

CISCO *Live!*



#CiscoLive



The bridge to possible

# ACI Multi-Site Architecture and Deployment

## Part 2

Max Ardica, Distinguished Engineer  
@maxardica  
BRKDCN-2480b



#CiscoLive

# 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

Webex spaces will be moderated by the speaker until June 17, 2022.



<https://ciscolive.ciscoevents.com/ciscolivebot/#BRKDCN-2480a>

# 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.)



# 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

BRKDCN-2480a

# ACI Multi-Site Control and Data Plane



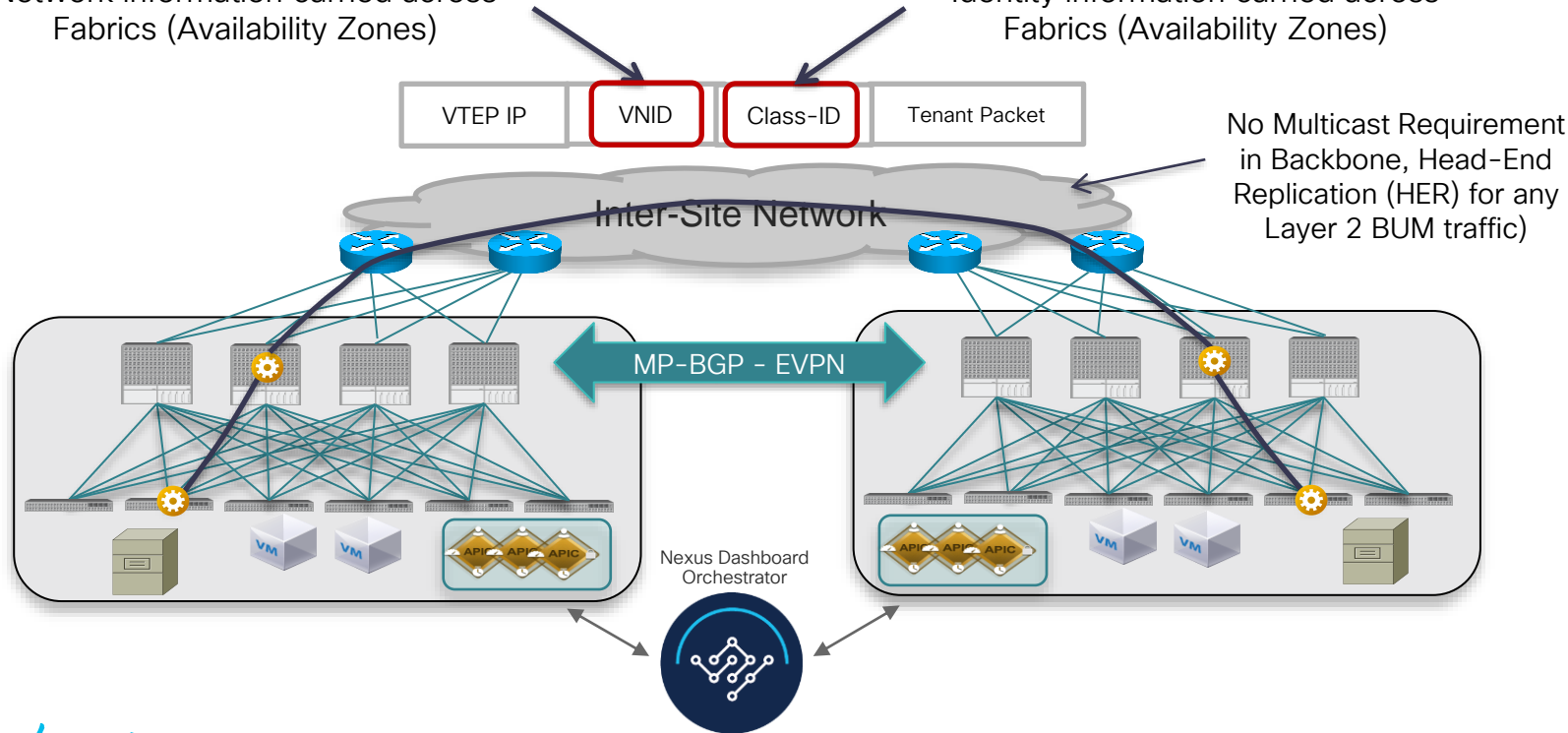
# Namespace Normalization and Shadow Objects

# ACI Multi-Site

## Network and Identity Extended between Fabrics

Network information carried across Fabrics (Availability Zones)

Identity information carried across Fabrics (Availability Zones)

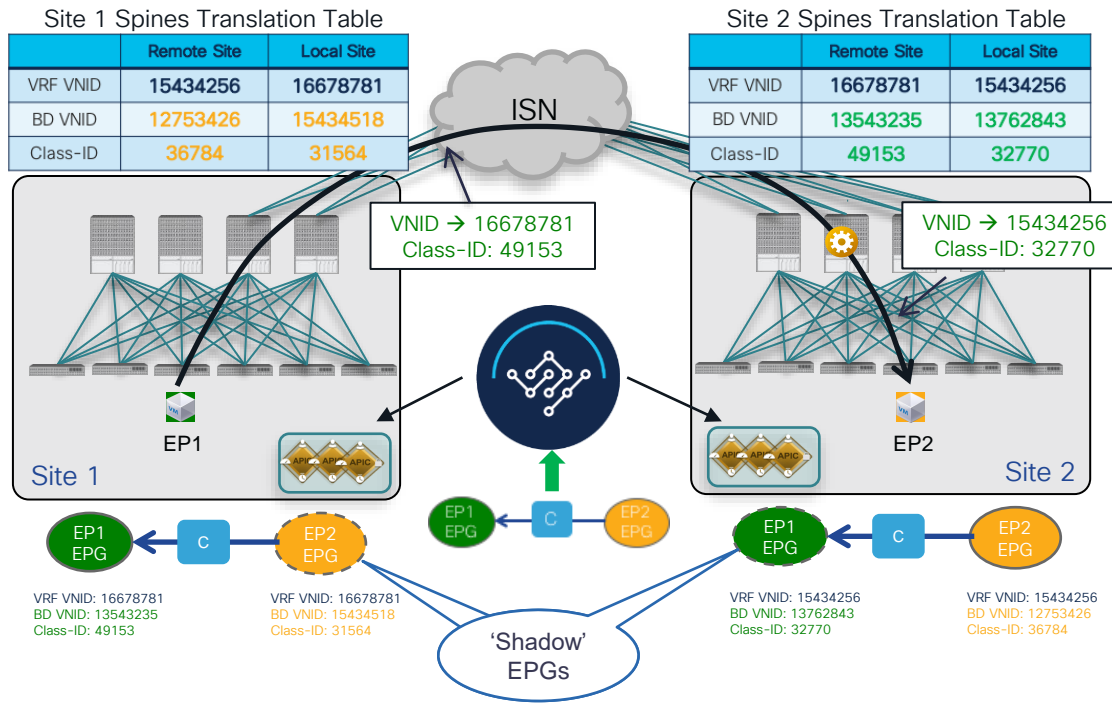




# ACI Multi-Site

## Inter-Site Policies and Spines' Translation Tables

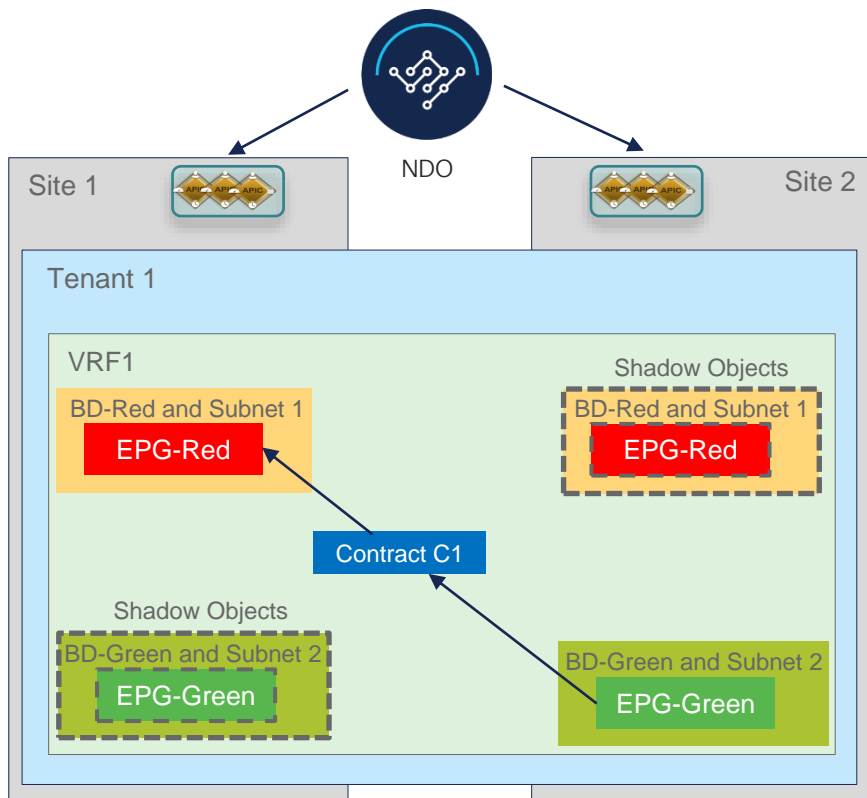
- Inter-Site policies defined on the ACI Nexus Dashboard Orchestrator are pushed to the respective APIC domains
  - End-to-end policy consistency
  - Creation of 'Shadow' EPGs to locally represent the policies
- Inter-site communication requires the installation of translation table entries on the spines (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



# Per Bridge Domain Behavior

# ACI Multi-Site

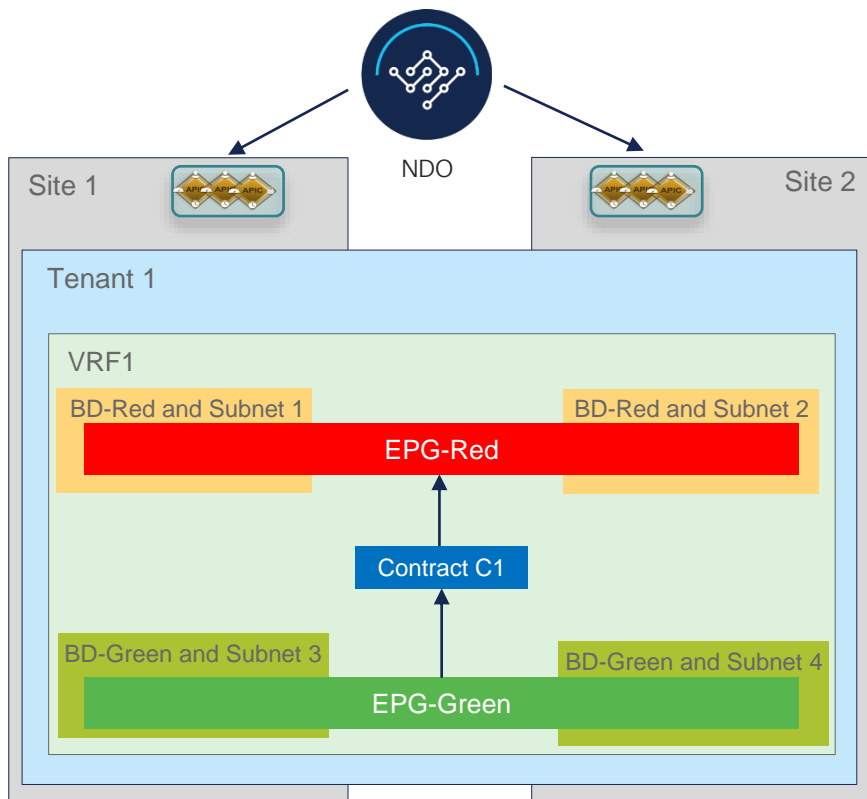
## Intra-VRF Layer 3 Communication across Sites



- Stretch tenant/VRF across ACI fabrics
  - BDs/EPGs defined as site local objects
- ☐ L2 Stretch ← BD-Red and BD-Green
- Configuration of policy between EPGs in separate fabrics to enable intra-VRF Layer 3 inter-site connectivity
  - Creation of shadow BDs/EPGs in remote site(s)

# ACI Multi-Site

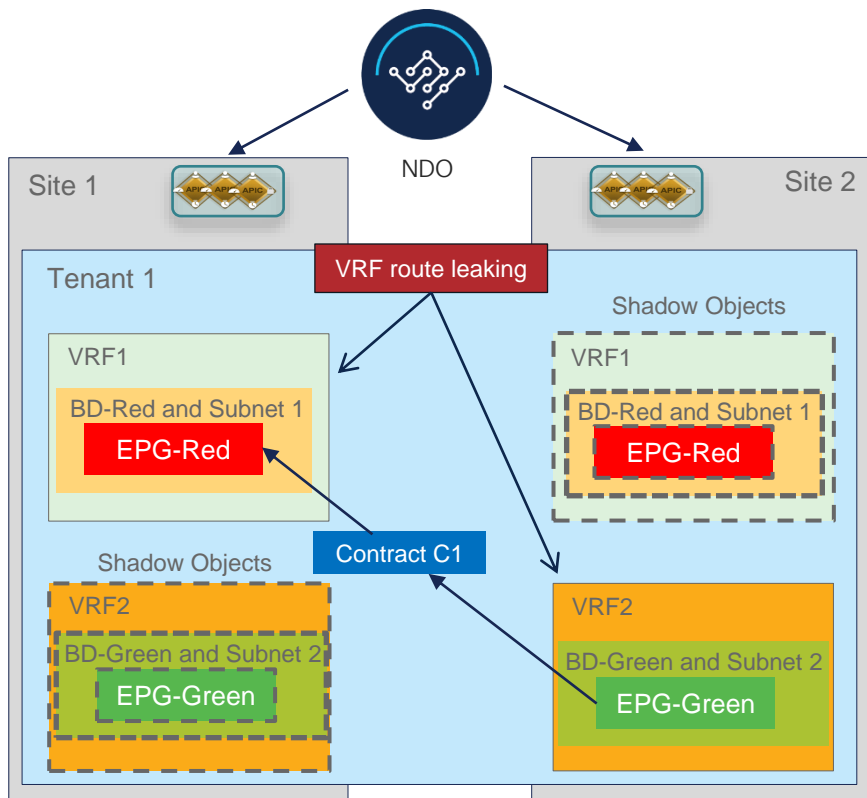
## Intra-VRF Layer 3 Communication across Sites (Stretched EPGs)



- Stretch tenant/VRF and EPG across ACI fabrics
  - BDs are also defined in a template associated to both sites but with “L2 Stretch” option disabled
    - Allows to assign a different IP subnet to the BDs in separate sites
- ☐ L2 Stretch ← BD-Red and BD-Green
- Layer 3 communication between sites for intra-EPG and inter-EPG flows
  - No requirement for shadow objects creation

# ACI Multi-Site

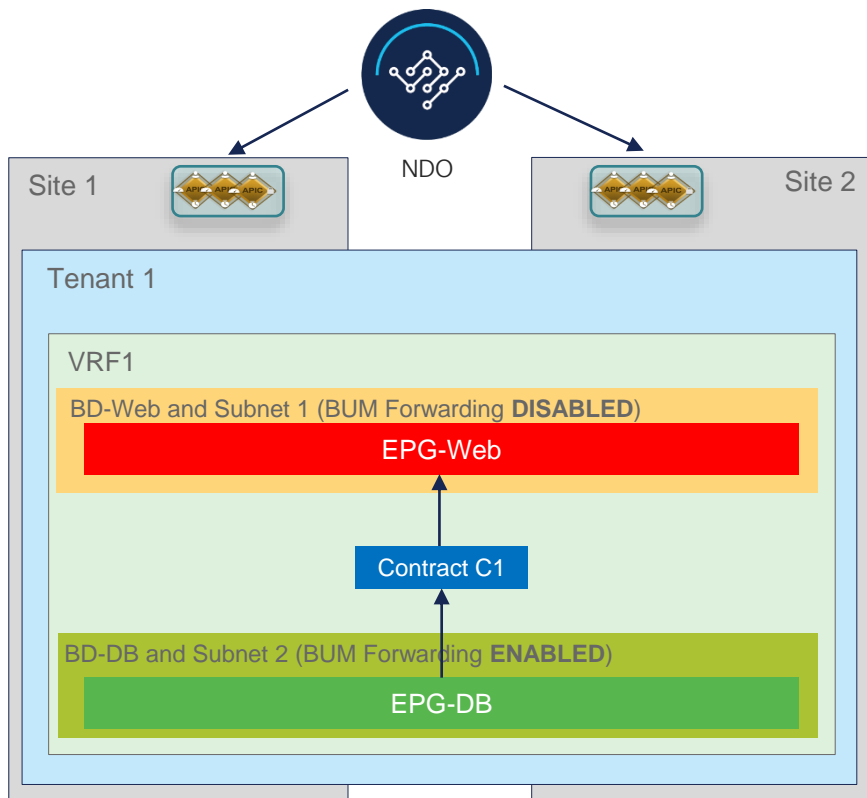
## Inter-VRF Layer 3 Communication across Sites (Shared Services)



- VRF/BD/EPG locally defined in each site
- Inter-VRF communication across sites (shared services)
- Route leaking between VRFs (requires subnet configured under the provider EPG)
- Supported within the same stretched tenant but also between different tenants
- Creation of shadow VRFs/BDs/EPGs in remote site(s)

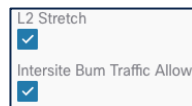
# ACI Multi-Site

## Layer 2 Extension across Sites



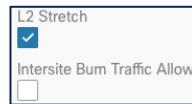
- Stretch tenant/VRF but also BDs/EPGs across ACI fabrics
- BUM forwarding can be controlled on a BD basis

Required only for establishing pure L2 communication across sites (DB clustering using L2 multicast or broadcast, for example)



← BD-DB

IP mobility (and live migration) can be supported without enabling BUM forwarding

