EVPN VXLAN Multi-Fabric

New features:
EVPN multi-hop (data-plane)
EVPN next-hop-self and route-map (control-plane)

AOS-CX 10.09

Vincent Giles / Daryl Wan

Technical Marketing Engineering

Objective of this presentation

- Provide a technical deep dive on EVPN based VXLAN multi-fabric solution for datacenter
- Schedule for 3 hours
- Detailed analysis of architecture
- Reference document to support new deployment
- Comprehensive list and analysis of show commands
- Provide configuration files for reference
- Provide packet capture files to serve as reference when debugging

Technology highlights:

- Dual-homing: logical VTEP = VSX VTEP
- Symmetric IRB and Distributed anycast GW
- ARP suppression
- DHCP-relay in overlay
- MP-BGP (iBGP and **eBGP**)
- **Next-hop-self**
- **Route-map**
- **Split Horizon and VXLAN tunnels bridging**

Agenda

1. Overview
2. Use Cases
3. Details and Caveats
4. Configuration
5. Best Practices
6. Troubleshooting
7. Demo
8. Additional Resources

Overview

Definitions

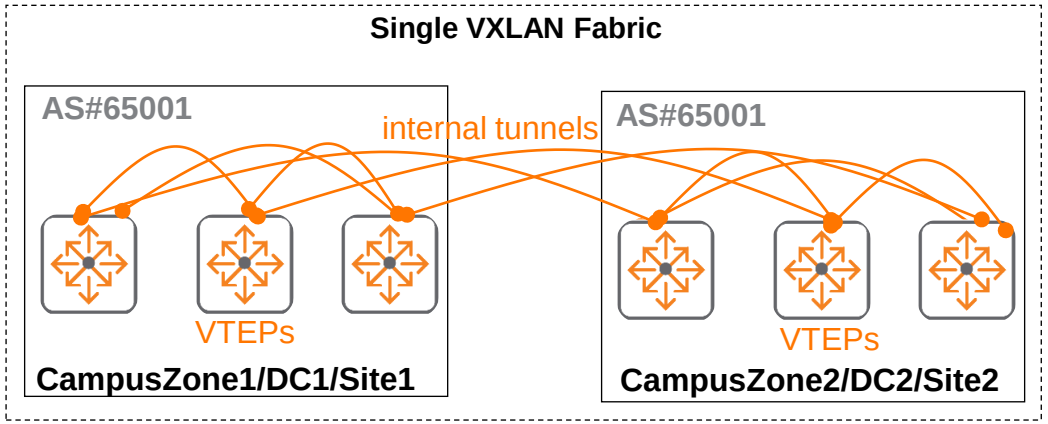
Acronyms

- VXLAN — **V**irtual e**X**tensible **LAN**
- VTEP — **V**XLAN **T**unnel **E**nd **P**oint
- VNI — **V**XLAN **N**etwork **I**dentifier
- L2VNI — **L**ayer2 **V**XLAN **N**etwork **I**dentifier (to extend L2 traffic)
- L3VNI — **L**ayer3 **V**XLAN **N**etwork **I**dentifier (to send routed traffic)
- EVPN — **E**thernet **V**irtual **P**rivate **N**etwork
- MP-BGP — **M**ulti-**P**rotocol **B**order **G**ateway **P**rotocol
- AF — **A**ddress **F**amily (Ex: IPv4, IPv6 or EVPN address families used in MP-BGP)
- MB-BGP EVPN — Refers to the EVPN AF in MP-BGP
- RT — Refers to EVPN **R**oute-**T**ype or Type of Route: (AOS-CX supports RT2, RT3, RT5)
- VRF — **V**irtual **R**outing and **F**orwarding
- IRB — **I**ntegrated **R**outing and **B**ridging (symmetric or asymmetric IRB used in VXLAN overlay)
- VSX — **V**irtual **S**witching e**X**tension
- ISL — **I**nter **S**witch **L**ink (link between VSX peers)
- AG — **A**ctive **G**ateway (anycast IP address used for default-gateway)
- VSX VTEP — VTEP function hosted on a VSX cluster for dual-homing capability.
- NHS — **N**ext-**H**op-**S**elf
- NHU — **N**ext-**H**op-**U**nchanged
- Border VTEP — VTEP acting as boundary for the Fabric
- BorderLeader — Border VTEP hosting BGP sessions with other Fabrics
- Local Fabric — internal Fabric (iBGP)
- Remote Fabric — external Fabric (eBGP)
- iBGP — internal BGP
- eBGP — external BGP
- ASN — Autonomous System Number (used in BGP)
- DCI — Data-Center-Interconnect
- POD — Point Of Delivery

AOS-CX 10.09 EVPN VXLAN Multi-Fabric Solution Overview

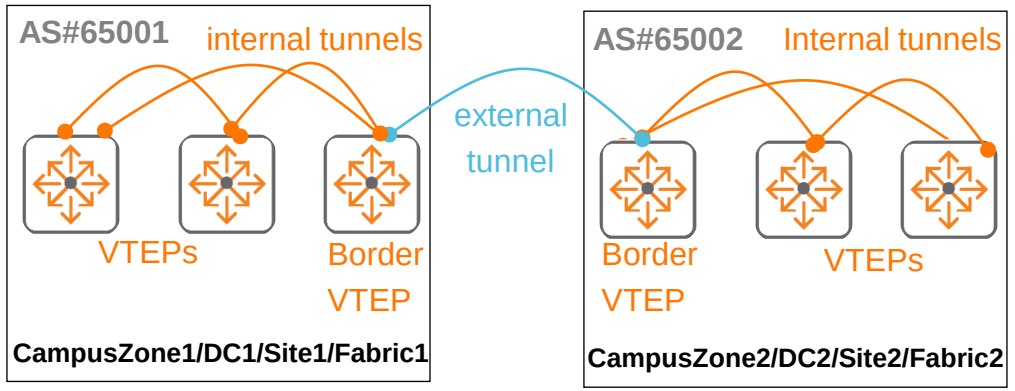
- 10.09 adds EVPN VXLAN Multi-Fabric to scale beyond single VXLAN fabric limits:
  - Supports tunnel to tunnel forwarding (intermediate VTEP hopping) for both IPv4 L2/L3 traffic between fabrics
  - Unidimensional scale limit of **256** BGP EVPN VTEP peers is now **per fabric**
  - Applicable to both campus and Data Center (DC) VXLAN deployments, refer to caveats section
- Supported platforms: 8325

Single-hop single VXLAN fabric



- Full mesh VXLAN tunnels between VTEPs within a VXLAN fabric
- A single VXLAN fabric with limited VTEP scale

Multi-hop VXLAN with multiple fabrics

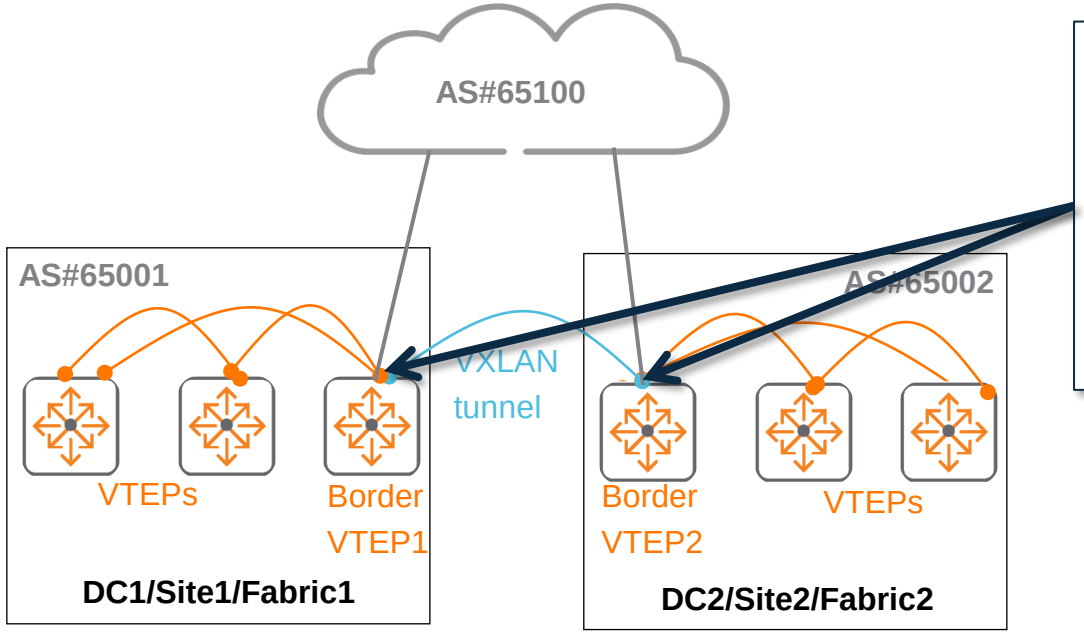


- Full mesh VXLAN tunnels between VTEPs within a VXLAN fabric
- VTEPs between different fabrics avoid full mesh tunnels and BGP peering between fabrics
- Border VTEPs establish VXLAN tunnels between fabrics
- Border VTEPs provide tunnel to tunnel forwarding and prevents loops between fabrics
- Multiple VXLAN fabrics provides improved VTEP scale

EVPN Route-Map Overview

AOS-CX 10.09 MP-BGP EVPN enhancement

- EVPN route-map support for VXLAN deployments
- Allows a VTEP to apply an inbound or/and outbound route-map to a BGP peer under EVPN address-family context
  - Match conditions: IPv4/IPv6 address prefix-list, aspath-list, VNI (L2/L3)
  - Set clauses: local-preference, as-path prepend/exclude, IP next-hop



Objective: Border VTEP2 should be preferred as outbound VTEP over Border VTEP1 for routes towards AS#65100

Example:
- On VTEP2, set higher local-preference for EVPN routes learnt from AS#65100 and, lower LP on VTEP1 for AS#65100 routes.
- Set AS-path prepending on VTEP1 for incoming traffic to reach VTEP2 instead of VTEP1.

- Required by multi-fabric VXLAN when multiple pods/fabrics/AS#s within each DC/site are used, hence including border-leaders.

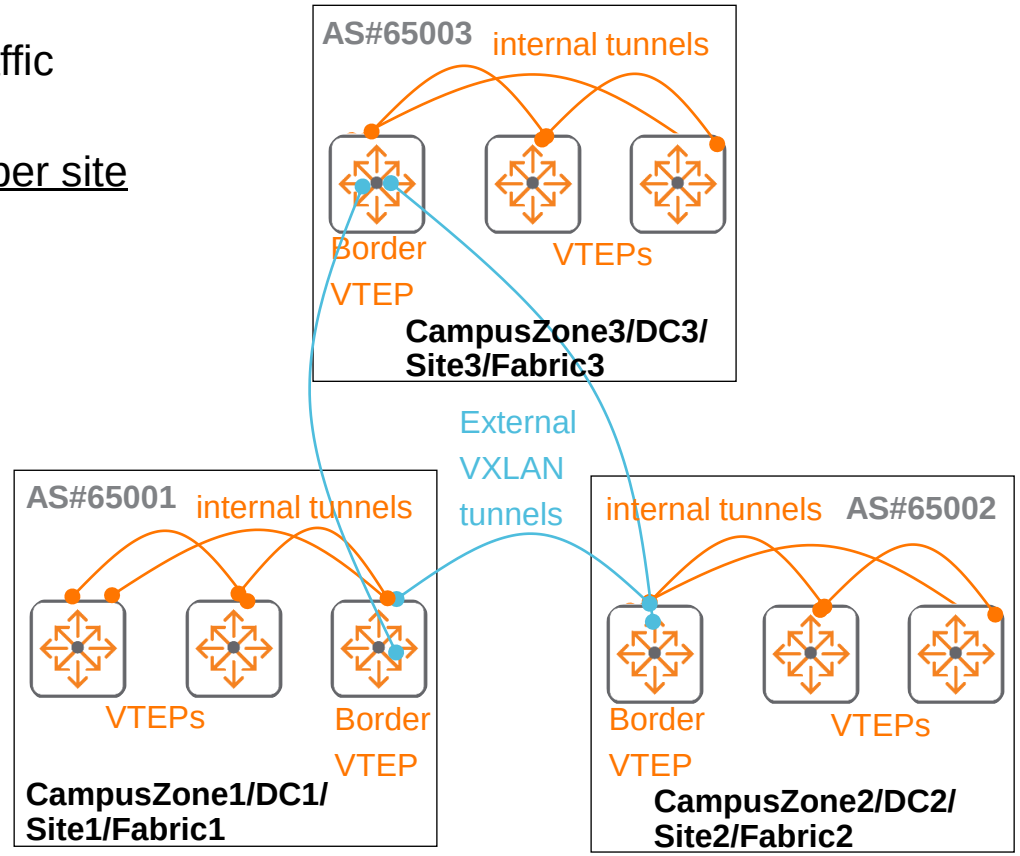
Use Cases

- Multi-Fabric use-cases for L2 and L3 DCI
- Other EVPN route-map use-cases

One Fabric (1 ASN) per Campus Zone or DC or Site

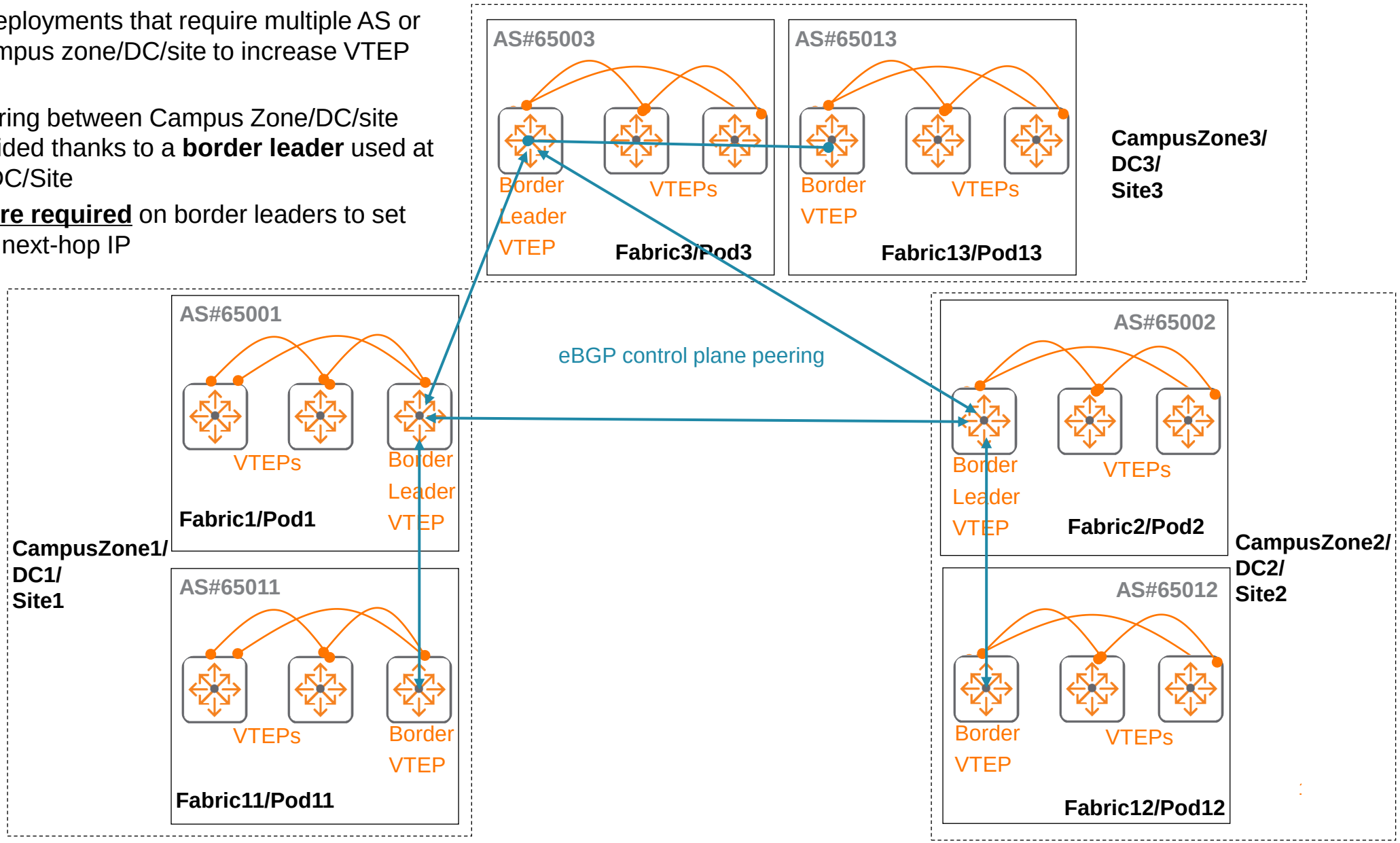
Recommended for majority of deployments

- To increase VTEPs scale beyond one single fabric
- Supports tunnel-to-tunnel forwarding for both L2/L3 traffic between fabrics
- Unidimensional scale of 256 BGP EVPN VTEP peers per site
- EVPN **route-map not required**



Multiple pods/fabrics/AS#s within each Campus Zone/DC/site

- Recommended for deployments that require multiple AS or pods within each Campus zone/DC/site to increase VTEP scale even further
- Full mesh eBGP peering between Campus Zone/DC/site border VTEPs is avoided thanks to a **border leader** used at each CampusZone/DC/Site
- EVPN **route-maps are required** on border leaders to set correct border VTEP next-hop IP

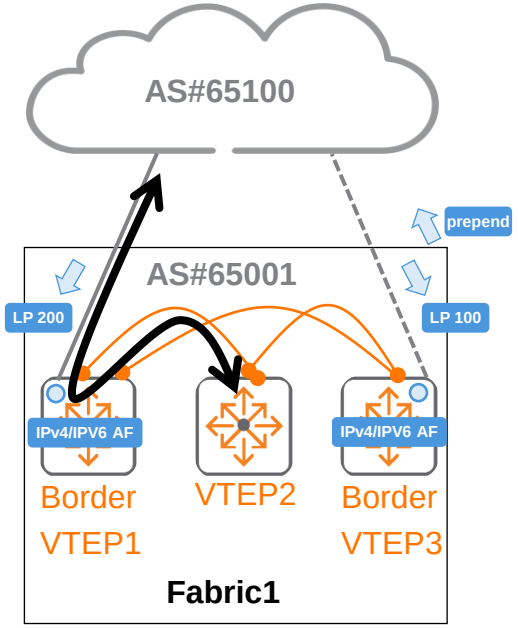


EVPN route-map use-case

Multi-path Routing Design

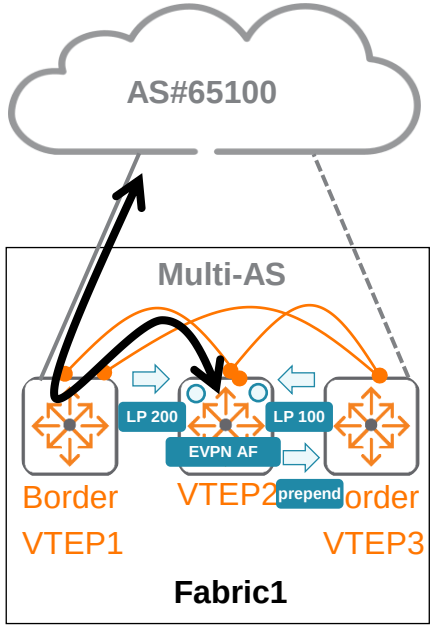
AOS-CX 10.08

- Limited to import of IPv4/IPv6 routes into EVPN BGP LP attribute retention.
- Routing preference set on **IPv4/IPv6 AF** routes.
- Limited to iBGP Fabric.



AOS-CX 10.09

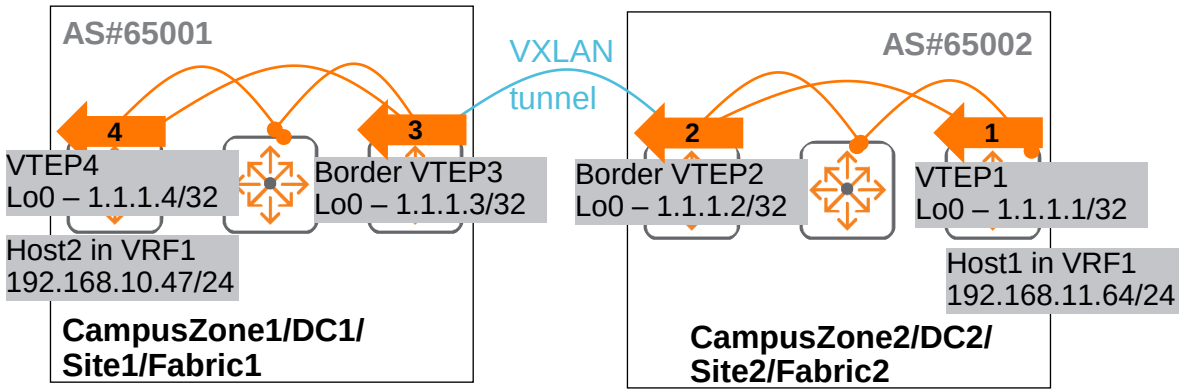
- Introduction of route-map to set routes preference on any VTEP.
- Routing preference set on **EVPN AF** routes.
- Enable multi-AS EVPN use-cases.



Details

with example

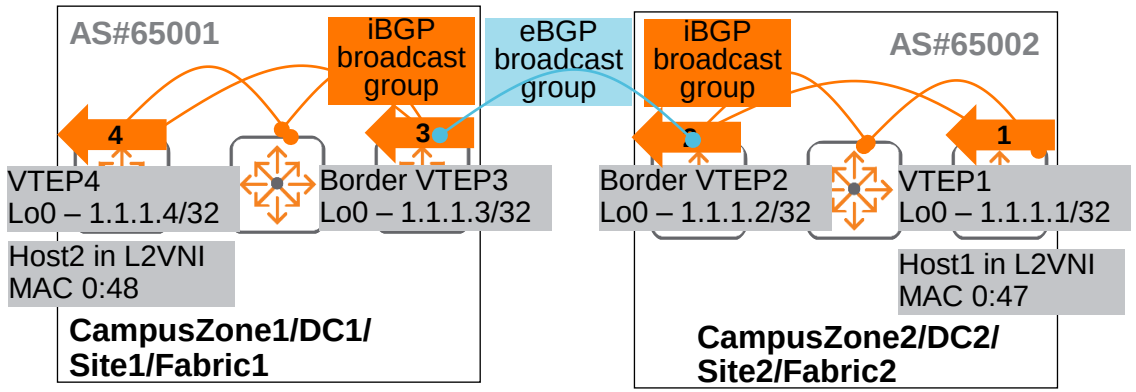
L3 Multi-Fabric: 1 fabric/1 AS# per Campus Zone/DC/site



VSX logical VTEPs for High Availability (HA) are supported and recommended

- Assuming EVPN control plane and VXLAN tunnels are established between all VTEPs, host routes are advertised and learnt.
- When Host1 connected to VTEP1 sends traffic to Host2 connected to VTEP4:
  1. VTEP1 sees destination IP in VRF1 route table with next-hop IP of Border-VTEP2 via L3VNI of VRF1.
     VTEP1 encapsulates and forwards traffic into VXLAN tunnel with next-hop IP of Border-VTEP2.
  2. Border-VTEP2 sees destination IP in VRF route table with next-hop IP of Border-VTEP3 via L3VNI of VRF1.
     Border-VTEP2 performs tunnel to tunnel forwarding with next-hop IP of Border-VTEP3.
  3. Border-VTEP3 sees destination IP in VRF route table with next-hop IP of VTEP4 via L3VNI of VRF1.
     Border-VTEP3 performs tunnel to tunnel forwarding with next-hop IP of VTEP4.
  4. VTEP4 sees destination IP in VRF route table with interface/port of directly connected host.
     VTEP4 performs VXLAN decapsulation and forwards traffic to Host2.
- L3 EVPN-VXLAN multi-fabric is only required on Border-VTEPs.
- VTEP1 and VTEP4 could be VTEPs that do not support EVPN-VXLAN multi-fabric.

L2 Multi-Fabric: 1 fabric/1 AS# per Campus Zone/DC/site

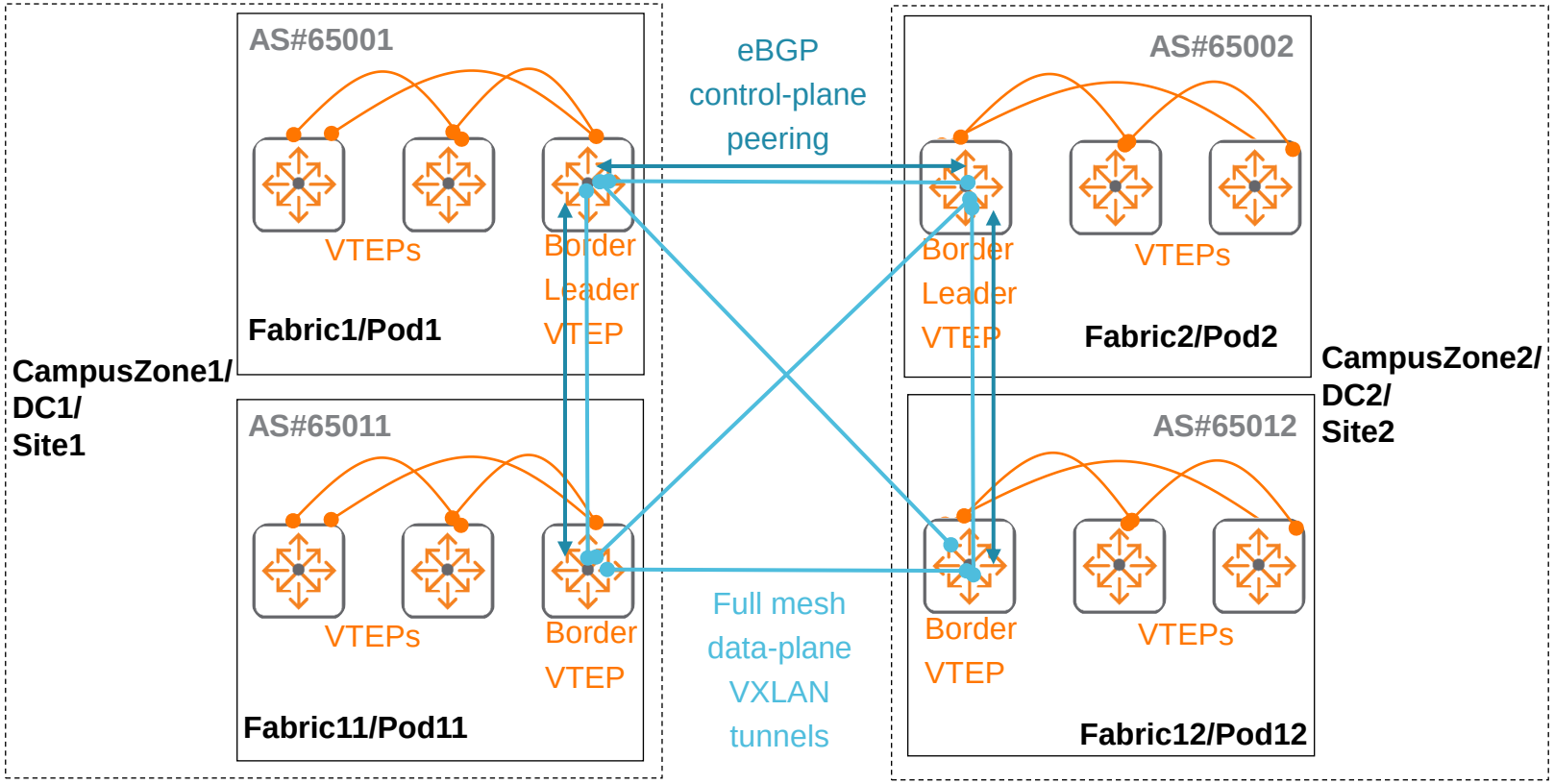


VSX logical VTEPs for High Availability (HA) are supported and recommended

- VXLAN split horizon enabled by default for BUM traffic.
  L2 tunnel to tunnel forwarding requires disabling VXLAN split-horizon between iBGP and eBGP broadcast groups on Border-VTEPs.
- Assuming EVPN control plane and VXLAN tunnels are established between all VTEPs, MAC addresses are advertised and learnt.
- When Host1 connected to VTEP1 sends traffic to Host2 connected to VTEP4:
  1. VTEP1 sees destination MAC in MAC table with next-hop IP of Border VTEP2 via L2VNI.
     VTEP1 encapsulates and forwards traffic into VXLAN tunnel with next-hop IP of Border-VTEP2.
  2. Border-VTEP2 sees destination MAC with next-hop IP of Border-VTEP3 via L2VNI.
     Border-VTEP2 performs tunnel to tunnel forwarding with next-hop IP of Border-VTEP3.
  3. Border-VTEP3 sees destination MAC with next-hop IP of VTEP4 via L2 VNI.
     Border-VTEP3 performs tunnel to tunnel forwarding with next-hop IP of VTEP4.
  4. VTEP4 sees destination MAC with interface/port of directly connected host.
     VTEP4 performs VXLAN decapsulation and forwards traffic to Host2.
- L2 EVPN-VXLAN multi-fabric is only required on Border-VTEPs.
- VTEP1 and VTEP4 could be VTEPs that do not support EVPN-VXLAN multi-fabric .

L2/L3 Multi-Fabric:
Multiple pods/fabrics/AS#s within each Campus Zone/DC/site

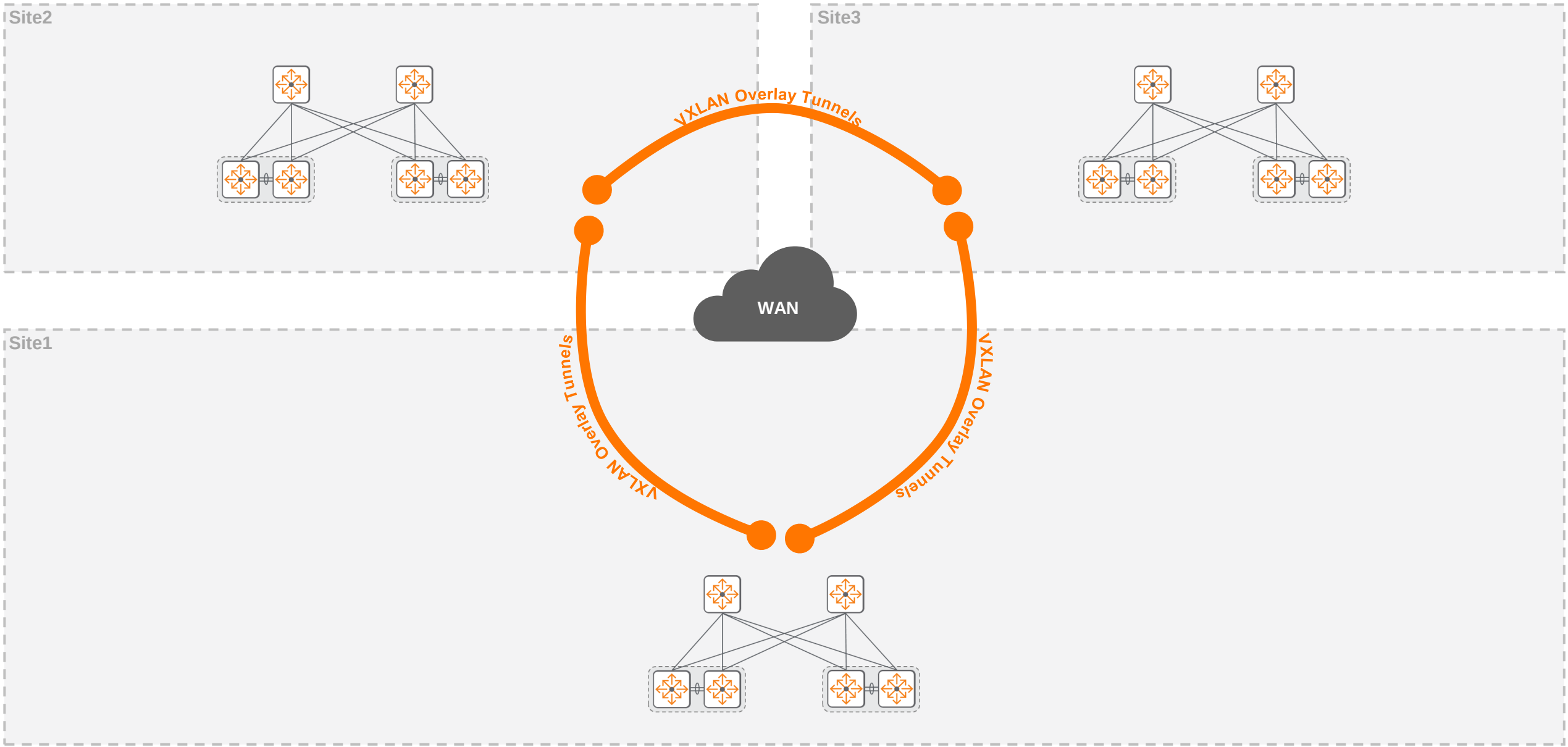
- Full mesh eBGP control-plane is not required
  - Border-VTEP in a fabric does not peer with another Border-VTEP in a different fabric.
  - Border-Leader-VTEP will peer with Border-VTEPs in the same Site and Border-Leader-VTEPs in other CampusZone/DC/Sites.
- Full mesh VXLAN data-plane is required between all Border-Leader-VTEPs and Border-VTEPs



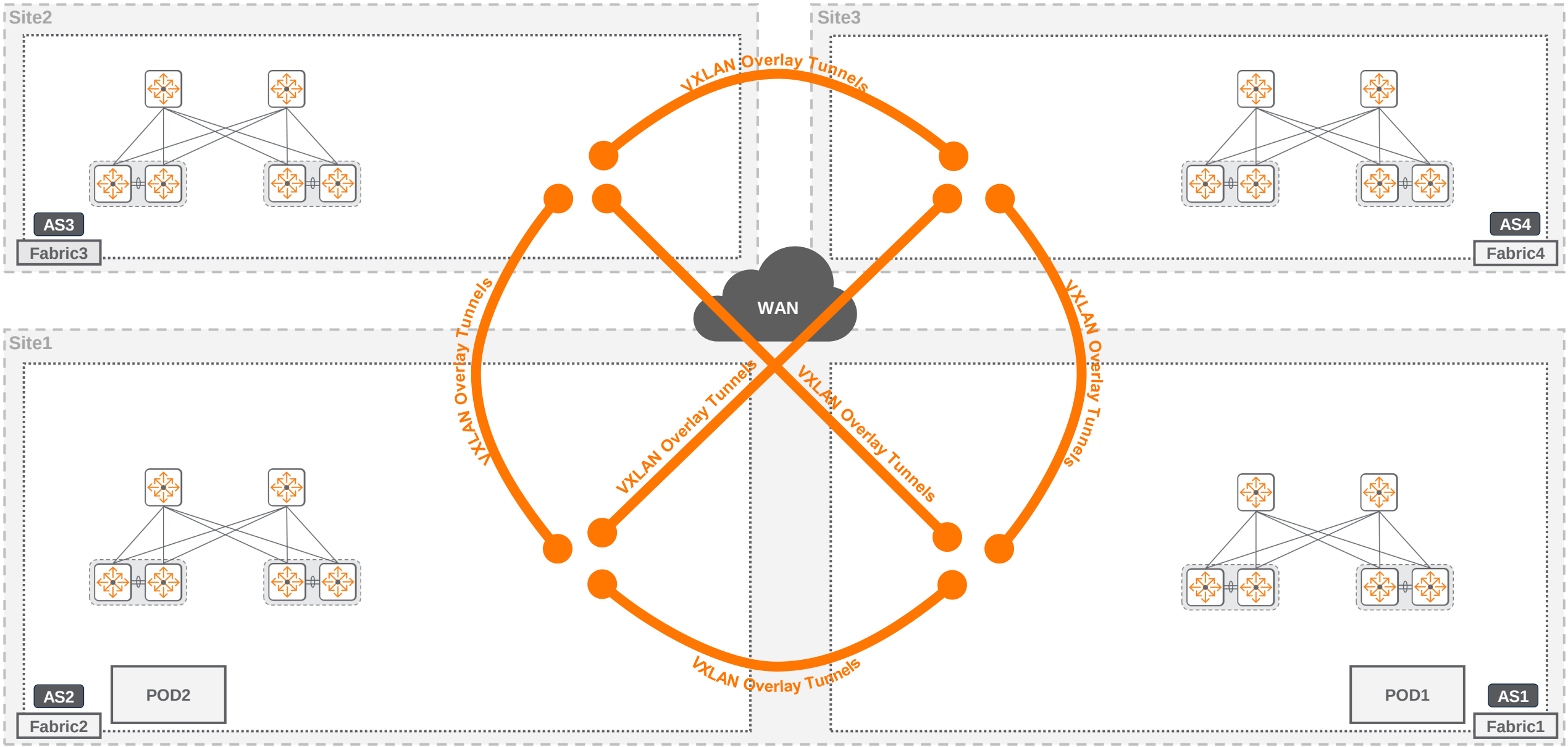
VSX logical VTEPs for High Availability (HA) are supported and recommended

Solution Example

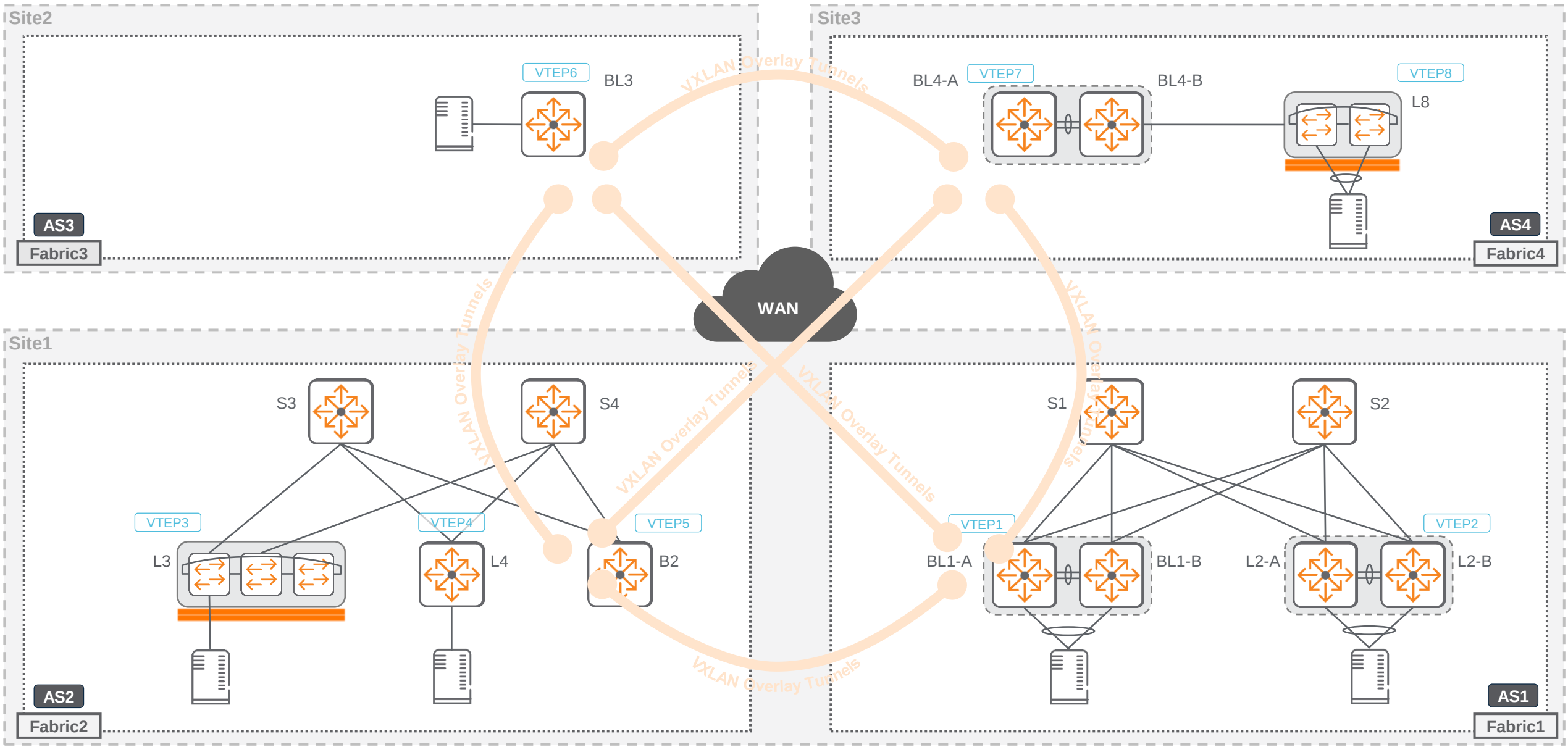
Solution Objective: Multi-Site VXLAN



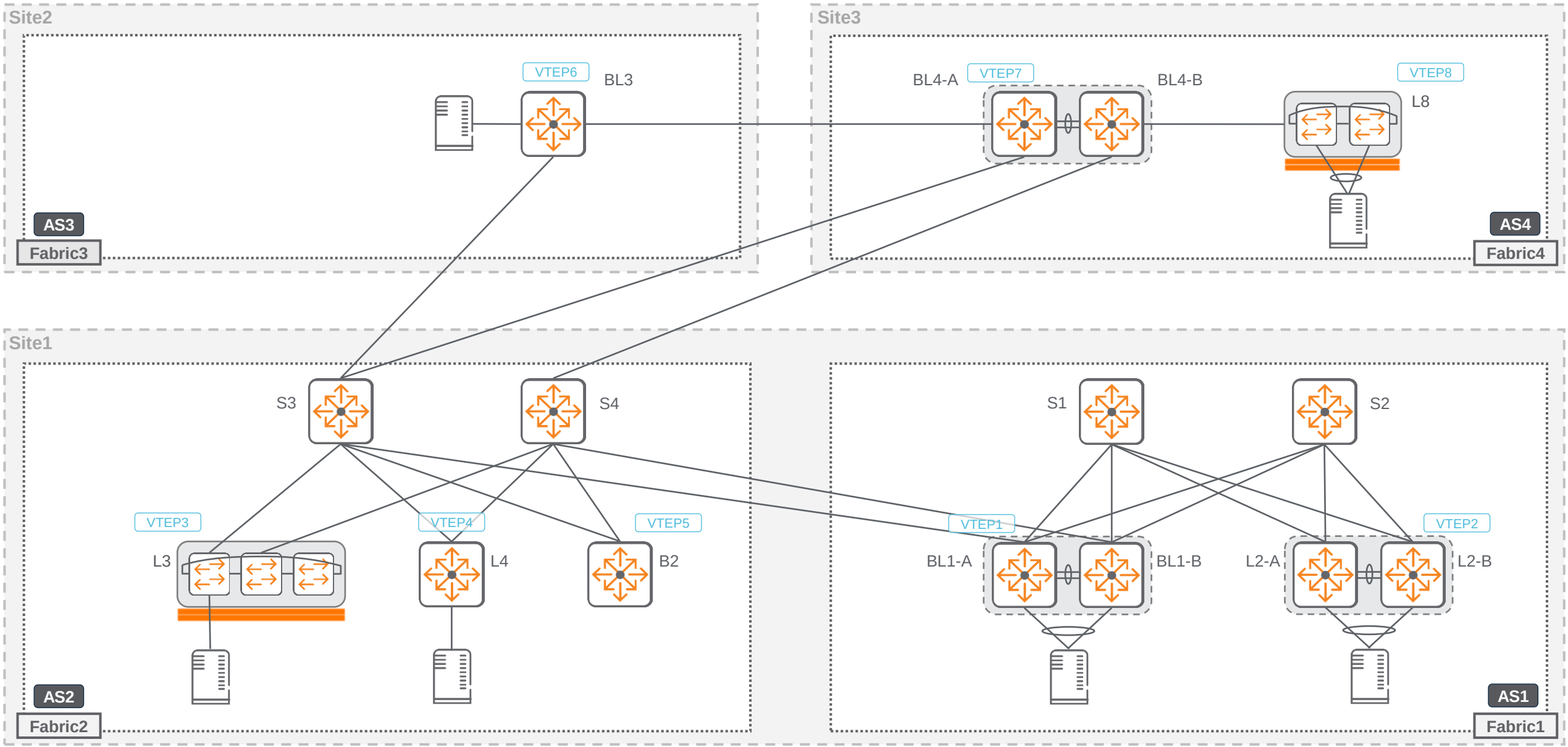
Solution Objective: Multi-Site + Multi-Fabric VXLAN



