cisco Live!

Let’s go

#CiscoLive

The bridge to possible

Next Generation DCI

EVPN/VXLAN Multisite A Way To Retire Old Technologies

Stephen McCabe, Technical Solutions Architect

BRKDCN-2933

#CiscoLive

Cisco Webex App

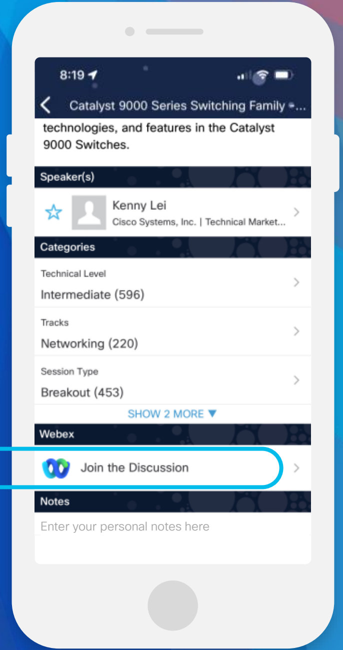
Questions?

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 9, 2023.



8:19

Catalyst 9000 Series Switching Family ...

technologies, and features in the Catalyst 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

https://ciscolive.ciscoevents.com/ciscolivebot/#BRKDCN-2933

Agenda

- Introduction
- What is EVPN Multi-Site?
- Use cases
- Multi-Site with DCI – A Deeper Look
- Migration from Legacy to new EVPN/VXLAN Fabric
- Failure Scenarios
- Automation and Observability with Nexus Dashboard
- Conclusion

Abstract

VXLAN is a widely adopted industry standard for encapsulation, and with MP-BGP, EVPN provides extensive capabilities as a control-plane. With VXLAN and EVPN, we have excellent capabilities for Data Center fabric deployments with an integrated Layer-2 and Layer-3 approach. With the maturity of the control and data planes, new capabilities for interconnecting multiple fabrics are experiencing growing interest with VXLAN BGP EVPN. The goal of the session is to provide a better understanding of how VXLAN EVPN Multi-Site architecture is a modern alternative to DCI technologies such as OTV, VPLS, or EoMPLS, especially for interconnecting data center networks that are solely built on legacy technologies (for example, STP, vPC, or Cisco FabricPath).

Important Note: The session is exclusively focused on NX-OS standalone VXLAN EVPN and does not discuss the multi-pod and multi-site solutions offered with Cisco ACI.

Introduction

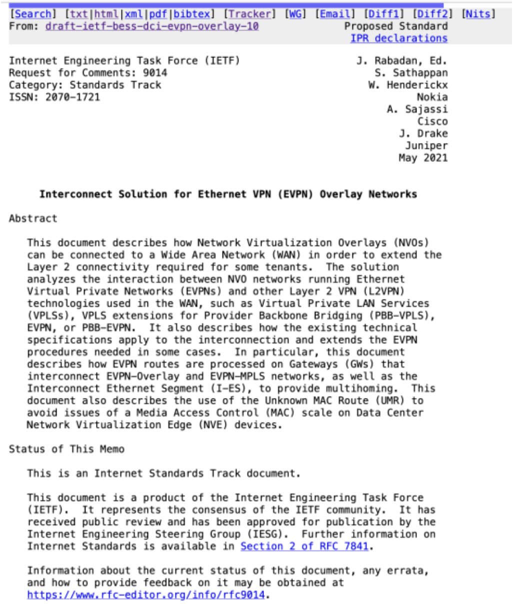
Introduction

- A brief touchpoint of the work at the IETF (Internet Engineering Task Force) and what RFC (Request for Comment) are Standard and what Informational
- What is VXLAN EVPN Multisite?
- Use Cases – Focus on Enabling Migration Off Legacy Technologies
  - Migration/Deployment Scenarios
- The Border Gateway (BGW)
- Automation and Observability

What is Multisite?

RFC 9014

By the Standards Body



[Search] [txt|html|xml|pdf|bibtex] [Tracker] [WG] [Email] [Diff1] [Diff2] [Nits]
From: draft-ietf-bess-dci-evpn-overlay-10 Proposed Standard
IPR declarations

Internet Engineering Task Force (IETF)
Request for Comments: 9014
Category: Standards Track
ISSN: 2070-1721

J. Rabadan, Ed.
S. Sathappan
W. Henderickx
Nokia
A. Sajassi
Cisco
J. Drake
Juniper
May 2021

Interconnect Solution for Ethernet VPN (EVPN) Overlay Networks

Abstract

This document describes how Network Virtualization Overlays (NVOs) can be connected to a Wide Area Network (WAN) in order to extend the Layer 2 connectivity required for some tenants. The solution analyzes the interaction between NVO networks running Ethernet Virtual Private Networks (EVPNs) and other Layer 2 VPN (L2VPN) technologies used in the WAN, such as Virtual Private LAN Services (VPLSs), VPLS extensions for Provider Backbone Bridging (PBB-VPLS), EVPN, or PBB-EVPN. It also describes how the existing technical specifications apply to the interconnection and extends the EVPN procedures needed in some cases. In particular, this document describes how EVPN routes are processed on Gateways (GWs) that interconnect EVPN-Overlay and EVPN-MPLS networks, as well as the Interconnect Ethernet Segment (I-ES), to provide multihoming. This document also describes the use of the Unknown MAC Route (UMR) to avoid issues of a Media Access Control (MAC) scale on Data Center Network Virtualization Edge (NVE) devices.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 7841.

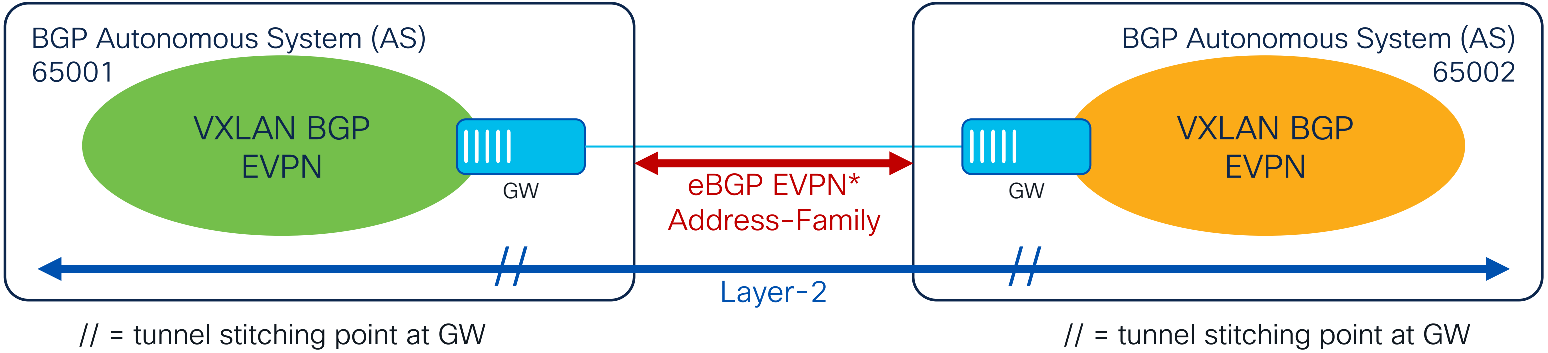
Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc9014.

- Internet Engineering Task Force (IETF) Request for Comment (RFC)
- Categorized for Standards Track
- Internet Standard since 2021
- Existing Industry Adoption
- Interconnect Solution for Ethernet VPN (EVPN) Overlay Networks
- Co-Authored by Cisco

- RFC 9014
  - https://datatracker.ietf.org/doc/html/rfc9014

RFC 9014 - Overview

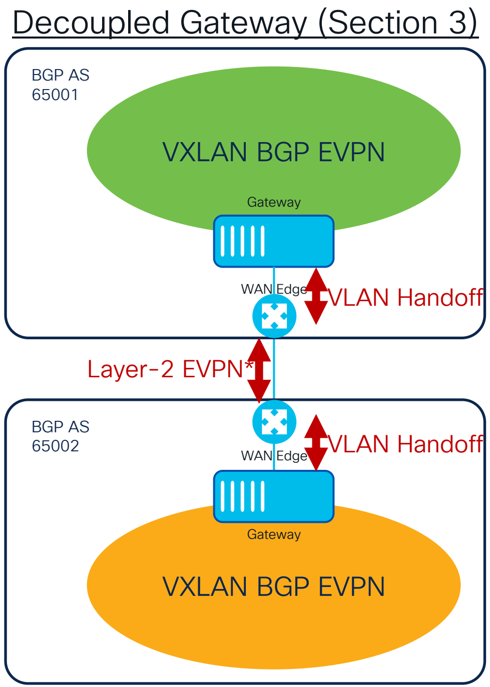
- DCI EVPN Overlay (aka RFC 9014)
- Interconnect Solution for Ethernet VPN (EVPN) Overlay Networks
- From the Abstract “*extend the Layer 2 connectivity required for some tenants.*”

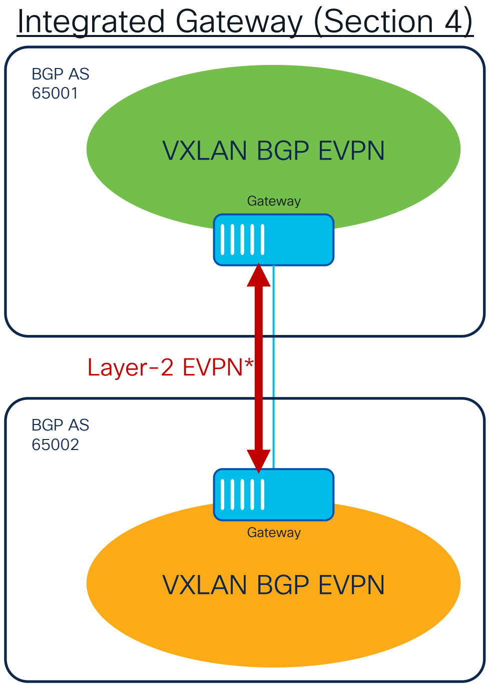

BGP Autonomous System (AS) 65001
VXLAN BGP EVPN
GW
eBGP EVPN* Address-Family
BGP Autonomous System (AS) 65002
VXLAN BGP EVPN
GW
//
//
Layer-2
// = tunnel stitching point at GW
// = tunnel stitching point at GW

*RFC 9014 supports more than just EVPN for the Interconnect Network

RFC 9014 Gateway Model Side-by-Side

Decoupled and Integrated Gateway



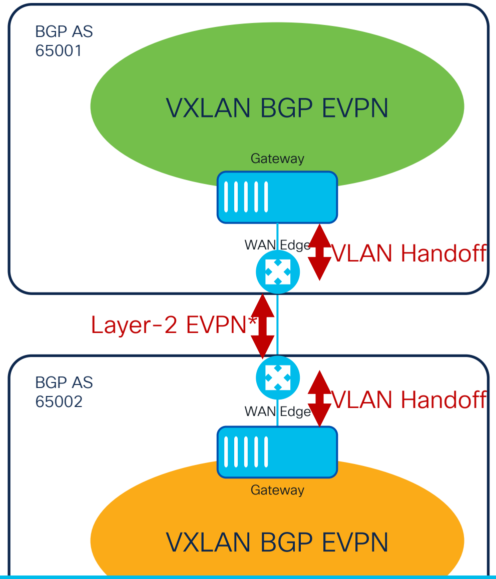


*RFC 9014 supports more than just EVPN for the Interconnect Network

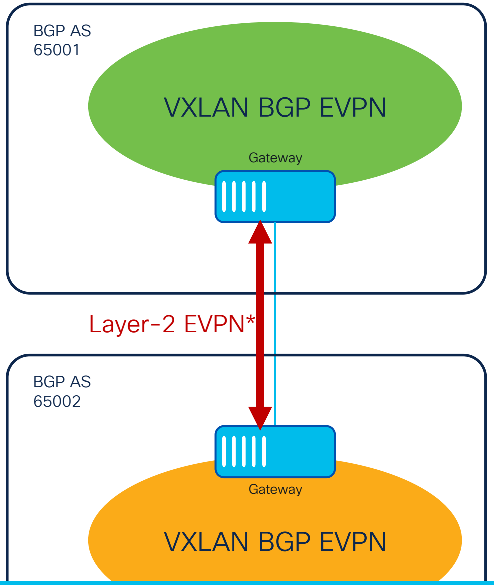
RFC 9014 Gateway Model Side-by-Side

Decoupled and Integrated Gateway

Decoupled Gateway (Section 3)



Integrated Gateway (Section 4)



What about Layer-3?

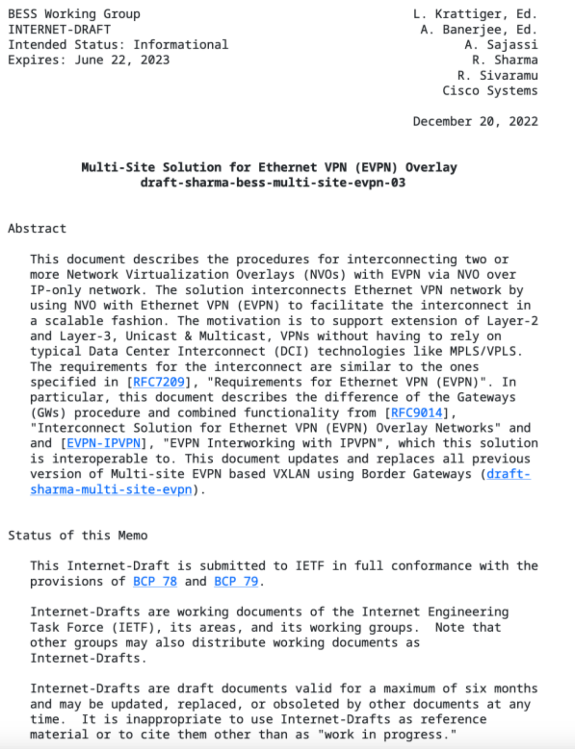
*RFC 9014 supports more than just EVPN for the Interconnect Network

Multi-Site Solution for Ethernet VPN (EVPN) Overlay

draft-sharma-bess-multi-site-evpn

What is Multi-Site?