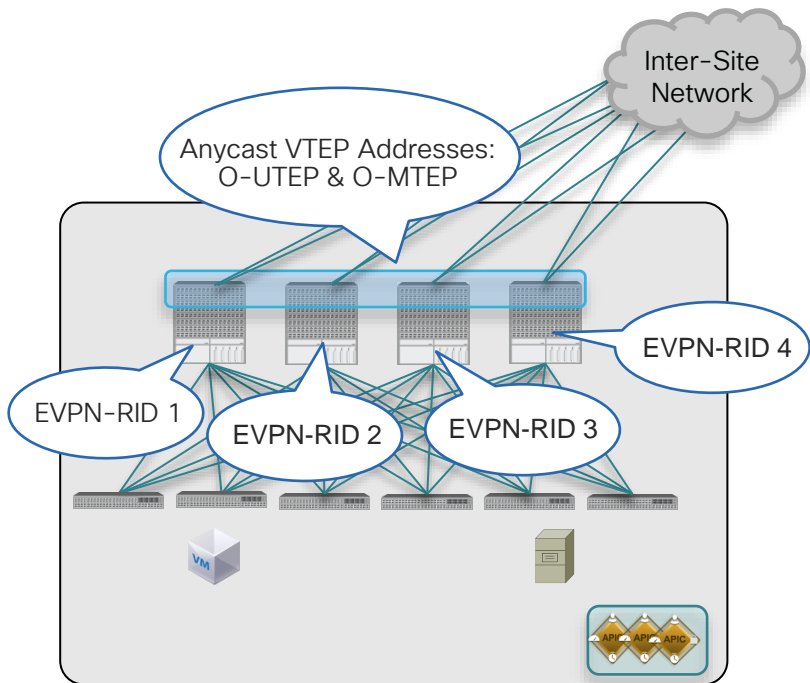


← BD-Web

# 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



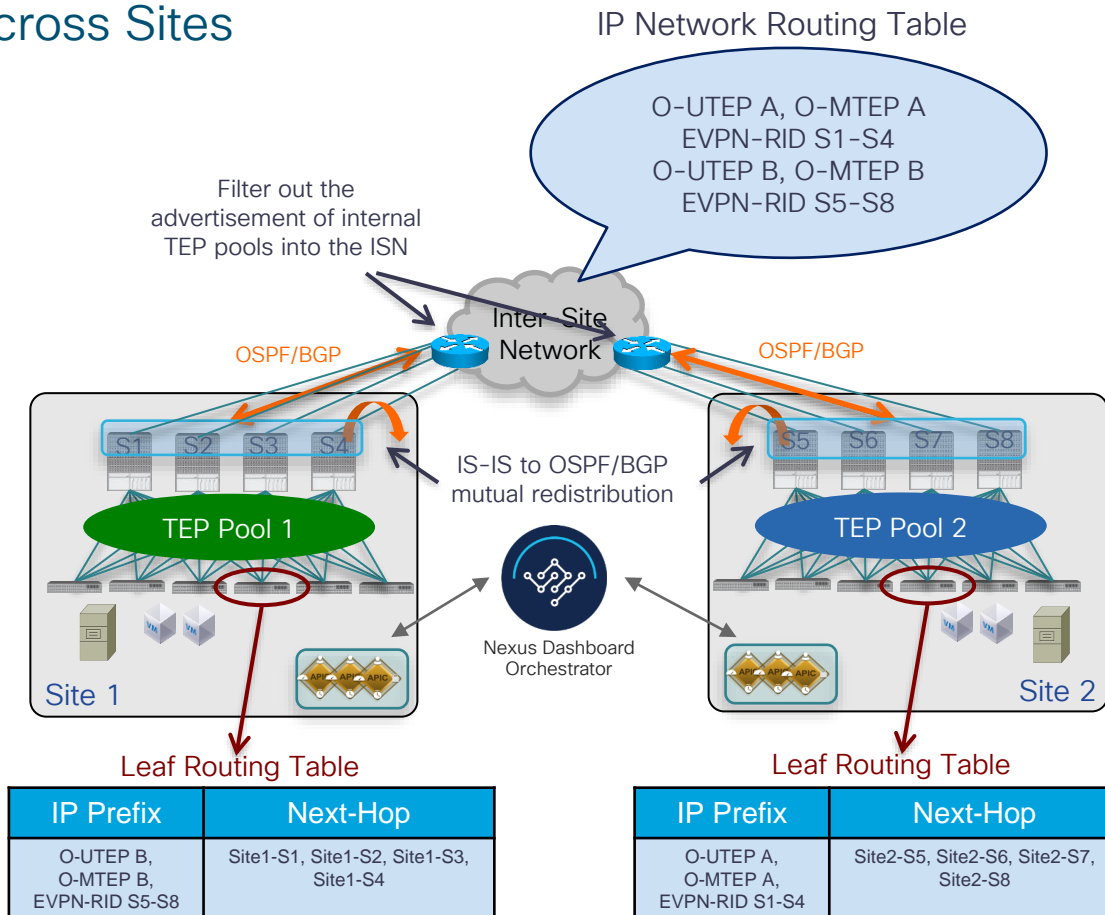
# ACI Multi-Site

## 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



# ACI Multi-Site

## Inter-Site MP-BGP EVPN Control Plane

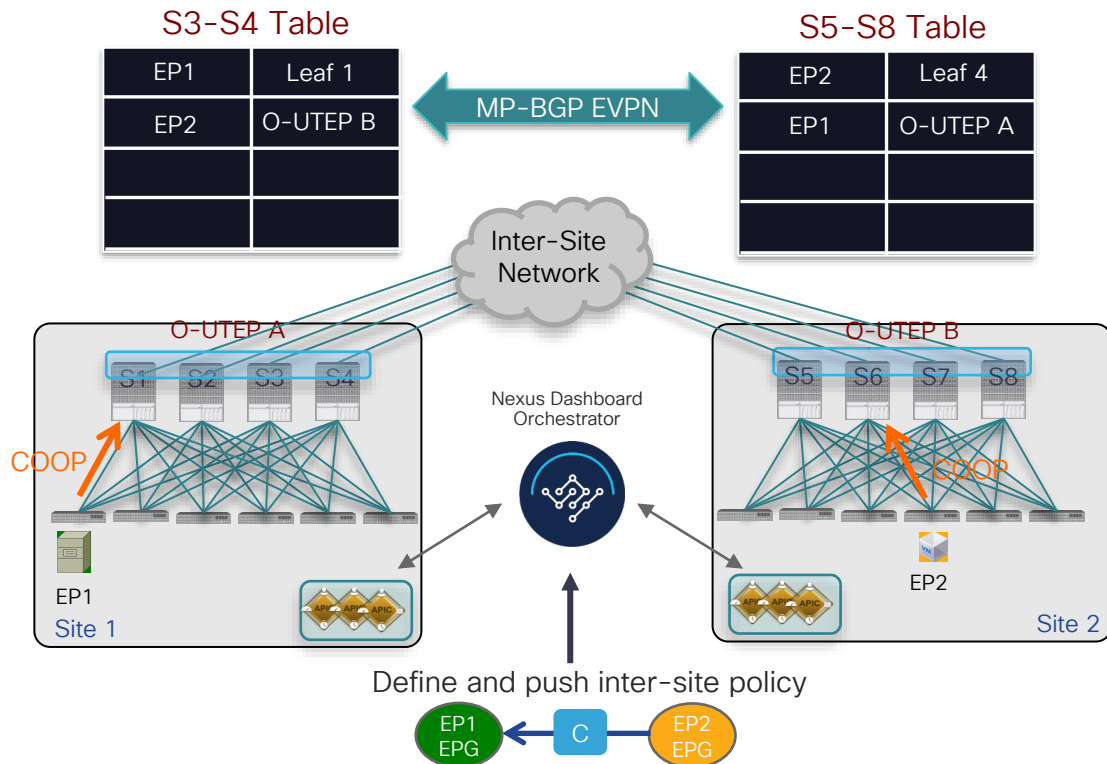
- MP-BGP EVPN used to communicate Endpoint (EP) information across Sites

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

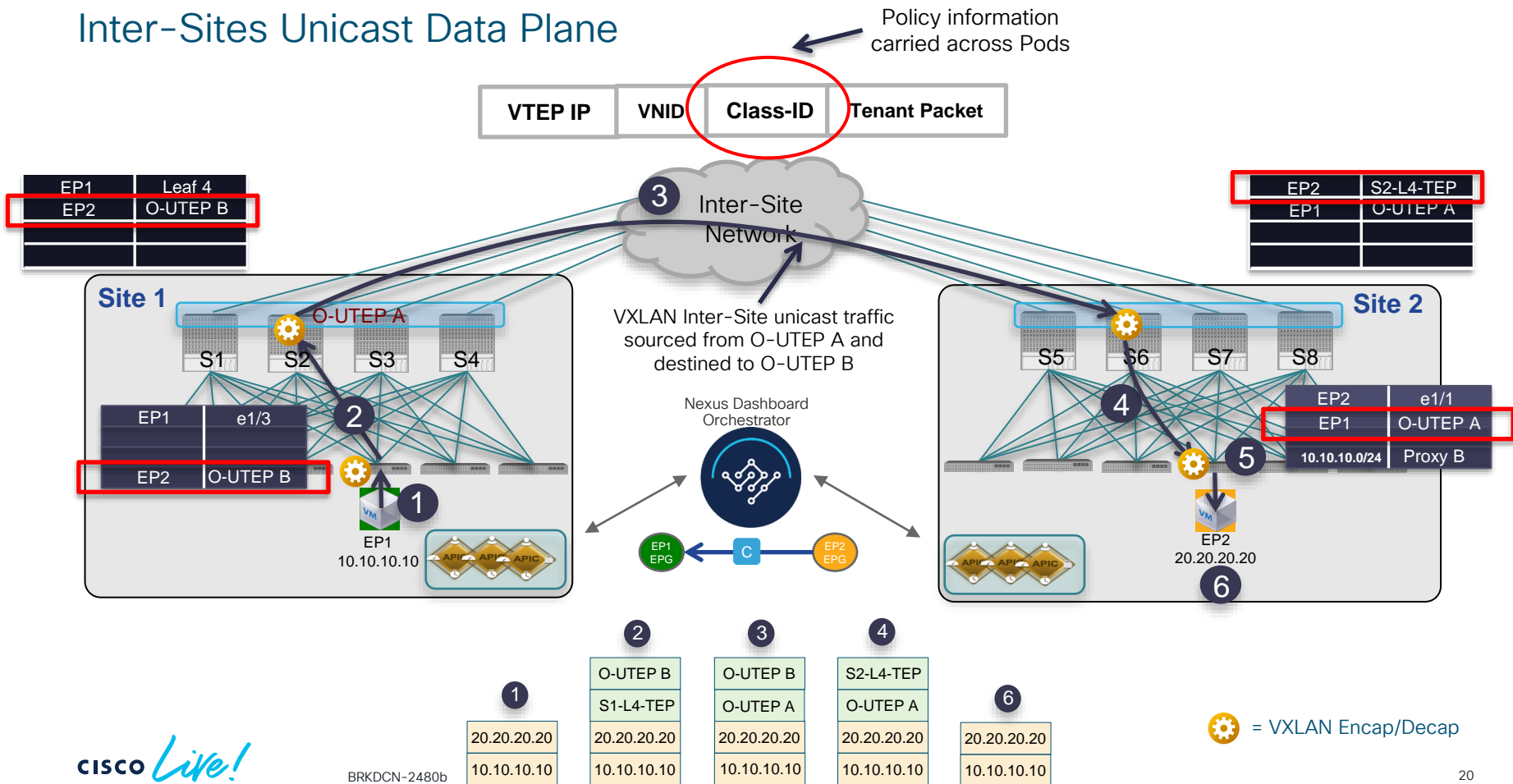
Host routes are exchanged across sites **only** if there is a cross-site contract requiring communication between endpoints



# Data Plane Communication across Sites

# ACI Multi-Site

## Inter-Sites Unicast Data Plane



# Layer 3 Only Communication across Sites

# ACI Multi-Site

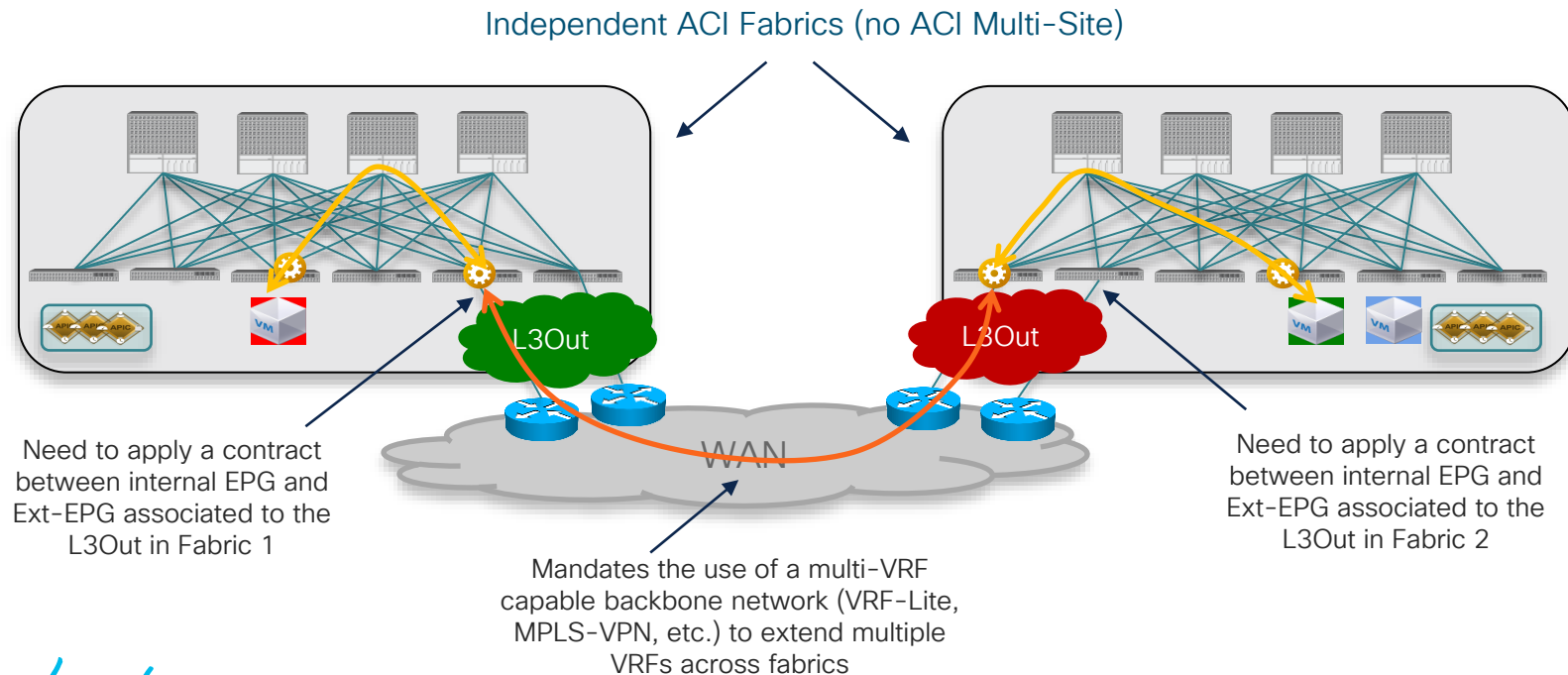
## Never Mixing ISN and L3Out Paths for E-W Routed Communication

- Routed communication between different ACI fabrics can be established in two ways:
  1. The first and recommended way is by using VXLAN tunnels established via the Inter-Site Network
  2. Through the L3Out connections deployed in each fabric and leveraging the connectivity services of the external Layer 3 network domain
- Mixing those two approaches is highly discouraged because it may lead to unexpected dropping of traffic

# ACI Multi-Site

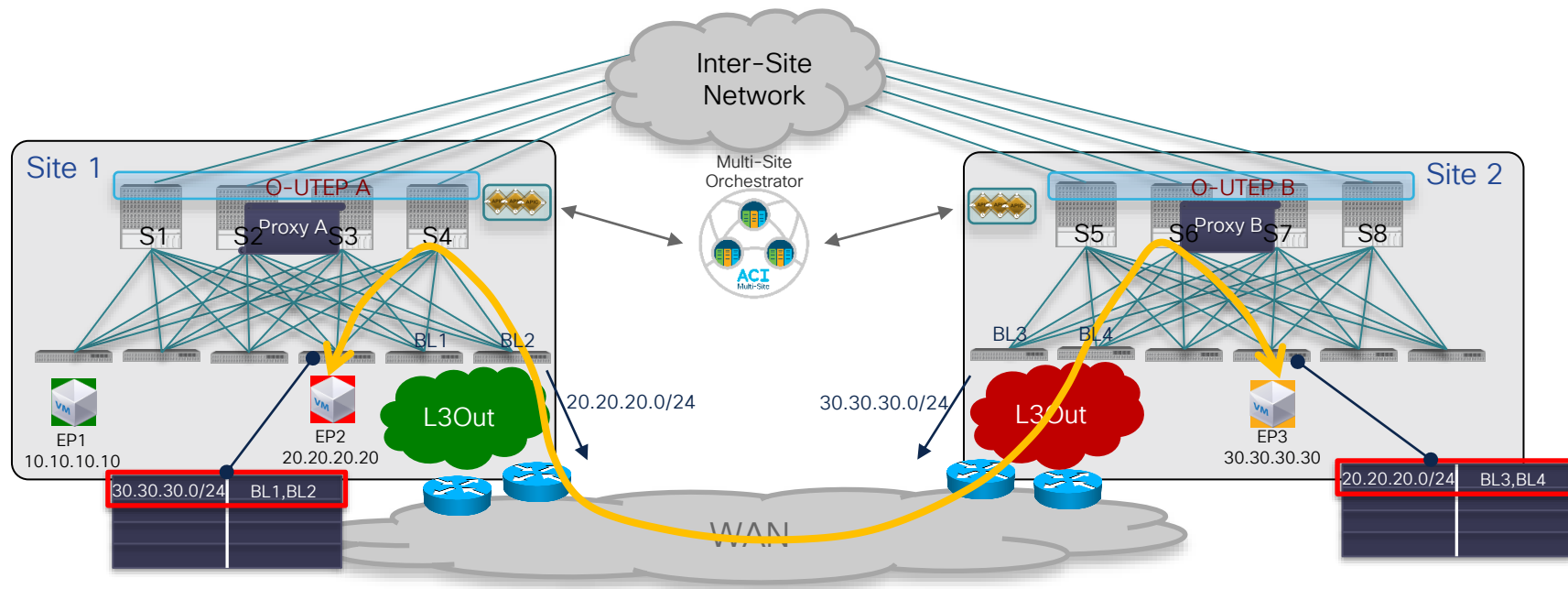
## L3 Only across Sites

Why not using just normal routing across independent fabrics?



