cisco Live!

GO BEYOND

#CiscoLiveAPJC

Deployment of VXLAN EVPN Gateways with Cisco ACI for the Interconnection of Heterogeneous Data Center Fabrics

Max Ardica, Distinguished TME
@maxardica

Hall of Fame
Elite Speaker
cisco Live!

BRKDCN-2634

Cisco Webex App

Questions?

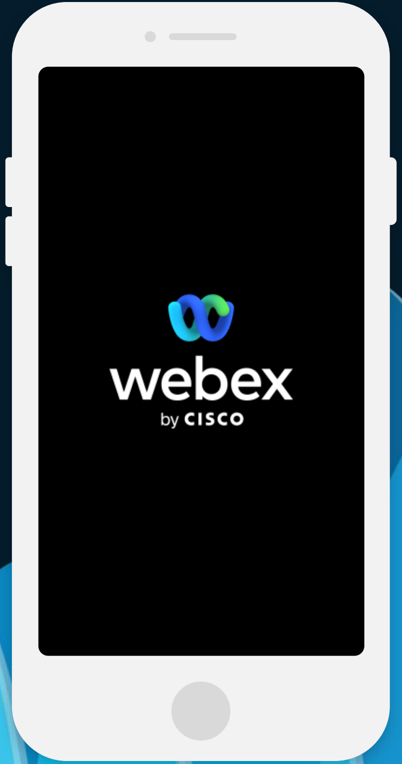
Use Cisco Webex App to chat
with the speaker after the session

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 November 15, 2024.

https://ciscolive.ciscoevents.com/
ciscolivebot/#BRKDCN-2634



Agenda

- Introducing Cisco Nexus ONE
- ACI Border Gateways (BGWs)
  - Introduction
  - Overview of Control-Plane and Data-Plane
  - Namespace Normalization
  - Workload Mobility across Domains
  - Policy Enforcement on ACI BGWs
- Secure Interconnection of Heterogeneous Fabrics

Introducing Cisco Nexus ONE

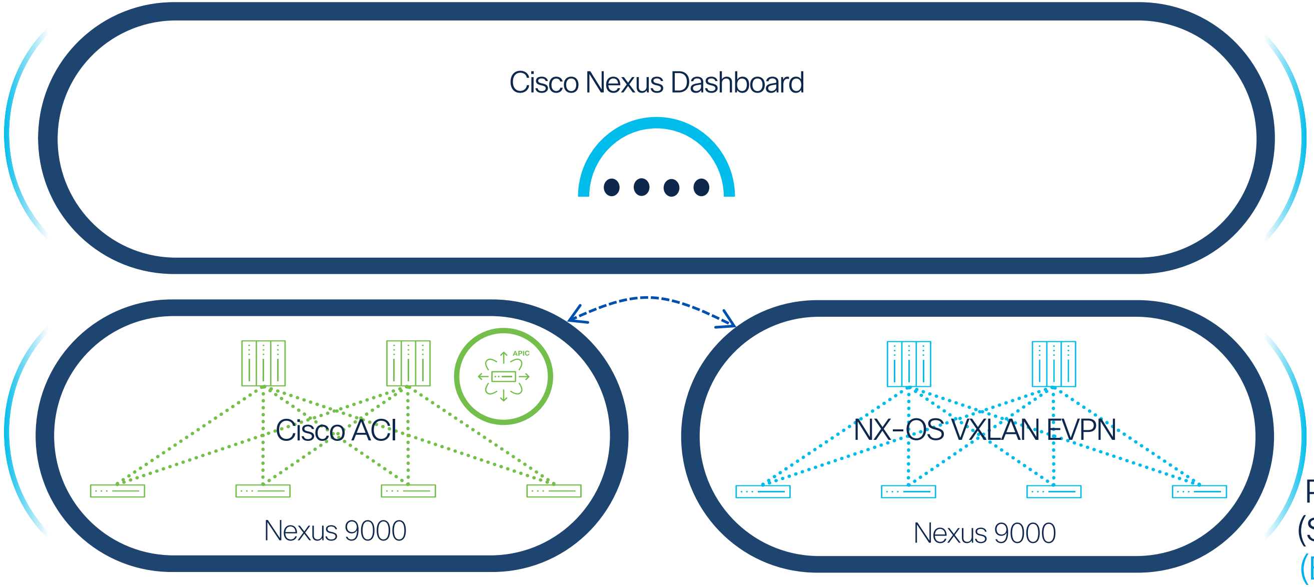
What is Cisco Nexus ONE?

Open NEtworking Fabric Experience

Evolve multiple DCN fabrics into a single user experience to deliver consistent use cases

Cisco Nexus ONE - Overview

3 Cisco Nexus Dashboard as single point of control and operations



1 ACI VXLAN EVPN Border Gateways

2 Policy in NX-OS (Security Groups) (BRKDCN-2633)

Different fabric architectures

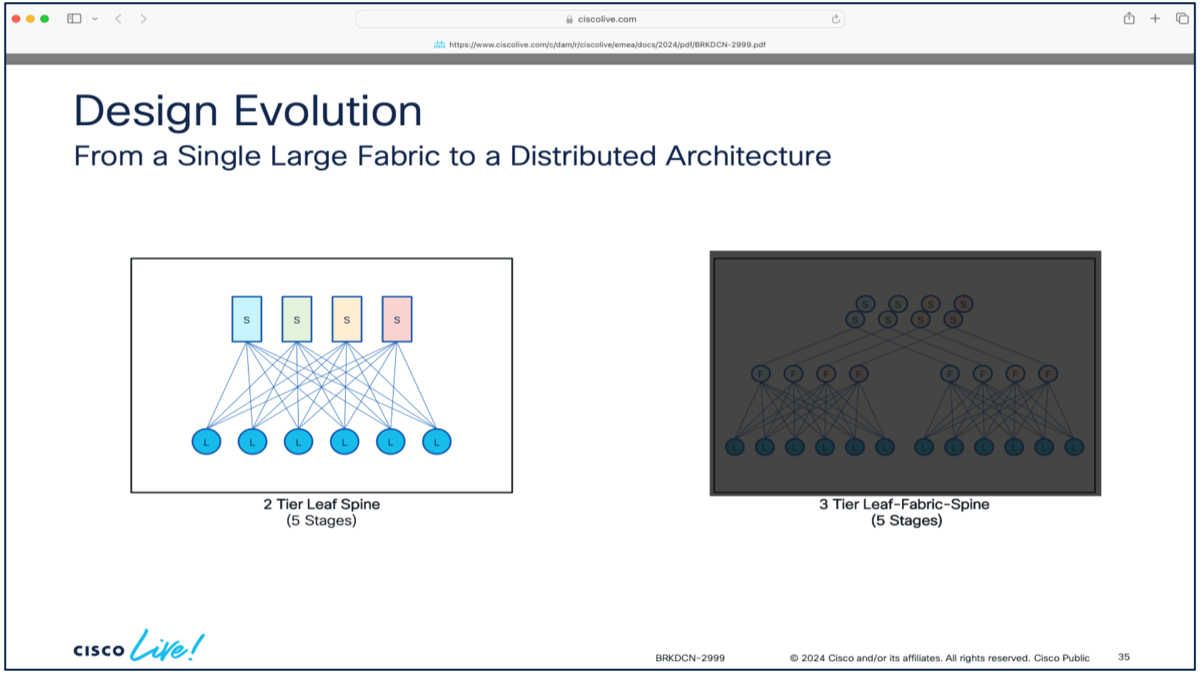
Same outcome with common experience

ACI Border Gateways

Introduction

DC Design Evolution

From a Single Large Fabric to a Distributed Architecture



ciscolive.com
https://www.ciscolive.com/c/dam/r/ciscolive/emea/docs/2024/pdf/BRKDCN-2999.pdf
S
L
F

Design Evolution

From a Single Large Fabric to a Distributed Architecture

2 Tier Leaf Spine
(5 Stages)

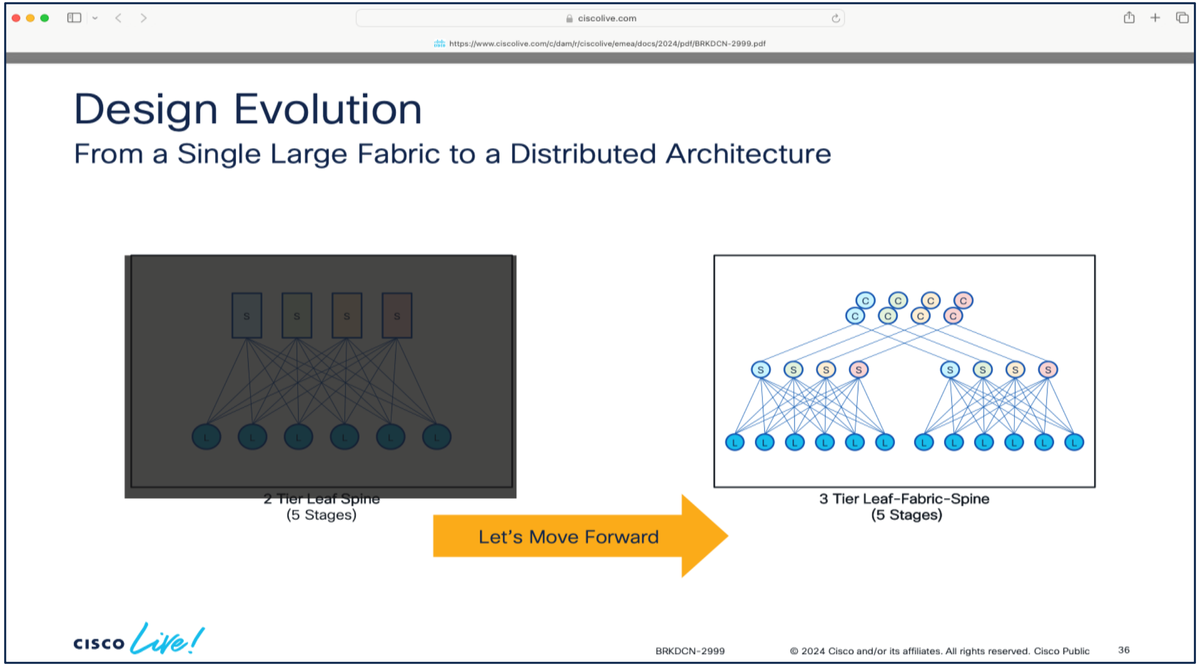
3 Tier Leaf-Fabric-Spine
(5 Stages)

BRKDCN-2999 © 2024 Cisco and/or its affiliates. All rights reserved. Cisco Public 35

For more information on DC Multi-Tier Design Evolution please refer to BRKDCN-2099

DC Design Evolution

From a Single Large Fabric to a Distributed Architecture



ciscolive.com

https://www.ciscolive.com/c/dam/r/ciscolive/emea/docs/2024/pdf/BRKDCN-2999.pdf

Design Evolution

From a Single Large Fabric to a Distributed Architecture

2 Tier Leaf Spine
(5 Stages)

Let's Move Forward

3 Tier Leaf-Fabric-Spine
(5 Stages)

cisco Live!

BRKDCN-2999

© 2024 Cisco and/or its affiliates. All rights reserved. Cisco Public

36

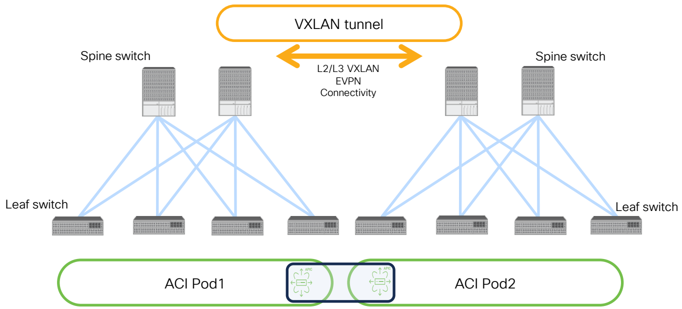
For more information on DC Multi-Tier Design Evolution please refer to BRKDCN-2099

Building Distributed DC Architectures

Homogeneous Options

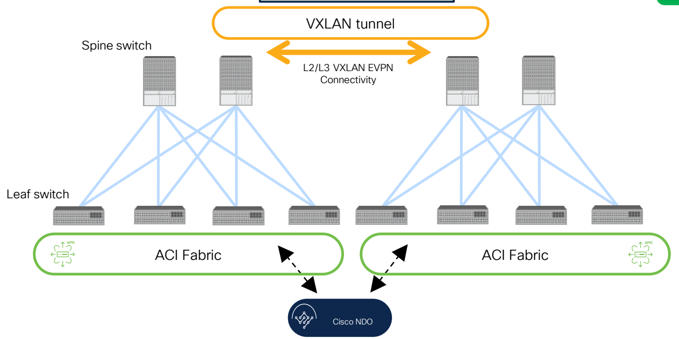
ACI Multi-Pod

Since 2017



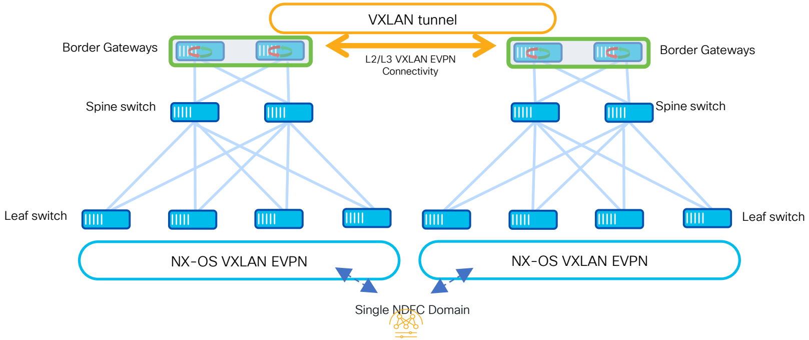
ACI Multi-Site

Since 2018



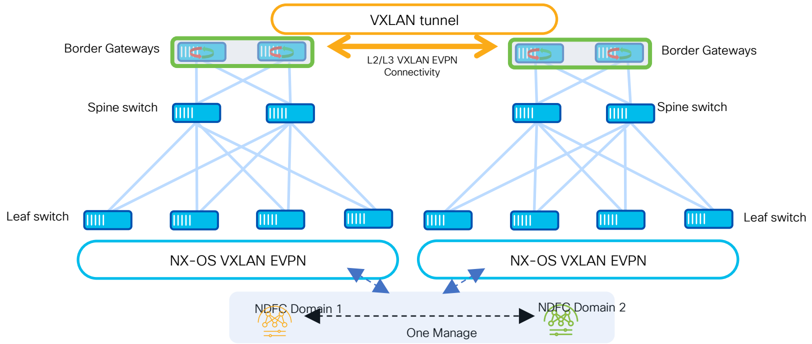
VXLAN EVPN Multi-Site

Since 2017



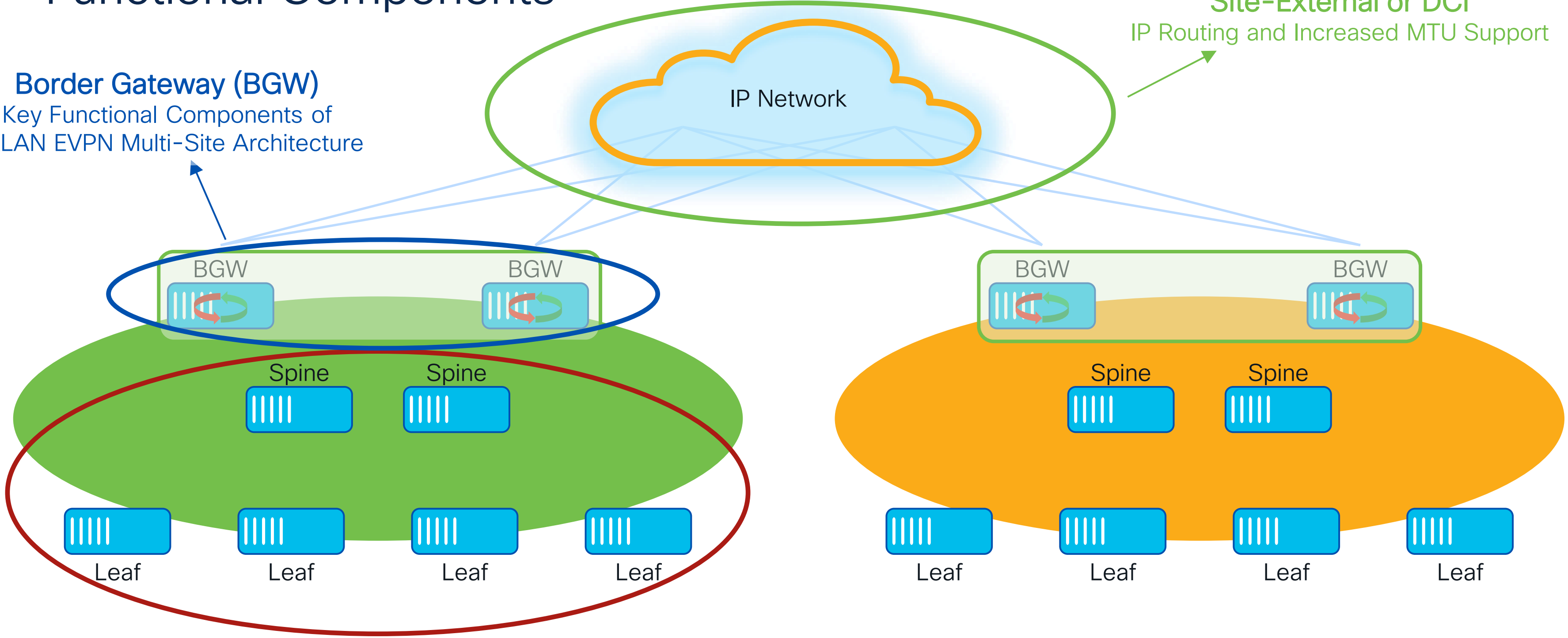
VXLAN EVPN Multi-Site

2024



VXLAN EVPN Multi-Site

Functional Components



Site-External or DCI
IP Routing and Increased MTU Support

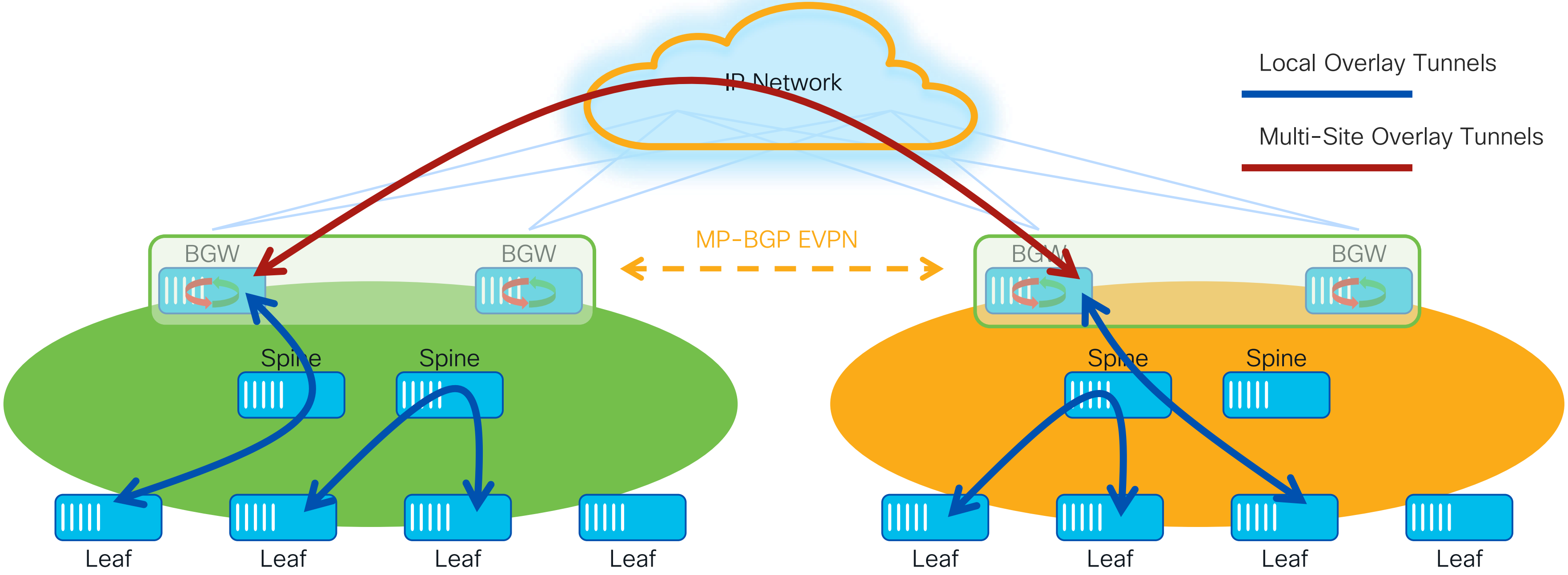
Border Gateway (BGW)
Key Functional Components of
VXLAN EVPN Multi-Site Architecture

