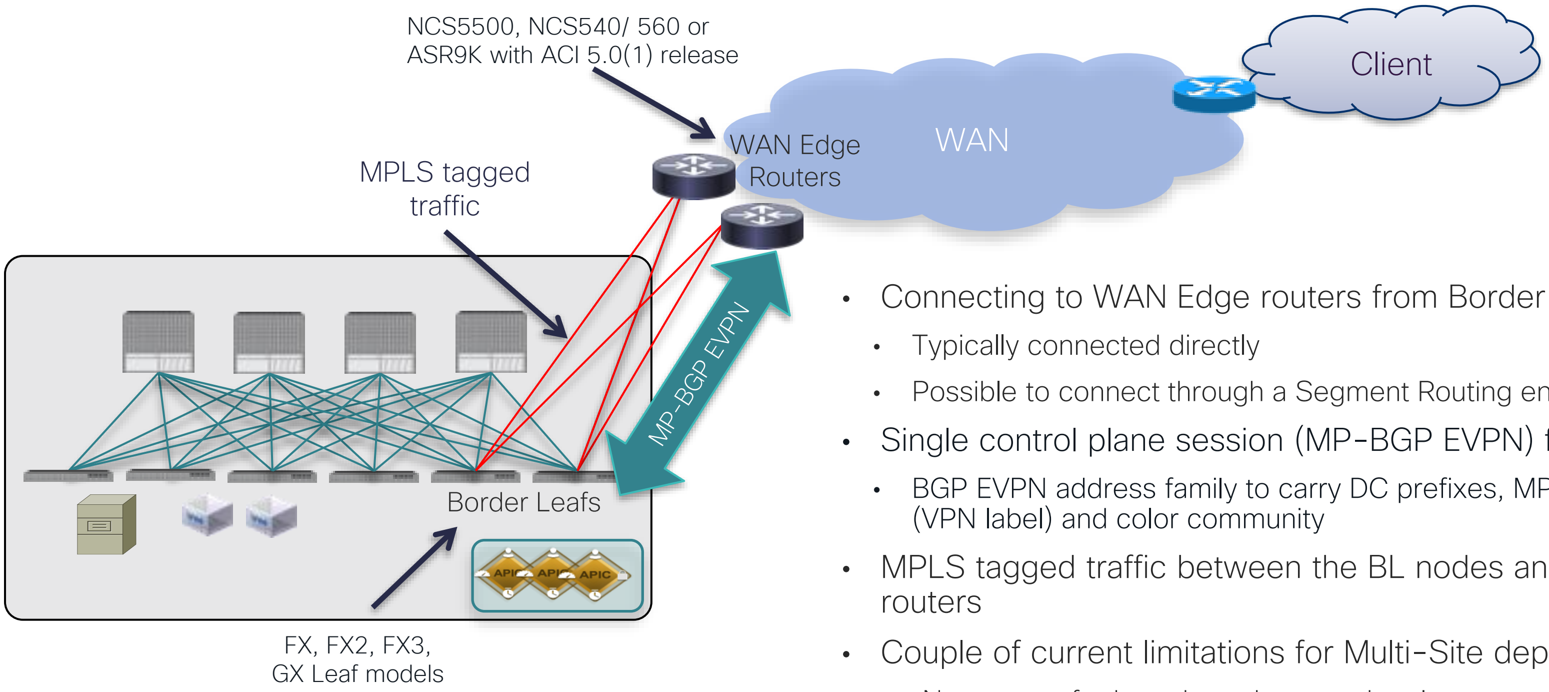


- Connecting to WAN Edge routers from Border Leaf nodes
- VRF-Lite hand-off for extending L3 multi-tenancy outside the ACI fabric
  - Up to 800 L3Outs/VRFs currently supported on the same BL nodes pair
- Support for host routes advertisement out of the ACI Fabric from ACI release 4.0(1)
  - Enabled at the BD level
- Support for L3 Multicast and Shared L3Out

Connecting to the External Layer 3 Domain

SR-MPLS/MPLS Hand-Off on the BL Nodes

NDO 4.0(2) Release

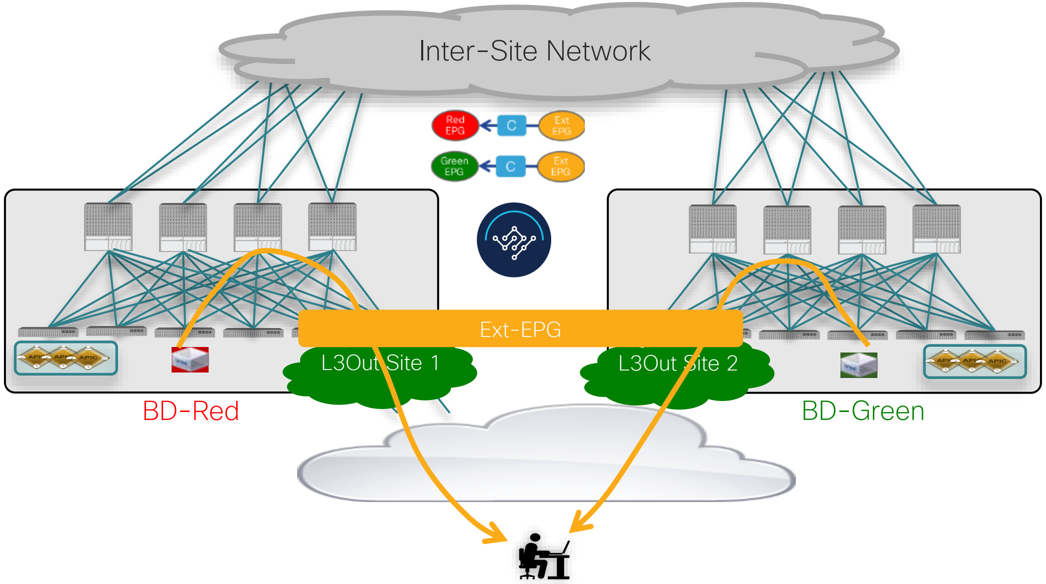


- Connecting to WAN Edge routers from Border Leaf nodes
  - Typically connected directly
  - Possible to connect through a Segment Routing enabled network
- Single control plane session (MP-BGP EVPN) for all tenant VRFs
  - BGP EVPN address family to carry DC prefixes, MPLS label for VRF (VPN label) and color community
- MPLS tagged traffic between the BL nodes and the WAN Edge routers
- Couple of current limitations for Multi-Site deployments
  - No support for host-based route advertisement
  - No current support for Layer 3 Multicast communication