By the Standards Body



```
BESS Working Group                                      L. Krattiger, Ed.
INTERNET-DRAFT                                           A. Banerjee, Ed.
Intended Status: Informational                                 A. Sajassi
Expires: June 22, 2023                                          R. Sharma
                                                              R. Sivaramu
                                                            Cisco Systems

                                                        December 20, 2022


           Multi-Site Solution for Ethernet VPN (EVPN) Overlay
                   draft-sharma-bess-multi-site-evpn-03


Abstract

   This document describes the procedures for interconnecting two or
   more Network Virtualization Overlays (NVOs) with EVPN via NVO over
   IP-only network. The solution interconnects Ethernet VPN network by
   using NVO with Ethernet VPN (EVPN) to facilitate the interconnect in
   a scalable fashion. The motivation is to support extension of Layer-2
   and Layer-3, Unicast & Multicast, VPNs without having to rely on
   typical Data Center Interconnect (DCI) technologies like MPLS/VPLS.
   The requirements for the interconnect are similar to the ones
   specified in [RFC7209], "Requirements for Ethernet VPN (EVPN)". In
   particular, this document describes the difference of the Gateways
   (GWs) procedure and combined functionality from [RFC9014],
   "Interconnect Solution for Ethernet VPN (EVPN) Overlay Networks" and
   and [EVPN-IPVPN], "EVPN Interworking with IPVPN", which this solution
   is interoperable to. This document updates and replaces all previous
   version of Multi-site EVPN based VXLAN using Border Gateways (draft-
   sharma-multi-site-evpn).


Status of this Memo

   This Internet-Draft is submitted to IETF in full conformance with the
   provisions of BCP 78 and BCP 79.

   Internet-Drafts are working documents of the Internet Engineering
   Task Force (IETF), its areas, and its working groups.  Note that
   other groups may also distribute working documents as
   Internet-Drafts.

   Internet-Drafts are draft documents valid for a maximum of six months
   and may be updated, replaced, or obsoleted by other documents at any
   time.  It is inappropriate to use Internet-Drafts as reference
   material or to cite them other than as "work in progress."
```

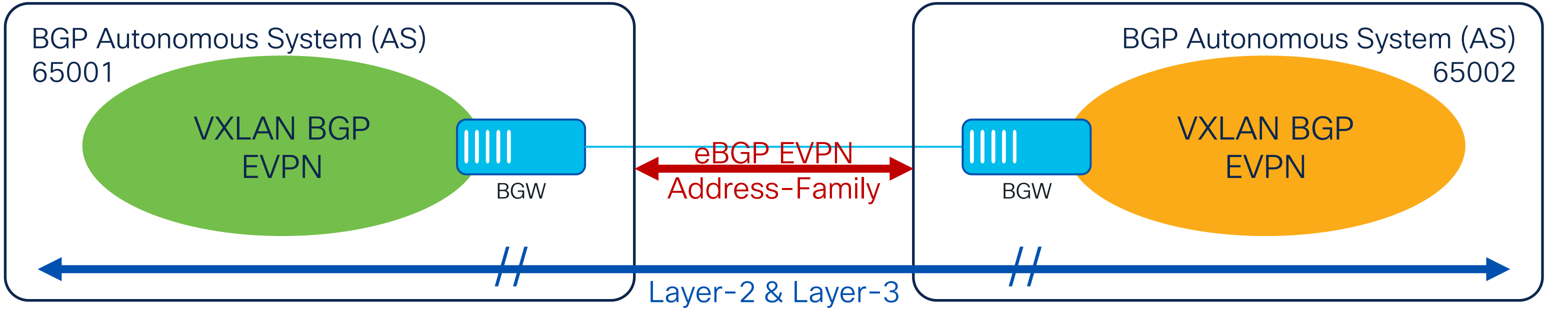
- Internet Engineering Task Force (IETF) Request for Comment (RFC)
- Categorized as Informational
- Internet Draft since 2016
  - Currently in Version 3
  - Overall, 7 versions
- Updated and Maintained by BESS version of draft
  - draft-sharma-bess-multi-site-evpn
- Shipping since 2017

- Multi-Site (BESS version)
  - https://datatracker.ietf.org/doc/html/draft-sharma-bess-multi-site-evpn
- Pre-Cursor Draft (replaced by BESS version)
  - https://datatracker.ietf.org/doc/html/draft-sharma-multi-site-evpn

Multi-Site

By the Standards Body

- Multi-Site Solution for Ethernet VPN (EVPN) Overlay (draft-sharma-bess-multi-site-evpn)
- Interconnect Solution for Ethernet VPN (EVPN) Overlay Networks
- From the Abstract “*support extension of Layer-2 and Layer-3, Unicast & Multicast, VPNs*”



// = tunnel stitching point at BGW

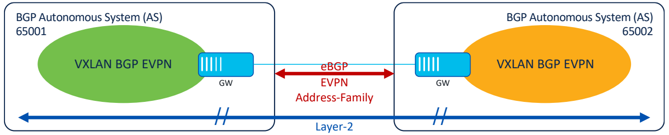
// = tunnel stitching point at BGW

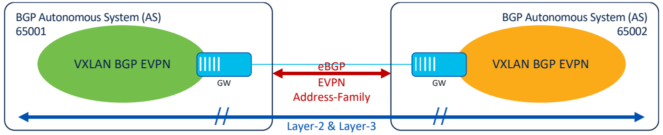
RFC9014 and Multi-Site - Side by Side

	DCI-EVPN-Overlay (RFC 9014)	Multi-Site EVPN (draft-sharma-bess-multi-site-evpn)	
Interconnect	Integrated (1-Box), Decoupled (2-Box)	Integrated (1-Box)	
DCI Encap	VPLS, PBB-VPLS, EVPN-MPLS, PBB-EVPN, VXLAN	VXLAN	
Gateway Mode	Multipath PIP	Anycast VIP	Multipath PIP
ECMP	Underlay and Overlay	Underlay	Underlay and Overlay
EVPN RT-1	Consumed and Generated	None	Consumed and Generated
EVPN RT-2	Re-Originated with I-ESI	Re-Originated with ESI 0	Re-Originated with I-ESI
EVPN RT-3	Consumed and Generated	Consumed and Generated	Consumed and Generated
EVPN RT-4	Consumed and Generated	Consumed and Generated	Consumed and Generated
EVPN RT-5	- (not part of RFC)	Re-Originated	Re-Originated
Route Distinguisher (RD)	Separate RD for Intra and Inter DC	Separate RD for VIP and PIP	
Route-Target (RT)	Separate RT for Intra and Inter DC	Same RT for Intra and Inter DC	
VNI Allocation	Global and Downstream	Global and Downstream	
DF Election	Based on EVPN RT-4	Based on EVPN RT-4	
Identifier	I-ESI	I-ESI (= Site-ID)	
Split Horizon	Local Bias	Local Bias	
ESI-Type	Type 0 (Operator Managed)	Type 3 (MAC Based) or Type 5 (AS based)	
BUM Tree #	2, GW stitched (Intra and Inter DC)	2, GW stitched (Intra and Inter DC)	

RFC9014 and Multi-Site – Side by Side

In a Nutshell





RFC 9014

Base Standard for Interconnecting EVPN
Defines the Layer-2 Stitching
Two Gateway Model
Multiple Encapsulations
Leverages Overlay and Underlay ECMP

Multi-Site

Extends RFC 9014 for Interconnecting EVPN
Describes Layer-2 and Layer-3 Stitching
Single Gateway Model (Two BGW* Model)
Focus only on VXLAN Encapsulation
Different ECMP model depending on BGW Model

*BGW – Border Gateway (BGW); Cisco's name for the VXLAN EVPN to VXLAN EVPN Gateway

EVPN Multisite

Use Cases

Use Cases - Overview

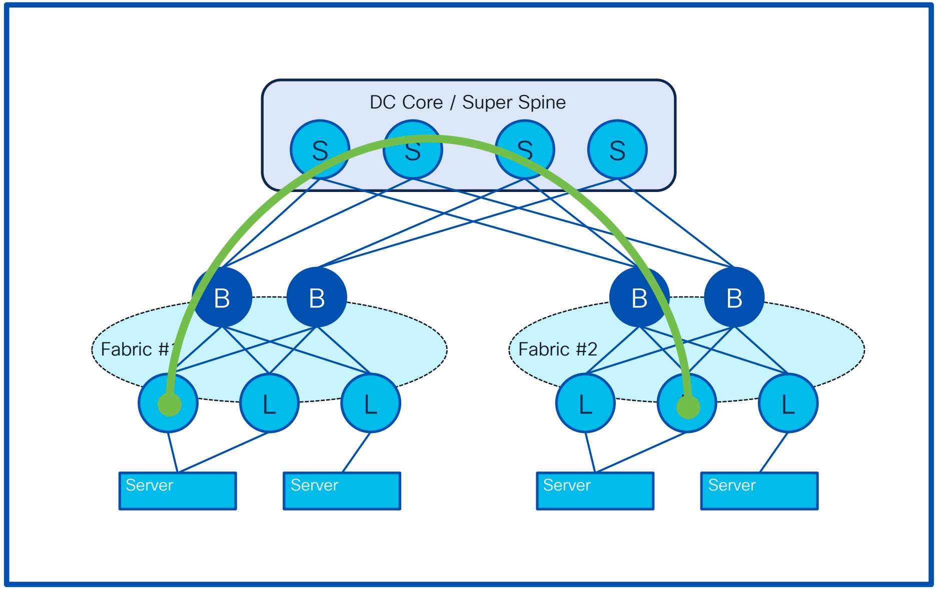
VXLAN EVPN Multi-Site architecture is a design for VXLAN BGP EVPN-based overlay networks. It allows interconnection of multiple distinct VXLAN BGP EVPN fabrics or overlay domains, and it allows new approaches to fabric scaling, compartmentalization, and DCI.

Use cases for EVPN Multisite:

- Compartmentalization
- Hierarchical scale-out approaches
- DCI
- Integration of legacy networks

Areas of Focus

Use Case #1: Compartmentalization

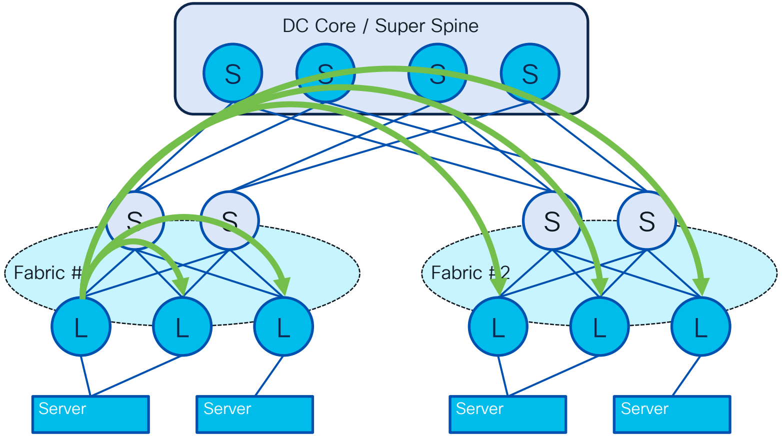


- Multiple Fabrics, single Data Center
  - Single or Multiple Data Halls
  - Within a Geographic Locations
- Control at BGW (Border Gateway)
  - Allows Extension of Layer-2
  - Allows Extension of Layer-3
  - Allows Extension of Layer-2 and Layer-3
  - Allows Traffic Control (BUM*)
  - Defines VNI allocation and stitching
  - Optimizes BUM* Replication

BUM Optimization

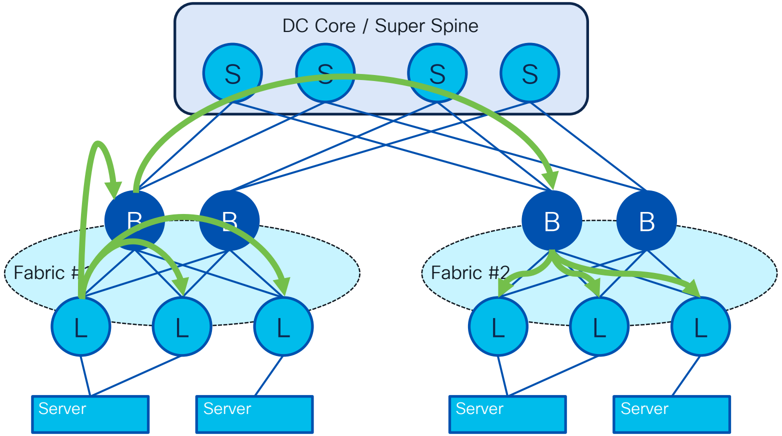
Use Case #1 – Compartmentalization

Single Fabric BUM with Ingress Replication



Single BUM Packet, 5x Replicated
3 Replication over DC Core / Super Spine (Between)
2 Replication for Fabric #1 (Local)

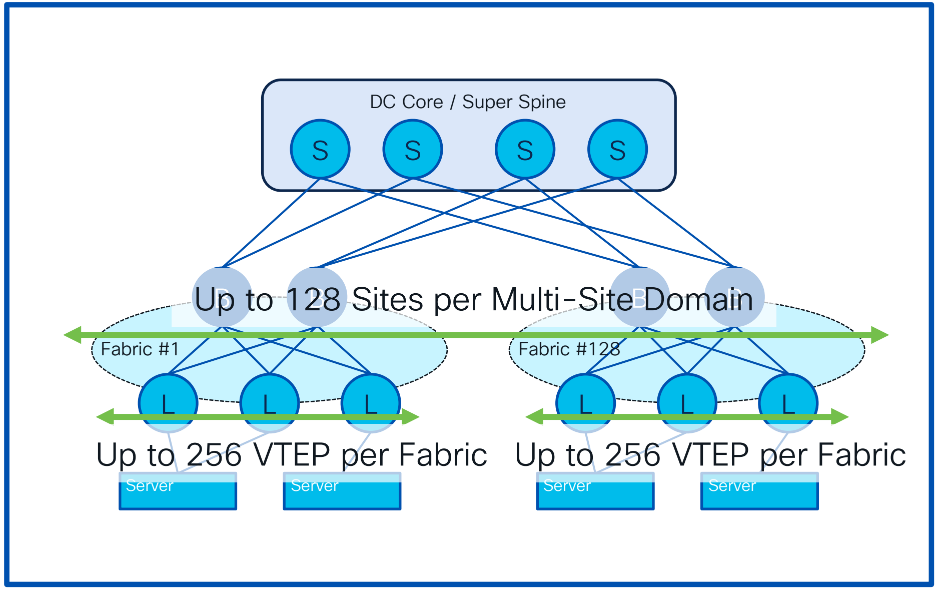
Multi-Site BUM with Ingress Replication



Single BUM Packet, 3x Replicated
1 Replication over DC Core / Super Spine (Between)
3 Replication for Fabric #1 (Local)
3 Replication for Fabric #2 (Local)

*BUM – Broadcast, Unknown Unicast, Multicast

Use Case #2 - Scale



DC Core / Super Spine
S
S
S
S
Up to 128 Sites per Multi-Site Domain
Fabric #1
Fabric #128
L
L
L
L
L
L
Up to 256 VTEP per Fabric
Up to 256 VTEP per Fabric
Server
Server
Server
Server

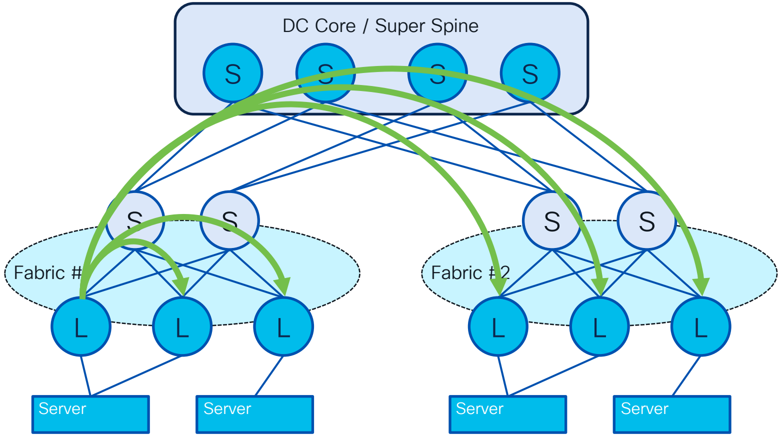
- Multiple Fabrics , single or multiple Data Center
  - Single or Multiple Data Halls
  - Within or between Geographic Locations
- Control at BGW (Border Gateway)
  - Reduces Remote VTEP Count
  - Expands VTEP scale
- Scale through Hierarchy
  - Multiply VTEP with Sites

*TRM upto 15 sites
*Number of BGWs per site 6 (Anycast), 2 (vPC)

32’768 VTEP to extend Layer-2 or/and Layer-3 segments to

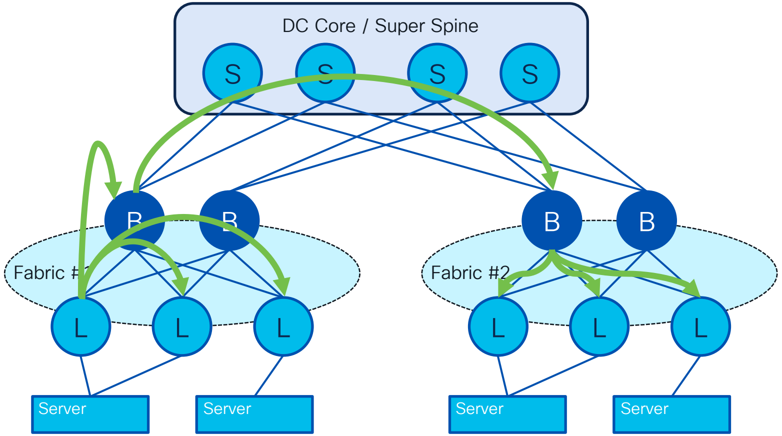
VTEP Scale
Use Case #2 - Scale

Single Fabric or Multi-POD



Leaf #1 sees every VTEP, 5 VTEP Peer
3 VTEP Peer for Fabric #2 (Between)
2 VTEP Peer for Fabric #1 (Local)