Site-Internal or Fabric
A Simple VXLAN EVPN Fabric

draft-sharma-bess-multi-site-evpn

VXLAN EVPN Multi-Site

Hierarchical Encapsulation



Local Overlay Tunnels

Multi-Site Overlay Tunnels

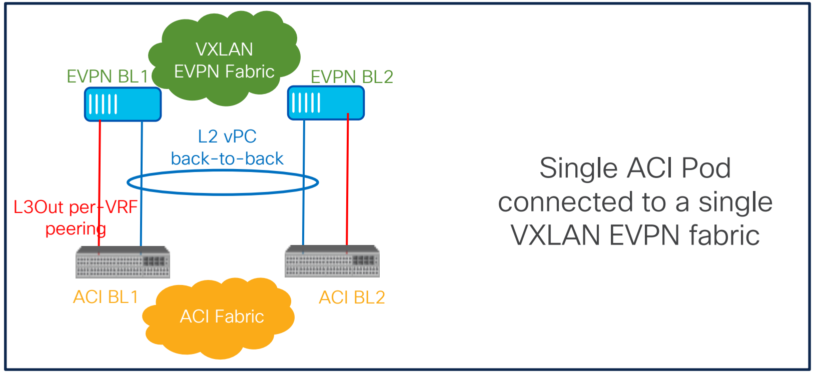
VXLAN EVPN Multi-Site Design and Deployment White Paper

https://www.cisco.com/c/en/us/products/collateral/switches/nexus-9000-series-switches/white-paper-c11-739942.html

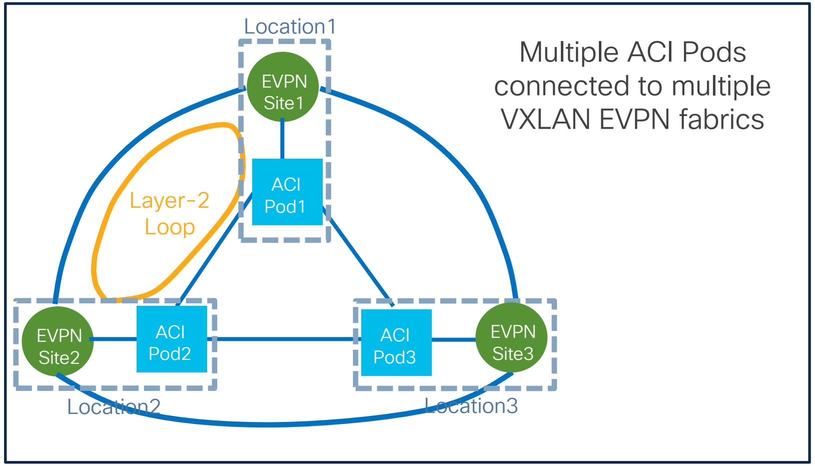
ACI Border Gateway

Problem Statement

- The traditional approach of interconnecting different types of DC fabrics (ACI and VXLAN EVPN, for example) calls for the creation of dedicated L2 and L3 connections between the domains to facilitate the extension of L2/L3 connectivity
- This is still manageable when a single ACI Pod is connected to a single VXLAN EVPN fabric
- When multiple ACI Pods (part of the same Multi-Pod fabric) needs to be interconnected to multiple VXLAN EVPN fabric (part of the same VXLAN Multi-Site domain), the establishment of those connections may lead to the creation of Layer 2 loops



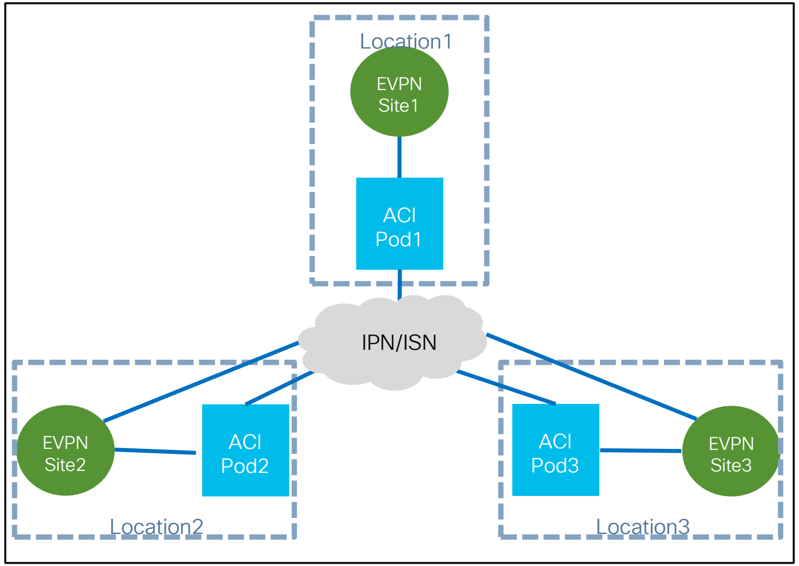
Single ACI Pod connected to a single VXLAN EVPN fabric



Multiple ACI Pods connected to multiple VXLAN EVPN fabrics

ACI Border Gateway
Solution and Use Cases

- The introduction of ACI Border Gateways allows to interconnect different ACI Pods and VXLAN EVPN fabrics through a generic Layer 3 infrastructure (sometimes referred to as IPN or ISN)
  - Standard VXLAN EVPN technology is used to extend Layer 2 ad Layer 3 connectivity between the ACI and the VXLAN EVPN domains
- Main use cases:
  - Migration/Coexistence of heterogeneous DC fabric types
  - Multi-Domain integration between DC and Campus domains
  - Autonomous Remote Leaf deployments

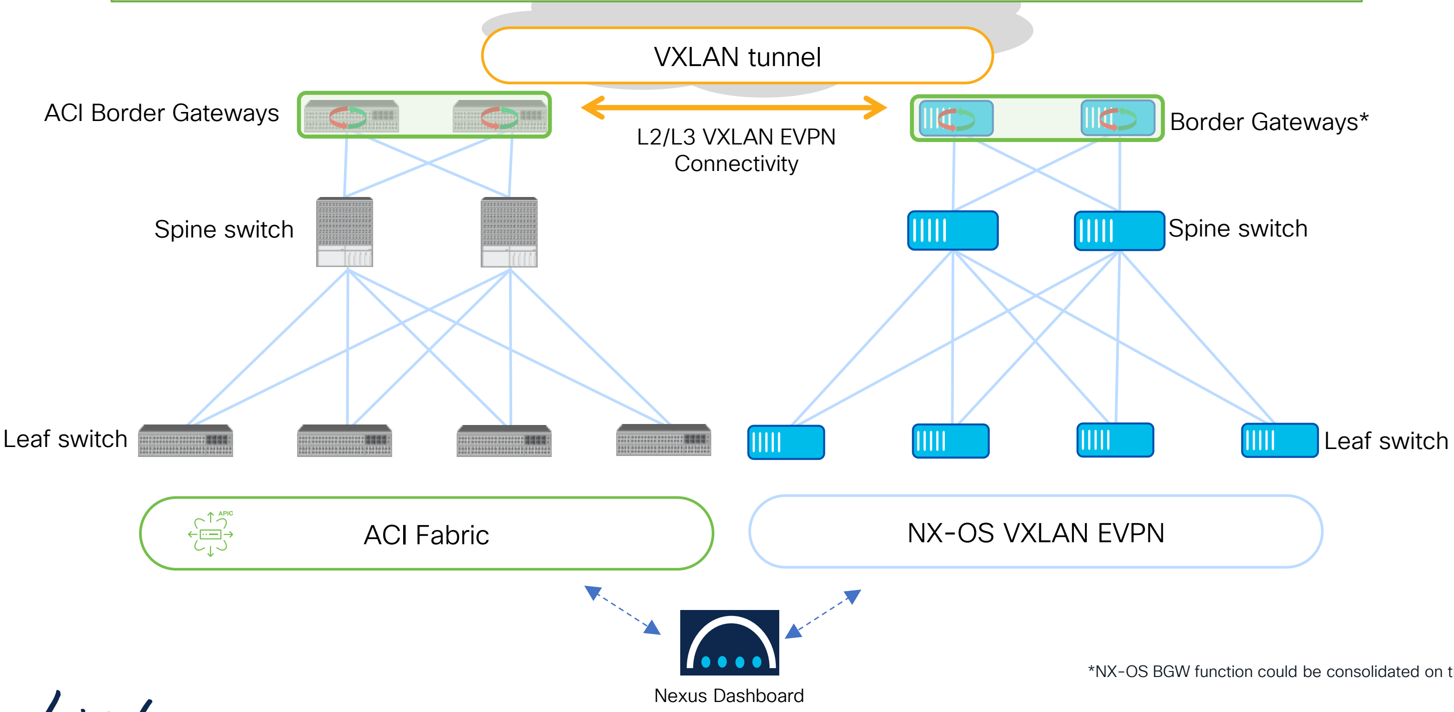


Heterogeneous Fabrics

Introducing ACI Border Gateways

ACI 6.1(x)

“Opening Up” L2/L3 Connectivity between ACI and VXLAN EVPN Fabrics

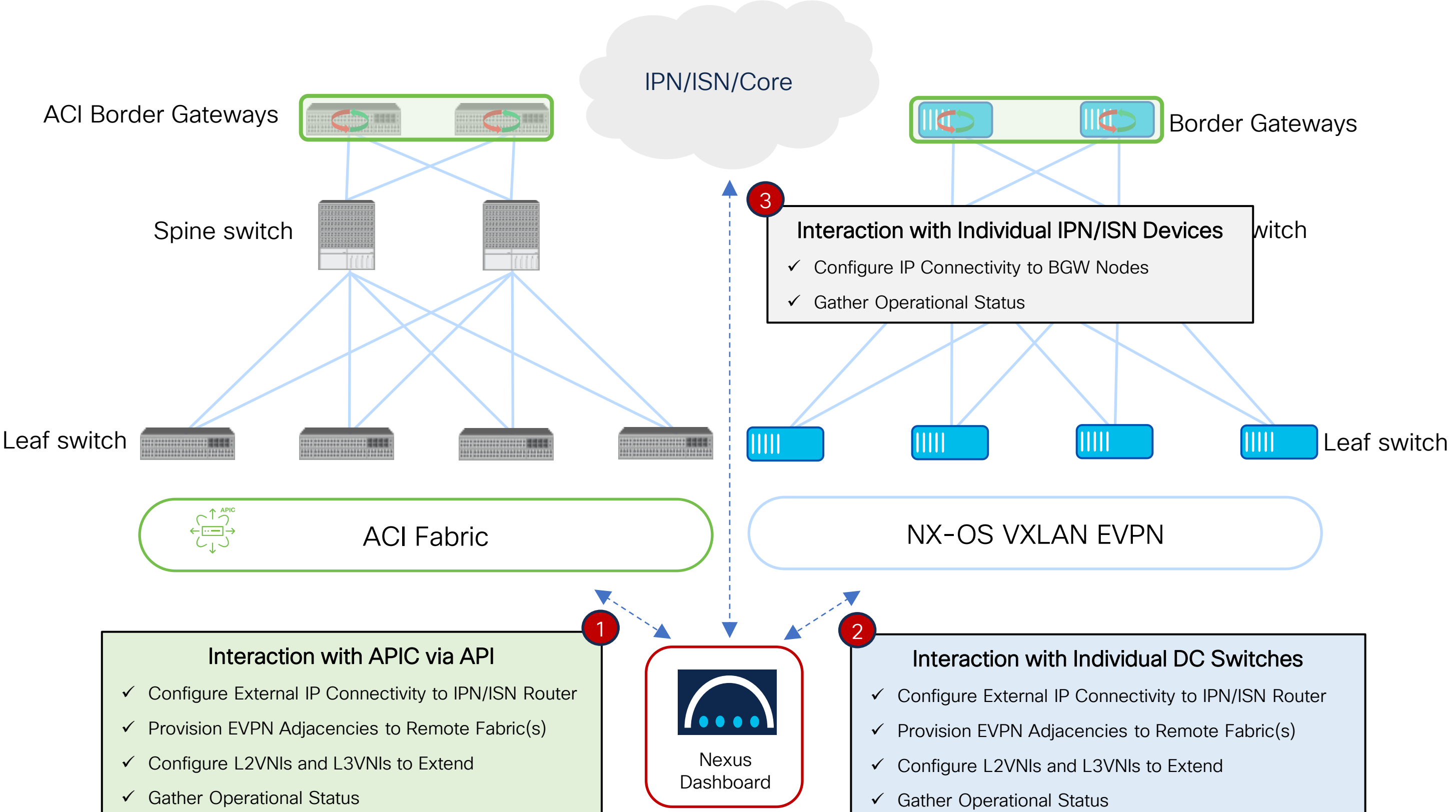


*NX-OS BGW function could be consolidated on the spines if desired

CY2025

ACI Border Gateway

Nexus Dashboard as Single Point of Management and Operation



Interaction with Individual IPN/ISN Devices

- ✓ Configure IP Connectivity to BGW Nodes
- ✓ Gather Operational Status

Interaction with APIC via API

- ✓ Configure External IP Connectivity to IPN/ISN Router
- ✓ Provision EVPN Adjacencies to Remote Fabric(s)
- ✓ Configure L2VNIs and L3VNIs to Extend
- ✓ Gather Operational Status

Interaction with Individual DC Switches

- ✓ Configure External IP Connectivity to IPN/ISN Router
- ✓ Provision EVPN Adjacencies to Remote Fabric(s)
- ✓ Configure L2VNIs and L3VNIs to Extend
- ✓ Gather Operational Status

ACI Border Gateway

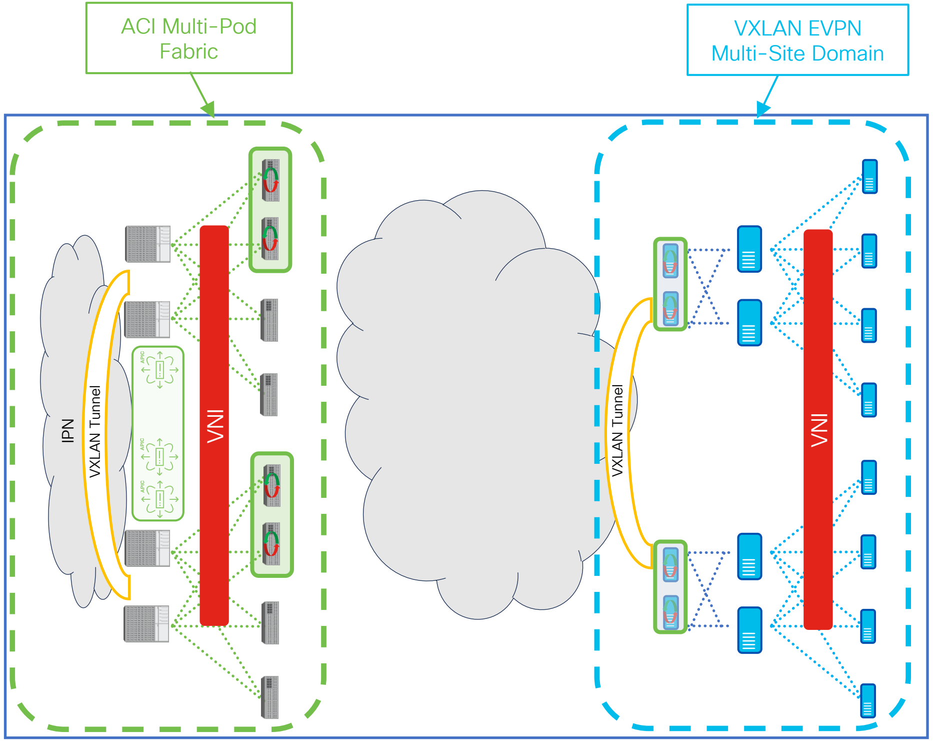
Deployment Considerations

- Hardware support for ACI BGWs: Nexus 9000 FX2 and above
- Dedicated leaf nodes for Border Gateway functionality
  - Coexistence with Border Leaf functions (L3Outs) planned for a future release
- IGMP snooping and L3 Multicast traffic not supported across domains
  - L2 Multicast traffic forwarded as BUM
- Symmetric namespace between ACI and VXLAN EVPN domains
  - VNIs must be defined in the VXLAN EVPN domain to match the APIC assigned VNIDs
- Support for a single ACI fabric (can be Multi-Pod)

Heterogeneous Fabrics

ACI Multi-Pod Fabric Support

ACI 6.1(1)

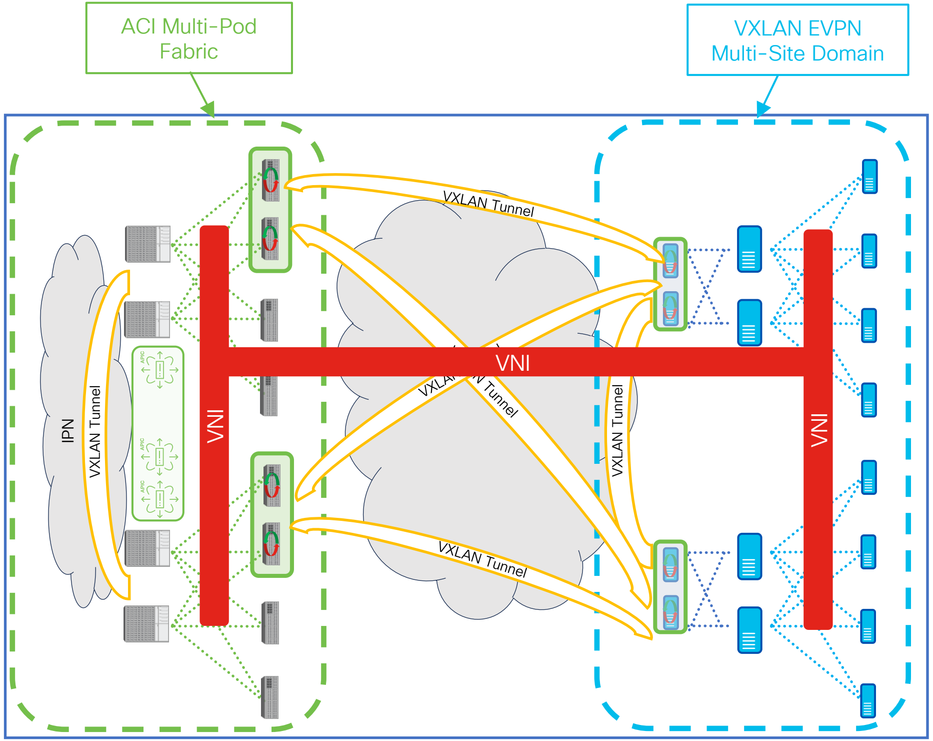


- L2/L3 VXLAN connectivity between ACI Pods part of the same fabric achieved via the spine-to-spine data path (through the IPN)
  - No VXLAN EVPN connectivity between ACI BGWs of different ACI Pods
- Local instance of ACI BGWs mandatory in each Pod
- For each BD extended across domains, a specific ACI BGW is elected as DF (across all the BGWs in all the Pods)

Heterogeneous Fabrics

ACI Multi-Pod Fabric Support

ACI 6.1(1)