Deploying External EPG(s) Associated to the L3Out

ACI Multi-Site and L3Out

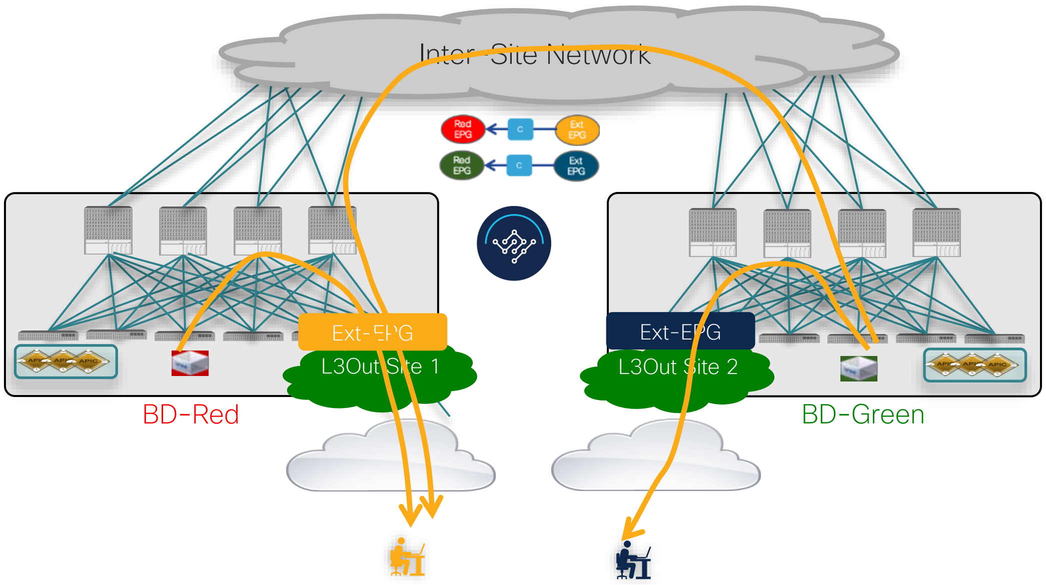
Stretching or Not Stretching the Ext-EPG?



- The Ext-EPG can be defined in a template associated to multiple sites (stretched object)
  - The Ext-EPG must then be mapped to the local L3Outs in the “site level” section of the template configuration
  - L3Outs remain independent objects defined in each site
- Recommended when the L3Outs in the separate sites provide access to a common set of external resources (as the WAN)
  - Simplifies the policy definition and external traffic classification
  - Still allows to apply route-map polices on each L3Out (since we have independent APIC domains)

ACI Multi-Site and L3Out

Stretching or Not Stretching the Ext-EPG?



- Separate Ext-EPGs can be defined in templates mapped to separate sites (non stretched objects)

  Each Ext-EPG can be mapped to the local L3Out in the "global" or "site level" section of the template configuration

- Allows to apply different policies to each Ext-EPGs at different time

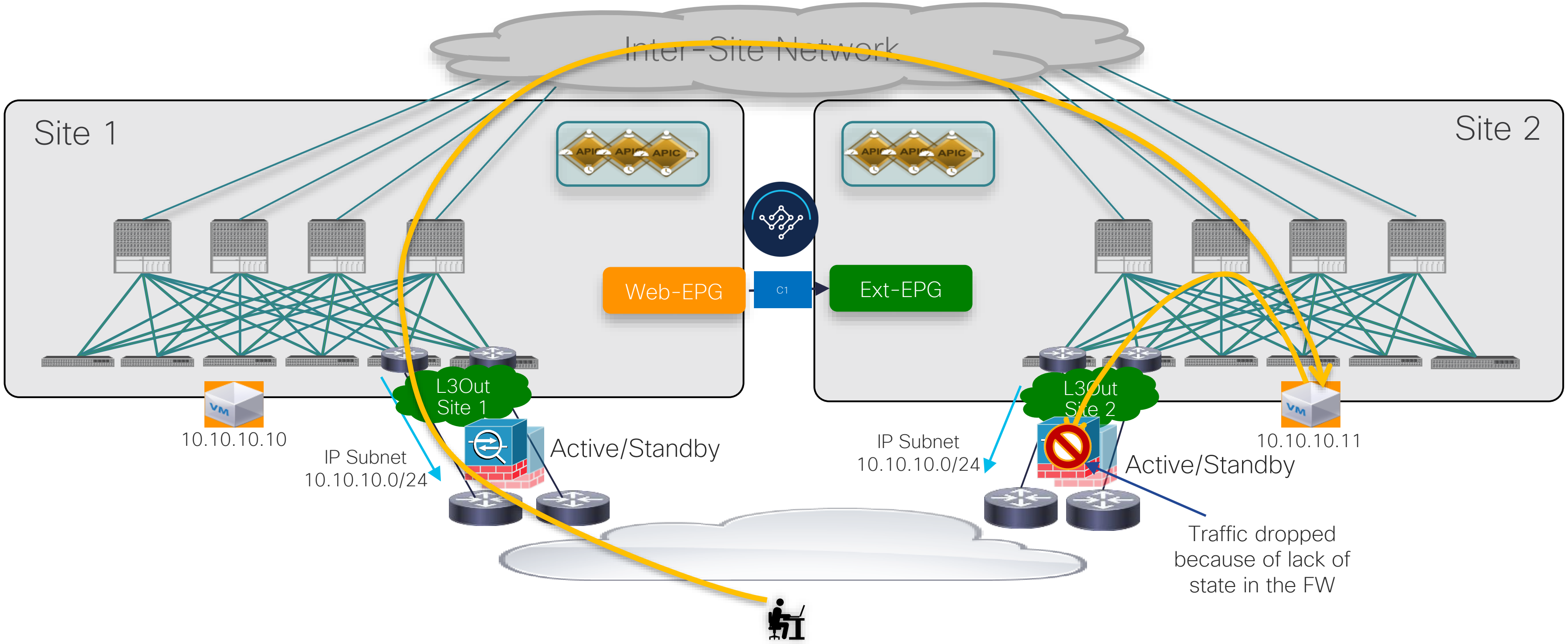
- Can still use the same 0.0.0.0/0 network configuration for classification on both sites