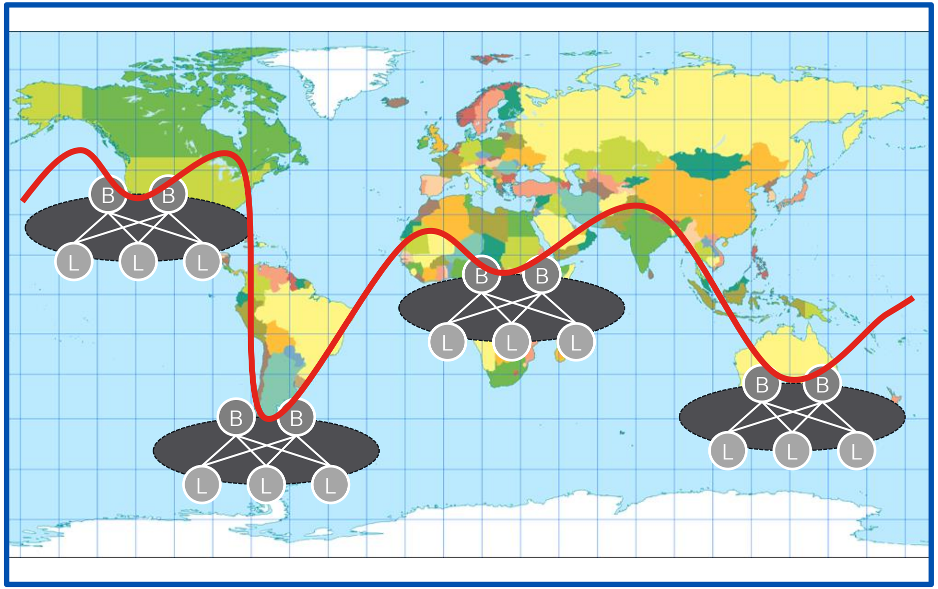
Multiple Fabric with Multi-Site



Leaf #1 sees only local VTEP, 3 VTEP Peer
1 VTEP Peer for Exit, BGW (Between)
2 VTEP Peer for Fabric #1 (Local)

*BUM – Broadcast, Unknown Unicast, Multicast

Use Case #3 – Data Center Interconnect (DCI)

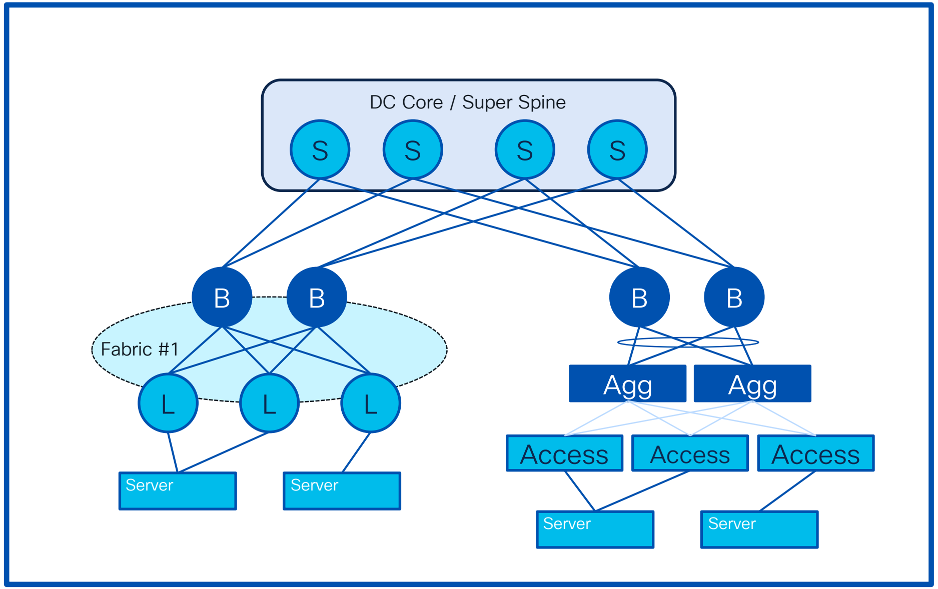


- Multiple Fabrics, Geographically Dispersed
- Classic DCI Use Case
  - Allows Extension of Layer-2
  - Allows Extension of Layer-3
  - Allows Extension of Layer-2 and Layer-3
  - Allows Traffic Control (BUM*)
  - Defines VNI allocation and stitching
  - Optimizes BUM* Replication

Works Within a Geographic Location – Works Between Geographic Locations

*BUM – Broadcast, Unknown Unicast, Multicast

Use Case #4 – Integration with Legacy Networks



- Integrating Fabrics with Legacy Networks
  - BGW Frontends Legacy Network
  - BGW Frontends New Network
- Host Mobility and Migration
  - Provides Distributed Default Gateway
  - Allows Layer-2 Extension where needed
- Benefits from all Multi-Site functions
  - Layer-2, Layer-3 Multicast and Unicast VPNs between different Networks for Migration or Co-Existance

Much more on this shortly!

Multisite and the Role of the Border Gateway

A Deeper Look

As we Talk about Scale

Hardware Support

Minimum Hardware and Software Requirements for BGW (Border Gateway)	
Cisco Nexus Hardware	Cisco Nexus 9300 EX platform Cisco Nexus 9300 FX platform Cisco Nexus 9300 FX2 platform Cisco Nexus 9300 FX3 platform Cisco Nexus 9300 GX platform Cisco Nexus 9300 GX2 platform Cisco Nexus 9364C platform Cisco Nexus 9332C platform Cisco Nexus 9500 platform with X9700-EX line card Cisco Nexus 9500 platform with X9700-FX line card Cisco Nexus 9500 platform with X9700-GX line card
Cisco Nexus Software (NX-OS)	Cisco NX-OS Software Release 7.0(3)I7(1) or later*

*Check for Hardware Specific Support Releases

As we Talk about Scale

Scalability Values as of NX-OS 10.2(5)M

Multi-Site Scale	
Number of Sites	128
Number of BGW per Site	6
Number of VTEP per Site (internal)	256

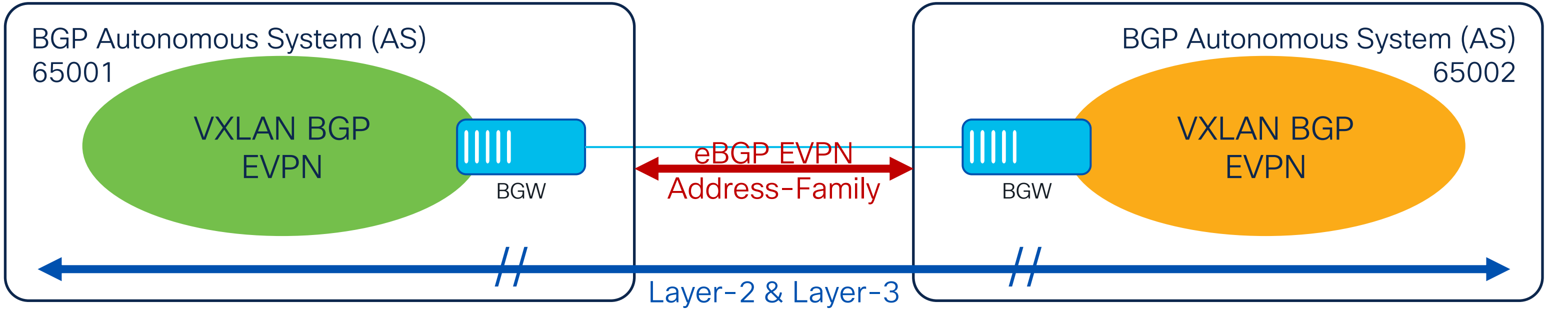
Border Gateway (BGW) Scale	EX	FX2	FX,FX3,GX,GX2	N9364C & N9332C
Number of Layer-2 VNI (VLAN)	3900			
Number of Layer-3 VNI (VRF)	2000			
MAC per BGW	92k			
IPv4 Host Routes per BGW*	450k	450k	1.1M	96k
IPv4 Network Routes per BGW*	450k	450k	1.1M	8k
IPv6 Host Routes per BGW*	24k	260k	620k	48k
IPv6 Network Routes per BGW*	200k	290k	620k	2k

*The values provided in these tables focus on the scalability of one particular Route scale at a time

Some Notes on BGW and VXLAN Tunnels

Multi-Site

- Tunnels are Stitched at the BGW (Border Gateway)
- Intra Fabric Tunnel goes from Leaf to Leaf or Leaf to BGW
- Inter Fabric Tunnel goes from BGW to BGW



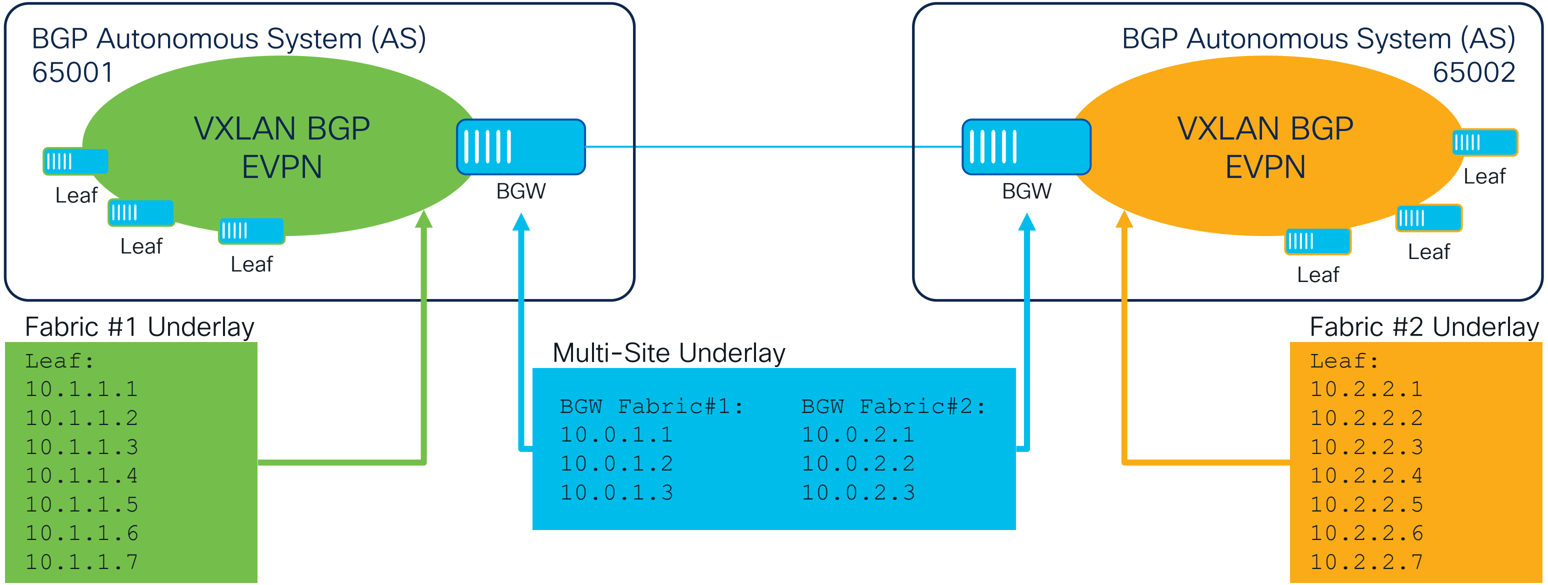
// = tunnel stitching point at BGW

// = tunnel stitching point at BGW

Some Notes on the Interconnect and Underlay

Multi-Site

- Fabric #1 Underlay (VTEP, Point-2-Point, Loopback etc) is not aware of Fabric #2
- Each Fabric maintains their Unique Network Topology, Protocols and IP Addressing
- Only BGW IP Addressing must be Unique and Aligned between Sites



Fabric #1 Underlay

```
Leaf:
10.1.1.1
10.1.1.2
10.1.1.3
10.1.1.4
10.1.1.5
10.1.1.6
10.1.1.7
```

Multi-Site Underlay

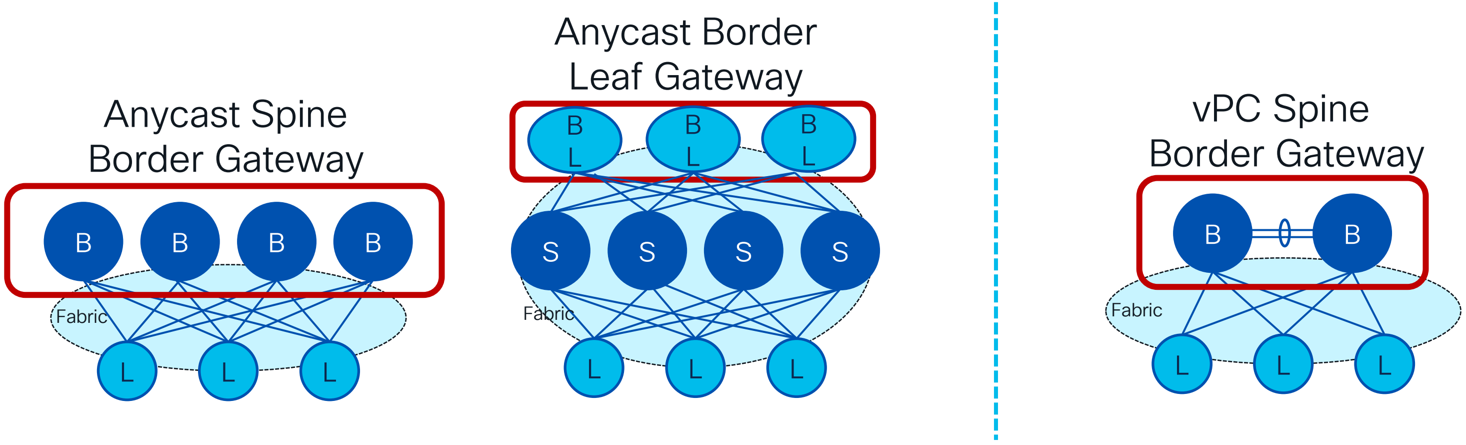
```
BGW Fabric#1:    BGW Fabric#2:
10.0.1.1         10.0.2.1
10.0.1.2         10.0.2.2
10.0.1.3         10.0.2.3
```

Fabric #2 Underlay

```
Leaf:
10.2.2.1
10.2.2.2
10.2.2.3
10.2.2.4
10.2.2.5
10.2.2.6
10.2.2.7
```

Border Gateway Details

Border Gateways Deployment Considerations



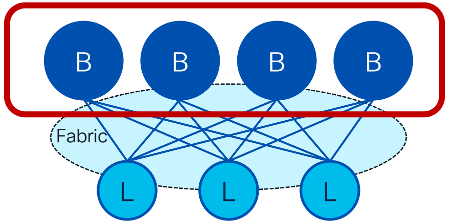
Anycast Spine Border Gateway

Anycast Border Leaf Gateway

vPC Spine Border Gateway

- Border Gateways used for two main functions:
    - Interconnecting each site to the Inter-Site network (for East-West traffic flows)
    - Connecting each site to the external Layer 3 domain (for North-South traffic flows)
    - NOTE: May also be used to connect endpoints and/or network service nodes (FWs, ADCs)
- Possible deployment models:
    - Anycast Border Gateways
    - vPC Border Gateways
- BGW function enablement in the VXLAN EVPN fabric:
    - BGWs on Leaf node (Border Gateway Leaf)
    - BGWs on Spine node (Border Gateway Spine)

Anycast Border Gateway

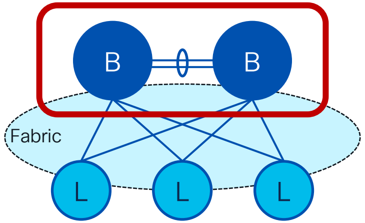


Anycast Border Gateway

- Up to 6 Border Gateways
- Border Gateway
  - Deploying as a Leaf node since release 7.0(3)I7(1)
  - Deploying as a Spine node since release 7.0(3)I7(2)

- Two Modes of Operation:
  - Can Operate as Multi-Site Anycast BGW with VIP
    - Focuses on Scale and Convergence
    - Using Virtual IP (VIP) for Tunnel Stitching
    - Uses Overlay ECMP
  - Can Operate in RFC 9014 BGW Mode with PIP
    - Focuses on 3rd Party Interop
    - Using Primary IP (PIP) for Tunnel Stitching
    - Uses Underlay and Overlay ECMP