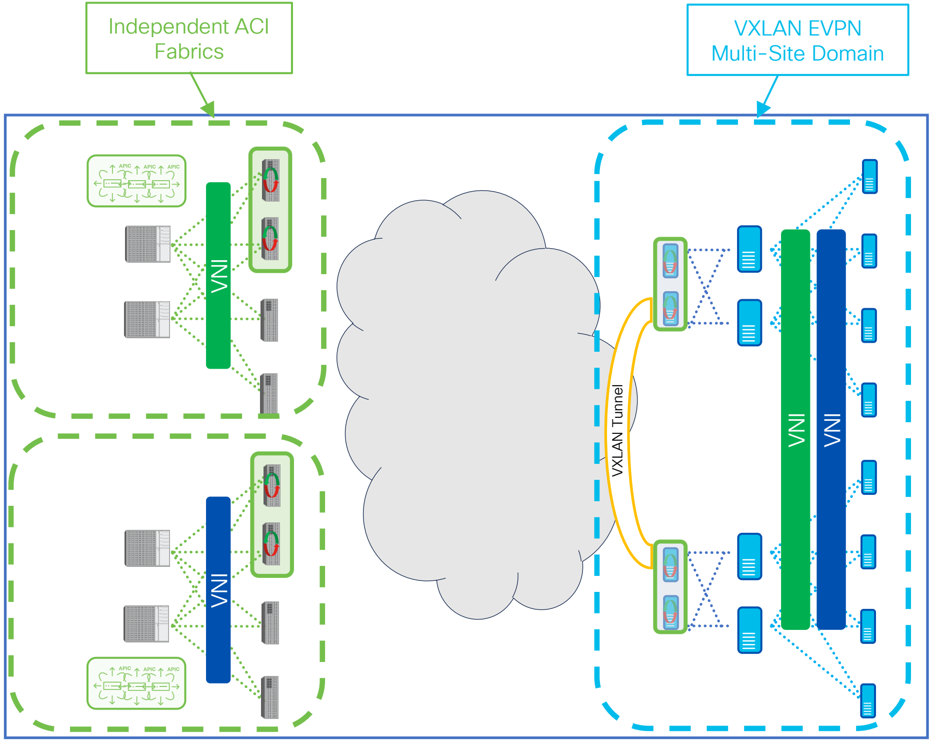


- L2/L3 VXLAN connectivity between ACI Pods part of the same fabric achieved via the spine-to-spine data path (through the IPN)
  - No VXLAN EVPN connectivity between ACI BGWs of different ACI Pods
- Local instance of ACI BGWs mandatory in each Pod
- For each BD extended across domains, a specific ACI BGW is elected as DF (across all the BGWs in all the Pods)

Heterogeneous Fabrics

Independent ACI Fabrics Support

Future

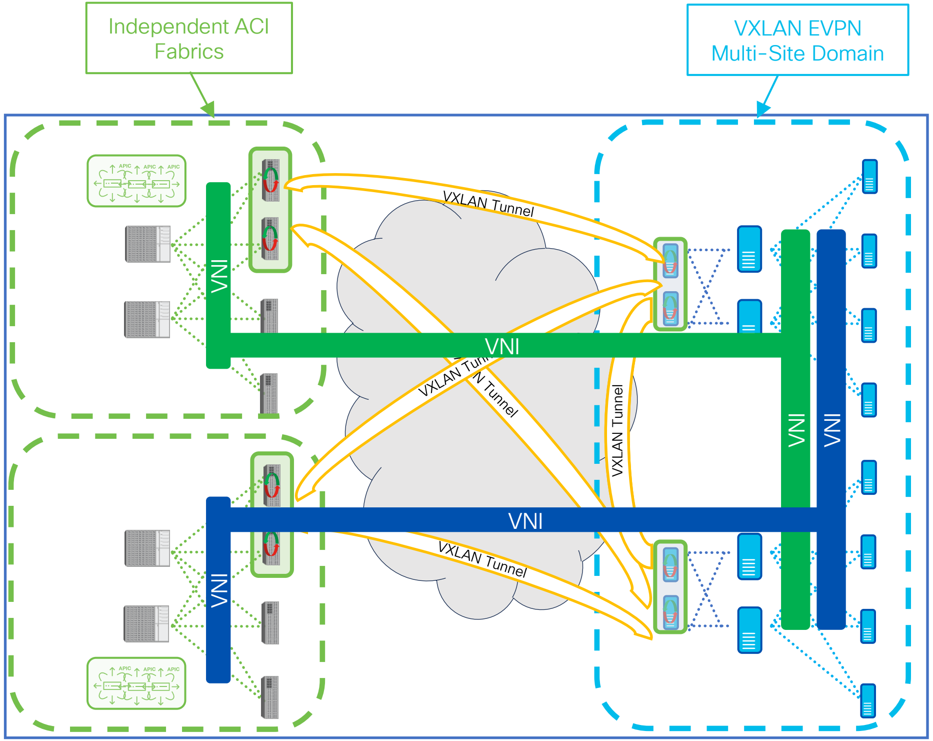


- Routed communications only via L3Out path possible between independent ACI fabrics
  - No VXLAN EVPN connectivity between ACI BGWs of different ACI Fabrics
- Different sets of VRFs/BDs can be extended between each ACI fabric and the VXLAN EVPN domain

Heterogeneous Fabrics

Independent ACI Fabrics Support

Future

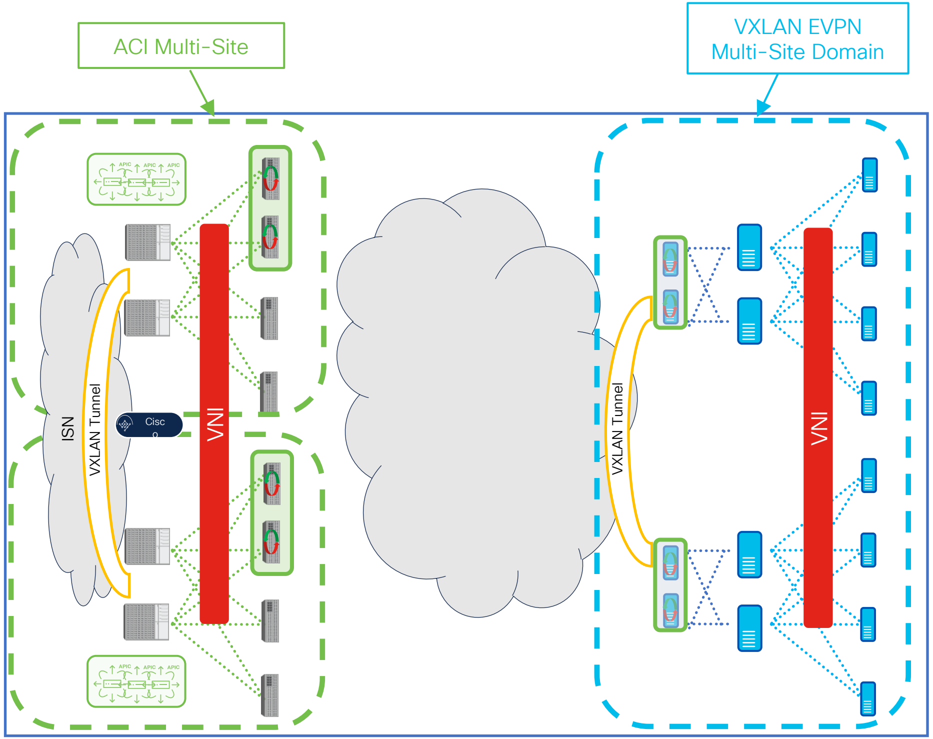


- Routed communications only via L3Out path possible between independent ACI fabrics
  - No VXLAN EVPN connectivity between ACI BGWs of different ACI Fabrics
- Different sets of VRFs/BDs can be extended between each ACI fabric and the VXLAN EVPN domain

Heterogeneous Fabrics

ACI Multi-Site Support

Future

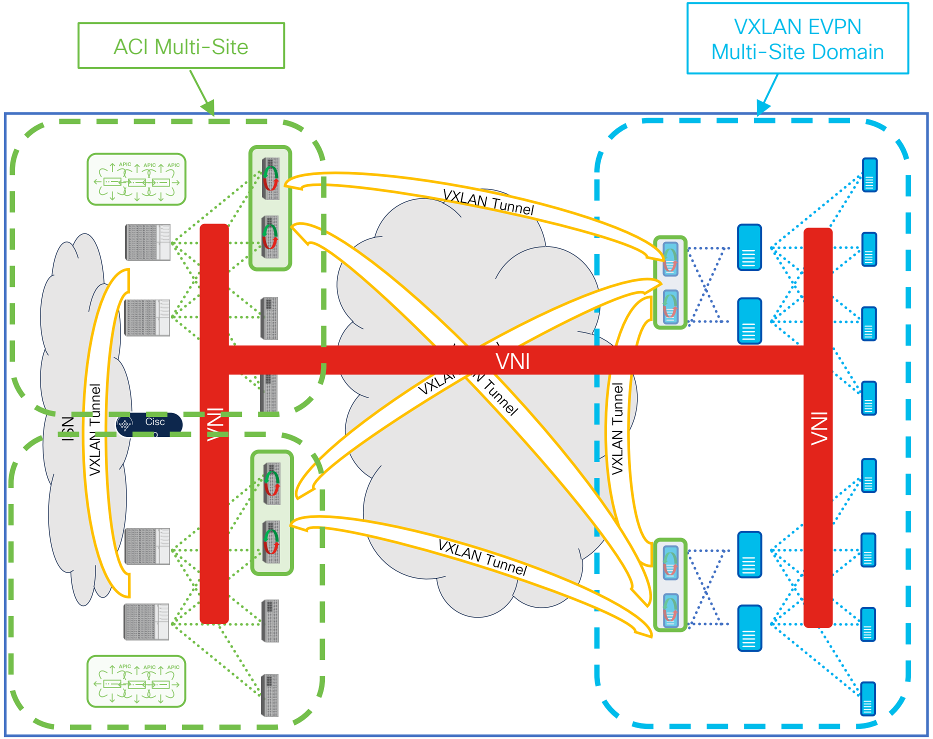


- L2/L3 VXLAN connectivity between ACI fabrics achieved via the spine-to-spine data path
  - No VXLAN EVPN connectivity between ACI BGWs of different ACI fabrics
- Each ACI fabric leverages a local instance of ACI BGWs to establish VXLAN EVPN connectivity with other domains
- NDO used for extending connectivity between ACI fabrics

Heterogeneous Fabrics

ACI Multi-Site Support

Future



- L2/L3 VXLAN connectivity between ACI fabrics achieved via the spine-to-spine data path
  - No VXLAN EVPN connectivity between ACI BGWs of different ACI fabrics
- Each ACI fabric leverages a local instance of ACI BGWs to establish VXLAN EVPN connectivity with other domains
- NDO used for extending connectivity between ACI fabrics

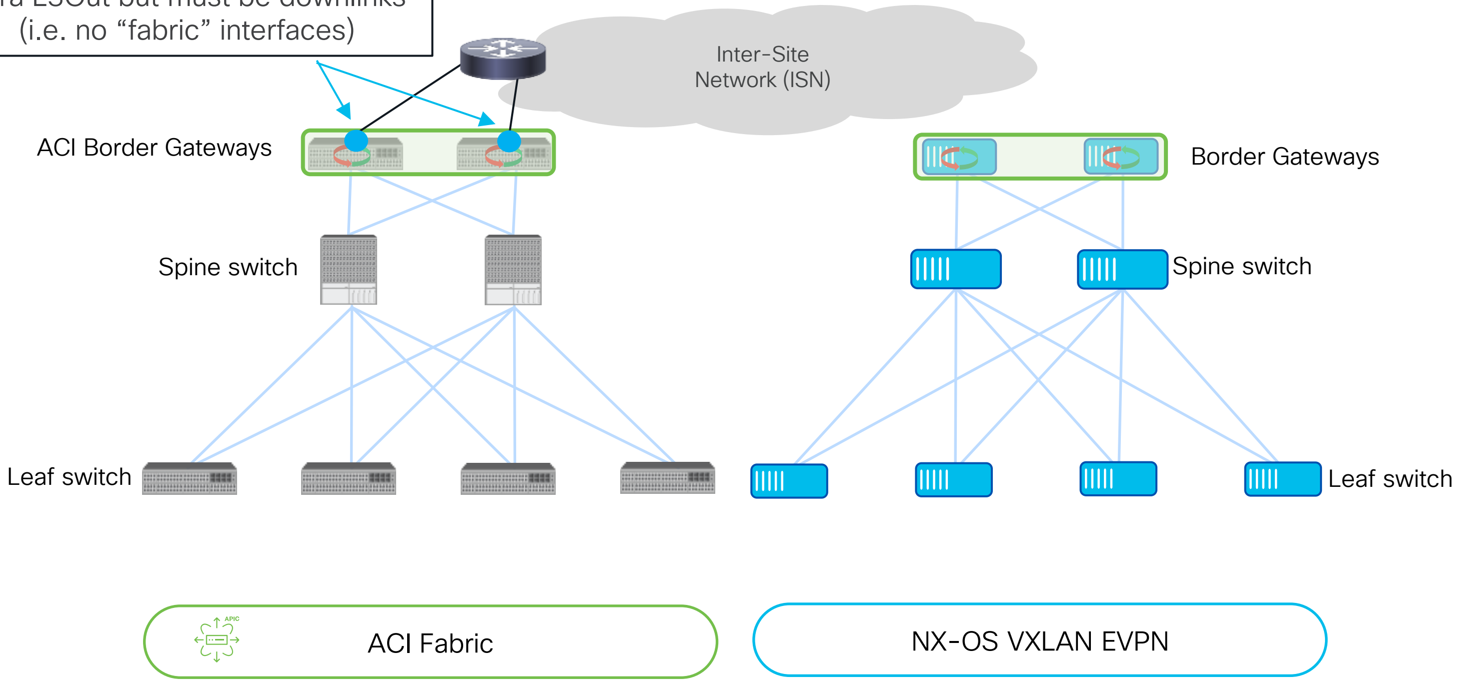
ACI Border Gateways

Overview of Control-Plane and Data-Plane

ACI Border Gateways

Definition of an Infra L3Out to Connect ACI BGWs to the ISN

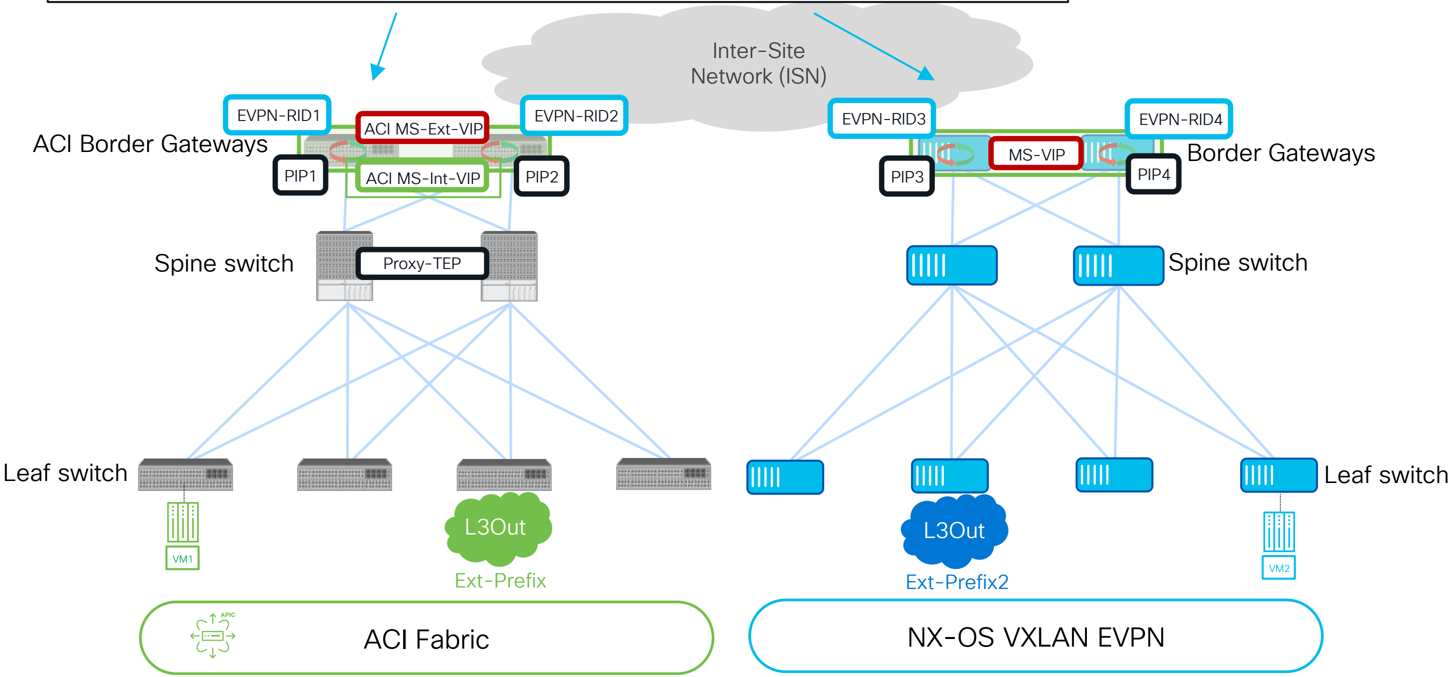
ACI BGWs interfaces are part of an Infra L3Out but must be downlinks (i.e. no “fabric” interfaces)



ACI Border Gateways

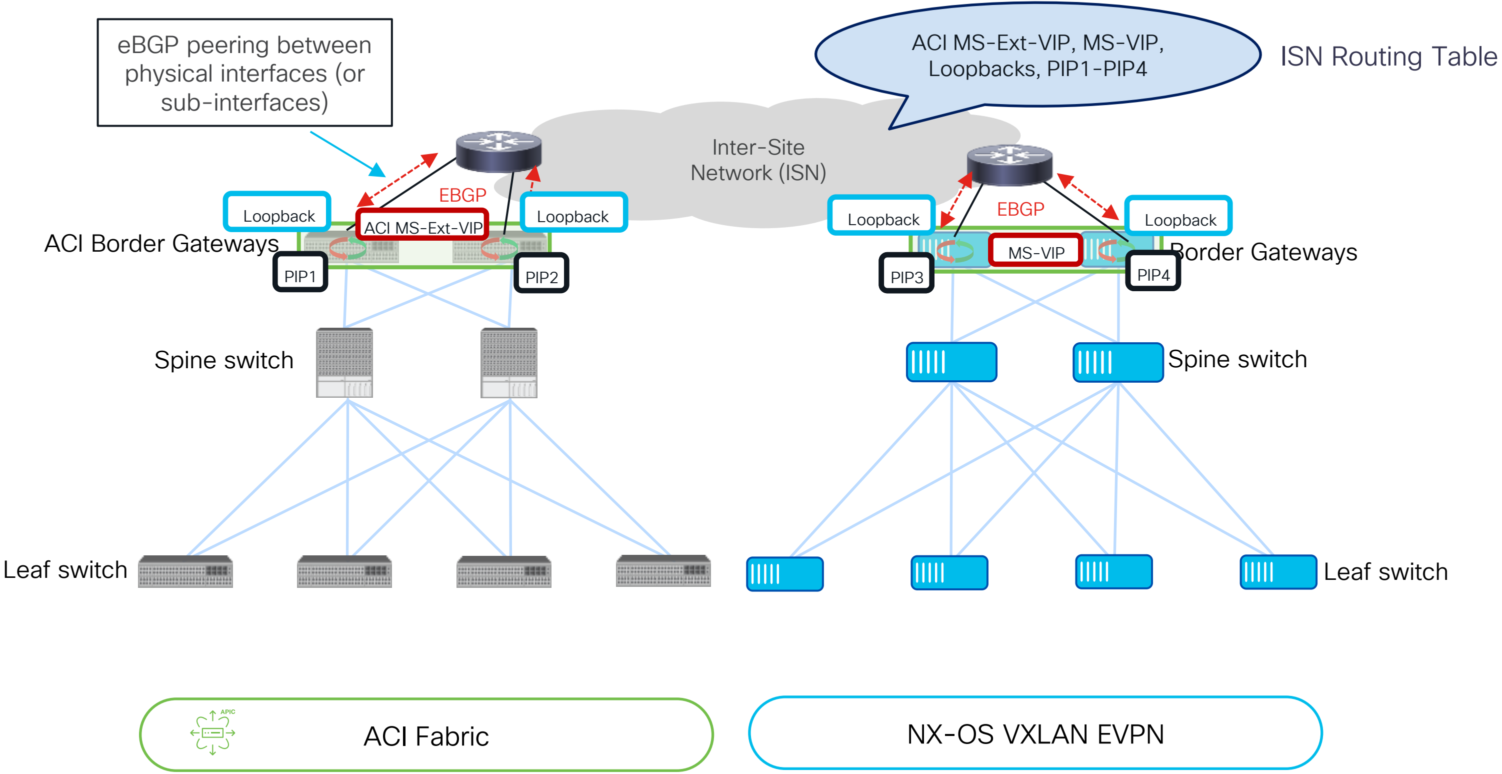
External and Internal Multi-Site VIP Addresses