- May require enablement of Intersite L3Out

Solving Asymmetric Routing Issues with the External Network

ACI Multi-Site and L3Out

Typical Deployment of Perimeter FWs

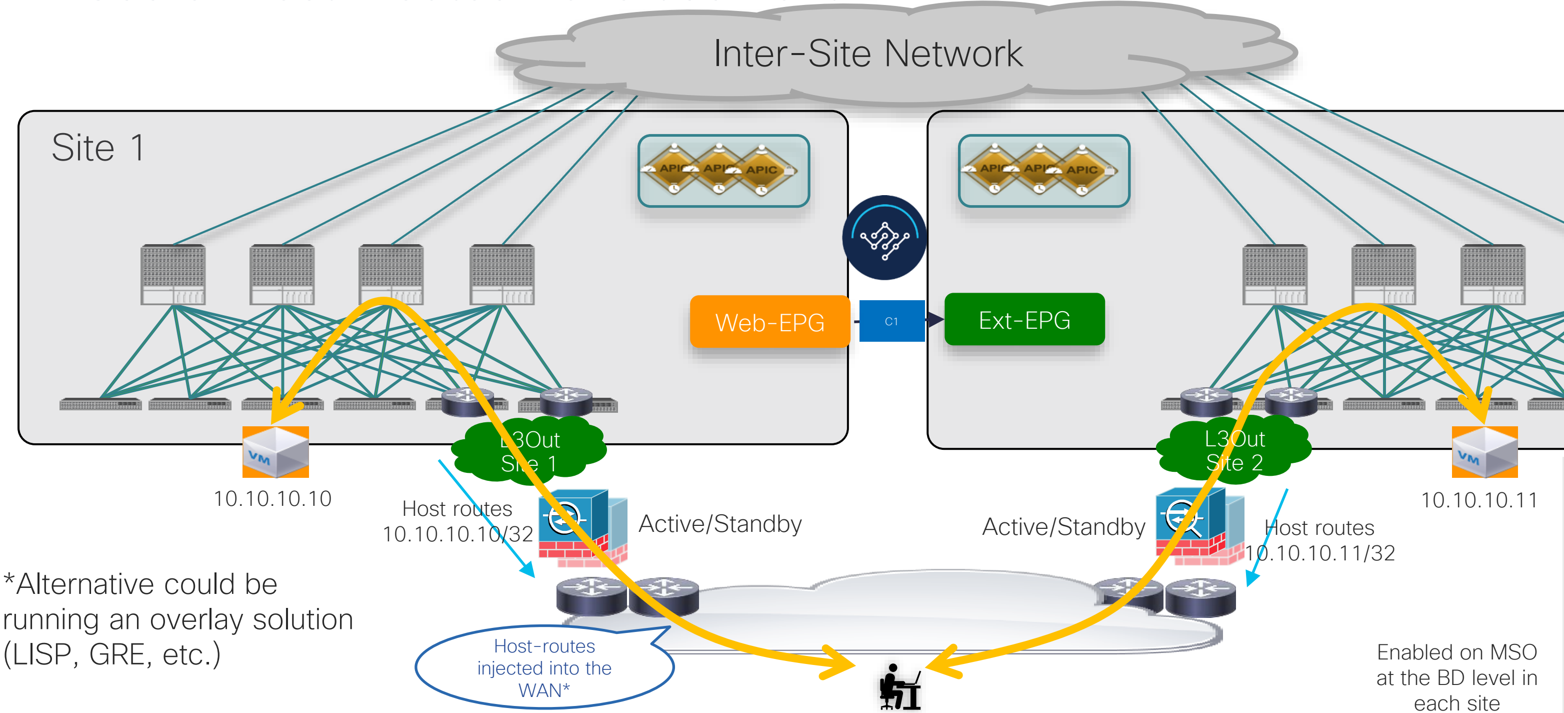


Traffic dropped because of lack of state in the FW

Solving Asymmetric Routing Issues

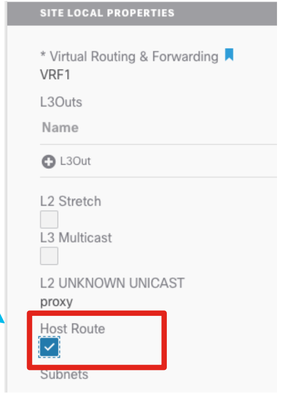
Use of Host-Routes Advertisement

ACI 4.0(1) Release



*Alternative could be running an overlay solution (LISP, GRE, etc.)