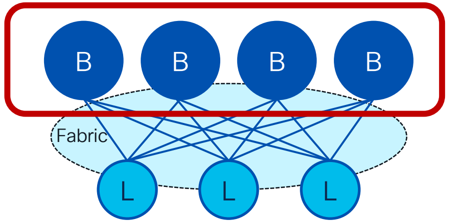
vPC Border Gateway

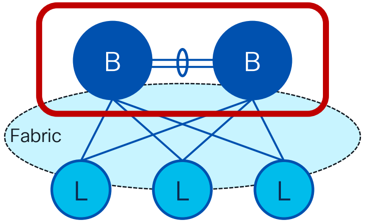


vPC Border Gateway

- Up to 2 Border Gateways
- Border Gateway
  - Deploying as a Leaf node since 9.2(1)
- Common Use Case
  - Legacy Network Integration or Migration
    - Provides Multi-Chassis Link Aggregation
    - Integrates with Ethernet and FabricPath
    - Hosts the Distributed Anycast Gateway
  - Attachment of Network Services
    - Dual-Attachment of Firewalls and ADCs
    - Acts like a vPC when it comes to Routing

When to use what BGW





Anycast Border Gateway

- Up to 6 BGW
  - Share Nothing
  - Simple Failure Scenarios
- Any Deployments
  - No End-Point or Network Services Connectivity on BGW
- Greenfield Deployments

vPC Border Gateway

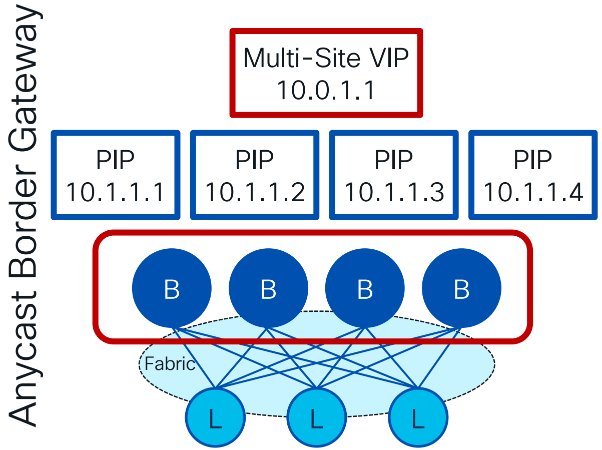
- 2 BGW with physical vPC Peer-Link
- Small Deployments
  - End-Point or Network Services Connectivity on BGW
- Migration Use-Cases (Brownfield)
  - Classic Ethernet/FabricPath to VXLAN EVPN

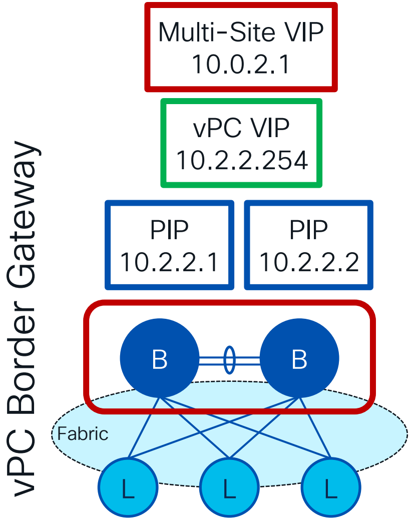
vPC Border Gateways

The Details

Details on the Different BGW

- Both Anycast and vPC Border Gateway needs to be configured with a common Multi-Site VIP address and an individual Primary IP (PIP) address
- vPC Border Gateways share a secondary IP address to be used as vPC virtual IP (vPC VIP)





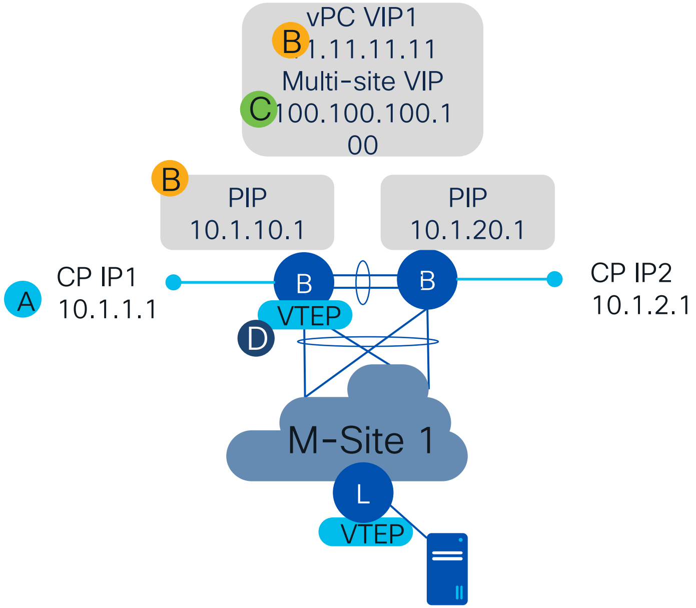
VXLAN EVPN Multi-Site with vPC BGW considerations

What’s What?

vPC BGWs’ logical interfaces:

- Unique logical interfaces (for example, loopback interfaces) must be defined on the vPC BGW devices to perform their duties

```
interface loopback0
 description CP IP or RID                              A
 ip address 10.1.1.1/32 tag 54321
!
interface loopback1
 description PIP1
 ip address 10.1.10.1/32 tag 54321                     B
 ip address 11.11.11.11/32 secondary tag 54321
!
interface loopback100
 description Multi-Site VIP1                           C
 ip address 100.100.100.100/32 tag 54321
!
interface nve1
 host-reachability protocol bgp                        D
 source-interface loopback1
 multisite border-gateway interface loopback100
```



VXLAN EVPN Multi-Site with vPC BGW considerations

What are the used for?

Control Plane IP address (CP IP):

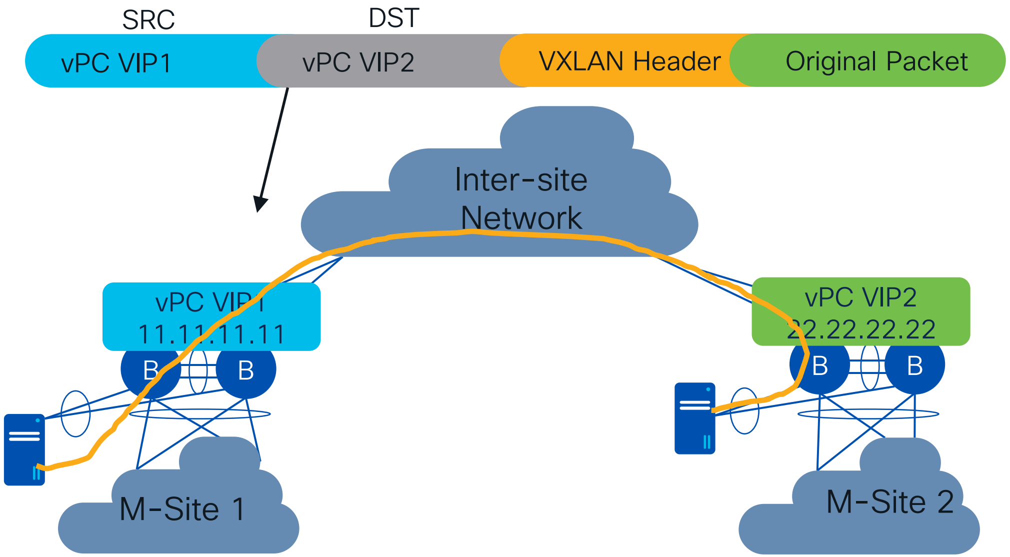
- Used for control plane adjacencies for the MP-BGP EVPN overlay with the remote BGW devices.

Primary IP address (PIP):

- Unique IPs per BGW used to source traffic originated from devices connected via Layer 3 and used to receive traffic from remote sites. North-South Traffic

vPC Virtual IP address (vPC VIP):

- Secondary IP defined on both BGWs part of the same vPC Domain used for two purposes:
  1. Sourcing BUM traffic for Layer 2 networks stretched to remote site(s)
  2. Sourcing/receiving traffic for single- or dual-attached endpoints locally connected at Layer 2 to the BGWs

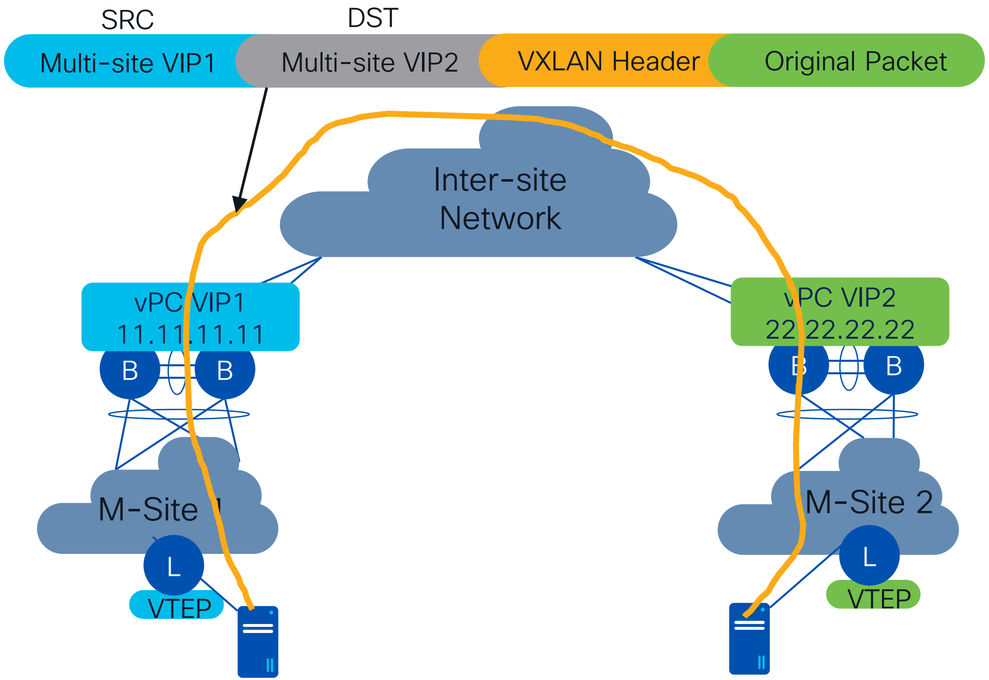


VXLAN EVPN Multi-Site with vPC BGW considerations

Things to Think About

Multi-Site Virtual IP address (Multi-Site VIP):

- IP address on a dedicated loopback defined on both BGW nodes that are part of the same vPC domain.
- IP address is used to source traffic destined to remote sites and originated from endpoints connected behind a leaf node in the local site. The same IP address is also used to receive traffic originating from remote sites and destined to endpoints connected behind a leaf node in the local site



DCI & vPC Border Gateways Connectivity and Migration

A deeper look

Architectural Benefits of Introducing vPC Border Gateways

1. Common Control plane & Data plane
2. Integrated Layer 2 and Layer 3 extension
3. Fault Containment
4. Transport Agnostic
5. Multihoming
6. Multipath Load Sharing
7. Loop Prevention and STP Isolation
8. Support for Multiple Sites

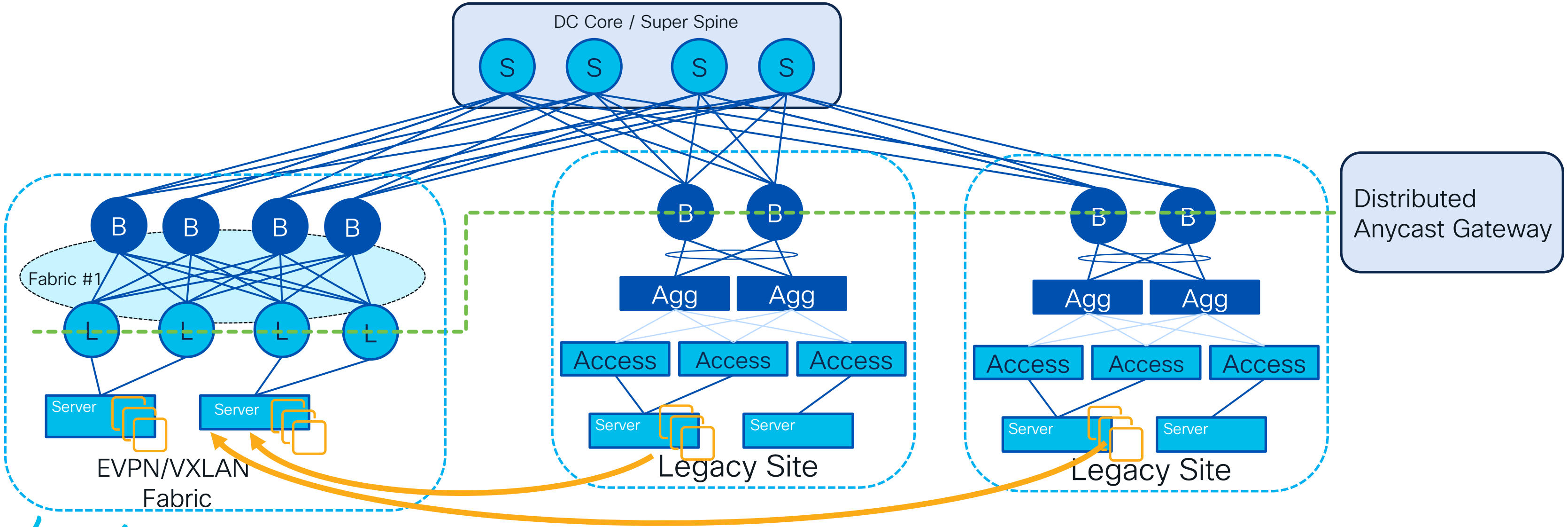
vPC Border Gateway Use Cases

Integration with Legacy Networks

Distributed Anycast Gateway

Primary Use cases

- vPC BGW attached to the existing legacy network providing interconnect with a remote network
- Enabling migration of Legacy fabric workloads to a modern fabric built with VXLAN EVPN (DCI Multisite)
    - The vPC BGWs use a Distributed Anycast Gateway (DAG) to provide a consistent first-hop gateway. This coupled with new EVPN/VXLAN fabric we can extended the anycast GWs to be available across each fabric

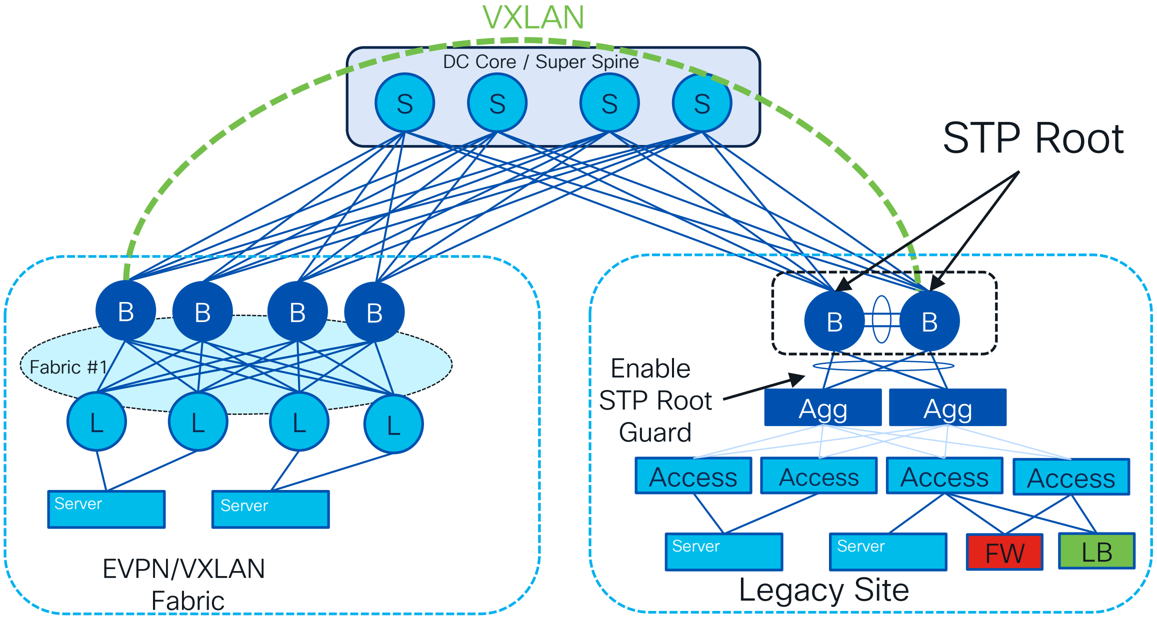


VXLAN EVPN Multi-Site with vPC BGWs

vPC BGW Use Case: #1 Legacy Site to VXLAN/EVPN Fabric

Capabilities/Benefits Achieved

- Integration/coexistence of a legacy site with a VXLAN BGP EVPN site with EVPN Multi-Site
- Provides ability to migration workloads to EVPN/VXLAN Fabrics
- STP Configurations
  - vPC BGW should be configured as STP Root
  - Best Practice is to configure STP Root-Guard on VPC Connections between BGWs and Legacy Network