Multi-Hop Example: Multi-Site + Multi-Fabric VXLAN

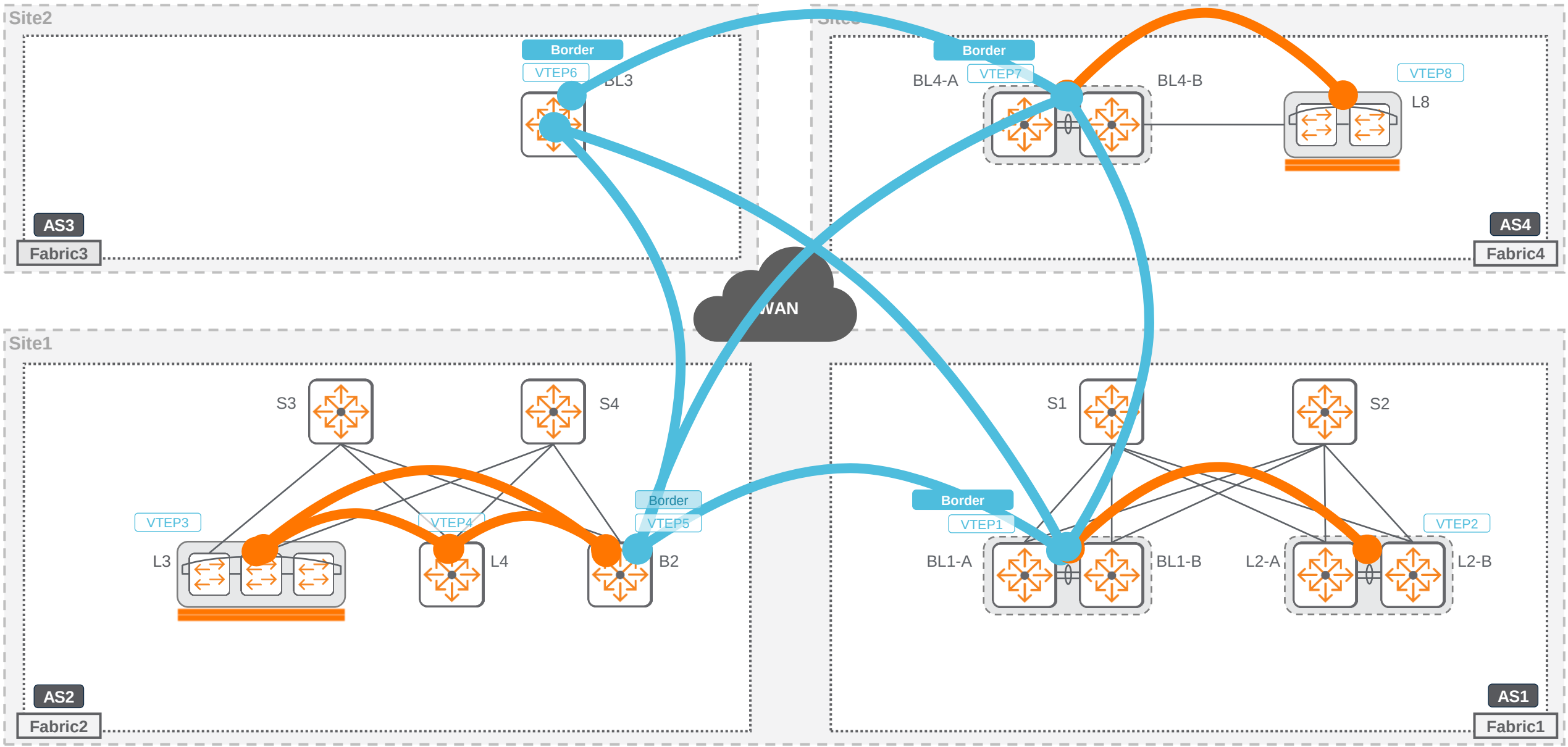


Physical Topology



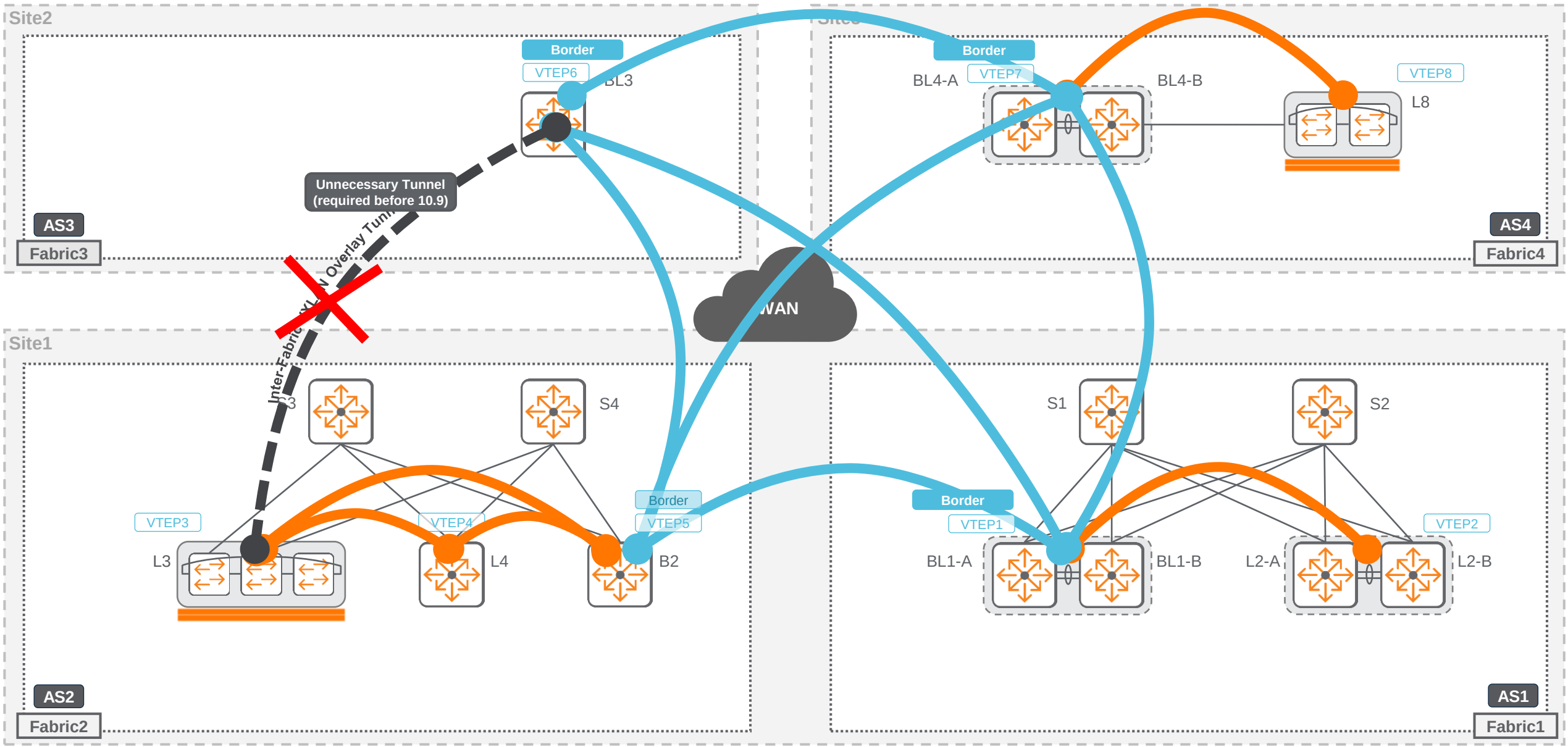
VXLAN Dataplane

intra-Fabric VTEP full-mesh, inter-Fabric border full-mesh



VXLAN Dataplane

no meshing between inter-Fabric non-border Leaves



Inter-Fabric Traffic

Max 3 VXLAN tunnels - One Inter-Fabric tunnel in the overlay path

Site2

Border

VTEP6

BL3

AS3

Fabric3

Site3

Border

BL4-A

VTEP7

BL4-B

VTEP8

L8

AS4

Fabric4

WAN

Site1

S3

S4

VTEP3

VTEP4

Border

VTEP5

L3

L4

B2

AS2

Fabric2

S1

S2

Border

VTEP1

VTEP2

BL1-A

BL1-B

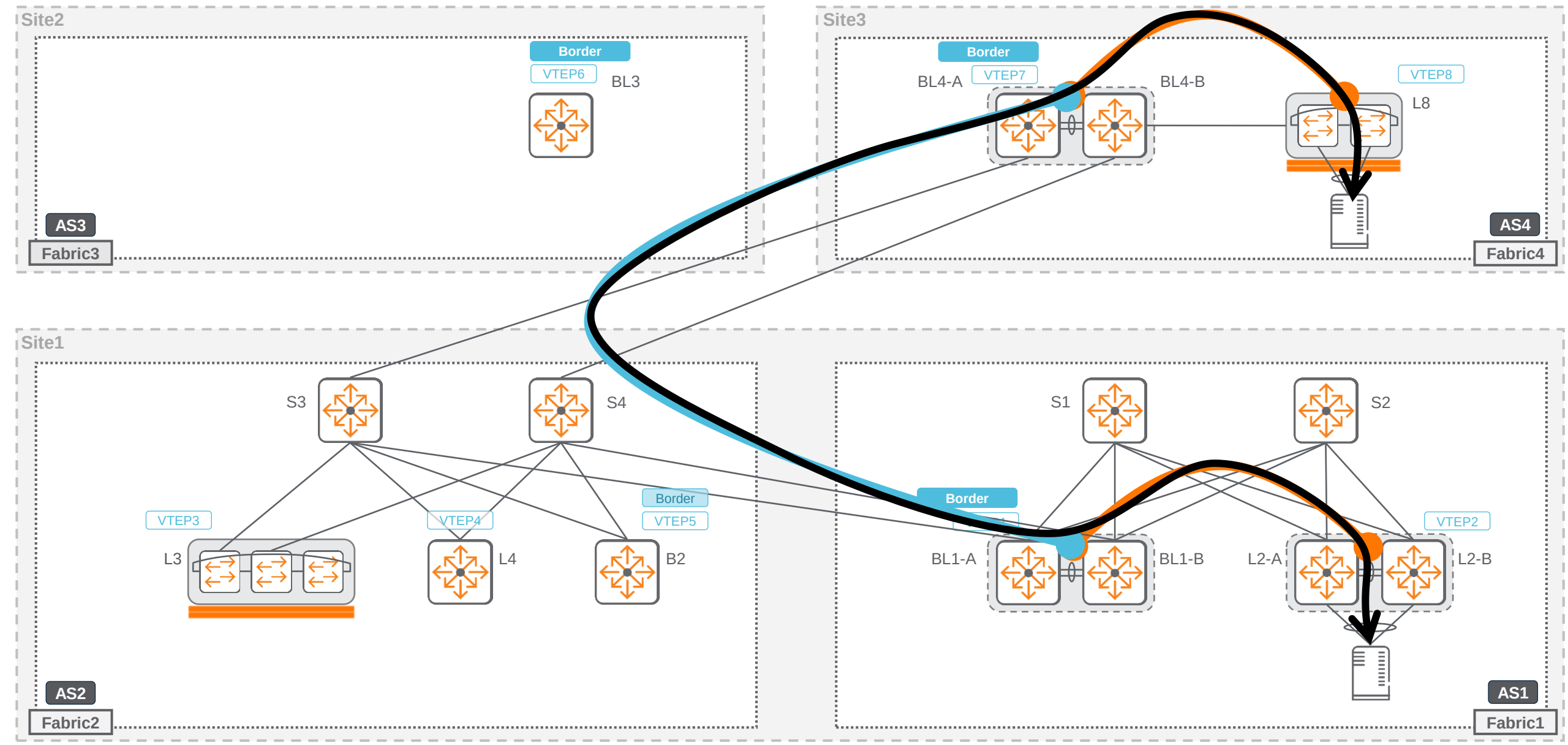
L2-A

L2-B

AS1

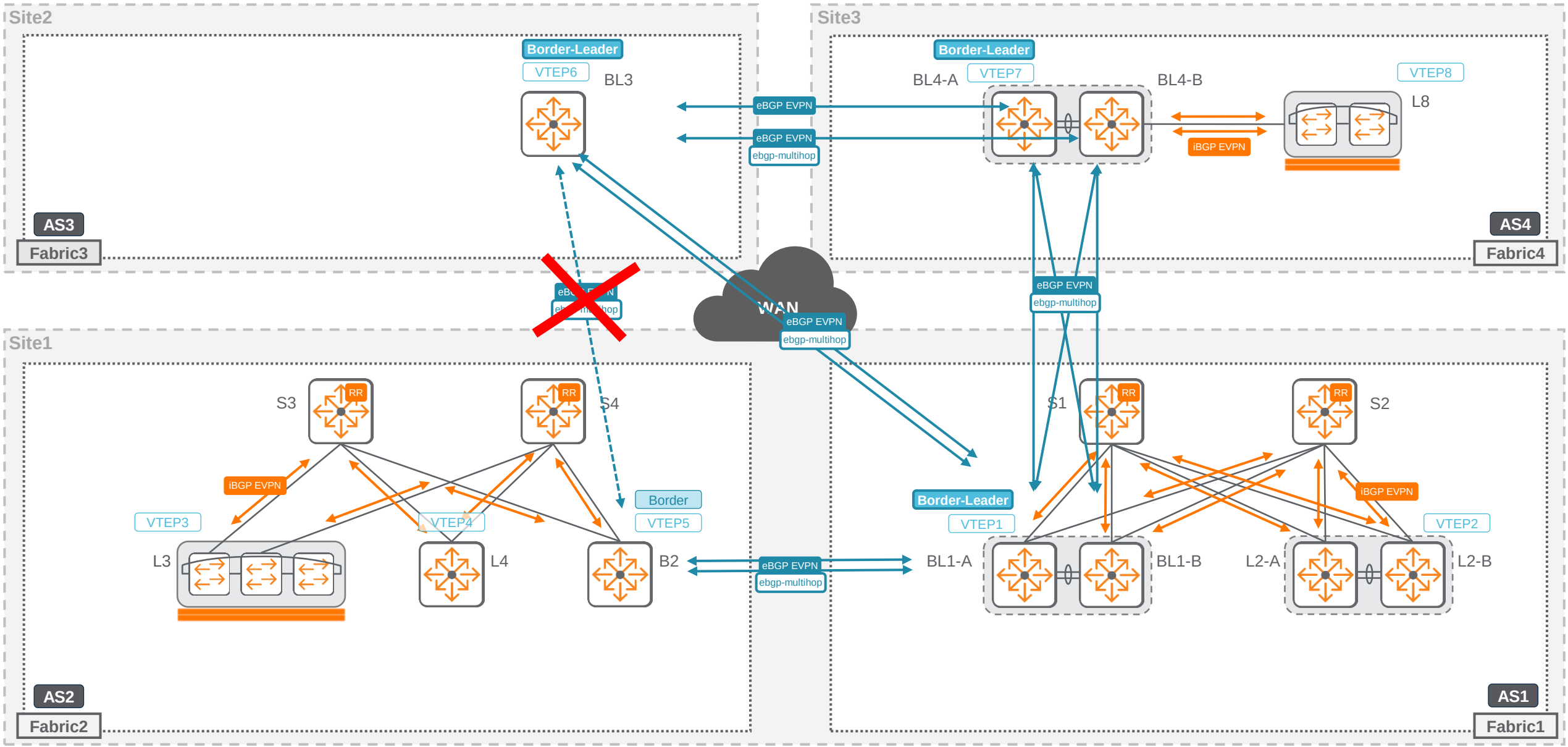
Fabric1

Overlay path over underlay network



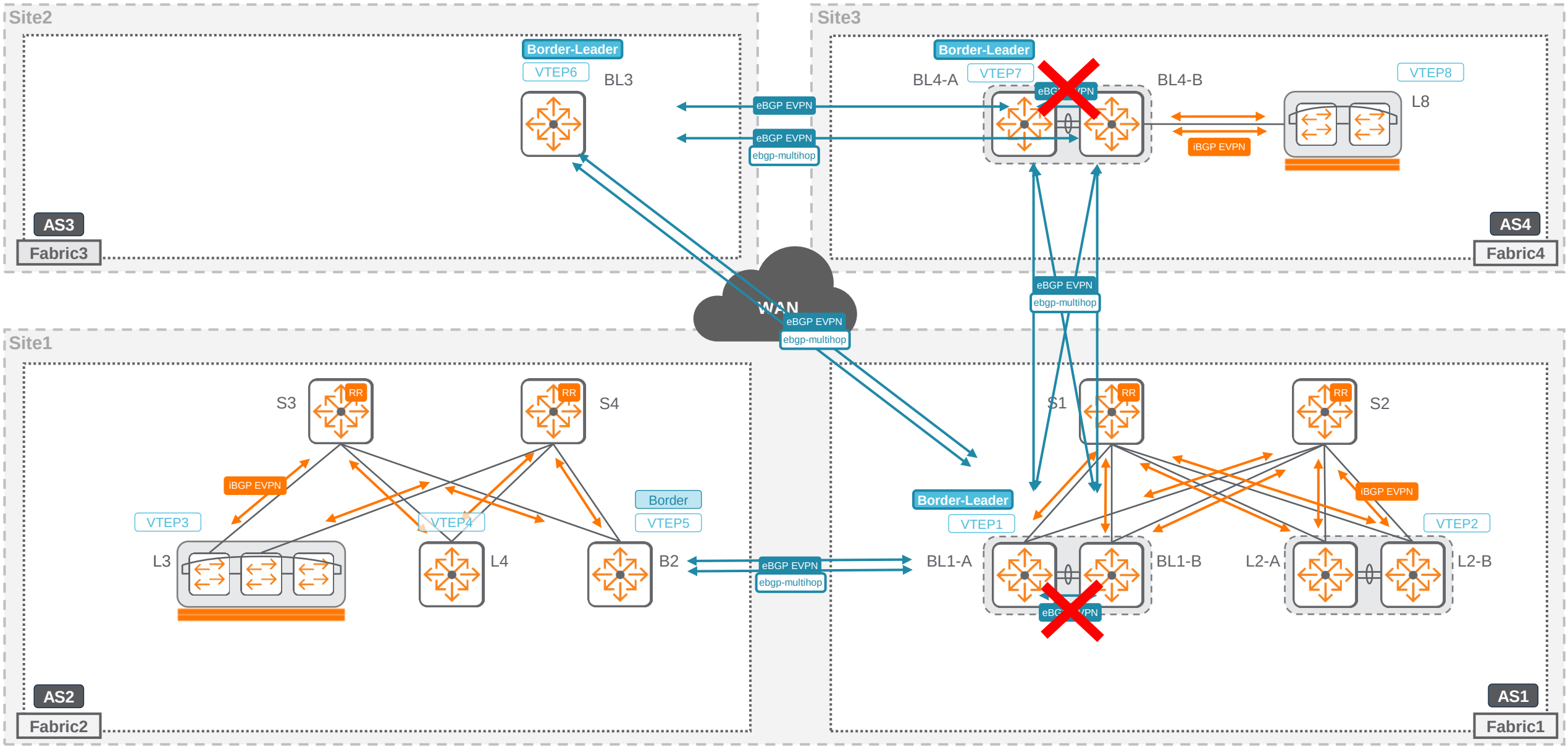
MP-BGP EVPN iBGP / eBGP sessions

Intra-Fabric Leaves/Spines, Border-Leader: Intra-Site B-to-BL, Inter-Site BL-to-BL



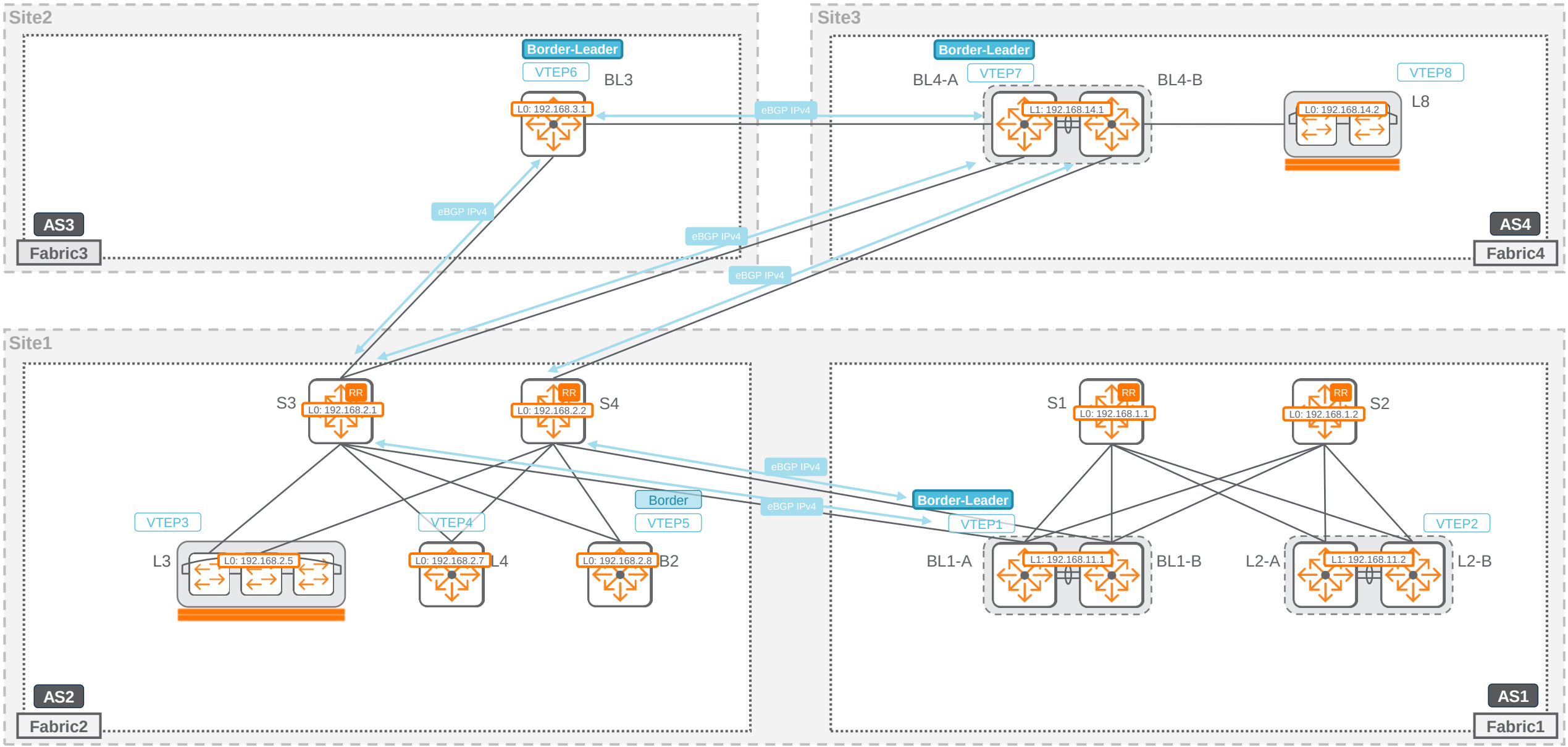
MP-BGP EVPN iBGP / eBGP sessions

No intra-VSX BGP EVPN session



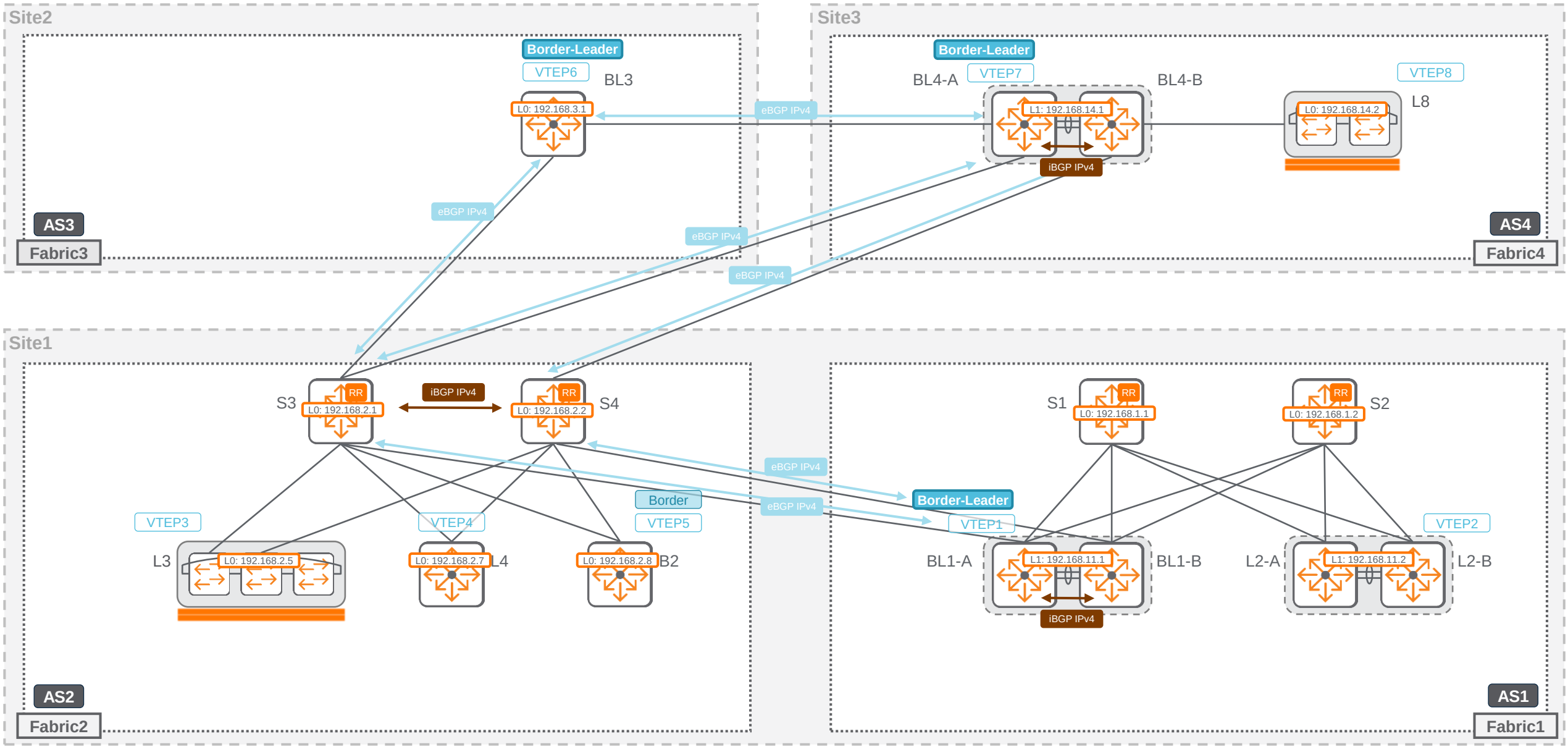
eBGP IPv4 Underlay Control-Plane

For VTEP IP address reachability (VSX VTEP L1, VSF/Standalone VTEP L0)

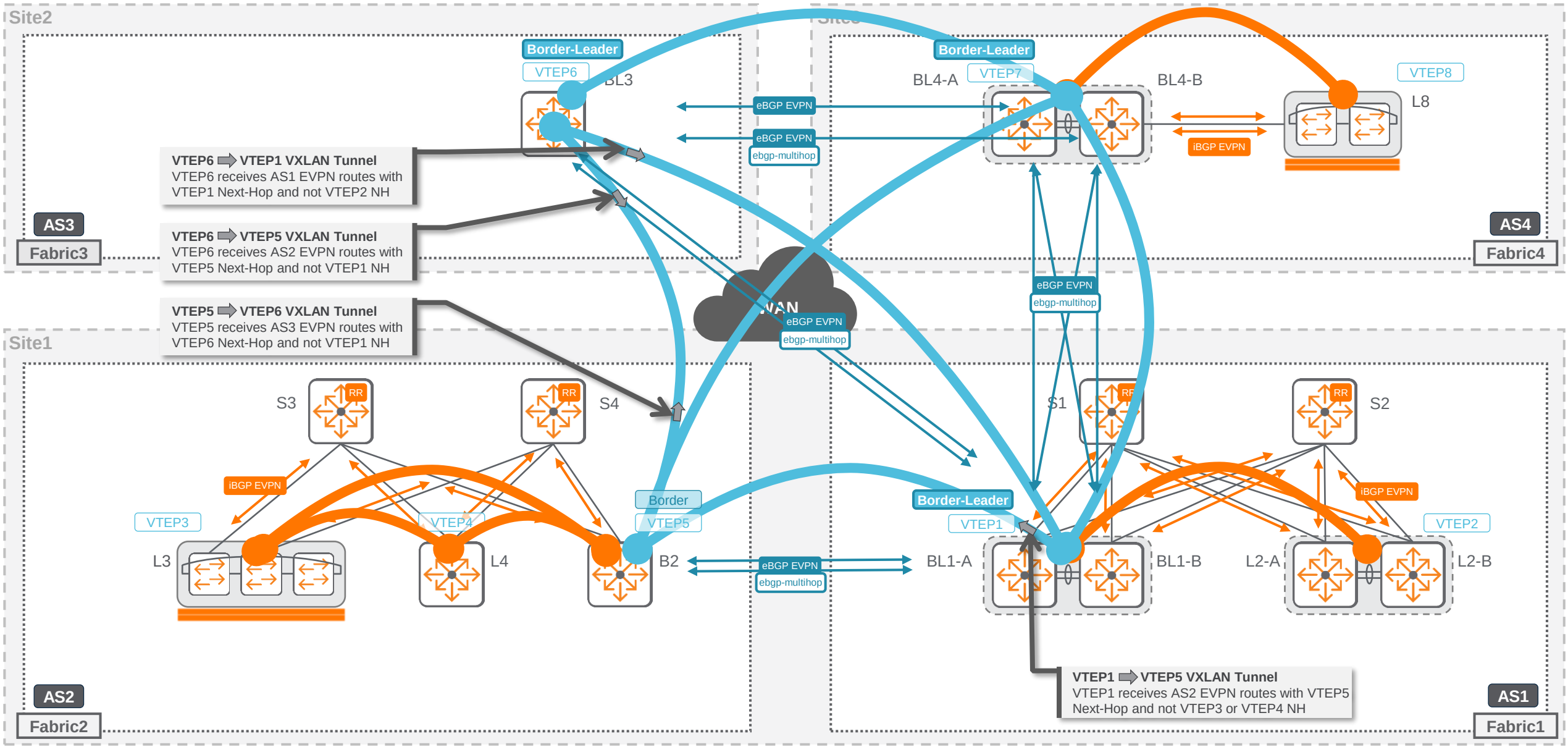


BGP IPv4 Underlay Control-Plane

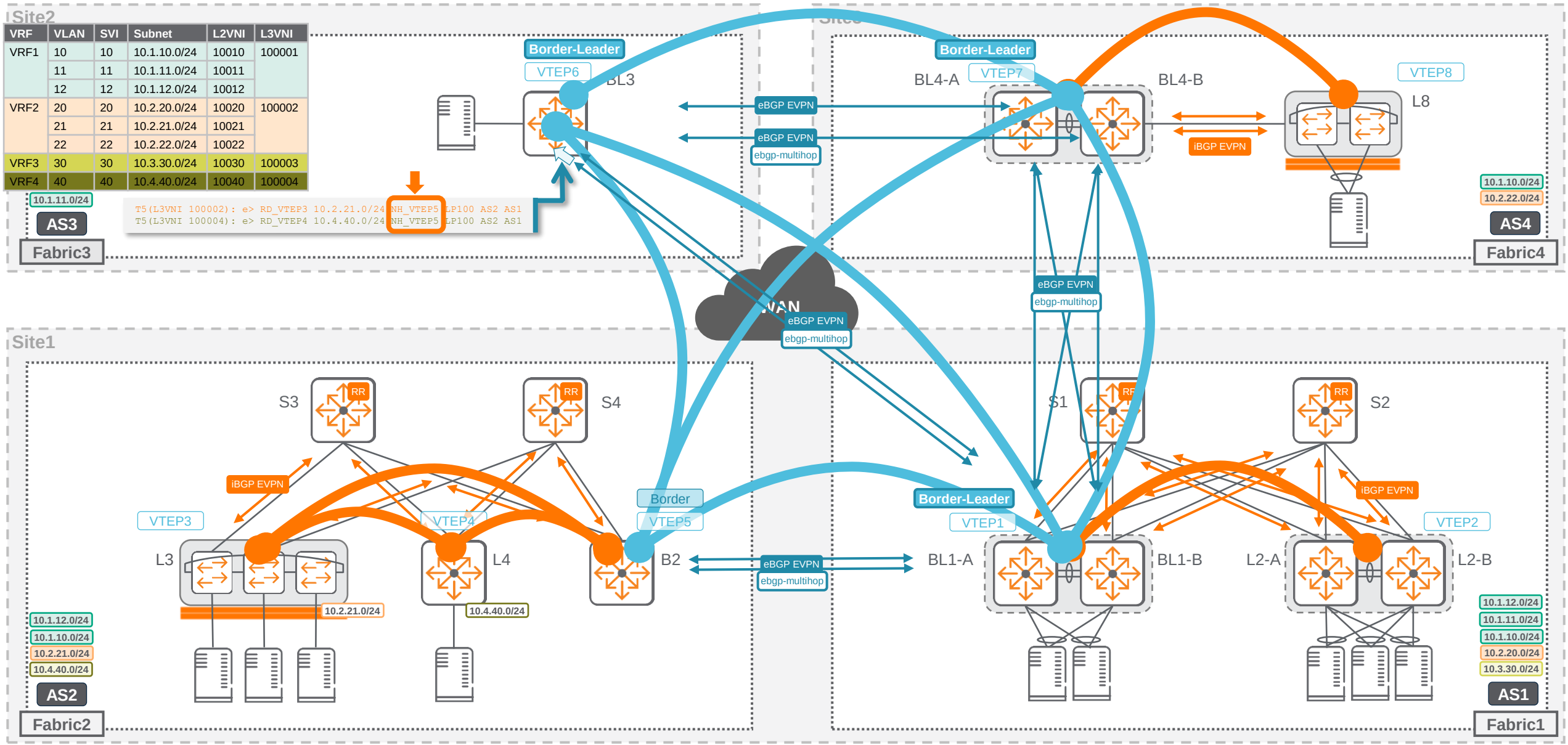
What about internal BGP sessions ? Decision depends on underlay topology and routing design



VTEP Next-Hop

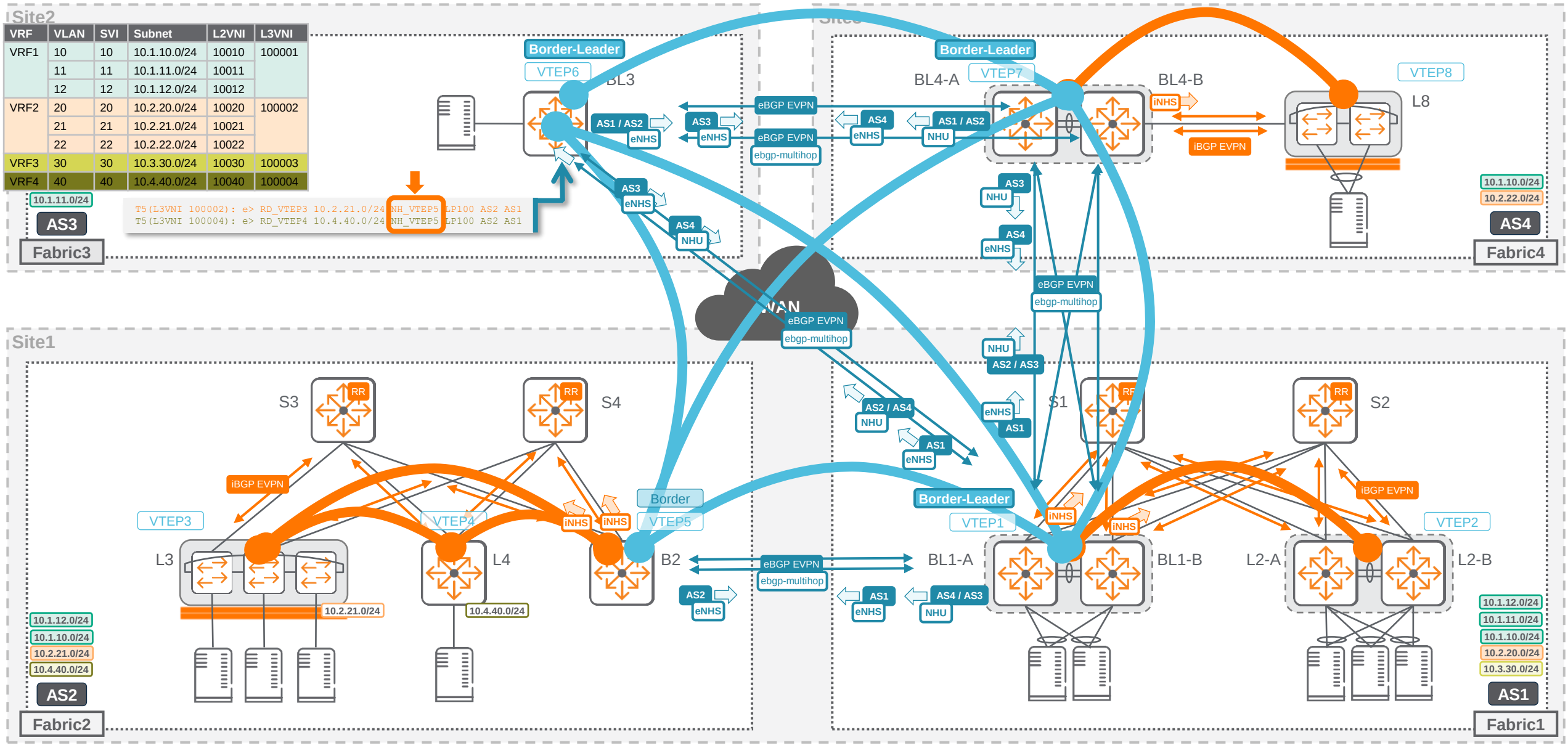


VTEP Next-Hop advertised from Borders and Border-Leaders



Control-Plane objective:
Next-Hop-Self (iNHS/eNHS) and Next-Hop-Unchanged (NHU)

VRF	VLAN	SVI	Subnet	L2VNI	L3VNI
VRF1	10	10	10.1.10.0/24	10010	100001
	11	11	10.1.11.0/24	10011	
	12	12	10.1.12.0/24	10012	
VRF2	20	20	10.2.20.0/24	10020	100002
	21	21	10.2.21.0/24	10021	
	22	22	10.2.22.0/24	10022	
VRF3	30	30	10.3.30.0/24	10030	100003
VRF4	40	40	10.4.40.0/24	10040	100004



EVPN VXLAN Multi-Fabric

AOS-CX 10.09 **Caveats** – **D**ata-**C**enter-**I**nterconnect Use-case validation

1. No support for:
   - L2/L3 multicast in the VXLAN overlay between fabrics
   - IPv6 in the VXLAN overlay between fabrics
   - VXLAN GBP between fabrics
2. Only iBGP is supported within each fabric (one ASN per fabric). eBGP inside a fabric is not yet supported.
3. External-BGP (eBGP) must be used between fabrics. No support for iBGP between Fabrics.
4. Super-Spine, as underlay hub for EVPN control-plane interconnection of fabrics, is not validated for 10.09.
5. Combined functions of Border VTEP and Spine on the same switch is not validated/supported in 10.09.
6. Only distributed L3 gateways (symmetric IRB) that support both L2/L3 connectivity or pure L2 VXLAN is supported between fabrics. Centralized L3 gateways are not supported.
7. One border-VTEP must be deployed per Fabric, and can not be shared between Fabrics.
8. Only a single border-VTEP (standalone or VSX logical VTEP) can be deployed per fabric. (No support of RT-1/4 for L2VNI-DCI.)
9. Overlay paths with two consecutive inter-fabric VXLAN tunnels are not supported:
   it does not work for L2VNI, and is not a validated scenario for L3VNI, although it might work.
10. Same L2VNI, L3VNI, VRF across fabrics must be used if in the scope of the Fabric.
11. As ARP suppression is supported only for L3 VTEP, ARP suppression can not happen for pure L2VNI DCI.
12. Border VTEPs should not have any static VXLAN tunnels configured as disabling VXLAN split-horizon between iBGP and eBGP broadcast groups will also disable split-horizon on static VXLAN tunnels.

EVPN VXLAN Multi-Fabric

AOS-CX 10.09 - **Additional Caveats**

1. Inter-Fabric L3VNI through Firewall (overlay aware) is not supported. (requiring bi-directional PBR).
2. Inter-Fabric L2VNI through active-active Firewall is not supported. (considered as a MAC move)
3. MTU on the underlay network should be at least 1550 bytes. (No MTU path discovery in the solution).
4. Caution on loopbacks routing to avoid uncontrolled cross-redistribution or routing loop.
5. eBGP multi-hop command for EVPN AF
6. Route-target (global versus local scope)
7. Outbound route-map not supported on specific peer of a peer-group (annoying for eBGP peer in multiple AFs)

```
8325-1(config-bgp-l2vpn-evpn)# neighbor 192.168.3.1  route-map test out
Neighbor is part of a peer-group, This configuration will not take effect
```

8. Existing solution for maintaining EVPN Next-Hop is subject for future improvement.

10.09 Platform Support

Next-hop-self and route-map support for EVPN address-family

Platform	4100 6000 6100	6200	6300	6400	8320	8325	8360	8400	10000	Simulator
EVPN AF route-map	No	No	**Yes**	**Yes**	No	**Yes**	**Yes**	**Yes**	**Yes**	No
EVPN AF next-hop-self	No	No	**Yes**	**Yes**	No	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**

10.09 Platform Support

Disabling VXLAN Split-Horizon with VXLAN Tunnels Bridging configuration

Platform	4100 6000 6100	6200	6300	6400	8320	8325	8360	8400	10000	Simulator
inter-vxlan-bridging-mode [static-dyn]	No	No	**Yes**	**Yes**	No	No	**Yes**	No	No	No
dyn-vxlan-tunnel-bridging-mode	No	No	No (POC*)	No (POC*)	No	**Yes**	**Yes**	No	**Yes**	No

POC*: Ask PLM for Proof-Of-Concept for future requests as the command is not available in CLI in 10.09 release.

- Multi-hop EVPN requires Dynamic VXLAN Tunnels. Static VXLAN is not in the scope of the border-VTEPs.
- Inter-vxlan-bridging-mode is not available for 8325/10000 to support different broadcast groups, as there is no static VXLAN for multi-hop VXLAN. This command is not required for EVPN VXLAN multi-fabric.
- If **dyn-vxlan-tunnel** is enabled then the traffic is bridged between EVPN and STATIC VXLAN tunnels (if static tunnels are configured).

```
8325-1(config-evpn)# dyn-vxlan-tunnel-bridging-mode  ibgp-ebgp
WARNING!! Enabling this command would bridge the traffic between EVPN and STATIC VxLAN tunnels.
```

10.09 Platform Support for EVPN-VXLAN multi-fabric

Platform	4100 6000 6100	6200	6300	6400	8320	8325	8360	8400	10000	Simulator
EVPN AF route-map	No	No	**Yes**	**Yes**	No	**Yes**	**Yes**	**Yes**	**Yes**	No
EVPN AF next-hop-self	No	No	**Yes**	**Yes**	No	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
dyn-vxlan-tunnel-bridging-mode	No	No	No	No	No	**Yes**	**Yes**	No	**Yes**	No
border-VTEP	No	No	No	No	No	**Yes**	**Yes**	No	No	No
border-leader-VTEP	No	No	No	No	No	**Yes**	**No***	No	No	No

* POC only
limited scale (3 Fabrics, 18 VTEPs per Fabric)

10.09 Platform Support

Validated Multi-Dimensional Scale

	Border Leader VTEP		Border VTEP	
	8325	8360 (POC only)	8325	8360
HW profile	Leaf	Agg-Leaf	Leaf	Agg-Leaf
VTEPs per Fabric (standalone or VSX logical VTEP pair)	64	16	64	64
Sites (Number of VSX border-leader VTEPs)	32	3		
Fabrics across sites (Number of VSX border-VTEPs, VXLAN full-mesh)	32	3		
L3 routes across all VRFs and all sites (including host routes)	22K	TBD (POC only)		
Overlay hosts (MAC / ARP) across sites Notes: - ND out of scope - Remote VTEPs share the same 32K UD limit for overlay neighbors - MD test-case: some VNIs are L2 only (MAC only)	Within 32K UD limit, tested MD use-case: MAC (local site): 6,000 ARP (local site): 6,000 MAC (remote site): 18,000 ARP (remote site): 10,000	TBD (POC only)		
VLANs local to the Fabric	512	512		
Stretched VLANs among all Fabrics	386	386		
VRFs shared among all Fabrics	16	16		

UD: Uni-Dimensional
MD: Multi-Dimensional

Configuration

Configuration

New Features Configuration Summary

VXLAN Tunnels Bridging

Data-plane

- For L2 connectivity across fabrics
- Must be used/applied on:
  - border-VTEPs
  - border-leader-VTEPs

```
evpn
    arp-suppression
    nd-suppression
    dyn-vxlan-tunnel-bridging-mode ibgp-ebgp
    vlan 10
        rd auto
        route-target export 1:10
        route-target import 1:10
```

EVPN next-hop-self + route-map

Control-plane

- Next-Hop-Self is required on all borders for peering to iBGP RRs.
- route-map is required on border-leader-VTEPs of sites featuring multiple fabrics

```
ip aspath-list fabric2 seq 10 permit _65002$
ip aspath-list fabric3 seq 10 permit _65003$
ip aspath-list fabricid seq 10 permit _<ASN>$
!
route-map to-borders permit seq 10
     match aspath-list fabric2
     set ip next-hop <remote-vtep-ip-fabric2>
route-map to-borders permit seq 20
     match aspath-list fabric3
     set ip next-hop <remote-vtep-ip-fabric3>
route-map to-borders permit seq <n>
     match aspath-list fabricid
     set ip next-hop <remote-vtep-ip-fabricid>
route-map to-borders permit seq 1000
!
router bgp 65001
…
address-family l2vpn evpn
   neighbor borders route-map to-borders out
   neighbor borders send-community both
   neighbor spine-RR next-hop-self
```

New commands in EVPN AF

VTEP next-hop IP address rewrite

Route-MAC consideration for L3VNI

- Rewriting VTEP next-hop IP induces rewriting router-MAC for EVPN NLRI attribute for L3VNI routing.
- If not performed, the VXLAN source IP and associated virtual-MAC might not be consistent in the BGP RIB cache used to build the adjacency-RIB to reach the remote VTEP:

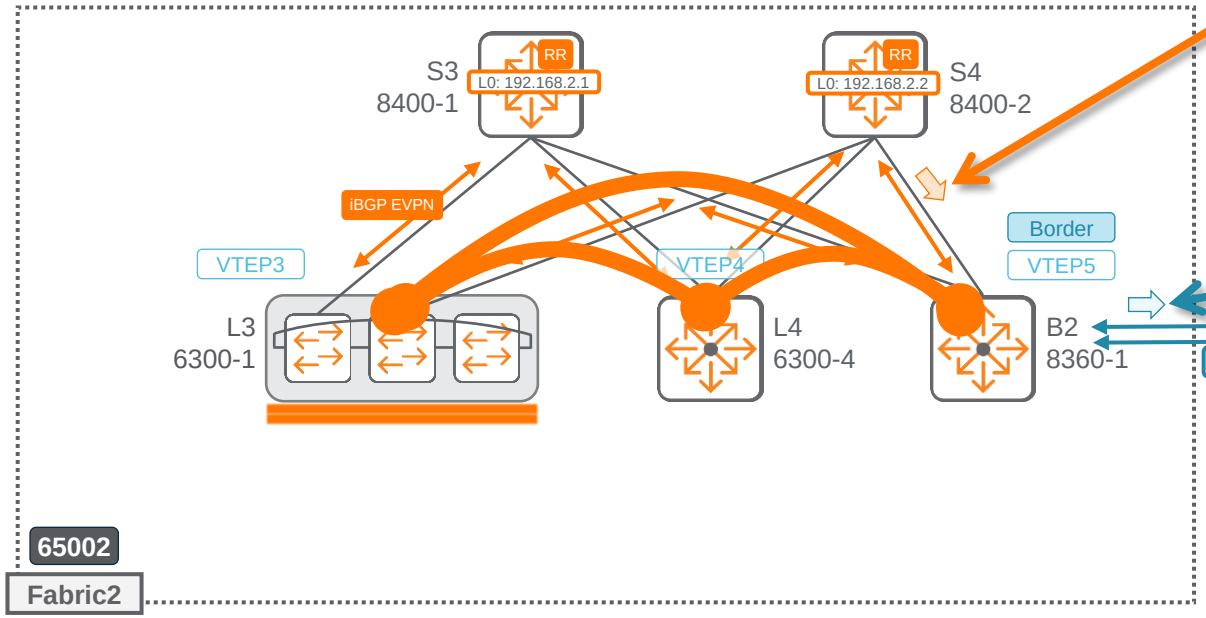
```
8360-1# show evpn vtep-neighbor all-vrfs
VTEP-IP         L3VNI    MAC               VRF                                  State
-------------------------------------------------------------------------------------
192.168.2.7     100001   02:00:00:00:04:00 VRF1                                 Up
192.168.10.3    100002   02:00:00:00:01:00 VRF2                                 Up
192.168.2.5     100002   02:00:00:00:03:00 VRF2                                 Up
192.168.40.1    100002   02:00:00:00:08:00 VRF2                                 Up
192.168.40.1    100001   02:00:00:00:07:00 VRF1                                 Up
192.168.2.5     100001   02:00:00:00:03:00 VRF1                                 Up
192.168.3.1     100001   02:00:00:00:06:00 VRF1                                 Up
192.168.10.3    100001   02:00:00:00:01:00 VRF1                                 Up
```

- This table is constructed based on the last received NLRI L3VNI update, indicating, per VRF, the router-MAC used per remote VTEP IP. (any type-2 or type-5 route update coming from the remote border VTEP)
- For L3VNI, this would lead to packet drops as the remote VTEP IP would receive an inner packet with an invalid destination MAC address.
- **The next-hop-self feature (eBGP and iBGP), in addition to replacing the next-hop IP of EVPN routes with self-IP, replaces the router-mac extended community of EVPN routes with self-MAC.**

EVPN Type-3 routes

Optimization: avoiding unnecessary VXLAN tunnels

- There is no need for re-advertising EVPN Type-3 routes that are received from intra-fabric VTEPs to other fabrics.
- Similarly, there is no need for re-advertising EVPN Type-3 routes that are received from external-fabric border-VTEPs into the intra-fabric VTEPs.
- Such optimization process is triggered through eBGP **next-hop-self**, and not through route-map (in 10.09).




```
8400-2# sh bgp l2 e neighbors 192.168.2.8 advertised-routes route-type 3
...

     Network                                          Nexthop         Metric     LocPrf     Weight     Path
-------------------------------------------------------------------------------------------------------------------
Route Distinguisher: 192.168.2.5:10       (L2VNI 10010)
*>e [3]:[0]:[192.168.2.5]                             192.168.2.5     0          100        0          ?

Route Distinguisher: 192.168.2.7:12       (L2VNI 10012)
*>e [3]:[0]:[192.168.2.7]                             192.168.2.7     0          100        0          ?

Route Distinguisher: 192.168.2.5:21       (L2VNI 10021)
*>e [3]:[0]:[192.168.2.5]                             192.168.2.5     0          100        0          ?
Total number of entries 3
```

```
8360-1# sh bgp l2 e neighbors 192.168.1.3 advertised-routes route-type 3
...

     Network                                          Nexthop         Metric     LocPrf     Weight     Path
-------------------------------------------------------------------------------------------------------------------
Route Distinguisher: 192.168.2.8:10       (L2VNI 10010)
*>e [3]:[0]:[192.168.2.8]                             192.168.2.8     0          0          0          65002
?

Route Distinguisher: 192.168.3.1:11       (L2VNI 10011)
*>e [3]:[0]:[192.168.3.1]                             192.168.2.8     0          0          0          65002
65001 65003 ?
Total number of entries 2
```