Enabled on MSO at the BD level in each site



SITE LOCAL PROPERTIES

* Virtual Routing & Forwarding
VRF1

L3Outs

Name

L3Out

L2 Stretch
□

L3 Multicast
□

L2 UNKNOWN UNICAST
proxy

Host Route
☑

Subnets

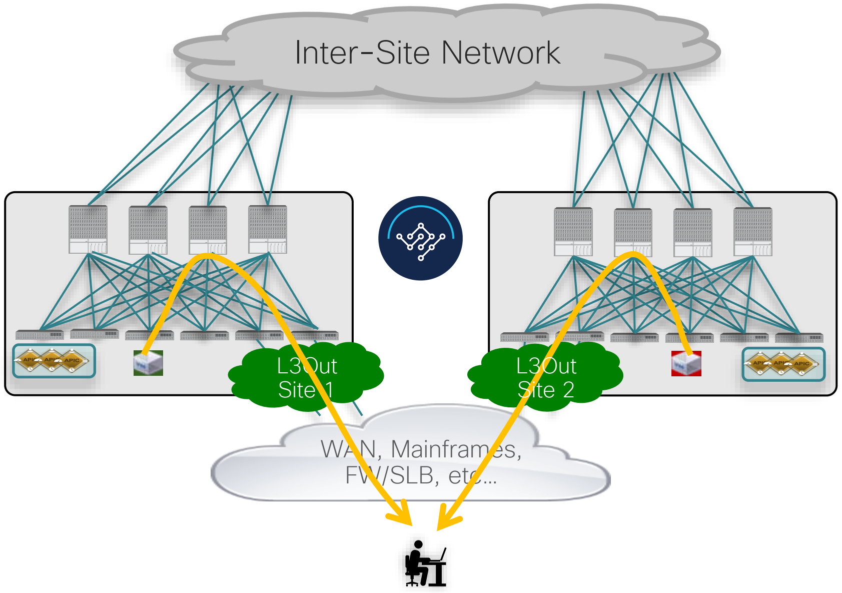
- Ingress optimization requires host-routes advertisement on the L3Out
  - Native support on ACI Border Leaf nodes available from ACI release 4.0(1)
  - Not currently supported for SR-MPLS L3Outs

Intersite L3Out Support

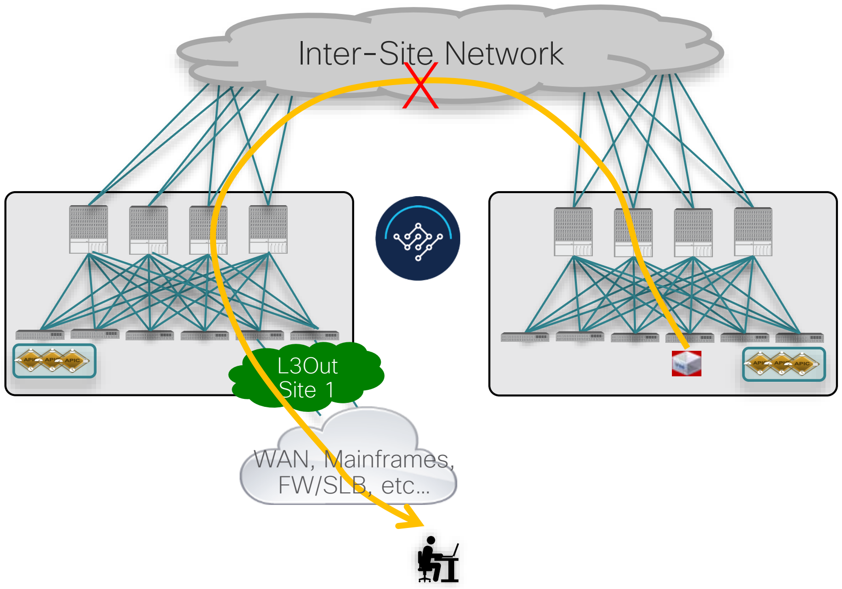
Problem Statement

Behavior before ACI Release 4.2(1)

Supported Design ✓



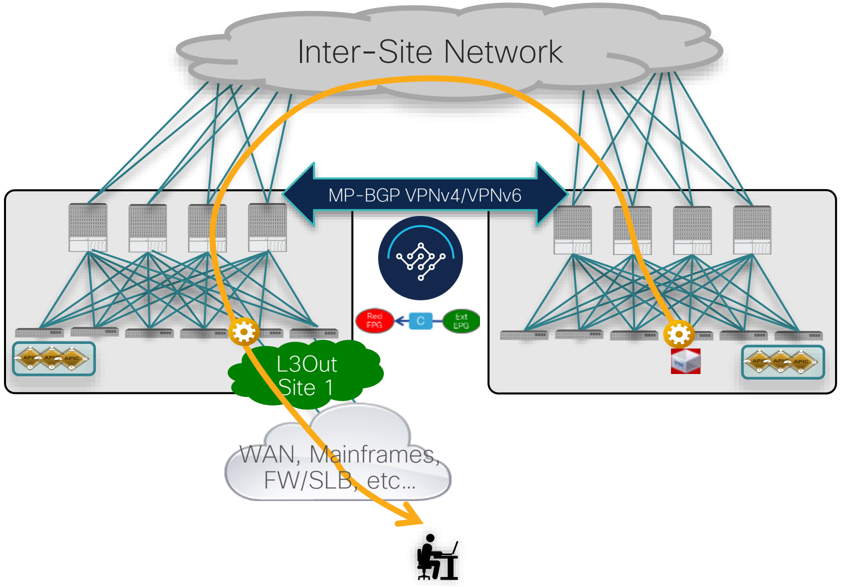
Not Supported Design ✗



Note: the same consideration applies to both IP-Based L3Outs and SR-MPLS L3Outs

ACI Multi-Site and L3Out

Support of Intersite L3Out

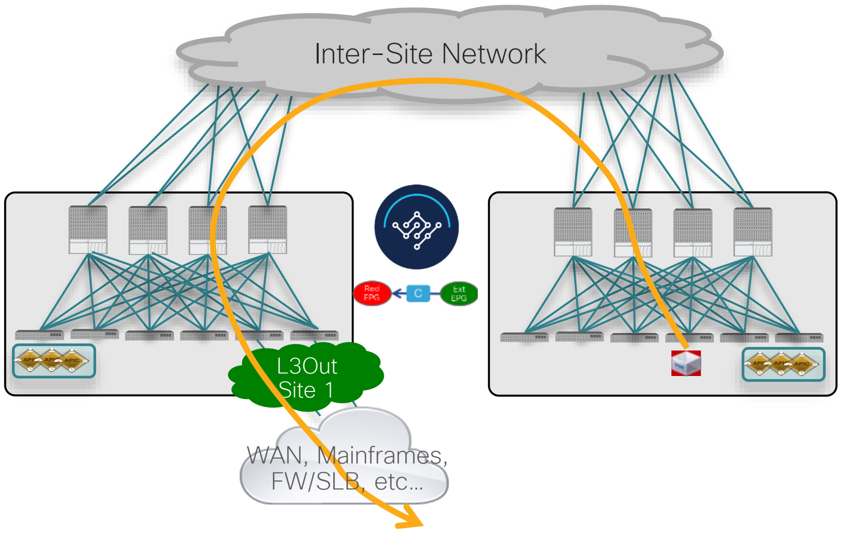


- Starting with ACI Release 4.2(1) it is possible for endpoints in a site to send traffic to resources (WAN, Mainframes, FWs/SLBs, etc.) accessible via a remote L3Out connection
- External prefixes are exchanged across sites via MP-BGP VPNV4/VPNv6 sessions between spines
- Traffic will be directly encapsulated to the TEP of the remote BL nodes
  - The BL nodes will get assigned an address part of an additional (configurable) prefix that must be routable across the ISN
- Same solution will also support transit routing across sites (L3Out to L3Out)