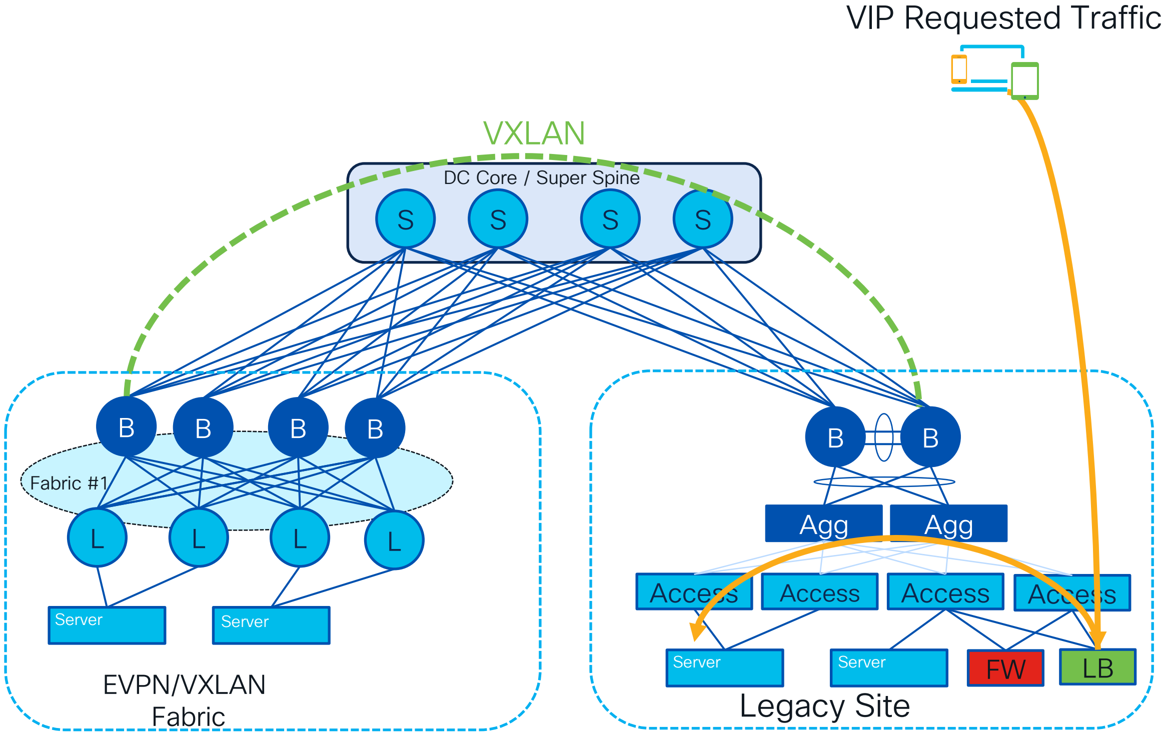
*Use case targets small-fabric deployments

VXLAN EVPN Multi-Site with vPC BGWs

vPC BGW Use Case: #1 Services Considerations

Capabilities/Benefits Achieved

- Integration/coexistence of a legacy site with a VXLAN BGP EVPN site with EVPN Multi-Site
- Provides ability to migration workloads to EVPN/VXLAN Fabrics
- Considerations for Services

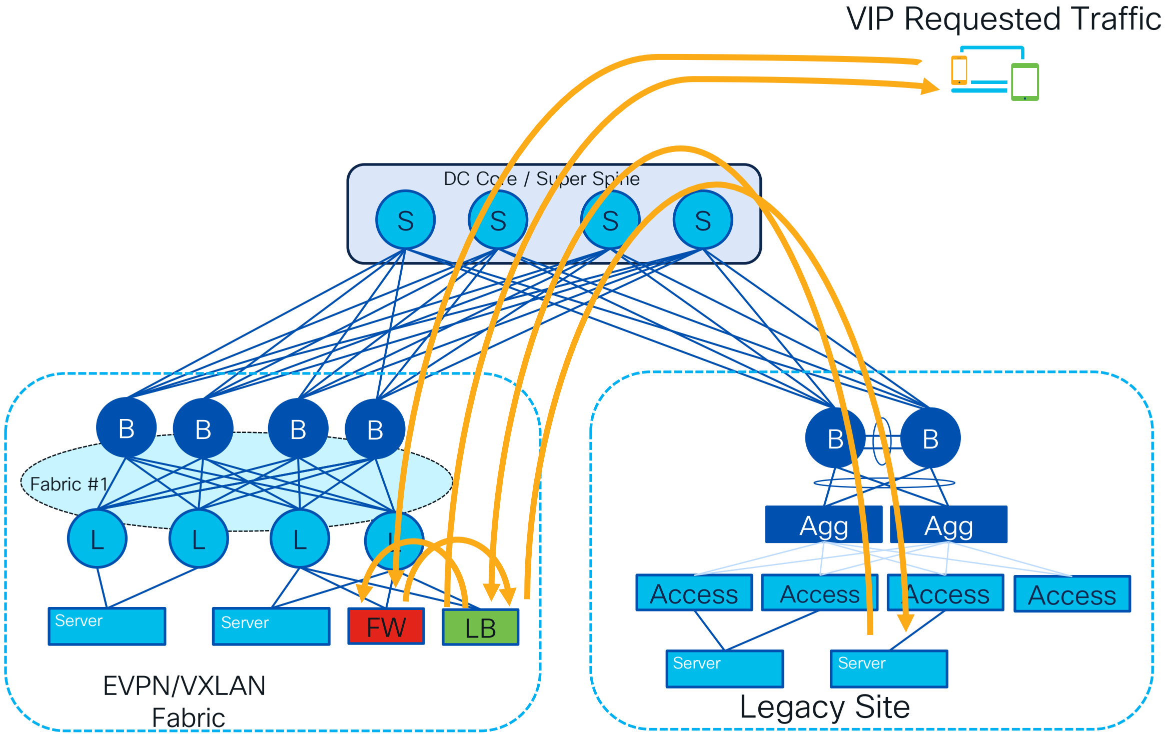


*Use case targets small-fabric deployments

VXLAN EVPN Multi-Site with vPC BGWs

vPC BGW Use Case: #1 Service Migration

- Load Balancer VIP/server migration
- DNS
- Stateful firewalls
- PBR (Policy Based Routing)
- Elastic Service Redirection
    - ePBR and ITD
- Is FHRP for hosts the Load Balancer or FW? Options..
    - Stretch Cluster
    - Migrate FHRP



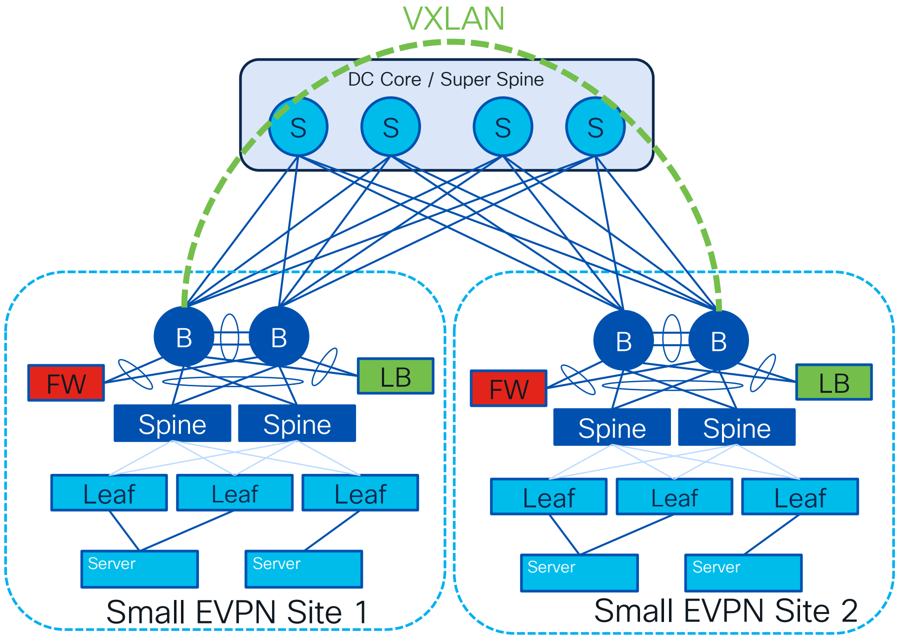
*Use case targets small-fabric deployments

VXLAN EVPN Multi-Site with vPC BGWs

Use Case #2 Small Site connectivity

Use Cases:

- Multisite connectivity for smaller EVPN/VXLAN sites
- Cost effective vs. deploying dedicated Anycast BGWs



*Use case targets small-fabric deployments

Migrating Away From Legacy

Using vPC Border Gateways

Migrating legacy to VXLAN EVPN fabrics using vPC BGWs

Steps involved

Step 1: Insert a pair of vPC BGWs in each legacy site, using Layer 2 double-sided vPC

Step 2: Configure vPC BGWs DCI underlay network

Step 3: Configure vPC BGWs DCI overlay network

Step 4: Configure vPC BGWs for DCI Layer 2 extension across sites

Step 5: Enable Anycast Gateway on vPC BGWs and keep it in shutdown state

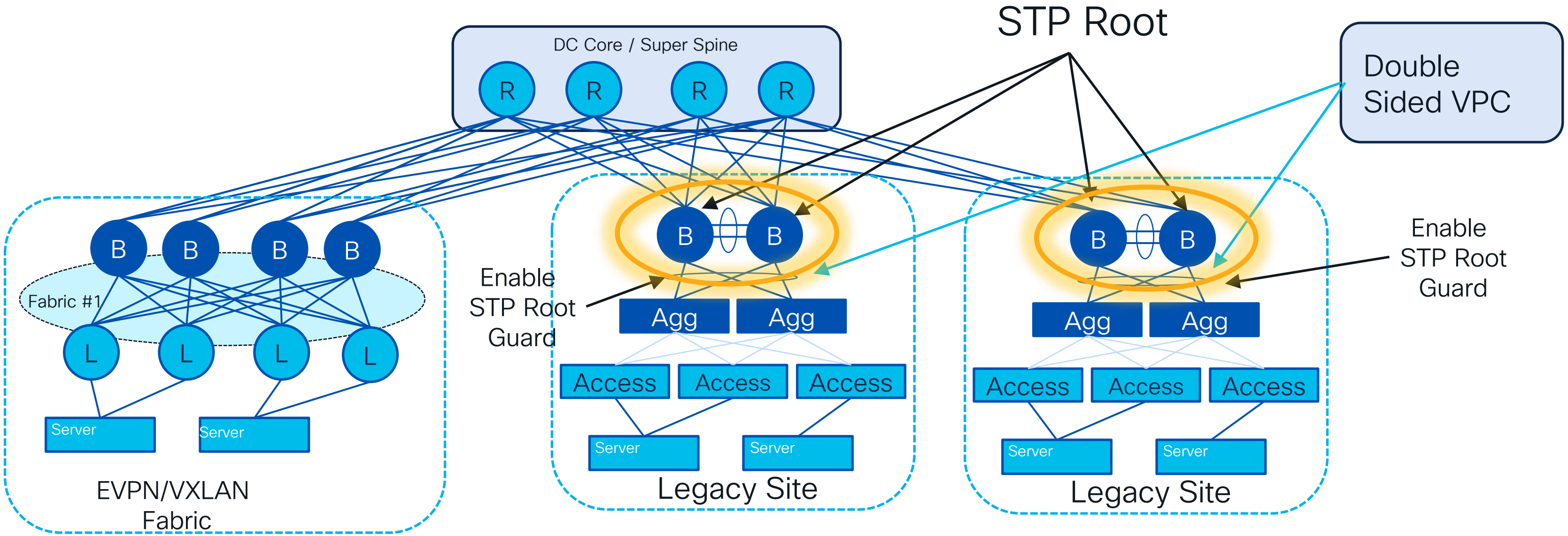
Step 6: Migrate first-hop FHRP Gateway in the legacy site to the vPC BGW Anycast Gateway

Step 7: Transition legacy data centers to new Data Center Fabric

Migrating legacy to VXLAN EVPN fabrics using vPC BGWs

Step 1: Insert Pair of BGWs into Each Legacy Site

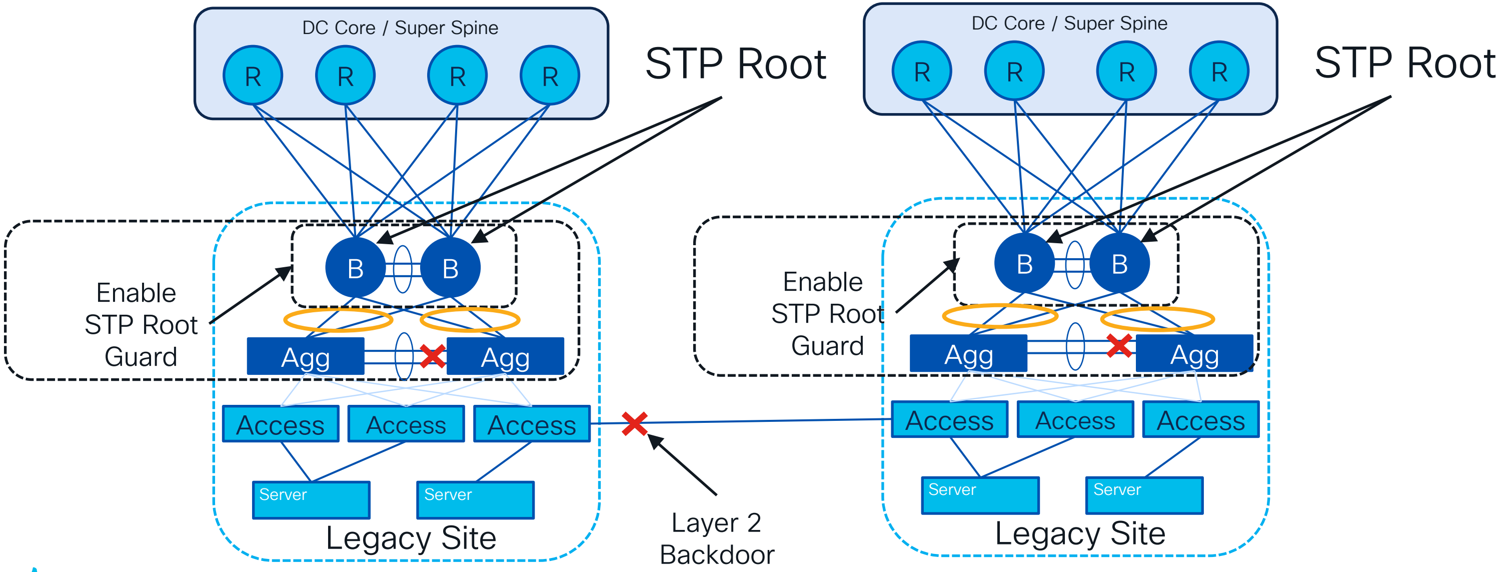
- If existing DC Aggregation devices support VPC/mLAG configure with double-sided VPC
- Double-sided VPC provides for active/active paths and removes need for STP to block paths
  - *NOTE: When the aggregation switches do not support vPC or MLAG, local port-channels can be created from each aggregation switch and the pair of vPC BGW nodes



Migrating legacy to VXLAN EVPN fabrics using vPC BGWs

Step 1 – Cont’d: If Legacy Devices Don’t support VPC/mLAG

- If aggregation switches don’t support vPC/MLAG, local port-channels can be created from each aggregation switch and the pair of vPC BGW nodes
  - STP block the Layer 2 loop created between the aggregation switches and the BGWs – STP root will be on the vPC BGWs