# ACI Multi-Site

## Never Mixing ISN and L3Out Paths for E-W Routed Communication

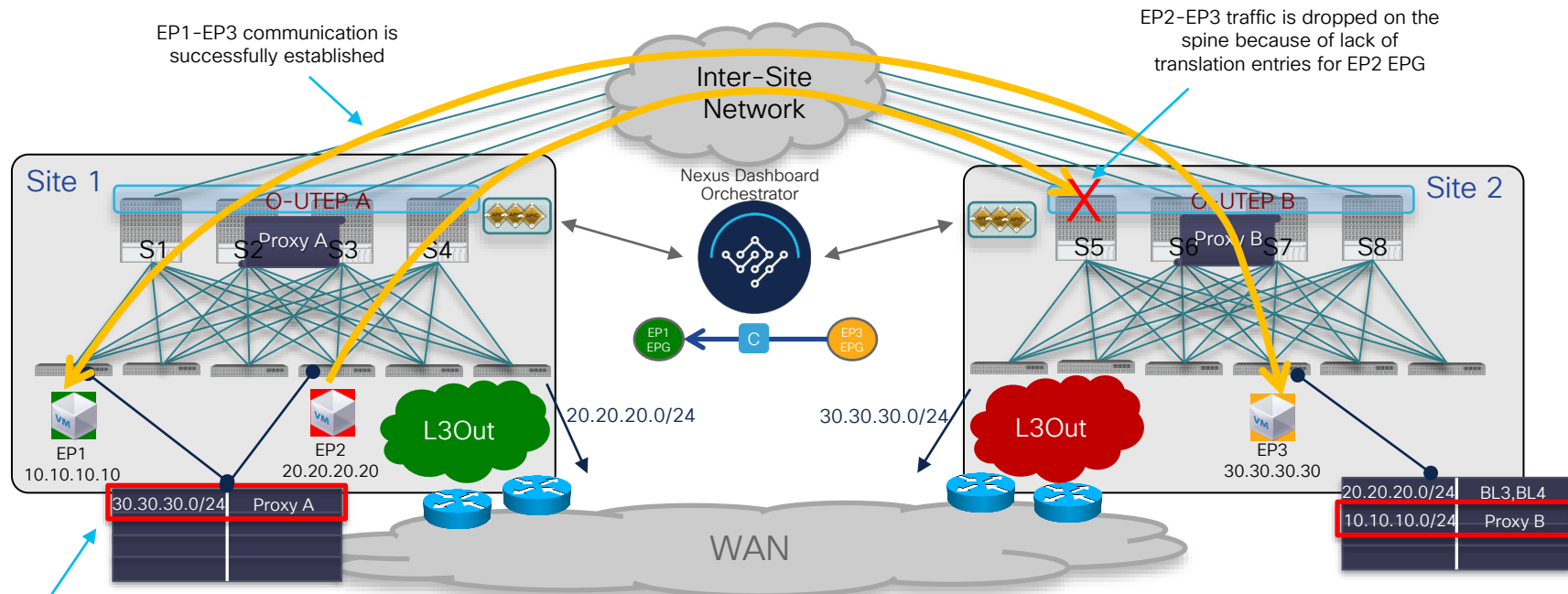


- 1 Initial Condition: EP2 is communicating to EP3 via the L3Out path



# ACI Multi-Site

## Never Mixing ISN and L3Out Paths for E-W Routed Communication



EP3 EPG subnet route pointing to spine proxy is installed on all the leaf where the VRF is deployed

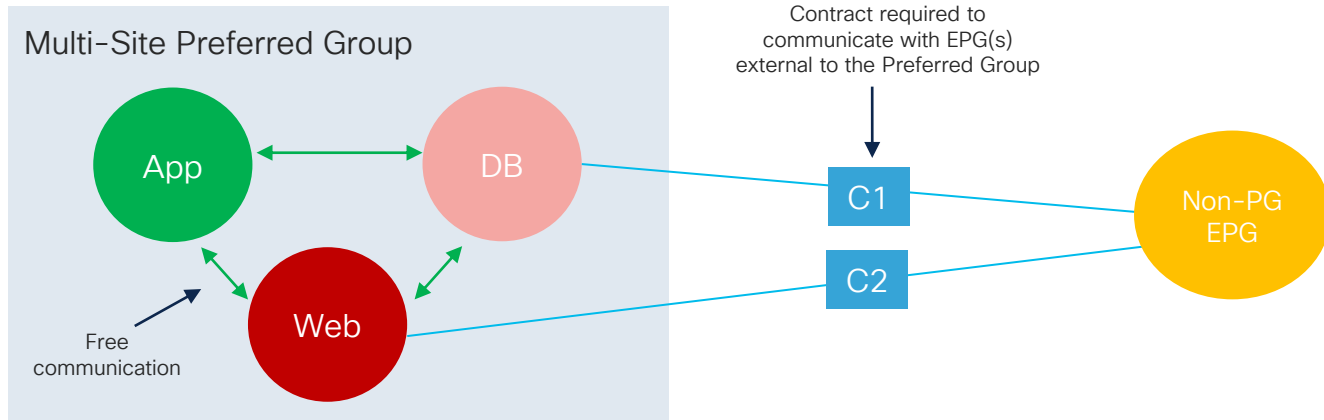
- 2 A contract between EP1 and EP3 EPG is configured on MSO to allow E-W communication via the ISN

# Simplify Policy Application

Preferred Group and vzAny

# ACI Multi-Site

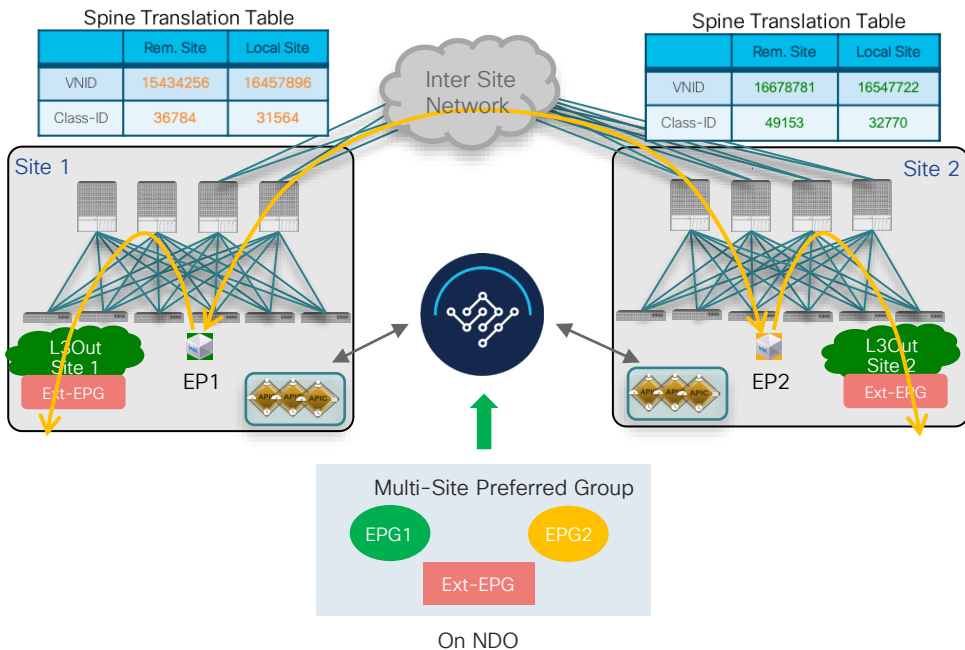
## Simplify Policy Enforcement: Preferred Groups



- "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



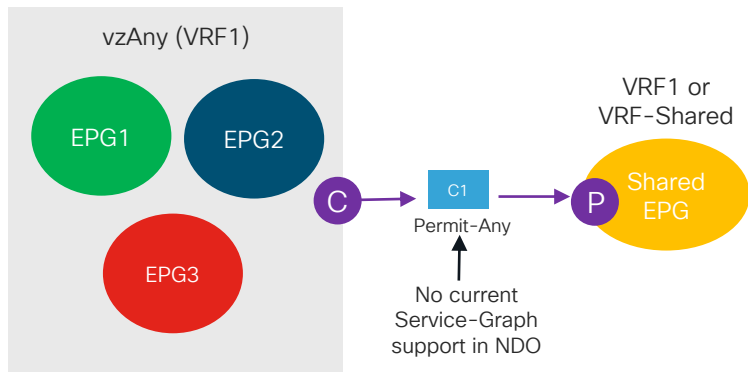
- Adding internal EPGs and External EPGs (associated to L3Outs) to the Preferred Group allows to enable free east-west and north-south connectivity
- When adding the Ext-EPG to the Preferred Group:
  - Can't use 0.0.0.0/0 for classification, needs more specific prefixes
  - As workaround it is possible 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 3.7(1))

# Simplify Policy Enforcement

## vzAny Support

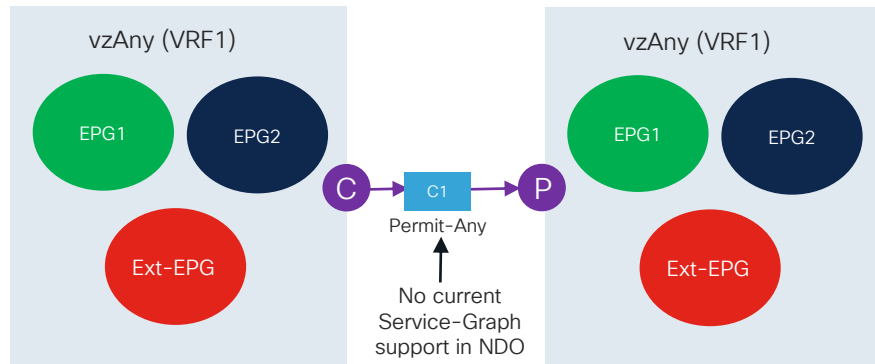
What is vzAny? Logical object representing all the EPGs in a VRF

### 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

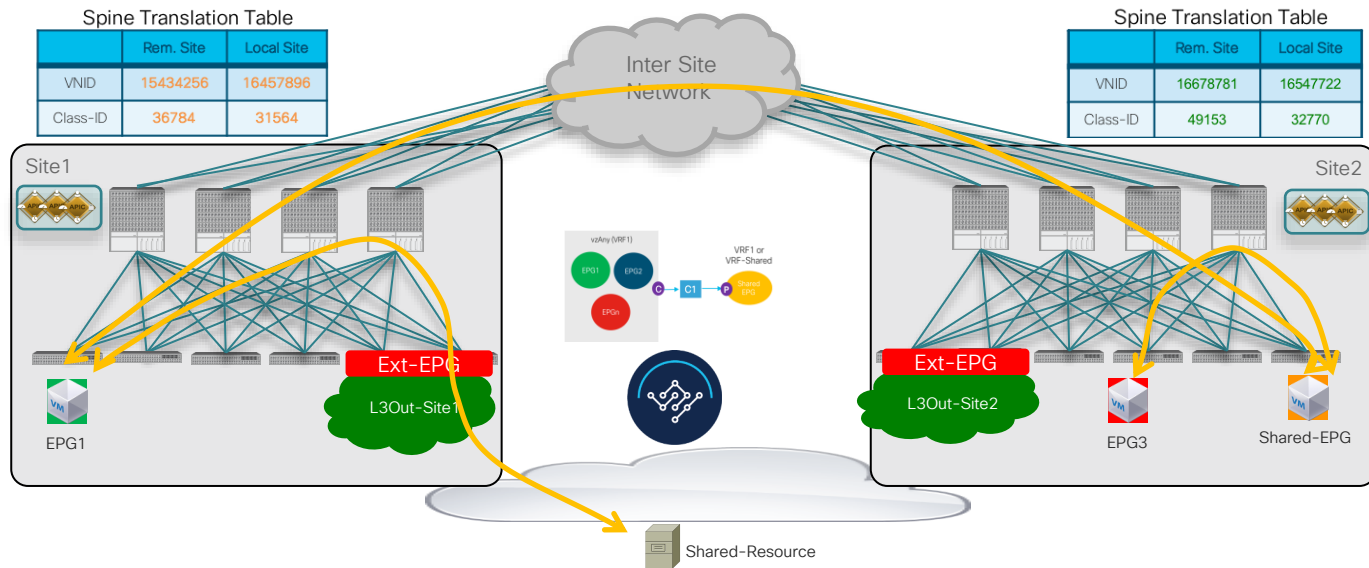
### Use case 2: Enable free communication inside a VRF



- 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”

# ACI Multi-Site and vzAny

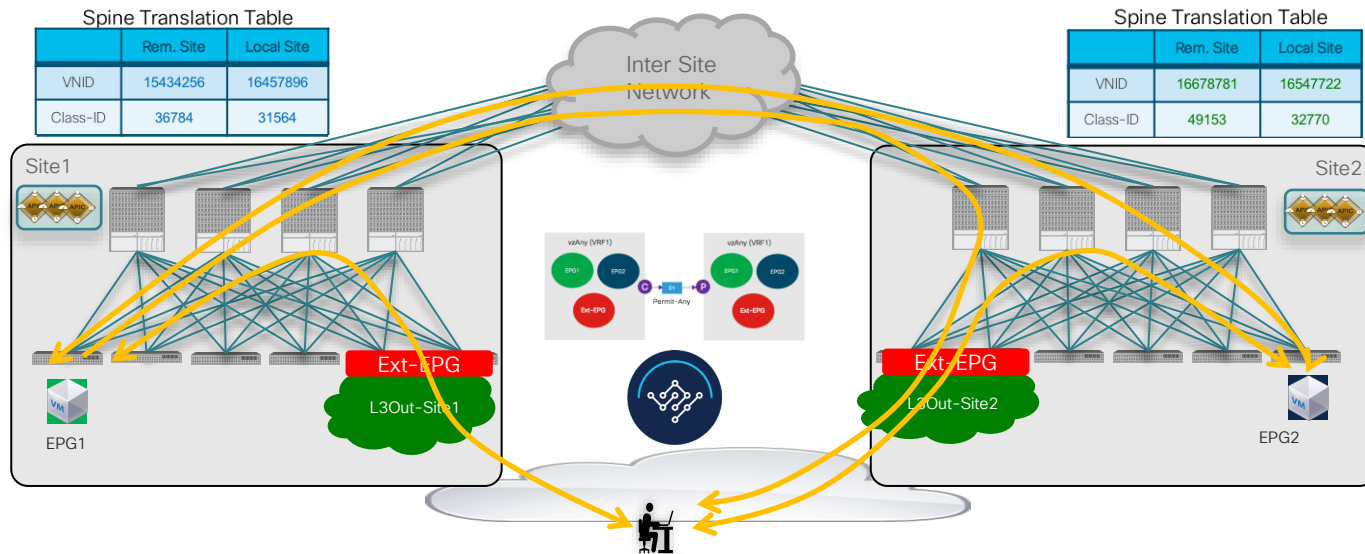
## Many-to-One Communication (Shared Services)



- Proper translation entries are created on the spines of both fabrics to enable east-west communication
- Supported also for Shared Services behind an L3Out

# ACI Multi-Site and vzAny

## Enable Inter-Site Free Communication Inside a VRF



- 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

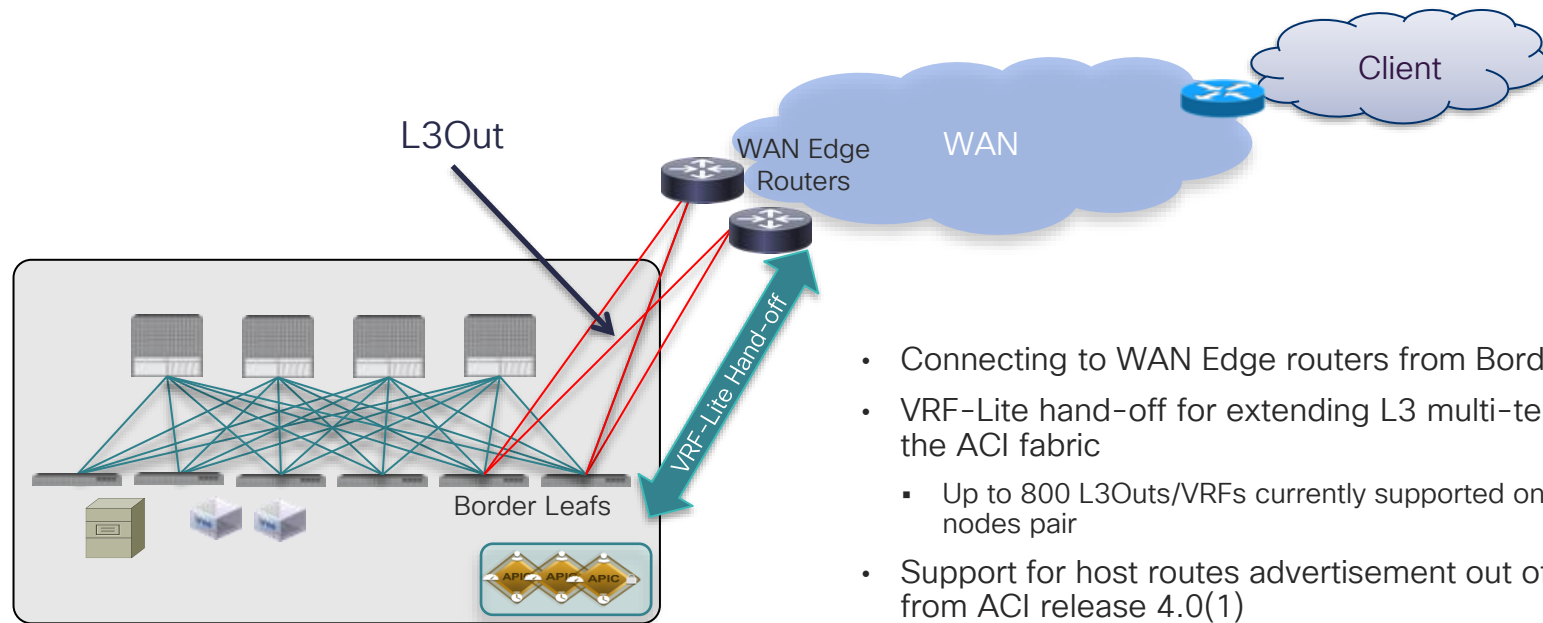




# Different Types of L3Outs

# Connecting to the External Layer 3 Domain

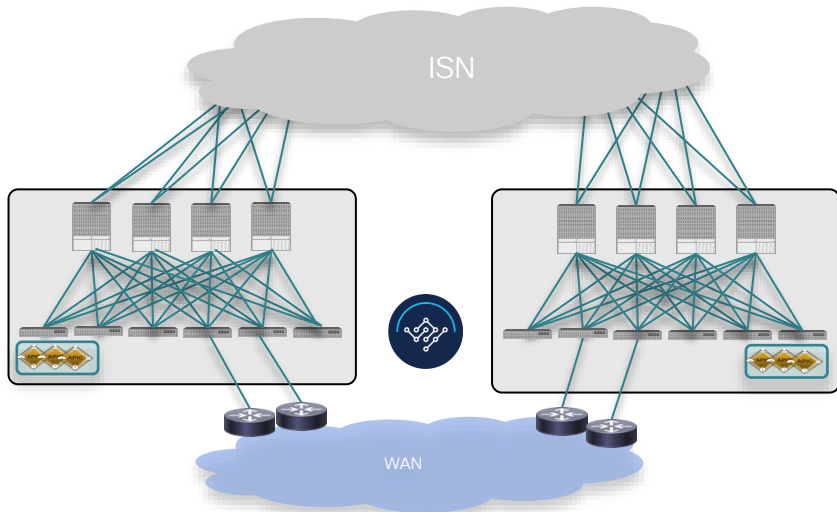
## 'Traditional' L3Outs on the BL Nodes (Recommended Option)