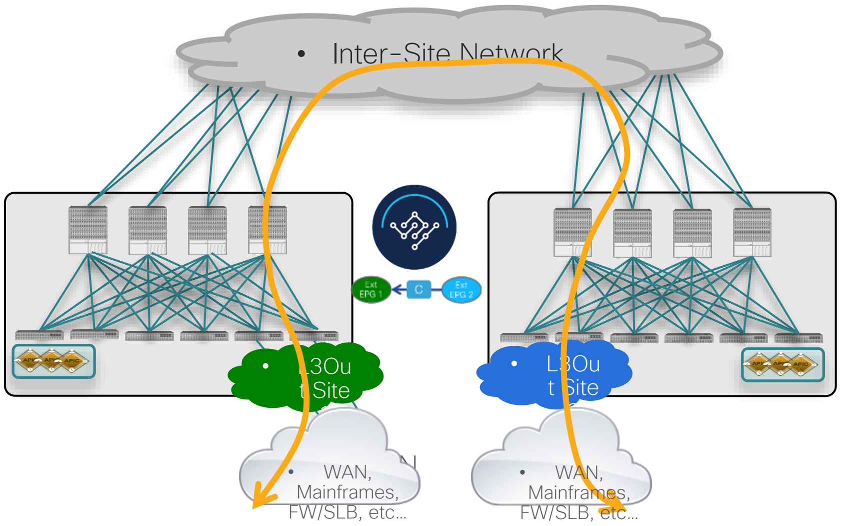
ACI Multi-Site and Intersite L3Out

Supported Scenarios

ACI 4.2(1) Release



- Endpoint to remote L3Out communication (intra-VRF)
- Endpoint to remote L3Out communication (inter-VRF)



- Inter-site transit routing (intra-VRF)
- Inter-site transit routing (inter-VRF)

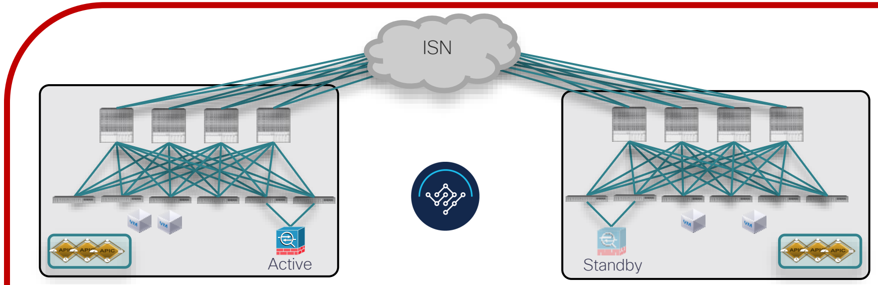
Network Services Integration

Integration Models

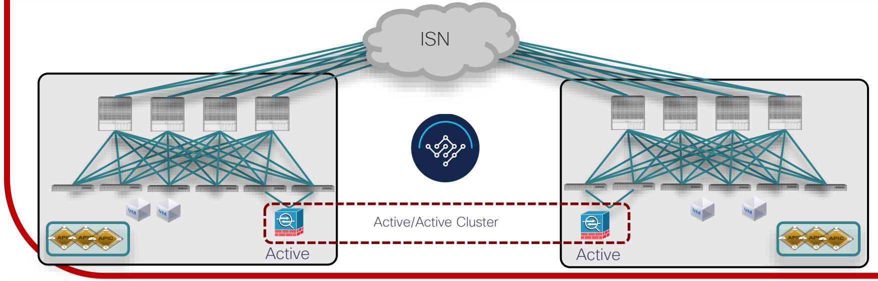
ACI Multi-Site and Network Services

Integration Models

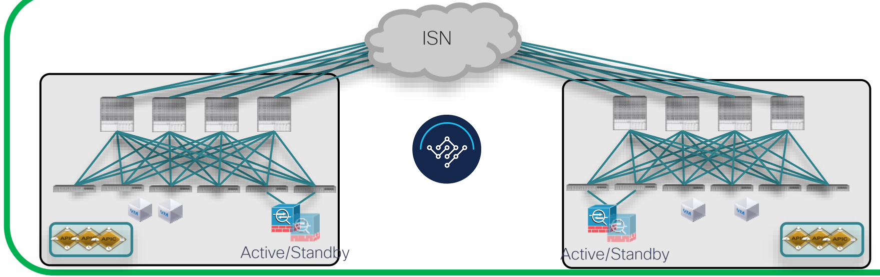
Deployment options fully supported with ACI Multi-Pod



- Active and Standby pair deployed across Pods
- Limited supported options



- Active/Active FW cluster nodes stretched across Sites (single logical FW)
- Limited supported options



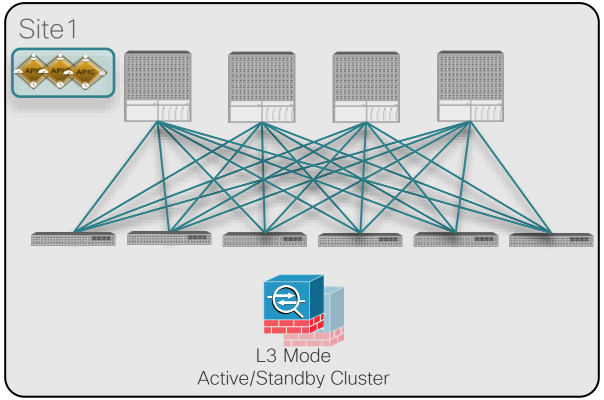
- Typical deployment model for ACI Multi-Site, each fabric leverages a dedicated service node function
- Use of PBR to avoid creating asymmetric paths through stateful devices (FWs, LBs, etc.) for both North-South and East-West communication