Handful of /32 prefixes that need to be exchanged across domains



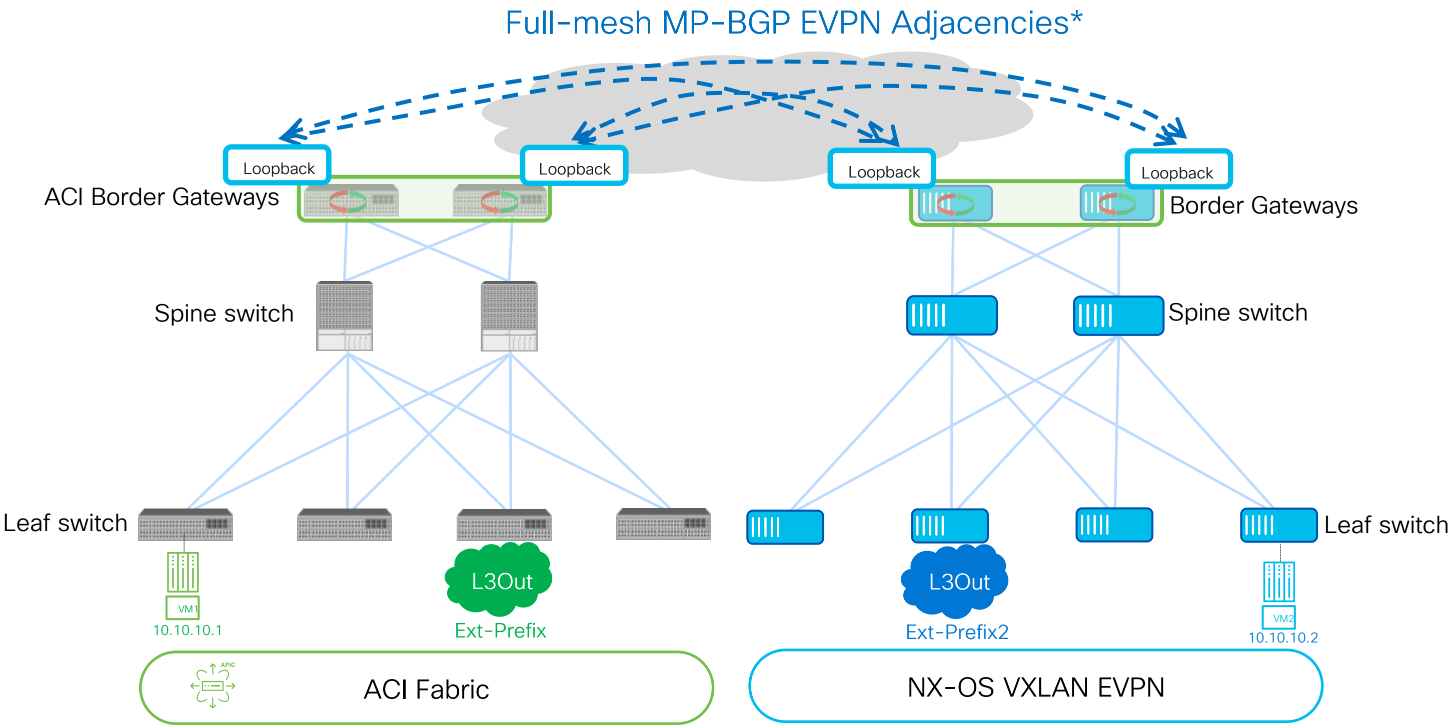
ACI Border Gateways

Underlay Control Plane Adjacencies to the ISN



ACI Border Gateways

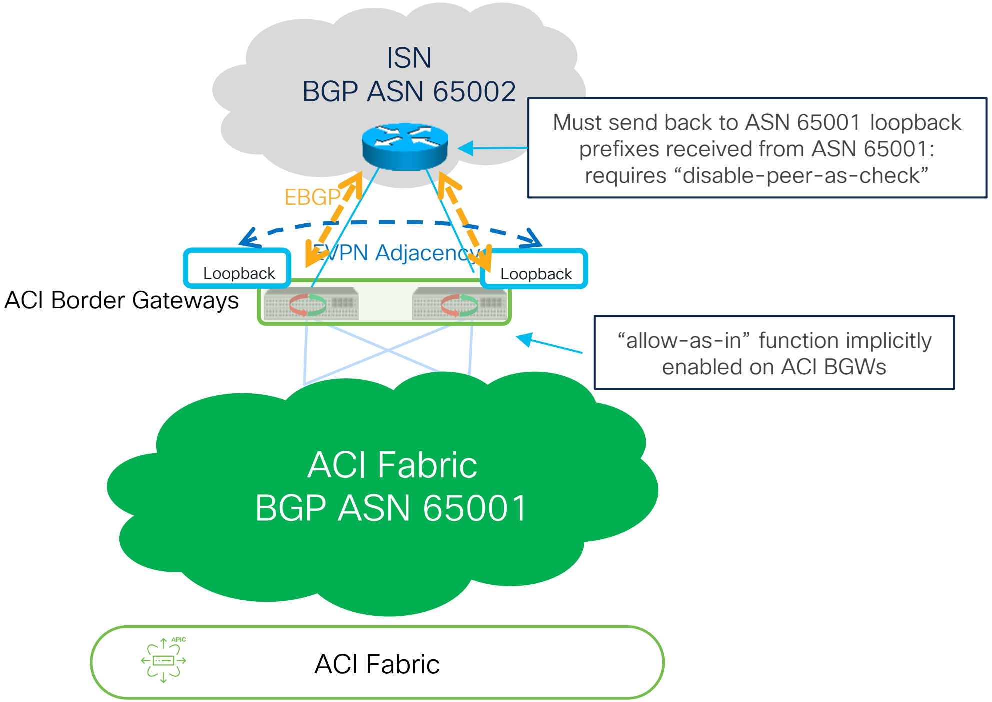
Overlay EVPN Connectivity across Domains



*No current Route-Server support

ACI Border Gateways

Overlay EVPN Connectivity between Local ACI BGWs



Must send back to ASN 65001 loopback prefixes received from ASN 65001: requires “disable-peer-as-check”

“allow-as-in” function implicitly enabled on ACI BGWs

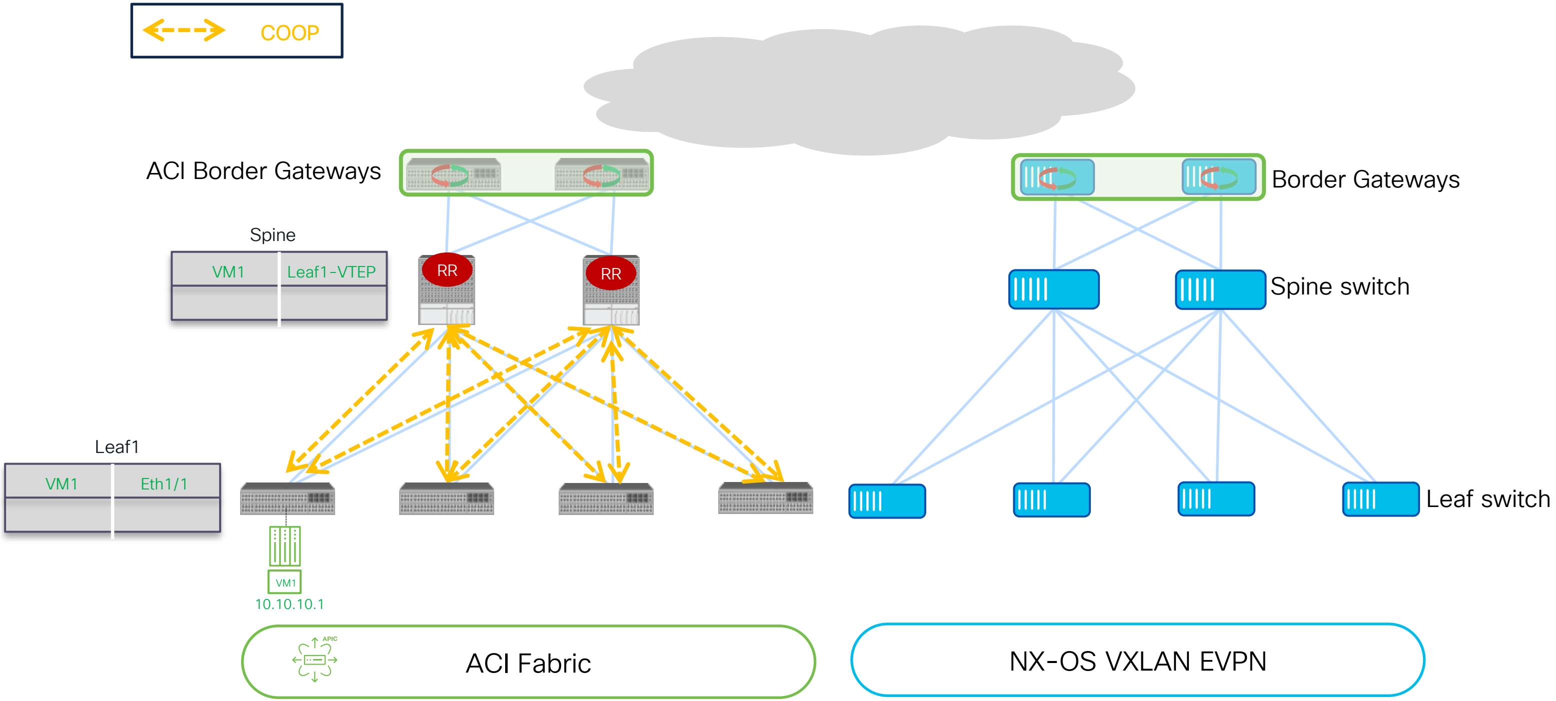
- ACI BGWs establish also local iBGP EVPN adjacencies between them
  - Used only to exchange Type-4 prefixes needed for the DF election (required for BUM forwarding toward the remote VXLAN EVPN fabrics)
- Local iBGP EVPN adjacencies are established using “external” loopback addresses as neighbor address
  - External loopback addresses need to be reachable through the ISN
  - Mandates specific configuration on the ISN device to ensure successful exchange of EVPN prefixes between the BGWs part of the same BGP ASN

```
router bgp 65002
  neighbor 192.168.3.5
      remote-as 65001
      address-family ipv4 unicast
        disable-peer-as-check
```

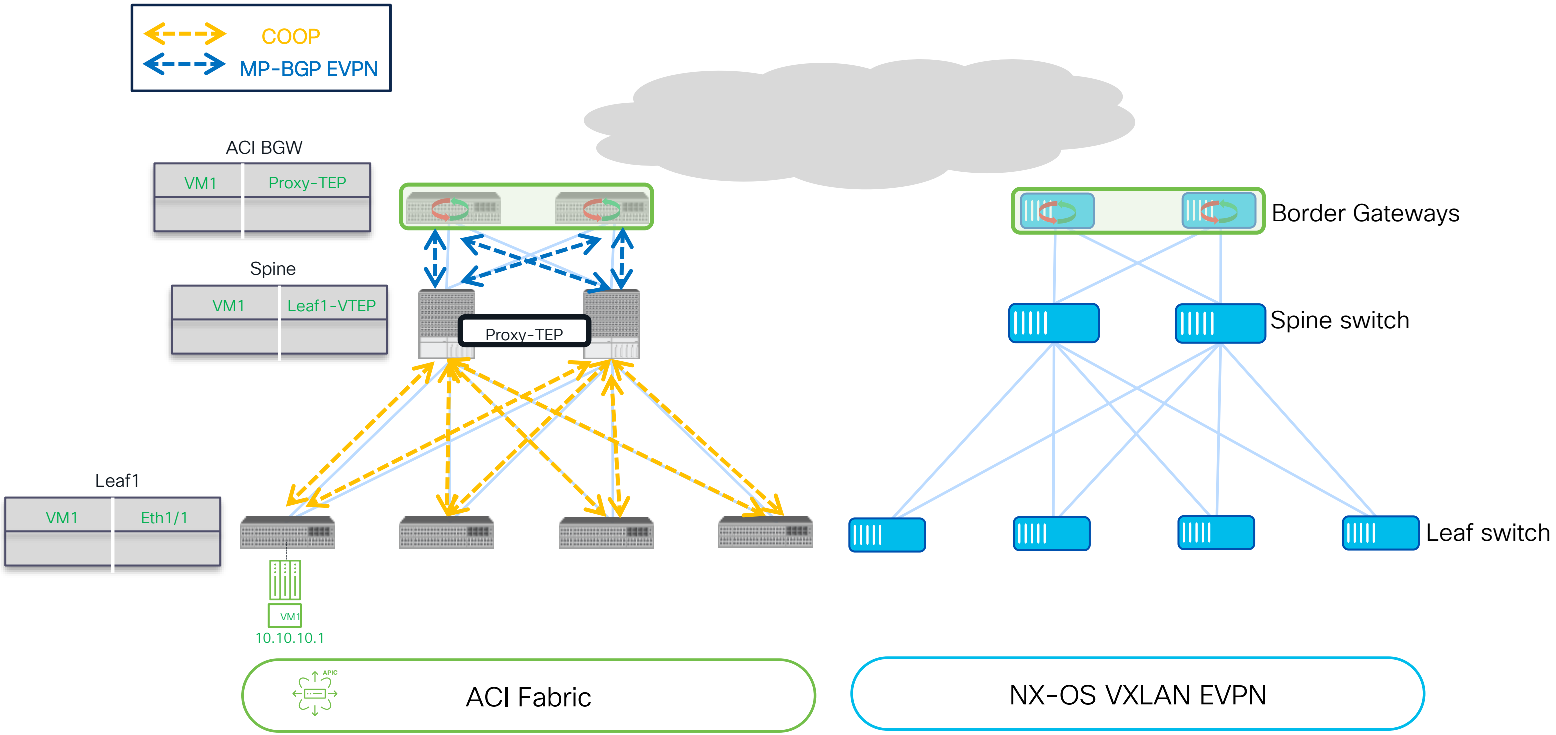
  - Same considerations apply to ACI single Pod or Multi-Pod fabric deployments

ACI Border Gateways

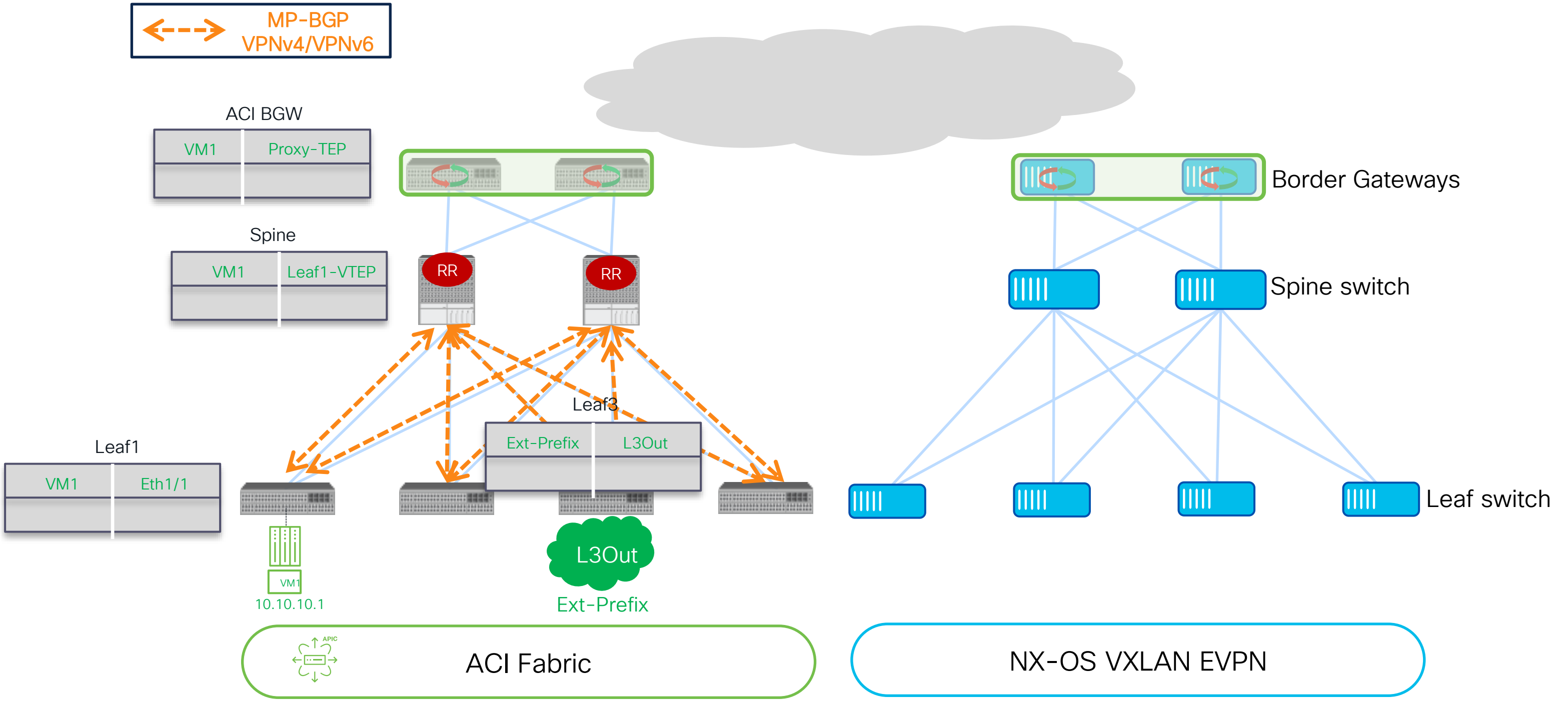
Control-Plane Overview



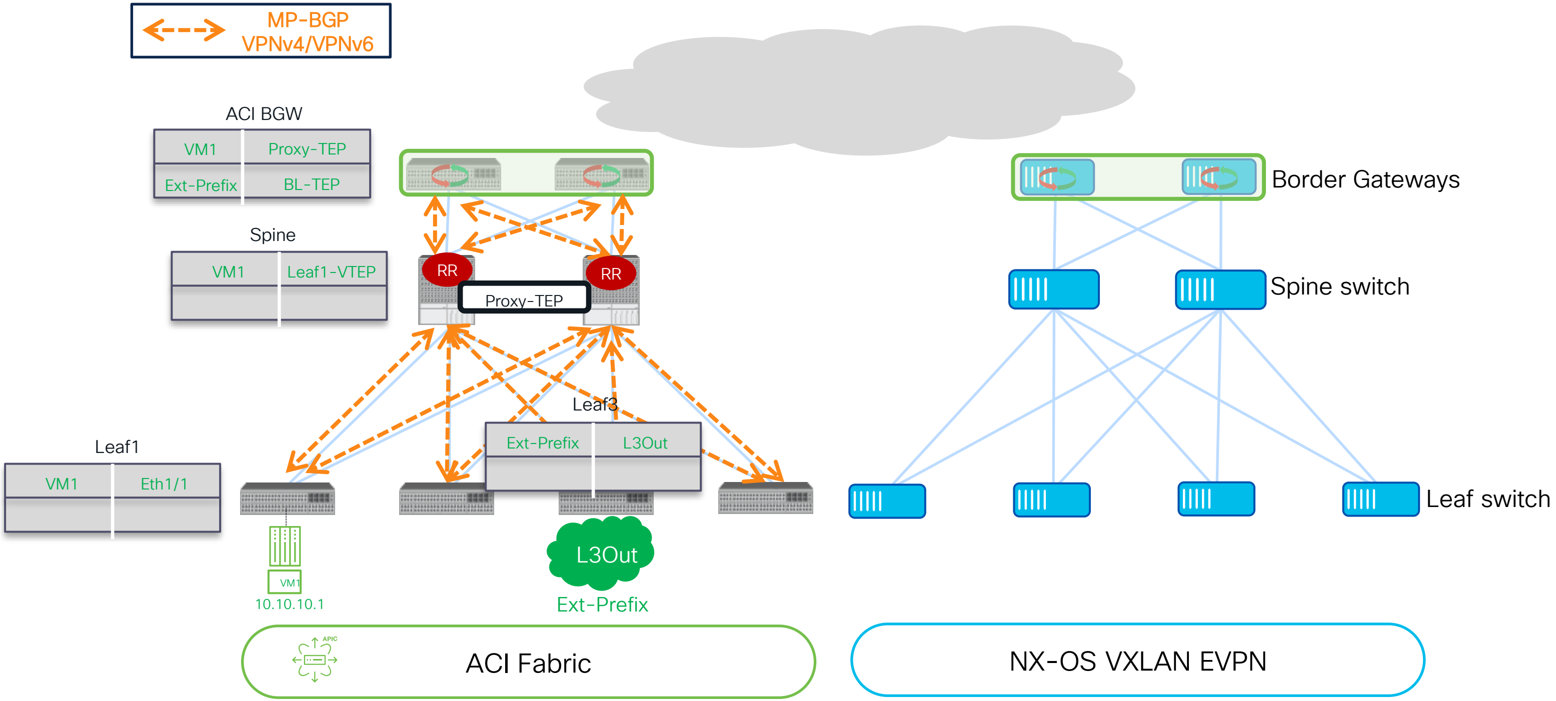
ACI Border Gateways
Control-Plane Overview