Undesired Type-3 routes

Without next-hop-self

```
8360-1# show bgp l2vpn evpn route-type 3
Status codes: s suppressed, d damped, h history, * valid, > best, = multipath,
              i internal, e external S Stale, R Removed, a additional-paths
Origin codes: i - IGP, e - EGP, ? - incomplete

EVPN Route-Type 2 prefix: [2]:[ESI]:[EthTag]:[MAC]:[OrigIP]
EVPN Route-Type 3 prefix: [3]:[EthTag]:[OrigIP]
EVPN Route-Type 5 prefix: [5]:[ESI]:[EthTag]:[IPAddrLen]:[IPAddr]
VRF : default
Local Router-ID 192.168.2.8

     Network                                          Nexthop         Metric     LocPrf     Weight   Path
-------------------------------------------------------------------------------------------------------------
Route Distinguisher: 192.168.10.3:10      (L2VNI 10010)
* e [3]:[0]:[192.168.10.3]                           192.168.10.3    0          100        0       65001 ?
*>e [3]:[0]:[192.168.10.3]                           192.168.10.3    0          100        0       65001 ?

Route Distinguisher: 192.168.10.5:10      (L2VNI 10010)
* e [3]:[0]:[192.168.10.5]                           192.168.10.3    0          100        0       65001 ?
*>e [3]:[0]:[192.168.10.5]                           192.168.10.3    0          100        0       65001 ?

Route Distinguisher: 192.168.2.5:10       (L2VNI 10010)
*>i [3]:[0]:[192.168.2.5]                            192.168.2.5     0          100        0       ?
* i [3]:[0]:[192.168.2.5]                            192.168.2.5     0          100        0       ?

Route Distinguisher: 192.168.2.8:10       (L2VNI 10010)
*>  [3]:[0]:[192.168.2.8]                            192.168.2.8     0          100        0       ?

Route Distinguisher: 192.168.4.3:10       (L2VNI 10010)
* e [3]:[0]:[192.168.4.3]                           192.168.40.1    0          100        0       65001 65004 ?
*>e [3]:[0]:[192.168.4.3]                           192.168.40.1    0          100        0       65001 65004 ?

Route Distinguisher: 192.168.40.1:10      (L2VNI 10010)
* e [3]:[0]:[192.168.40.1]                           192.168.40.1    0          100        0       65001 65004 ?
*>e [3]:[0]:[192.168.40.1]                           192.168.40.1    0          100        0       65001 65004 ?

Route Distinguisher: 192.168.10.3:11      (L2VNI 10011)
* e [3]:[0]:[192.168.10.3]                           192.168.10.3    0          100        0       65001 ?
*>e [3]:[0]:[192.168.10.3]                           192.168.10.3    0          100        0       65001 ?

Route Distinguisher: 192.168.3.1:11       (L2VNI 10011)
* e [3]:[0]:[192.168.3.1]                            192.168.3.1     0          100        0       65001 65003 ?
*>e [3]:[0]:[192.168.3.1]                            192.168.3.1     0          100        0       65001 65003 ?

Route Distinguisher: 192.168.10.5:12      (L2VNI 10012)
* e [3]:[0]:[192.168.10.5]                           192.168.10.3    0          100        0       65001 ?
*>e [3]:[0]:[192.168.10.5]                           192.168.10.3    0          100        0       65001 ?
```

```
Route Distinguisher: 192.168.2.7:12       (L2VNI 10012)
*>i [3]:[0]:[192.168.2.7]                            192.168.2.7     0          100        0       ?
* i [3]:[0]:[192.168.2.7]                            192.168.2.7     0          100        0       ?

Route Distinguisher: 192.168.10.3:20      (L2VNI 10020)
* e [3]:[0]:[192.168.10.3]                           192.168.10.3    0          100        0       65001 ?
*>e [3]:[0]:[192.168.10.3]                           192.168.10.3    0          100        0       65001 ?

Route Distinguisher: 192.168.10.5:20      (L2VNI 10020)
* e [3]:[0]:[192.168.10.5]                           192.168.10.3    0          100        0       65001 ?
*>e [3]:[0]:[192.168.10.5]                           192.168.10.3    0          100        0       65001 ?

Route Distinguisher: 192.168.2.5:21       (L2VNI 10021)
*>i [3]:[0]:[192.168.2.5]                            192.168.2.5     0          100        0       ?
* i [3]:[0]:[192.168.2.5]                            192.168.2.5     0          100        0       ?

Route Distinguisher: 192.168.4.3:22       (L2VNI 10022)
* e [3]:[0]:[192.168.4.3]                           192.168.40.1    0          100        0       65001 65004 ?
*>e [3]:[0]:[192.168.4.3]                           192.168.40.1    0          100        0       65001 65004 ?
Total number of entries 27
```

Optimized Type-3 routes

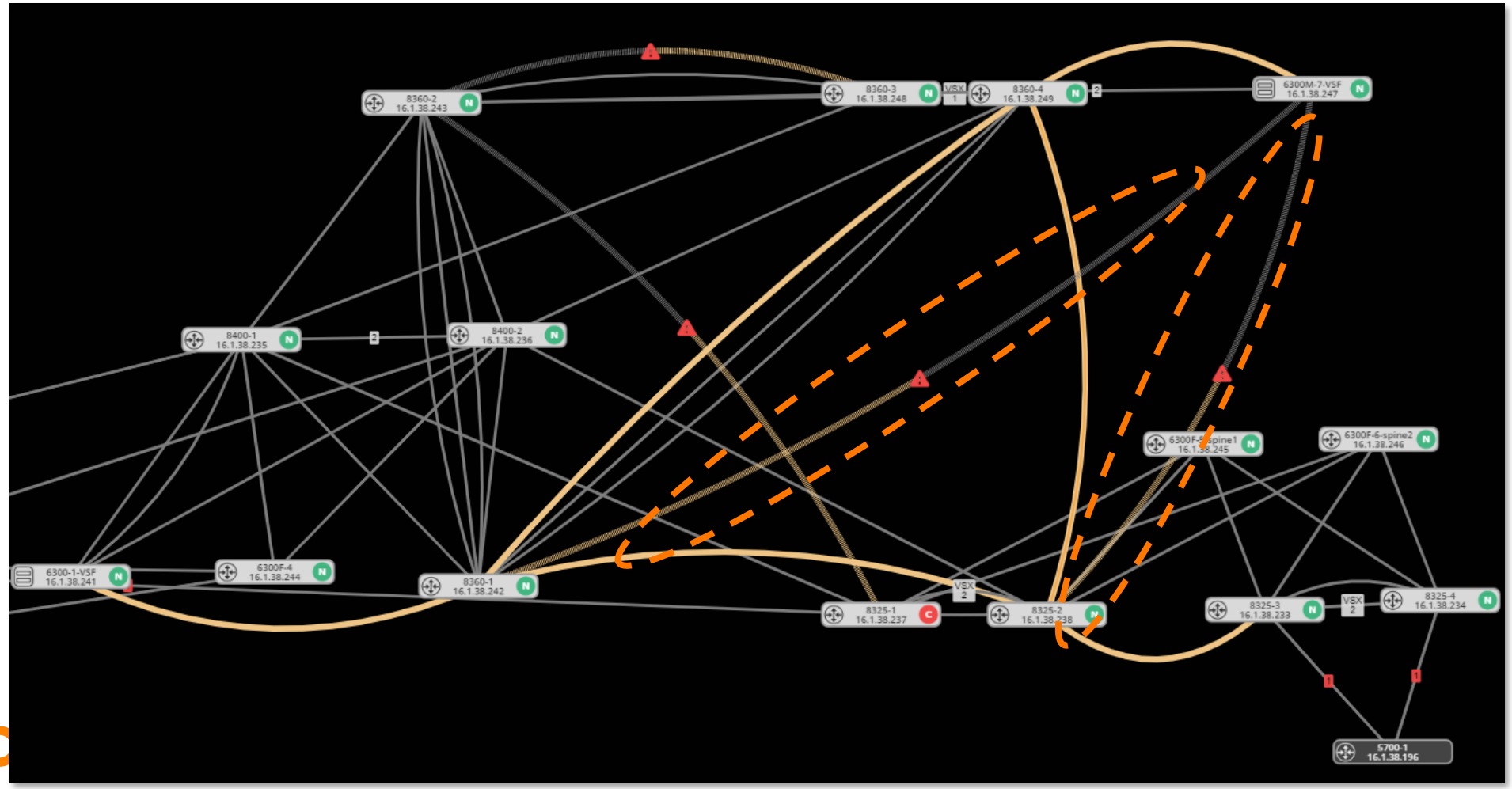
Through eBGP next-hop-self

```
8360-1# sh bgp l2 e route-type 3
Status codes: s suppressed, d damped, h history, * valid, > best, = multipath,
              i internal, e external S Stale, R Removed, a additional-paths
Origin codes: i - IGP, e - EGP, ? - incomplete

EVPN Route-Type 2 prefix: [2]:[ESI]:[EthTag]:[MAC]:[OrigIP]
EVPN Route-Type 3 prefix: [3]:[EthTag]:[OrigIP]
EVPN Route-Type 5 prefix: [5]:[ESI]:[EthTag]:[IPAddrLen]:[IPAddr]
VRF : default
Local Router-ID 192.168.2.8

     Network                                          Nexthop         Metric     LocPrf     Weight   Path
-------------------------------------------------------------------------------------------------------------
Route Distinguisher: 192.168.10.3:10      (L2VNI 10010)
*>e [3]:[0]:[192.168.10.3]                            192.168.10.3    0          100        0       65001 ?
* e [3]:[0]:[192.168.10.3]                            192.168.10.3    0          100        0       65001 ?

Route Distinguisher: 192.168.2.5:10       (L2VNI 10010)
*>i [3]:[0]:[192.168.2.5]                             192.168.2.5     0          100        0       ?
* i [3]:[0]:[192.168.2.5]                             192.168.2.5     0          100        0       ?

Route Distinguisher: 192.168.2.8:10       (L2VNI 10010)
*>  [3]:[0]:[192.168.2.8]                             192.168.2.8     0          100        0       ?

Route Distinguisher: 192.168.40.1:10      (L2VNI 10010)
*>e [3]:[0]:[192.168.40.1]                            192.168.40.1    0          100        0       65001 65004 ?
* e [3]:[0]:[192.168.40.1]                            192.168.40.1    0          100        0       65001 65004 ?

Route Distinguisher: 192.168.10.3:11      (L2VNI 10011)
*>e [3]:[0]:[192.168.10.3]                            192.168.10.3    0          100        0       65001 ?
* e [3]:[0]:[192.168.10.3]                            192.168.10.3    0          100        0       65001 ?

Route Distinguisher: 192.168.3.1:11       (L2VNI 10011)
* e [3]:[0]:[192.168.3.1]                             192.168.3.1     0          100        0       65001 65003 ?
*>e [3]:[0]:[192.168.3.1]                             192.168.3.1     0          100        0       65001 65003 ?

Route Distinguisher: 192.168.2.7:12       (L2VNI 10012)
*>i [3]:[0]:[192.168.2.7]                             192.168.2.7     0          100        0       ?
* i [3]:[0]:[192.168.2.7]                             192.168.2.7     0          100        0       ?

Route Distinguisher: 192.168.10.3:20      (L2VNI 10020)
*>e [3]:[0]:[192.168.10.3]                            192.168.10.3    0          100        0       65001 ?
* e [3]:[0]:[192.168.10.3]                            192.168.10.3    0          100        0       65001 ?

Route Distinguisher: 192.168.2.5:21       (L2VNI 10021)
*>i [3]:[0]:[192.168.2.5]                             192.168.2.5     0          100        0       ?
* i [3]:[0]:[192.168.2.5]                             192.168.2.5     0          100        0       ?
Total number of entries 17
```

Undesired VXLAN tunnels

Type-3 routes create dynamic Tunnels

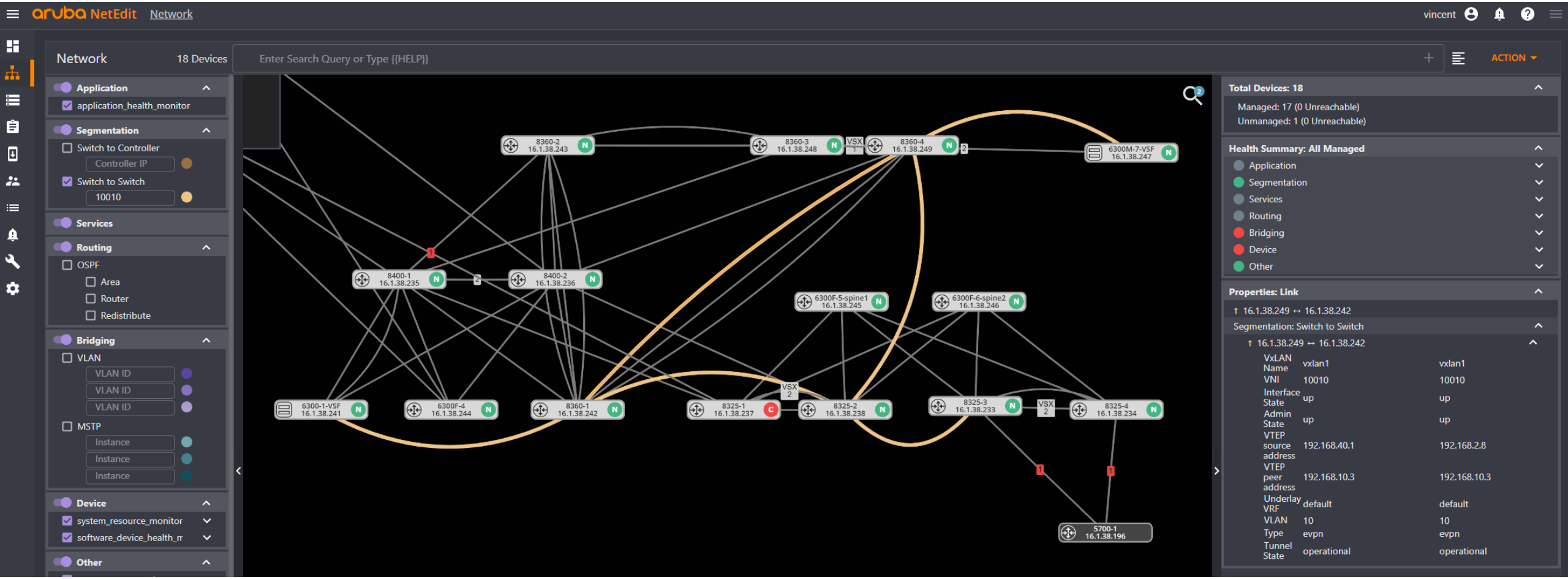
```
8360-1# sh interface vxlan vteps
Source           Destination      Origin       Status              VNI       Routing   VLAN  VRF
---------------- ---------------- ------------ ------------------- --------- --------- ----- ----------
192.168.2.8      192.168.2.5      evpn         operational         10010     disabled  10    --
192.168.2.8      192.168.2.5      evpn         operational         100001    enabled   --    VRF1
192.168.2.8      192.168.2.5      evpn         operational         100002    enabled   --    VRF2
192.168.2.8      192.168.2.7      evpn         operational         100001    enabled   --    VRF1
192.168.2.8      192.168.3.1      evpn         operational         100001    enabled   --    VRF1
192.168.2.8      192.168.4.3      evpn         operational         10010     disabled  10    --
192.168.2.8      192.168.10.3     evpn         operational         10010     disabled  10    --
192.168.2.8      192.168.10.3     evpn         operational         100001    enabled   --    VRF1
192.168.2.8      192.168.10.3     evpn         operational         100002    enabled   --    VRF2
192.168.2.8      192.168.40.1     evpn         operational         10010     disabled  10    --
192.168.2.8      192.168.40.1     evpn         operational         100001    enabled   --    VRF1
192.168.2.8      192.168.40.1     evpn         operational         100002    enabled   --    VRF2
```



Optimized VXLAN tunnels

Thanks to Type-3 optimization

```
8360-1# sh interface vxlan vteps
Source           Destination      Origin       Status              VNI       Routing   VLAN  VRF
---------------- ---------------- ------------ ------------------- --------- --------- ----- ----------
192.168.2.8      192.168.2.5      evpn         operational         10010     disabled  10    --
192.168.2.8      192.168.2.5      evpn         operational         100001    enabled   --    VRF1
192.168.2.8      192.168.2.5      evpn         operational         100002    enabled   --    VRF2
192.168.2.8      192.168.2.7      evpn         operational         100001    enabled   --    VRF1
192.168.2.8      192.168.3.1      evpn         operational         100001    enabled   --    VRF1
192.168.2.8      192.168.10.3     evpn         operational         10010     disabled  10    --
192.168.2.8      192.168.10.3     evpn         operational         100001    enabled   --    VRF1
192.168.2.8      192.168.10.3     evpn         operational         100002    enabled   --    VRF2
192.168.2.8      192.168.40.1     evpn         operational         10010     disabled  10    --
192.168.2.8      192.168.40.1     evpn         operational         100001    enabled   --    VRF1
192.168.2.8      192.168.40.1     evpn         operational         100002    enabled   --    VRF2
```



EVPN route-map

Applied **on border-leader-VTEP** in multi-fabric site

Method#1 (10.09)

Sourcing ASN & Remote VTEP IP

```
ip aspath-list fabric2 seq 10 permit _65002$
ip aspath-list fabric3 seq 10 permit _65003$
ip aspath-list fabric4 seq 10 permit _65004$
ip aspath-list fabricid seq 10 permit _<ASN>$
!
route-map to-borders permit seq 10
     match aspath-list fabric2
     set ip next-hop <remote-vtep-ip-fabric2>
route-map to-borders permit seq 20
     match aspath-list fabric3
     set ip next-hop <remote-vtep-ip-fabric3>
route-map to-borders permit seq 30
     match aspath-list fabric4
     set ip next-hop <remote-vtep-ip-fabric4>
route-map to-borders permit seq <n>
     match aspath-list fabricid
     set ip next-hop <remote-vtep-ip-fabricid>
route-map to-borders permit seq 1000
!
router bgp <ASN>
…
address-family l2vpn evpn
   neighbor borders route-map to-borders out
   neighbor borders send-community both
```

- Configuration intensive when number of sites increases.

Method#2 (not supported)

Local ASN & local VTEP IP

```
ip aspath-list local-AS seq 10 permit ^$
!
route-map to-borders permit seq 10
   match aspath-list local-AS
   set ip next-hop <local-vtep-ip>
route-map to-borders permit seq 100
!
router bgp <ASN>
…
address-family l2vpn evpn
   neighbor borders next-hop-unchanged
   neighbor borders route-map to-borders out
```

- Router-MAC replacement is achieved through:
  - regular eBGP peering
  - next-hop-self iBGP command
- This processing is not yet supported through route-map.

Method#3 (not supported)

Based on local community

```
ip community-list standard local-fabrics seq 10 permit 1:0
!
route-map from-local-fabrics permit seq 10
   set community 1:0 additive
!
route-map to-borders permit seq 10
   match community-list local-fabrics
   set ip next-hop-unchanged
   set community 1:0 delete
!
router bgp <ASN>
…
address-family l2vpn evpn
   neighbor borders send-community both
   neighbor borders route-map to-borders out
   neighbor <local-fabric-vtep> route-map from-local-fabrics in
```

- Need support of set clause for next-hop-unchanged (not yet supported)
- No change on existing route-MAC rewrite process.
- High flexibility

EVPN route-map for border-leader in 10.09

Applied **on border-leader-VTEP** in multi-fabric site

Method#1 (10.09)

Sourcing ASN & Remote VTEP IP

```
ip aspath-list fabric2 seq 10 permit _65002$
ip aspath-list fabric3 seq 10 permit _65003$
ip aspath-list fabric4 seq 10 permit _65004$
ip aspath-list fabricid seq 10 permit _<ASN>$
!
route-map to-borders permit seq 10
     match aspath-list fabric2
     set ip next-hop <remote-vtep-ip-fabric2>
route-map to-borders permit seq 20
     match aspath-list fabric3
     set ip next-hop <remote-vtep-ip-fabric3>
route-map to-borders permit seq 30
     match aspath-list fabric4
     set ip next-hop <remote-vtep-ip-fabric4>
route-map to-borders permit seq <n>
     match aspath-list fabricid
     set ip next-hop <remote-vtep-ip-fabricid>
route-map to-borders permit seq 1000
!
router bgp <ASN>
…
address-family l2vpn evpn
   neighbor borders route-map to-borders out
   neighbor borders send-community both
```

- Configuration intensive when number of sites increases.

Method#2 (not supported)

Local ASN & local VTEP IP

```
ip aspath-list local-AS seq 10 permit _$
!
route-map to-borders permit seq 10
   match aspath-list local-AS
   set ip next-hop <local-vtep-ip>
route-map to-borders permit seq 1000
!
router bgp <ASN>
…
address-family l2vpn evpn
   neighbor borders next-hop-unchanged
   neighbor borders route-map to-borders out
```

- Router-MAC replacement is achieved through:
  - regular eBGP peering
  - next-hop-self iBGP command
- This processing is not yet supported through route-map.

Method#3 (not supported)

Based on local community

```
ip community-list standard local-fabrics seq 10 permit 1:0
!
route-map from-local-fabrics permit seq 10
   set community 1:0 additive
!
route-map to-borders permit seq 10
   match community-list local-fabrics
   set ip next-hop-unchanged
   set community 1:0 delete
!
router bgp <ASN>
…
address-family l2vpn evpn
   neighbor borders send-community both
   neighbor borders route-map to-borders out
   neighbor <local-fabric-vtep> route-map from-local-fabrics in
```

- Need support of set clause for next-hop-unchanged (not yet supported)
- No change on existing route-MAC rewrite process.
- High flexibility

AS path Filtering

Reminder on Regular Expressions for AS-path filter

.	Match any single character including space
*	Match 0 or more sequences of the pattern
+	Match 1 or more sequences of the pattern
?	Match 0 or 1 occurrence of the pattern
^	Match the beginning of the input string
$	Match the End of the input string
\c	Match the specific character ‘c’
[]	Designate a range for a single character
()	Designate a pattern of multiple characters
[^]	Matches every character except the ones inside brackets.
-	Dash sign separates value within a range expression [1-9]
_	Underscore match the following characters: comma, left or right brace or parenthesis

AS path Filtering

Examples

```
ip aspath-list <name> seq 10 permit _65001_
```

Match any AS-path containing AS 65001

```
ip aspath-list <name> seq 10 permit _65001$
```

Match any AS-path ending with AS 65001 (routes sourced from AS 65001)

Used for 10.09 EVPN VXLAN Multi-Fabric

```
ip aspath-list <name> seq 10 permit ^65001.*
```

Match any AS-path advertised by directly attached peer AS 65001 and AS's directly attached to it.

```
ip aspath-list <name> seq 10 permit ^65001( [0-9]+)?$
```

Match any AS-path advertised by directly attached peer AS 65001 and AS's directly attached to it or one AS away.

```
ip aspath-list <name> seq 10 permit (_65001_|_65001$)
```

Match containing AS 65001 either in the middle or at the end.

Best Practices

Multi-hop VXLAN

Reminder on Existing Best Practices

- Ensure sufficient MTU (9K MTU recommended) for VXLAN overhead, especially across WAN.
- If VSX logical VTEP is used for HA, both switches should use the same “virtual-mac” (router MAC). The same value than VSX system-MAC can be use for ease of provisioning:

```
vsx
    system-mac 02:00:00:00:01:00
…
virtual-mac 02:00:00:00:01:00
```

- For distributed L3 gateway deployments, a unique “**virtual-mac**” should be used on every VTEP, “Border VTEP”, “Border leader VTEP”.
- Manual EVPN route targets must be used as eBGP EVPN is used between fabrics.
  Auto EVPN route targets can only be used for iBGP EVPN deployments.
  (RT auto option would require to know the calculated auto RT value on remote VTEPs which is very cumbersome in operations.)
- VSX-sync usage: **vsx-sync route-map** is useful, as it synchronizes from the VSX primary the aspath-list, the prefix list, the community list and the route-map configuration (except ip next-hop).
  Warning: “set ip next-hop” is not synced from the VSX primary to the VSX secondary (by feature) and must be manually configured on the VSX secondary with same IP address than on the VSX primary.

Multi-Fabric EVPN Solution

Best Practices Summary

- Routes optimization: aggregation, routes filtering.
  L2VNI routes removal implication on host entries in VRF routing table.
- Prefer Common RD against Distinct RD
- Existing solution for maintaining EVPN Next-Hop is subject to improvement.
- BGP communities: use both (extended + standard)
- DHCP-relay VRF option:
  - Prefer DHCP-sourcing in the DHCP-server VRF
  - Rather than route-leaking between tenant VRF and DHCP-server VRF
- IPAM: tenant VRF loopbacks, tenant VRF IP interconnectivity

Route-Target

Global / Local

```
vrf VRF1
    rd 192.168.1.3:1
    route-target export 65001:1 evpn
    route-target import 65001:1 evpn
    route-target export 1:1 evpn
    route-target import 1:1 evpn
vrf VRF2
    rd 192.168.1.3:2
    route-target export 65001:2 evpn
    route-target import 65001:2 evpn
    route-target export 1:2 evpn
    route-target import 1:2 evpn
```

local filter (route-target) per fabric

global filter (route-target) common to all fabrics

local filter (route-target) per fabric

global filter (route-target) common to all fabrics

Global scope

- Defined per VRF.
- Not all VRFs might require a global scope. Some network may not have a global scope.
- Most of the time, a global scope route-target is defined on all local-fabric and remote-fabric VTEPs in the scope of the VRF. Some identified VTEPs may be out of the global scope, with local scope only.
- Define a unique global scope for simplicity.
  Multiple global values might be possible to create groups of fabrics (more complex design).

Local scope

- Defined per VRF
- A local scope route-target is defined on every VTEPs of the local fabric.
- Must be set on all local-fabric VTEPs in the scope of the VRF

- **Recommendation**: use local scope and global scope for greater filtering flexibility
  Example: a route can be populated on all local-fabric VTEPs but not to other external fabrics

Underlay Loopbacks Routing

Include all VTEP loopbacks versus VTEP-border loopbacks only

All VTEPs

- Easier troubleshooting to be able to:
  - ping any VTEP Loopback from any VTEP Loopback
  - Traceroute any VTEP Loopback from any VTEP Loopback
- IP must be unique across all fabrics: IPAM extra work

Border-VTEPs only

- Ping and traceroute can still be performed between borders but no longer to/from a non-border leaf VTEP. Less practical for underlay troubleshooting.
- Loopback IP addresses for non-border VTEP can overlap between fabrics: less work for IPAM.

10.09 Multi-Fabric EVPN routes

Full routes / Optimized routes

All routes

- Granular routes come with associated complexity due to the volume of routes.
- For 10.09.0001: redistribute local-SVI
- For 10.09 CPE (supporting SVI_IP = AG_IP): avoid redistribute local-SVI if possible
- VSX Common Route-Distinguisher

Optimized routes (10.09)

- Routes aggregation helps optimizing routing tables (minimize CPU load).
- Routes filtering:
  - **Remove site-transit functionality** for EVPN routes
  - host routes filter with IP prefix-list (deny of /32)
  - host routes filter with L2VNI matching
  - No support yet for route-type filtering
- VSX Common Route-Distinguisher

L2VNI, L3VNI, VRF Configuration Normalization

Across multiple EVPN fabrics

- For L3 VXLAN multi-fabric:
  - Configure the same L3 VNI, VRF and EVPN Route-Targets (global scope) on all VTEPs in the data-plane path

- For L2 VXLAN multi-fabric:
  - Configure the same L2 VNI, EVPN Route-Targets (global scope) on all VTEPs in the data-plane path.
  - Same VLAN per L2VNI is also recommended for simplicity.

BGP send-community

Extended or both (extended + standard)

- Extended is required for EVPN AF.
- Standard might be used in future release for route-engineering.
- Consequently, it is recommended to use “both” as enabler for later support of community based route-engineering.

AFC

Support for multi-fabrics

- AFC can manage multiple isolated fabrics.
- AFC 6.2 can not yet manage Fabrics EVPN interconnectivity.
- Some changes must be processed in addition of AFC workflow per Fabric:
  - On all VTEPs: loopback routing (OSPF route-map and cross-redistribution between OSPF and BGP)
  - On fabric-borders:
    - eBGP IPv4 AF peering
    - eBGP EVPN AF multi-hop peering
    - split-horizon
    - route-map for site-transit removal and support of multi-PODs

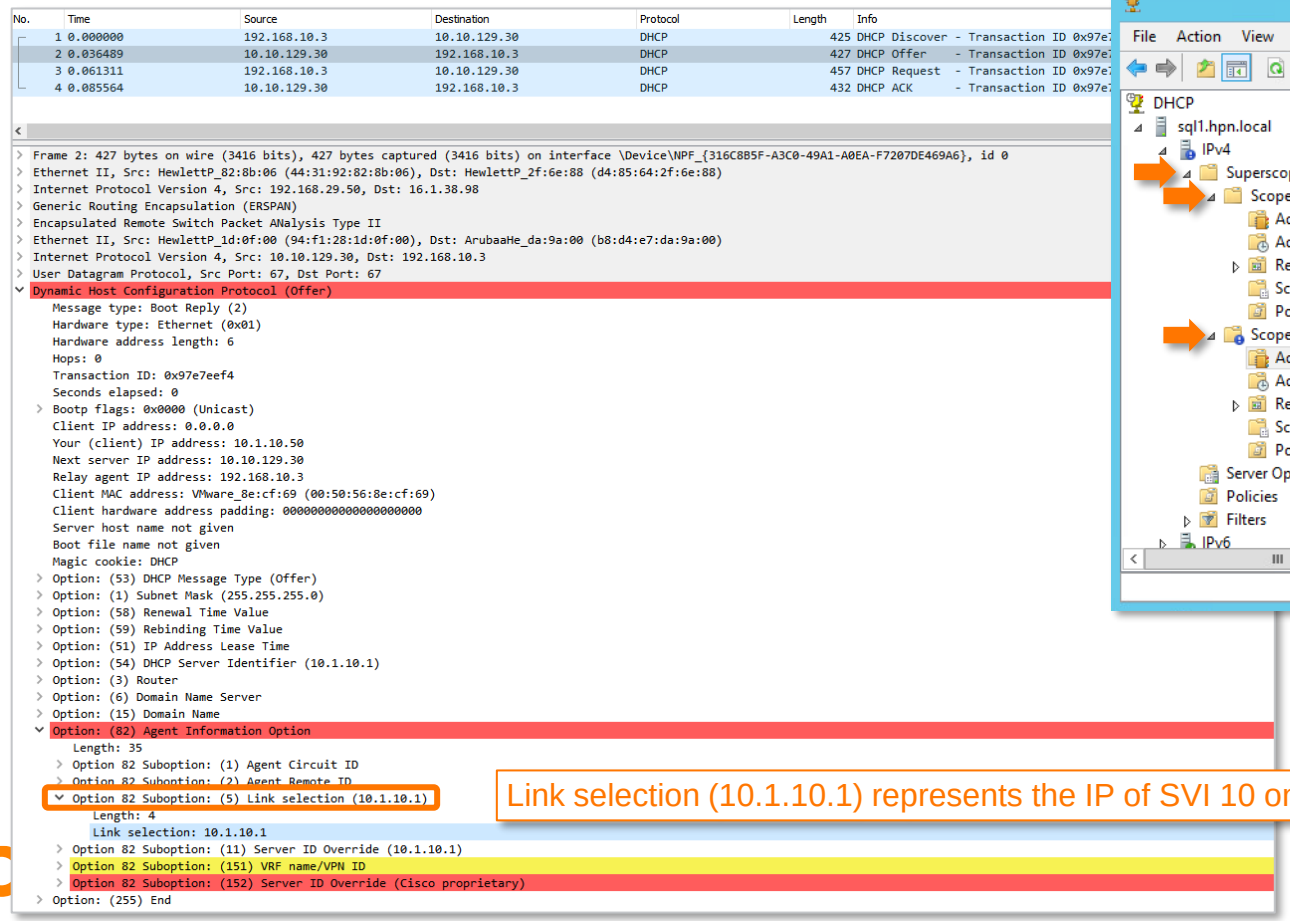
Windows DHCP server

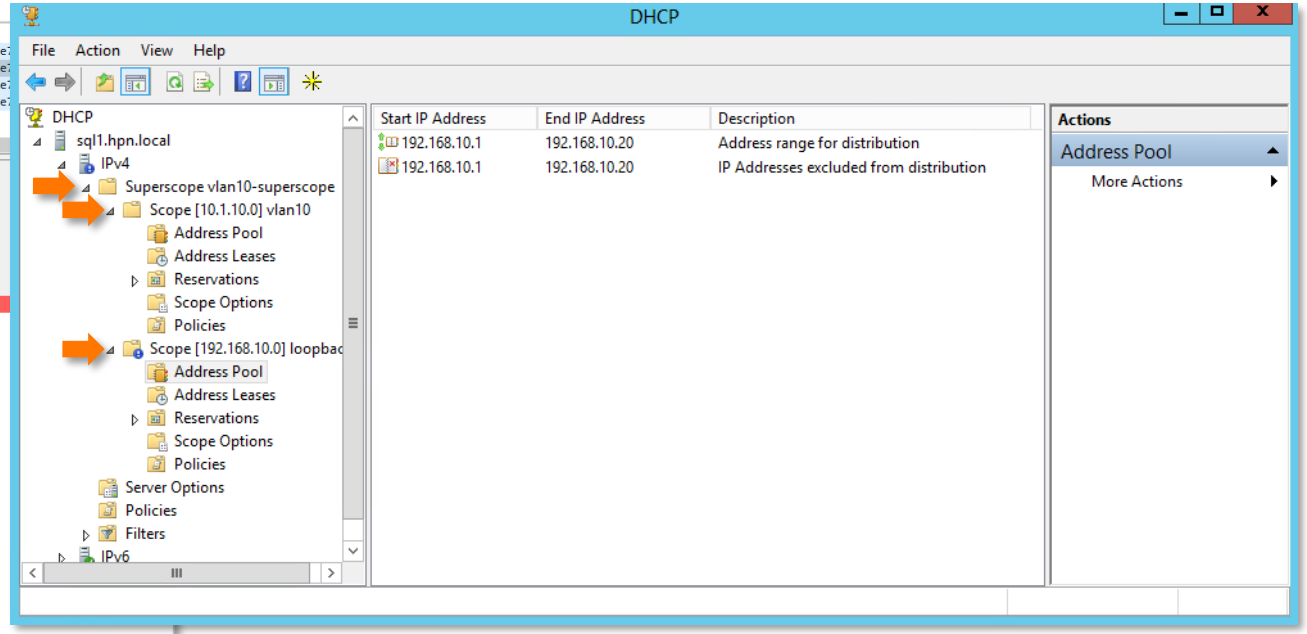
Option 82 - Sub-option 5 (SVI IP)

- Windows Server 2016 or later
- Superscope

- Windows DHCP server does not support VRF (sub-option 151)
- Source IP of DHCP request is the VTEP loopback instead of SVI IP.
- DHCP server will respond with NAK if source is not in the same range as the requested subnet range
- Adding the dummy scope with all loopbacks from all VTEPs to the DHCP pool is a way to inform the DHCP server to treat the request as authorized. (Exclude the loopback range).

https://datatracker.ietf.org/doc/html/rfc3527
https://docs.microsoft.com/en-us/windows-server/networking/technologies/dhcp/dhcp-subnet-options





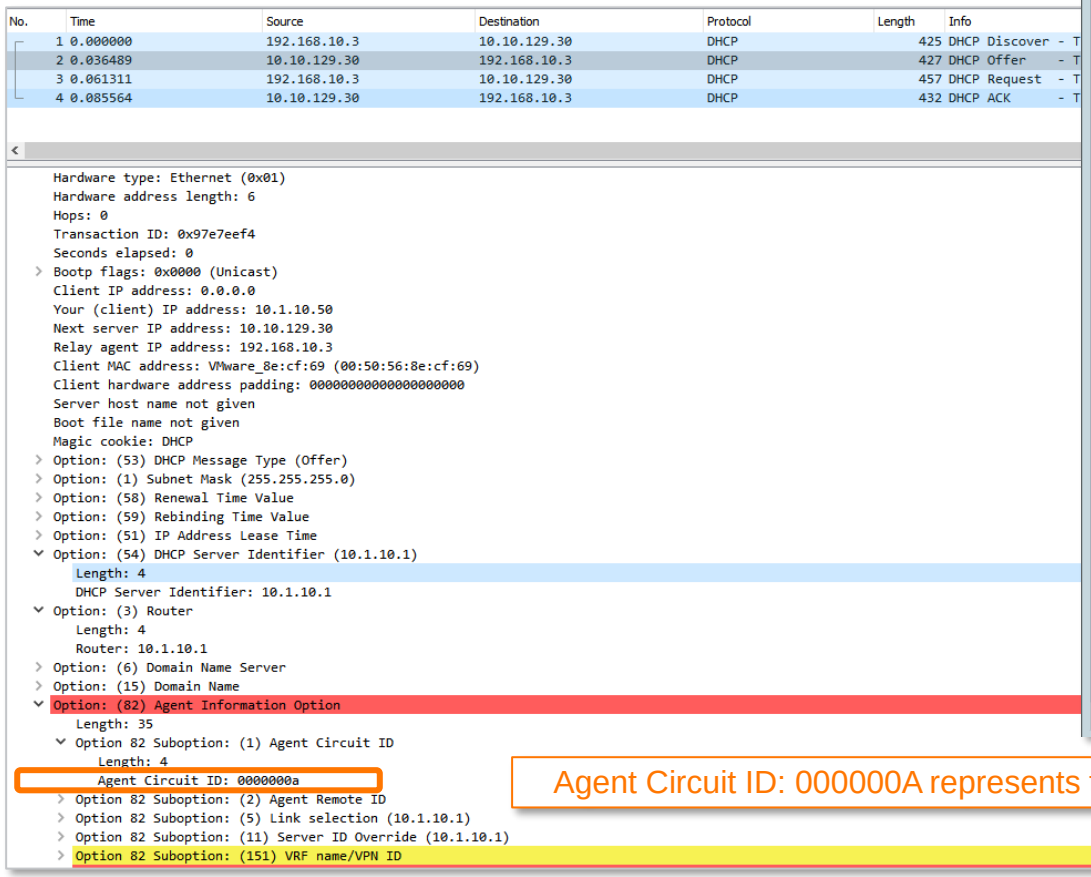
Link selection (10.1.10.1) represents the IP of SVI 10 on the DHCP relay VTEP

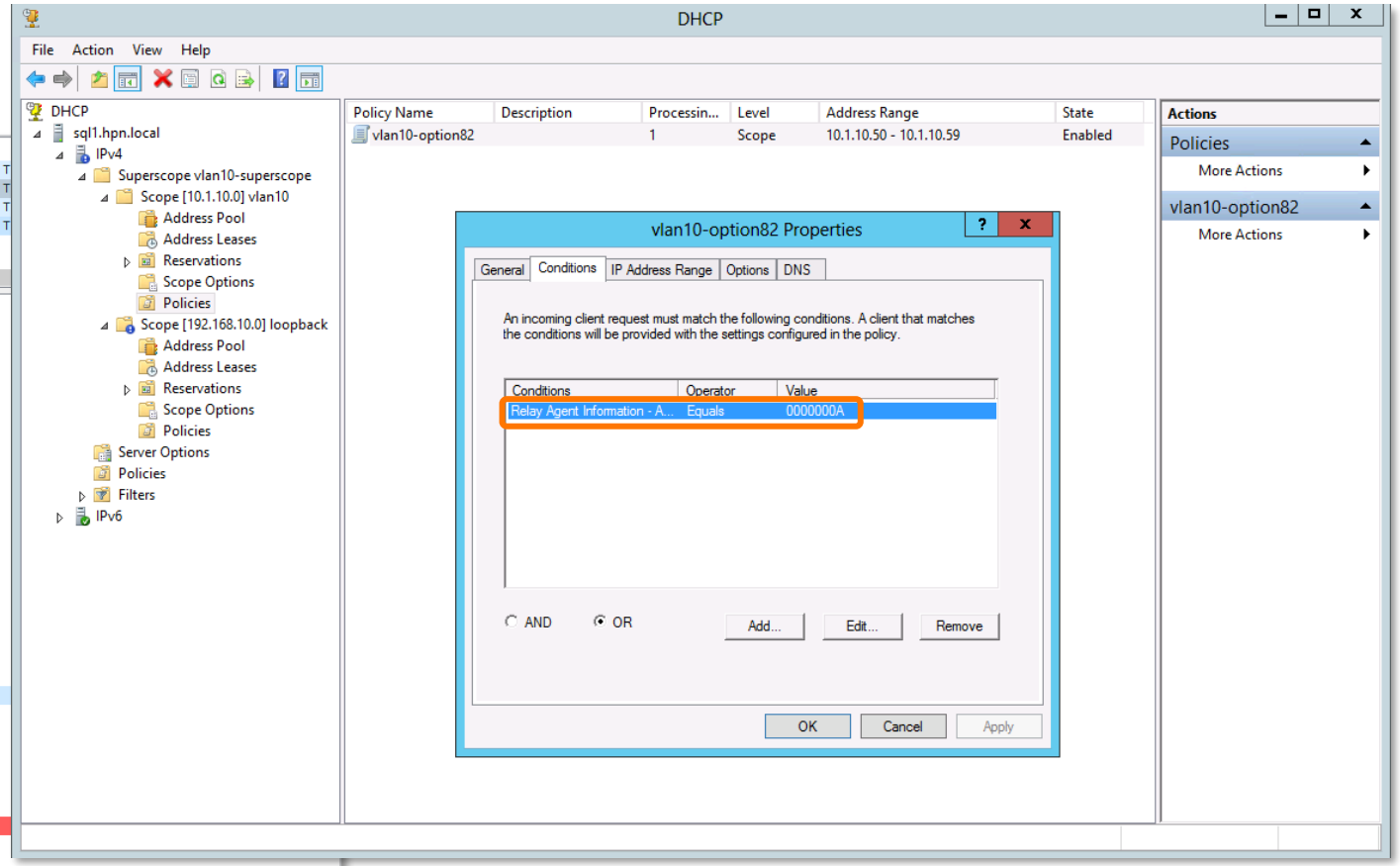
Windows DHCP server

Option 82 - Sub-option 1 (SVI ID)

- Windows Server 2012 (or later)
- Superscope + Windows DHCP server policy

- Windows DHCP server does not support VRF (sub-option 151)
- Source IP of DHCP request is the VTEP loopback instead of SVI IP.
- DHCP server will respond with NAK if source is not in the same range as the requested subnet range
- Adding the dummy scope with all loopbacks from all VTEPs to the DHCP pool is a way to inform the DHCP server to treat the request as authorized. (Exclude the loopback range).
- Use Agent Circuit ID (= SVI ID). It requires having homogenous SVI ID among DCs…





Agent Circuit ID: 000000A represents the SVI 10 on the DHCP relay VTEP

Troubleshooting

EVPN VXLAN Multi-Fabric Troubleshooting

- Have a topology diagram with interface, IP and AS# details ready
- Check physical cabling and generate “show tech” when opening a TAC case
- Check underlay network: show LLDP neighbor, ping and traceroute between loopbacks, fix any issues found
  - The Overlay network is dependent on a working underlay network

Recommended L3 multi-fabric troubleshooting flow

1. Check L3 multi-fabric and EVPN route-maps are correctly configured
2. Verify EVPN peers are up between expected VTEPs
3. Verify VXLAN tunnels are up with expected next-hop IP and router-MAC seen
4. Verify destination host routes are in VRF routing table
5. Verify hosts have L3 connectivity across fabrics

Same initial steps for both L3 and L2 multi-fabric troubleshooting

Recommended L2 multi-fabric troubleshooting flow

1. Check L2 multi-fabric and EVPN route-maps are correctly configured
2. Verify EVPN peers are up between expected VTEPs
3. Verify VXLAN tunnels are up with expected next-hop IP seen
4. Verify destination MAC addresses are learnt
5. Verify hosts have L2 connectivity across fabrics

1. Check multi-fabric and EVPN route-maps are correctly configured

- Refer to config and demo section for L3 and L2 multi-fabric and EVPN route-map sample configs

2. Verify EVPN peers are up between expected VTEPs

- Check that your expected EVPN peers are established
- Within the same AS and to other AS (if expected)

```
8325-1# show bgp l2vpn evpn sum
VRF : default
BGP Summary
-----------
 Local AS               : 65001        BGP Router Identifier  : 192.168.1.3
 Peers                  : 6            Log Neighbor Changes   : Yes
 Cfg. Hold Time         : 180          Cfg. Keep Alive        : 60
 Confederation Id       : 0

 Neighbor        Remote-AS MsgRcvd MsgSent   Up/Down Time State       AdminStatus
 192.168.1.1     65001       1075    896     04h:14m:38s  Established   Up
 192.168.1.2     65001       1040    906     04h:14m:38s  Established   Up
 192.168.2.8     65002       779     948     04h:14m:37s  Established   Up
 192.168.3.1     65003       949     981     04h:14m:24s  Established   Up
 192.168.4.1     65004       973     979     04h:14m:38s  Established   Up
 192.168.4.2     65004       957     972     04h:14m:38s  Established   Up
```

2. Verify EVPN peers are up between expected VTEPs

- You can also verify VNI/VLAN/VRF mappings between EVPN peers
- EVI = EVPN instance

```
8325-2# show evpn evi summary
   L2VNI     VLAN                 Status
   ------------------------------------
   10010     10                   Up
   10011     11                   Up
   10020     20                   Up

   L3VNI     VRF                  Status
   ------------------------------------
   100001    VRF1                 Up
   100002    VRF2                 Up

   EVPN instances    : 5
   EVPN instances Up : 5
```

- Check on the border-VTEP that non-directly attached VLANs have L2VNI-VLAN entry in the EVPN configuration, if these said VLANs must be extended between Fabrics.
- Check on all VTEPs that Route-Target values are set as expected, especially the L2VNI and L3VNI requiring a **Global Scope**

```
8325-2# show evpn vtep-neighbor all-vrfs
VTEP-IP         L3VNI    MAC               VRF                                  State
-------------------------------------------------------------------------------------
192.168.3.1     100001   02:00:00:00:06:00 VRF1                                 Up
192.168.10.5    100001   02:00:00:00:02:00 VRF1                                 Up
192.168.40.1    100002   02:00:00:00:07:00 VRF2                                 Up
192.168.40.1    100001   02:00:00:00:07:00 VRF1                                 Up
192.168.2.8     100002   02:00:00:00:05:00 VRF2                                 Up
192.168.2.8     100001   02:00:00:00:05:00 VRF1                                 Up
192.168.10.5    100002   02:00:00:00:02:00 VRF2                                 Up
```

Important verification for L3VNI:
Check binding between VTEP-IP and router-MAC

2. Verify EVPN peers are up between expected VTEPs

– Another way to verify EVPN peers with RD/RT/MAC info

```
8325-2# sh evpn evi detail
L2VNI : 10010
    Route Distinguisher         : 192.168.10.3:10
    VLAN                        : 10
    Status                      : up
    RT Import                   : 1:10, 65001:10
    RT Export                   : 1:10, 65001:10
    Local MACs                  : 10
    Remote MACs                 : 23
    Peer VTEPs                  : 3

    Peer-VTEP-Address                           Remote-MACs
    ----------------------------------------------------
    192.168.10.5                                8
    192.168.2.8                                 7
    192.168.40.1                                8

L2VNI : 10011
    Route Distinguisher         : 192.168.10.3:11
    VLAN                        : 11
    Status                      : up
    RT Import                   : 1:11, 65001:11
    RT Export                   : 1:11, 65001:11
    Local MACs                  : 7
    Remote MACs                 : 4
    Peer VTEPs                  : 1

    Peer-VTEP-Address                           Remote-MACs
    ----------------------------------------------------
    192.168.3.1                                 4

L2VNI : 10020
    Route Distinguisher         : 192.168.10.3:20
    VLAN                        : 20
    Status                      : up
    RT Import                   : 65001:20
    RT Export                   : 65001:20
    Local MACs                  : 4
    Remote MACs                 : 5
    Peer VTEPs                  : 1

    Peer-VTEP-Address                           Remote-MACs
    ----------------------------------------------------
    192.168.10.5                                5
```

```
L3VNI : 100001
    Route Distinguisher         : 192.168.1.4:1
    VRF                         : VRF1
    Status                      : up
    RT Import                   : 1:1, 65001:1
    RT Export                   : 1:1, 65001:1
    Local Type-5 Routes         : 6
    Remote Type-5 Routes        : 47
    Peer VTEPs                  : 4

    Peer-VTEP-Address                           Remote-Routes
    ----------------------------------------------------
    192.168.10.5                                22
    192.168.2.8                                 13
    192.168.40.1                                9
    192.168.3.1                                 3

L3VNI : 100002
    Route Distinguisher         : 192.168.1.4:2
    VRF                         : VRF2
    Status                      : up
    RT Import                   : 1:2, 65001:2
    RT Export                   : 1:2, 65001:2
    Local Type-5 Routes         : 1
    Remote Type-5 Routes        : 14
    Peer VTEPs                  : 4

    Peer-VTEP-Address                           Remote-Routes
    ----------------------------------------------------
    192.168.10.5                                5
    192.168.2.8                                 5
    192.168.40.1                                4
    192.168.3.1                                 0
```

3. Verify VXLAN tunnels are up with expected next hop seen

– VXLAN tunnels should be up for the desired VNIs with expected next hop seen

```
8325-1# show interface vxlan
Interface vxlan1 is up
Admin state is up
Description:
Underlay VRF: default
Destination UDP port: 4789
VTEP source IPv4 address: 192.168.10.3

VNI        Routing      VLAN   VRF          VTEP Peers        Origin
---------- ------------ ------ ------------ ----------------- --------
10010      disabled     10     --           192.168.2.8       evpn
10010      disabled     10     --           192.168.10.5      evpn
10010      disabled     10     --           192.168.40.1      evpn
10011      disabled     11     --           192.168.3.1       evpn
10020      disabled     20     --           192.168.10.5      evpn
100001     enabled      --     VRF1         192.168.2.8       evpn
100001     enabled      --     VRF1         192.168.3.1       evpn
100001     enabled      --     VRF1         192.168.10.5      evpn
100001     enabled      --     VRF1         192.168.40.1      evpn
100002     enabled      --     VRF2         192.168.2.8       evpn
100002     enabled      --     VRF2         192.168.10.5      evpn
100002     enabled      --     VRF2         192.168.40.1      evpn
```

```
Aggregate Statistics
--------------------
 Decap:
           8 input packets                1168 bytes
        4053 broadcast packets          666916 bytes
           0 drop packets
 Encap:
       47449 output packets            3929299 bytes
        3839 BUM packets                356512 bytes
           0 drop packets
```

Check on the border-VTEP that non-directly attached VLANs have L2VNI-VLAN entry in the VXLAN interface, if these said VLANs must be extended between Fabrics.