- 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

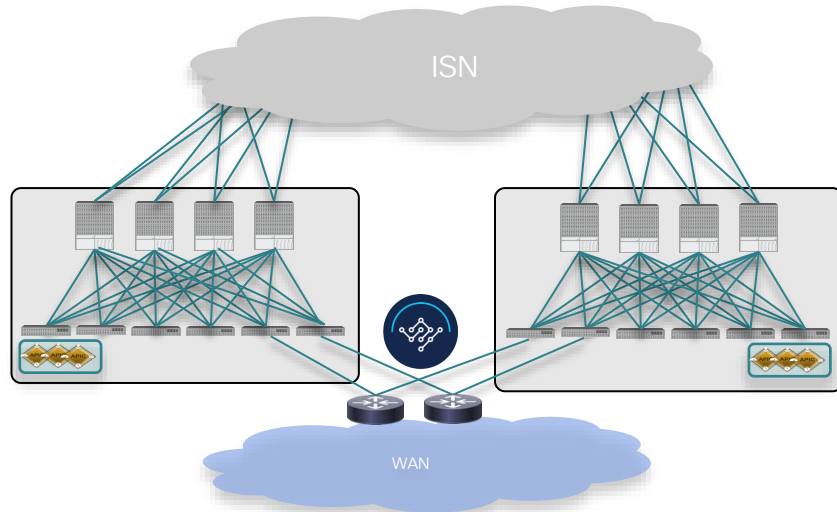
# ACI Multi-Site and Border Leaf L3Outs Deployment Options

## Dedicated pair of WAN edge routers



- BLs on each ACI site connect to a separate pair of WAN edge routers for communication with the WAN
- Most common deployment model for ACI fabrics geographically dispersed

## Shared pair of WAN edge routers

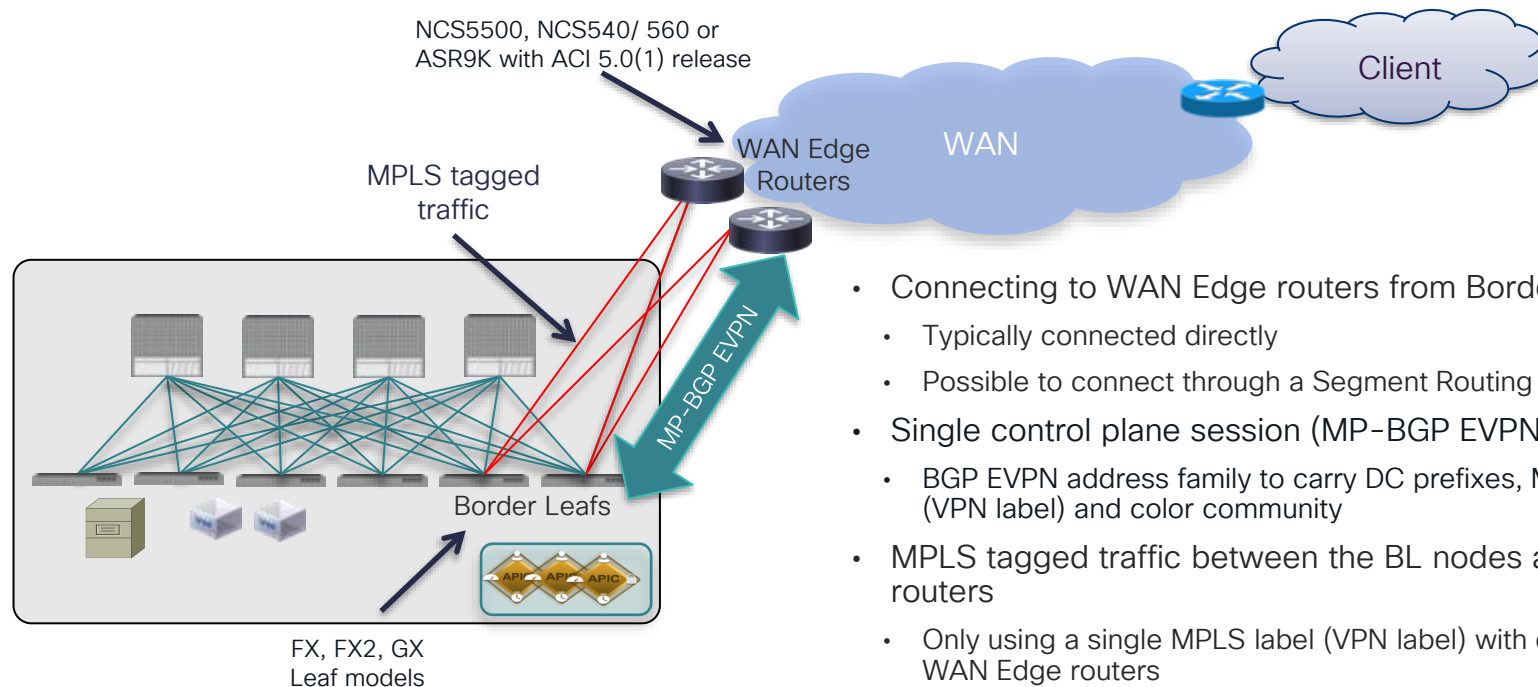


- BLs of different sites connect to a common pair of WAN edge routers for communication with the WAN
- Typical deployment model when Multi-Site is used for scaling up the fabric in a single DC location

# Connecting to the External Layer 3 Domain

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

ACI 5.0(1)  
Release

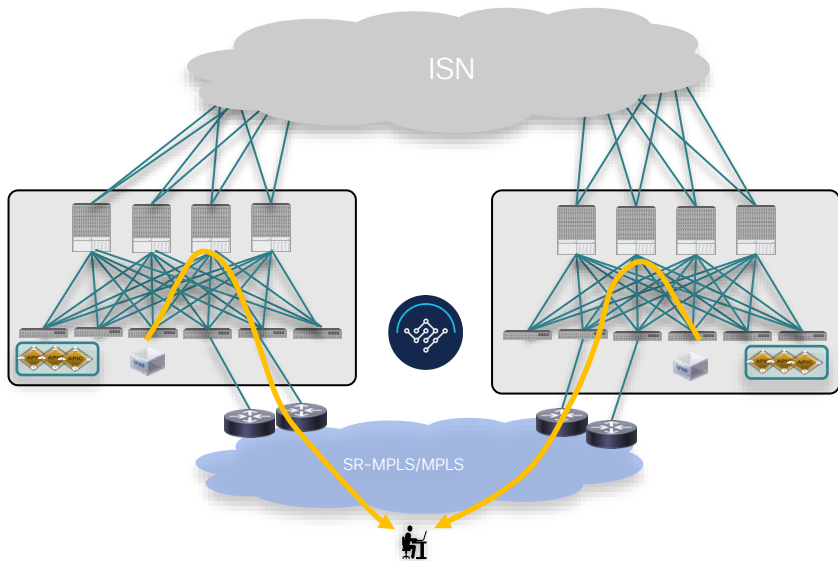


- Connecting to WAN Edge routers from Border Leaf nodes
  - Typically connected directly
  - Possible to connect through a Segment Routing enabled network
- Single control plane session (MP-BGP EVPN) for all tenant VRFs
  - BGP EVPN address family to carry DC prefixes, MPLS label for VRF (VPN label) and color community
- MPLS tagged traffic between the BL nodes and the WAN Edge routers
  - Only using a single MPLS label (VPN label) with directly connected WAN Edge routers
  - VPN + SR labels used with a SR enabled network in the middle
- No current support for Layer 3 Multicast communication

# ACI Multi-Site and Border Leaf L3Outs Deployment Options

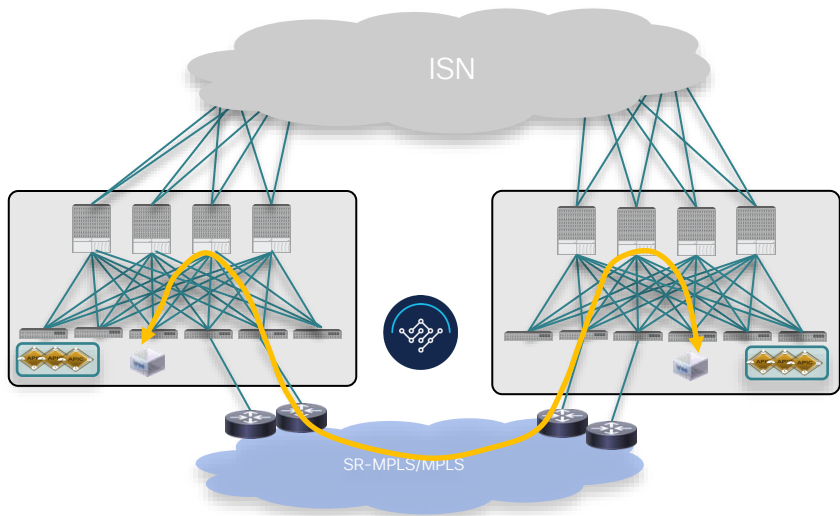
ACI 5.0(1)  
Release

## SR-MPLS/MPLS hand-Off for N-S Traffic



- SR-MPLS/MPLS Hand-Off used in each site to communicate with external resources (through an MPLS enabled backbone)

## SR-MPLS/MPLS hand-Off for Inter-DC Flows

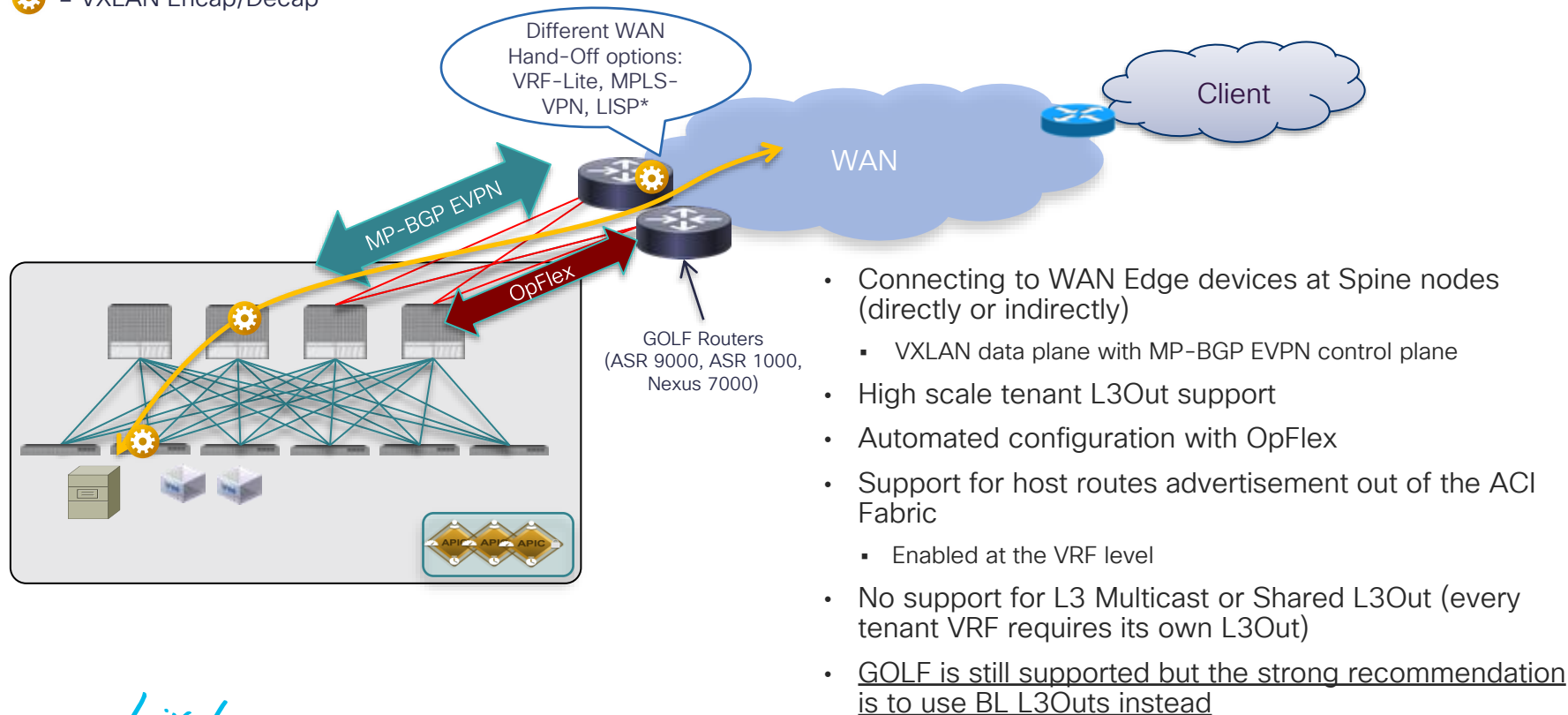


- No EPGs, BDs or VRFs are stretched across ACI sites  
Different VRFs in each site are mapped to a common VRF in the MPLS enabled backbone
- Use of SR-TE/Flex-Algo for inter-DC flows
- WAN teams can monitor inter-DC flows using existing IP/SR based monitoring tools

# Connecting to the External Layer 3 Domain

## 'GOLF' L3Outs (VRF High Scale Use Cases)

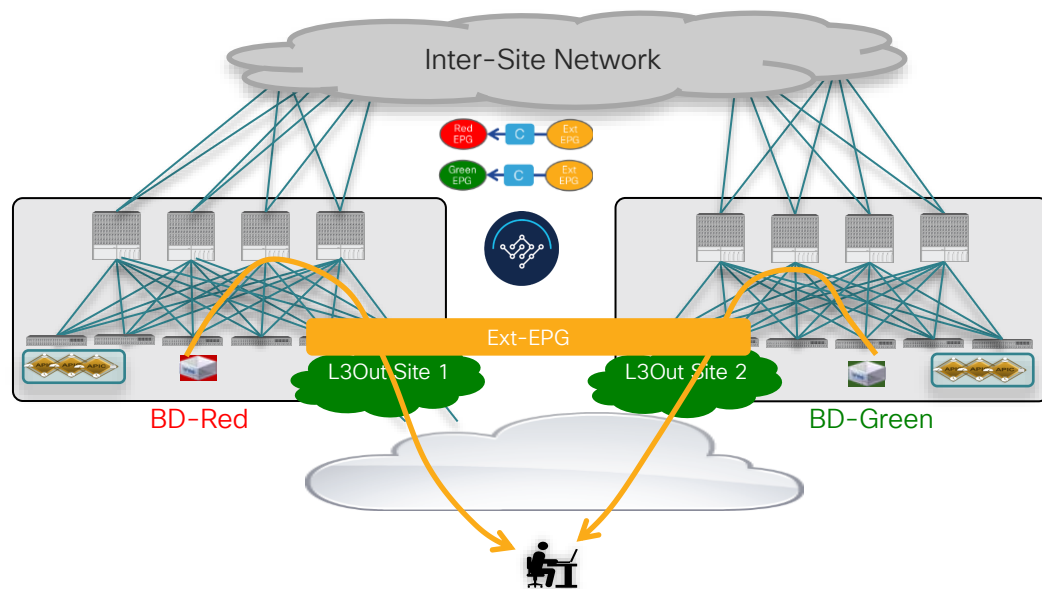
⚙️ = VXLAN Encap/Decap



# 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)

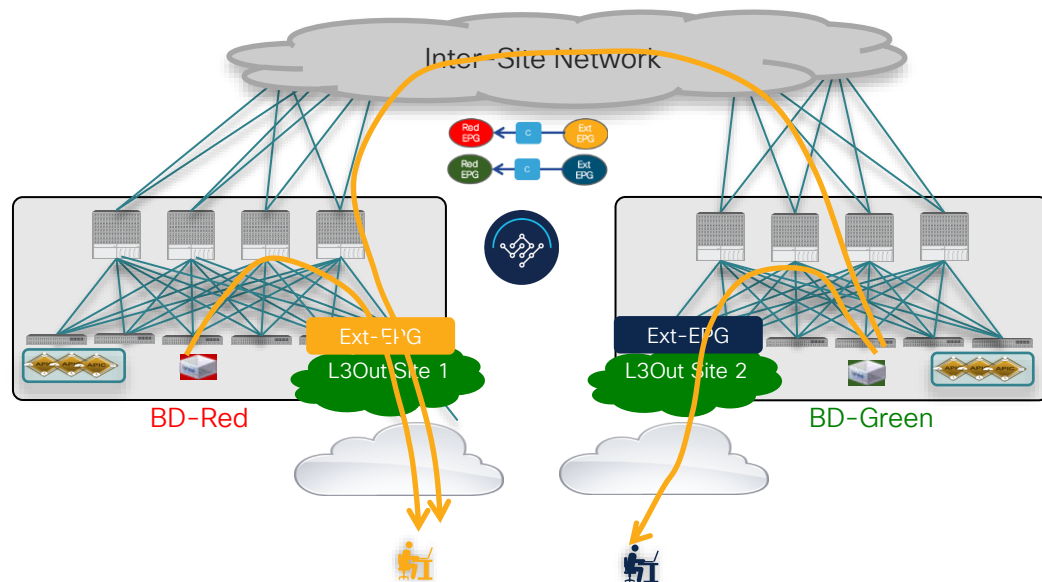
Simplifies the policy definition and external traffic classification

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?