ACI Border Gateways
Control-Plane Overview

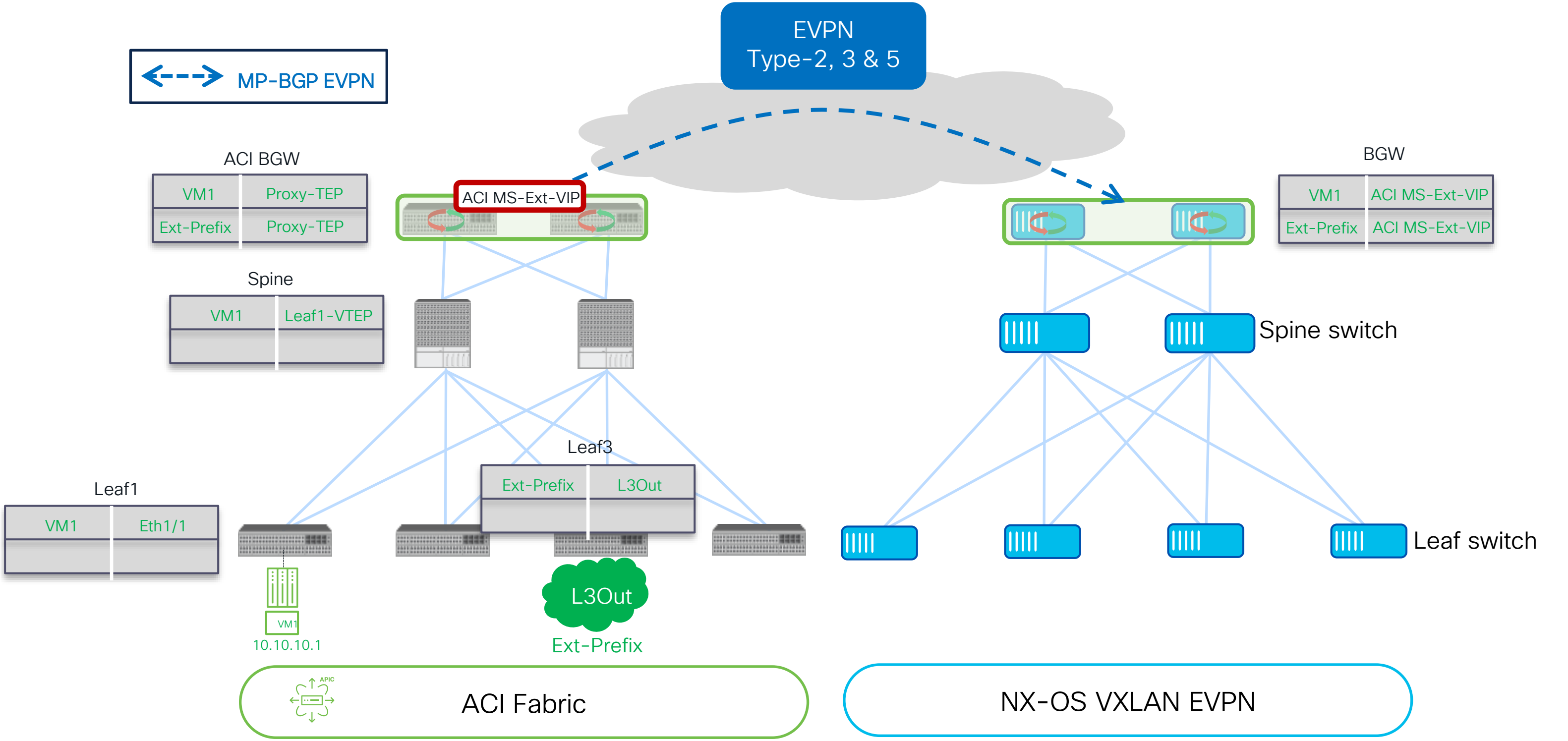


ACI Border Gateways
Control-Plane Overview



ACI Border Gateways

Control-Plane Overview



ACI BGW

VM1	Proxy-TEP
Ext-Prefix	Proxy-TEP

Spine

VM1	Leaf1-VTEP

Leaf1

VM1	Eth1/1

Leaf3

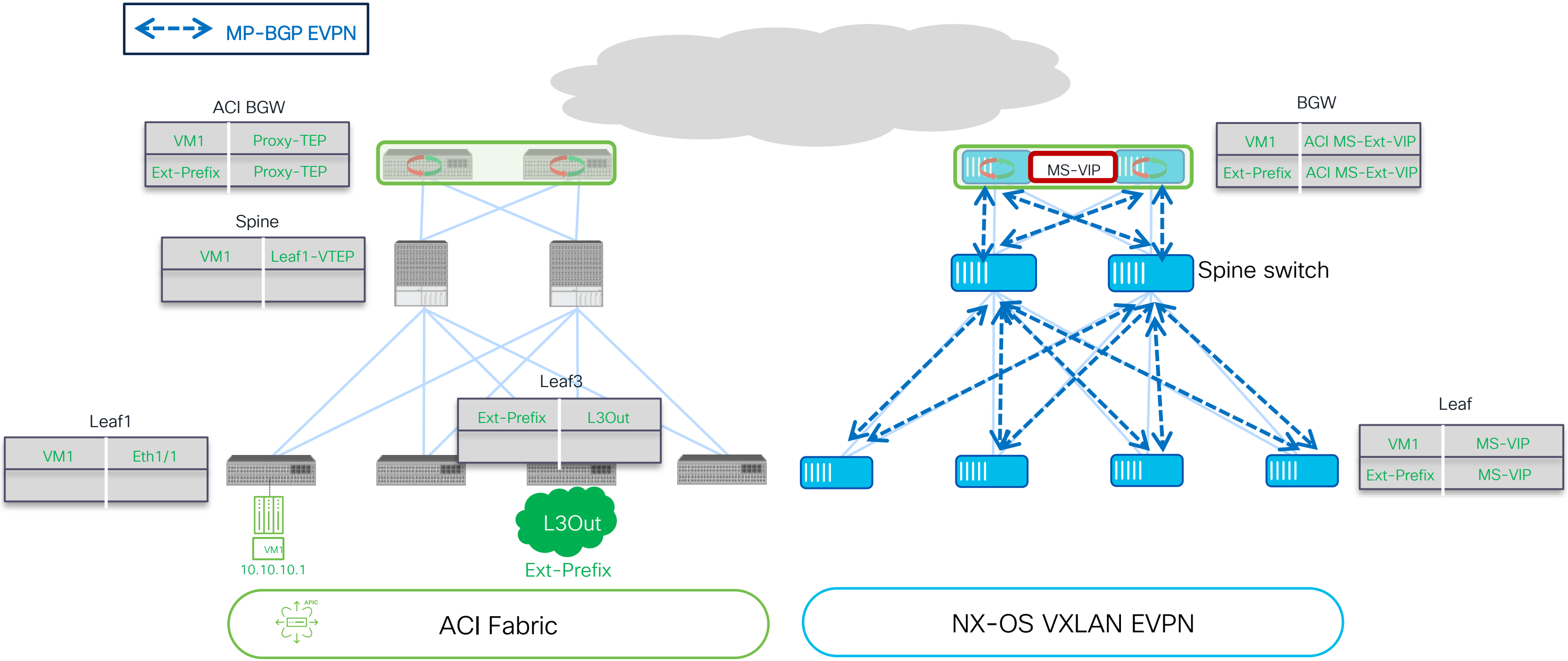
Ext-Prefix	L3Out

BGW

VM1	ACI MS-Ext-VIP
Ext-Prefix	ACI MS-Ext-VIP

ACI Border Gateways

Control-Plane Overview



ACI BGW

VM1	Proxy-TEP
Ext-Prefix	Proxy-TEP

Spine

VM1	Leaf1-VTEP

Leaf1

VM1	Eth1/1

Leaf3

Ext-Prefix	L3Out

BGW

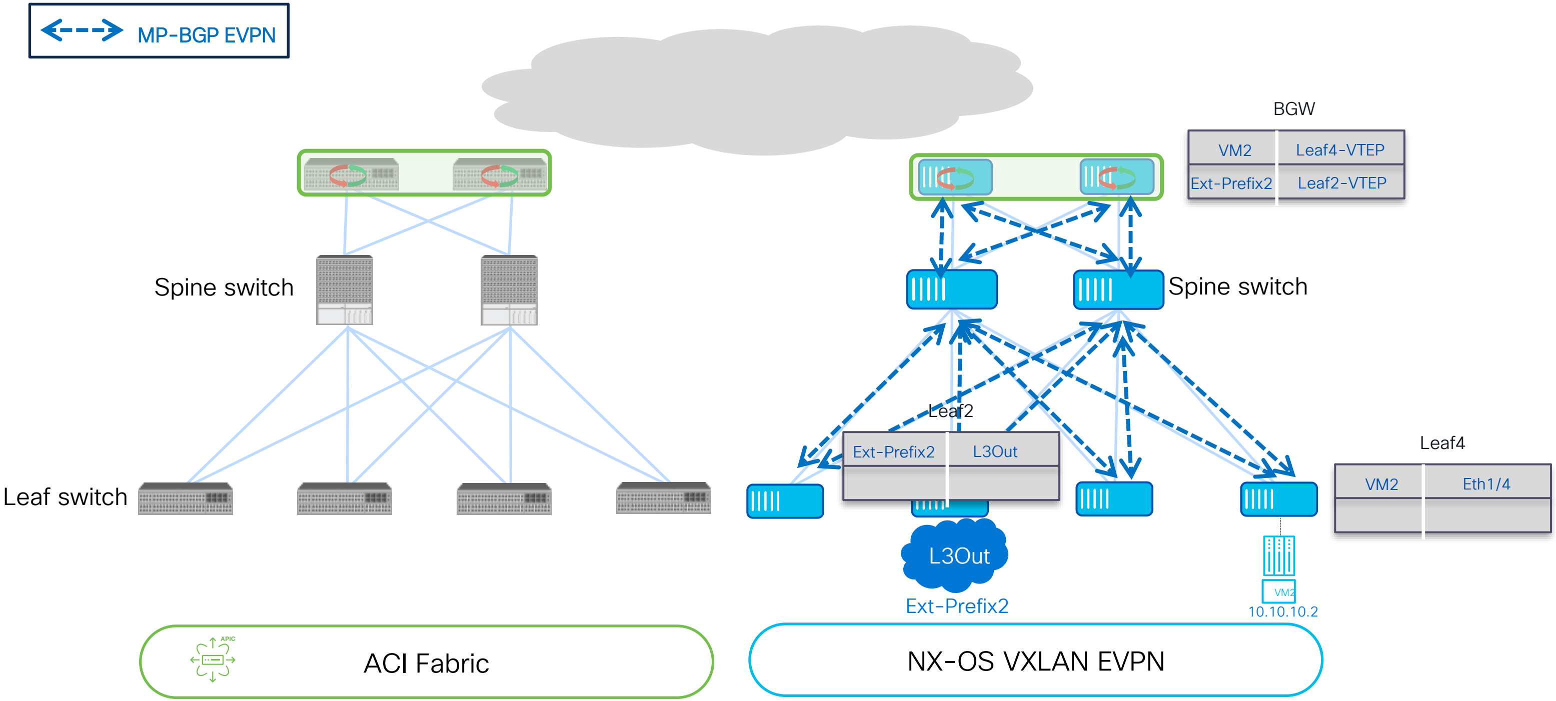
VM1	ACI MS-Ext-VIP
Ext-Prefix	ACI MS-Ext-VIP