3. Verify VXLAN tunnels are up with expected next hop seen

– With more details

```
8325-1# show interface vxlan vni vteps
VNI            : 10010       VLAN : 10
Routing        : disabled    VRF  : --
VNI-Status     : operational
VTEPS
=====
    Origin      Source           Destination      VRF         VTEP-STATUS
    ----------- ---------------- ---------------- ----------- ----------------
    evpn        192.168.10.3     192.168.2.8      default     operational
    evpn        192.168.10.3     192.168.10.5     default     operational
    evpn        192.168.10.3     192.168.40.1     default     operational

VNI            : 10011       VLAN : 11
Routing        : disabled    VRF  : --
VNI-Status     : operational
VTEPS
=====
    Origin      Source           Destination      VRF         VTEP-STATUS
    ----------- ---------------- ---------------- ----------- ----------------
    evpn        192.168.10.3     192.168.3.1      default     operational

VNI            : 10020       VLAN : 20
Routing        : disabled    VRF  : --
VNI-Status     : operational
VTEPS
=====
    Origin      Source           Destination      VRF         VTEP-STATUS
    ----------- ---------------- ---------------- ----------- ----------------
    evpn        192.168.10.3     192.168.10.5     default     operational

VNI            : 100001      VLAN : --
Routing        : enabled     VRF  : VRF1
VNI-Status     : operational
VTEPS
=====
    Origin      Source           Destination      VRF         VTEP-STATUS
    ----------- ---------------- ---------------- ----------- ----------------
    evpn        192.168.10.3     192.168.2.8      default     operational
    evpn        192.168.10.3     192.168.3.1      default     operational
    evpn        192.168.10.3     192.168.10.5     default     operational
    evpn        192.168.10.3     192.168.40.1     default     operational
```

```
8325-1# sh int vxlan vteps detail
Destination     :  192.168.2.8
Source          :  192.168.10.3
Origin          :  evpn
VRF             :  default
Status          :  operational
Nexthops
========
    IP-ADDRESS       INTERFACE   NEXTHOP-MAC
    ---------------  ----------  -----------------
    192.168.29.6     1/1/51      94:f1:28:1d:0f:00
Destination     :  192.168.3.1
Source          :  192.168.10.3
Origin          :  evpn
VRF             :  default
Status          :  operational
Nexthops
========
    IP-ADDRESS       INTERFACE   NEXTHOP-MAC
    ---------------  ----------  -----------------
    192.168.29.6     1/1/51      94:f1:28:1d:0f:00
Destination     :  192.168.10.5
Source          :  192.168.10.3
Origin          :  evpn
VRF             :  default
Status          :  operational
Nexthops
========
    IP-ADDRESS       INTERFACE   NEXTHOP-MAC
    ---------------  ----------  -----------------
    192.168.19.8     1/1/24      88:3a:30:93:bc:00
    192.168.19.0     1/1/23      88:3a:30:93:ca:40
Destination     :  192.168.40.1
Source          :  192.168.10.3
Origin          :  evpn
VRF             :  default
Status          :  operational
Nexthops
========
    IP-ADDRESS       INTERFACE   NEXTHOP-MAC
    ---------------  ----------  -----------------
    192.168.29.6     1/1/51      94:f1:28:1d:0f:00
```

4. Verify destination host routes are in VRF routing table

- Applicable to L3 multi-fabric
- /32 host route should be seen with expected next-hop

```
8325-1# sh ip route vrf VRF1
!snip
VRF: VRF1
Prefix                Nexthop           Interface     VRF(egress)    Origin/   Distance/    Age
                                                                     Type      Metric
-------------------------------------------------------------------------------------------
10.1.10.0/24          -                 vlan10        -                C         [0/0]         -
10.1.10.1/32          -                 vlan10        -                L         [0/0]         -
10.1.10.6/32          192.168.2.8       -             -                B/EV      [200/0]       05h:00m:35s
10.1.10.13/32         192.168.10.5      -             -                B/EV      [200/0]       00h:05m:22s
10.1.10.15/32         192.168.2.8       -             -                B/EV      [200/0]       05h:00m:35s
10.1.10.18/32         192.168.40.1      -             -                B/EV      [200/0]       00h:02m:30s
10.1.11.0/24          -                 vlan11        -                C         [0/0]         -
10.1.11.1/32          -                 vlan11        -                L         [0/0]         -
10.1.11.17/32         192.168.3.1       -             -                B/EV      [200/0]       00h:00m:34s
10.1.12.0/24          192.168.10.5      -             -                B/EV      [200/0]       05h:00m:36s
10.1.12.1/32          192.168.10.5      -             -                B/EV      [200/0]       05h:00m:36s
10.1.12.14/32         192.168.10.5      -             -                B/EV      [200/0]       00h:05m:19s
10.1.12.16/32         192.168.2.8       -             -                B/EV      [200/0]       05h:00m:35s
192.168.11.3/32       -                 loopback12    -                L         [0/0]         -
192.168.11.5/32       192.168.10.5      -             -                B/EV      [200/0]       05h:00m:36s
192.168.11.6/32       192.168.10.5      -             -                B/EV      [200/0]       05h:00m:36s
192.168.11.103/32     -                 loopback11    -                L         [0/0]         -
```

4. Verify destination host routes are in VRF routing table

– ARP table should also be populated with remote host info if the same VLANs exist

```
8325-1# sh arp vrf VRF1

IPv4 Address     MAC                Port         Physical Port              State      VRF
-----------------------------------------------------------------------------------------------------
10.1.10.15       00:50:56:8e:d7:96  vlan10       vxlan1(192.168.2.8)        permanent  VRF1
10.1.10.6        88:3a:30:9a:7a:00  vlan10       vxlan1(192.168.2.8)        permanent  VRF1
10.1.11.11       00:50:56:8e:4d:9c  vlan11       lag1                       reachable  VRF1
10.1.10.50       00:50:56:8e:cf:69  vlan10       lag2                       reachable  VRF1
10.1.10.10       00:50:56:8e:61:91  vlan10       lag1                       reachable  VRF1
10.1.10.13       00:50:56:86:2d:79  vlan10       vxlan1(192.168.10.5)       permanent  VRF1
10.1.11.17       00:50:56:8e:92:44  vlan11       vxlan1(192.168.3.1)        permanent  VRF1
10.1.10.18       00:50:56:8e:4e:88  vlan10       vxlan1(192.168.40.1)       permanent  VRF1

Total Number Of ARP Entries Listed: 8.
-----------------------------------------------------------------------------------------------------
```

4. Verify destination MAC addresses are learnt

- Applicable to L2 multi-fabric
- Destination MAC address should be seen with expected next-hop

```
8325-1# sh mac-add
MAC age-time             : 300 seconds
Number of MAC addresses : 20

MAC Address          VLAN     Type                        Port
-------------------------------------------------------------
b8:d4:e7:da:28:00    10       evpn                        vxlan1(192.168.10.5)
88:3a:30:9a:7a:00    10       evpn                        vxlan1(192.168.2.8)
54:80:28:fc:5c:00    10       dynamic                     lag256
00:50:56:8e:cf:69    10       dynamic                     lag2
00:50:56:8e:61:91    10       dynamic                     lag1
00:50:56:86:2d:79    10       evpn                        vxlan1(192.168.10.5)
00:50:56:8e:4e:88    10       evpn                        vxlan1(192.168.40.1)
00:50:56:8e:d7:96    10       evpn                        vxlan1(192.168.2.8)
54:80:28:fd:f3:00    10       evpn                        vxlan1(192.168.10.5)
88:3a:30:98:8a:00    10       evpn                        vxlan1(192.168.40.1)
```

4. Verify destination EVPN MAC/IP are learnt

– Another method to verify MACs are learnt across L2 VNIs

```
8325-2# sh evpn mac-ip
Flags:  Local(L), Remote(R), Sticky bit(S)

EVI    MAC                IP                 Next-hop                Seq-Num   Flags
-------------------------------------------------------------------------------------
10010  00:50:56:8e:61:91                                             0         L
10010  00:50:56:8e:61:91  10.1.10.10                                 0         L
10010  00:50:56:8e:cf:69                                             0         L
10010  00:50:56:8e:cf:69  10.1.10.50                                 0         L
10010  00:50:56:8e:cf:69  fe80::f4af:f785:51cb:403c
                                                                     0         L
10010  00:50:56:8e:d7:96                     vxlan1(192.168.2.8)     0         R
10010  00:50:56:8e:d7:96  10.1.10.15         vxlan1(192.168.2.8)     0         R
10010  12:00:00:00:01:00  10.1.10.1          vxlan1(192.168.2.8)     0         R,S
10010  12:00:00:00:01:00  fe80:10:1:10::1    vxlan1(192.168.2.8)     0         R,S
10010  54:80:28:fc:5c:00  10.1.10.1                                  0         L,S
10010  54:80:28:fc:5c:00  fd00:10:1:10::1                            0         L,S
10010  54:80:28:fc:5c:00  fe80:10:1:10::1                            0         L,S
10010  54:80:28:fd:f3:00  10.1.10.1          vxlan1(192.168.10.5)    0         R,S
10010  88:3a:30:98:8a:00  10.1.10.1          vxlan1(192.168.40.1)    0         R,S
10010  88:3a:30:98:8a:00  fd00:10:1:10::1    vxlan1(192.168.40.1)    0         R,S
10010  88:3a:30:98:8a:00  fe80:10:1:10::1    vxlan1(192.168.40.1)    0         R,S
10010  88:3a:30:9a:7a:00  10.1.10.6          vxlan1(192.168.2.8)     0         R,S
10010  88:3a:30:9a:7a:00  fd00:10:1:10::1    vxlan1(192.168.2.8)     0         R,S
10010  88:3a:30:9a:7a:00  fe80:10:1:10::1    vxlan1(192.168.2.8)     0         R,S
10010  b8:d4:e7:da:28:00  10.1.10.1          vxlan1(192.168.10.5)    0         R,S
10010  b8:d4:e7:da:28:00  fd00:10:1:10::1    vxlan1(192.168.10.5)    0         R,S
10010  b8:d4:e7:da:28:00  fe80:10:1:10::1    vxlan1(192.168.10.5)    0         R,S
10011  00:50:56:8e:4d:9c                                             0         L
10011  00:50:56:8e:4d:9c  10.1.11.11                                 0         L
10011  00:50:56:8e:92:44                     vxlan1(192.168.3.1)     0         R
10011  00:50:56:8e:92:44  10.1.11.17         vxlan1(192.168.3.1)     0         R
10011  12:00:00:00:01:00  10.1.11.1          vxlan1(192.168.3.1)     0         R,S
10011  12:00:00:00:01:00  fe80:10:1:11::1    vxlan1(192.168.3.1)     0         R,S
10011  54:80:28:fc:5c:00  10.1.11.1                                  0         L,S
10011  54:80:28:fc:5c:00  fd00:10:1:11::1                            0         L,S
10011  54:80:28:fc:5c:00  fe80:10:1:11::1                            0         L,S
10011  b4:99:ba:54:8b:60  10.1.11.1          vxlan1(192.168.3.1)     0         R,S
10020  00:50:56:8e:32:e8                                             0         L
10020  00:50:56:8e:32:e8  10.2.20.20                                 0         L
10020  00:50:56:8e:6b:d8                     vxlan1(192.168.10.5)    0         R
10020  00:50:56:8e:6b:d8  10.2.20.21         vxlan1(192.168.10.5)    0         R
10020  12:00:00:00:01:00  10.2.20.1                                  0         L,S
10020  54:80:28:fc:5c:00  10.2.20.3                                  0         L,S
10020  54:80:28:fd:f3:00  10.2.20.5          vxlan1(192.168.10.5)    0         R,S
10020  b8:d4:e7:da:28:00  10.2.20.4          vxlan1(192.168.10.5)    0         R,S

MACs           : 20
Remote MACs    : 12
```

5. Verify hosts have L2/L3 connectivity across fabrics

– Ping/traceroute between hosts and verify L2/L3 connectivity across fabrics

```
root@vlinux28:/home# ifconfig eth0
eth0      Link encap:Ethernet  HWaddr 00:50:56:8e:4e:88
          inet addr:10.1.10.18  Bcast:10.1.10.255  Mask:255.255.255.0
```

```
root@vlinux28:/home# ping 10.1.10.15
PING 10.1.10.15 (10.1.10.15) 56(84) bytes of data.
64 bytes from 10.1.10.15: icmp_req=1 ttl=64 time=2.30 ms
64 bytes from 10.1.10.15: icmp_req=2 ttl=64 time=0.601 ms
```

```
root@vlinux28:/home# ping 10.1.11.11
PING 10.1.11.11 (10.1.11.11) 56(84) bytes of data.
64 bytes from 10.1.11.11: icmp_req=1 ttl=61 time=2.98 ms
64 bytes from 10.1.11.11: icmp_req=2 ttl=61 time=0.597 ms
```

– Mirror traffic and packet capture if required

```
mirror session 1
    enable
    destination interface 1/1/40
    source interface 1/1/51 both
```

EVPN-VXLAN Multi-Fabric Solution Demonstration

AOS-CX 10.09

EVPN-VXLAN Multi-Fabric - Solution Demo

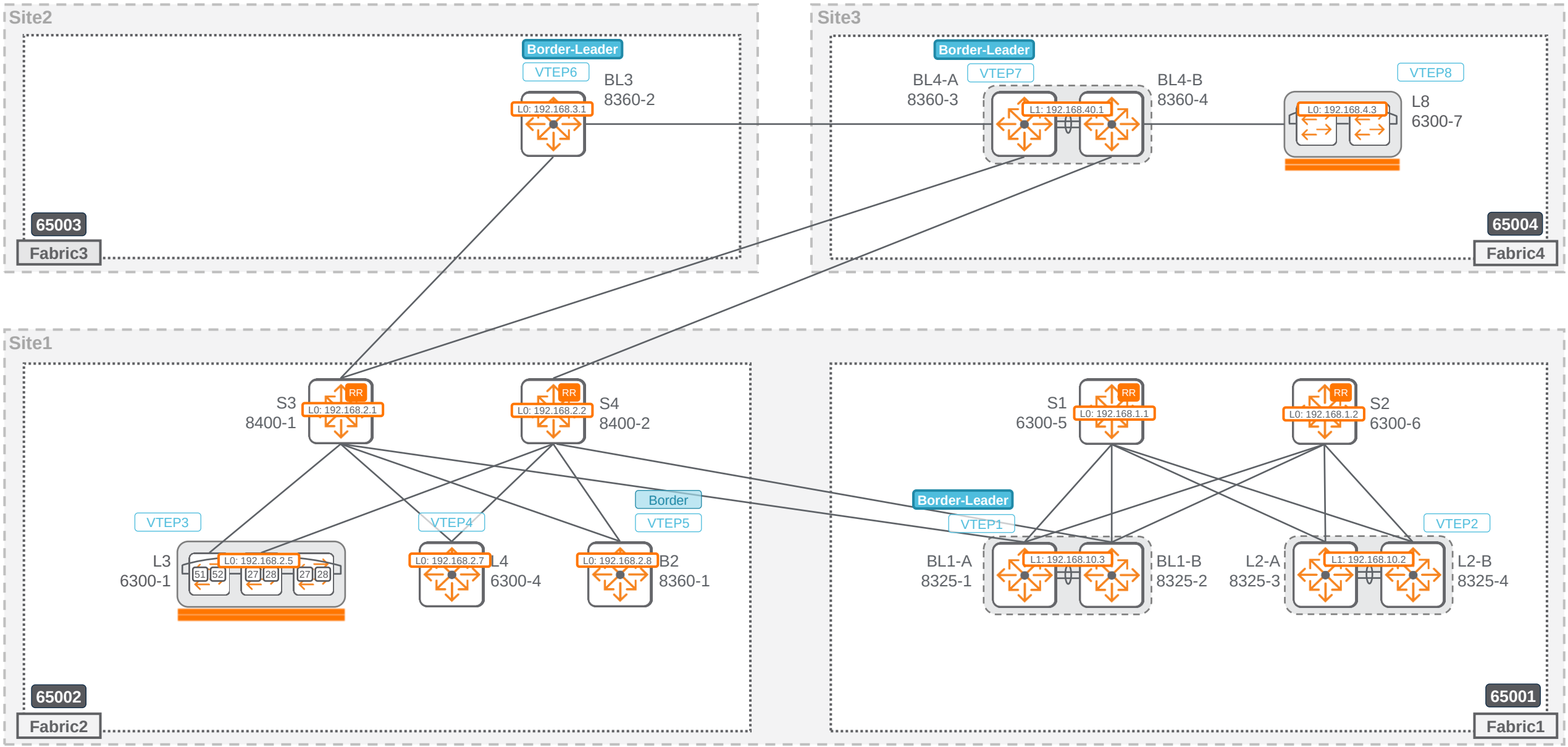
Agenda

- Overall set-up
- Brief review of the set-up of a local fabric:
  - underlay routing
  - overlay set-up
  - Reminder on best practices for intra-fabric
- Additional configuration for inter-Fabric communication: without route-map requirements
- Additional configuration for inter-Fabric communication with route-map requirements
- Benefits of route-map usage:
  - EVPN Next-hop rewrite
  - Site-transit function removal
  - Routes aggregation
  - Routes filtering
  - Path optimization
- Troubleshooting

8360s are used in this demonstration environment, although not yet supported as border-leader in 10.09

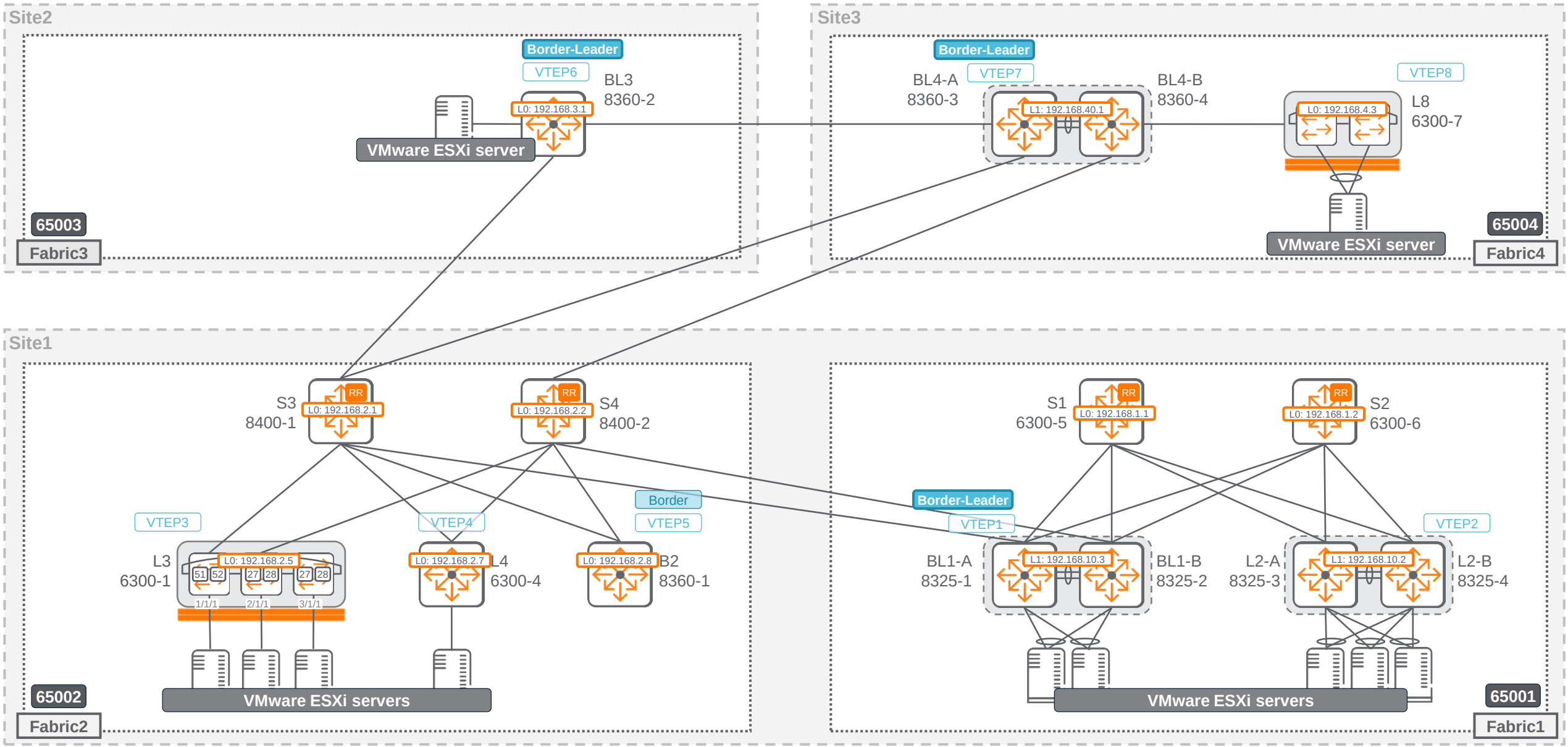
Demo set-up

Topology



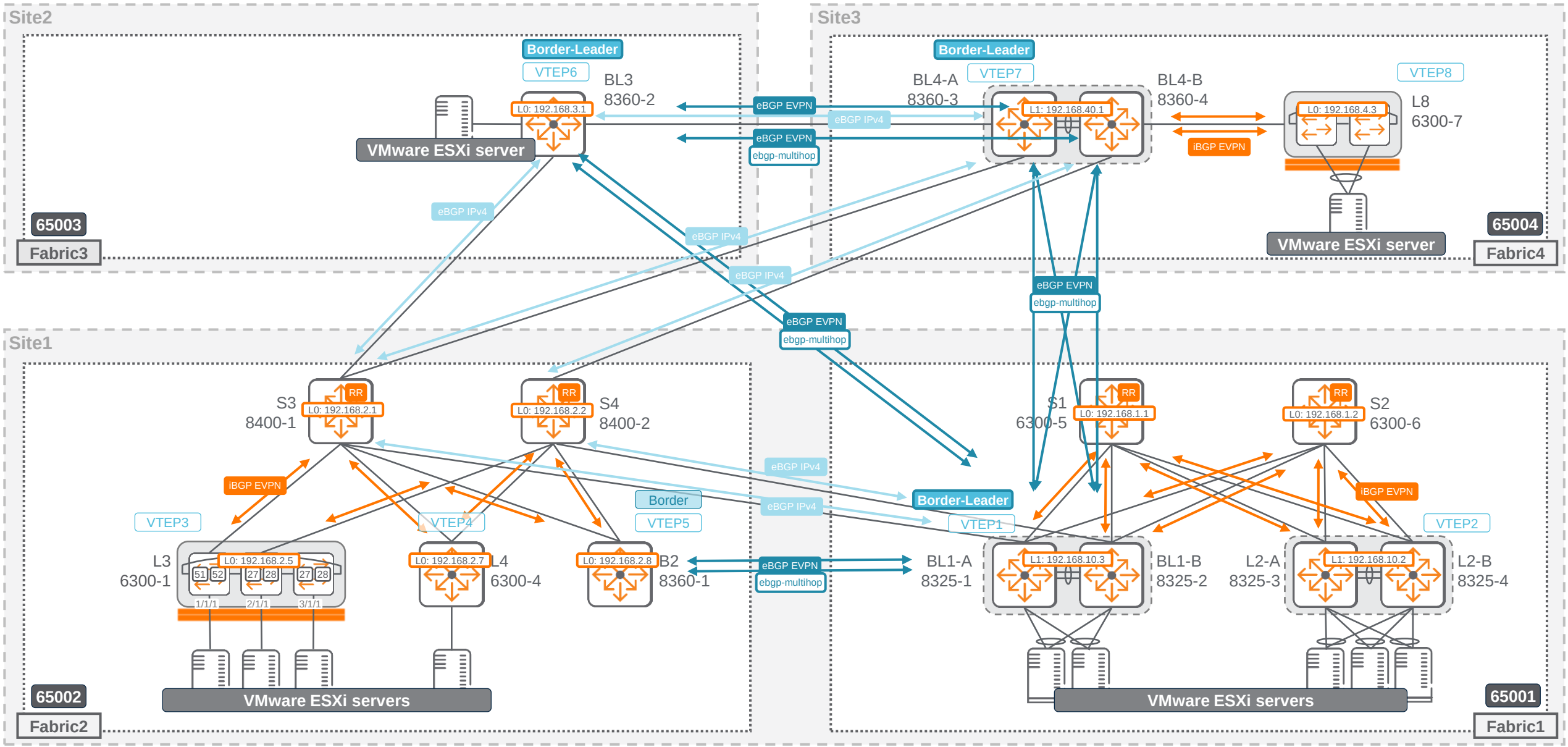
Demo set-up

Topology + hosts-servers-VMs



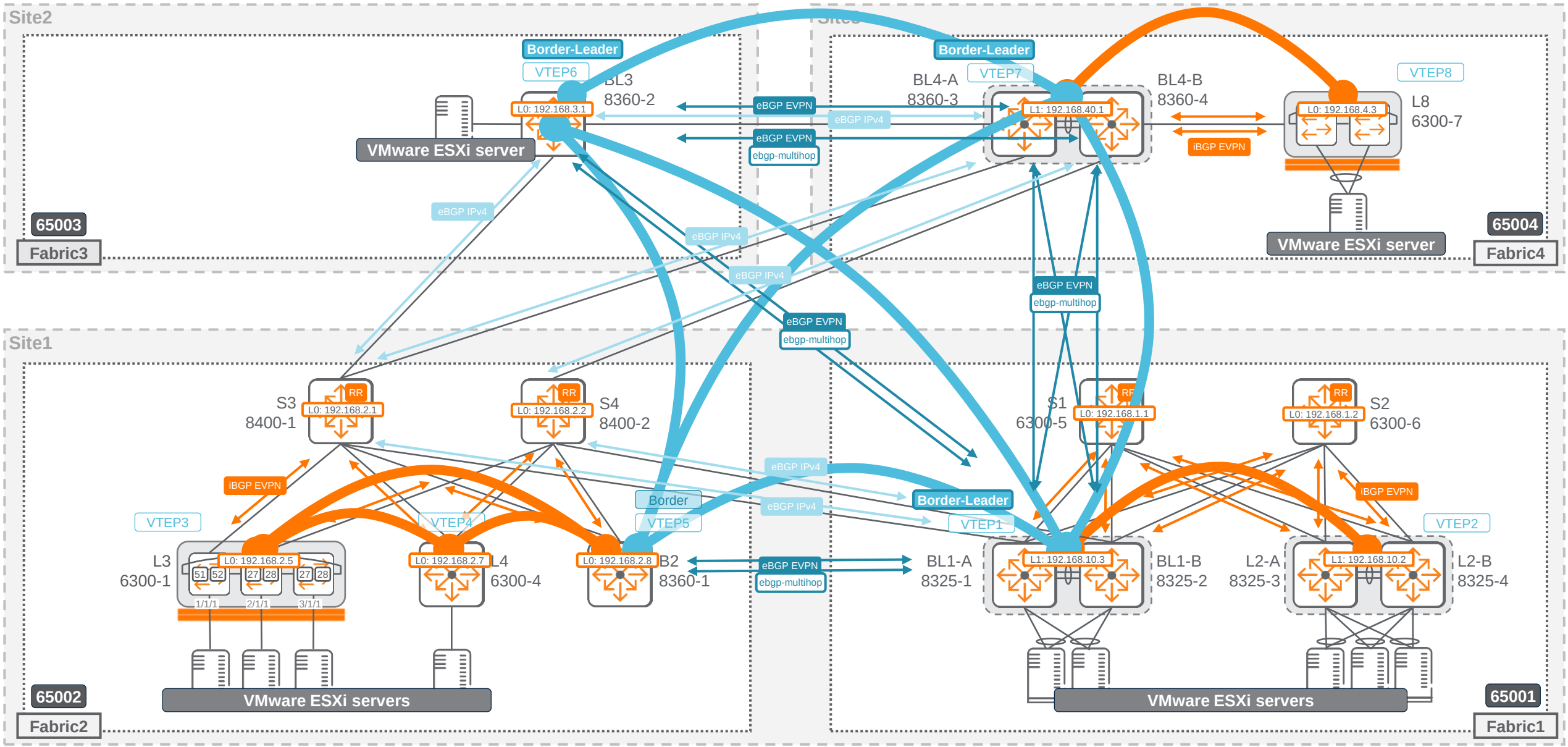
Demo set-up

Control-plane summary



Demo set-up

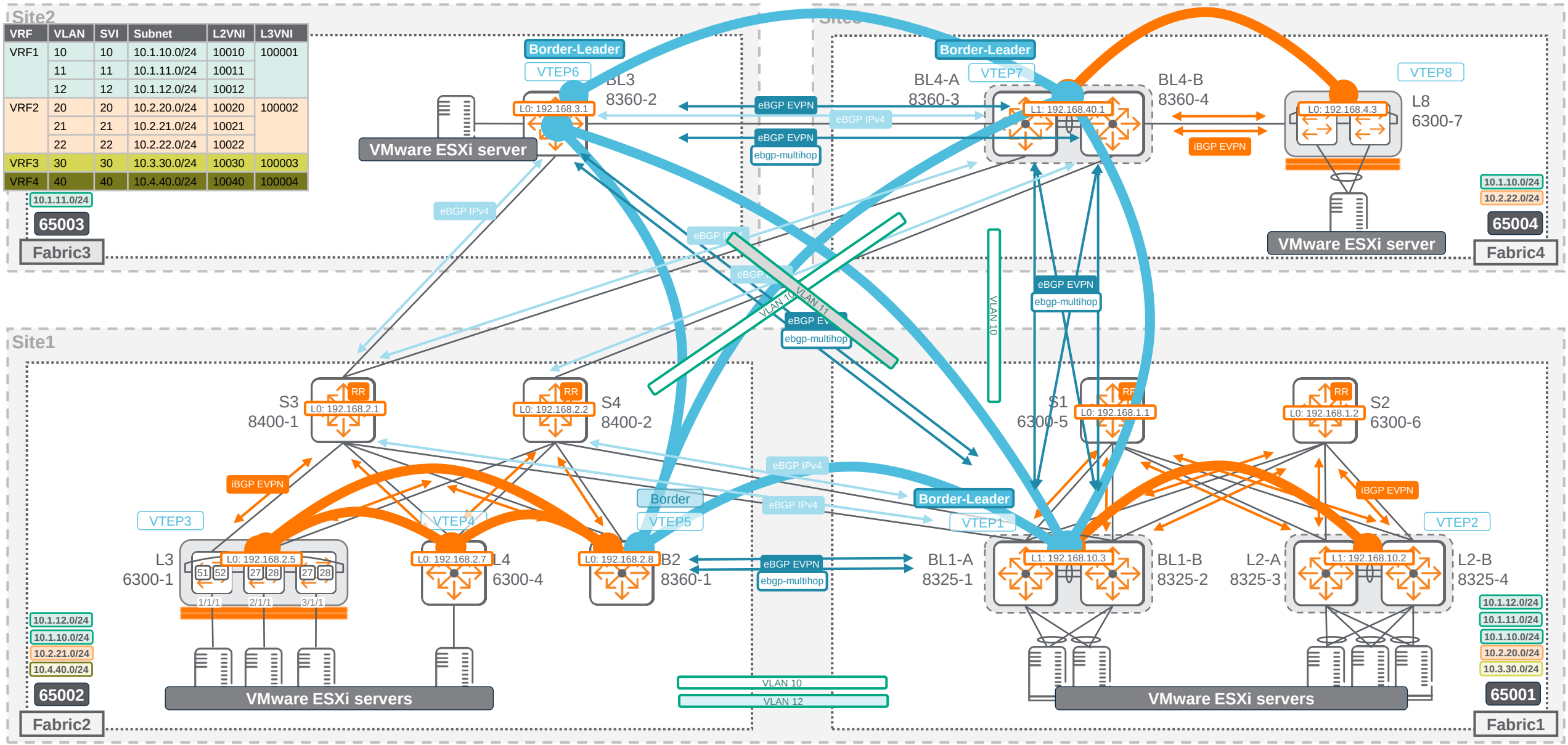
Data-plane



Demo set-up

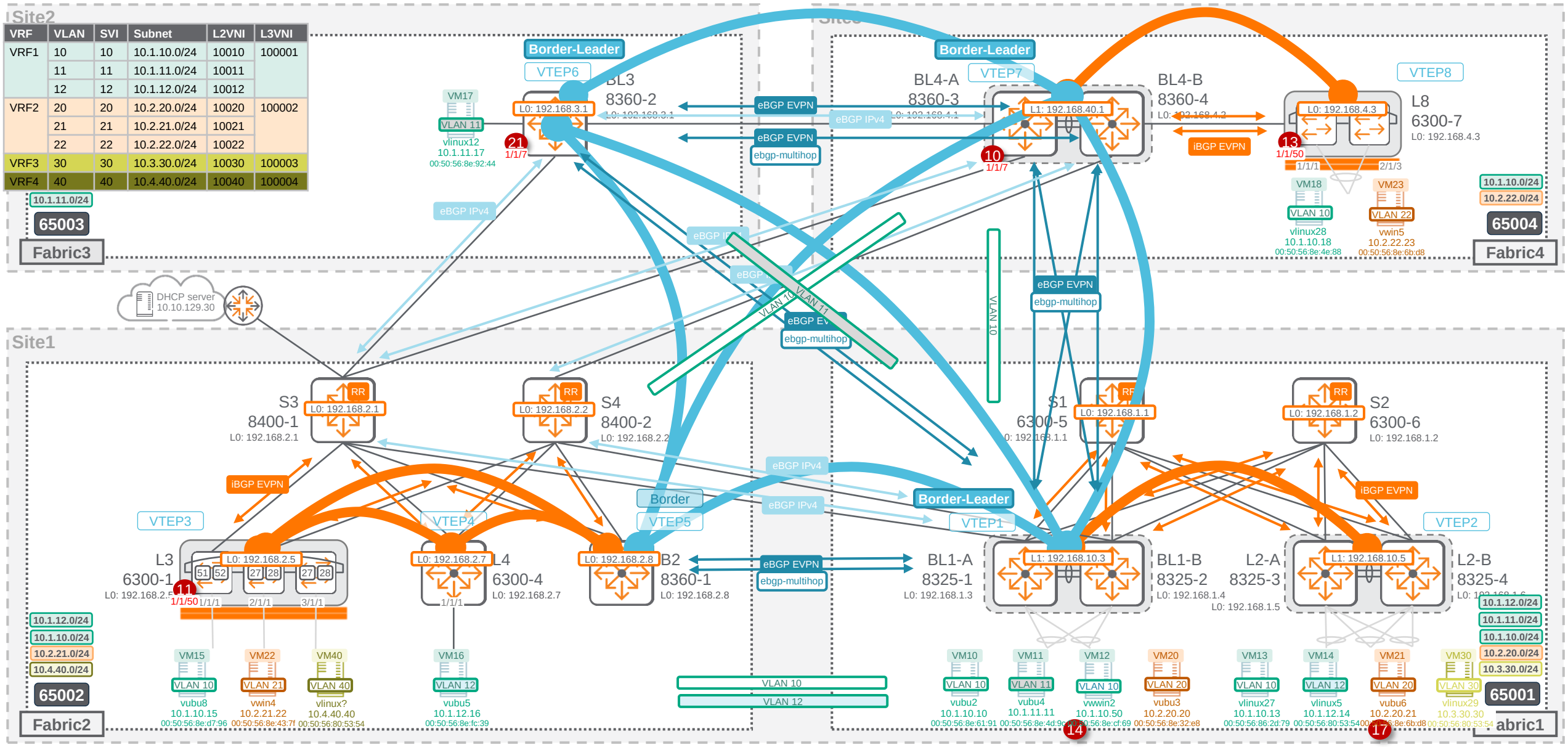
Tenant VRFs, VLANs, stretched VLANs

VRF	VLAN	SVI	Subnet	L2VNI	L3VNI
VRF1	10	10	10.1.10.0/24	10010	100001
	11	11	10.1.11.0/24	10011	
	12	12	10.1.12.0/24	10012	
VRF2	20	20	10.2.20.0/24	10020	100002
	21	21	10.2.21.0/24	10021	
	22	22	10.2.22.0/24	10022	
VRF3	30	30	10.3.30.0/24	10030	100003
VRF4	40	40	10.4.40.0/24	10040	100004



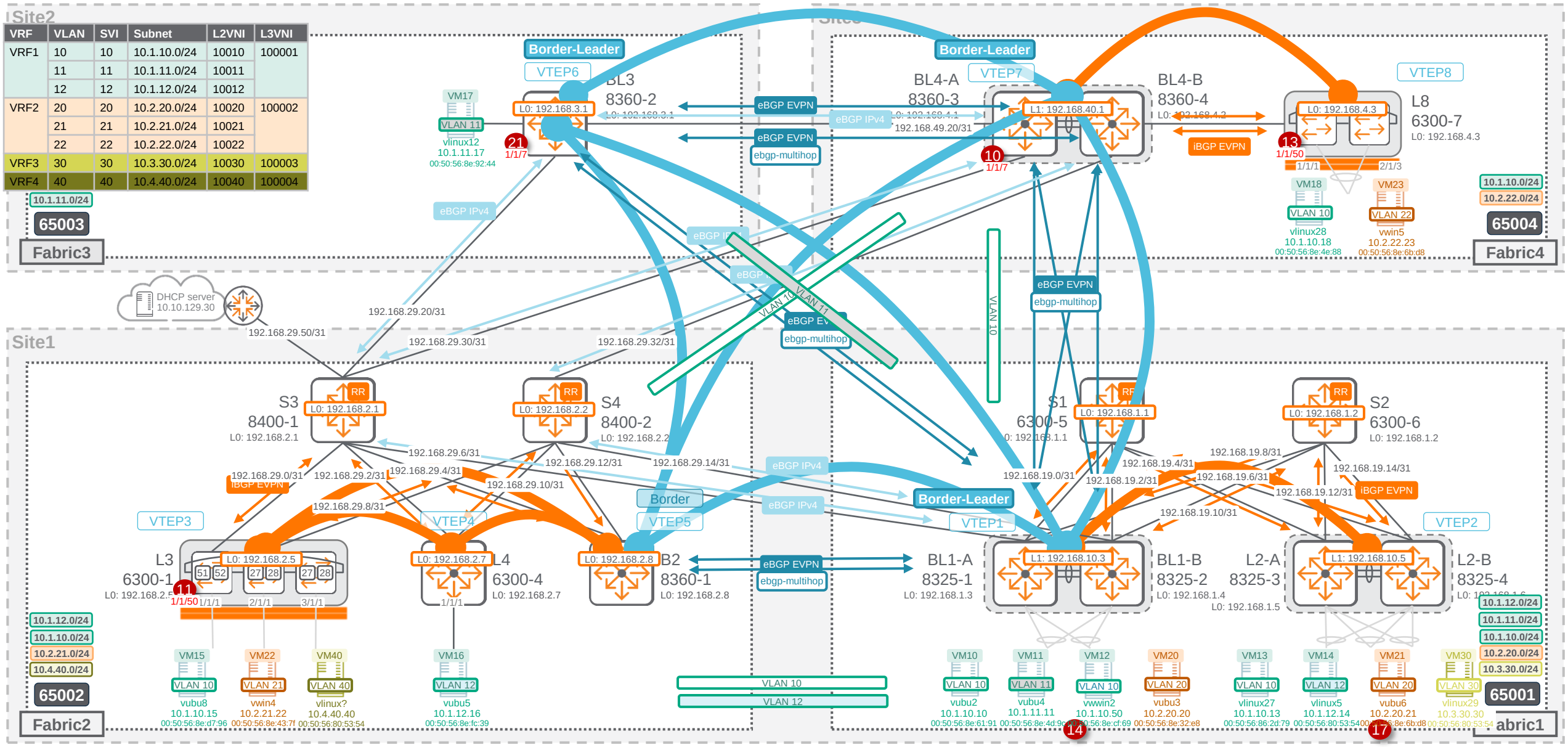
Demo Details

VRF	VLAN	SVI	Subnet	L2VNI	L3VNI
VRF1	10	10	10.1.10.0/24	10010	100001
	11	11	10.1.11.0/24	10011	
	12	12	10.1.12.0/24	10012	
VRF2	20	20	10.2.20.0/24	10020	100002
	21	21	10.2.21.0/24	10021	
	22	22	10.2.22.0/24	10022	
VRF3	30	30	10.3.30.0/24	10030	100003
VRF4	40	40	10.4.40.0/24	10040	100004



Demo Details
with ROP IP

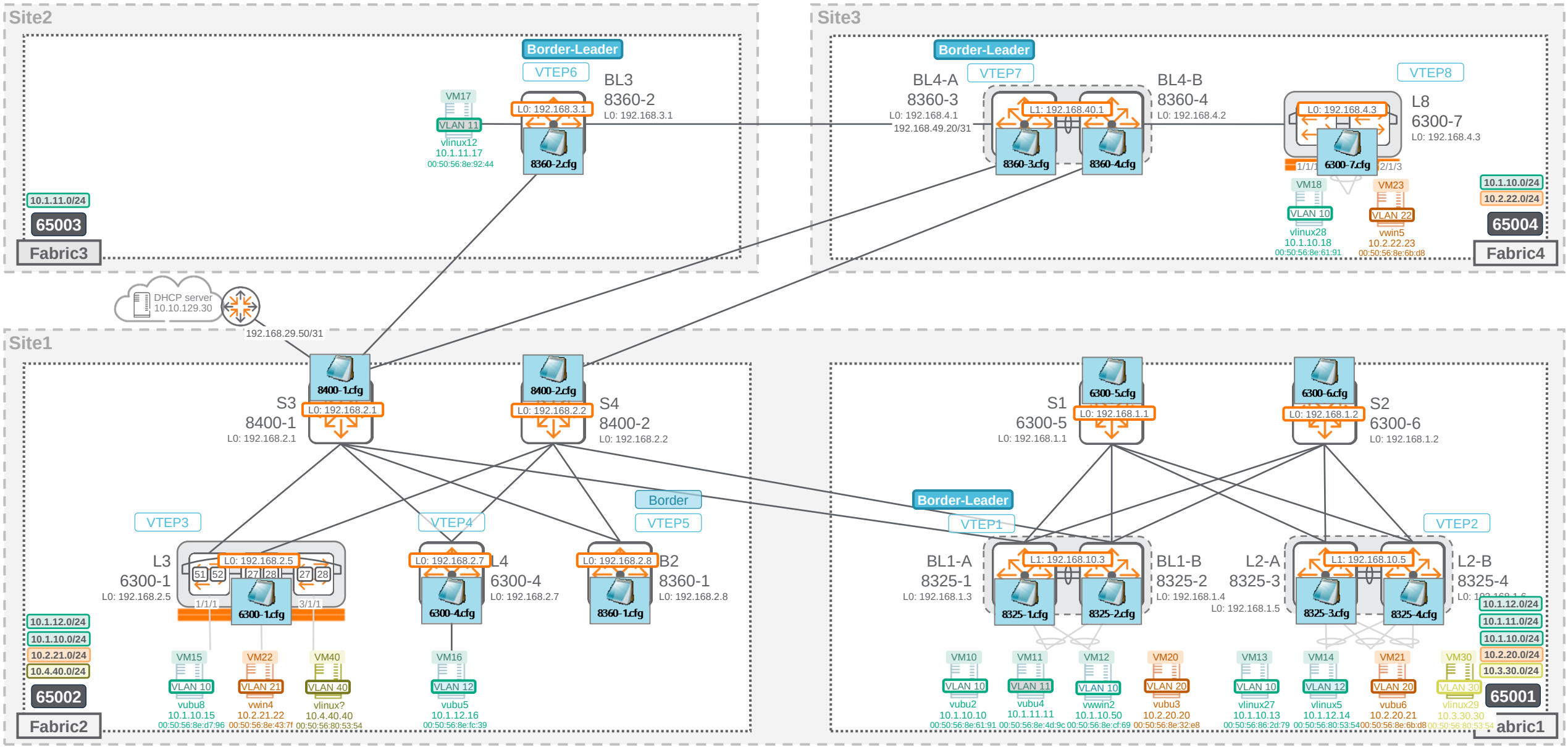
VRF	VLAN	SVI	Subnet	L2VNI	L3VNI
VRF1	10	10	10.1.10.0/24	10010	100001
	11	11	10.1.11.0/24	10011	
	12	12	10.1.12.0/24	10012	
VRF2	20	20	10.2.20.0/24	10020	100002
	21	21	10.2.21.0/24	10021	
	22	22	10.2.22.0/24	10022	
VRF3	30	30	10.3.30.0/24	10030	100003
VRF4	40	40	10.4.40.0/24	10040	100004



Switch configuration

Double-click on .cfg light-blue box

Some configuration parts do not reflect entirely operational configuration for demo/education purpose addressed during troubleshooting section



IP Address Management

For Lab, POC, Documentation

- Must provide **Global IP scope**: IP uniqueness among all sites and all Fabrics:
  - For default VRF
  - For SERVICES VRF that is used for interconnectivity between switches (VRFF-lite). This could be required by customer for the “in-bound Management” VRF.
  - For each endpoint subnet as they can span across all DC sites.
  - No overlapping between VRFs as many customers used VRF for security and not for solving IP space overlapping.
  - Associated Transit VLAN-ID should be as much as possible unique as they might be used to interconnect PODs. (reminder for multiple VRFs, as some platforms do not have sub-interfaces, transit VLANs are a must).
  - For all loopbacks, including VRF loopback used for DHCP request source and troubleshooting from VRF.
  - Up to 5 DCs, 8 VRFs + default
- Must provide **Local IP scope**: IP is unique on a given DC, but can be reused in an other DC:
  - For underlay interconnection (unlikely but might be used during migration) for VRF1, VRF2, VRF3. Niche but has to be planned.
  - It can be used for local Endpoint, but not so useful in our lab or documentation.
  - Associated Transit VLAN-ID can be re-used among DCs as there are no interconnect for those VRFs between sites. (reminder for multiple VRFs, as some platforms do not have sub-interfaces, transit VLANs are a must).