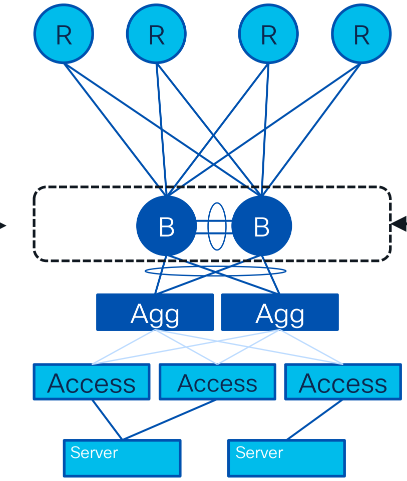


Migrating legacy to VXLAN EVPN fabrics using vPC BGWs

Step 1: Configuration

- Define the vPC domain and properly tune the delay-restore and the reload-delay timers to optimize convergence after a vPC peer reload event.
- Establish iBGP peering relationship along with associated IGP peering (OSPF, ISIS, etc.)

```
feature vpc

 vpc domain 1
  peer-switch
  peer-keepalive destination 172.19.217.122
source 172.19.217.123
  delay-restore 150
  peer-gateway
  auto-recovery reload-delay 360
  ipv6 nd synchronize
  ip arp synchronize

interface port-channel10
  vpc peer-link
```

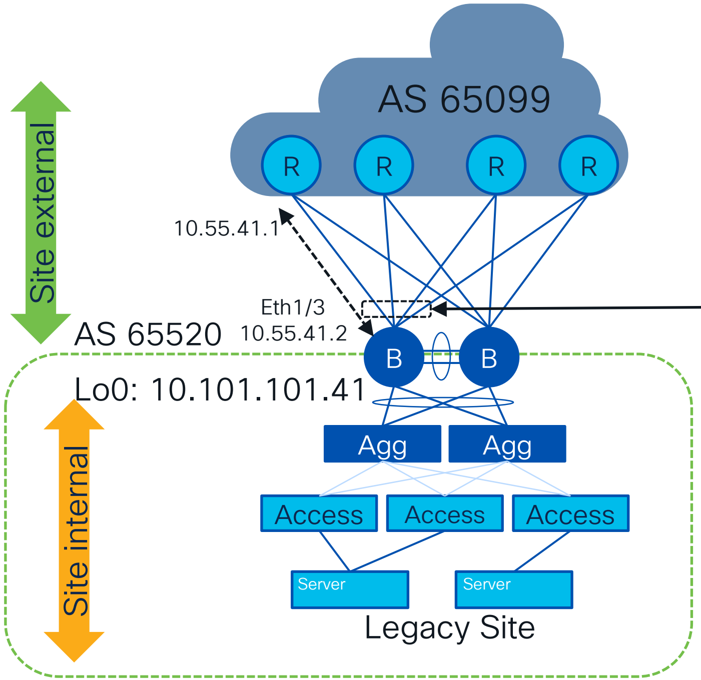


Legacy Site

```
vlan 3600

interface Vlan3600
  description VPC-Peer-Link SVI
  no shutdown
  mtu 9216
  no ip redirects
  ip address 10.1.10.49/30
  no ipv6 redirects
  ip ospf network point-to-point
  ip router ospf UNDERLAY area 0.0.0.0
  ip pim sparse-mode

system nve infra-vlans 3600

router bgp 65501
  neighbor 10.1.10.50
    remote-as 65501
    address-family ipv4 unicast
```

Migrating legacy to VXLAN EVPN fabrics using vPC BGWs

Step 2: Configure vPC BGWs DCI underlay network

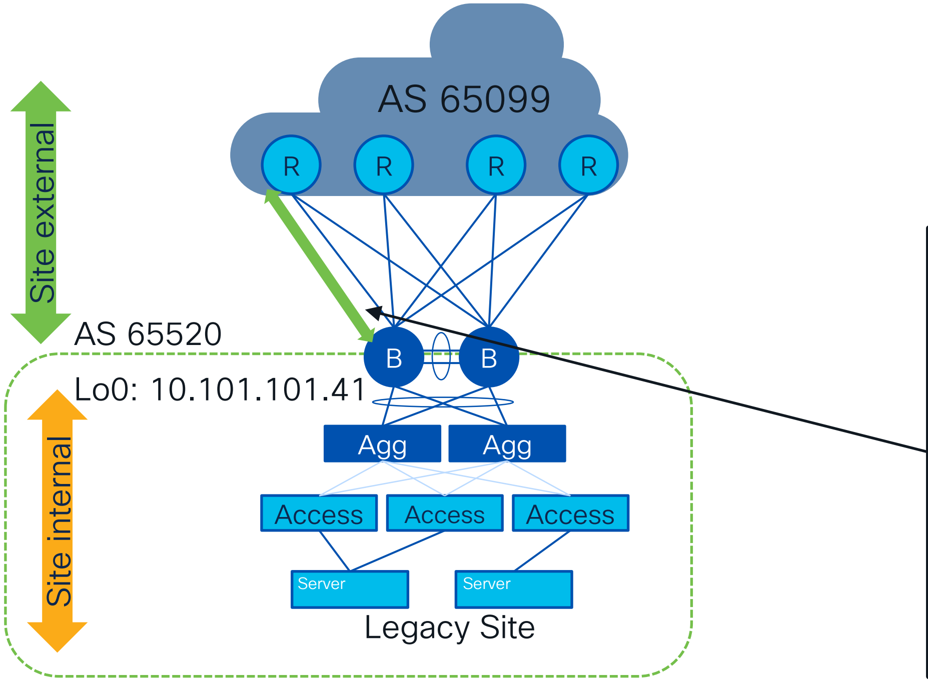


- EVPN Multi-Site interface tracking is required on the interface(s) connecting to the external Layer 3 core to detect the scenario where a given vPC BGW node gets isolated from the external network

```
interface Ethernet1/3
  no switchport
  mtu 9216
  ip address 10.55.41.2/30 tag 54321
  evpn multisite dci-tracking
```

Migrating legacy to VXLAN EVPN fabrics using vPC BGWs

Step 2: Configure vPC BGWs DCI underlay network



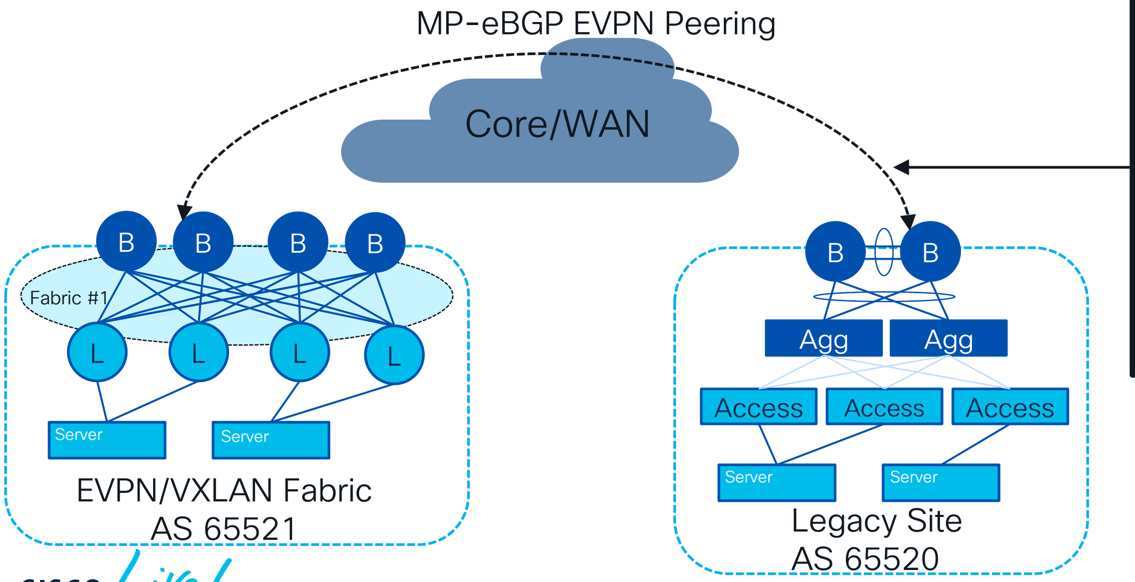
- Configure eBGP peering relationships between each vPC BGW and external peers
- Activate the IPv4 unicast family (VRF default) to redistribute required loopback prefixes and directly connected interfaces

```
router bgp 65520
  router-id 10.101.101.41
  log-neighbor-changes
  address-family ipv4 unicast
    redistribute direct route-map RMAP-REDIST-DIRECT
    maximum-paths 4
    neighbor 10.55.41.1
      remote-as 65099
      update-source Ethernet1/3
      address-family ipv4 unicast

route-map RMAP-REDIST-DIRECT permit 10
  match tag 54321
```

Migrating legacy to VXLAN EVPN fabrics using vPC BGWs

Step 3: Configure vPC BGWs DCI Overlay network

- Configure the remote BGW neighbor(s) with the EVPN address family type L2VPN EVPN enabled
  - The IP address specified for the neighbor represents its loopback0 CP IP address
  - ebgp-multihop command will likely be required to support remote BGW devices
  - The ***peer-type fabric-external*** configuration is required for each remote Multi-Site BGW(s)
  - The ***rewrite-evpn-rt-asn*** configuration is required to enable the rewriting of Route-Target values for prefixes advertised to remote BGWs



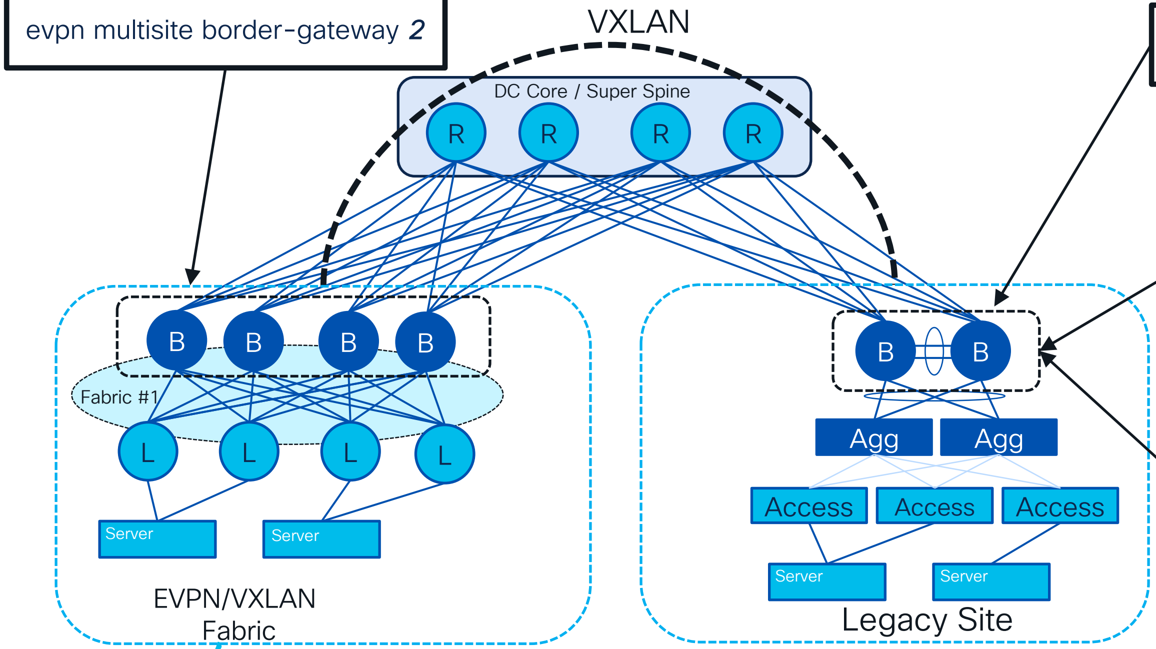
```
router bgp 65520
  router-id 10.101.101.41
  log-neighbor-changes
  neighbor 10.101.201.41
    remote-as 65521
    update-source loopback0
    ebgp-multihop 5
    peer-type fabric-external
    address-family l2vpn evpn
      send-community
      send-community extended
      rewrite-evpn-rt-asn
```

Migrating legacy to VXLAN EVPN fabrics using vPC BGWs

Step 4: Configure vPC BGWs for DCI Layer 2 extension across sites

- Define the site-id on each vPC BGW - the pair of vPC BGWs at the same site must use the same ***site-id*** value
- Define the loopback interface to be used as Multi-Site virtual IP address (Multi-Site VIP), and the loopback interface to be used as Primary IP address (PIP) and vPC virtual IP address (vPC VIP)
- Map the VLANs to the corresponding Layer 2 VNIs.

```
evpn multisite border-gateway 2
```



```
evpn multisite border-gateway 1
  delay-restore time 300
```

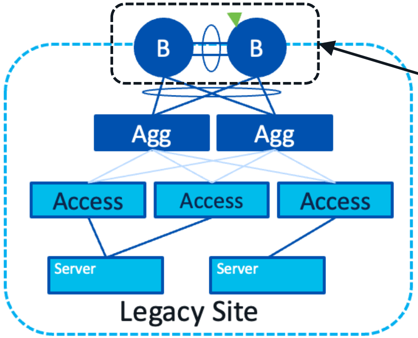
```
interface loopback100
  description Multi-Site VIP
  ip address 10.10.12.1/32 tag 54321
  ip pim sparse-mode
!
interface loopback1
  ip address 10.10.10.1/24 tag 54321 <-- The first
IP is each BGW's PIP and is unique in the pair
  ip address 10.10.11.1/24 secondary tag 54321
```

```
vlan 5
  vn-segment 30005
vlan 6
  vn-segment 30006
```

Migrating legacy to VXLAN EVPN fabrics using vPC BGWs

Step 4 – Con't: Configure vPC BGWs for DCI Layer 2 extension across sites

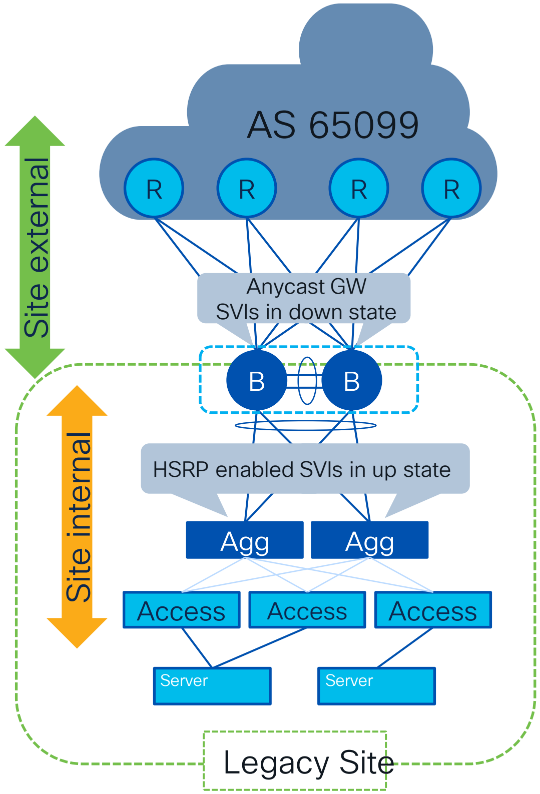
- Associate the Layer 2 VNIs with the NVE interface (VTEP) for selective advertisement. Only the associated Layer 2 VNIs are extended across the DCI.
- NOTE: If VLANs being extended in VXLAN are already extended via a traditional DCI solution (OTV, VPLS), it is critical to avoid the creation of an end-to-end Layer 2 loop between data center sites. This can be achieved in a couple of different ways (on a VLAN-by-VLAN basis):
  - "Flip the switch" - Disable the VLAN extension in traditional DCI solution and start using VXLAN, or;
  - Keep the VLAN extension function via the traditional DCI solution and avoid trunking the VLAN on one of the two vPC connections between the legacy networks and the vPC BGW nodes.



```
interface nve1
  no shutdown
  host-reachability protocol bgp
  source-interface loopback1
  multisite border-gateway interface loopback100
  global ingress-replication protocol bgp
  member vni 30005
    multisite ingress-replication
    ingress-replication protocol bgp
  member vni 30006
    multisite ingress-replication
    mcast-group 239.1.1.1
```

Migrating legacy to VXLAN EVPN fabrics using vPC BGWs

Step 5: Enable Anycast Gateway on vPC BGWs and keep it in shutdown state



- Define the Anycast Gateway MAC address (2020.0000.00AA in this example) for all the defined tenant SVIs
- Map one of the reserved VLANs to the L3 VNI to be used for a given VRF (tenant-1)
- Associate L3VNI to NVE interface (VTEP on BGW)