Leaf

VM1	MS-VIP
Ext-Prefix	MS-VIP

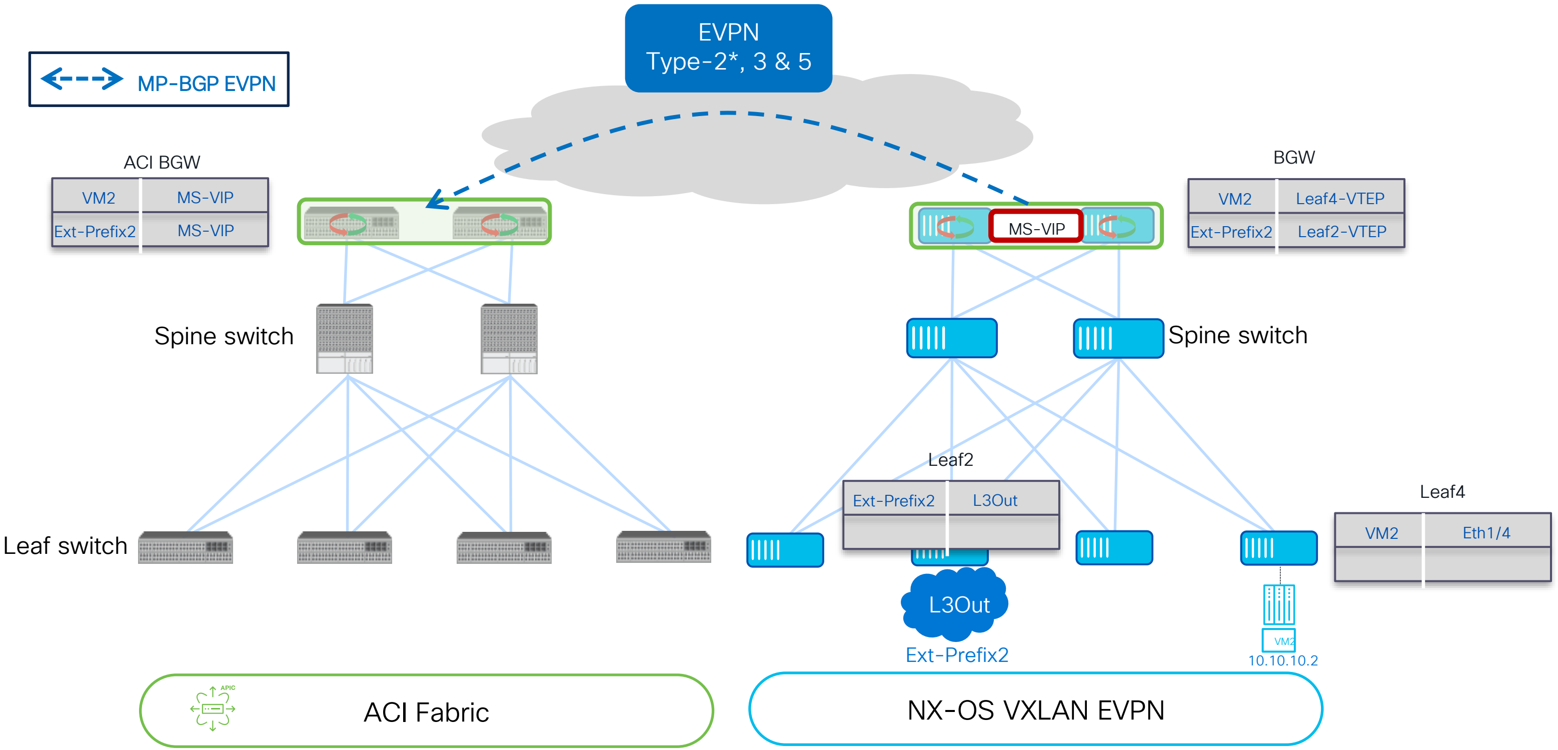
ACI Border Gateways

Control-Plane Overview



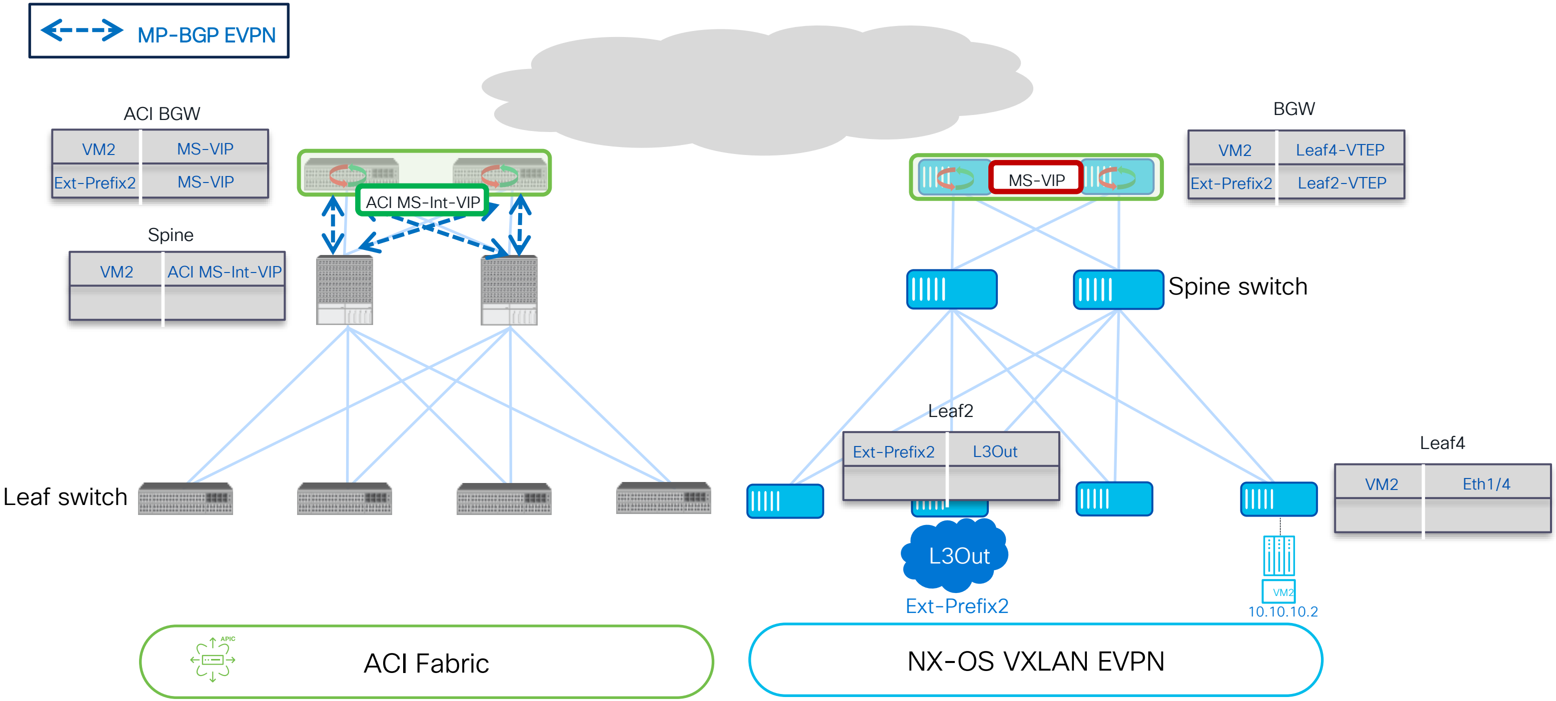
ACI Border Gateways

Control-Plane Overview



*Type-2 not advertised if VM2's network is not stretched across domains

ACI Border Gateways
Control-Plane Overview



ACI BGW

VM2	MS-VIP
Ext-Prefix2	MS-VIP

Spine

VM2	ACI MS-Int-VIP

BGW

VM2	Leaf4-VTEP
Ext-Prefix2	Leaf2-VTEP

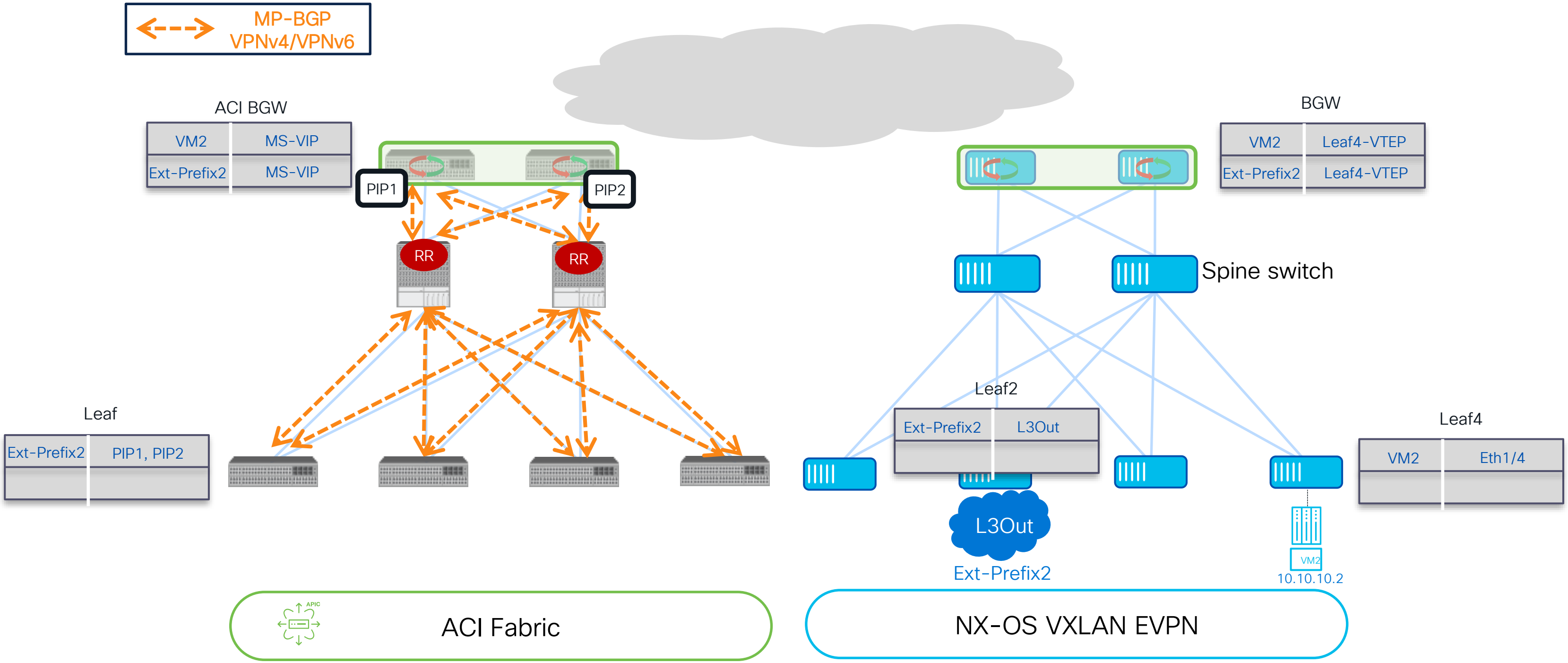
Leaf2

Ext-Prefix2	L3Out

Leaf4

VM2	Eth1/4

ACI Border Gateways
Control-Plane Overview



ACI BGW

VM2	MS-VIP
Ext-Prefix2	MS-VIP

BGW

VM2	Leaf4-VTEP
Ext-Prefix2	Leaf4-VTEP

Leaf

Ext-Prefix2	PIP1, PIP2

Leaf2

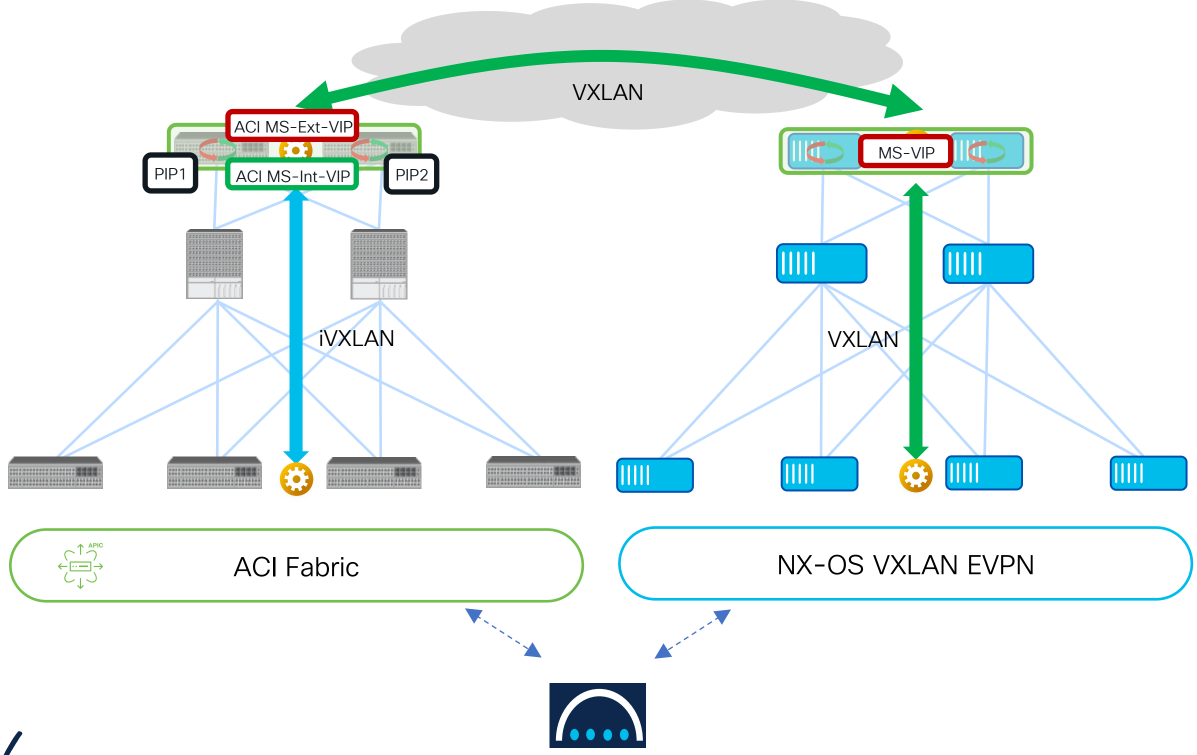
Ext-Prefix2	L3Out

Leaf4

VM2	Eth1/4

ACI Border Gateways
Data-Plane Overview

Cross-fabrics End-to-End Connectivity through Tunnel Stitching



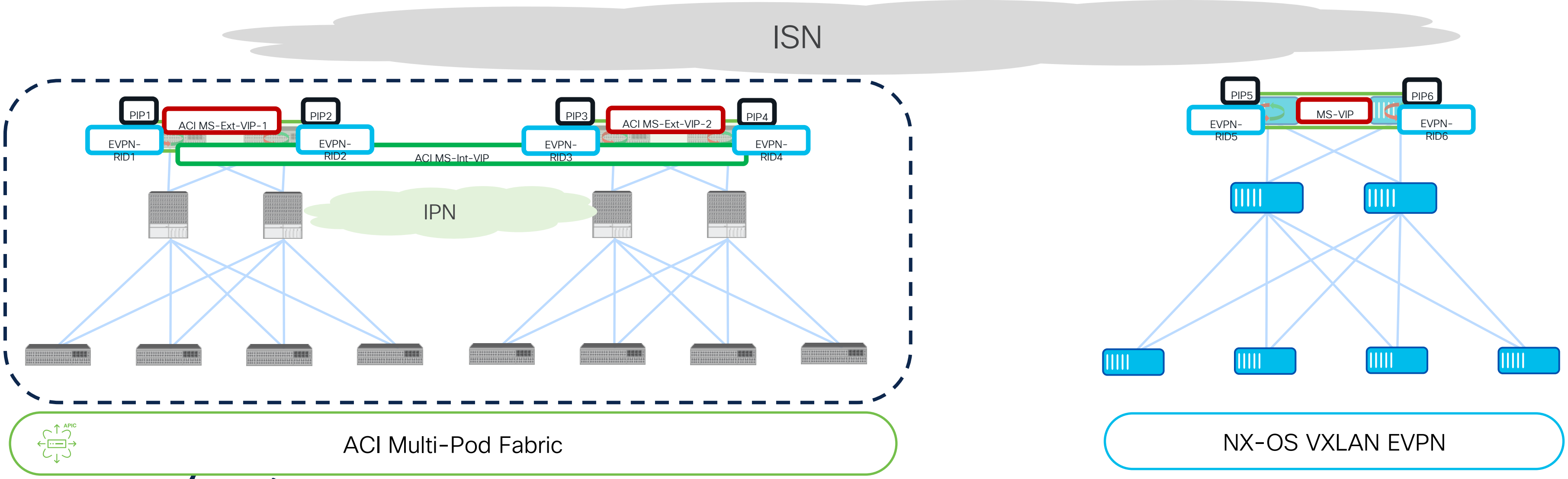
ACI Multi-Pod Fabric

ACI Multi-Pod Fabric with BGWs

Deployment Considerations

ACI 6.1(1)

- ACI Multi-Site External VIP unique per Pod, ACI Multi-Site Internal VIP common across Pods
- Pod transit not supported → all Pods must locally deploy BGW nodes
- If IPN and ISN are the same network infrastructure, separate VRFs must be used for Multi-Pod and cross-domains traffic
  - MS-EXT-VIPs, MS-VIP, EVPN-RIDs and PIPs info must only be exchanged via ACI BGWs (and not via ACI spines)



ACI Border Gateways

Summary of TEP Addresses and Their Purpose

TEP	Use
Multi-Site Internal VIP	• Used as source IP for all traffic received from the VXLAN EVPN domain and re-encapsulated to be sent into the ACI fabric • Used as destination IP for all traffic destined to remote endpoints part of stretched BDs • Common value assigned to all the spines part of the same fabric (single pod or Multi-Pod)
Multi-Site External VIP	• Anycast IP used as next-hop for Type-2 and Type-5 EVPN routes advertised to remote VXLAN EVPN fabrics • Used as destination IP for L2/L3 communications initiated from the VXLAN EVPN domain toward endpoints or external networks reachable via ACI • Unique value per Pod
BGW PIP	• Used as source IP for all traffic sent toward the remote VXLAN EVPN domain • Used as next-hop for Type-5 prefixes received from the VXLAN EVPN domain and injected into the ACI fabric using VPNv4/VPNv6 • Unique per BGW

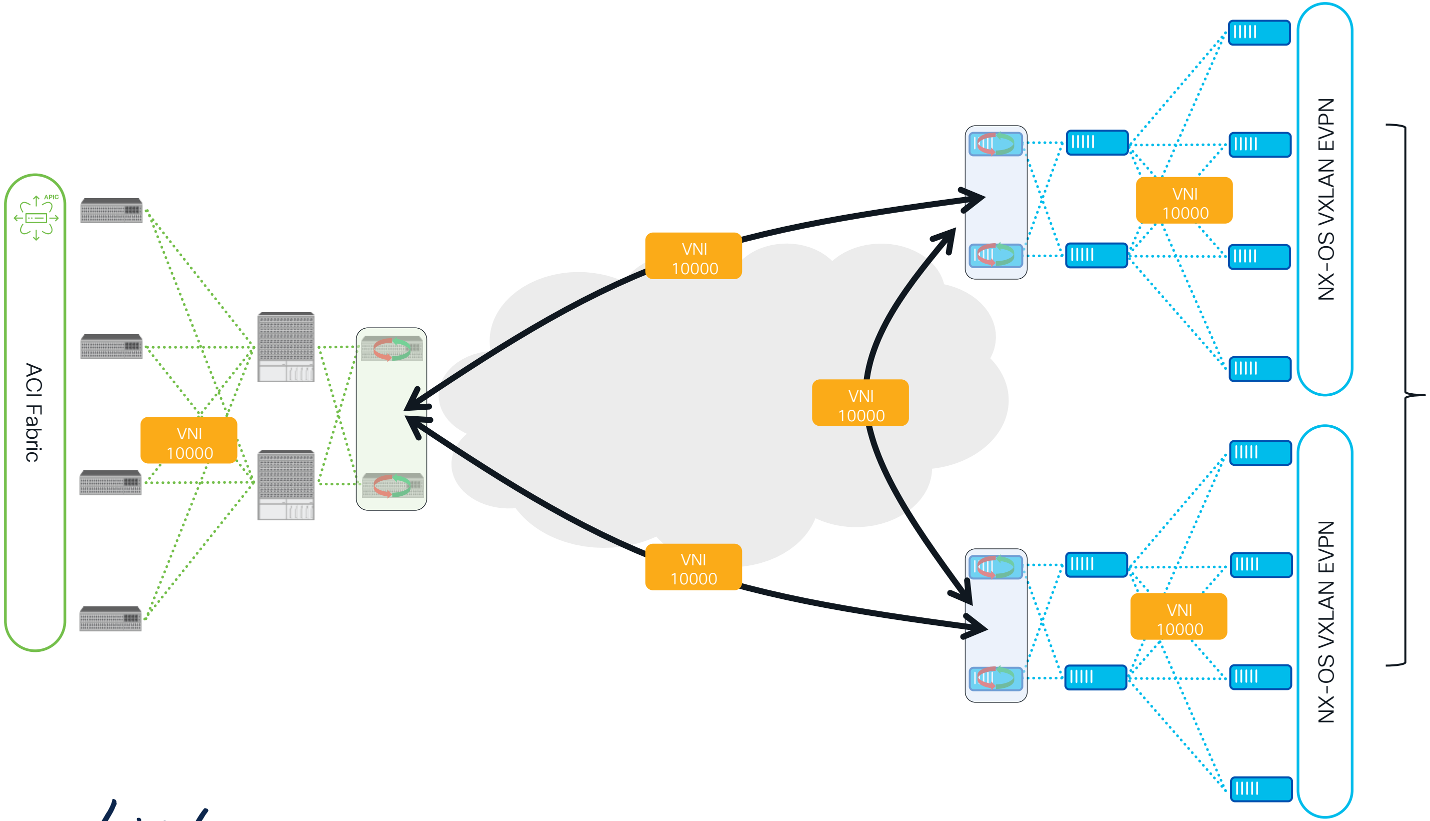
ACI Border Gateways

Namespace Normalization

ACI Border Gateways

Symmetric Namespace for Stretched BDs (or VRFs)

ACI 6.1(1)

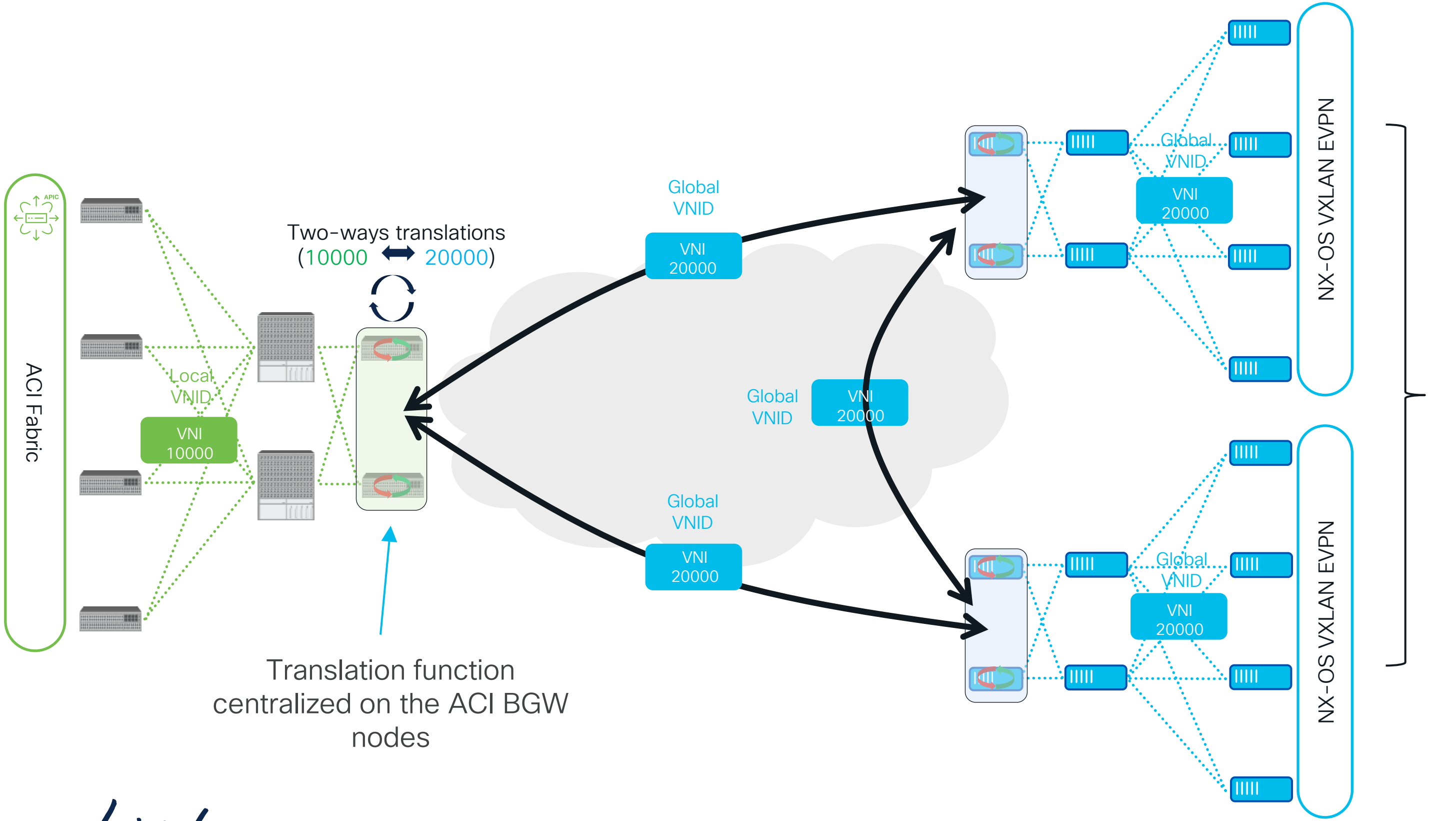


VNIDs for stretched BDs/VRFs must be provisioned on the VXLAN EVPN fabrics to **match the values assigned by APIC**

ACI Border Gateways

Centralized Namespace Normalization for Stretched BDs (or VRFs)

ACI 6.1(4)



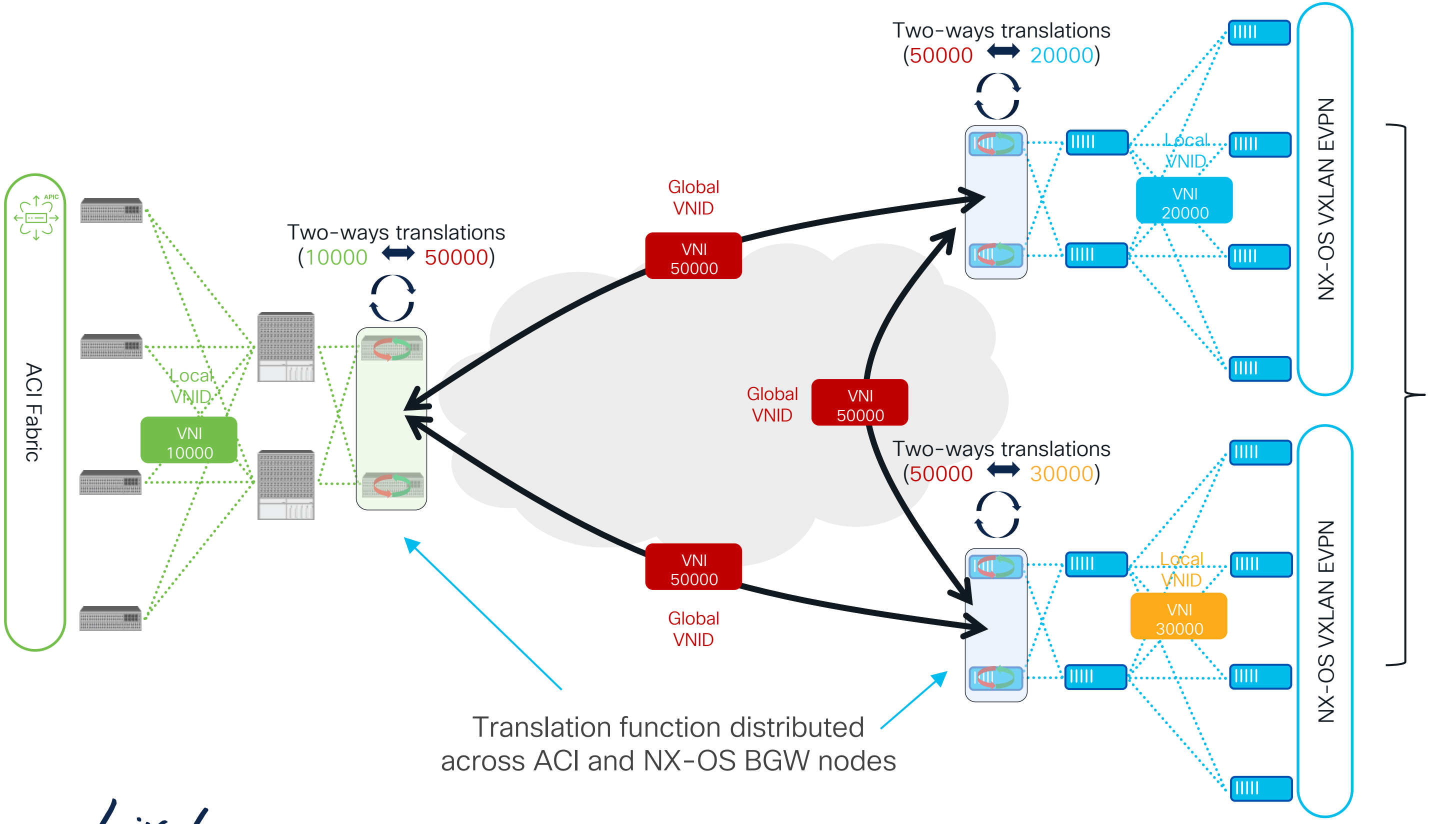
Translation function centralized on the ACI BGW nodes