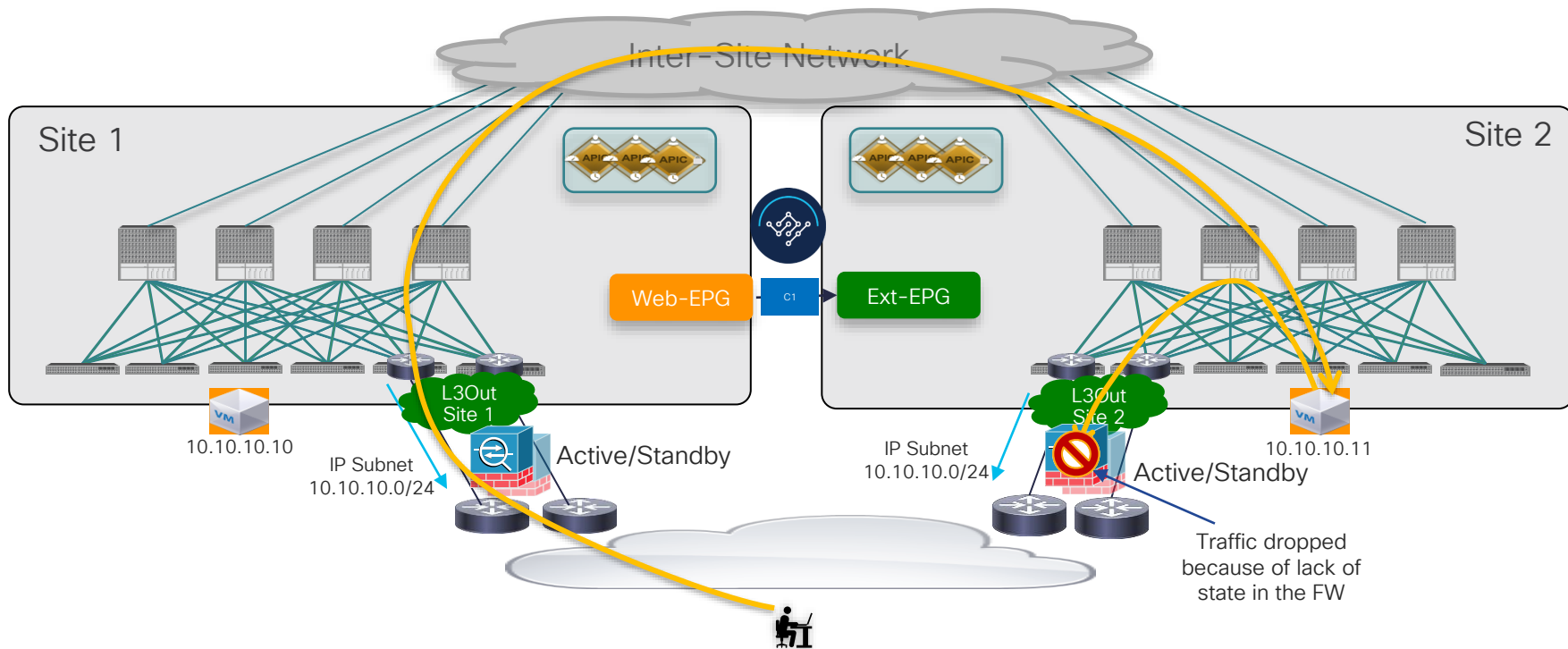


- Separate Ext-EPGs can be defined in templates mapped to separate sites (non stretched objects)
  - Each Ext-EPG can be mapped to the local L3Out in the “global” or “site level” section of the template configuration
- Allows to apply different policies to each Ext-EPGs at different time
- Can still 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

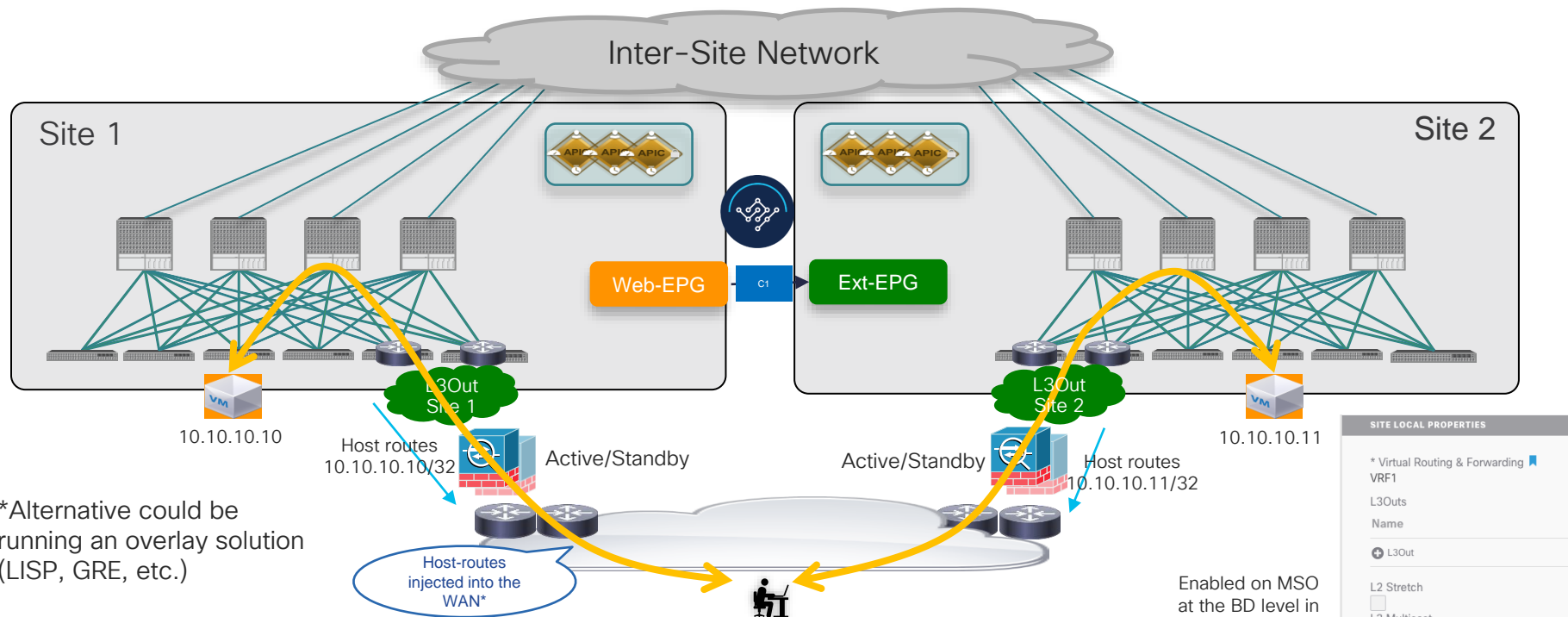
## Endpoints Normally Use Local L3Outs for Outbound Traffic



# Solving Asymmetric Routing Issues

## Use of Host-Routes Advertisement

ACI 4.0(1)  
Release



- Ingress optimization requires host-routes advertisement on the L3Out
  - Native support on ACI Border Leaf nodes available from ACI release 4.0(1)
  - Supported also on GOLF L3Outs (enabled at the VRF level)

BRKDCN-2480b

SITE LOCAL PROPERTIES

\* Virtual Routing & Forwarding

VRF1

L3Outs

Name

+ L3Out

L2 Stretch

☐ L3 Multicast

☐ L2 UNKNOWN UNICAST proxy

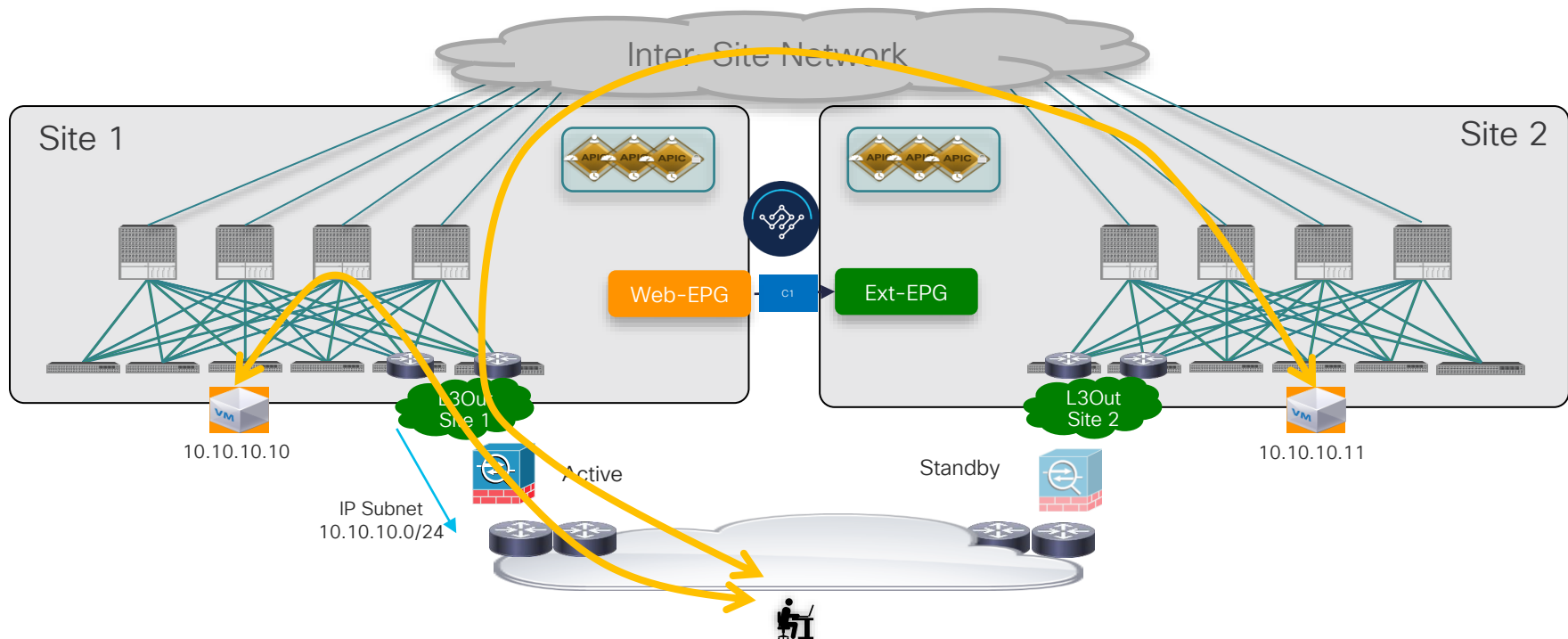
Host Route

☒ Subnets

# Solving Asymmetric Routing Issues

## Use of Active/Standby FW Pair Deployed across Sites

ACI 4.2(1)  
Release



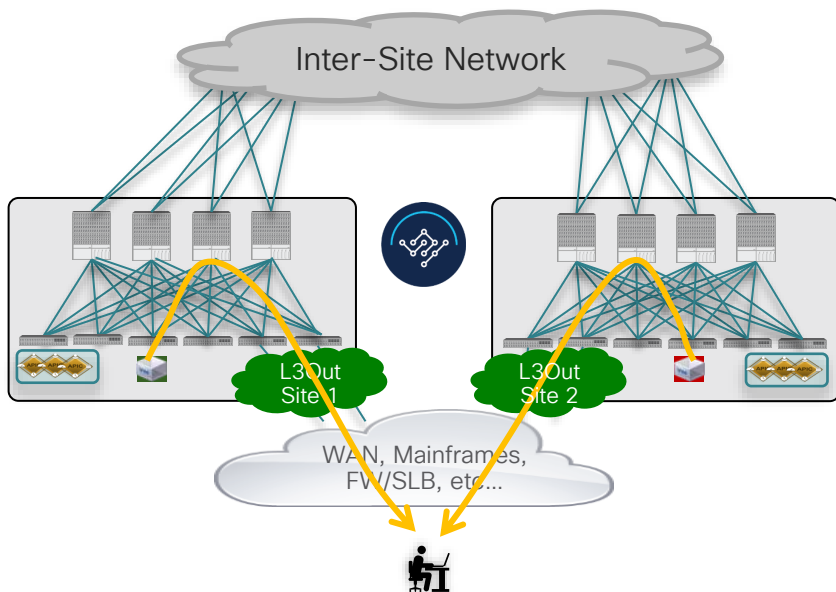
- Inbound and outbound flows are forced through the site with the active perimeter FW node
  - Common scenario in a Multi-Pod deployment, less desirable with Multi-Site
- Requires Intersite L3Out support (ACI release 4.2(1))

# Intersite L3Out Support

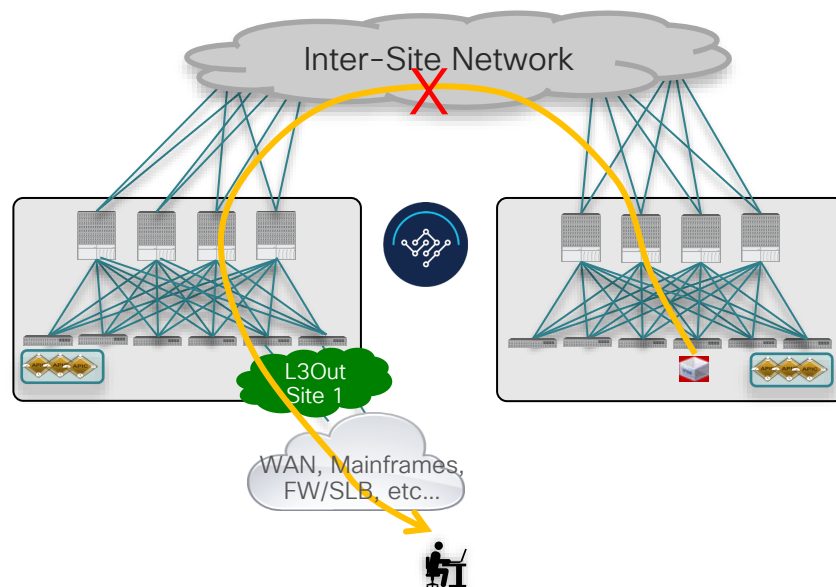
# Problem Statement

## Behavior before ACI Release 4.2(1)

### Supported Design



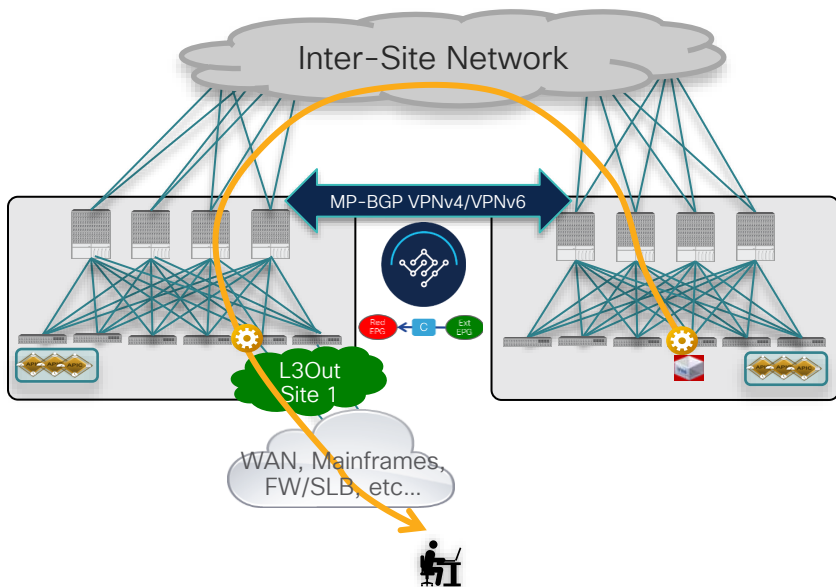
### Not Supported Design



Note: the same consideration applies to both Border Leaf L3Outs and SR-MPLS L3Outs

# ACI Multi-Site and L3Out

## Support of Intersite L3Out

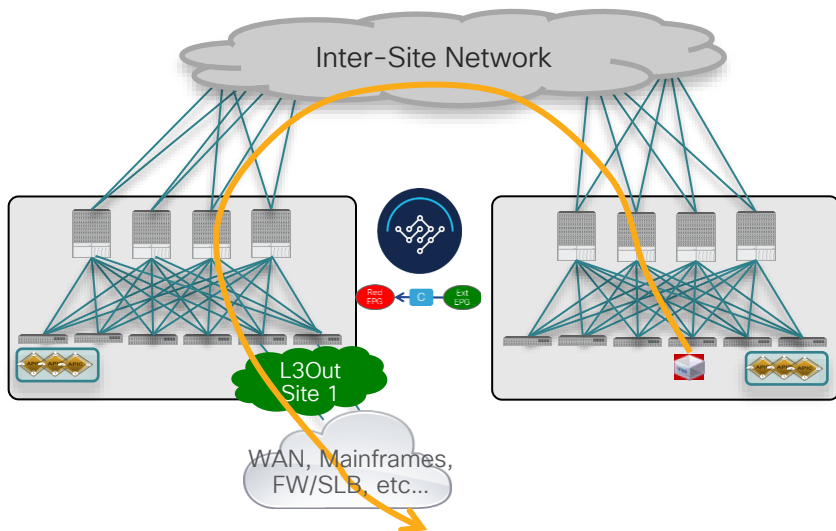


- Starting with ACI Release 4.2(1) it is possible for endpoints in a site to send traffic to resources (WAN, Mainframes, FWs/SLBs, etc.) accessible via a remote L3Out connection
- External prefixes are exchanged across sites via MP-BGP VPNv4/VPNv6 sessions between spines
- Traffic will be directly encapsulated 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)