IPAM Example

IPAM Usage	Scope	IPv4 Range	IPv4 size	Example (1rst)	IPv6 Range	IPv6 size	Example (1rst)	VLANs
VSX keepalive	Local	192.168.0.0 – 192.168.0.255	/31	192.168.0.2/31	FD00:192:168:0::0/64	/127	FD00:192:168:0::0/127	2
DC1 – vrf default Loopback0	**Global**	192.168.1.1 – 192.168.1.255	/32	192.168.1.1/32	FD00:192:168:1::0/64	/128	FD00:192:168:1::1/128	
DC1 – vrf default Loopback1	**Global**	192.168.10.1 – 192.168.10.255	/32	192.168.10.5/32	FD00:192:168:10::0/64	/128	FD00:192:168:10::1/128	
DC1 – vrf VRF1 Loopback11 (VSX anycast DHCP) – Loopback 12 (unicast troublesh.)	**Global**	192.168.11.1 – 192.168.11.255	/32	192.168.11.5/32	FD00:192:168:11::0/64	/128	FD00:192:168:11::1/128	
DC1 – vrf VRF2 Loopback21 (VSX anycast DHCP) – Loopback 22 (unicast troublesh.)	**Global**	192.168.12.1 – 192.168.12.255	/32	192.168.12.5/32	FD00:192:168:12::0/64	/128	FD00:192:168:12::1/128	
DC1 – vrf VRF3 Loopback31 (VSX anycast DHCP) – Loopback 32 (unicast troublesh.)	**Global**	192.168.13.1 – 192.168.13.255	/32	192.168.13.5/32	FD00:192:168:13::0/64	/128	FD00:192:168:13::1/128	
DC1 – vrf default – Interco/transit SVI	**Global**	192.168.19.0 – 192.168.19.255	ROP: /31 or SVI: /29	192.168.19.0/31	FD00:192:168:19::/64	ROP: /127 or SVI: /125	FD00:192:168:19::0/127	1011-1019
DC1 – vrf VRF1 – Interco/transit SVI	Local1	192.168.110.0 – 192.168.114.254	ROP: /31 or SVI: /29	192.168.110.0/31	FD00:192:168:110::/64	ROP: /127 or SVI: /125	FD00:192:168:110::0/127	1110-1114
DC1 – vrf VRF2 – Interco/transit SVI	Local1	192.168.120.0 – 192.168.124.254	ROP: /31 or SVI: /29	192.168.120.0/31	FD00:192:168:120::/64	ROP: /127 or SVI: /125	FD00:192:168:120::0/127	1120-1124
DC1 – vrf VRF3 – Interco/transit SVI	Local1	192.168.130.0 – 192.168.134.254	ROP: /31 or SVI: /29	192.168.130.0/31	FD00:192:168:130::/64	ROP: /127 or SVI: /125	FD00:192:168:130::0/127	1130-1134
DC1 – vrf SERVICES (VRF5) – Interco/transit SVI	**Global**	192.168.210.0 – 192.168.219.254	ROP: /31 or SVI: /29	192.168.210.0/31	FD00:192:168:210::/64	ROP: /127 or SVI: /125	FD00:192:168:210::0/127	1210-1219
DC2 – vrf default Loopback0	**Global**	192.168.2.1 – 192.168.2.255	/32	192.168.2.1/32	FD00:192:168:2::0/64	/128	FD00:192:168:2::1/128	
DC2 – vrf default Loopback1	**Global**	192.168.20.1 – 192.168.20.255	/32	192.168.20.5/32	FD00:192:168:20::0/64	/128	FD00:192:168:20::1/128	
DC2 – vrf VRF1 Loopback11 (VSX anycast DHCP) – Loopback 12 (unicast troublesh.)	**Global**	192.168.21.1 – 192.168.21.255	/32	192.168.21.5/32	FD00:192:168:21::0/64	/128	FD00:192:168:21::1/128	
DC2 – vrf VRF2 Loopback21 (VSX anycast DHCP) – Loopback 22 (unicast troublesh.)	**Global**	192.168.22.1 – 192.168.22.255	/32	192.168.22.5/32	FD00:192:168:22::0/64	/128	FD00:192:168:22::1/128	
DC2 – vrf default – Interco/transit SVI	**Global**	192.168.29.0 – 192.168.29.255	ROP: /31 or SVI: /29	192.168.29.0/31	FD00:192:168:29::/64	ROP: /127 or SVI: /125	FD00:192:168:29::0/127	1021-1029
DC2 – vrf VRF1 – Interco/transit SVI	Local2	192.168.110.0 – 192.168.114.254	ROP: /31 or SVI: /29	192.168.110.0/31	FD00:192:168:110::/64	ROP: /127 or SVI: /125	FD00:192:168:110::0/127	1110-1114
DC2 – vrf VRF2 – Interco/transit SVI	Local2	192.168.120.0 – 192.168.124.254	ROP: /31 or SVI: /29	192.168.120.0/31	FD00:192:168:120::/64	ROP: /127 or SVI: /125	FD00:192:168:120::0/127	1120-1124
DC2 – vrf SERVICES (VRF5) – Interco/transit SVI	**Global**	192.168.220.0 – 192.168.229.254	ROP: /31 or SVI: /29	192.168.220.0/31	FD00:192:168:220::/64	ROP: /127 or SVI: /125	FD00:192:168:220::0/127	1220-1229
DC3 – vrf default Loopback0	**Global**	192.168.3.1 – 192.168.3.255	/32	192.168.3.1/32	FD00:192:168:3::0/64	/128	FD00:192:168:3::1/128	
DC3 – vrf default Loopback1	**Global**	192.168.30.1 – 192.168.30.255	/32	192.168.30.5/32	FD00:192:168:30::0/64	/128	FD00:192:168:30::1/128	
DC3 – vrf VRF1 Loopback11 (VSX anycast DHCP) – Loopback 12 (unicast troublesh.)	**Global**	192.168.31.1 – 192.168.31.255	/32	192.168.31.5/32	FD00:192:168:31::0/64	/128	FD00:192:168:31::1/128	
DC3 – vrf default – Interco/transit SVI	**Global**	192.168.39.0 – 192.168.39.255	ROP: /31 or SVI: /29	192.168.39.0/31	FD00:192:168:39::/64	ROP: /127 or SVI: /125	FD00:192:168:39::0/127	1031-1139
DC3 – vrf VRF1 – Interco/transit SVI	Local3	192.168.110.0 – 192.168.114.254	ROP: /31 or SVI: /29	192.168.110.0/31	FD00:192:168:110::/64	ROP: /127 or SVI: /125	FD00:192:168:110::0/127	1110-1114
DC3 – vrf VRF2 – Interco/transit SVI	Local3	192.168.120.0 – 192.168.124.254	ROP: /31 or SVI: /29	192.168.120.0/31	FD00:192:168:120::/64	ROP: /127 or SVI: /125	FD00:192:168:120::0/127	1120-1124
DC3 – vrf SERVICES (VRF5) – Interco/transit SVI	**Global**	192.168.230.0 – 192.168.239.254	ROP: /31 or SVI: /29	192.168.230.0/31	FD00:192:168:230::/64	ROP: /127 or SVI: /125	FD00:192:168:230::0/127	1230-1239
DC - vrf default – endpoint/client/VMs subnet	**Global**	10.0.0.0/16	/24	10.0.5.0/24	FD00:10:0:0::/48	/64	FD00:10:0:5::/64	**5-9**
DC - vrf VRF1 – endpoint/client/VMs subnet	**Global**	10.1.0.0/16	/24	10.1.10.0/24	FD00:10:1:0::/48	/64	FD00:10:1:10::/64	**10-19**
DC - vrf VRF2 – endpoint/client/VMs subnet	**Global**	10.2.0.0/16	/24	10.2.20.0/24	FD00:10:2:0::/48	/64	FD00:10:2:20::/64	**20-29**
DC - vrf VRF3 – endpoint/client/VMs subnet	**Global**	10.3.0.0/16	/24	10.3.30.0/24	FD00:10:3:0::/48	/64	FD00:10:3:30::/64	**30-39**
DC - vrf VRF4 – endpoint/client/VMs subnet	**Global**	10.4.0.0/16	/24	10.4.40.0/24	FD00:10:4:0::/48	/64	FD00:10:4:40::/64	**40-49**
DC - vrf SERVICES (VRF5) – endpoint/client/VMs subnet	**Global**	10.5.0.0/16	/24	10.5.50.0/24	FD00:10:5:0::/48	/64	FD00:10:5:50::/64	**50-59**

VLAN ID = 1000+ 3rd digit group

REMINDER

Fabric1 Spines Configuration

- Underlay ROPs, loopbacks
- OSPF underlay routing
- iBGP Route-Reflectors for EVPN AF

S1 (Spine-1) Configuration

ROPs to Leafs / Loopback

```
interface 1/1/25
    no shutdown
    mtu 9198
    routing
    ip mtu 9198
    ip address 192.168.19.0/31
    ip ospf 1 area 0.0.0.0
    no ip ospf passive
    ip ospf network point-to-point
    ip ospf authentication message-digest
    ip ospf message-digest-key 1 md5 ciphertext AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
interface 1/1/26
    no shutdown
    mtu 9198
    routing
    ip mtu 9198
    ip address 192.168.19.2/31
    ip ospf 1 area 0.0.0.0
    no ip ospf passive
    ip ospf network point-to-point
    ip ospf authentication message-digest
    ip ospf message-digest-key 1 md5 ciphertext AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
interface 1/1/27
    no shutdown
    mtu 9198
    routing
    ip mtu 9198
    ip address 192.168.19.4/31
    ip ospf 1 area 0.0.0.0
    no ip ospf passive
    ip ospf network point-to-point
    ip ospf authentication message-digest
    ip ospf message-digest-key 1 md5 ciphertext AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
interface 1/1/28
    no shutdown
    mtu 9198
    routing
    ip mtu 9198
    ip address 192.168.19.6/31
    ip ospf 1 area 0.0.0.0
    no ip ospf passive
    ip ospf network point-to-point
    ip ospf authentication message-digest
    ip ospf message-digest-key 1 md5 ciphertext AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
!
interface loopback 0
    ip address 192.168.1.1/32
```

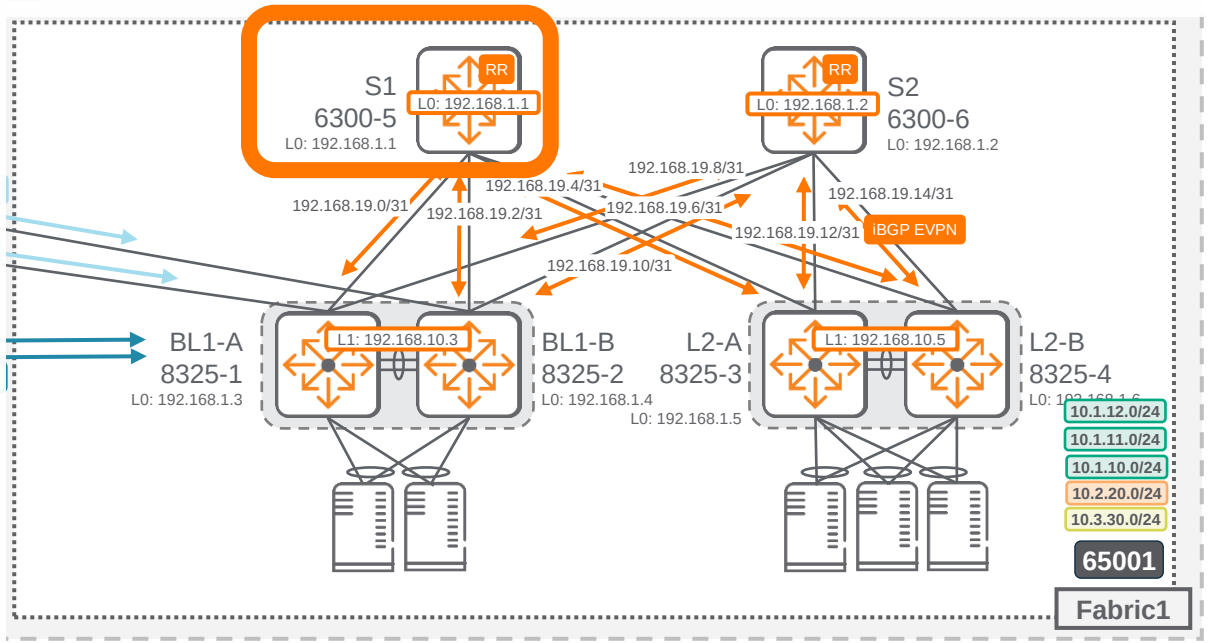
L3 Link to BL1-A (8325-1)

L3 Link to BL1-B (8325-2)

L3 Link to L2-A (8325-3)

L3 Link to L2-B (8325-4)

Loopback0



S1 (Spine-1) Configuration

Routing

```
route-map connected-ospf permit seq 10
     set tag 1000
!
router ospf 1
    router-id 192.168.1.1
    ! optional
    max-metric router-lsa include-stub on-startup 300
    passive-interface default
    !
    redistribute local loopback route-map connected-ospf
    area 0.0.0.0

router bgp 65001
    bgp router-id 192.168.1.1
    ! optional
    trap-enable
    bgp log-neighbor-changes
    !
    bgp fast-external-fallover
    !
    bgp deterministic-med
    bgp always-compare-med
    !
    neighbor leaf peer-group
    neighbor leaf remote-as 65001
    neighbor leaf description Leaf RR clients
    ! optional
    neighbor leaf password ciphertext
AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
    neighbor leaf fall-over
    !
    neighbor leaf update-source loopback 0
    neighbor 192.168.1.3 peer-group leaf
    neighbor 192.168.1.4 peer-group leaf
    neighbor 192.168.1.5 peer-group leaf
    neighbor 192.168.1.6 peer-group leaf
    address-family l2vpn evpn
        neighbor leaf route-reflector-client
        neighbor leaf send-community both
        neighbor 192.168.1.3 activate
        neighbor 192.168.1.4 activate
        neighbor 192.168.1.5 activate
        neighbor 192.168.1.6 activate
    exit-address-family
```

tag value defined for local Fabric loopback

Optional best practices

loopback redist. in OSPF

Optional best practices

Enabled by default

If used, must be the same in the BGP domain

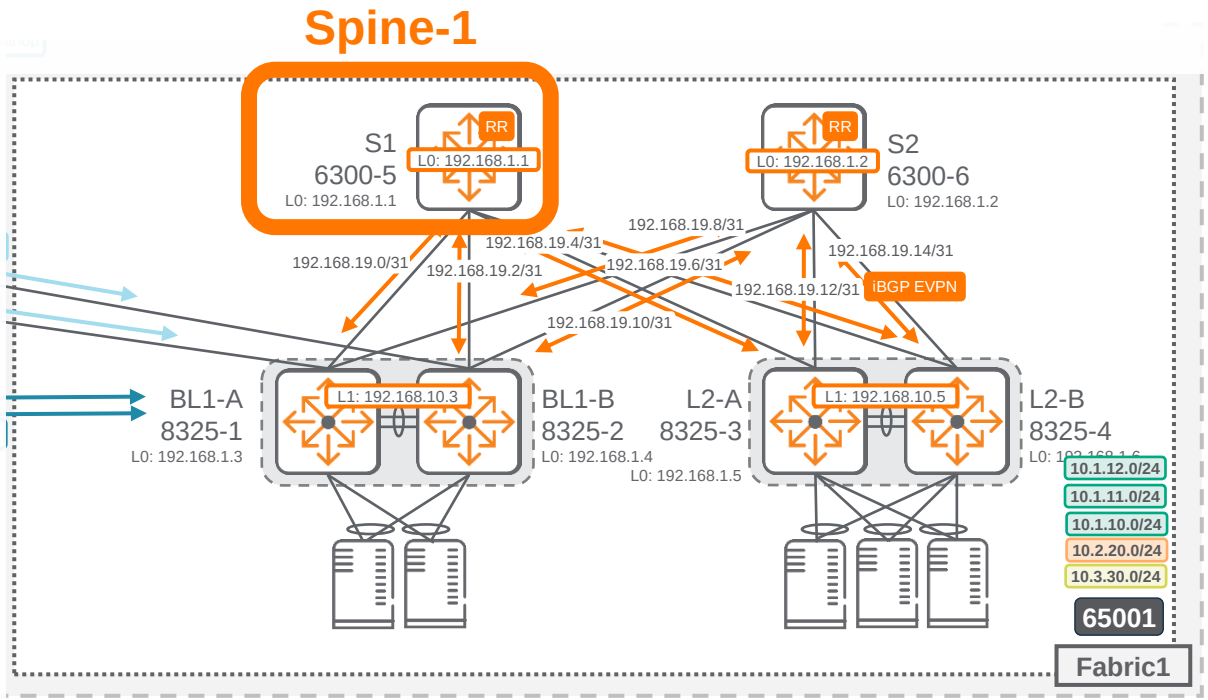
Spines are iBGP RR

Fall-over notification from FIB to BGP

iBGP peering between loopbacks

Extended community is required for EVPN NLRI (both includes extended)

- Reminder: tag value is carried in OSPF LSA.



S2 (Spine-2) Configuration

ROPs to Leafs / Loopback

```
interface 1/1/25
    no shutdown
    mtu 9198
    routing
    ip mtu 9198
    ip address 192.168.19.8/31
    ip ospf 1 area 0.0.0.0
    no ip ospf passive
    ip ospf network point-to-point
    ip ospf authentication message-digest
    ip ospf message-digest-key 1 md5 ciphertext AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
interface 1/1/26
    no shutdown
    mtu 9198
    routing
    ip mtu 9198
    ip address 192.168.19.10/31
    ip ospf 1 area 0.0.0.0
    no ip ospf passive
    ip ospf network point-to-point
    ip ospf authentication message-digest
    ip ospf message-digest-key 1 md5 ciphertext AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
interface 1/1/27
    no shutdown
    mtu 9198
    routing
    ip mtu 9198
    ip address 192.168.19.12/31
    ip ospf 1 area 0.0.0.0
    no ip ospf passive
    ip ospf network point-to-point
    ip ospf authentication message-digest
    ip ospf message-digest-key 1 md5 ciphertext AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
interface 1/1/28
    no shutdown
    mtu 9198
    routing
    ip mtu 9198
    ip address 192.168.19.14/31
    ip ospf 1 area 0.0.0.0
    no ip ospf passive
    ip ospf network point-to-point
    ip ospf authentication message-digest
    ip ospf message-digest-key 1 md5 ciphertext AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
!
interface loopback 0
    ip address 192.168.1.2/32
```

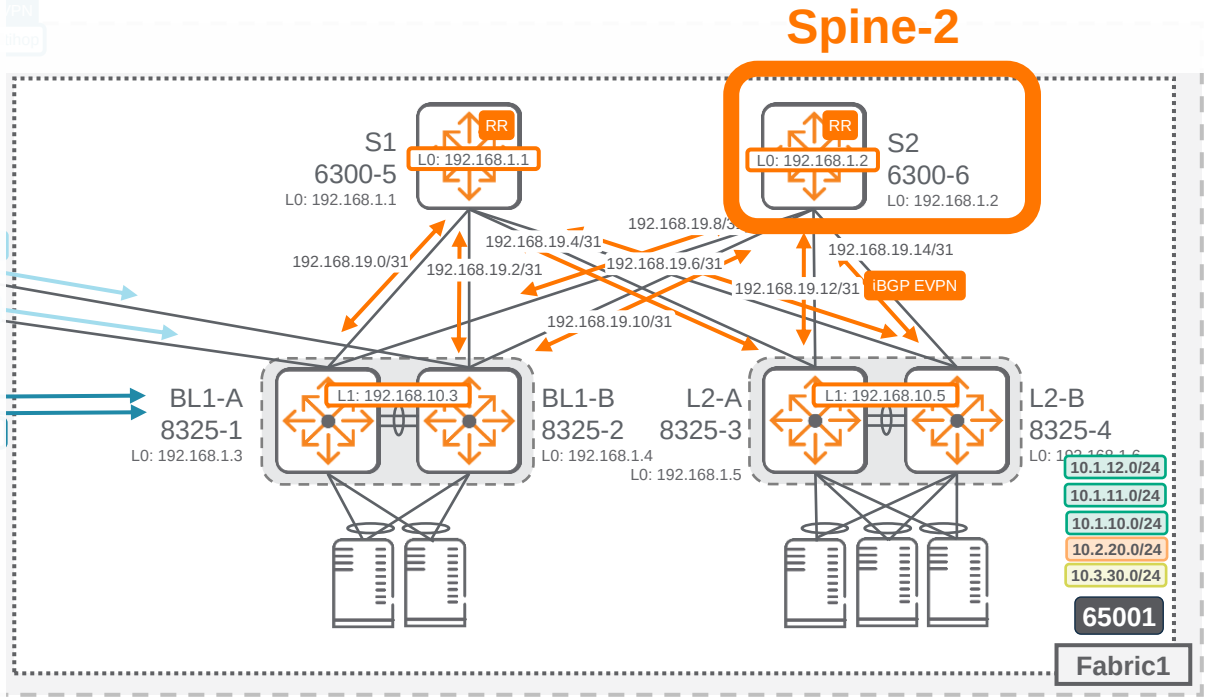
L3 Link to BL1-A (8325-1)

L3 Link to BL1-B (8325-2)

L3 Link to L2-A (8325-3)

L3 Link to L2-B (8325-4)

Loopback0



S2 (Spine-2) Configuration

Routing

```
route-map connected-ospf permit seq 10
     set tag 1000
!
router ospf 1
    router-id 192.168.1.2
    ! optional
    max-metric router-lsa include-stub on-startup 300
    passive-interface default
    !
    redistribute local loopback route-map connected-ospf
    area 0.0.0.0

router bgp 65001
    bgp router-id 192.168.1.2
    ! optional
    trap-enable
    bgp log-neighbor-changes
    !
    bgp fast-external-fallover
    !
    bgp deterministic-med
    bgp always-compare-med
    !
    neighbor leaf peer-group
    neighbor leaf remote-as 65001
    neighbor leaf description Leaf RR clients
    ! optional
    neighbor leaf password ciphertext
AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
    neighbor leaf fall-over
    !
    neighbor leaf update-source loopback 0
    neighbor 192.168.1.3 peer-group leaf
    neighbor 192.168.1.4 peer-group leaf
    neighbor 192.168.1.5 peer-group leaf
    neighbor 192.168.1.6 peer-group leaf
    address-family l2vpn evpn
        neighbor leaf route-reflector-client
        neighbor leaf send-community both
        neighbor 192.168.1.3 activate
        neighbor 192.168.1.4 activate
        neighbor 192.168.1.5 activate
        neighbor 192.168.1.6 activate
    exit-address-family
```

tag value defined for local Fabric loopback

Optional best practices

loopback redist. in OSPF

Optional best practices

Enabled by default

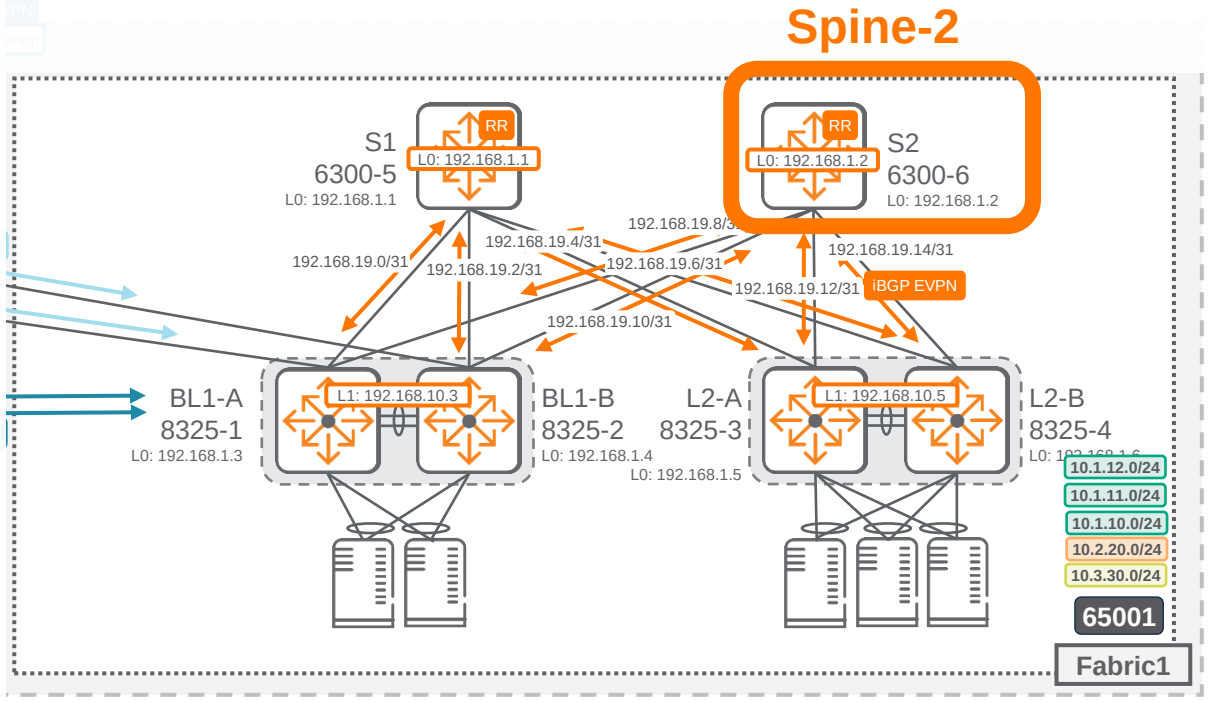
If used, must be the same in the BGP domain

Spines are iBGP RR

Fall-over notification from FIB to BGP

iBGP peering between loopbacks

Extended community is required for EVPN NLRI (both includes extended)



REMINDER

Fabric1 Leaves Configuration

- VSX / Servers VSX LAGs
- Underlay ROPs, Loopbacks, transit VLAN
- VRFs and EVPN
- Servers SVIs, DHCP
- VXLAN Tunnel, VLAN-to-VNI mapping, L3VNI per VRF
- OSPF underlay routing
- iBGP EVPN AF
- iBGP IPv4 AF for Route-Type 5

BL1-A (Leaf-1 VSX primary) Configuration

VSX Configuration / VLANs

```
vrf KA
!
interface lag 256
    vsx-sync vlans
    no shutdown
    description ISL
    no routing
    vlan trunk native 1 tag
    vlan trunk allowed all
    lacp mode active
interface 1/1/41
    no shutdown
    vrf attach KA
    description 8325-2 1/1/41 for keepalive
    ip address 192.168.0.0/31
interface 1/1/49
    no shutdown
    mtu 9198
    description VSX ISL link
    lag 256
interface 1/1/50
    no shutdown
    mtu 9198
    description VSX ISL link
    lag 256
!
vsx
    system-mac 02:00:00:00:01:00
    inter-switch-link lag 256
    role primary
    keepalive peer 192.168.0.1 source 192.168.0.0 vrf KA
    vsx-sync aaa bgp copp-policy dhcp-relay dns evpn l2-vlan-mac-cfg-mode mclag-interfaces
qos-global route-map sflow-global snmp ssh stp-global time vsx-global
!
vlan 1019
    vsx-sync
vlan 10
    vsx-sync
vlan 11
    vsx-sync
vlan 20
    vsx-sync
```

ISL LAG for VSX ISLP

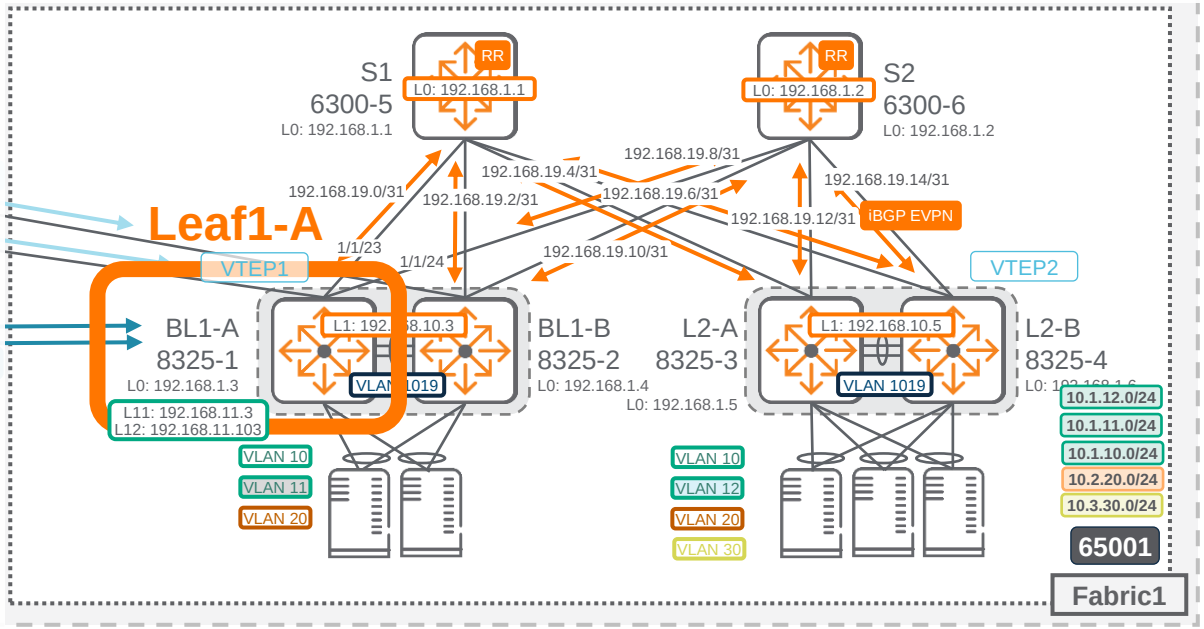
Keepalive circuit (optional, alternative: routed KA)

ISL physical links

Follow best practices for system-mac value

Transit VLAN between VSX nodes for IGP continuity

Servers VLANs



BL1-A (Leaf-1 VSX primary) Configuration

VSX LAGs to Servers

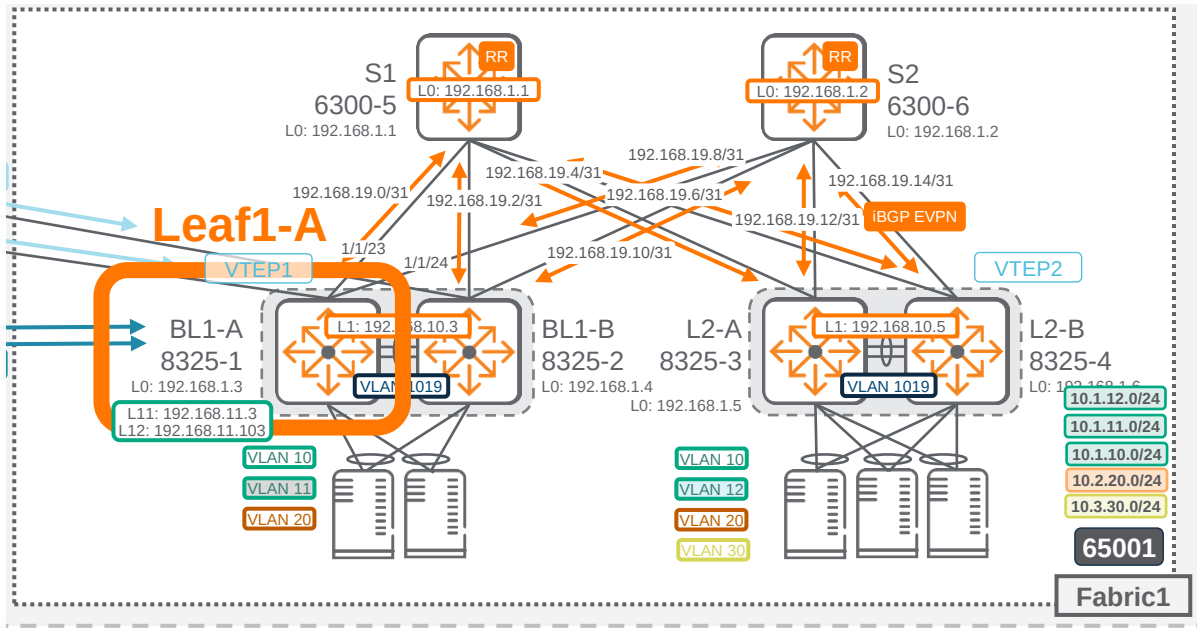
```
interface lag 1 multi-chassis
    no shutdown
    description 2930-3
    no routing
    vlan trunk native 1
    vlan trunk allowed 10-11
    lacp mode active
    hash l4-src-dst
!
interface lag 2 multi-chassis
    no shutdown
    description 2930-4
    no routing
    vlan trunk native 1
    vlan trunk allowed 20
    lacp mode active
!
interface 1/1/1
    no shutdown
    mtu 9000
    description Link to 2930-3 (49) - LAG 1 member
    lag 1
!
interface 1/1/2
    no shutdown
    mtu 9000
    description Link to 2930-3 (49) - LAG 2 member
    lag 2
```

VSX LAG to server

Optional L4 hashing

VSX LAG to server

Links to servers



BL1-A (Leaf-1 VSX primary) Configuration

ROPs to Spines / Loopbacks / Transit VLAN

```
interface 1/1/23
    no shutdown
    mtu 9198
    ip mtu 9198
    ip address 192.168.19.1/31
    ip ospf 1 area 0.0.0.0
    no ip ospf passive
    ip ospf network point-to-point
    ip ospf authentication message-digest
    ip ospf message-digest-key 1 md5 ciphertext
AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
!
interface 1/1/24
    no shutdown
    mtu 9198
    ip mtu 9198
    ip address 192.168.19.9/31
    ip ospf 1 area 0.0.0.0
    no ip ospf passive
    ip ospf network point-to-point
    ip ospf authentication message-digest
    ip ospf message-digest-key 1 md5 ciphertext
AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
!
interface loopback 0
    description unicast default loopback for control-plane
    ip address 192.168.1.3/32
!
interface loopback 1
    description VSX anycast loopback for VXLAN sourcing in default VRF
    ip address 192.168.10.3/32
!
interface vlan 1019
    ip mtu 9198
    ip address 192.168.19.200/31
    ip ospf 1 area 0.0.0.0
    no ip ospf passive
    ip ospf cost 50
    ip ospf network point-to-point
    ip ospf authentication message-digest
    ip ospf message-digest-key 1 md5 ciphertext
AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
```

L3 Link to S1 (6300-5)

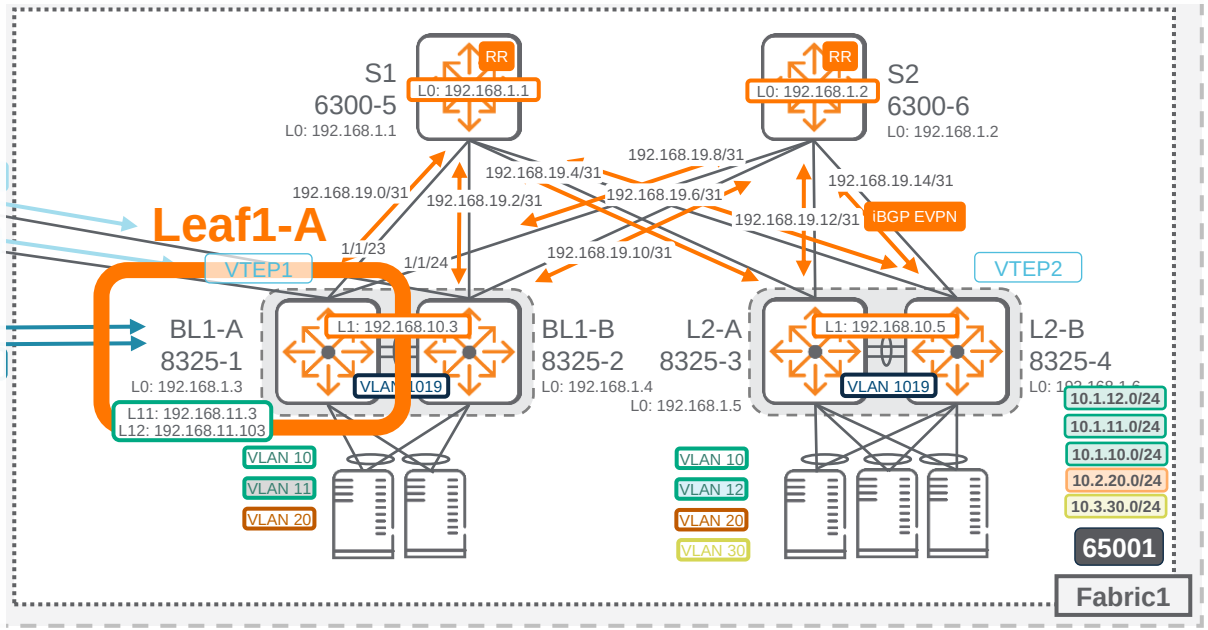
L3 Link to S2 (6300-6)

Unique Loopback0 on each node

Shared anycast Loopback1 of the VSX logical VTEP pair

Transit VLAN SVI between VSX nodes for IGP continuity

Higher OSPF cost than uplinks



BL1-A (Leaf-1 VSX primary) Configuration

VRFs / EVPN / Distinct VRF RD option

```
vrf VRF1
    rd 192.168.1.3:1
    route-target export 65001:1 evpn
    route-target import 65001:1 evpn
vrf VRF2
    rd 192.168.1.3:2
    route-target export 65001:2 evpn
    route-target import 65001:2 evpn
!
evpn
    arp-suppression
    nd-suppression
    redistribute local-svi
    vlan 10
        rd auto
        route-target export 65001:10
        route-target import 65001:10
        redistribute host-route
    vlan 11
        rd auto
        route-target export 65001:11
        route-target import 65001:11
        redistribute host-route
    vlan 20
        rd auto
        route-target export 65001:20
        route-target import 65001:20
        redistribute host-route
```

RD is unique per VRF for each VTEP
Recommendation value: [L1:VRF_ID]
in this demo [L0:VRF_ID] (educative purpose)

RT proposal: [AS_number:VRF_ID]

ARP / ND caching to limit broadcast traffic

local-svi is useful for troubleshooting when SVI IP != AG IP

RD auto is: [VTEP_L1:VLAN_ID]

RT can be set to [AS_number:VLAN_ID]

Host-route injection

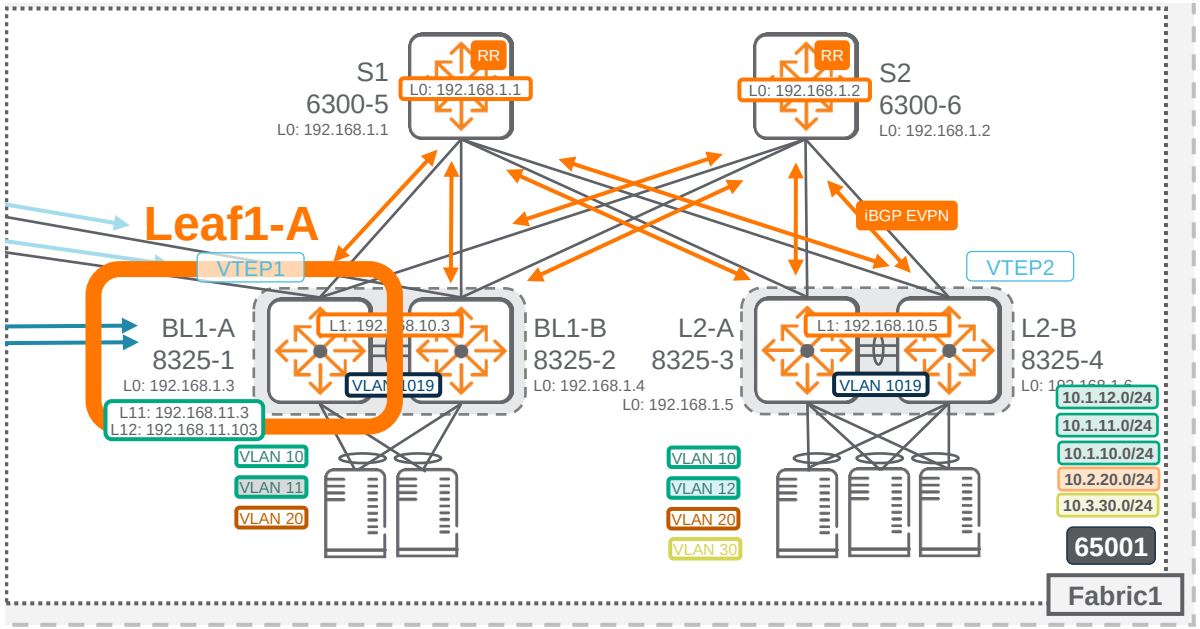
Warning: RT should not be set to auto
- RT auto works only with 2-byte ASN
- RT auto is not suitable for eBGP model

redistribute host-route is used to advertise /32 host routes to remote VTEPs

Example: 10.1.10.10/32 is connected to Leaf1-A/Leaf1-B
10.1.12.14/32 is connected to Leaf2-A/Leaf2-B

This command is required as the same subnet is spread across multiple VTEPs.
Without host routes, a remote VTEP would not be aware of the correct destination VTEP to send traffic to.

- **Distinct RD** per CX switch in VSX cluster or **common RD** in the VSX cluster.
  Both options were validated.
- **Distinct RD** offers more granular troubleshooting.
- **Common RD is preferred** as it optimizes the RIB table size.



BL1-A (Leaf-1 VSX primary) Configuration

Endpoint SVIs / DHCP relay

```
interface vlan 10
    vsx-sync active-gateways
    vrf attach VRF1
    ip mtu 9000
    ip address 10.1.10.1/24
    active-gateway ip mac 12:00:00:00:01:00
    active-gateway ip 10.1.10.1
    ipv6 address fd00:10:1:10::1/64
    active-gateway ipv6 mac 12:00:00:00:01:00
    active-gateway ipv6 fe80:10:1:10::1
    no ipv6 nd suppress-ra
    ipv6 nd router-preference high
    ip helper-address 10.10.129.30 vrf default
!
interface vlan 11
    vsx-sync active-gateways
    vrf attach VRF1
    ip mtu 9000
    ip address 10.1.11.1/24
    active-gateway ip mac 12:00:00:00:01:00
    active-gateway ip 10.1.11.1
    ipv6 address fd00:10:1:11::1/64
    active-gateway ipv6 mac 12:00:00:00:01:00
    active-gateway ipv6 fe80:10:1:11::1
    no ipv6 nd suppress-ra
    ipv6 nd router-preference high
!
interface vlan 20
    vsx-sync active-gateways
    vrf attach VRF2
    ip mtu 9000
    ip addess 10.2.20.2/24
    active-gateway ip mac 12:00:00:00:01:00
    active-gateway ip 10.2.20.1
!
ip source-interface dhcp_relay interface loopback1
dhcp-relay option 82 source-interface
dhcp-relay option 82 replace
```

SVI IP = AG IP best practice

Follow best practices for AG MAC value

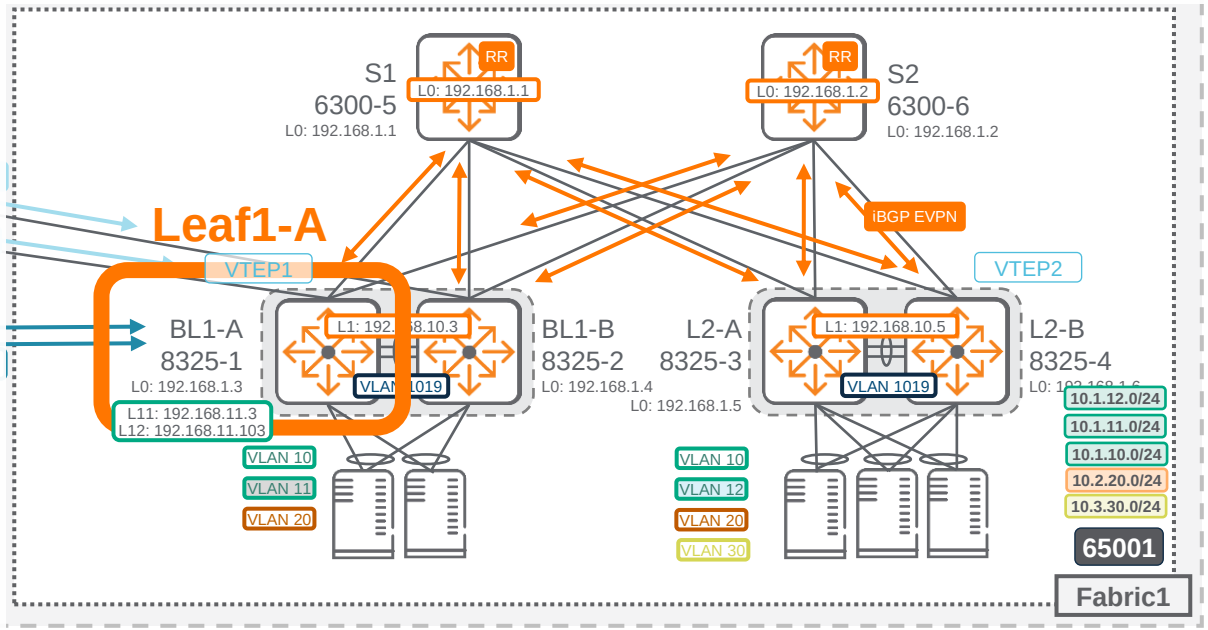
IPv6 AG must be a link-local address as this is the default gateway used by roaming clients.