Homogeneous namespace across all NX-OS VXLAN EVPN fabrics

ACI Border Gateways

Distributed Namespace Normalization for Stretched BDs/VRFs

Future



Heterogeneous namespace across NX-OS VXLAN EVPN fabrics

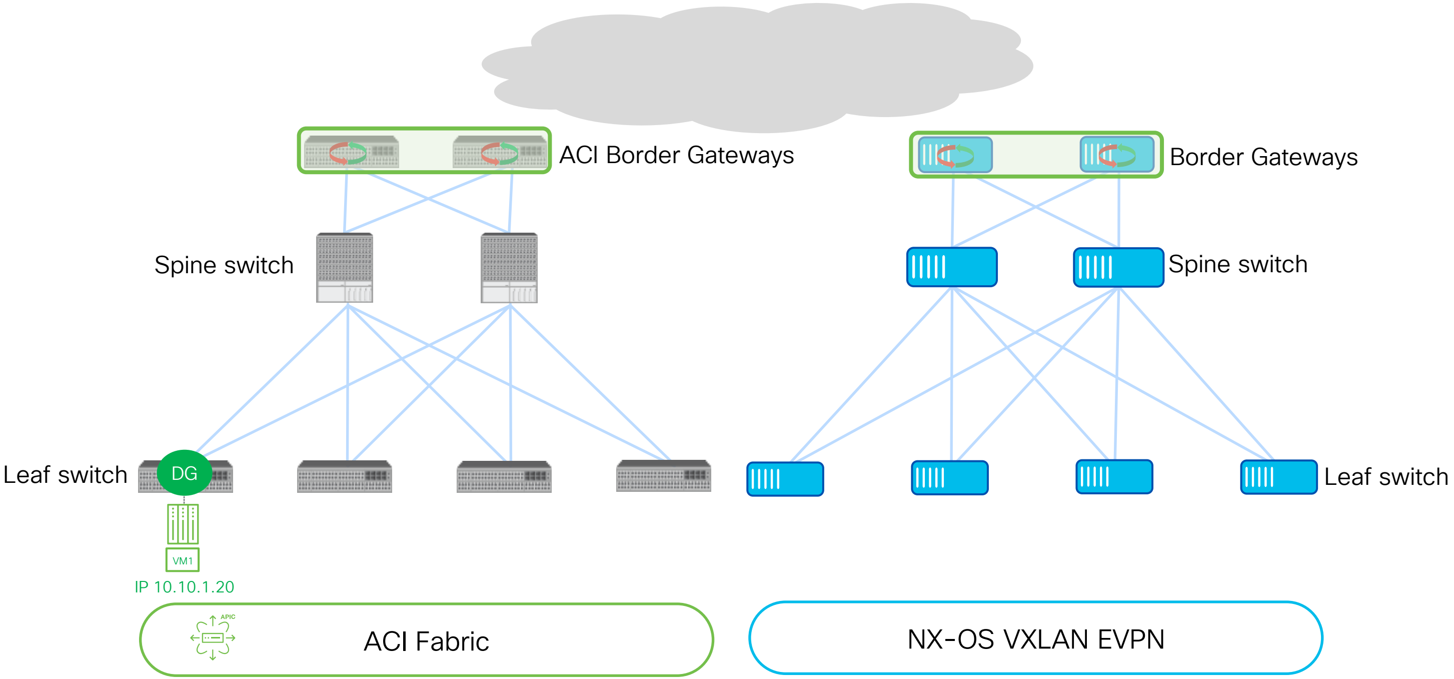
Translation function distributed across ACI and NX-OS BGW nodes

ACI Border Gateways

Workload Mobility across Domains

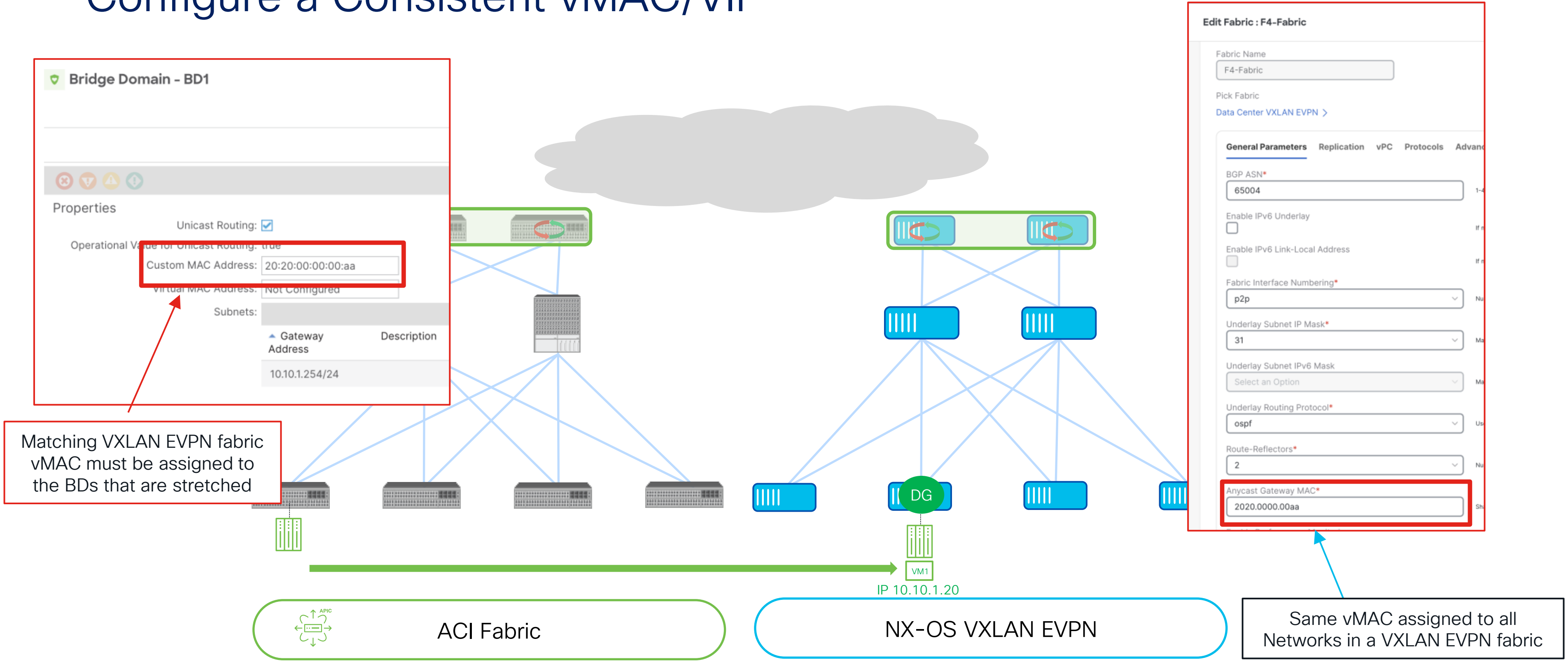
Workload Mobility

Configure a Consistent vMAC/VIP



Workload Mobility

Configure a Consistent vMAC/VIP



Bridge Domain - BD1

Properties

Unicast Routing: ☑

Operational Value for Unicast Routing: true

Custom MAC Address: 20:20:00:00:00:aa

Virtual MAC Address: Not Configured

Subnets:

Gateway Address	Description
10.10.1.254/24	

Matching VXLAN EVPN fabric vMAC must be assigned to the BDs that are stretched

Edit Fabric : F4-Fabric

Fabric Name: F4-Fabric

Pick Fabric: Data Center VXLAN EVPN >

General Parameters | Replication | vPC | Protocols | Advanc

BGP ASN*: 65004

Enable IPv6 Underlay: □

Enable IPv6 Link-Local Address: □

Fabric Interface Numbering*: p2p

Underlay Subnet IP Mask*: 31

Underlay Subnet IPv6 Mask: Select an Option

Underlay Routing Protocol*: ospf

Route-Reflectors*: 2

Anycast Gateway MAC*: 2020.0000.00aa

Same vMAC assigned to all Networks in a VXLAN EVPN fabric

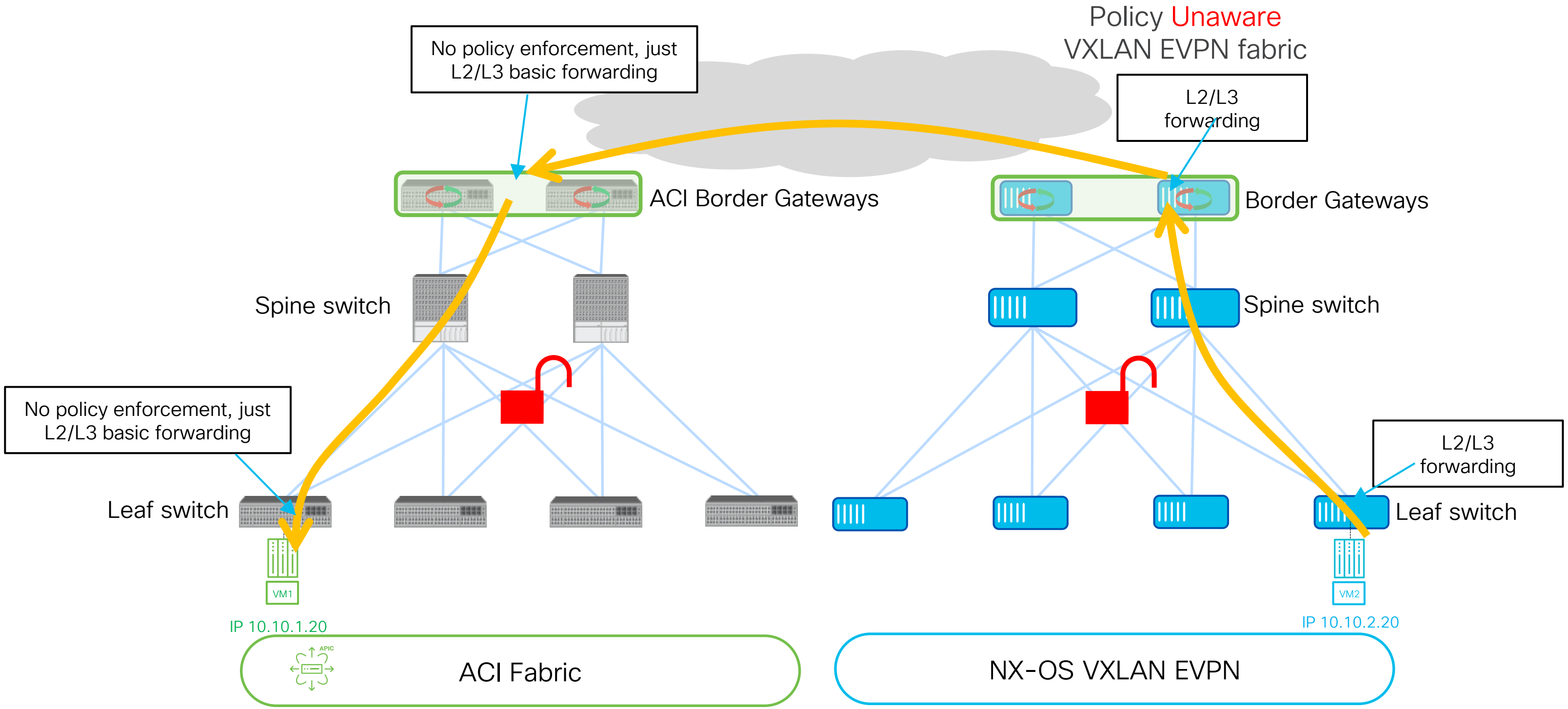
ACI Border Gateways

Policy Enforcement on ACI BGWs

Heterogeneous Fabrics

VRF Unenforced in ACI 6.1(1) Release

ACI 6.1(1)



ACI 6.1(2)

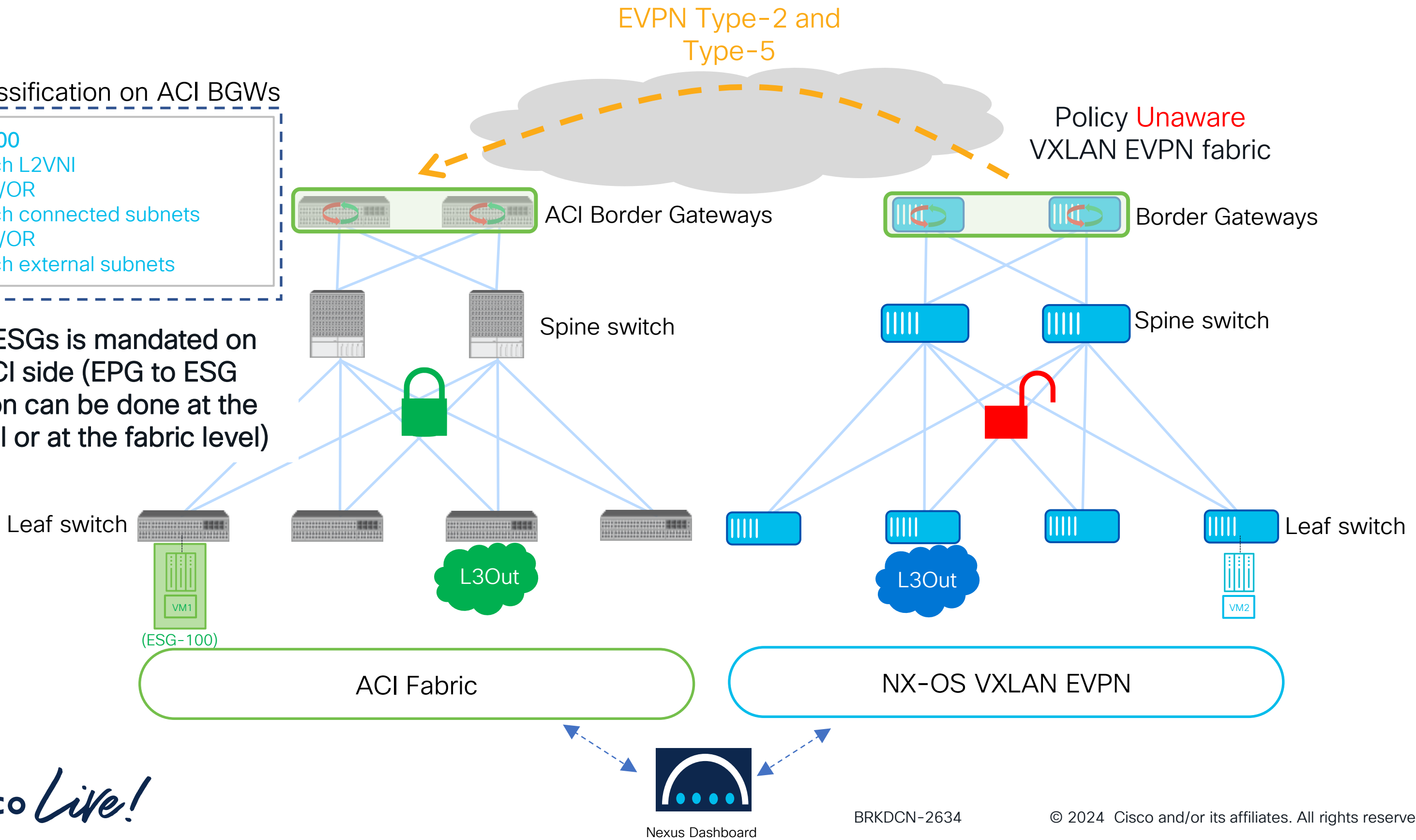
Heterogeneous Fabrics

Classification and Policy Enforcement in ACI 6.1(2) Release

ESG Classification on ACI BGWs

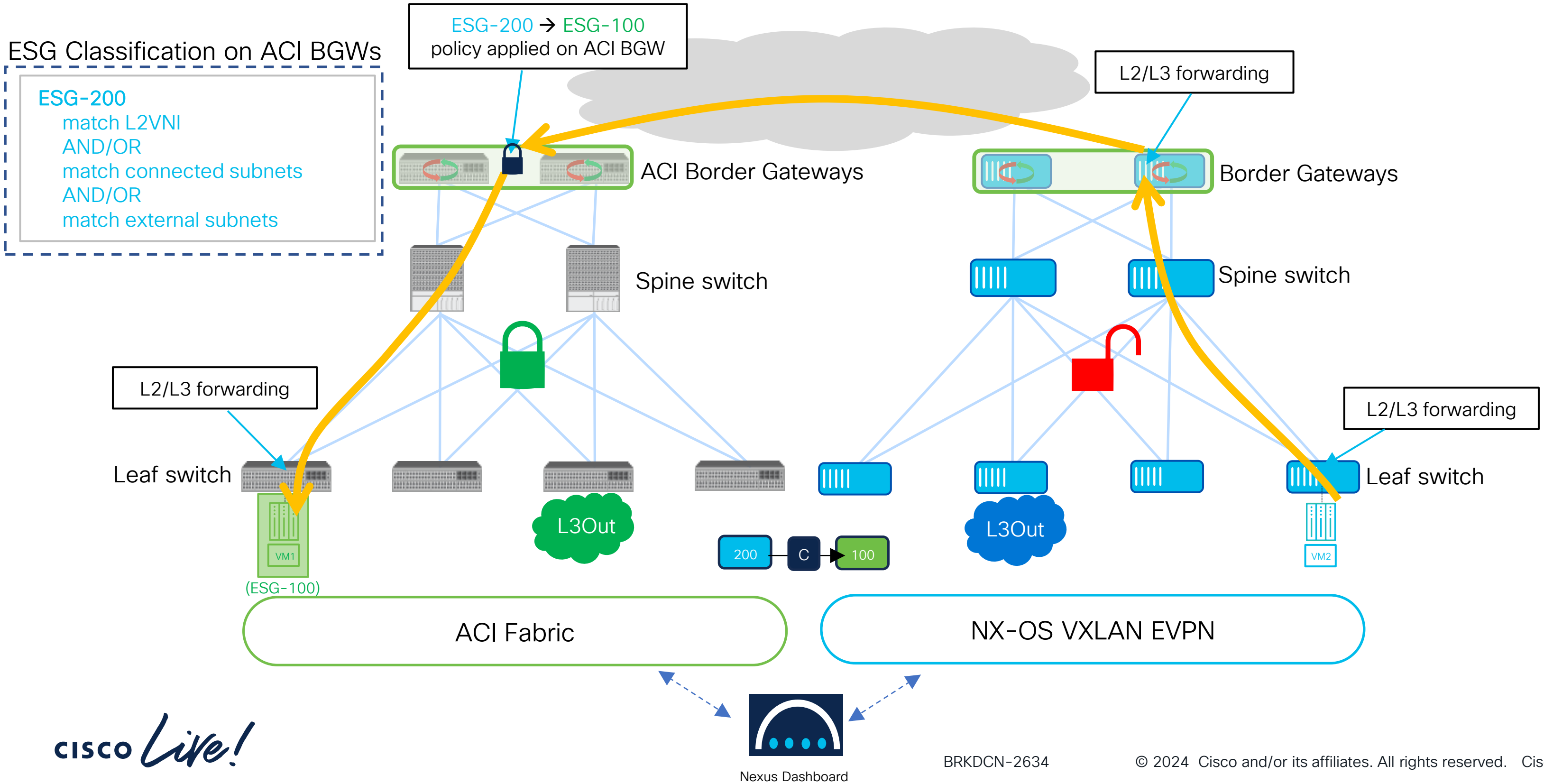
ESG-200
match L2VNI
AND/OR
match connected subnets
AND/OR
match external subnets