Use of Service Graph and Policy Based Redirection

Resilient Service Node Deployment in Each Site

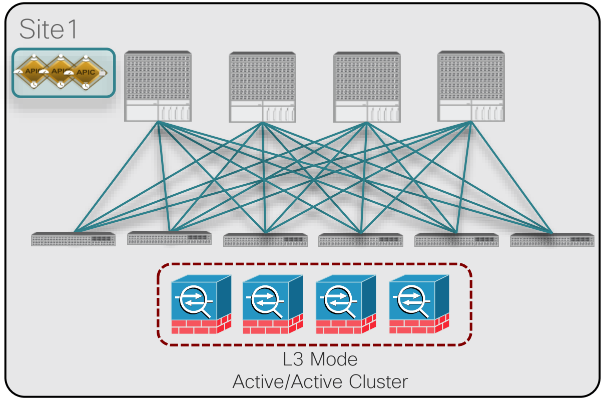
PBR redirection only supported to a local service function, hence it is important to deploy such function in a resilient way

Active/Standby Cluster



- The Active/Standby pair represents a single MAC/IP entry in the PBR policy

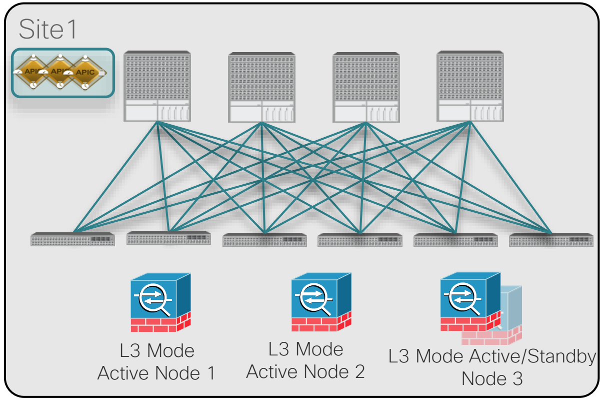
Active/Active Cluster



- The Active/Active cluster represents a single MAC/IP entry in the PBR policy
- Spanned Ether-Channel Mode supported with Cisco ASA/FTD platforms

  All ASA/FTD nodes must be connected to the same leaf nodes pair

Independent Active Nodes

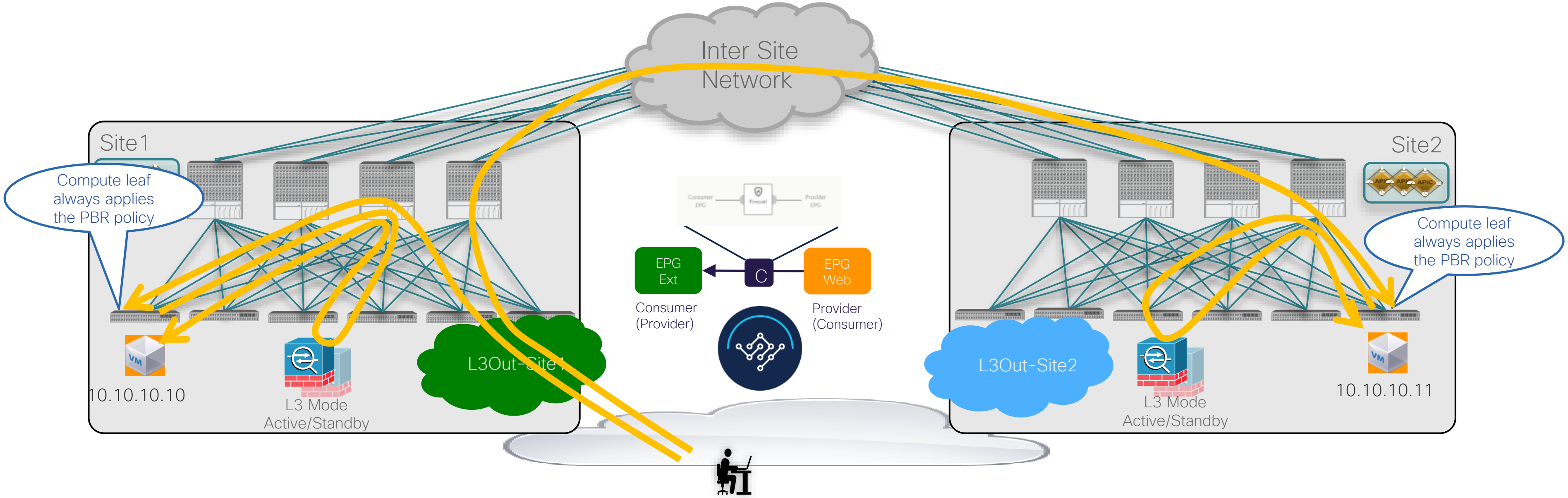


- Each Active node represent a unique MAC/IP entry in the PBR policy
- Use of Symmetric PBR to ensure each flow is handled by the same Active node in both directions

Use of Service Graph and PBR North-South and East-West

Use of Service Graph and Policy Based Redirection

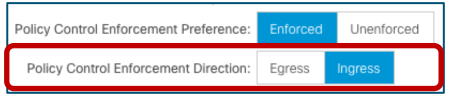
North-South Communication – Inbound Traffic



- Inbound traffic can enter any site when destined to a stretched subnet (if ingress optimization is not deployed or possible)
- PBR policy is always applied on the compute leaf node where the destination endpoint is connected

  Requires the VRF to have the default policies for enforcement preference and direction