- Define the SVI to be used as Anycast Gateway and *keep it in shutdown mode*
- Configure the VRF under the BGP process to be able to start exchanging L3 prefixes with the remote BGW nodes:
  - Associate route-map used to redistribute IP subnet information into the EVPN control plane – match on TAG

```
fabric forwarding anycast-gateway-mac
2020.0000.00AA
!
vlan 2001
  vn-segment 50001
vrf context tenant-1
    vni 50001  <-- Maps the tenant/VRF to L3VNI
!
interface nve1
  member vni 50001 associate-vrf
```

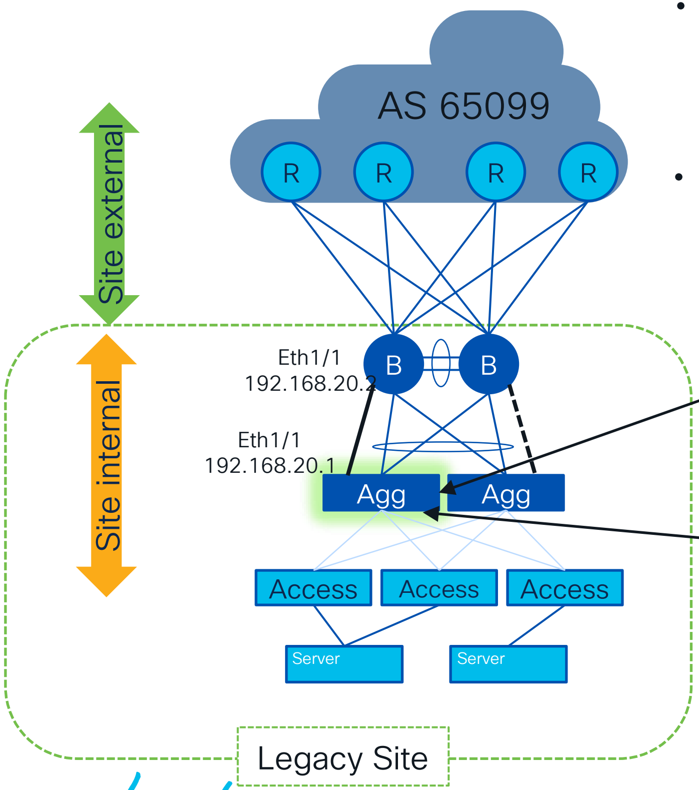
```
interface Vlan5
  shutdown
  vrf member tenant1
  ip address 10.1.5.1/24 tag 12345 ←NOTE: Tag to facilitate
redistribution
  fabric forwarding mode anycast-gateway
```

```
router bgp 65520
    vrf tenant-1
      address-family ipv4 unicast
        redistribute direct route-map FABRIC-RMAP-REDIST-SUBNET
        maximum-paths ibgp 2← only needed for local fabric
      address-family ipv6 unicast
        redistribute direct route-map FABRIC-RMAP-REDIST-SUBNET
        maximum-paths ibgp 2 ← only needed for local fabric
!
route-map FABRIC-RMAP-REDIST-SUBNET permit 10
  match tag 12345
```

Migrating legacy to VXLAN EVPN fabrics using vPC BGWs

Step 6: Migrate first-hop FHRP Gateway in the legacy site to the vPC BGW Anycast Gateway

- Align all FHRP Gateway MAC and IP addresses with the Multi-Site vPC BGW distributed IP Anycast Gateway configuration. You must use the same virtual MAC address for all of the different IP subnets, because the Anycast Gateway virtual MAC address is a global configuration parameter on VXLAN EVPN VTEPs.
- Create a sub-interface per tenant and enable exchange of IPv4 routes with the BGP neighbor.



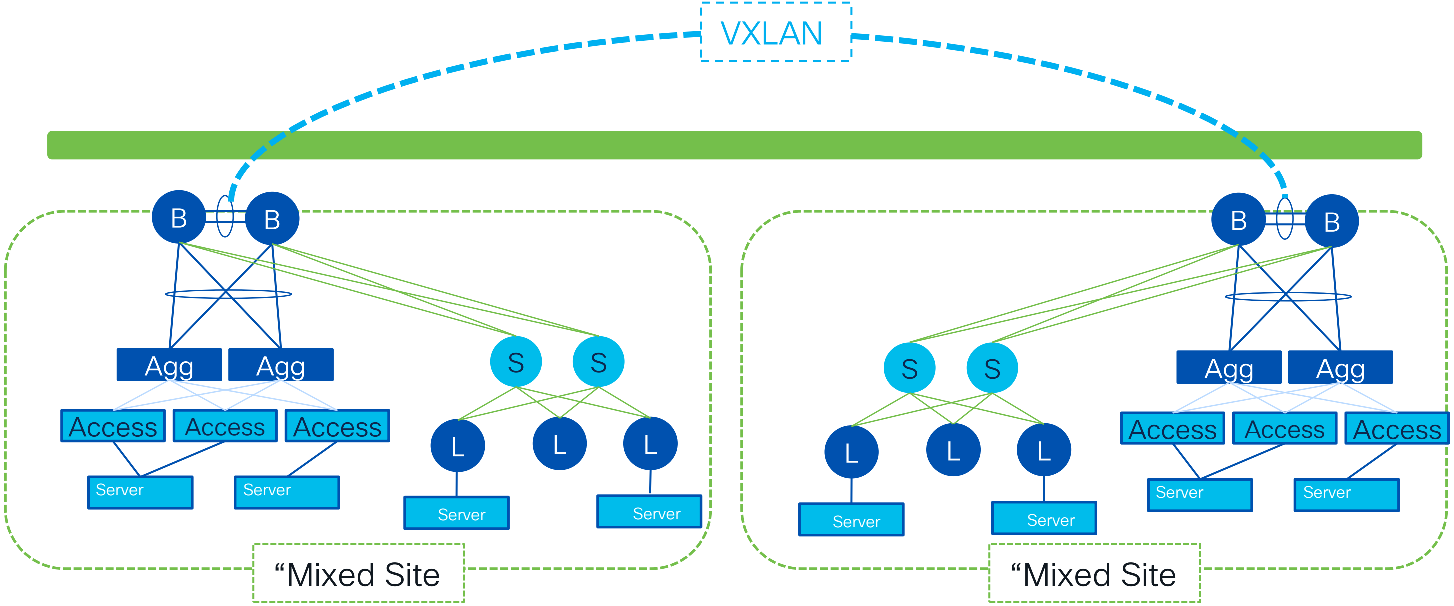
```
interface vlan 20
  vrf member Tenant-A
  ip address 192.168.20.201/24
  hsrp 10
  ip 192.168.20.1
  mac-address 2020.0000.00aa
```

```
interface Ethernet1/1.20
  description L3 Link to vPC BGW1 (T1)
  encapsulation dot1q 20
  vrf member Tenant-A
  ip address 192.168.20.4/31

router bgp 65520
  router-id 100.100.100.1
  vrf Tenant-A
    neighbor 192.168.20.5
      remote-as 65520
      address-family ipv4 unicast
```

Migrating legacy to VXLAN EVPN fabrics using vPC BGWs

Step 7: Transition legacy data centers to new Nexus 9000 EVPN/VXLAN fabric

Connect the new fabric spines to the pair of vPC BGWs with point-to-point Layer 3 links. Modify the configuration on the vPC BGWs to integrate with the new VXLAN EVPN fabric. Those changes do not affect the existing connectivity between the legacy networks.

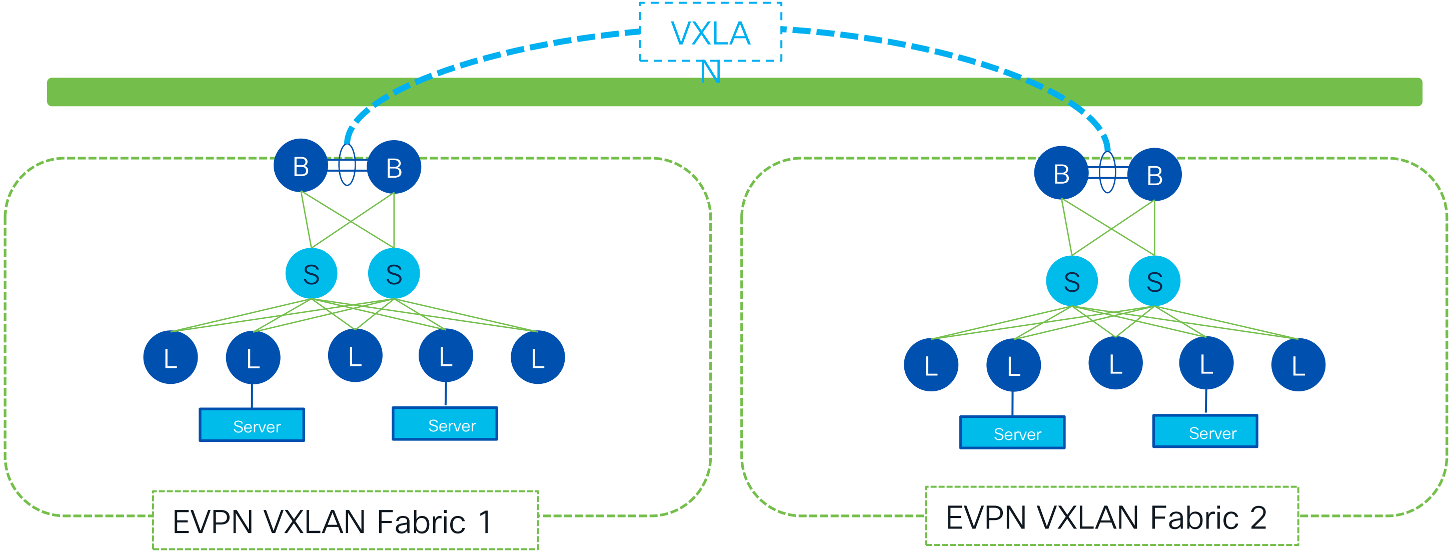


Migrating legacy to VXLAN EVPN fabrics using vPC BGWs

Step 7: Continued

End state of the legacy data center migration to VXLAN EVPN fabrics with vPC BGW nodes

- Getting to this point - Migration of services (Firewall, Load Balancing, DNS, etc.), application workloads and associated dependences have migrated to EVPN fabric
- Notice that the vPC BGW nodes perform the full BGW duties as they allow extending connectivity between endpoints connected to local and remote VTEP devices. This is in contrast with original state in the “legacy” zones, where there was no presence of VTEP nodes inside the local sites.

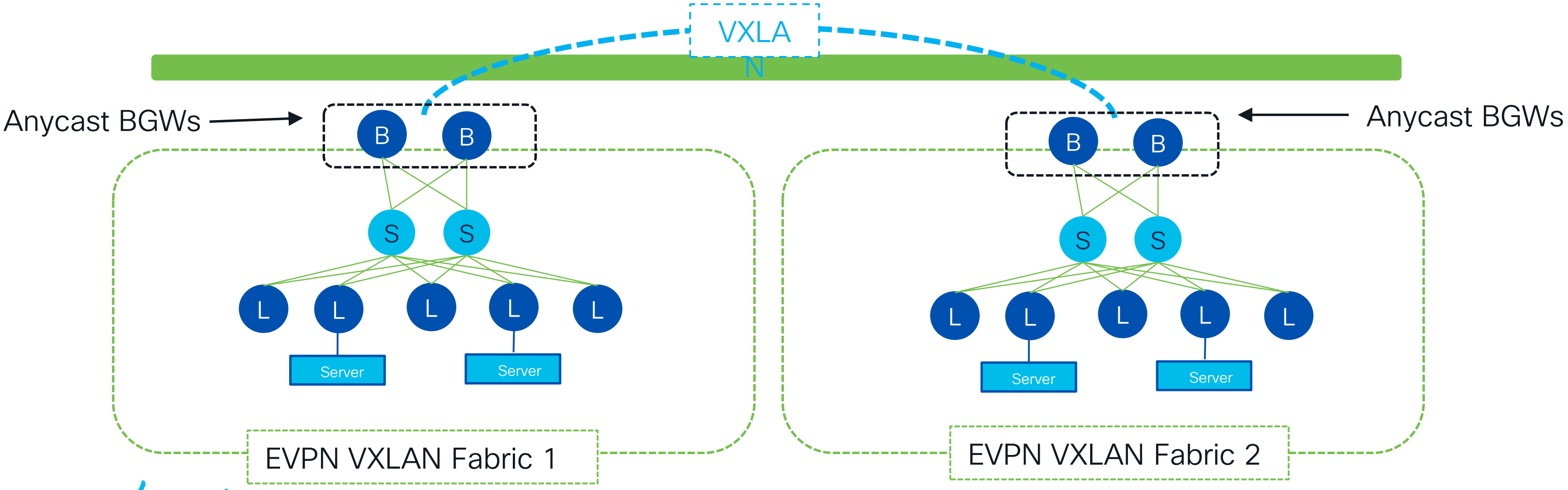


Migrating legacy to VXLAN EVPN fabrics using vPC BGWs

Step 7: Continued

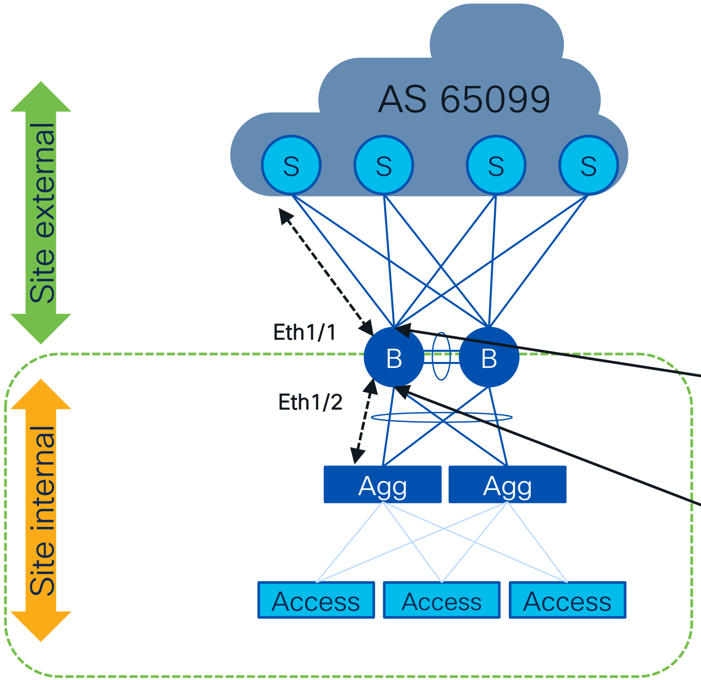
Converting vPC BGWs to Anycast BGWs (Optional, but recommended Last step)

- This is the recommended deployment model for interconnecting VXLAN EVPN fabrics, but it is only possible if there are no endpoints connected to the original vPC BGWs that are using them as their default gateway.
  - **Note:** The conversion to Anycast mode can be performed one BGW at the time, in order not to disrupt the Layer 2 and L3 connectivity between sites.