Overlay DHCP relay with DHCP server in underlay VRF

SVI IP != AG IP for demo purpose in VRF2

DHCP relay sourcing from anycast loopback in default VRF

- Configuration simplification with same IPv4 address for both SVI and Active Gateway is intended for 10.09 maintenance release. (under validation)
- For IPv6 (under validation):
  - same Link-Local-Address for AG IPv6 on all VTEPs
  - same Global-Unicast-Address SVI IPv6 on all VTEPs (and common between VSX primary and secondary)



BL1-A (Leaf-1 VSX primary) Configuration

Tenant VRF Loopbacks / Tenant VRF Transit VLAN

Optional: VSX Anycast IP for DHCP sourcing in the DHCP server tenant VRF

```
interface loopback 11
    vrf attach VRF1
    description for sourcing DHCP-relay to DHCP-server located in VRF1
    ip address 192.168.11.103/32
interface loopback 12
    vrf attach VRF1
    description for troubleshooting in VRF1
    ip address 192.168.11.3/32
interface loopback 21
    vrf attach VRF2
    description for sourcing DHCP-relay to DHCP-server located in VRF2
    ip address 192.168.12.103/32
interface loopback 22
    vrf attach VRF2
    description for troubleshooting in VRF2
    ip address 192.168.12.3/32
!
ip source-interface dhcp_relay interface loopback11 vrf VRF1
vlan 1110
    vsx-sync
interface vlan 1110
    vrf attach VRF1
    description VRF1 transit for loopbacks in VSX
    ip mtu 9198
    ip address 192.168.110.0/31
ip route 192.168.11.4/32 192.168.110.1 vrf VRF1
```

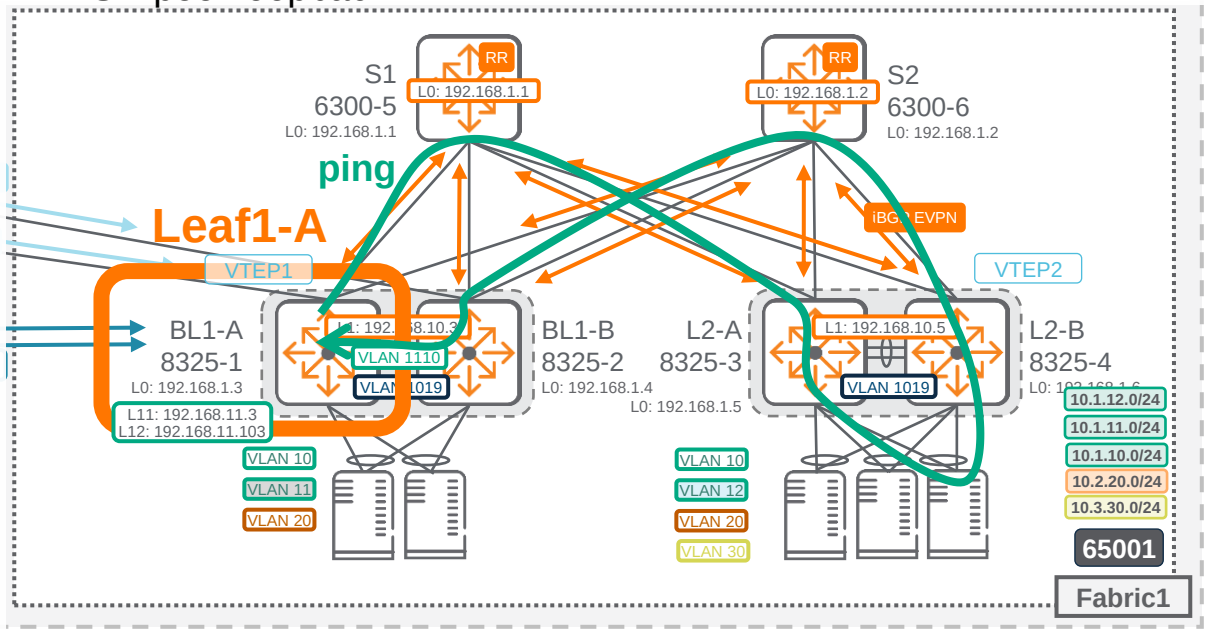
Per VRF, unicast IP for troubleshooting

Optional: for inter-VRF DHCP-relay use-case with DHCP server in tenant VRF

Similar config on Leaf-2 VSX secondary

```
interface loopback 11
    vrf attach VRF1
    description for sourcing DHCP-relay to DHCP-server located in VRF1
    ip address 192.168.11.103/32
interface loopback 12
    vrf attach VRF1
    description for troubleshooting in VRF1
    ip address 192.168.11.4/32
...
ip source-interface dhcp_relay interface loopback11 vrf VRF1
vlan 1110
    vsx-sync
interface vlan 1110
    vrf attach VRF1
    description VRF1 transit for loopbacks in VSX
    ip mtu 9198
    ip address 192.168.110.1/31
ip route 192.168.11.3/32 192.168.110.0 vrf VRF1
```

- On VSX VTEPs, one additional anycast loopback is recommended in case of inter-VRF DHCP relay use-case. This anycast source IP address should be configured in the same VRF than the DHCP server located in the tenant VRF.
- **On all VTEPs**, one additional unicast loopback is recommended per VRF for troubleshooting (ping & traceroute from unicast IP) in the tenant VRF, as the SVI IP is common on all VTEPs.
- On VSX VTEPs, one **additional transit SVI per VRF** is needed for the reachability of the VSX-peer VRF unicast loopback so that the troubleshooting return packet can hit the proper VSX node, thanks to the associated transit VLAN carried over the ISL.
  In addition, an associated static route is set-up per VRF for the VSX-peer loopback.



BL1-A (Leaf-1 VSX primary) Configuration

VXLAN interface / Virtual-MAC

```
interface vxlan 1
    source ip 192.168.10.3
    vxlan-counters aggregate
    no shutdown
    vni 10010
        vlan 10
    vni 10011
        vlan 11
    vni 10020
        vlan 20
    vni 100001
        vrf VRF1
        routing
    vni 100002
        vrf VRF2
        routing
!
virtual-mac 02:00:00:00:01:00
```

VXLAN tunnel source IP = anycast shared Loopback1

Optional counters for VXLAN statistics

1-to-1 VLAN-VNI mapping

L3 VNI associated to vrf VRF1

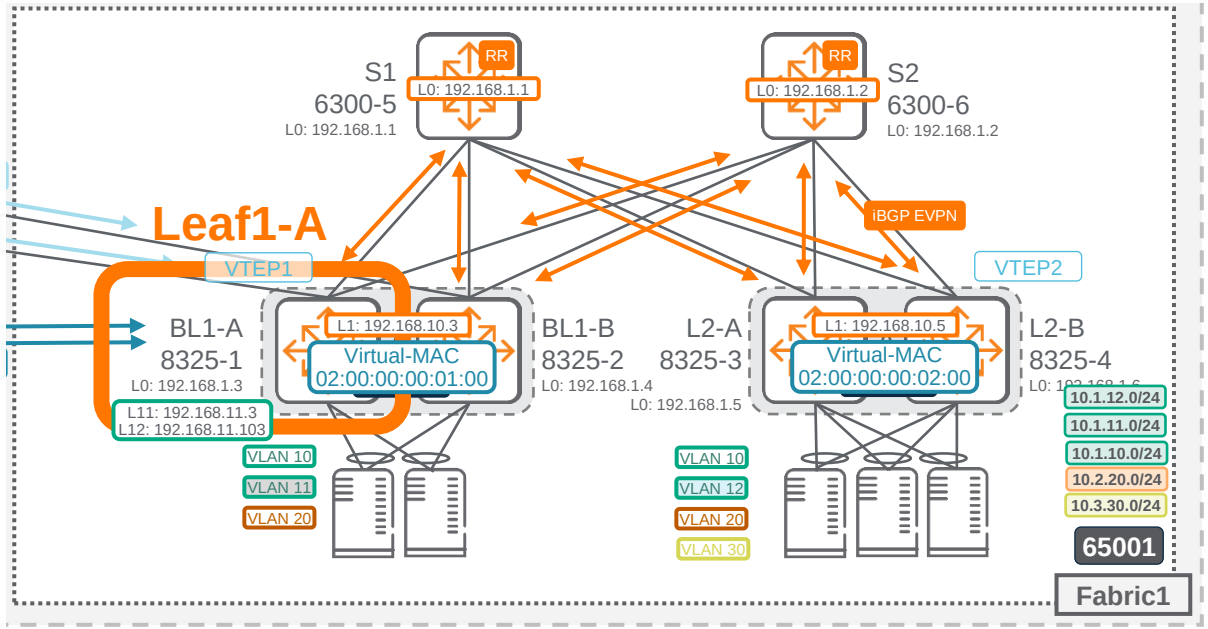
L3 VNI associated to vrf VRF2

VTEP virtual-MAC for L3VNI functionality

- For single fabric: configure local VLANs per VTEP
- For multi-fabric: add remote VLANs, non-local to the border-VTEP, for VLAN extension over multiple fabrics (Ex: vni 10012 / vlan 12 binding to be added here)

virtual-mac is <u>required</u> for symmetric IRB Distributed L3 Gateways deployments.
This is used as the Destination Router MAC for L3VNI routing to reach the egress VTEP router.
On VSX logical VTEPs, **the same virtual-mac should be used on both nodes of the same VSX logical VTEP pair**.
Different virtual-macs should be used on every VTEP or VSX logical VTEP.

L3VNI is <u>optional</u>. Deployment might not necessary required routing. For instance, L3 overlay Routing might be done in some cases by a pair of Firewalls (pure L2 EVPN Fabric).



BL1-A (Leaf-1 VSX primary) Configuration

Routing

```
ip prefix-list endpoint-VRF1 seq 10 permit 10.1.0.0/16 le 32
ip prefix-list endpoint-VRF2 seq 10 permit 10.2.0.0/16 le 32
ipv6 prefix-list v6-endpoint-VRF1 seq 10 permit fd00:10:1:10::/64 le 128
!
route-map connected-bgp-VRF1 permit seq 10
     match ip address prefix-list endpoint-VRF1
route-map connected-bgp-VRF1 permit seq 20
     match ipv6 address prefix-list v6-endpoint-VRF1
route-map connected-bgp-VRF2 permit seq 10
     match ip address prefix-list endpoint-VRF2
!
route-map connected-ospf permit seq 10
     set tag 1000
!
router bgp 65001
    bgp router-id 192.168.1.3
    trap-enable
    bgp fast-external-fallover
    bgp log-neighbor-changes
    bgp deterministic-med
    bgp always-compare-med
    neighbor spine-RR peer-group
    neighbor spine-RR remote-as 65001
    neighbor spine-RR description Spine and RR peer-group
    neighbor spine-RR password ciphertext
AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
    neighbor spine-RR fall-over
    neighbor spine-RR update-source loopback 0
    neighbor 192.168.1.1 peer-group spine-RR
    neighbor 192.168.1.2 peer-group spine-RR
    address-family l2vpn evpn
        neighbor spine-RR send-community both
        neighbor 192.168.1.1 activate
        neighbor 192.168.1.2 activate
    exit-address-family
!
    vrf VRF1
        bgp router-id 192.168.1.3
        address-family ipv4 unicast
            redistribute local loopback
            redistribute connected route-map connected-bgp-VRF1
        exit-address-family
        address-family ipv6 unicast
            redistribute local loopback
            redistribute connected route-map connected-bgp-VRF1
        exit-address-family
!
    vrf VRF2
        bgp router-id 192.168.1.3
        address-family ipv4 unicast
            redistribute local loopback
            redistribute connected route-map connected-bgp-VRF2
        exit-address-family
```

Control for route injection

tag value defined for local Fabric loopback

If used, must be the same in the BGP domain

Fall-over notification from FIB to BGP

iBGP peering between loopbacks

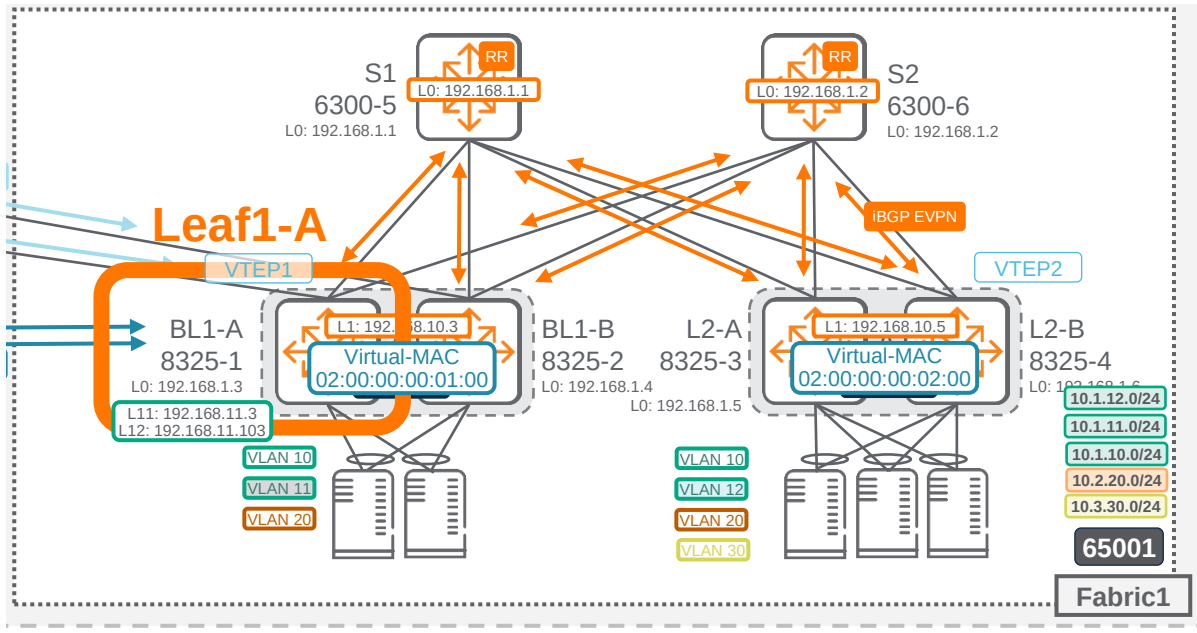
Extended community is required

Used to populate EVPN Type-5 IP prefix routes

```
router ospf 1
    router-id 192.168.1.3
    max-metric router-lsa include-stub on-startup 300
    passive-interface default
    redistribute local loopback route-map connected-ospf
    trap-enable
    area 0.0.0.0
```

best practice for HA during reboot

loopback redistribution in OSPF



NEW

Fabric1 Border Configuration

Additional Configuration for intra-site communication (2 Fabrics)

- ROPs, underlay eBGP IPv4 AF and loopbacks routing
- eBGP EVPN AF
- VXLAN Tunnels Bridging

BL1-A (Border-Leader-1 VSX Primary) Configuration

ROPs, underlay eBGP IPv4 AF and loopbacks routing

```
interface 1/1/51
    no shutdown
    mtu 9198
    description 8400-1 1/3/3
    ip mtu 9198
    ip address 192.168.29.7/31
!
route-map BGP-OSPF deny seq 10
     match tag 1000
route-map BGP-OSPF permit seq 20
!
route-map OSPF-BGP permit seq 10
     match tag 1000
!
router ospf 1
    router-id 192.168.1.3
    max-metric router-lsa include-stub on-startup 240
    passive-interface default
    redistribute bgp route-map BGP-OSPF
    redistribute local loopback route-map connected-ospf
    area 0.0.0.0
!
router bgp 65001
    bgp router-id 192.168.1.3
    ...
    neighbor 192.168.29.6 remote-as 65002
    neighbor 192.168.29.6 description Fabric2 S3
    neighbor 192.168.29.6 password ciphertext
AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
    address-family ipv4 unicast
        neighbor 192.168.29.6 activate
        redistribute local loopback
        redistribute ospf 1 route-map OSPF-BGP
    exit-address-family
```

L3 Link to S3 (8400-1)

Control of BGP routes injected to OSPF: to avoid cross-injection on the border-VTEP

Control of OSPF routes injected to BGP: only local-Fabric loopbacks

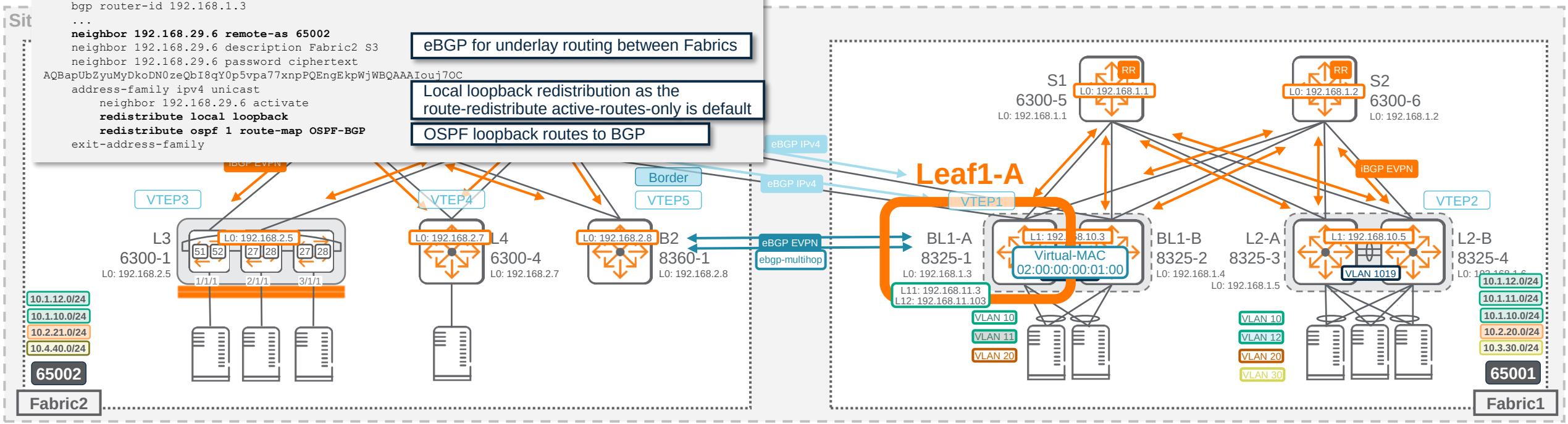
BGP underlay routes to OSPF underlay

eBGP for underlay routing between Fabrics

Local loopback redistribution as the route-redistribute active-routes-only is default

OSPF loopback routes to BGP

- Only local-AS loopbacks are redistributed from OSPF into BGP IPv4 AF (in underlay default VRF).
- Local-border-VTEP loopback must be redistributed directly into BGP IPV4 AF (in underlay default VRF) due to "route-redistribute active-routes-only".
- BGP IPv4 underlay routes are injected into OSPF except the local loopbacks.
- Similar configuration is done on remote Fabrics. (here Fabric2 as underlay interconnectivity path)



BL1-A (Border-Leader-1 VSX Primary) Configuration

eBGP EVPN AF: next-hop-self, peering with other border-VTEPs

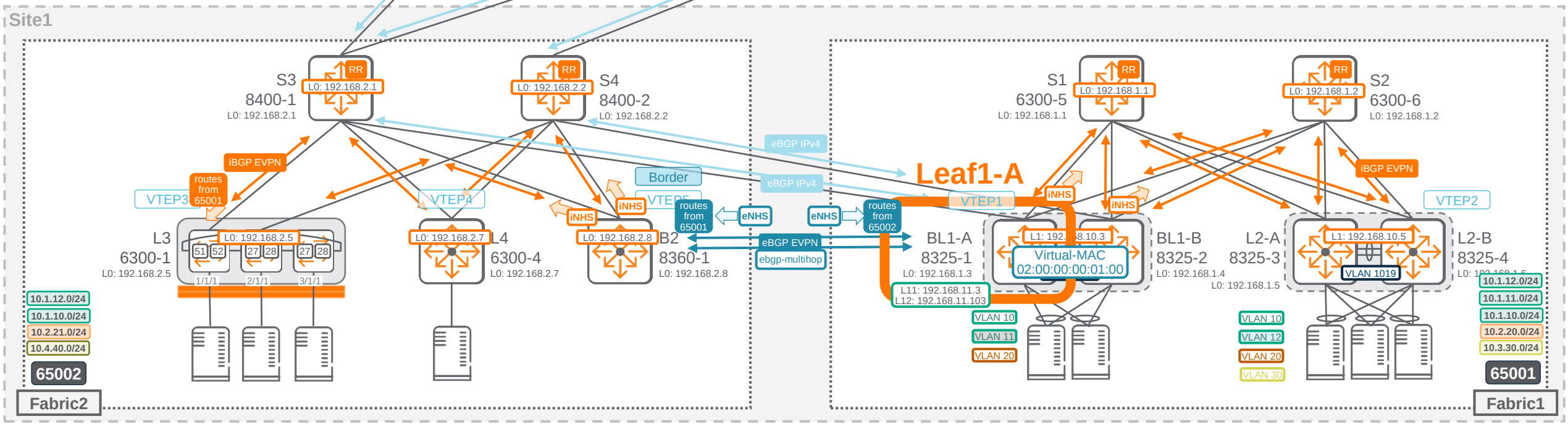
```
router bgp 65001
    bgp router-id 192.168.1.3
    ...
    neighbor borders peer-group
    neighbor borders description eBGP EVPN peering with Fabrics
    neighbor borders password ciphertext AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
    neighbor borders fall-over
    neighbor borders update-source loopback 0
    neighbor 192.168.2.8 remote-as 65002
    neighbor 192.168.2.8 peer-group borders
    neighbor 192.168.2.8 description Fabric2 Border2
    neighbor 192.168.2.8 ebgp-multihop 10
!
    address-family l2vpn evpn
        neighbor spine-RR next-hop-self
        neighbor spine-RR send-community both
        neighbor borders send-community both
        neighbor 192.168.2.8 activate
exit-address-family
```

local border VTEP eBGP peering over loopbacks (knowns from eBGP IPv4 AF)

ebgp-multihop has to be set per neighbor as the remote-as is not defined in the peer-group

Next-Hop for eBGP routes is reset to Border2 towards Fabric2 RRs — iNHS

- Number of intermediate/transit hops shall be adjusted



BL1-A (Border-Leader-1 VSX Primar

NHS outcome: eBGP routes to iBGP routes

```
router bgp 65001
    bgp router-id 192.168.1.3
    ...
    neighbor borders peer-group
    neighbor borders description eBGP EVPN peering with Fabrics
    neighbor borders password ciphertext AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xn...PQEngEkpWjWBQAAAIouj7OC
    neighbor borders fall-over
    neighbor borders update-source loopback 0
    neighbor 192.168.2.8 remote-as 65002
    neighbor 192.168.2.8 peer-group borders
    neighbor 192.168.2.8 description Fabric2 Border2
    neighbor 192.168.2.8 ebgp-multihop 10
!
    address-family l2vpn evpn
        neighbor spine-RR next-hop-self
        neighbor spine-RR send-community both
        neighbor borders send-community both
        neighbor 192.168.2.8 activate
exit-address-family
```

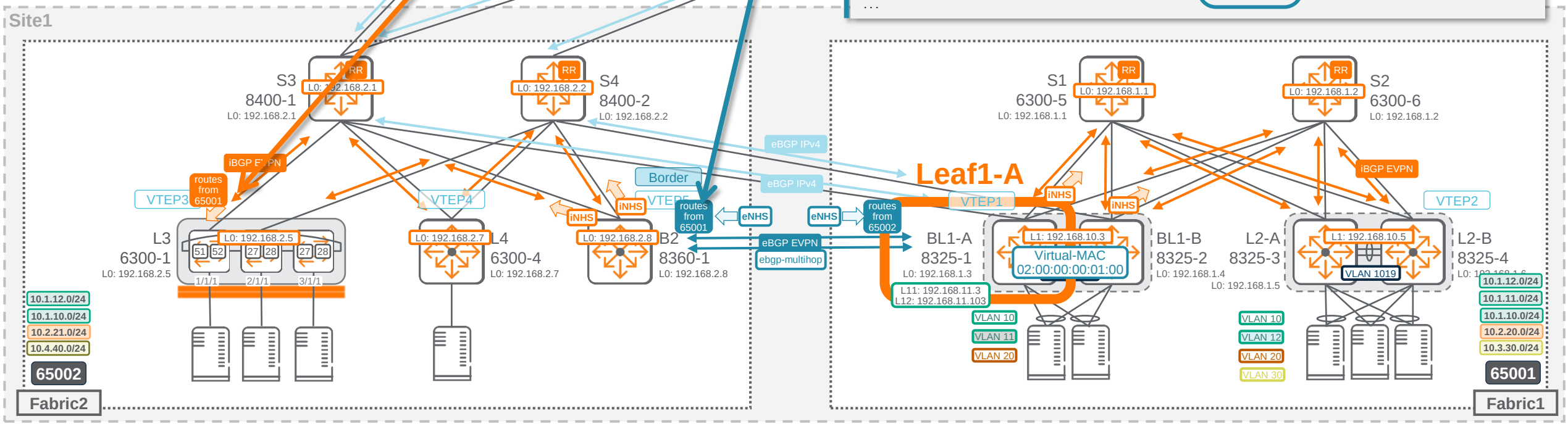
local border VTEP eBGP peering over loopbacks (known as from eBGP IPv4 AF)

ebgp-multihop has to be set per neighbor as the remote-as is not defined in the peer-group

Next-Hop for eBGP routes is reset to Border2 towards Fabric2 RRs iNHS

```
6300-1-VSF# sh bgp l2 evpn neighbors 192.168.2.1 routes route-type 5
Status codes: s suppressed, d damped, h history, * valid, > best, = multipath,
              i internal, e external S Stale, R Removed, a additional-paths
Origin codes: i - IGP, e - EGP, ? - incomplete

EVPN Route-Type 5 prefix: [5]:[ESI]:[EthTag]:[IPAddrLen]:[IPAddr]
VRF : default
Local Router-ID 192.168.2.5

     Network                                                    Nexthop          Metric     LocPrf     Weight     Path
-------------------------------------------------------------------------------------------------------------------------
...
Route Distinguisher: 192.168.1.5:1        (L3VNI 100001)
*>i [5]:[0]:[0]:[24]:[10.1.10.0]                                192.168.2.8      0          100        0          65001 ?
*>i [5]:[0]:[0]:[24]:[10.1.12.0]                                192.168.2.8      0          100        0          65001 ?
*>i [5]:[0]:[0]:[32]:[192.168.11.105]                           192.168.2.8      0          100        0          65001 ?
*>i [5]:[0]:[0]:[32]:[192.168.11.5]                             192.168.2.8      0          100        0          65001 ?
*>i [5]:[0]:[0]:[64]:[fd00:10:1:12::]                           192.168.2.8      0          100        0          65001 ?
...
```

```
8360-1# sh bgp l2 evpn neighbors 192.168.1.3 routes route-type 5
Status codes: s suppressed, d damped, h history, * valid, > best, = multipath,
              i internal, e external S Stale, R Removed, a additional-paths
Origin codes: i - IGP, e - EGP, ? - incomplete

EVPN Route-Type 5 prefix: [5]:[ESI]:[EthTag]:[IPAddrLen]:[IPAddr]
VRF : default
Local Router-ID 192.168.2.8

     Network                                                    Nexthop          Metric     LocPrf     Weight     Path
-------------------------------------------------------------------------------------------------------------------------
...
Route Distinguisher: 192.168.1.5:1        (L3VNI 100001)
*>e [5]:[0]:[0]:[24]:[10.1.10.0]                                192.168.10.3     0          100        0          65001 ?
*>e [5]:[0]:[0]:[24]:[10.1.12.0]                                192.168.10.3     0          100        0          65001 ?
*>e [5]:[0]:[0]:[32]:[192.168.11.105]                           192.168.10.3     0          100        0          65001 ?
*>e [5]:[0]:[0]:[32]:[192.168.11.5]                             192.168.10.3     0          100        0          65001 ?
*>e [5]:[0]:[0]:[64]:[fd00:10:1:12::]                           192.168.10.3     0          100        0          65001 ?
...
```



BL1-A (Border-Leader-1 VSX Primary) Configuration

Global scope for Route-Target for L3VNI

```
vrf VRF1
    rd 192.168.1.3:1
    route-target export 65001:1 evpn
    route-target import 65001:1 evpn
    route-target export 1:1 evpn
    route-target import 1:1 evpn
vrf VRF2
    rd 192.168.1.3:2
    route-target export 65001:2 evpn
    route-target import 65001:2 evpn
    route-target export 1:2 evpn
    route-target import 1:2 evpn
```

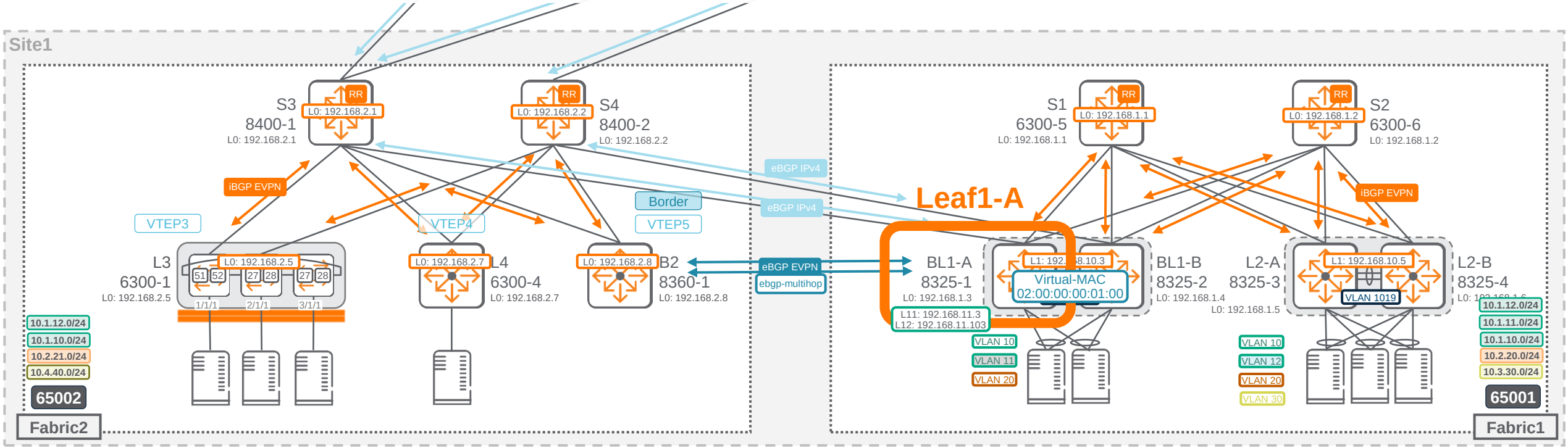
local filter (route-target) per fabric

global filter (route-target) common to all fabrics

local filter (route-target) per fabric

global filter (route-target) common to all fabrics

- Recommendation: use local scope and global scope for greater flexibility
- If only one should be used (for simplicity): use global route-target
- Per VRF: not every VRFs might require a global scope
- Recommendation to have a unique global scope for simplicity. Multiple global values could be defined to create groups of sites with associated higher complexity.
- Must be set on all VTEPs in the scope of the VRF



BL1-A (Border-Leader-1 VSX Primary) Configuration

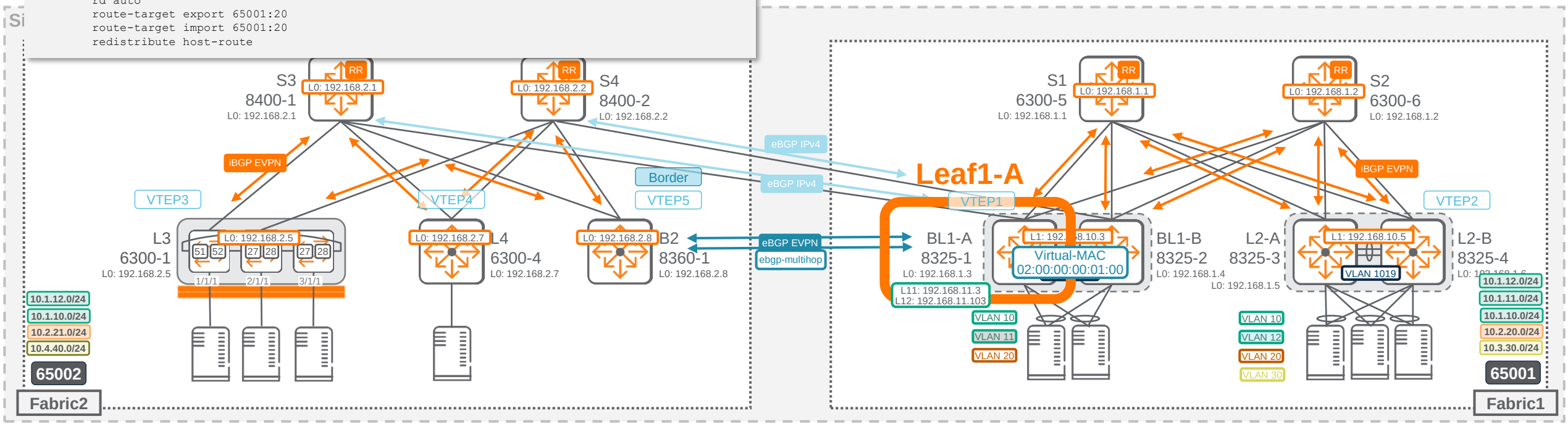
Global scope for Route-Target for L2VNI (stretched VLAN over multiple fabrics)

```
evpn
    ...
    vlan 10
        rd auto
        route-target export 65001:10
        route-target import 65001:10
        route-target export 1:10
        route-target import 1:10
        redistribute host-route
    vlan 11
        rd auto
        route-target export 65001:11
        route-target import 65001:11
        redistribute host-route
    vlan 12
        rd auto
        route-target export 65001:12
        route-target import 65001:12
        redistribute host-route
    vlan 20
        rd auto
        route-target export 65001:20
        route-target import 65001:20
        redistribute host-route
```

local filter (route-target) per fabric

global filter (route-target) common to all fabrics

- Only for stretched VLANs over multiple fabrics
- Must be set on the border-VTEP to enable this VLAN to participate to bridging between the iBGP VTEPs and the eBGP VTEPs. L2VNI-VLAN mapping must be set in **VXLAN interface as well** !
- Global scope route-target value must **also** be set on all leaves that require to sent/receive L2 traffic on stretched VLANs over multiple fabrics



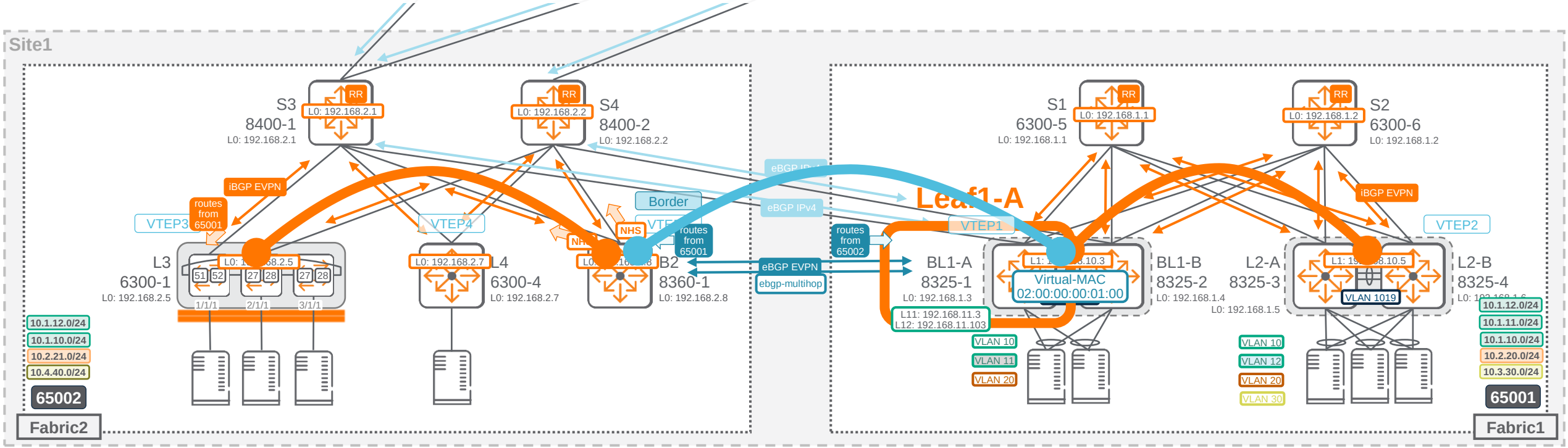
BL1-A (Border-Leader-1 VSX Primary) Configuration

EVPN VXLAN Tunnels Bridging (adjusting Split-Horizon rule)

```
evpn
    arp-suppression
    nd-suppression
    redistribute local-svi
!
    dyn-vxlan-tunnel-bridging-mode ibgp-ebgp
!
    vlan 10
        ...
```

Allows VXLAN traffic bridging between iBGP VTEPs and eBGP VTEPs

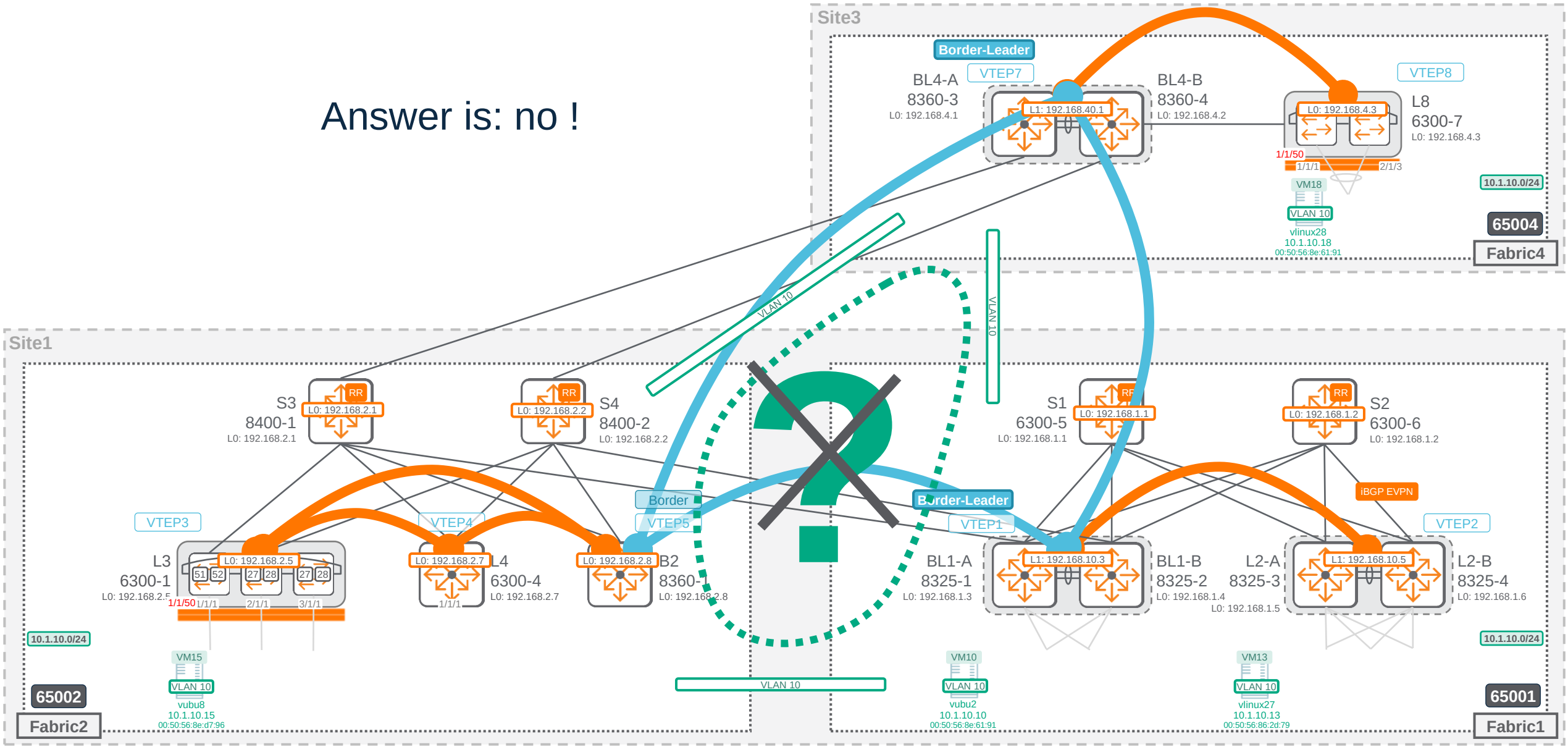
8325-1(config-evpn)# dyn-vxlan-tunnel-bridging-mode ibgp-ebgp
WARNING!! Enabling this command would bridge the traffic between EVPN and STATIC VxLAN tunnels.



Breaking VXLAN Split-Horizon

Any L2 loop ?

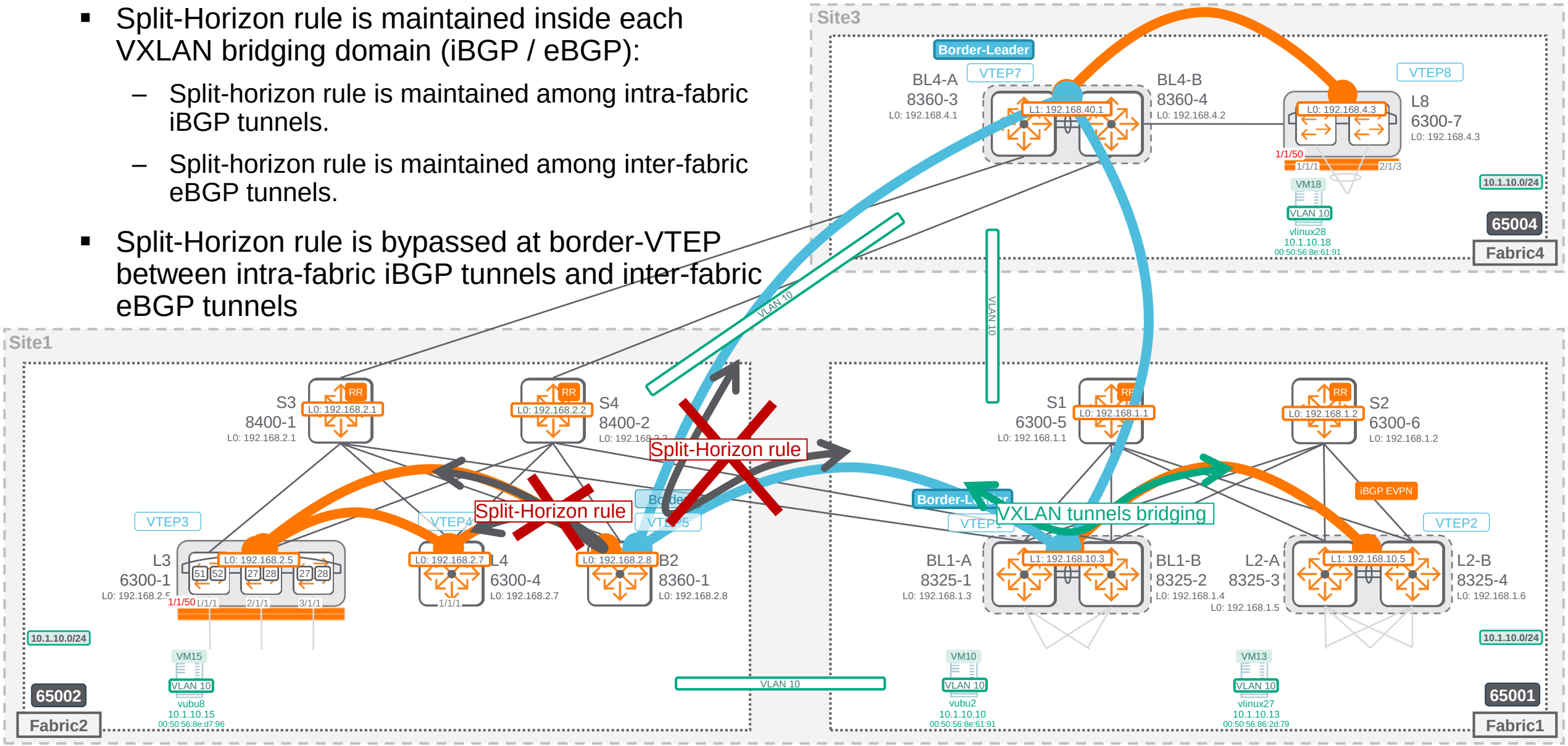
Answer is: no !