  Ext-EPG and Web EPG can indifferently be provider or consumer of the contract

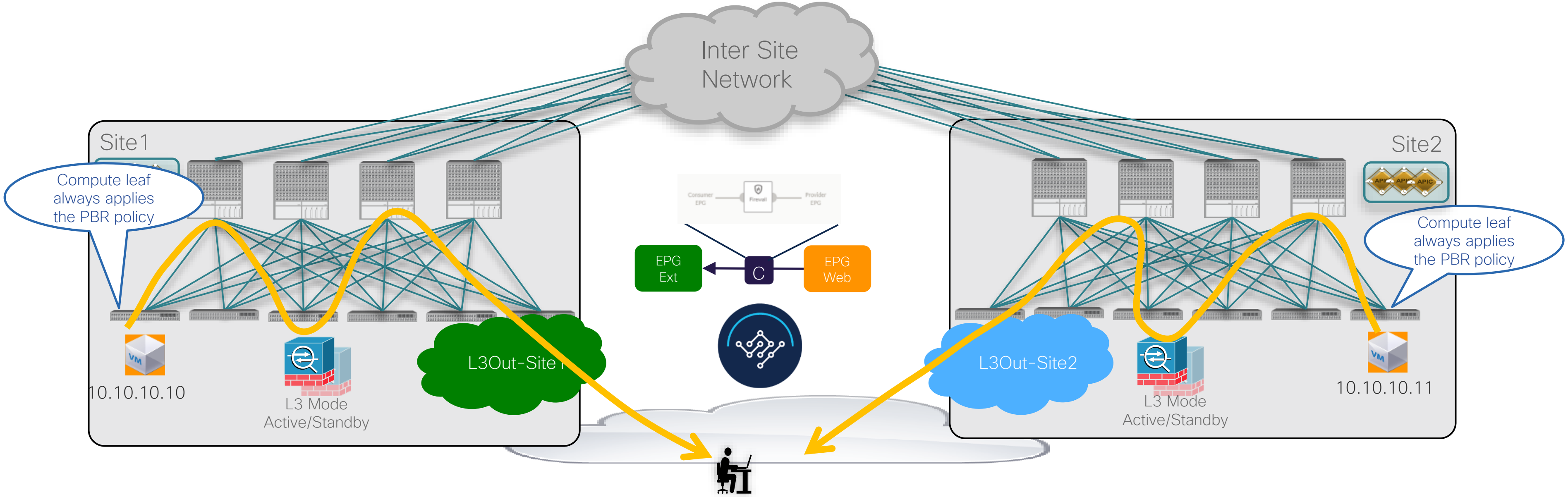


Policy Control Enforcement Preference: [✓ Enforced] Unenforced

Policy Control Enforcement Direction: Egress [✓ Ingress]

Use of Service Graph and Policy Based Redirection

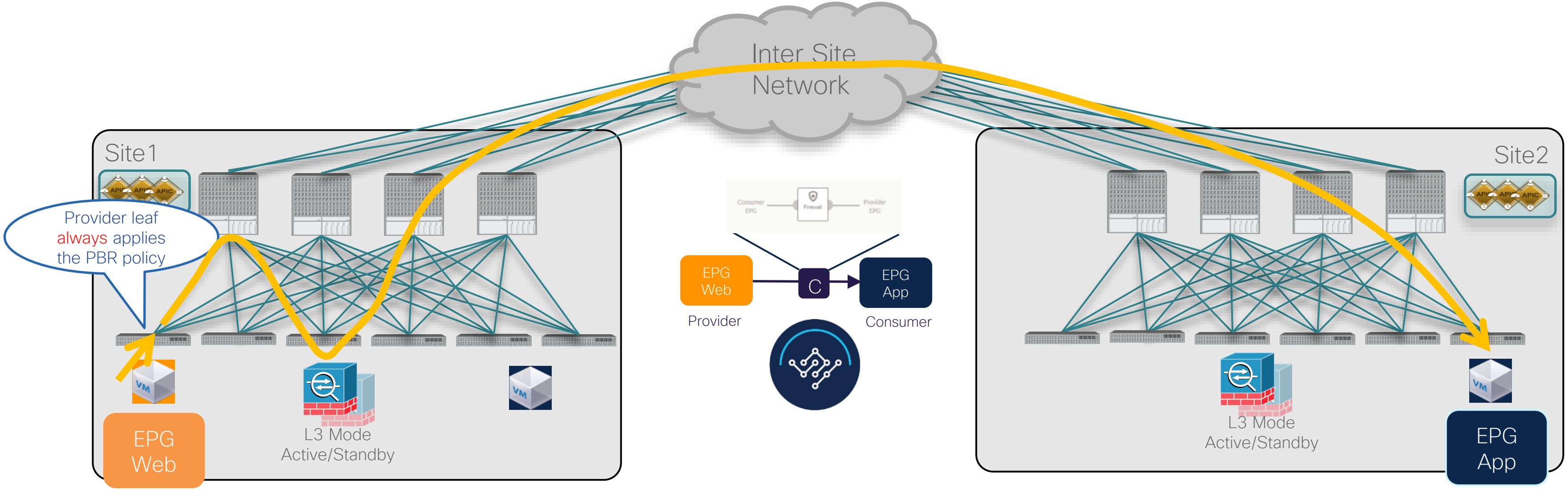
North-South Communication – Outbound Traffic



- PBR policy always applied on the same compute leaf where it was applied for inbound traffic
- Ensures the same service node is selected for both legs of the flow
- Different L3Outs can be used for inbound and outbound directions of the same flow