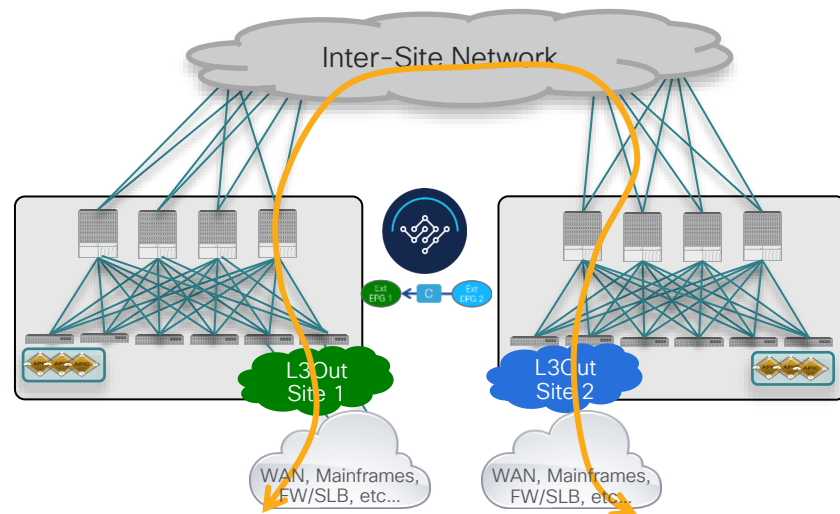


# ACI Multi-Site and Intersite L3Out Supported Scenarios

ACI 4.2(1)  
Release



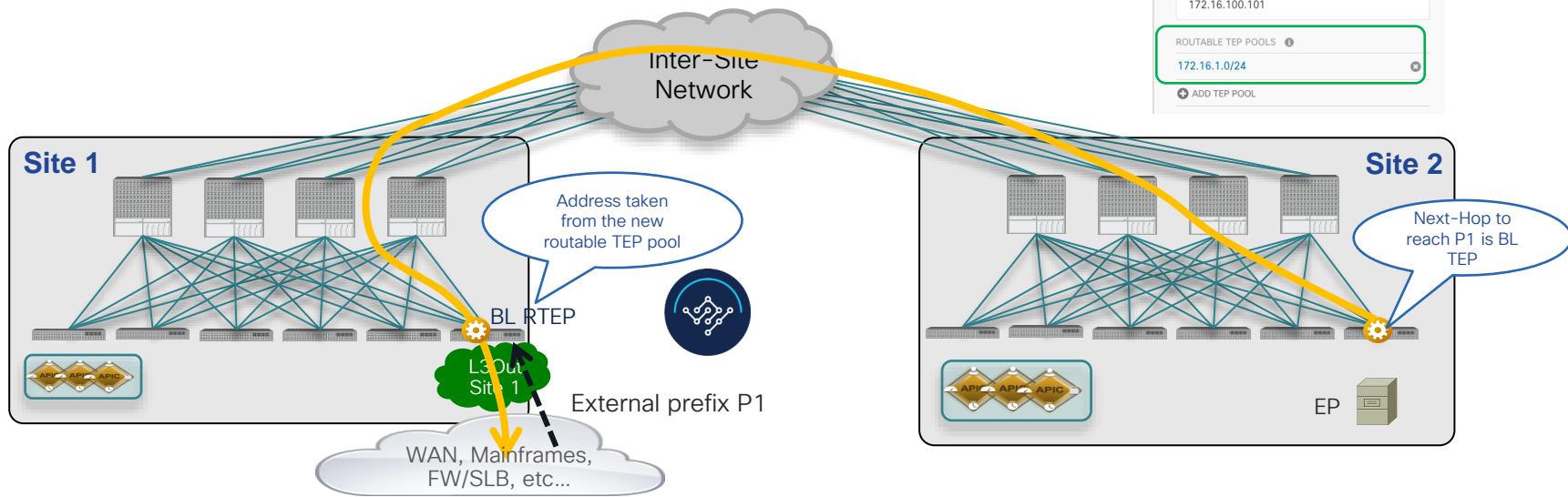
- Endpoint to remote L3Out communication (intra-VRF)
- Endpoint to remote L3Out communication (inter-VRF)



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

# ACI Multi-Site and Intersite L3Out

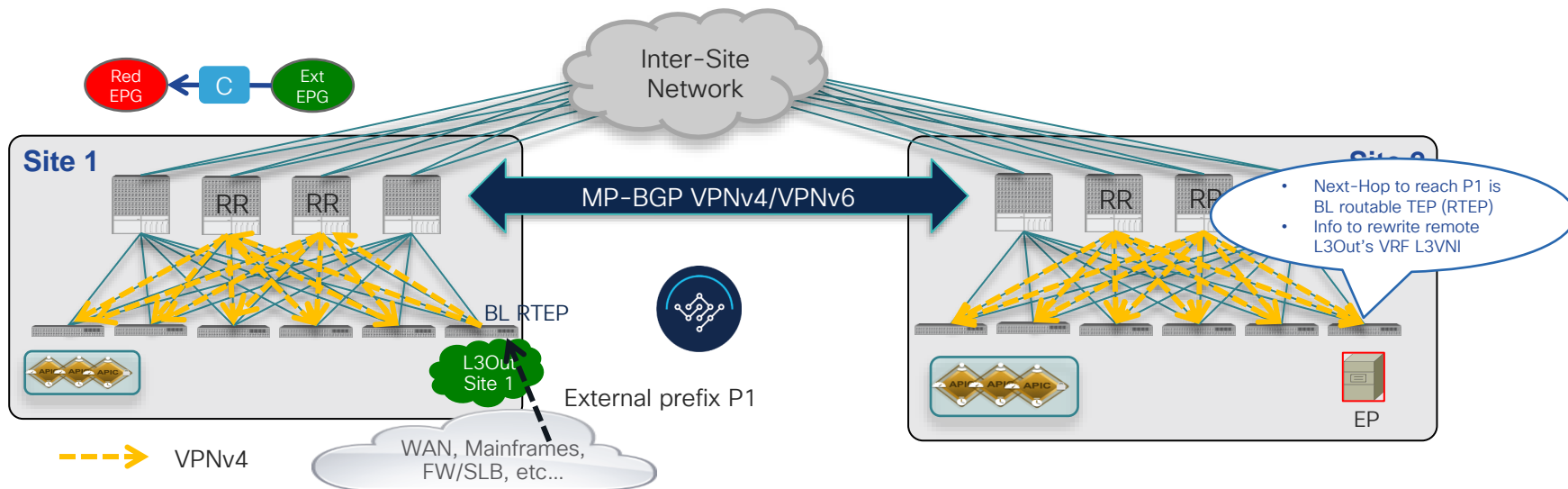
## Introduction of a Routable TEP Pool



- The BL TEP is normally taken from the original TEP pool assigned during the fabric bring-up procedure
- Since we don't want to assume that the original TEP pool can be reached across the ISN, a separate routable TEP pool is introduced to support intersite L3Out
  - The routable TEP pool can be directly configured on the Multi-Site Orchestrator
  - One or more routable TEP pools can be configured (pool size is /22 to /29)

# ACI Multi-Site and Intersite L3Out

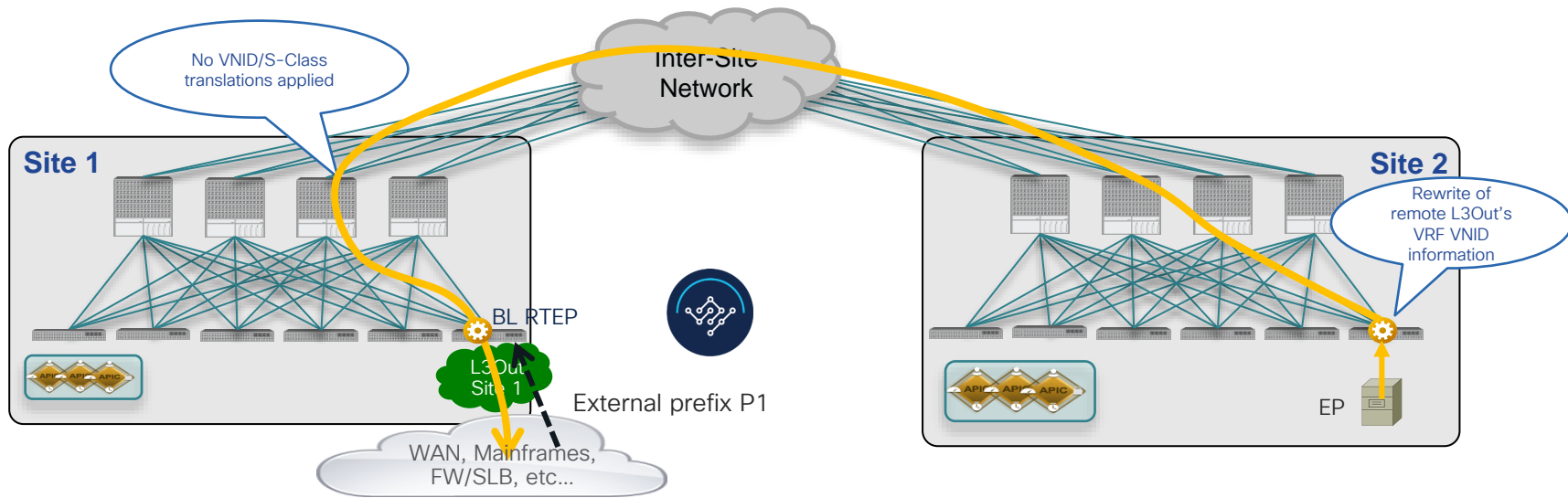
## Control Plane



- External prefix advertisements received via the L3Out are redistributed to the leaf nodes in the local site via MP-BGP VPNv4/VPNv6 through the RRs in the spines (normal ACI intra-fabric behavior)
- MP-BGP VPNv4 advertisements are also used to distribute this information to the remote sites
- The prefixes are then redistributed inside the remote sites via VPNv4/VPNv6 by the RR spines
  - The next-hop VTEP for the prefixes is the BL routable TEP (RTEP) that received the routes from the external network
  - Associated to the prefix information are the info to rewrite the VRF L3VNI value to match the one in the remote site

# ACI Multi-Site and Intersite L3Out

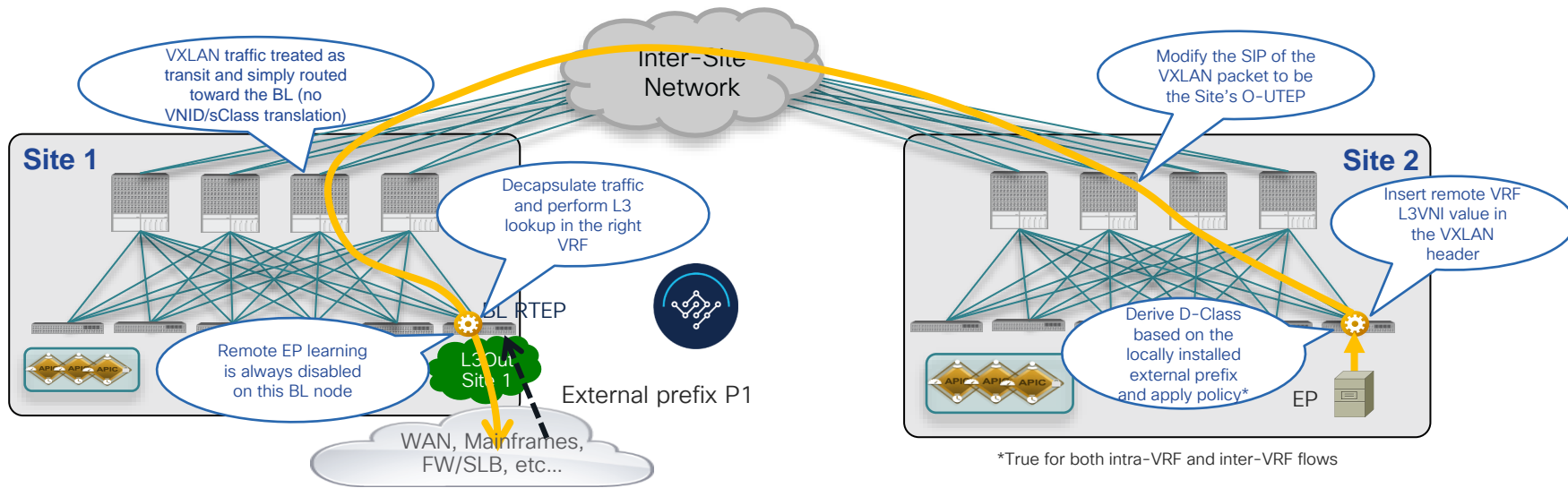
## No VNID/S-Class Translations on Receiving Spines



- Differently from regular inter-site (east-west) communication between endpoints, the spines on the receiving site are not involved in VNID/Class-ID translations for communication to external prefixes
- BGP program the rewrite of VNIDs of remote site routes directly on ToRs

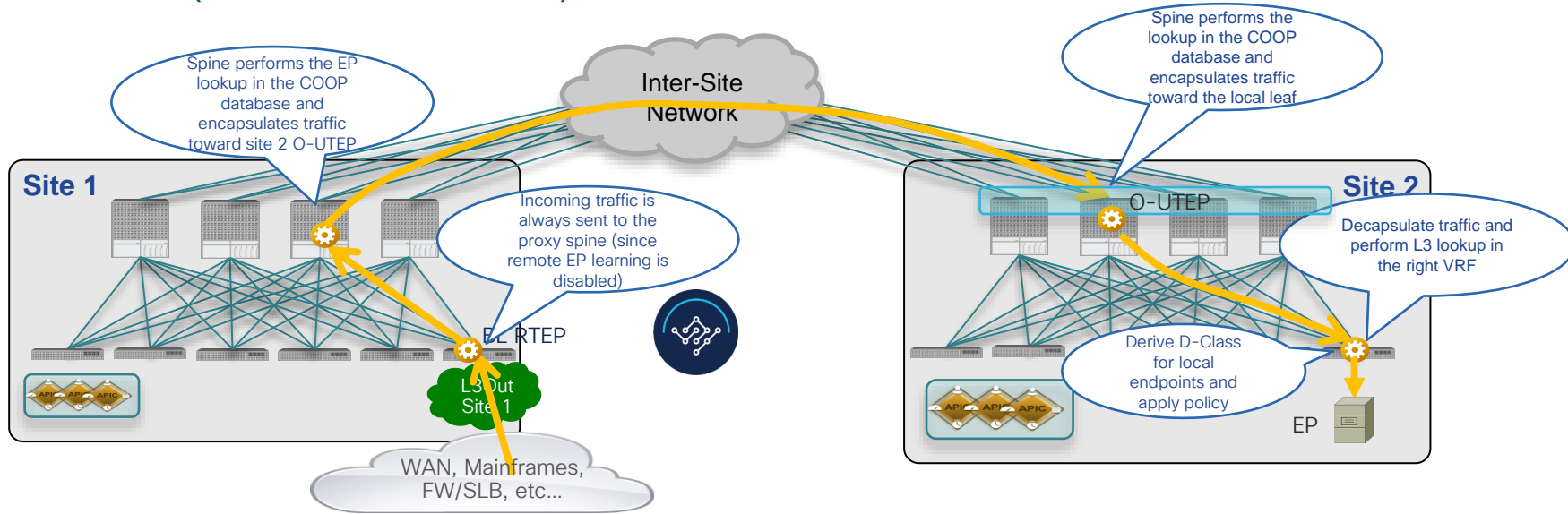
# ACI Multi-Site and Intersite L3Out

## Data Plane (EndPoint to L3Out)



- VXLAN tunnel is established directly between the leaf in site 2 and the BL in site 1
  - The spines in the source site still translate the SIP to be the site's O-UTEP
  - The spines in the destination site simply route the VXLAN packet toward the destination BL node
- The BL node uses the L3VNI info in the VXLAN header to perform the lookup in the right VRF
- Remote endpoint learning always disabled on the BL nodes to avoid learning the wrong sClass info (as there is no translation happening on the receiving spines)

# ACI Multi-Site and Intersite L3Out Data Plane (L3Out to EndPoint)



- Traffic received from the WAN hit the BL node and is always sent to the spine proxy
- The local spine performs the lookup for the destination endpoint and encapsulates to the O-UTEP of the destination site (after changing the SIP in the VXLAN header to match the local O-UTEP)
- The receiving spine performs the lookup in the COOP DB and S-Class/VNID translations (as in regular Multi-Site data-plane) → the traffic is encapsulated to the local leaf
- The local leaf decapsulates the packet, performs the L3 lookup, applies the policy and sends traffic to the endpoint (if allowed)

# Intersite L3Out Deployment Considerations

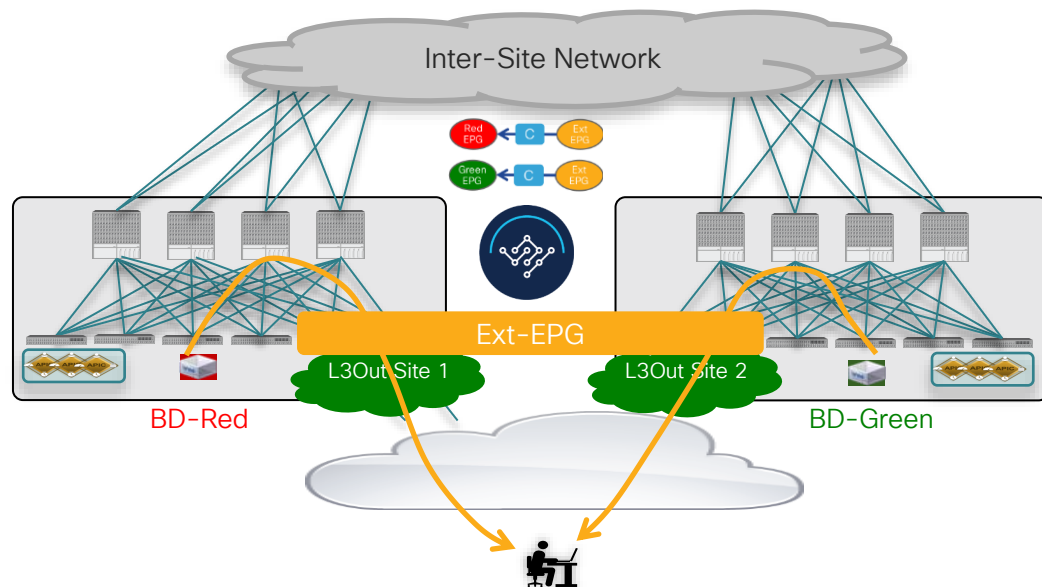
# ACI Multi-Site and Intersite L3Out

## Deployment Considerations

- Before ACI release 4.2(1), the outbound and inbound traffic flows take always a deterministic path
  - For BDs that are only locally defined in a site (i.e. not stretched), outbound communication is only possible via local L3Outs
  - Inbound communication is also only possible via the local L3Out, as it is not possible to advertise the BD subnet(s) out of a remote L3Out connection
  - For stretched BDs, the option of enabling host-based routing advertisement has been made available from ACI release 4.0(1) to ensure inbound flows take always an optimal path
- The enablement of the Inter-site L3Out functionality may change this behavior, so it is important to keep in mind some specific deployment considerations discussed in the following slide