Use of ESGs is mandated on the ACI side (EPG to ESG migration can be done at the VRF level or at the fabric level)



Heterogeneous Fabrics

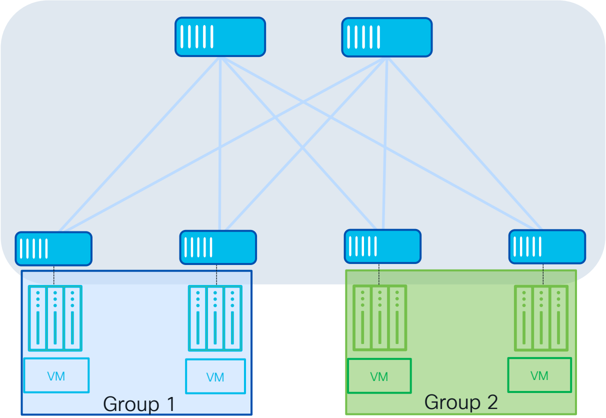
Classification and Policy Enforcement in ACI 6.1(2) Release



Secure Interconnection of Heterogeneous Fabrics

NX-OS 10.4(3)F

VXLAN GPO with NX-OS



NX-OS VXLAN EVPN

VXLAN GPO with NX-OS

- Group Policy Object carried in standard VXLAN header
- Decoupling network connectivity and security

Grouping

- Classify endpoints to create security groups
- Based on IP, VLAN, VM attributes, etc. across VRFs

Policy enforcement

- Create contracts/SGACLs between security groups
- Possible actions: permit, deny, redirect (service chaining)

Automation

- Automate using NDFC or Open APIs

Benefits

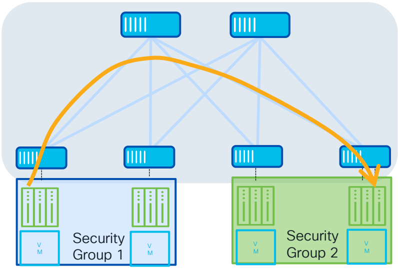
Segment East-West traffic	Flexible security isolation	Reduce attack surface	Automate your way

VXLAN GPO with NX-OS

Main Use Cases

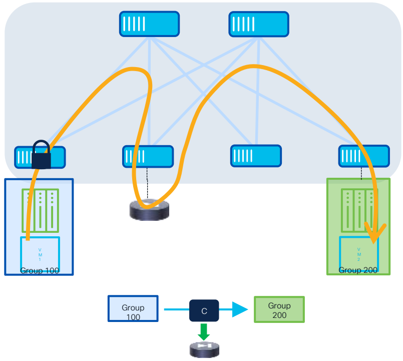
Creation of Security Zones

- VXLAN GPO allows to define policies for enforcing security policies (SGACLs) between security groups (SGs)
- SGACLs are a simpler, more flexible and more scalable policy enforcement mechanism compared to traditional ACLs
- Provides better control over the flow of network traffic (both east-west and north-south)



Service Chaining

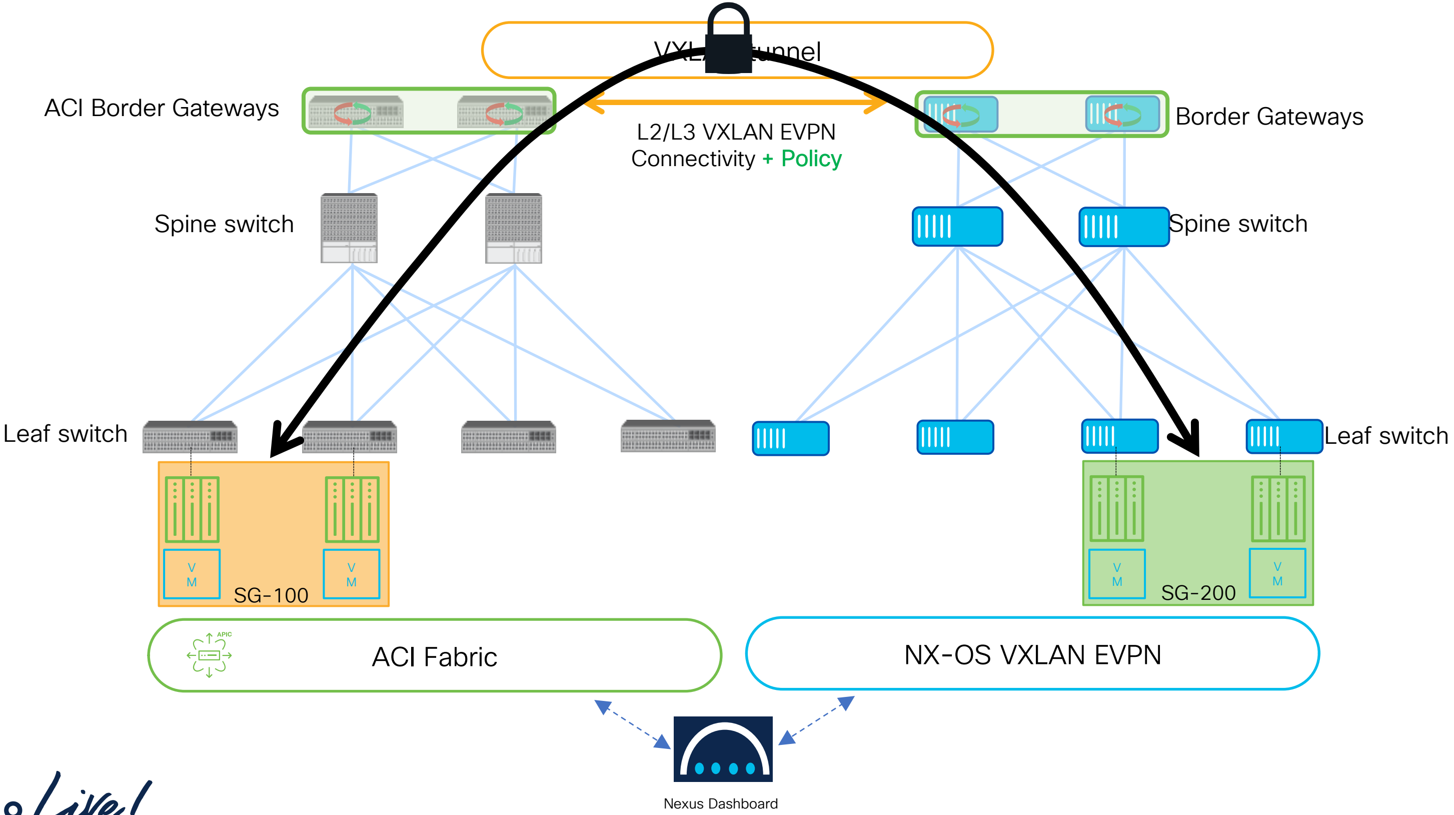
- VXLAN GPO can be used to insert network services into a packet flow based on specific policy criteria
- Service chaining steers flows through the appropriate network services functions (such as firewalls, load balancers, or intrusion detection systems)



Heterogeneous Fabrics

Policy Enforcement End-to-End

Future



VXLAN tunnel

ACI Border Gateways

L2/L3 VXLAN EVPN
Connectivity + Policy

Border Gateways

Spine switch

Spine switch

Leaf switch

Leaf switch

SG-100

SG-200

ACI Fabric

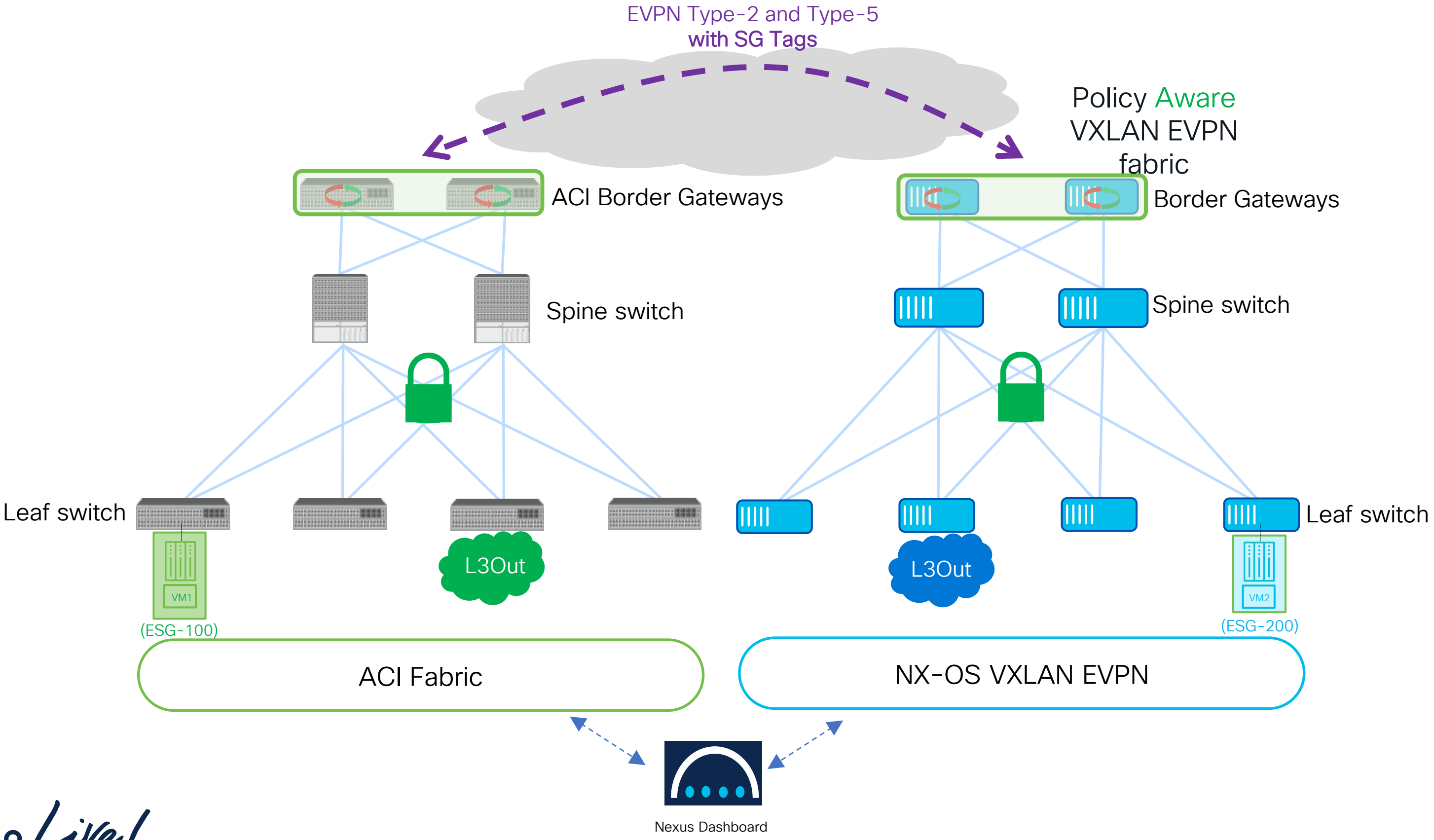
NX-OS VXLAN EVPN

Nexus Dashboard

Future

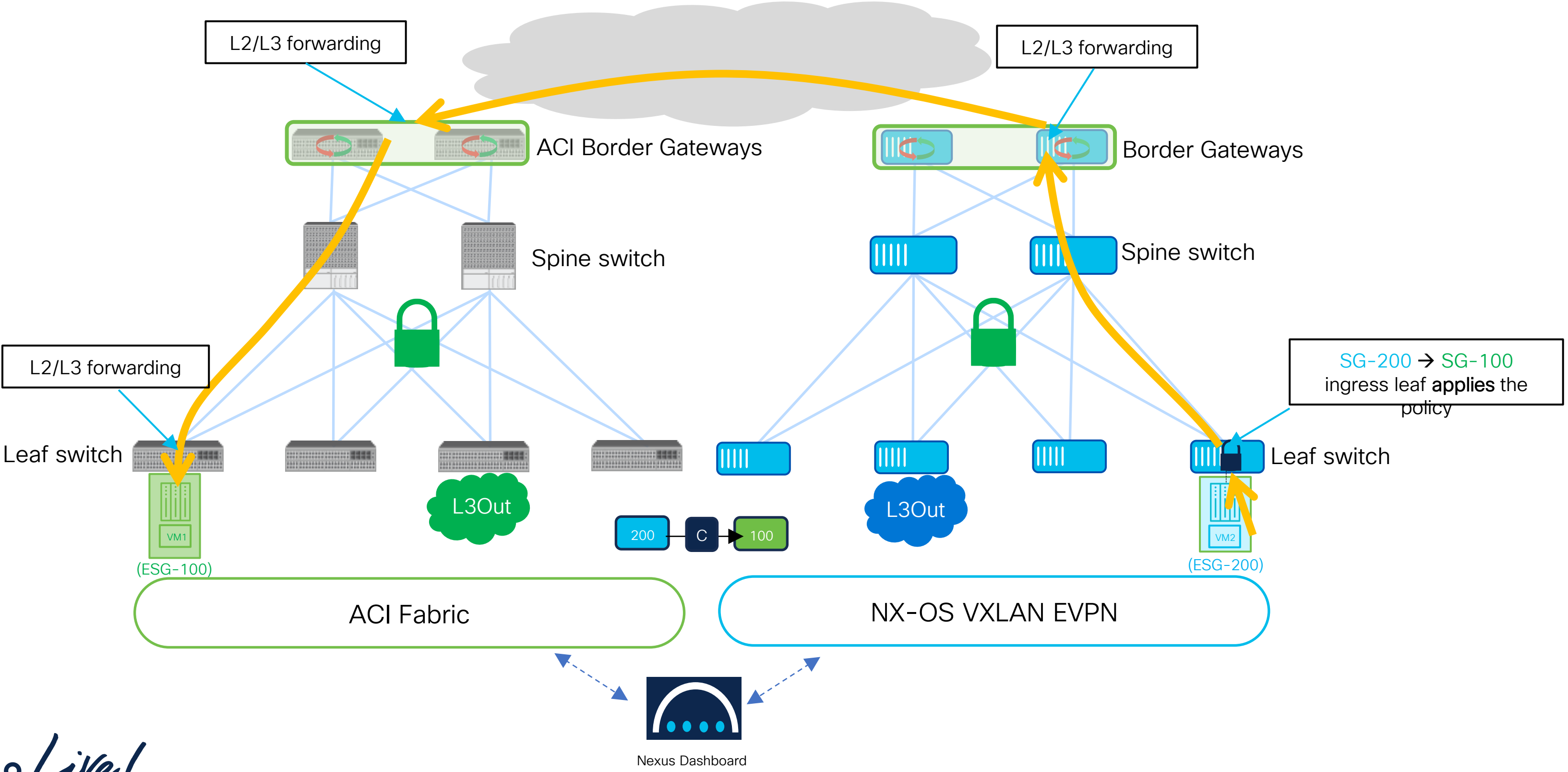
Heterogeneous Fabrics

EVPN Control Plane between ACI and NX-OS BGWs



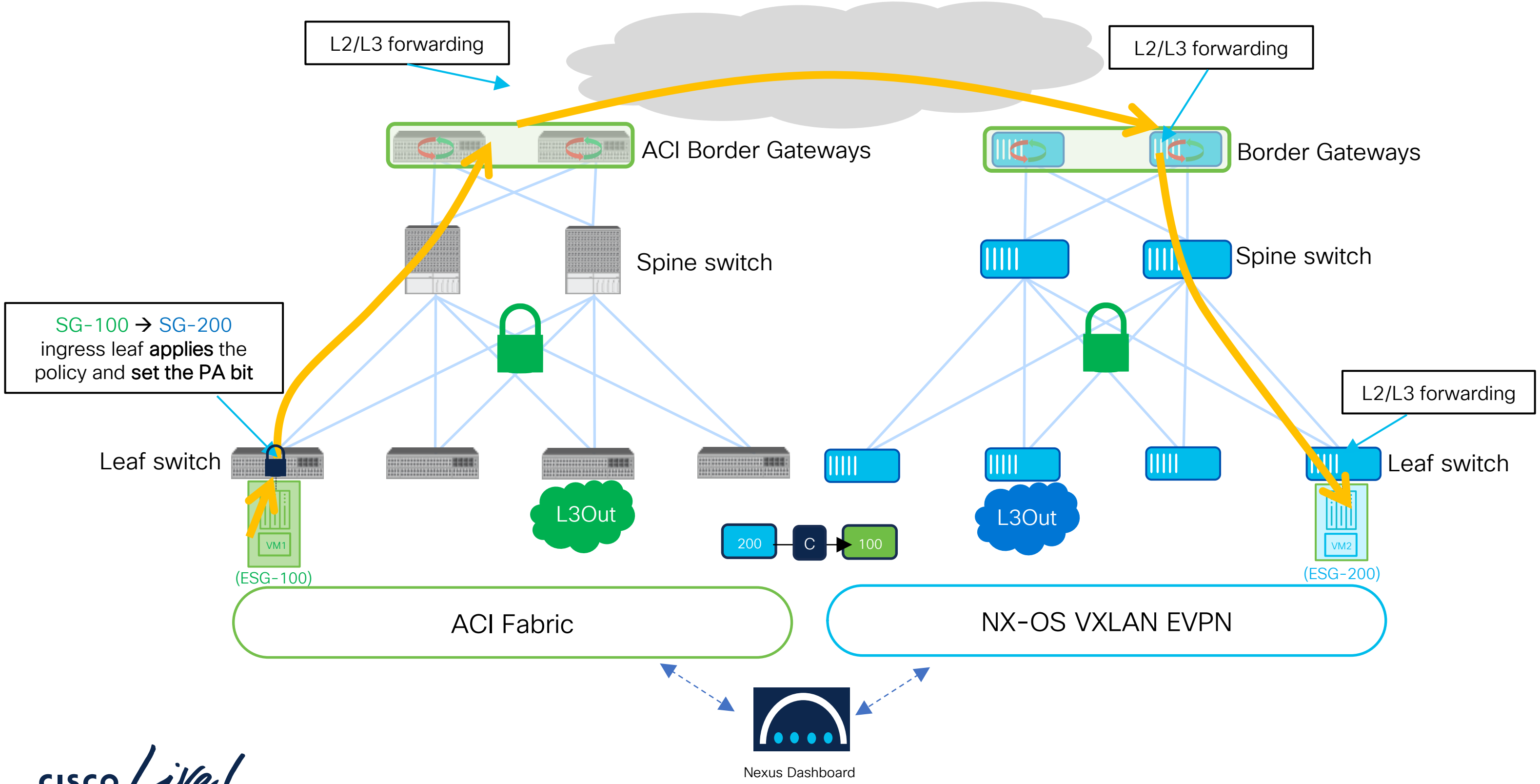
Heterogeneous Fabrics
VXLAN Data Plane between ACI and NX-OS BGWs

Future



Future

Heterogeneous Fabrics
Classification and Policy Enforcement on ACI BGWs



Conclusions

Conclusions

- Building distributed infrastructures is key to the deployment of resilient and scalable designs
- Cisco Nexus ONE aims to seamlessly interconnect and operate a mix of heterogeneous fabrics (ACI and VXLAN EVPN)
- The three main pillars to realize the Nexus ONE vision are:
  1. BGW function for ACI fabrics
  2. Security policies in VXLAN EVPN fabrics (GPO)
  3. Introduction of centralized management and operation platforms for heterogeneous fabric on Nexus Dashboard

Complete Your Session Evaluations

Complete a minimum of 4 session surveys and the Overall Event Survey to claim a **Cisco Live T-Shirt**.

Complete your surveys in the **Cisco Live mobile app**.

Continue your education

- Visit the Cisco Stand for related demos
- Book your one-on-one Meet the Expert 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

Thank you

cisco Live!

GO **BEYOND**

#CiscoLiveAPJC