EVPN Multi-Site vPC BGW failure scenarios

EVPN VXLAN Multi-site BGW Failure Scenarios



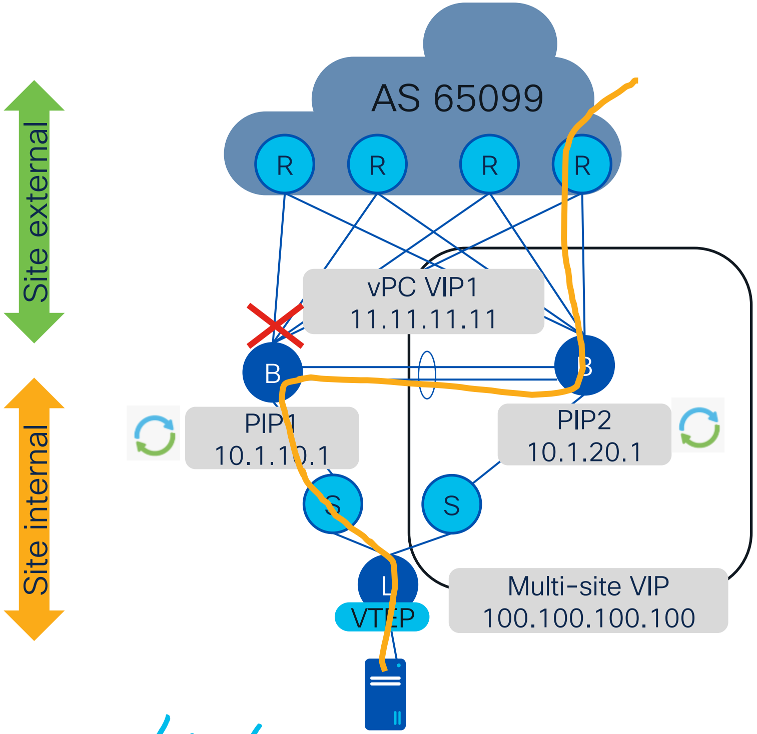
- **EVPN Multi-Site dci-tracking:** interface tracking is required on the interface(s) connecting to the external Layer 3 core to detect the scenario where a given vPC BGW node gets isolated from the external network (Site External)

- **EVPN multi-site fabric tracking:** Interface tracking is the mechanism implemented on each BGW node to detect a potential loss of connectivity toward the site-internal or site-external network, and be able to properly react to those events

```
interface Ethernet1/1
 description L3 Link to Site-External Network
 ip address 10.111.111.1/30
 evpn multisite dci-tracking
```

```
interface Ethernet1/2
 description L3 Link to Site-Internal Network
 ip address 10.0.1.5/30
 evpn multisite fabric-tracking
```

vPC BGW isolation from the site-external network

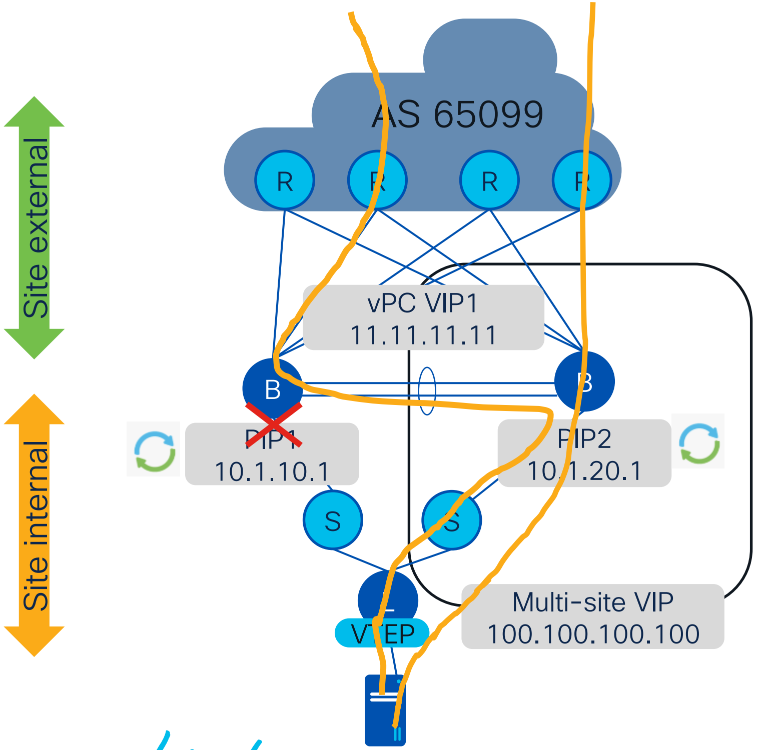


vPC BGW isolation from the site-external network

- Under these circumstances, the following sequence of events will happen on the vPC BGW node isolated from the site-external network:

- The PIP1 and vPC VIP addresses continue to be advertised toward the site-internal network and to the peer BGW via the Layer 3 adjacency established on the vPC peer-link. This is required to allow connectivity to the external network and to local endpoints (only reachable via the isolated BGW node) both from endpoints connected to the local site and in remote sites.

vPC BGW isolation from the site-internal network



vPC BGW isolation from the site-internal network

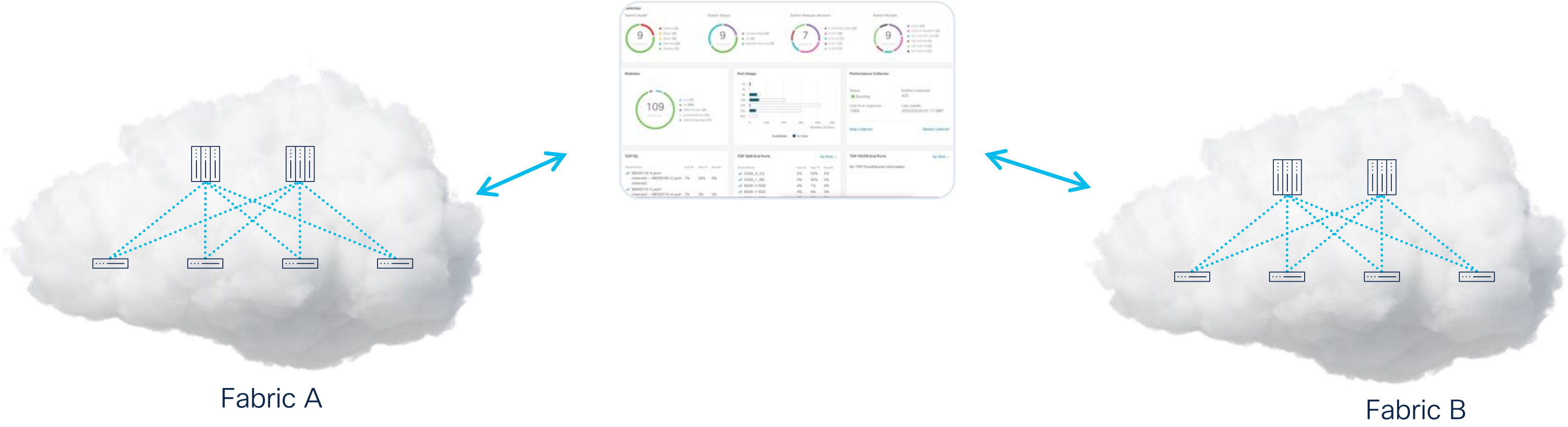
- Under these circumstances, all the logical interfaces on the isolated BGW (PIP, vPC VIP, and Multi-Site VIP) remain active and their addresses are still advertised toward the site-external network (and to the peer BGW via the Layer 3 adjacency established on the vPC peer-link)
- This implies that 50 percent of the traffic flows incoming from remote sites will need to be forwarded via the vPC peer-link, together with the totality of flows originated from endpoints or networks directly connected to the isolated BGW node

Automation and Observability

Nexus Dashboard Fabric Controller and Insights

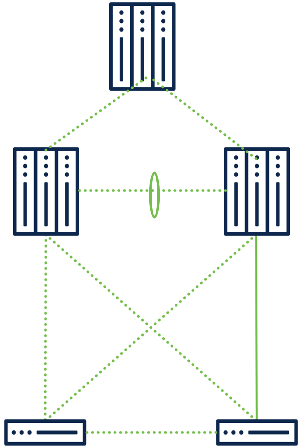
Nexus Dashboard Fabric Controller (NDFC)
Solution Benefits

- Streamlined lifecycle management
- Automate and configure your networks with ease
- Maintain compliance and detect errors
- Extensive visibility, monitoring and modernized topology views
- Expand your network with integrations with NDO and NDI



Enhanced Classic LAN

Profile for Automating Migration of Legacy to EVPN/VXLAN



Classic LAN Fabric

Fully automated fabric - Enhanced Classic LAN

Support for greenfield and brownfield deployments

Provisioning of 3tier architecture/ L2/L3 Networks and VRFs

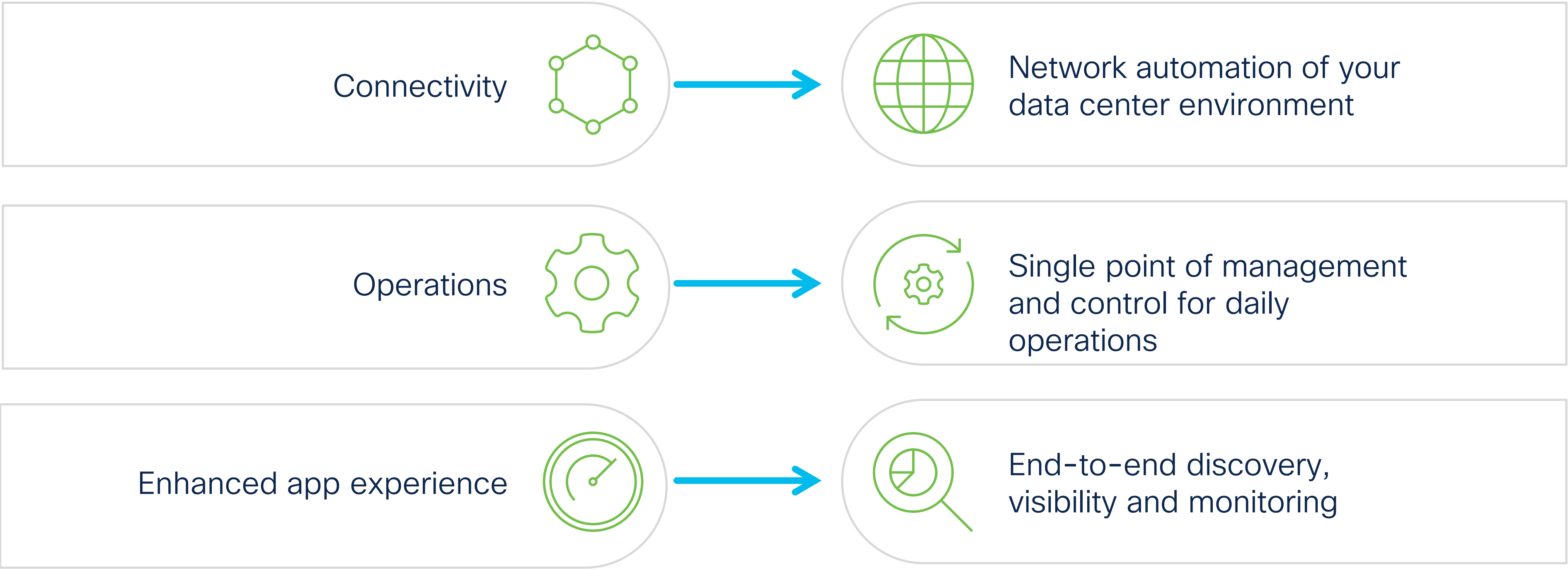
VRF-Lite Between Agg and Core

Benefits

Best Practice Templates | Simplified workflows | Flexibility based on customer needs

Cisco NDFC & Nexus Insights

Seamless integration with Day 2 operations for in depth telemetry analytics



Connectivity → Network automation of your data center environment

Operations → Single point of management and control for daily operations

Enhanced app experience → End-to-end discovery, visibility and monitoring

Conclusion

Conclusion – Key Take-Aways

#1

vPC Border Gateways

Provides an Industry Standard method to migrate off Legacy DC Tech
Flexible Integration model with older Network Gear
Proven technology with documented Migration Plans
Coordination with Application Teams once Migration Path is ready
Nexus Dashboard for Automation, Management and Visibility

#2

VXLAN BGP EVPN Multi-Site

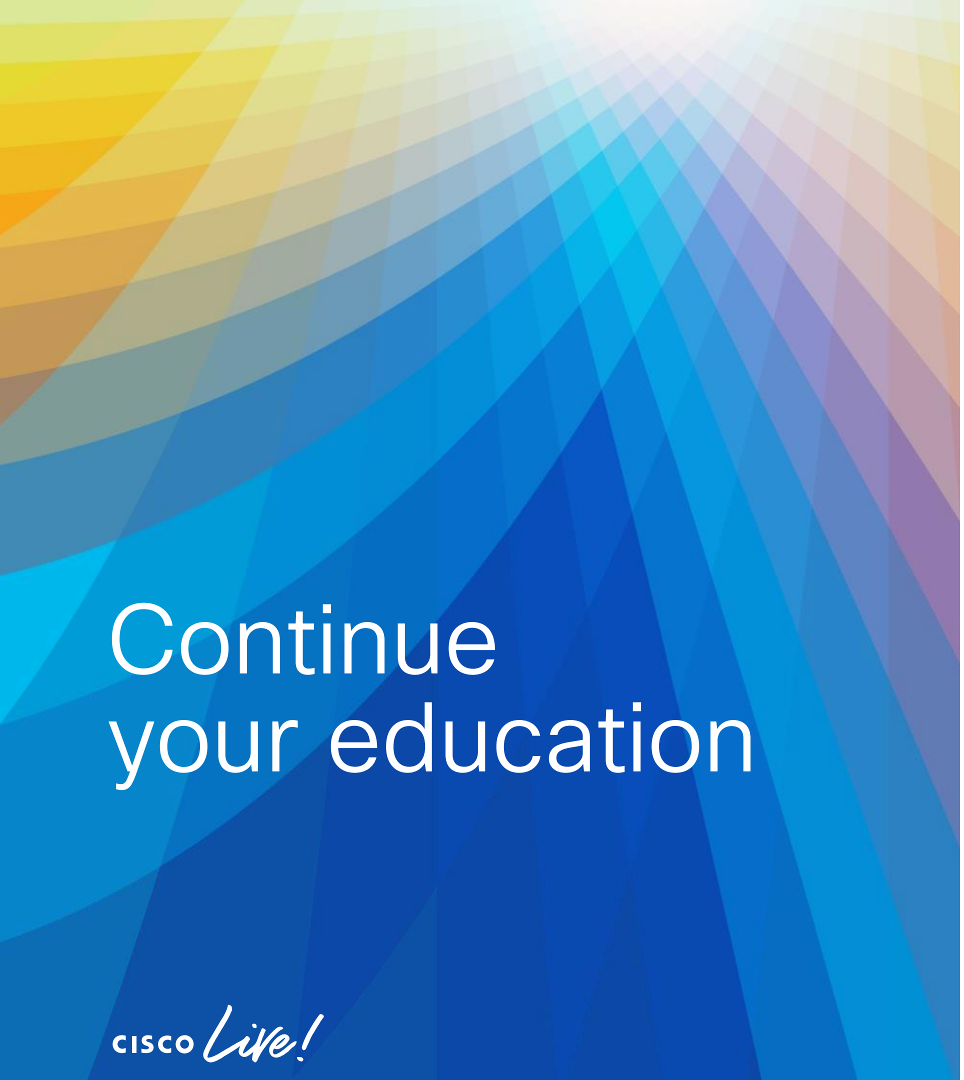
A Simple add or drop-in
First introduced in September 2017 – proven and deployed
A Solution beyond EVPN DCI Overlay (RFC9014)
Provides Layer-2 and Layer-3 extension
Wide Hardware Support
Flexible Deployment Option - Not just for VXLAN Fabrics
Nexus Dashboard for Automation, Management and Visibility

Fill out your session surveys!

Attendees who fill out a minimum of four session surveys and the overall event survey will get **Cisco Live-branded socks** (while supplies last)!

Attendees will also earn 100 points in the **Cisco Live Challenge** for every survey completed.

These points help you get on the leaderboard and increase your chances of winning daily and grand prizes



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

Thank you

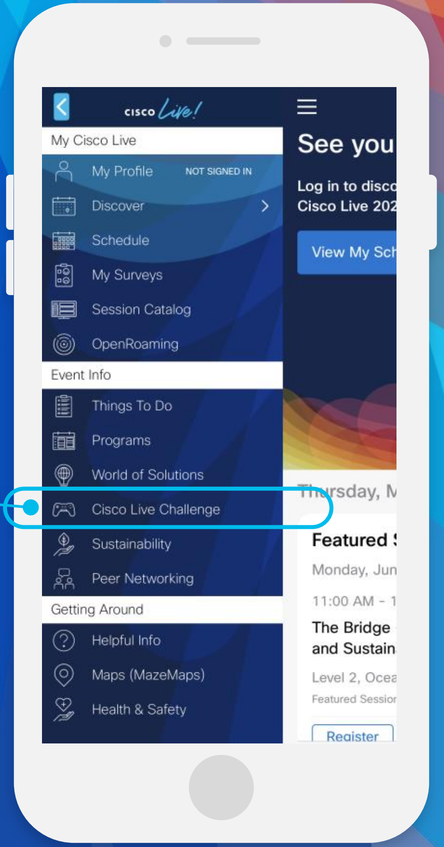
Cisco Live **Challenge**

Gamify your Cisco Live experience!

Get points for attending this session!

How:

1. Open the Cisco Events App.
2. Click on 'Cisco Live Challenge' in the side menu.
3. Click on View Your Badges at the top.
4. Click the + at the bottom of the screen and scan the QR code:



cisco Live!

Let's go

#CiscoLive