# ACI Multi-Site and Intersite L3Out Stretched Ext-EPG - Outbound Flows



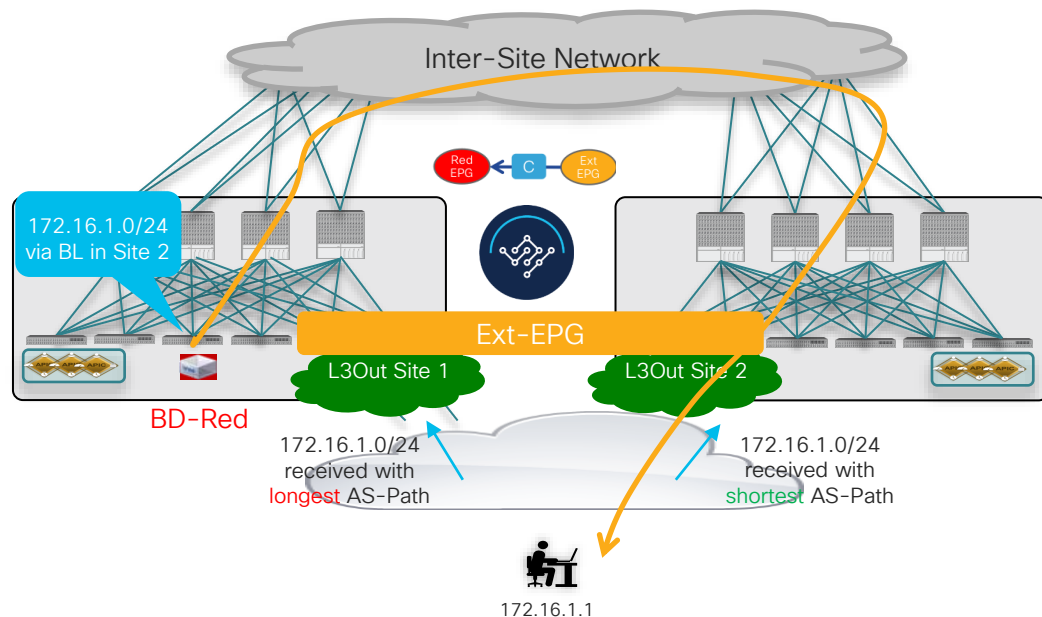
- When peering BGP with the external router, if all BGP parameters are the same, the local L3Out is preferred by default for outbound flows

The default behavior can be modified by tuning a BGP parameter (Local-Preference, MED, ...) for the received external prefixes

Notice that Local-Preference is only propagated inside a BGP AS, so can't be used if separate sites are peering EBGP

- When peering OSPF or EIGRP with the external router, local L3Out is preferred by default (best IS-IS metric to the local BL nodes)

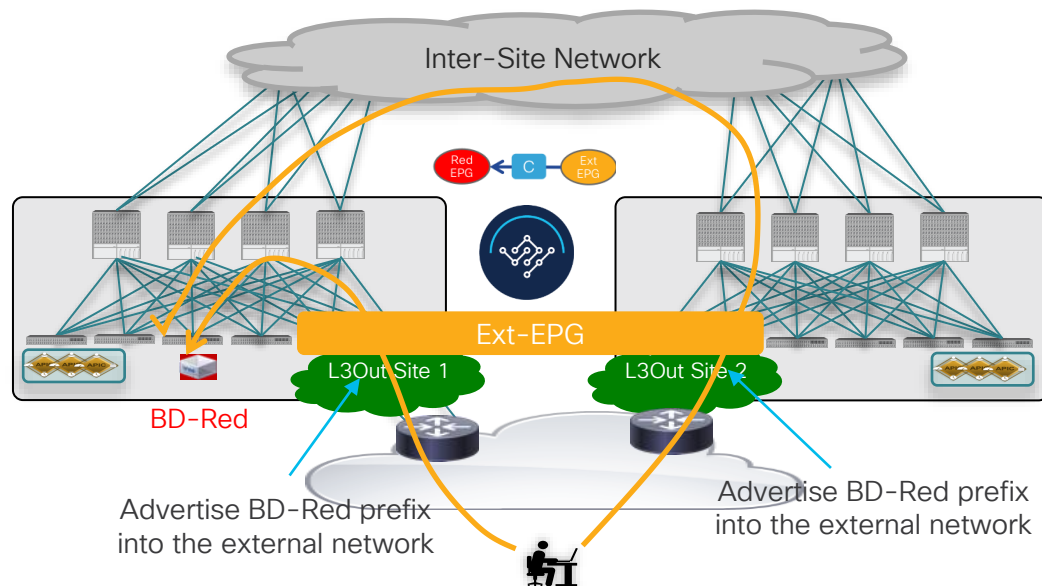
# ACI Multi-Site and Intersite L3Out Stretched Ext-EPG - Outbound Flows



- When peering BGP with the external router, if all BGP parameters are the same, local L3Out is preferred by default
- This is not the case if the BGP attributes (like AS-Path for example) are better for a prefix received in a given site
- A user may not be able to modify this behavior by applying inbound route-map on the BL nodes and would lose control on outbound traffic path

AS-Path is considered before Local-Preference and MED in the BGP route selection algorithm

# ACI Multi-Site and Intersite L3Out Stretched Ext-EPG - Inbound Flows



- If Intersite L3Out is enabled for the WAN isolation use case, it is required to advertise the BD Red subnet out of the local and remote L3Outs
- Without additional tuning it may happen that inbound traffic is steered toward the site where the BD is not deployed
  - This may cause asymmetric traffic path also for not stretched BD
- Enabling host-based advertisement is a possible solution, if there are not scalability concerns on the amount of host routes advertised externally
- Policies can also be applied on the L3Out of each site to make one path preferable compared to the other

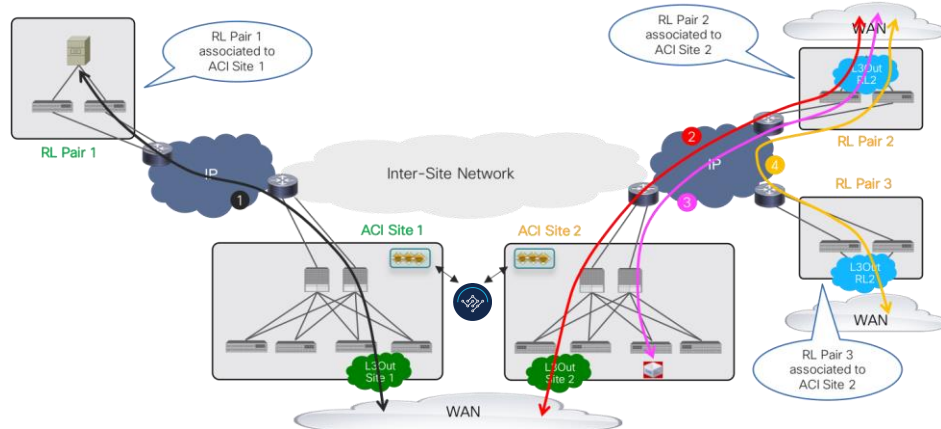
# ACI Multi-Site and Intersite L3Out

## Current Restrictions

- The following restrictions currently apply to the Inter-Site L3Out functionality:
  - Not supported when deploying GOLF L3Outs: this implies that a local GOLF L3Out connection is always required to enable outbound connectivity for endpoints deployed in a given site
  - Not supported between Remote Leaf pairs associated to separate sites
  - Not supported for some specific L3 multicast deployment scenarios (external source with PIM ASM/SSM or external receiver with PIM ASM)
  - Use of a remote L3Out connection after redirecting traffic through a local service node via PBR is not supported as of ACI 5.0(1) release
  - Currently CloudSec and the configuration of Routable TEP pools are mutually exclusive  
No support for Remote Leaf and/or Intersite L3Out deployments

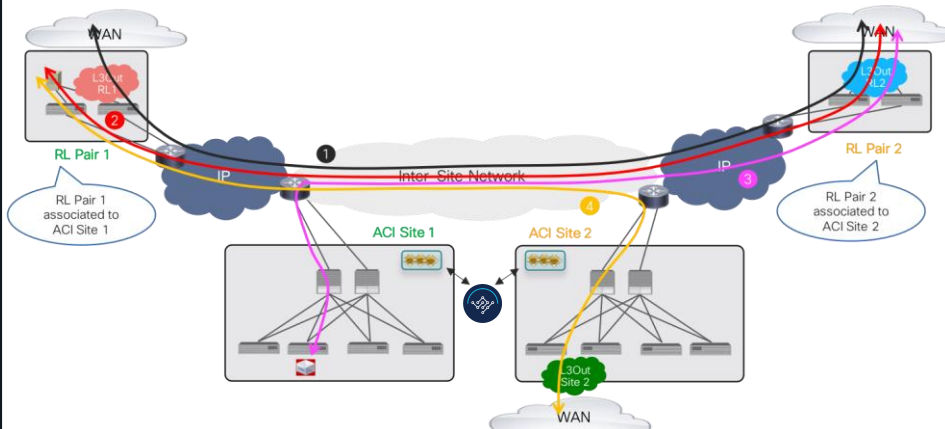
# ACI Multi-Site and Intersite L3Out Integration with Remote Leaf Nodes

## Supported Traffic Flows



- The traffic flows working with ACI release 4.1(2) will continue to be supported in 4.2(1)
  - Endpoint connected to the RL pair associated to a site communicating with L3Out(s) deployed in the same site
  - Transit routing between L3Outs deployed in the main site and the RL pair associated to the same site
  - Endpoint connected to the main site communicating with the L3Out deployed on the RL pair associated to the same site
  - Transit routing between L3Outs deployed on RL pairs associated to the same site

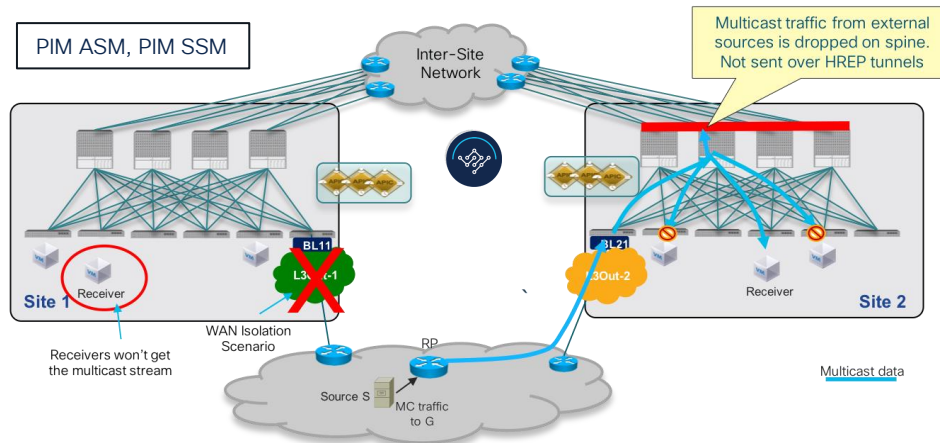
## Not Supported Traffic Flows



- The following traffic flows are not supported in ACI release 4.2(1)
  - Transit routing between L3Outs deployed on RL pairs associated to separate sites
  - Endpoint connected to a RL pair associated to a site communicating with the L3Out deployed on the RL pair associated to a remote site
  - Endpoint connected to the local site communicating with the L3Out deployed on the RL pair associated to a remote site
  - Endpoint connected to a RL pair associated to a site communicating with the L3Out deployed on a remote site

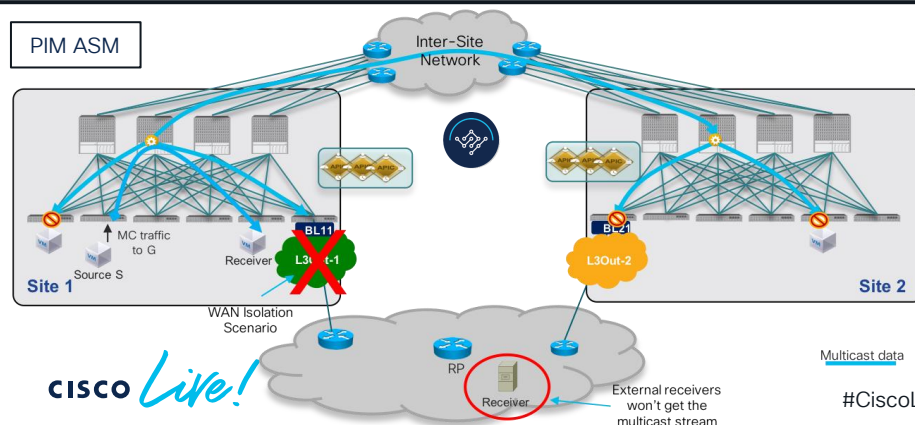
# ACI Multi-Site and Intersite L3Out

## Integration with L3 Multicast – Not Supported Scenarios



- Each site with local receivers must receive on a local L3Out connection the multicast stream originated by an external source

Multicast traffic received in a site from an external source will never be sent toward remote site via the Inter-Site network



- The site with the local source must have connectivity to the external RP via a local L3Out

In a WAN isolation scenario, external receivers won't be able to receive the multicast stream originated in a site that is isolated from the WAN until the local L3Out connection is recovered

# Network Services Integration

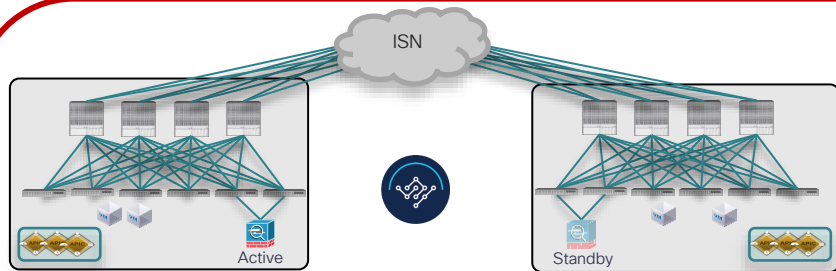


# Integration Models

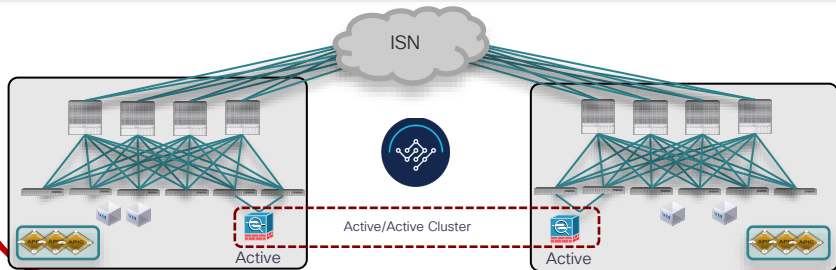


# ACI Multi-Site and Network Services Integration Models

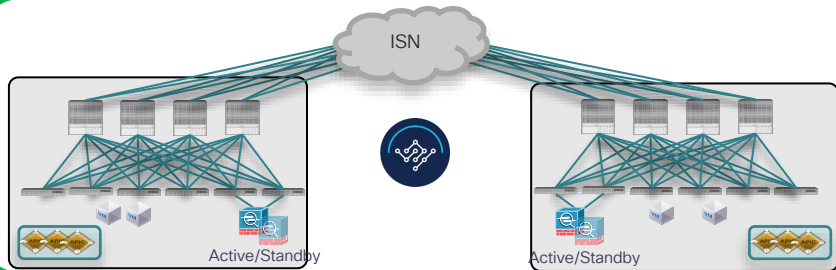
Deployment options fully supported with ACI Multi-Pod



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



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



- Recommended deployment model for ACI Multi-Site
- Option 1: supported for N-S traffic flows when the FW is connected in L3 mode to the fabric
- Option 2: supported for N-S and E-W traffic flows with the use of Service Graph with Policy Based Redirection (PBR)

# Independent Service Node Instances across Sites

## Use of Service Graph and Policy Based Redirection

- The PBR policy applied on a leaf switch can only redirect traffic to a service node deployed in the local site

Requires the deployment of independent service node function in each site

Various design options to increase resiliency for the service node function: per site Active/Standby pair, per site Active/Active cluster, per site multiple independent Active nodes

- HW dependencies:

Mandates the use of EX/FX or newer leaf nodes (both for compute and service leaf switches)

- SW dependencies:

ACI release 3.2(1)

- Only a single node PBR policy with FW supported
- Consumer side enforcement in 3.2 release (for E-W flows)

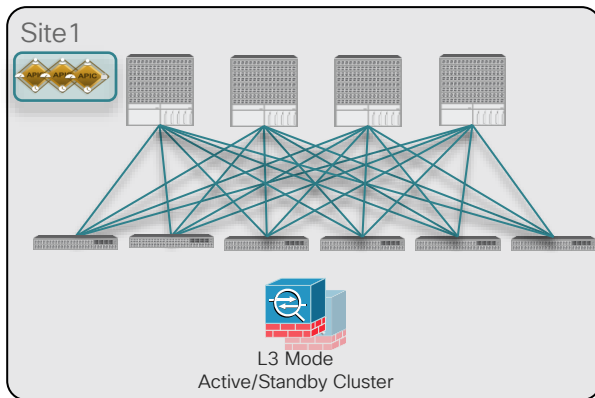
ACI release 4.0(1)

- Multiple node PBR policy with FW + LB supported in 4.0 release
- Provider side enforcement in 4.0 release

# Use of Service Graph and Policy Based Redirection

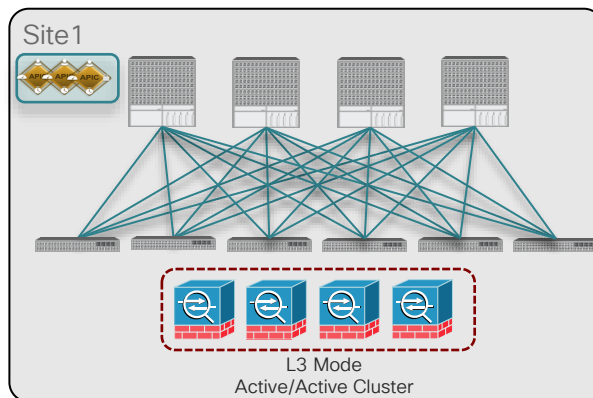
## Resilient Service Node Deployment in Each Site

### Active/Standby Cluster



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

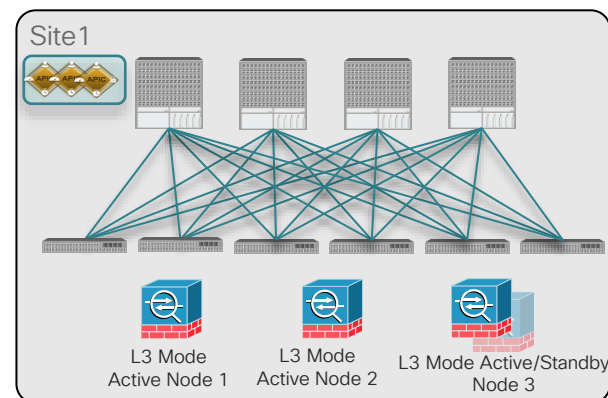
### Active/Active Cluster



- 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



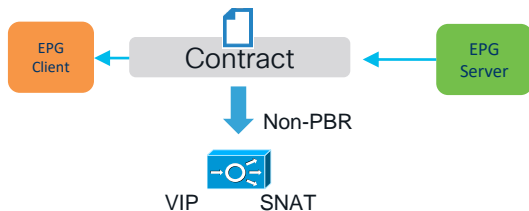
- Each Active node represent a unique MAC/IP entry in the PBR policy
- Use of Symmetric PBR to ensure each flow is handled by the same Active node in both directions