VXLAN Split-Horizon for L2 traffic

Intra-Fabric, inter-Fabric

- Split-Horizon rule is maintained inside each VXLAN bridging domain (iBGP / eBGP):
  - Split-horizon rule is maintained among intra-fabric iBGP tunnels.
  - Split-horizon rule is maintained among inter-fabric eBGP tunnels.
- Split-Horizon rule is bypassed at border-VTEP between intra-fabric iBGP tunnels and inter-fabric eBGP tunnels



NEW

Fabric1 Border Configuration

Additional Configuration for inter-site inter-fabric communication

- eBGP EVPN AF

Border-Leader-1A Configuration

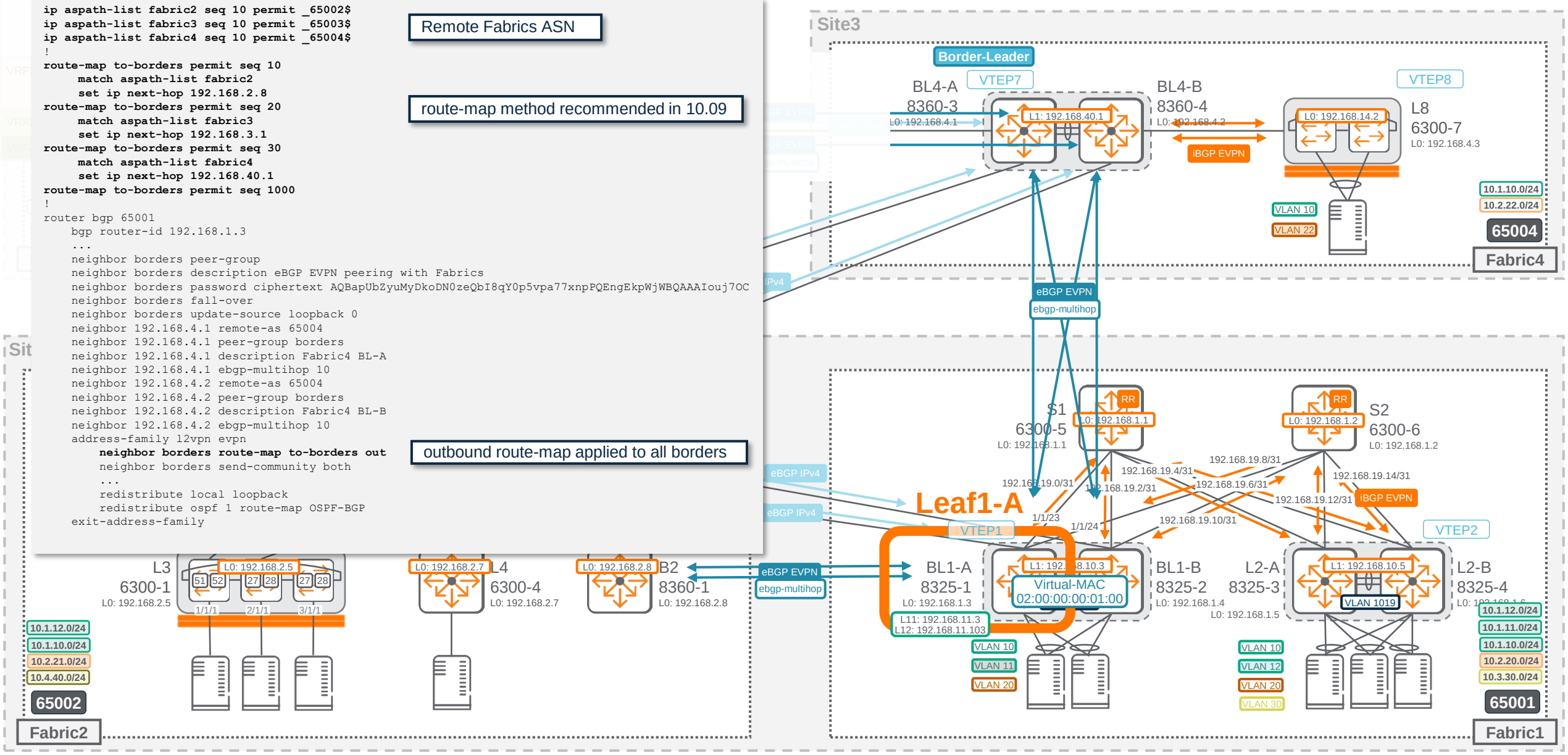
EVPN Route-map

```
ip aspath-list fabric2 seq 10 permit _65002$
ip aspath-list fabric3 seq 10 permit _65003$
ip aspath-list fabric4 seq 10 permit _65004$
!
route-map to-borders permit seq 10
     match aspath-list fabric2
     set ip next-hop 192.168.2.8
route-map to-borders permit seq 20
     match aspath-list fabric3
     set ip next-hop 192.168.3.1
route-map to-borders permit seq 30
     match aspath-list fabric4
     set ip next-hop 192.168.40.1
route-map to-borders permit seq 1000
!
router bgp 65001
    bgp router-id 192.168.1.3
    ...
    neighbor borders peer-group
    neighbor borders description eBGP EVPN peering with Fabrics
    neighbor borders password ciphertext AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
    neighbor borders fall-over
    neighbor borders update-source loopback 0
    neighbor 192.168.4.1 remote-as 65004
    neighbor 192.168.4.1 peer-group borders
    neighbor 192.168.4.1 description Fabric4 BL-A
    neighbor 192.168.4.1 ebgp-multihop 10
    neighbor 192.168.4.2 remote-as 65004
    neighbor 192.168.4.2 peer-group borders
    neighbor 192.168.4.2 description Fabric4 BL-B
    neighbor 192.168.4.2 ebgp-multihop 10
    address-family l2vpn evpn
        neighbor borders route-map to-borders out
        neighbor borders send-community both
        ...
        redistribute local loopback
        redistribute ospf 1 route-map OSPF-BGP
    exit-address-family
```

Remote Fabrics ASN

route-map method recommended in 10.09

outbound route-map applied to all borders



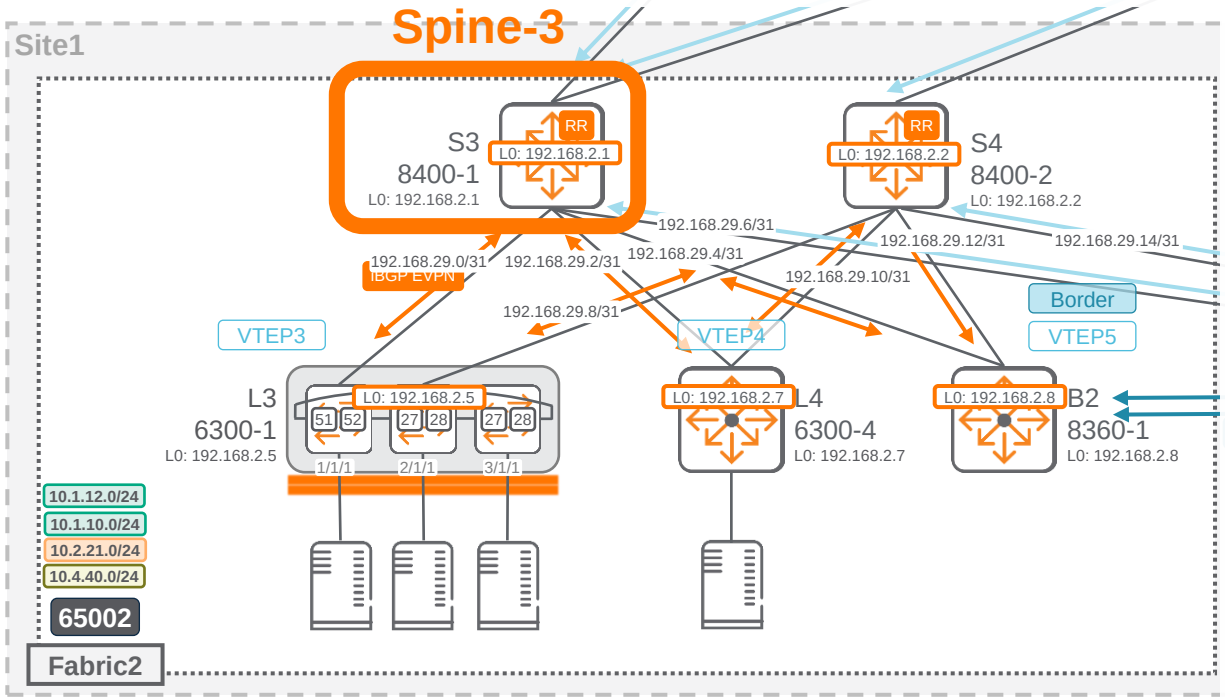
REMINDER

Fabric2 Spines Configuration

- Underlay ROPs, loopbacks
- OSPF underlay routing
- iBGP Route-Reflectors for EVPN AF
- ROPs, underlay eBGP IPv4 AF and loopbacks routing

S3 (Spine-3) Configuration

ROPs to Leafs / Loopback




```
interface 1/1/13
    no shutdown
    mtu 9198
    routing
    ip mtu 9198
    ip address 192.168.29.0/31
    ip ospf 1 area 0.0.0.0
    no ip ospf passive
    ip ospf network point-to-point
    ip ospf authentication message-digest
    ip ospf message-digest-key 1 md5 ciphertext AQBapUbZyuMyDkoDN0zeQbI8qY17dlU3LitKLwE+4EU4v8nuBQAAABt51nkM
interface 1/1/9
    no shutdown
    mtu 9198
    routing
    ip mtu 9198
    ip address 192.168.29.2/31
    ip ospf 1 area 0.0.0.0
    no ip ospf passive
    ip ospf network point-to-point
    ip ospf authentication message-digest
    ip ospf message-digest-key 1 md5 ciphertext AQBapUbZyuMyDkoDN0zeQbI8qY17dlU3LitKLwE+4EU4v8nuBQAAABt51nkM
interface 1/10/13
    no shutdown
    mtu 9198
    routing
    ip mtu 9198
    ip address 192.168.29.4/31
    ip ospf 1 area 0.0.0.0
    no ip ospf passive
    ip ospf network point-to-point
    ip ospf authentication message-digest
    ip ospf message-digest-key 1 md5 ciphertext AQBapUbZyuMyDkoDN0zeQbI8qY17dlU3LitKLwE+4EU4v8nuBQAAABt51nkM
interface 1/3/3
    no shutdown
    mtu 9198
    routing
    ip mtu 9198
    ip address 192.168.29.6/31
!
interface loopback 0
    ip address 192.168.1.1/32
```

L3 Link to L3 (6300-1)

L3 Link to L4 (6300-4)

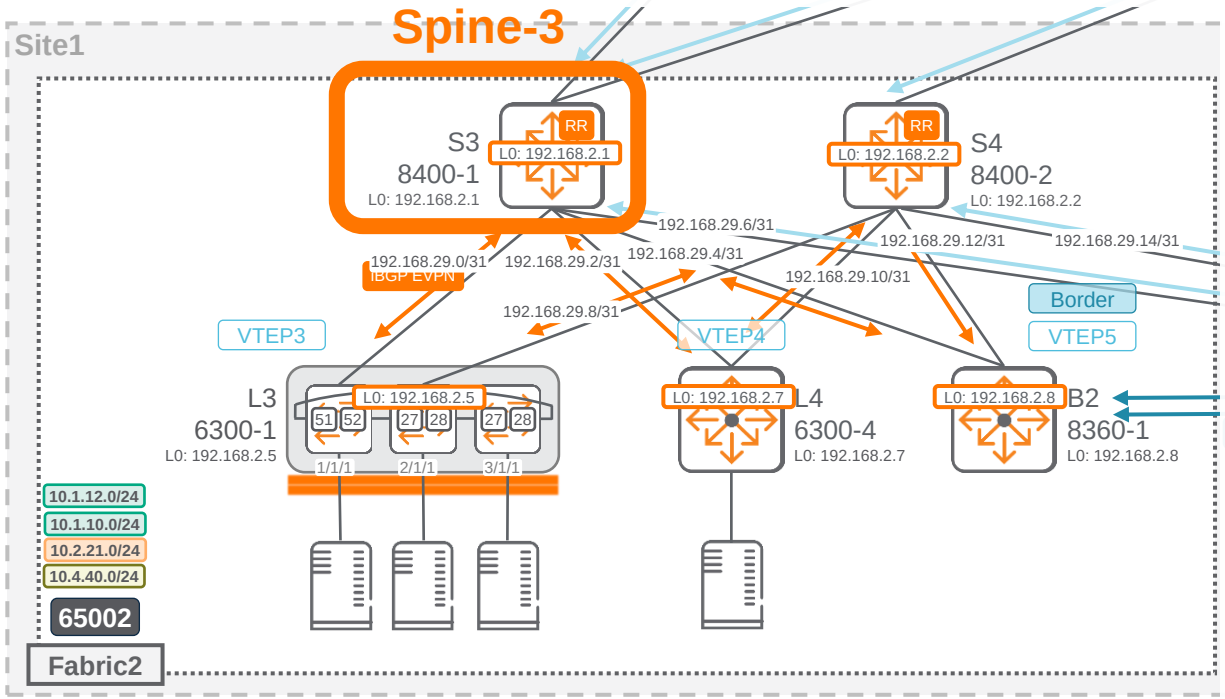
L3 Link to B2 (8360-1)

L3 Link to BL1-A (8325-1)

Loopback0

S3 (Spine-3) Configuration

Routing




```
route-map connected-ospf permit seq 10
     set tag 1000
!
router ospf 1
    router-id 192.168.2.1
    ! optional
    max-metric router-lsa include-stub on-startup 300
    passive-interface default
    !
    redistribute local loopback route-map connected-ospf
    area 0.0.0.0
!
router bgp 65002
    bgp router-id 192.168.2.1
    ! optional
    trap-enable
    bgp log-neighbor-changes
    !
    bgp fast-external-fallover
    !
    bgp deterministic-med
    bgp always-compare-med
    !
    neighbor leaf peer-group
    neighbor leaf remote-as 65002
    neighbor leaf description Leaf RR clients
    ! optional
    neighbor leaf password ciphertext
AQBapUbZyuMyDkoDN0zeQbI8qY17dlU3LitKLwE+4EU4v8nuBQAAABt51pkM
    neighbor leaf fall-over
    !
    neighbor leaf update-source loopback 0
    neighbor 192.168.2.5 peer-group leaf
    neighbor 192.168.2.7 peer-group leaf
    neighbor 192.168.2.8 peer-group leaf
    address-family l2vpn evpn
        neighbor leaf route-reflector-client
        neighbor leaf send-community both
        neighbor 192.168.2.5 activate
        neighbor 192.168.2.7 activate
        neighbor 192.168.2.8 activate
    exit-address-family
```

tag value defined for local Fabric loopback

Optional best practices

loopback redist. in OSPF

Optional best practices

Enabled by default

If used, must be the same in the BGP domain

Spines are iBGP RR

Fall-over notification from FIB to BGP

iBGP peering between loopbacks

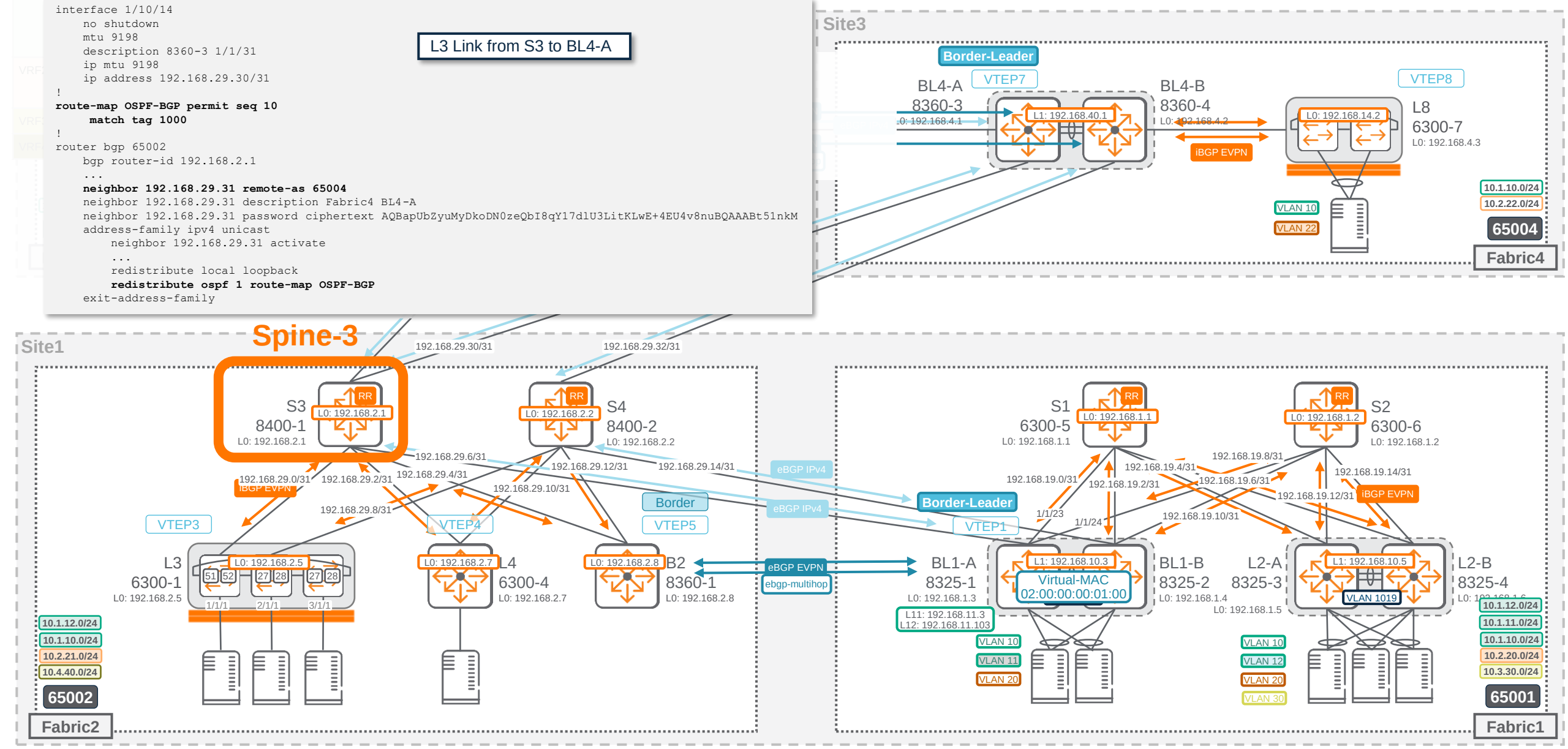
Extended community is required for EVPN NLRI

S3 (Spine-3) Additional Configuration

Underlay ROPs and eBGP IPv4 AF routing

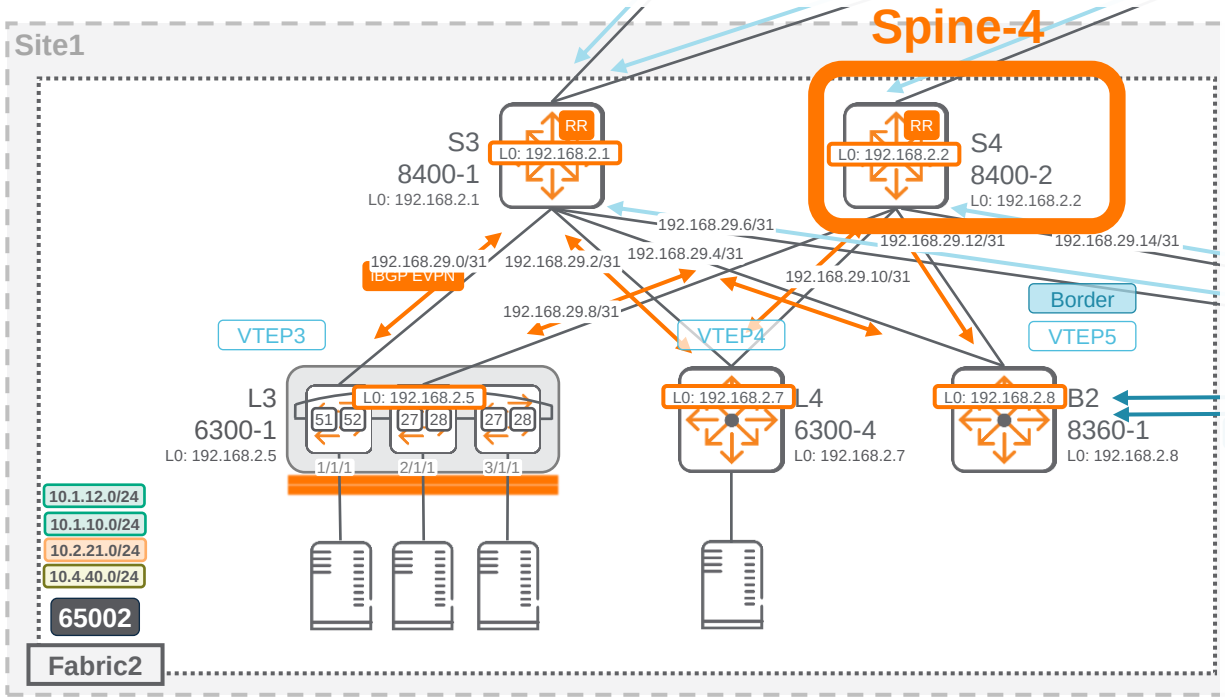
```
interface 1/10/14
    no shutdown
    mtu 9198
    description 8360-3 1/1/31
    ip mtu 9198
    ip address 192.168.29.30/31
!
route-map OSPF-BGP permit seq 10
     match tag 1000
!
router bgp 65002
    bgp router-id 192.168.2.1
    ...
    neighbor 192.168.29.31 remote-as 65004
    neighbor 192.168.29.31 description Fabric4 BL4-A
    neighbor 192.168.29.31 password ciphertext AQBapUbZyuMyDkoDN0zeQbI8qY17dlU3LitKLwE+4EU4v8nuBQAAABt51nkM
    address-family ipv4 unicast
        neighbor 192.168.29.31 activate
        ...
        redistribute local loopback
        redistribute ospf 1 route-map OSPF-BGP
    exit-address-family
```

L3 Link from S3 to BL4-A



S4 (Spine-4) Configuration

ROPs to Leafs / Loopback




```
interface 1/1/13
    no shutdown
    mtu 9198
    routing
    ip mtu 9198
    ip address 192.168.29.8/31
    ip ospf 1 area 0.0.0.0
    no ip ospf passive
    ip ospf network point-to-point
    ip ospf authentication message-digest
    ip ospf message-digest-key 1 md5 ciphertext AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
interface 1/1/9
    no shutdown
    mtu 9198
    routing
    ip mtu 9198
    ip address 192.168.29.10/31
    ip ospf 1 area 0.0.0.0
    no ip ospf passive
    ip ospf network point-to-point
    ip ospf authentication message-digest
    ip ospf message-digest-key 1 md5 ciphertext AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
interface 1/10/12
    no shutdown
    mtu 9198
    routing
    ip mtu 9198
    ip address 192.168.29.12/31
    ip ospf 1 area 0.0.0.0
    no ip ospf passive
    ip ospf network point-to-point
    ip ospf authentication message-digest
    ip ospf message-digest-key 1 md5 ciphertext AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
interface 1/3/3
    no shutdown
    mtu 9198
    routing
    ip mtu 9198
    ip address 192.168.29.14/31
!
interface loopback 0
    ip address 192.168.1.1/32
```

L3 Link to L3 (6300-1)

L3 Link to L4 (6300-4)

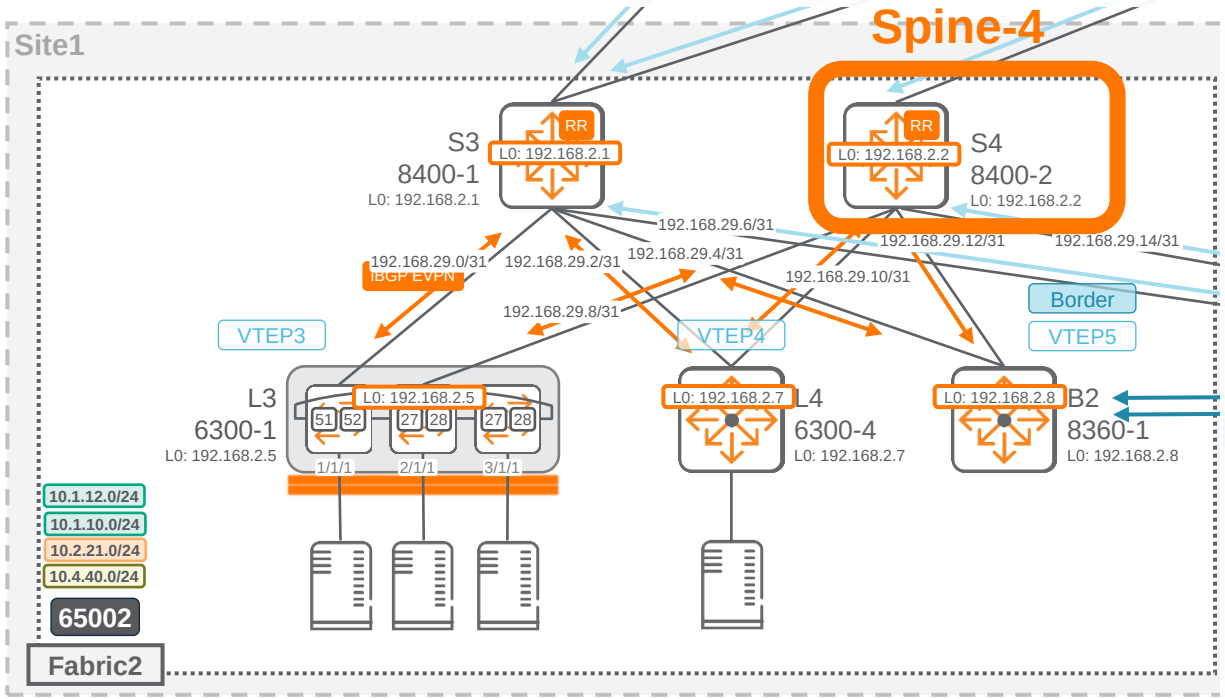
L3 Link to B2 (8360-1)

L3 Link to BL1-A (8325-1)

Loopback0

S4 (Spine-4) Configuration

Routing




```
route-map connected-ospf permit seq 10
     set tag 1000
!
router ospf 1
    router-id 192.168.2.2
    ! optional
    max-metric router-lsa include-stub on-startup 300
    passive-interface default
    !
    redistribute local loopback route-map connected-ospf
    area 0.0.0.0
!
router bgp 65002
    bgp router-id 192.168.2.2
    ! optional
    trap-enable
    bgp log-neighbor-changes
    !
    bgp fast-external-fallover
    !
    bgp deterministic-med
    bgp always-compare-med
    !
    neighbor leaf peer-group
    neighbor leaf remote-as 65002
    neighbor leaf description Leaf RR clients
    ! optional
    neighbor leaf password ciphertext
AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIoui7QC
    neighbor leaf fall-over
    !
    neighbor leaf update-source loopback 0
    neighbor 192.168.2.5 peer-group leaf
    neighbor 192.168.2.7 peer-group leaf
    neighbor 192.168.2.8 peer-group leaf
    address-family l2vpn evpn
        neighbor leaf route-reflector-client
        neighbor leaf send-community both
        neighbor 192.168.2.5 activate
        neighbor 192.168.2.7 activate
        neighbor 192.168.2.8 activate
    exit-address-family
```

tag value defined for local Fabric loopback

Optional best practices

loopback redist. in OSPF

Optional best practices

Enabled by default

If used, must be the same in the BGP domain

Spines are iBGP RR

Fall-over notification from FIB to BGP

iBGP peering between loopbacks

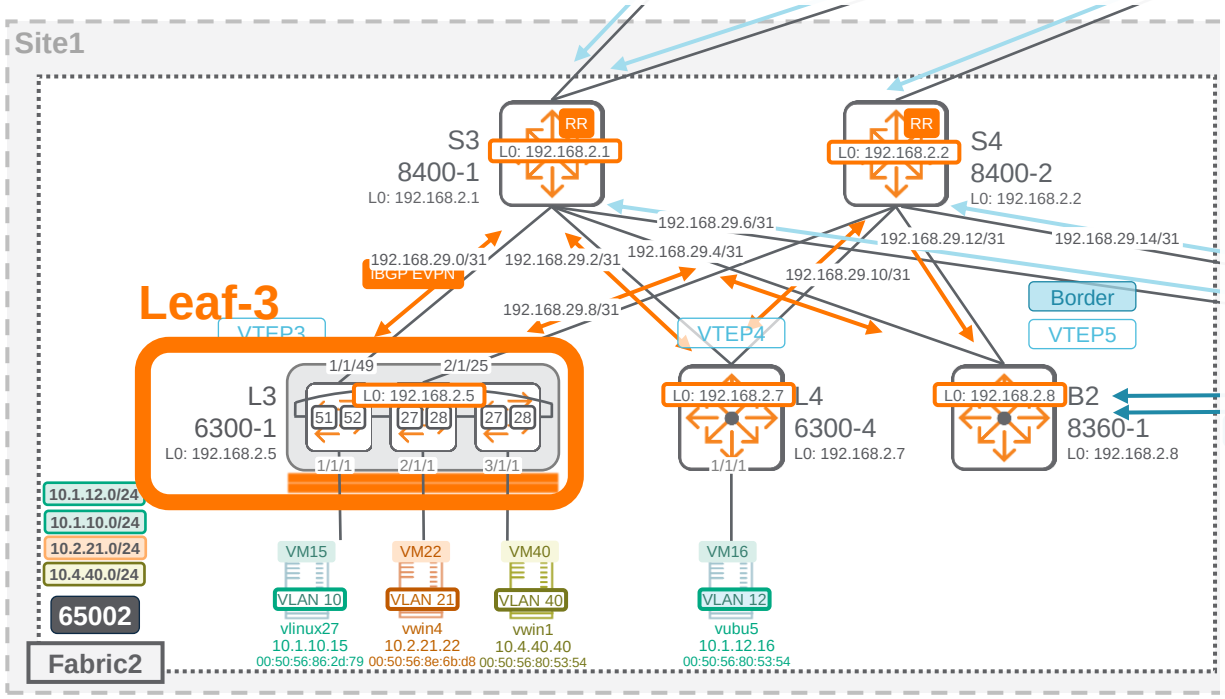
Extended community is required for EVPN NLRI

Fabric2 Leaves Configuration

- Summary

L3 (Leaf-3 VSF) Configuration

Servers Connectivity




```
vlan 10,21,40
!
interface 1/1/1
    no shutdown
    description VM15
    no routing
    vlan access 10
!
interface 2/1/1
    no shutdown
    description VM22
    no routing
    vlan access 21
!
interface 3/1/1
    no shutdown
    description VM40
    no routing
    vlan access 40
!
```

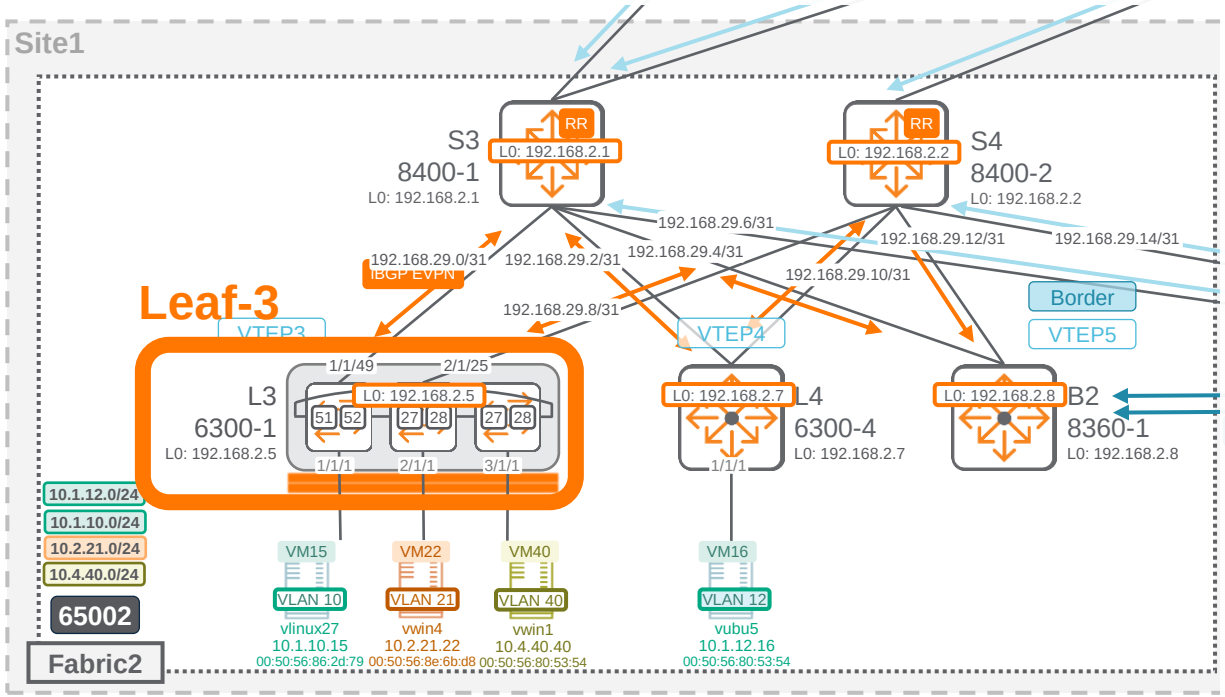
Servers VLANs

Servers Connectivity

(VMs)

L3 (Leaf-3 VSF) Configuration

ROPs to Spines / Loopback




```
interface 1/1/49
    no shutdown
    mtu 9198
    ip mtu 9198
    ip address 192.168.29.1/31
    ip ospf 1 area 0.0.0.0
    no ip ospf passive
    ip ospf network point-to-point
    ip ospf authentication message-digest
    ip ospf message-digest-key 1 md5 ciphertext
AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
!
interface 2/1/25
    no shutdown
    mtu 9198
    ip mtu 9198
    ip address 192.168.29.9/31
    ip ospf 1 area 0.0.0.0
    no ip ospf passive
    ip ospf network point-to-point
    ip ospf authentication message-digest
    ip ospf message-digest-key 1 md5 ciphertext
AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
!
interface loopback 0
    ip address 192.168.2.5/32
!
```

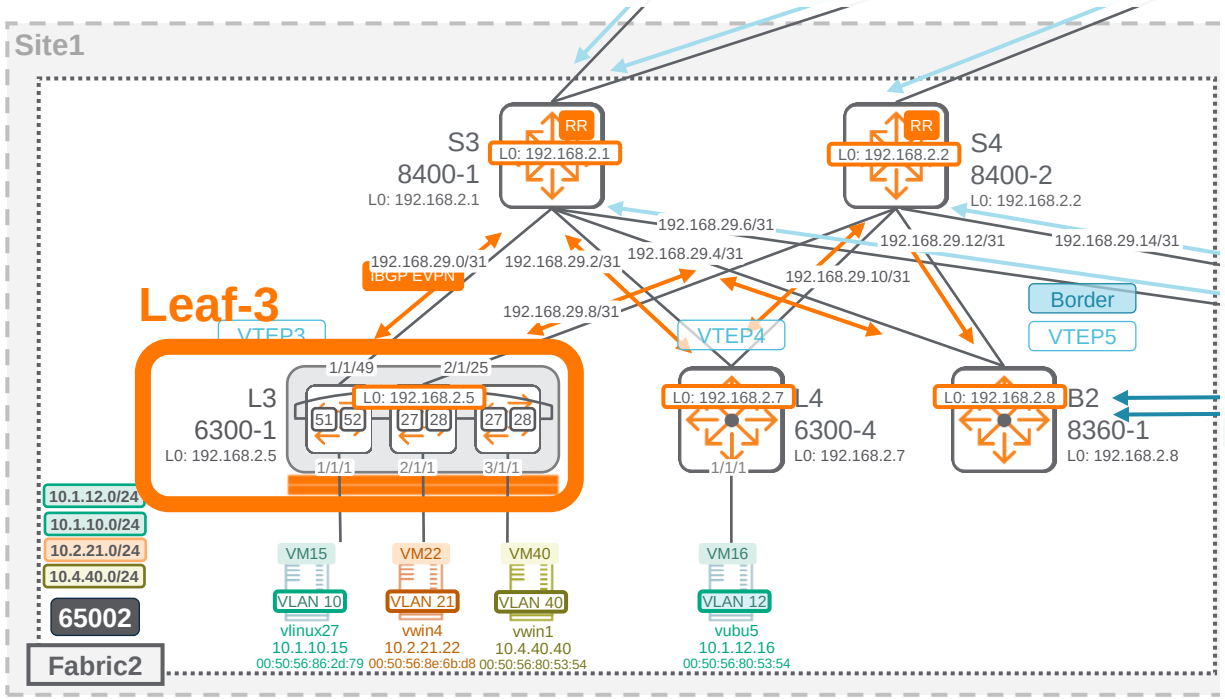
L3 Link to S3 (8400-1)

L3 Link to S4 (8400-2)

Unique Loopback0 for VSF or standalone

L3 (Leaf-3 VSF) Configuration

VRFs / EVPN




```
vrf VRF1
    rd 192.168.2.5:1
    route-target export 1:1 evpn
    route-target export 65002:1 evpn
    route-target import 1:1 evpn
    route-target import 65002:1 evpn
vrf VRF2
    rd 192.168.2.5:2
    route-target export 1:2 evpn
    route-target export 65002:2 evpn
    route-target import 1:2 evpn
    route-target import 65002:2 evpn
vrf VRF4
    rd 192.168.2.5:4
    route-target export 65002:4 evpn
    route-target import 65002:4 evpn
!
evpn
    arp-suppression
    nd-suppression
    redistribute local-svi
 vlan 10
        rd auto
        route-target export 1:10
        route-target export 65002:10
        route-target import 1:10
        route-target import 65002:10
        redistribute host-route
    vlan 21
        rd auto
        route-target export 65002:21
        route-target import 65002:21
        redistribute host-route
    vlan 40
        rd auto
        route-target export auto
        route-target import auto
        redistribute host-route
```

Recommendation value: [L1:VRF_ID]
in this demo [L0:VRF_ID] (educative purpose)

Local RT: 65002:1
+
Global RT: 1:1 **VRF1 is exported to other Fabrics**

Local RT: 65002:2
+
Global RT: 1:2 **VRF2 is exported to other Fabrics**

Local RT only: 65002:4
VRF4 is not exported out of Fabric2

RD auto is: [VTEP_L1:VLAN_ID]

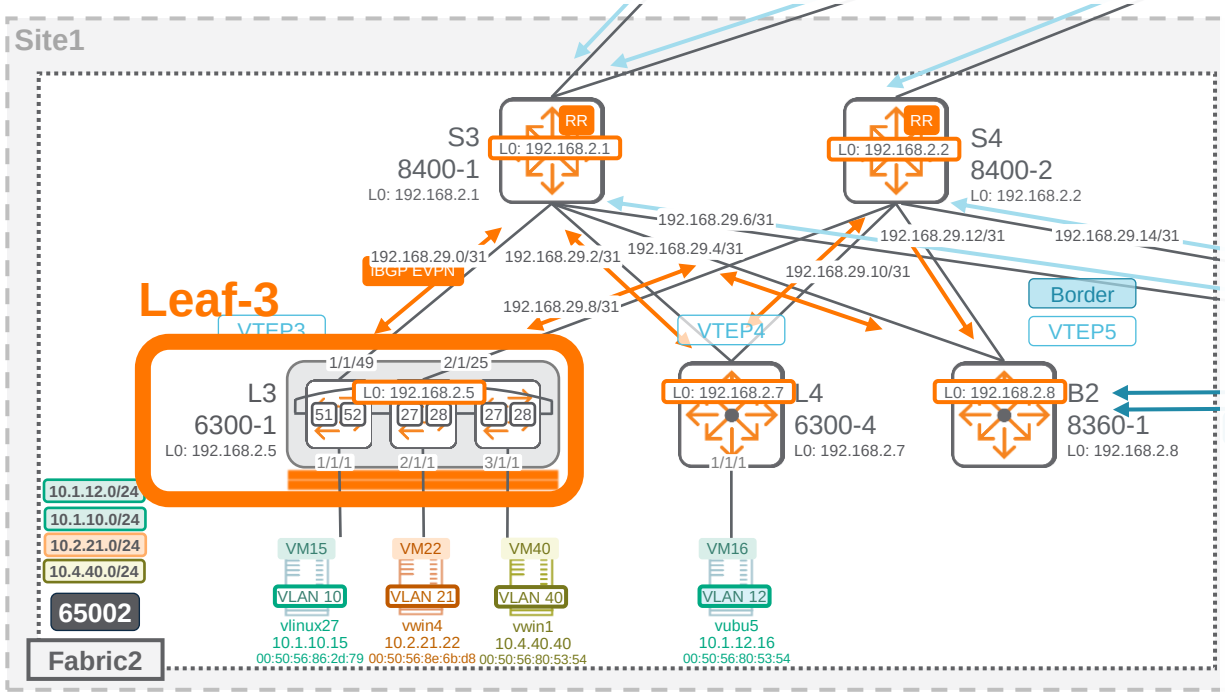
Local RT: 65002:10
+
Global RT: 1:10 **VLAN10 is extended to other Fabrics**

Local RT only: 65002:21
VLAN21 is not exported out of Fabric2

It is possible to use RT auto for local-only iBGP routes

L3 (Leaf-3 VSF) Configuration

Endpoint SVIs / VRF Loopbacks / DHCP relay




```
interface vlan 10
    vrf attach VRF1
    ip mtu 9000
    ip address 10.1.10.1/24
    active-gateway ip mac 12:00:00:00:01:00
    active-gateway ip 10.1.10.1
    ip helper-address 10.10.129.30 vrf default
!
interface vlan 21
    vrf attach VRF2
    ip mtu 9000
    ip address 10.2.21.5/24
    active-gateway ip mac 12:00:00:00:01:00
    active-gateway ip 10.2.21.1
!
interface vlan 40
    vrf attach VRF4
    ip mtu 9000
    ip address 10.4.40.1/24
    active-gateway ip mac 12:00:00:00:01:00
    active-gateway ip 10.4.40.1
!
interface loopback 12
    vrf attach VRF1
    description for troubleshooting in VRF1
    ip address 192.168.21.5/32
!
ip source-interface dhcp_relay interface loopback0
dhcp-relay option 82 source-interface
dhcp-relay option 82 replace
```

IPv4 only (no IPv6 multi-hop support in 10.09)

SVI IP = AG IP best practice

Follow best practices for AG MAC value

Overlay DHCP relay with DHCP server in underlay default VRF

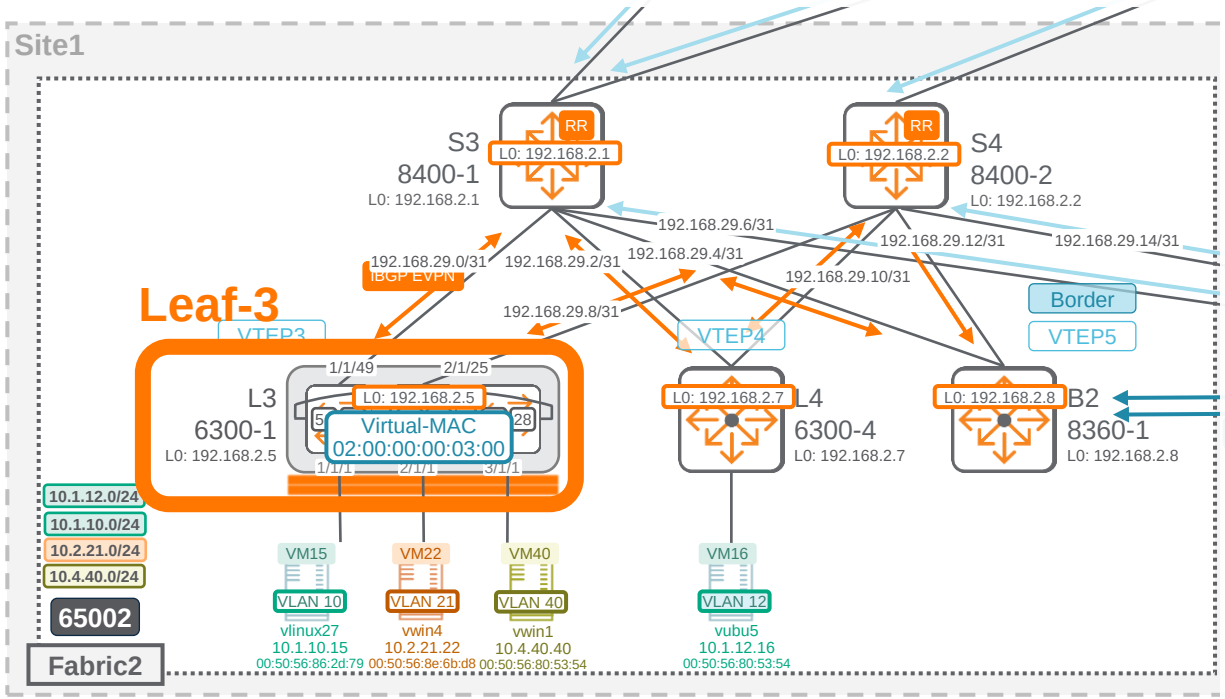
SVI IP != AG IP for demo purpose in VRF2

Per VRF, unicast IP for troubleshooting

DHCP relay sourcing

L3 (Leaf-3 VSF) Configuration

VXLAN interface / Virtual-MAC




```
interface vxlan 1
    source ip 192.168.2.5
    vxlan-counters aggregate
    no shutdown
    vni 10010
        vlan 10
    vni 10021
        vlan 21
    vni 100001
        vrf VRF1
        routing
    vni 100002
        vrf VRF2
        routing
    vni 100004
        vrf VRF4
        routing
!
virtual-mac 02:00:00:00:03:00
```

VXLAN tunnel source IP = VSF unicast Loopback0

Optional counters for VXLAN statistics

1-to-1 VLAN-VNI mapping

L3 VNI associated to VRFs

VTEP virtual-MAC for L3VNI functionality

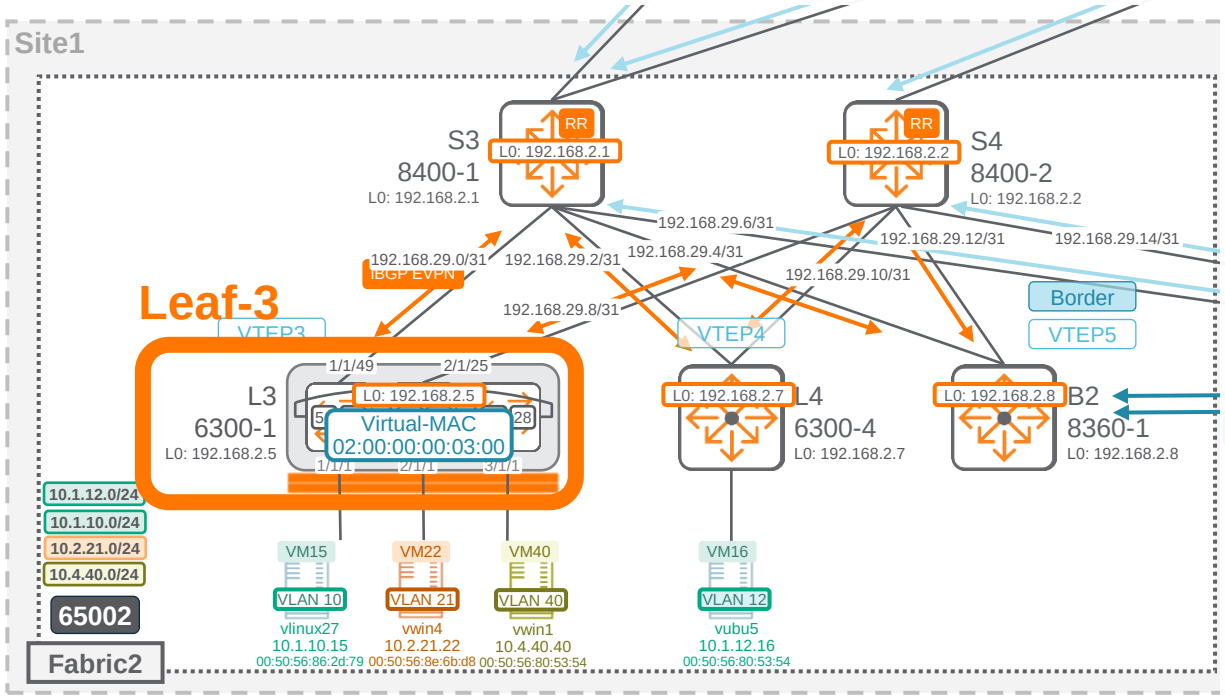
L3 (Leaf-3 VSF) Configuration

Routing

```
router ospf 1
    router-id 192.168.2.5
    max-metric router-lsa include-stub on-startup 300
    passive-interface default
    graceful-restart ignore-lost-interface
    trap-enable
    redistribute local loopback route-map connected-ospf
    area 0.0.0.0
```

best practice for HA during reboot

loopback redistribution in OSPF




```
ip prefix-list endpoint-VRF1 seq 10 permit 10.1.0.0/16 le 32
ip prefix-list endpoint-VRF2 seq 10 permit 10.2.0.0/16 le 32
ip prefix-list endpoint-VRF4 seq 10 permit 10.4.0.0/16 le 32
!
route-map connected-bgp-VRF1 permit seq 10
     match ip address prefix-list endpoint-VRF1
route-map connected-bgp-VRF2 permit seq 10
     match ip address prefix-list endpoint-VRF2
route-map connected-bgp-VRF4 permit seq 10
     match ip address prefix-list endpoint-VRF4
!
route-map connected-ospf permit seq 10
     set tag 1000
!
router bgp 65002
    bgp router-id 192.168.2.5
    trap-enable
    bgp fast-external-fallover
    bgp log-neighbor-changes
    bgp deterministic-med
    bgp always-compare-med
    neighbor spine-RR peer-group
    neighbor spine-RR remote-as 65002
    neighbor spine-RR description Spine and RR peer-group
    neighbor spine-RR password ciphertext
AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7QC
    neighbor spine-RR fall-over
    neighbor spine-RR update-source loopback 0
    neighbor 192.168.2.1 peer-group spine-RR
    neighbor 192.168.2.2 peer-group spine-RR
    address-family l2vpn evpn
        neighbor spine-RR send-community both
        neighbor 192.168.2.1 activate
        neighbor 192.168.2.2 activate
    exit-address-family
!
    vrf VRF1
        bgp router-id 192.168.2.5
        address-family ipv4 unicast
            redistribute local loopback
            redistribute connected route-map connected-bgp-VRF1
        exit-address-family
!
    vrf VRF2
        bgp router-id 192.168.2.5
        address-family ipv4 unicast
            redistribute local loopback
            redistribute connected route-map connected-bgp-VRF2
        exit-address-family
!
    vrf VRF4
        bgp router-id 192.168.2.5
        address-family ipv4 unicast
            redistribute local loopback
            redistribute connected route-map connected-bgp-VRF4
        exit-address-family
```