Use of Service Graph and Policy Based Redirection

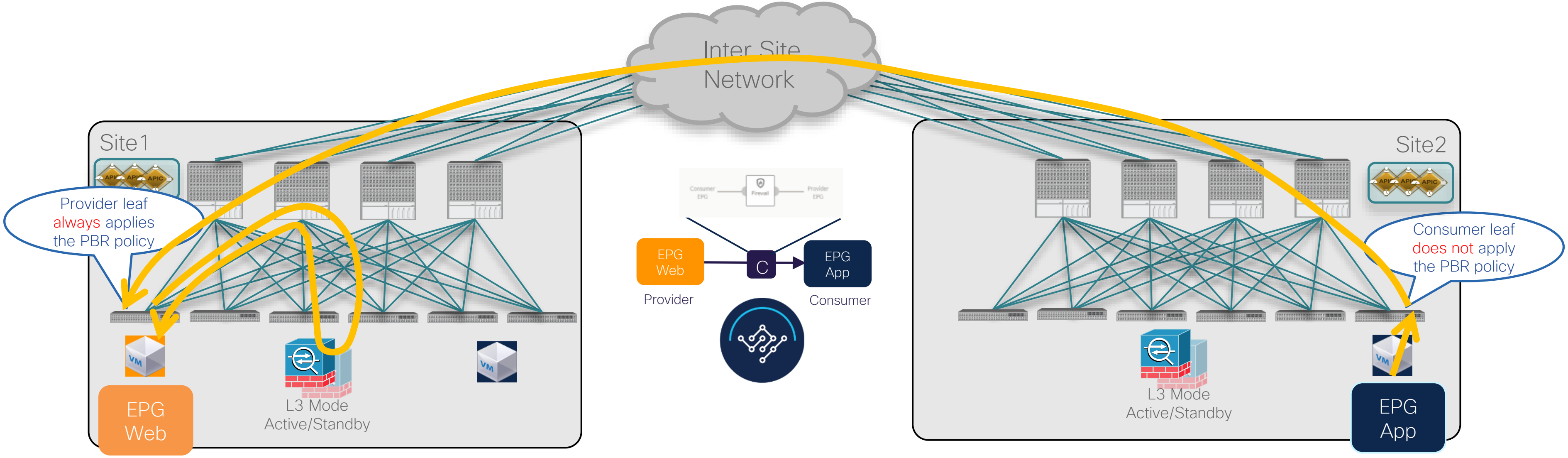
East-West Communication (2)



- EPGs can be locally defined or stretched across sites and can be part of the same VRF or in different VRFs (and/or Tenants)
- PBR policy is always applied on the leaf switch where the Provider endpoint is connected
  - The Provider leaf always redirects traffic to a local service node
  - Mandates to configure an IP Selector under the Consumer EPG

Use of Service Graph and Policy Based Redirection

East-West Communication (2)



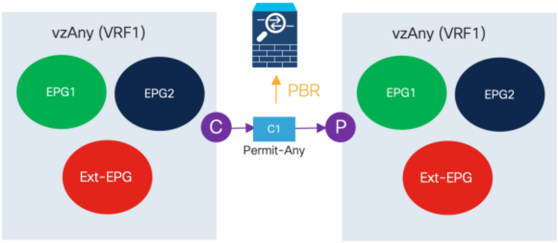
- The Consumer leaf must not apply PBR policy to ensure proper traffic stitching to the FW node that has built connection state
- Ensures both legs of the flow are handled by the same service node

ACI Multi-Site and PBR Enhancements

Orchestrator Radar (2HCY23)

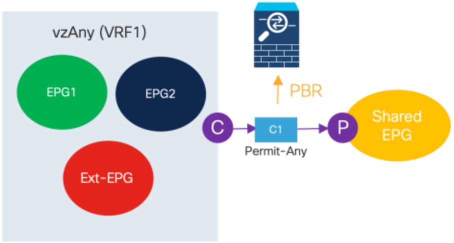
Future Supported Use Cases

Any-to-Any



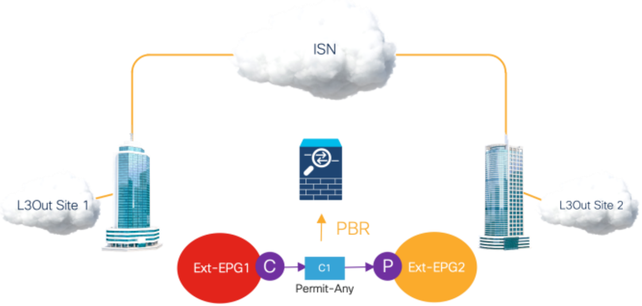
- Support only for single service node insertion
- Distributed deployment model (traffic is redirected via both local and remote service node)
- Works for both "network centric" and "app centric" designs

Many-to-One



- Support for one service node only (if the Provider is the Ext-EPG) or two service nodes (if the Provider is a regular EPG)
- Intra-VRF only
- Traffic redirected only through the service node on the provider's site
- Works for both "network centric" and "app centric" designs

Transit Intersite L3Out



- Redirect intersite transit routing traffic flows
- Traffic is redirected via both local and remote service node
- Support only for single service node insertion
- Intra-VRF and inter-VRF

ACI Multi-Site

Where to Go for More Information

- ✓ ACI Multi-Pod White Paper

  http://www.cisco.com/c/en/us/solutions/collateral/data-center-virtualization/application-centric-infrastructure/white-paper-c11-737855.html?cachemode=refresh

- ✓ 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 Multi-Site White Paper

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

- ✓ Cisco Multi-Site Deployment Guide for ACI Fabrics

  https://www.cisco.com/c/en/us/td/docs/dcn/whitepapers/cisco-multi-site-deployment-guide-for-aci-fabrics.html

- ✓ ACI Multi-Site and Service Node Integration White Paper

  https://www.cisco.com/c/en/us/solutions/collateral/data-center-virtualization/application-centric-infrastructure/white-paper-c11-743107.html

- ✓ ACI Multi-Site Training Sessions

  https://www.cisco.com/c/en/us/solutions/data-center/learning.html#~nexus-dashboard

Complete your Session Survey

- Please complete your session survey after each session. Your feedback is important.

- Complete a minimum of 4 session surveys and the Overall Conference survey (open from Thursday) to receive your Cisco Live t-shirt.

- All surveys can be taken in the Cisco Events Mobile App or by logging in to the Session Catalog and clicking the "Attendee Dashboard" at https://www.ciscolive.com/emea/learn/sessions/session-catalog.html

Continue Your Education

- Visit the Cisco Showcase for related demos.
- Book your one-on-one Meet the Engineer meeting.
- Attend any of the related sessions at the DevNet, Capture the Flag, and Walk-in Labs zones.
- Visit the On-Demand Library for more sessions at ciscolive.com/on-demand.

cisco The bridge to possible

Thank you

cisco Live!

cisco Live!

ALL IN