# Multi-Site and Network Services Integration

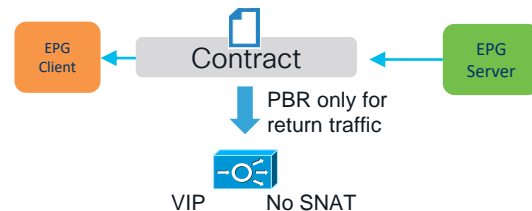
## Single-Node Service Graph with PBR



- Single-Node FW with PBR (used for both traffic directions)
- FW connected in one-arm or two-arms mode
- ACI 3.2(1): PBR policy applied on the compute leaf (N-S flows) or on the consumer leaf (E-W flows)
- ACI 4.0(1): PBR policy applied on the compute leaf (N-S flows) or on the provider leaf (E-W flows)



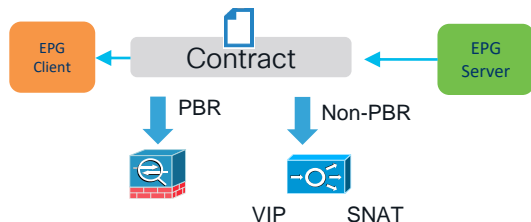
- LB connected in one-arm or two-arms mode
- No need for PBR at all, LB connected to BDs/EPGs as a regular endpoint
- LB could also be connected to an L3Out (may require intersite L3Out support)



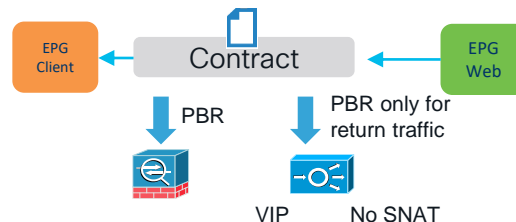
- LB with PBR for return traffic supported from ACI 4.0(1) release
- LB connected in one-arm or two-arms mode
- PBR policy applied on the compute leaf (N-S flows) or on the provider leaf (E-W flows)

# Multi-Site and Network Services Integration

## Dual-Node Service-Graph with PBR (Main Use Cases)



- PBR policy only to redirect traffic to the FW, no need for PBR for the LB (LB connected to BDs/EPGs as a regular endpoint)
- FW and LB in single-arm or dual-arms mode
- Supported from ACI release 4.0(1)



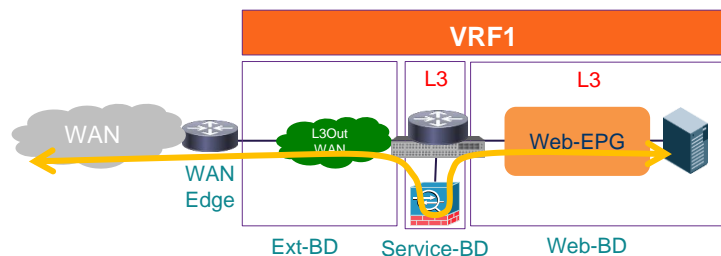
- PBR policy to redirect traffic to the FW, second PBR policies for the LB (for return traffic)
- FW and LB in single-arm or dual-arms mode
- Supported from ACI release 4.0(1)

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

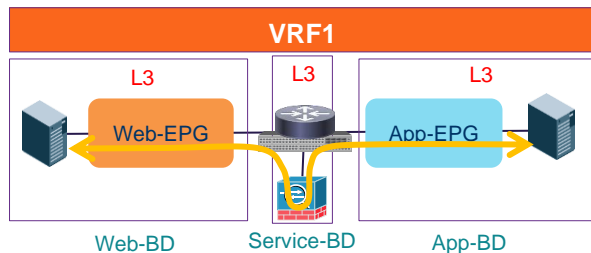
# Use of Service Graph and Policy Based Redirection

## North-South and East-West Use Cases

### North-South



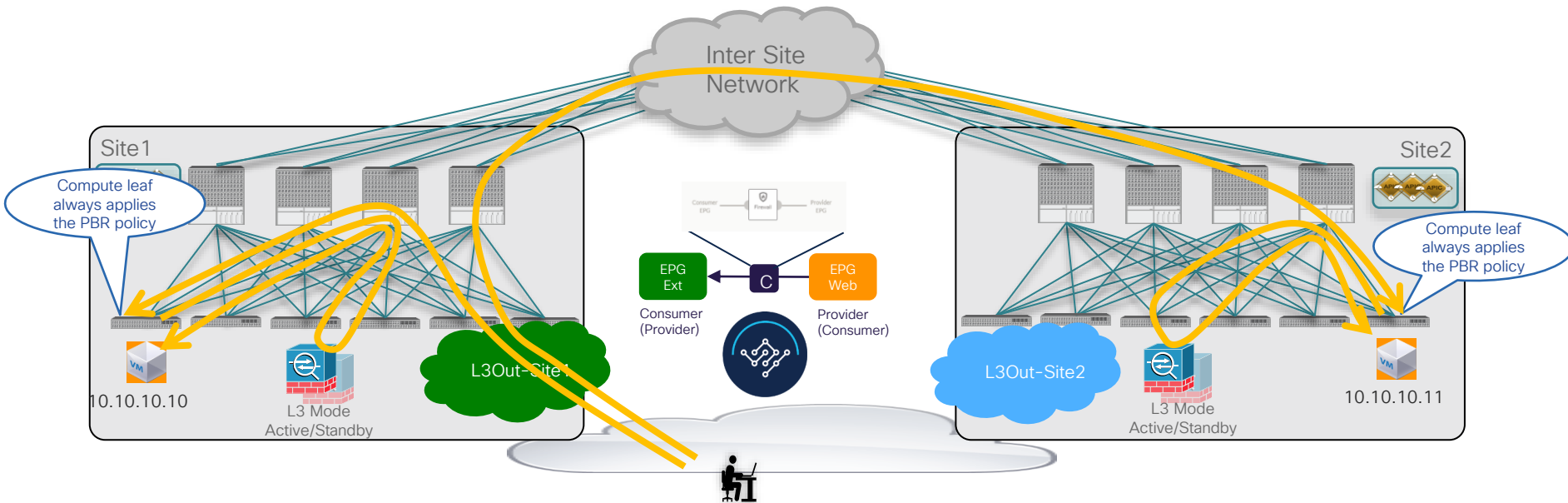
### East-West



- Best practice recommendations for both North-South and East-West use cases:
  - Service Node deployed in 'one arm' mode ('two-arms' mode also supported but not preferred)
  - Service-BD must be stretched across sites (BUM flooding can/should be disabled)
  - Ext-EPG must also be a stretched object, mapped to the individual L3Outs defined in each site
  - Web-BD and App-BD can be stretched across sites or locally defined in each site
- North-South use case
  - Intra-VRF only support as of ACI 5.0(1) release
- East-West use case
  - Supported intra-VRF or inter-VRFs/Tenants
  - Requires to configure the subnet under the Consumer EPG from ACI release 4.0(1) (under the Provider EPG in 3.2(x) releases)

# 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

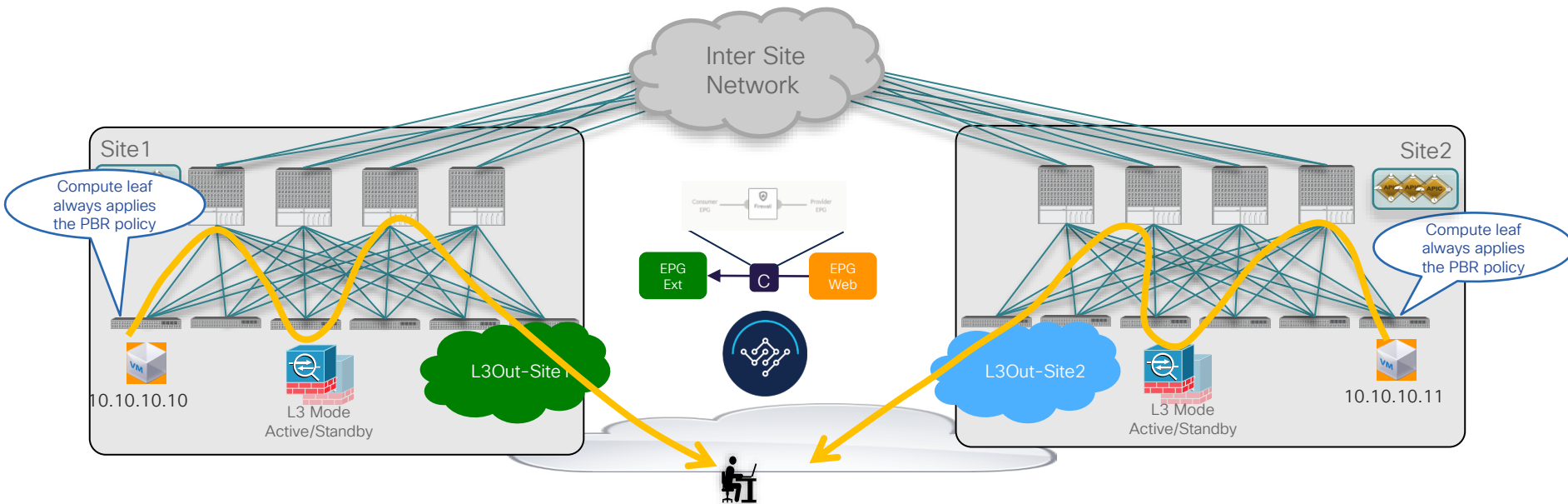
Requires the VRF to have the default policies for enforcement preference and direction  
Ext-EPG and Web EPG can indifferently be provider or consumer of the contract

Policy Control Enforcement Preference:	<input checked="" type="checkbox"/> Enforced	<input type="checkbox"/> Unenforced
Policy Control Enforcement Direction:	<input type="checkbox"/> Egress	<input checked="" type="checkbox"/> Ingress



# Use of Service Graph and Policy Based Redirection

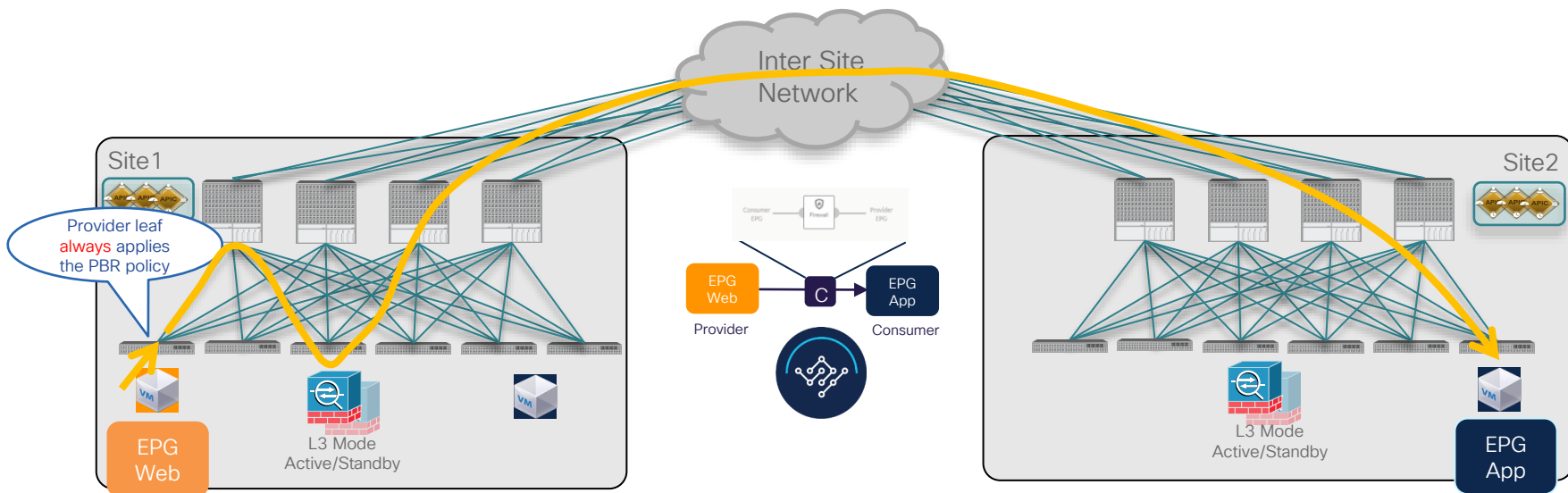
## North-South Communication – Outbound Traffic



- 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)



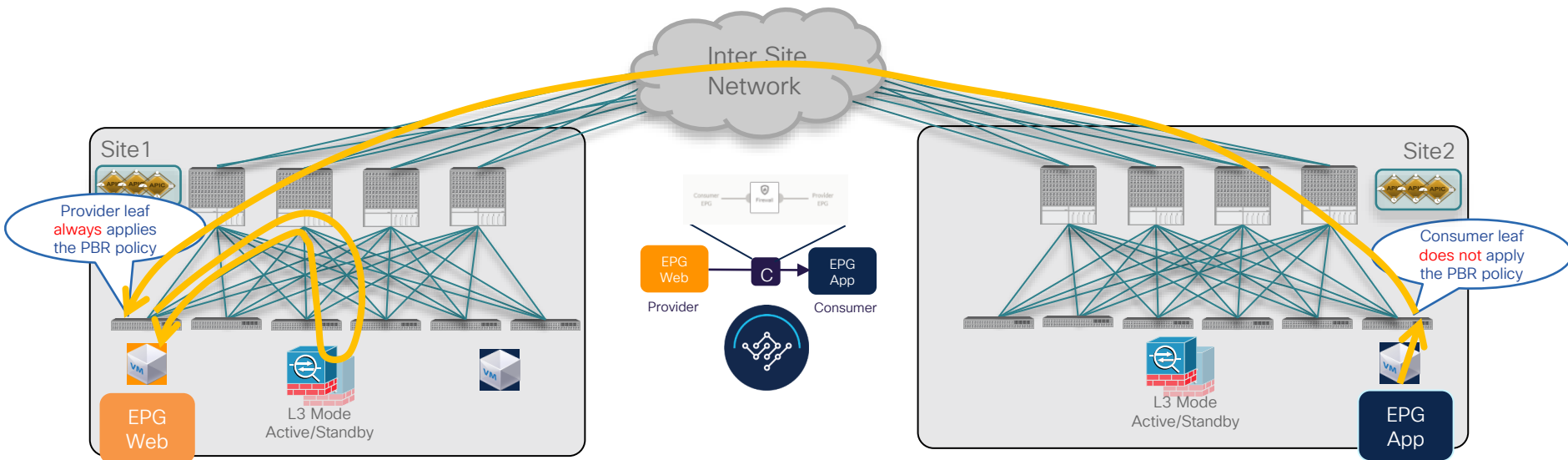
- EPGs can be locally defined or stretched across sites and can be part of the same VRF or in different VRFs (and/or Tenants)
- PBR policy is always applied on the leaf switch where the Provider endpoint is connected

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

## 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>

# Technical Session Surveys

- Attendees who fill out a minimum of four session surveys and the overall event survey will get Cisco Live branded socks!
- Attendees will also earn 100 points in the Cisco Live Game for every survey completed.
- These points help you get on the leaderboard and increase your chances of winning daily and grand prizes.



# Cisco Learning and Certifications

From technology training and team development to Cisco certifications and learning plans, let us help you empower your business and career. [www.cisco.com/go/certs](https://www.cisco.com/go/certs)

## Pay for Learning with Cisco Learning Credits

(CLCs) are prepaid training vouchers redeemed directly with Cisco.



## Learn

### Cisco U.

IT learning hub that guides teams and learners toward their goals

### Cisco Digital Learning

Subscription-based product, technology, and certification training

### Cisco Modeling Labs

Network simulation platform for design, testing, and troubleshooting

### Cisco Learning Network

Resource community portal for certifications and learning



## Train

### Cisco Training Bootcamps

Intensive team & individual automation and technology training programs

### Cisco Learning Partner Program

Authorized training partners supporting Cisco technology and career certifications

### Cisco Instructor-led and Virtual Instructor-led training

Accelerated curriculum of product, technology, and certification courses



## Certify

### Cisco Certifications and Specialist Certifications

Award-winning certification program empowers students and IT Professionals to advance their technical careers

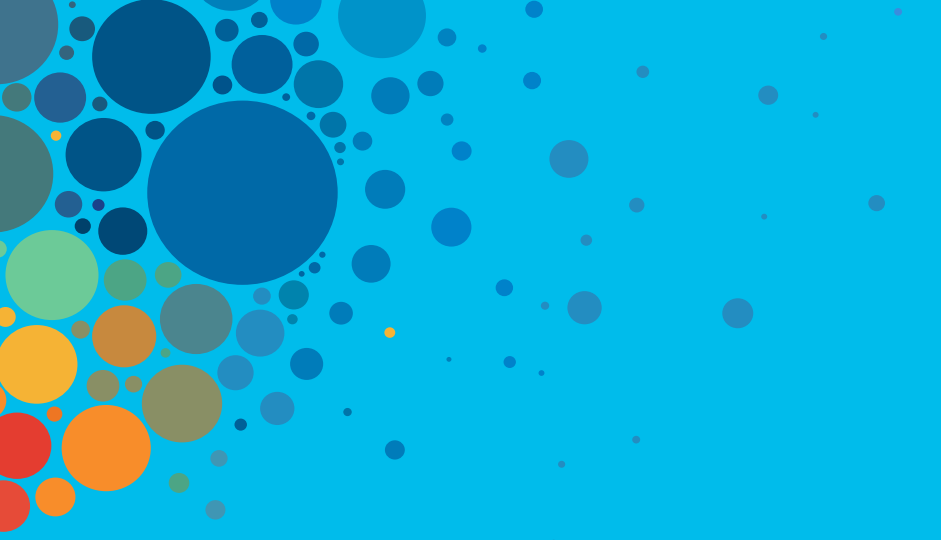
### Cisco Guided Study Groups

180-day certification prep program with learning and support

### Cisco Continuing Education Program

Recertification training options for Cisco certified individuals

Here at the event? Visit us at **The Learning and Certifications lounge at the World of Solutions**



# 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 [www.CiscoLive.com/on-demand](https://www.CiscoLive.com/on-demand)



The bridge to possible

# Thank you



CISCO *Live!*



#CiscoLive