Control for route injection

tag value defined for local Fabric loopback

If used, must be the same in the BGP domain

Fall-over notification from FIB to BGP

iBGP peering between loopbacks

Extended community is required

Used to advertise EVPN Type-5 IP prefix routes

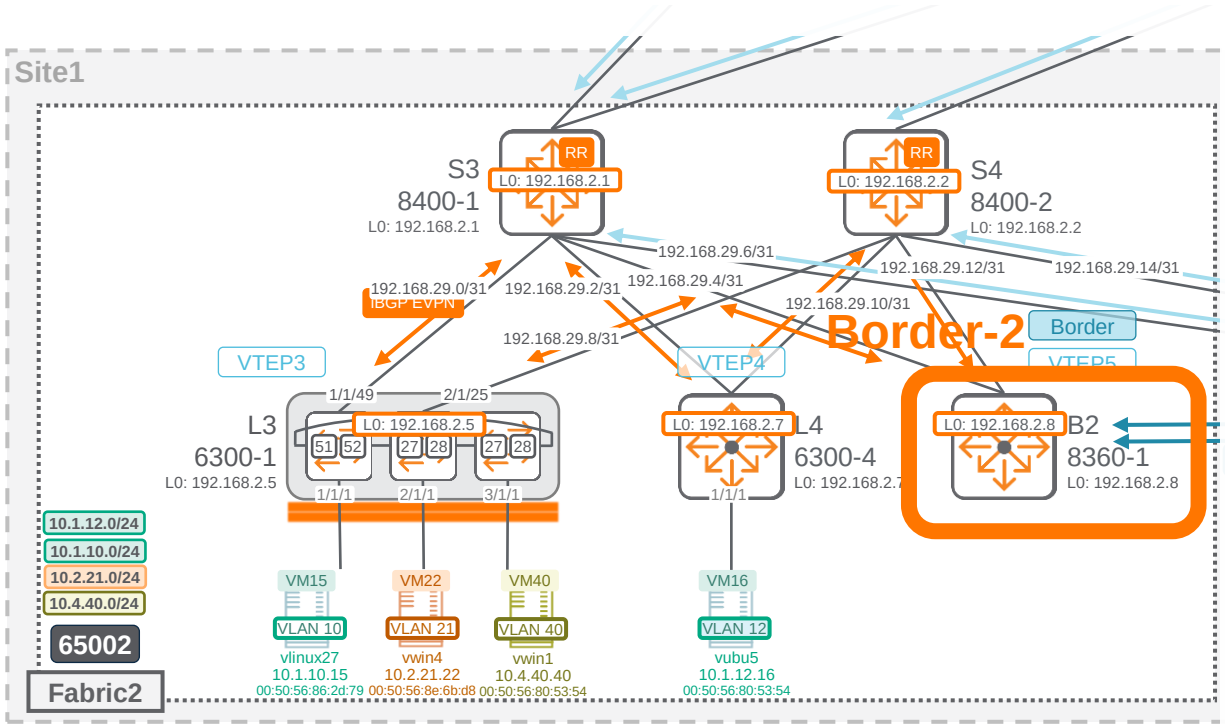
Fabric2 Border Configuration

- Summary

B2 (Border-2) Configuration

ROPs to Spines / Loopbacks

- Standalone unit (versus VSX in operation)
- No server/host connection (for demo purpose)
- No SVI configured per VRF (for demo purpose)




```
interface 1/1/10
    no shutdown
    mtu 9198
    description 8400-2 1/10/12
    ip mtu 9198
    ip address 192.168.29.13/31
    ip ospf 1 area 0.0.0.0
    no ip ospf passive
    ip ospf network point-to-point
    ip ospf authentication message-digest
    ip ospf message-digest-key 1 md5 ciphertext
AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
!
interface 1/1/13
    no shutdown
    mtu 9198
    description 8400-1 1/10/13
    ip mtu 9198
    ip address 192.168.29.5/31
    ip ospf 1 area 0.0.0.0
    no ip ospf passive
    ip ospf network point-to-point
    ip ospf authentication message-digest
    ip ospf message-digest-key 1 md5 ciphertext
AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
!
interface loopback 0
    ip address 192.168.2.8/32
!
interface loopback 12
    vrf attach VRF1
    description for troubleshooting in VRF1
    ip address 192.168.21.8/32
!
interface loopback 22
    vrf attach VRF2
    description for troubleshooting in VRF2
    ip address 192.168.22.8/32
```

L3 Link to S4 (8400-2)

L3 Link to S3 (8400-1)

Unique Loopback0 for VSF or standalone

Per VRF, unicast IP for troubleshooting

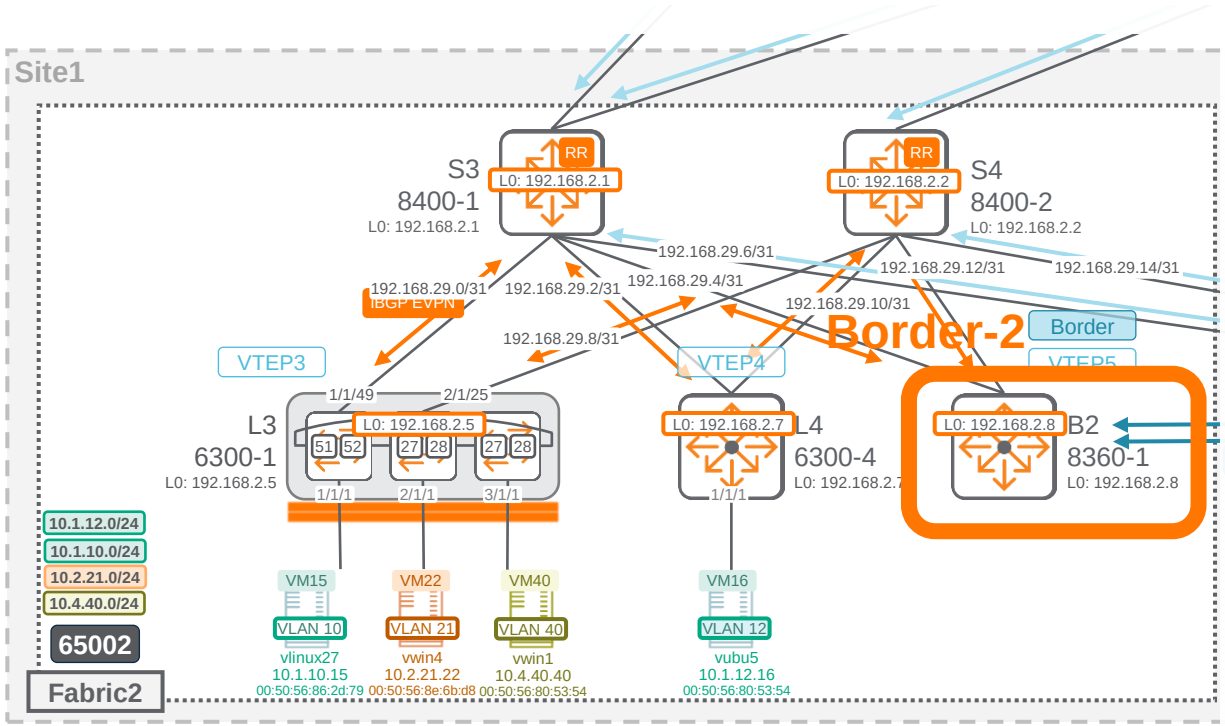
Per VRF, unicast IP for troubleshooting

B2 (Border-2) Configuration

VRFs / VLANs / EVPN

- VLAN 10 and VLAN 12 are extended between Fabrics
- Extended VLANs **MUST be configured** on border-VTEP even if there is no local ports transporting these extended VLANs

```
8360-1# show vlan 10

------------------------------------------------------------------------------------------------------
VLAN  Name                             Status  Reason              Type        Interfaces
------------------------------------------------------------------------------------------------------
10    VLAN10                           up      ok                  static      vxlan1
```




```
vrf VRF1
    rd 192.168.2.5:1
    route-target export 1:1 evpn
    route-target export 65002:1 evpn
    route-target import 1:1 evpn
    route-target import 65002:1 evpn
vrf VRF2
    rd 192.168.2.5:2
    route-target export 1:2 evpn
    route-target export 65002:2 evpn
    route-target import 1:2 evpn
    route-target import 65002:2 evpn
vrf VRF4
    rd 192.168.2.5:4
    route-target export 65002:4 evpn
    route-target import 65002:4 evpn
!
vlan 10,12
!
evpn
    arp-suppression
    nd-suppression
    dyn-vxlan-tunnel-bridging-mode ibgp-ebgp
    vlan 10
        rd auto
        route-target export 1:10
        route-target import 1:10
        redistribute host-route
    vlan 12
        rd auto
        route-target export 1:12
        route-target import 1:12
```

Recommendation value: [L1:VRF_ID]
in this demo [L0:VRF_ID] (educative purpose)

Local RT: 65002:1
+
Global RT: 1:1

VRF1 is exported to other Fabrics

Local RT: 65002:2
+
Global RT: 1:2

VRF2 is exported to other Fabrics

Local RT only: 65002:4
VRF4 is not exported out of Fabric2

Add extended VLANs in border

VXLAN tunnels bridging on Fabric border

Global RT only: 1:10 (for demo/education purpose)

VLAN10 is extended to other Fabrics

Global RT only: 1:12 (for demo/education purpose)

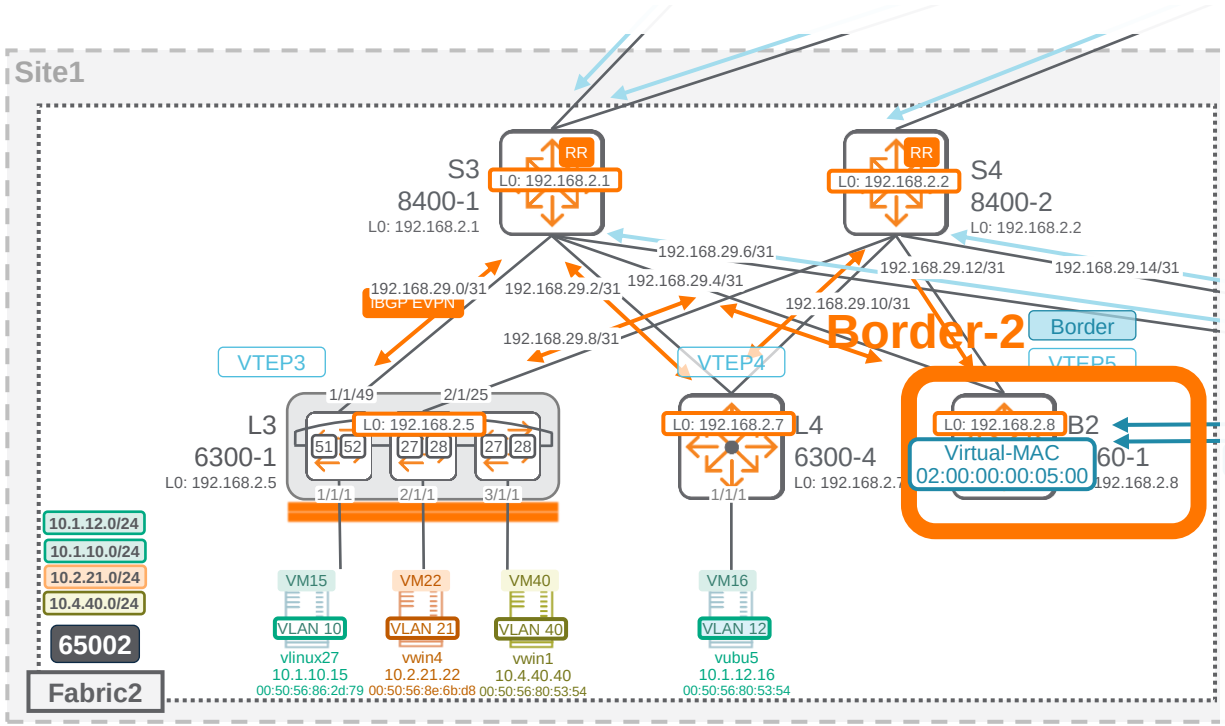
No host-route redistribution for demo purpose

B2 (Border-2) Configuration

VXLAN interface / Virtual-MAC

- VLAN 10 and VLAN 12 are extended between Fabrics
- Extended VLANs **MUST be configured** on border-VTEP even if there is no local ports transporting these extended VLANs

```
8360-1# show vlan 10

------------------------------------------------------------------------------------------------------
VLAN  Name                                Status  Reason                Type        Interfaces
------------------------------------------------------------------------------------------------------
10    VLAN10                              up      ok                    static      vxlan1
```




```
interface vxlan 1
    source ip 192.168.2.8
    vxlan-counters aggregate
    no shutdown
    vni 10010
        vlan 10
    vni 10012
        vlan 12
    vni 100001
        vrf VRF1
        routing
    vni 100002
        vrf VRF2
        routing
!
virtual-mac 02:00:00:00:05:00
```

VXLAN tunnel source IP = standalone unicast Loopback0

Optional counters for VXLAN statistics

For VLANs extended between Fabrics
1-to-1 VLAN-VNI mapping must be added on border-VTEP

L3 VNI associated to VRFs

VTEP virtual-MAC for L3VNI functionality

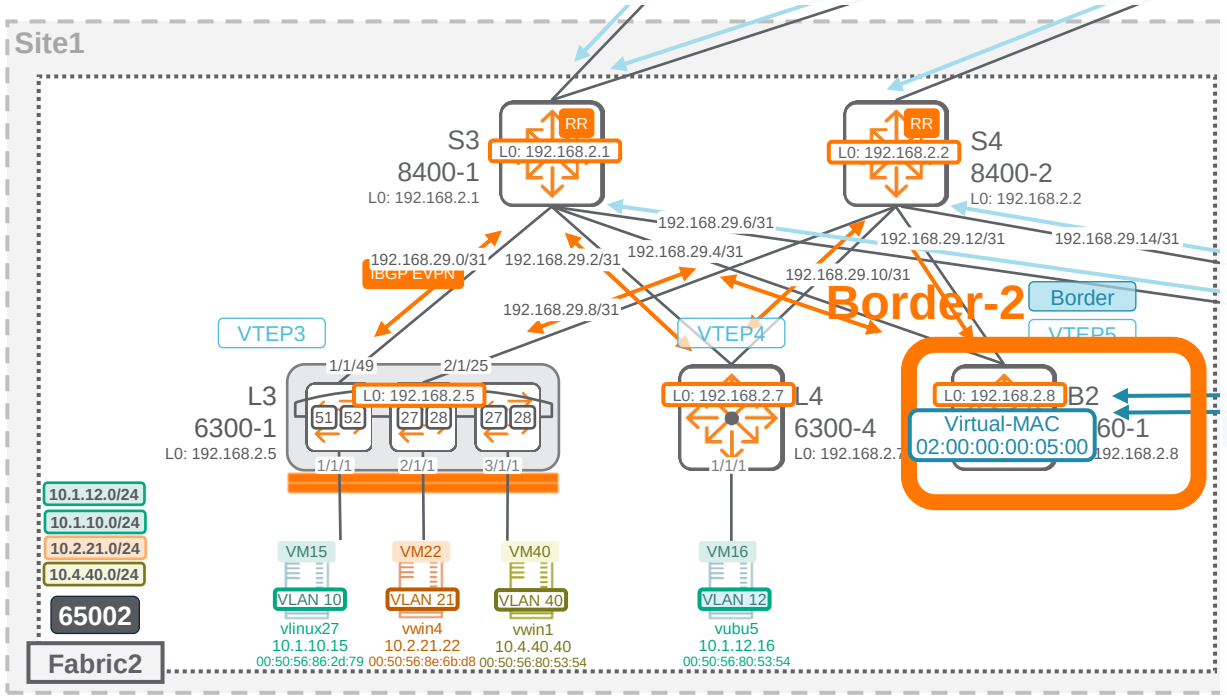
B2 (Border-2) Configuration

Routing

```
router ospf 1
    router-id 192.168.2.8
    max-metric router-lsa include-stub on-startup 300
    passive-interface default
    graceful-restart ignore-lost-interface
    trap-enable
    redistribute local loopback route-map connected-ospf
    area 0.0.0.0
```

best practice for HA during reboot

loopback redistribution in OSPF




```
ip prefix-list endpoint-VRF1 seq 10 permit 10.1.0.0/16 le 32
ip prefix-list endpoint-VRF2 seq 10 permit 10.2.0.0/16 le 32
ip prefix-list endpoint-VRF4 seq 10 permit 10.4.0.0/16 le 32
!
route-map connected-bgp-VRF1 permit seq 10
     match ip address prefix-list endpoint-VRF1
route-map connected-bgp-VRF2 permit seq 10
     match ip address prefix-list endpoint-VRF2
route-map connected-bgp-VRF4 permit seq 10
     match ip address prefix-list endpoint-VRF4
route-map connected-ospf permit seq 10
     set tag 1000
!
router bgp 65002
    bgp router-id 192.168.2.8
    trap-enable
    bgp fast-external-fallover
    bgp log-neighbor-changes
    bgp deterministic-med
    bgp always-compare-med
    neighbor border-leader peer-group
    neighbor border-leader remote-as 65001
    neighbor border-leader description Fabric1 BorderLeader1
    neighbor border-leader password ciphertext AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
    neighbor border-leader fall-over
    neighbor border-leader ebgp-multihop 10
    neighbor border-leader update-source loopback 0
    neighbor spine-RR peer-group
    neighbor spine-RR remote-as 65002
    neighbor spine-RR description Spine and RR peer-group
    neighbor spine-RR password ciphertext AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
    neighbor spine-RR fall-over
    neighbor spine-RR update-source loopback 0
    neighbor 192.168.1.3 peer-group border-leader
    neighbor 192.168.1.4 peer-group border-leader
    neighbor 192.168.2.1 peer-group spine-RR
    neighbor 192.168.2.2 peer-group spine-RR
    address-family l2vpn evpn
        neighbor border-leader send-community both
        neighbor spine-RR next-hop-self
        neighbor spine-RR send-community both
        neighbor 192.168.1.3 activate
        neighbor 192.168.1.4 activate
        neighbor 192.168.2.1 activate
        neighbor 192.168.2.2 activate
    exit-address-family
!
    vrf VRF1
        bgp router-id 192.168.2.8
        address-family ipv4 unicast
            redistribute local loopback
            redistribute connected route-map connected-bgp-VRF1
        exit-address-family
!
    vrf VRF2
        bgp router-id 192.168.2.8
        address-family ipv4 unicast
            redistribute local loopback
            redistribute connected route-map connected-bgp-VRF2
        exit-address-family
!
    vrf VRF4
        bgp router-id 192.168.2.8
        address-family ipv4 unicast
```

Control for route injection

tag value defined for local Fabric loopback

If used, must be the same in the BGP domain

Fall-over notification from FIB to BGP

ebgp-multihop (adjust # of intermediate hops)

iBGP peering between loopbacks

iBGP peering between loopbacks

Extended community is required

Used to populate EVPN Type-5 IP prefix routes

NEW

Route-map outcome

- Next-hop rewrite (and associated router MAC)

Next-Hop setting between sites with single-fabric

Example: Fabric4 to Fabric1

No need for EVPN route-map

Originated EVPN route from VTEP8 connecting VM18
Received iBGP routes:

```
8360-3# show bgp l2vpn evpn neighbors 192.168.4.3 routes route-type 2 | inc 10.1.10.18
*>i [2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]          192.168.4.3     0          100        0       ?
*>i [2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]          192.168.4.3     0          100        0       ?
```

L2VNI route
L3VNI route

Default iBGP to eBGP routes advertisement (eNHS):
From eBGP peers, border-leader-VTEP7 is seen as the next-hop for routes coming from AS 65004
Advertised eBGP routes:

```
8360-3# show bgp l2vpn evpn neighbors 192.168.1.3 advertised-route route-type 2 | include 10.1.10.18
*>e [2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]          192.168.40.1    0          0          0       65004 ?
*>e [2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]          192.168.40.1    0          0          0       65004 ?
```

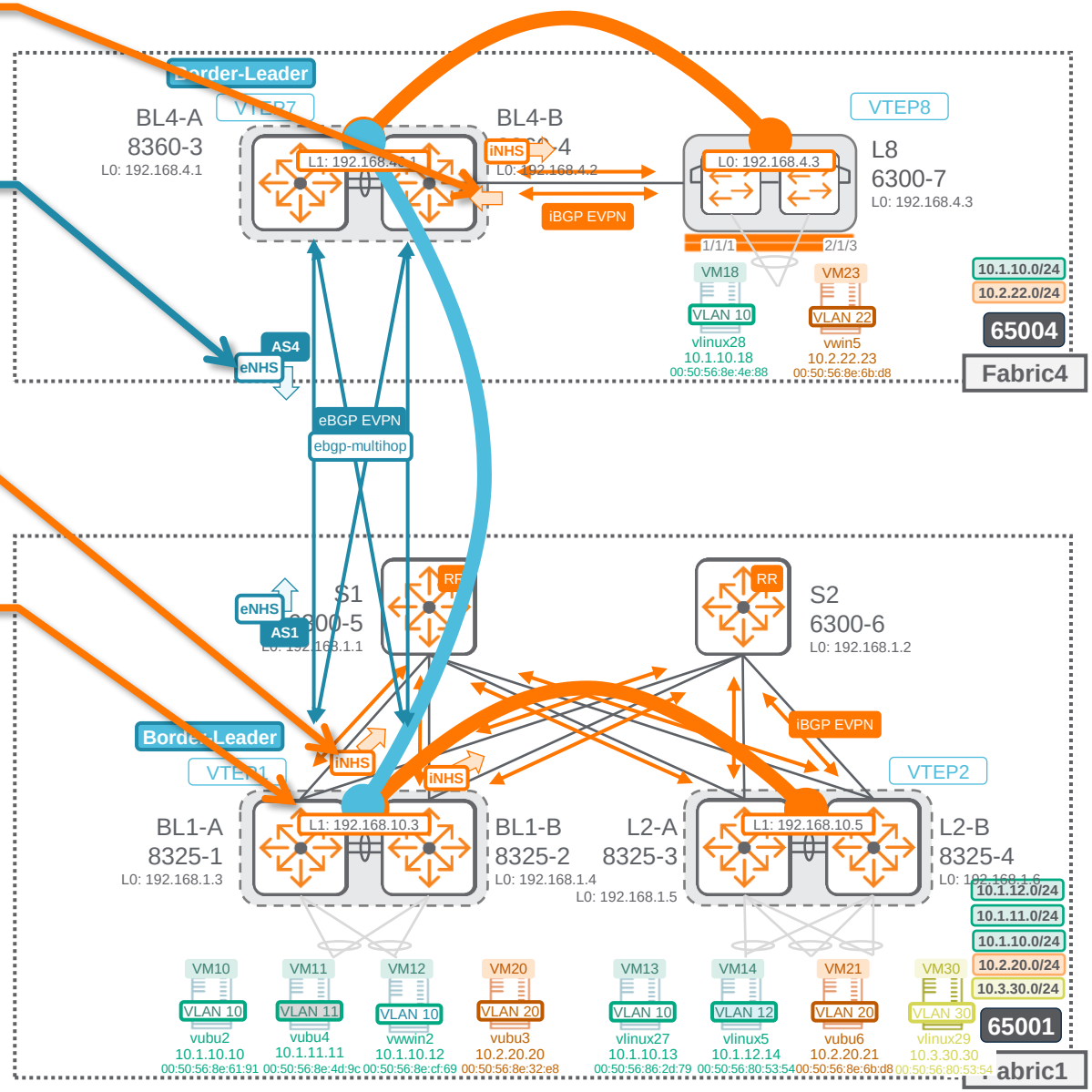
```
address-family l2vpn evpn
        ...
        neighbor spine-RR next-hop-self
```

Configurable eBGP to iBGP routes advertisement (iNHS):
From iBGP peers, border-leader-VTEP1 is seen as the next-hop for all external routes
Received eBGP routes:

```
8325-1# sh bgp l2 e nei 192.168.4.1 routes route-type 2 | include 10.1.10.18
* e [2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]          192.168.40.1    0          100        0       65004 ?
* e [2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]          192.168.40.1    0          100        0       65004 ?

8325-1# sh bgp l2 e nei 192.168.4.2 routes route-type 2 | include 10.1.10.18
*>e [2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]          192.168.40.1    0          100        0       65004 ?
*>e [2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]          192.168.40.1    0          100        0       65004 ?
```

Advertised eBGP route to iBGP peers:

```
8325-1# sh bgp l2 e nei 192.168.1.1 advertised-routes route-type 2 | incl 10.1.10.18
*>i [2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]          192.168.10.3    0          100        0       65004 ?
*>i [2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]          192.168.10.3    0          100        0       65004 ?
```



with Router MAC rewrite (for L3VNI routing)

Example: Fabric4 to Fabric1

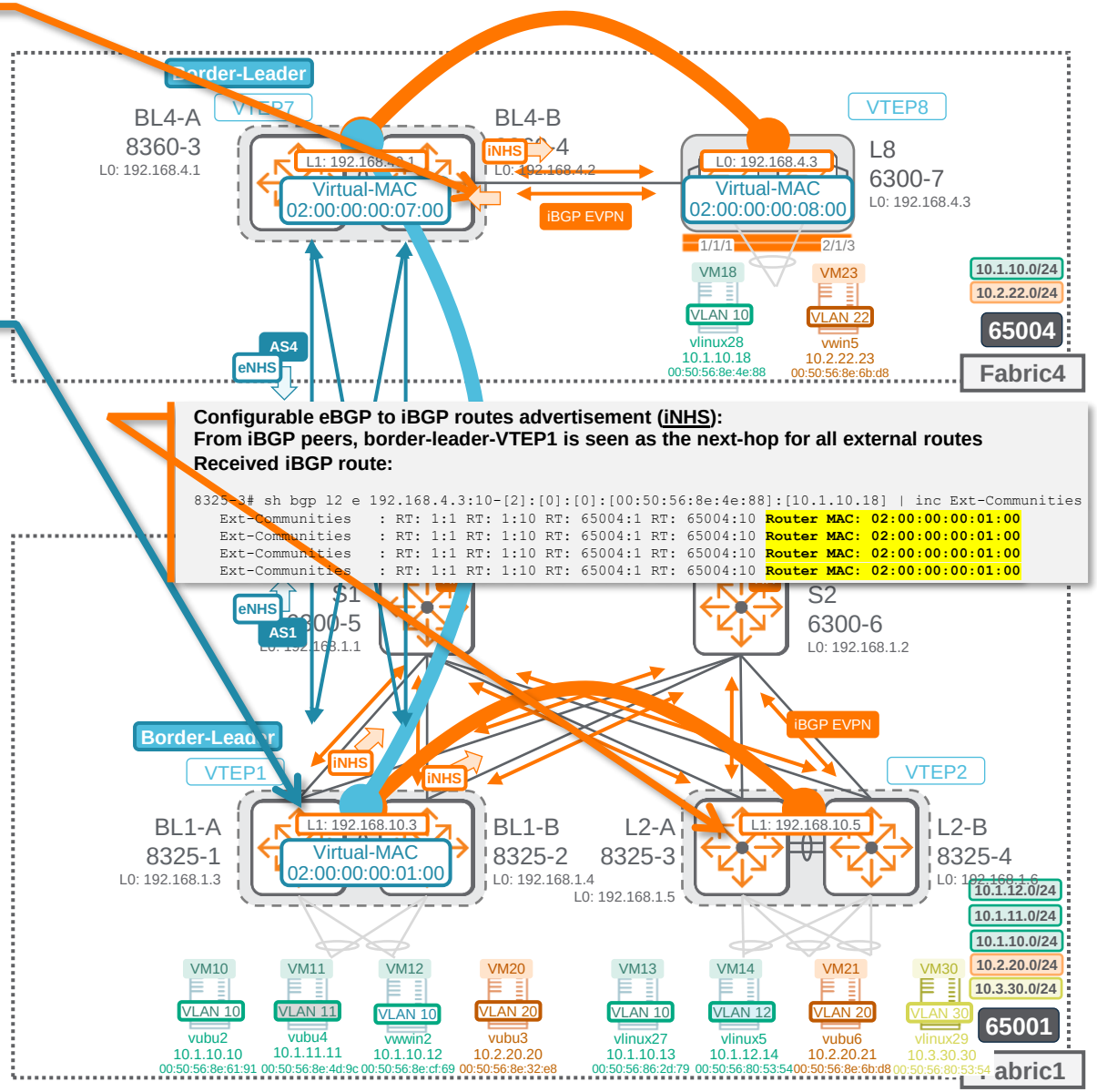
No CLI command needed

Originated EVPN route from VTEP8 connecting VM18
Received iBGP routes:

```
8360-3# sh bgp l2 e 192.168.4.3:10-[2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18] | inc Ext-Communities
   Ext-Communities    : RT: 1:1 RT: 1:10 RT: 65004:1 RT: 65004:10 Router MAC: 02:00:00:00:08:00
   Ext-Communities    : RT: 1:1 RT: 1:10 RT: 65004:1 RT: 65004:10 Router MAC: 02:00:00:00:08:00
```

Default iBGP to eBGP routes advertisement (eNHS):
From eBGP peers, border-leader-VTEP7 is seen as the next-hop for routes coming from AS 65004

```
8325-1# sh bgp l2 e 192.168.4.3:10-[2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]

VRF : default
BGP Local AS 65001       BGP Router-id 192.168.1.3

   Network           : 192.168.4.3:10-[2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]
   Nexthop           : 192.168.40.1
   vni               : 10010                    vni_type            : L2VNI
   Peer              : 192.168.4.1              Origin              : incomplete
   Metric            : 0                        Local Pref          : 100
   Weight            : 0                        Calc. Local Pref    : 100
   Best              : Yes                      Valid               : Yes
   Type              : external                 Stale               : No
   Originator ID     : 0.0.0.0
   Aggregator ID     :
   Aggregator AS     :
   Atomic Aggregate  :

   AS-Path           :  65004

   Cluster List      :
   Communities       :
   Ext-Communities   : RT: 1:1 RT: 1:10 RT: 65004:1 RT: 65004:10 Router MAC: 02:00:00:00:07:00


   Network           : 192.168.4.3:10-[2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]
   Nexthop           : 192.168.40.1
   vni               : 100001                   vni_type            : L3VNI
   Peer              : 192.168.4.1              Origin              : incomplete
   Metric            : 0                        Local Pref          : 100
   Weight            : 0                        Calc. Local Pref    : 100
   Best              : Yes                      Valid               : Yes
   Type              : external                 Stale               : No
   Originator ID     : 0.0.0.0
   Aggregator ID     :
   Aggregator AS     :
   Atomic Aggregate  :

   AS-Path           :  65004

   Cluster List      :
   Communities       :
   Ext-Communities   : RT: 1:1 RT: 1:10 RT: 65004:1 RT: 65004:10 Router MAC: 02:00:00:00:07:00
```



Next-Hop setting between Fabrics of same site

Example: Fabric1 to Fabric2

No need for EVPN route-map

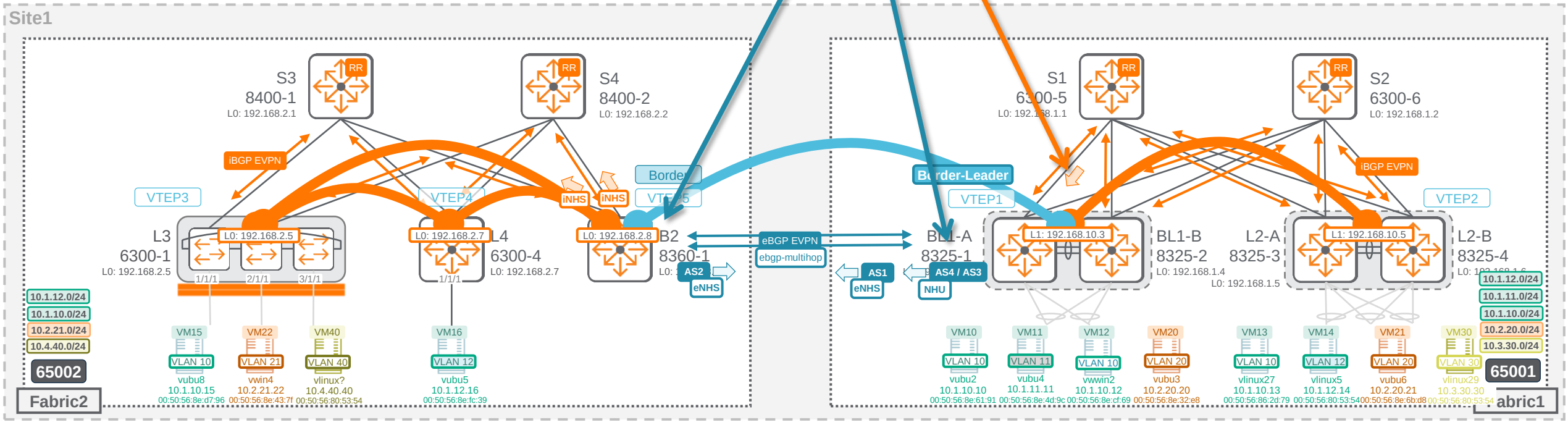
```
8325-1# show bgp l2vpn evpn nei 192.168.1.1 routes route-type 2 | incl 10.1.12.14
*>i [2]:[0]:[0]:[00:50:56:80:53:54]:[10.1.12.14]          192.168.10.5    0          100        0       ?
*>i [2]:[0]:[0]:[00:50:56:80:53:54]:[10.1.12.14]          192.168.10.5    0          100        0       ?
```

Default iBGP to eBGP routes advertisement (eNHS):
From eBGP peers, border-leader-VTEP7 is seen as the next-hop for routes coming from AS 65004

```
8325-1# show bgp l2vpn evpn neighbors 192.168.2.8 advertised-routes route-type 2 | inc 10.1.12.14
*>e [2]:[0]:[0]:[00:50:56:80:53:54]:[10.1.12.14]          192.168.10.3    0          0          0       65001 ?
*>e [2]:[0]:[0]:[00:50:56:80:53:54]:[10.1.12.14]          192.168.10.3    0          0          0       65001 ?
```

Received eBGP route:

```
8360-1# show bgp l2vpn evpn neighbors 192.168.1.3 routes route-type 2 | include 10.1.12.14
*>e [2]:[0]:[0]:[00:50:56:80:53:54]:[10.1.12.14]          192.168.10.3    0          100        0       65001 ?
*>e [2]:[0]:[0]:[00:50:56:80:53:54]:[10.1.12.14]          192.168.10.3    0          100        0       65001 ?
```



with Router MAC rewrite (for L3VNI routing)

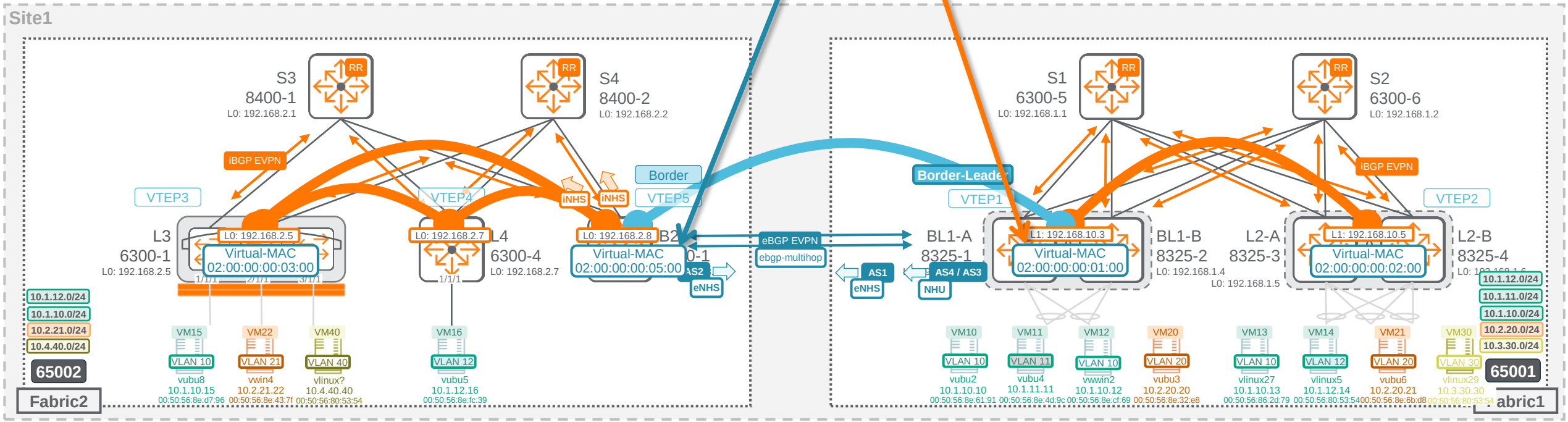
Example: Fabric1 to Fabric2

No CLI command

```
8325-1# show bgp l2vpn evpn 192.168.10.5:12-[2]:[0]:[0]:[00:50:56:80:53:54]:[10.1.12.14] | inc Ext-Communities
   Ext-Communities    : RT: 1:1 RT: 65001:1 RT: 65001:12 Router MAC: 02:00:00:00:02:00
   Ext-Communities    : RT: 1:1 RT: 65001:1 RT: 65001:12 Router MAC: 02:00:00:00:02:00
   Ext-Communities    : RT: 1:1 RT: 65001:1 RT: 65001:12 Router MAC: 02:00:00:00:02:00
   Ext-Communities    : RT: 1:1 RT: 65001:1 RT: 65001:12 Router MAC: 02:00:00:00:02:00
```

Received eBGP routes:

```
8360-1# show bgp l2vpn evpn 192.168.10.5:12-[2]:[0]:[0]:[00:50:56:80:53:54]:[10.1.12.14] | inc Ext-Communities
   Ext-Communities    : RT: 1:1 RT: 65001:1 RT: 65001:12 Router MAC: 02:00:00:00:01:00
   Ext-Communities    : RT: 1:1 RT: 65001:1 RT: 65001:12 Router MAC: 02:00:00:00:01:00
   Ext-Communities    : RT: 1:1 RT: 65001:1 RT: 65001:12 Router MAC: 02:00:00:00:01:00
   Ext-Communities    : RT: 1:1 RT: 65001:1 RT: 65001:12 Router MAC: 02:00:00:00:01:00
```



Next-Hop setting between sites with multiple fabrics

Example: Fabric2 to Fabric4

EVPN route-map is needed

Originated EVPN route from VTEP8 connecting VM18
Received iBGP routes:

```
8360-3# show bgp l2vpn evpn neighbors 192.168.4.3 routes route-type 2 | inc 10.1.10.18
*>i [2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]          192.168.4.3     0          100        0       ?
*>i [2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]          192.168.4.3     0          100        0       ?
```

```
8360-3# show bgp l2vpn evpn neighbors 192.168.1.3 advertised-route route-type 2 | include 10.1.10.18
*>e [2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]          192.168.40.1    0          0          0       65004 ?
*>e [2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]          192.168.40.1    0          0          0       65004 ?
```

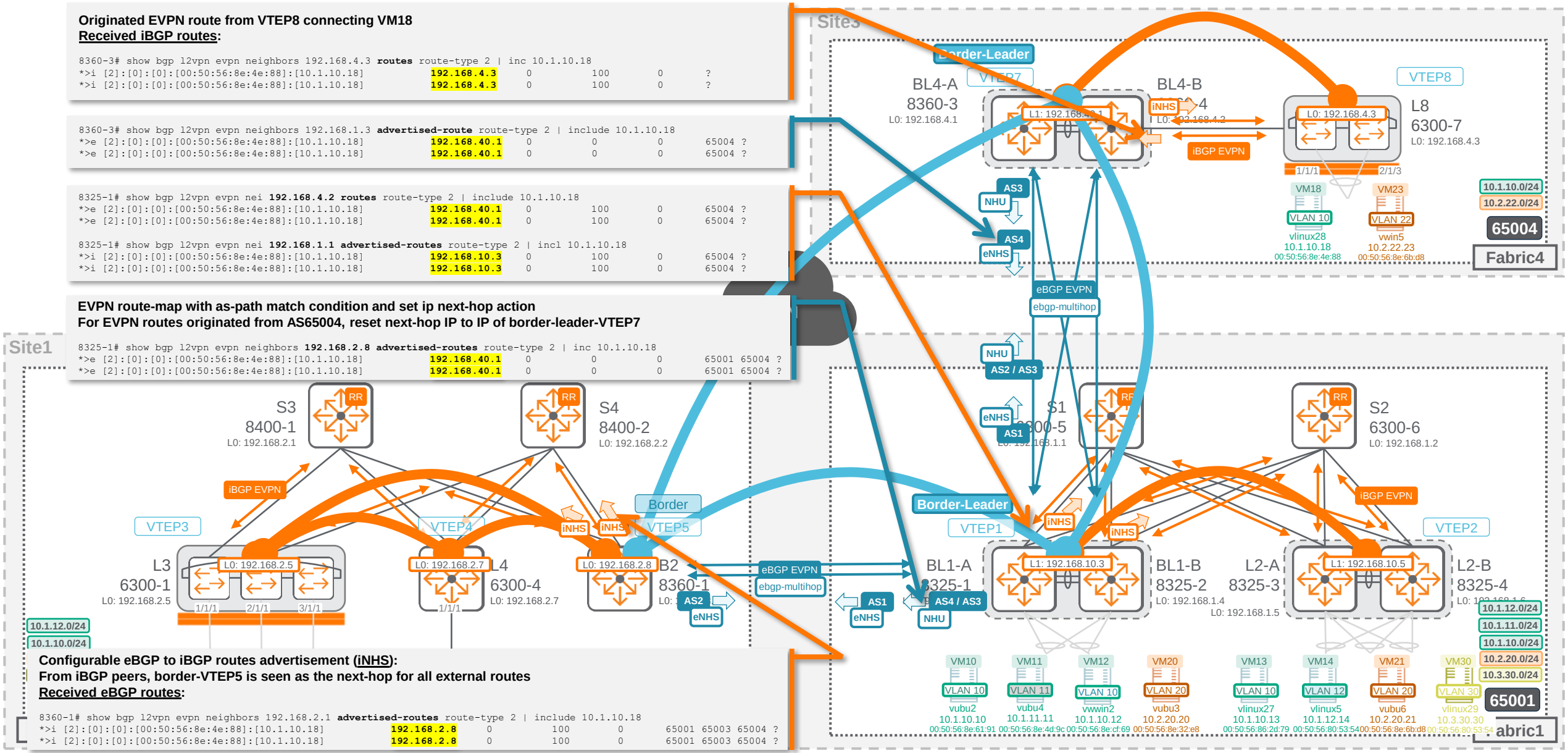
```
8325-1# show bgp l2vpn evpn nei 192.168.4.2 routes route-type 2 | include 10.1.10.18
*>e [2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]          192.168.40.1    0          100        0       65004 ?
*>e [2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]          192.168.40.1    0          100        0       65004 ?

8325-1# show bgp l2vpn evpn nei 192.168.1.1 advertised-routes route-type 2 | incl 10.1.10.18
*>i [2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]          192.168.10.3    0          100        0       65004 ?
*>i [2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]          192.168.10.3    0          100        0       65004 ?
```

EVPN route-map with as-path match condition and set ip next-hop action
For EVPN routes originated from AS65004, reset next-hop IP to IP of border-leader-VTEP7

```
8325-1# show bgp l2vpn evpn neighbors 192.168.2.8 advertised-routes route-type 2 | inc 10.1.10.18
*>e [2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]          192.168.40.1    0          0          0       65001 65004 ?
*>e [2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]          192.168.40.1    0          0          0       65001 65004 ?
```

Configurable eBGP to iBGP routes advertisement (iNHS):
From iBGP peers, border-VTEP5 is seen as the next-hop for all external routes
Received eBGP routes:

```
8360-1# show bgp l2vpn evpn neighbors 192.168.2.1 advertised-routes route-type 2 | include 10.1.10.18
*>i [2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]          192.168.2.8     0          100        0       65001 65003 65004 ?
*>i [2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]          192.168.2.8     0          100        0       65001 65003 65004 ?
```



with Router MAC rewrite (for L3VNI routing)

Example: Fabric2 to Fabric4

No CLI command

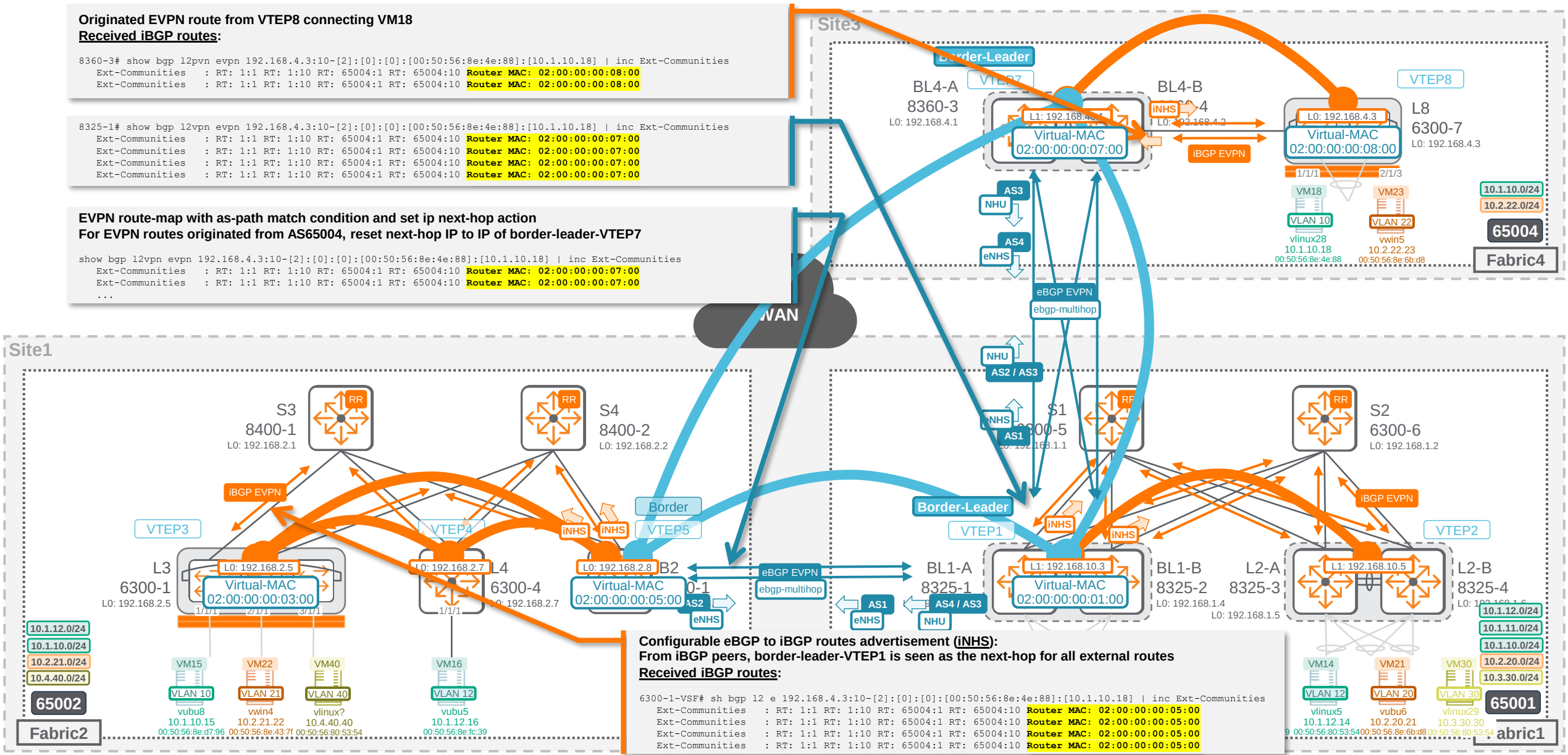
Originated EVPN route from VTEP8 connecting VM18
Received iBGP routes:

```
8360-3# show bgp l2pvn evpn 192.168.4.3:10-[2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18] | inc Ext-Communities
   Ext-Communities    : RT: 1:1 RT: 1:10 RT: 65004:1 RT: 65004:10 Router MAC: 02:00:00:00:08:00
   Ext-Communities    : RT: 1:1 RT: 1:10 RT: 65004:1 RT: 65004:10 Router MAC: 02:00:00:00:08:00
```

```
8325-1# show bgp l2vpn evpn 192.168.4.3:10-[2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18] | inc Ext-Communities
   Ext-Communities    : RT: 1:1 RT: 1:10 RT: 65004:1 RT: 65004:10 Router MAC: 02:00:00:00:07:00
   Ext-Communities    : RT: 1:1 RT: 1:10 RT: 65004:1 RT: 65004:10 Router MAC: 02:00:00:00:07:00
   Ext-Communities    : RT: 1:1 RT: 1:10 RT: 65004:1 RT: 65004:10 Router MAC: 02:00:00:00:07:00
   Ext-Communities    : RT: 1:1 RT: 1:10 RT: 65004:1 RT: 65004:10 Router MAC: 02:00:00:00:07:00
```

EVPN route-map with as-path match condition and set ip next-hop action
For EVPN routes originated from AS65004, reset next-hop IP to IP of border-leader-VTEP7

```
show bgp l2vpn evpn 192.168.4.3:10-[2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18] | inc Ext-Communities
   Ext-Communities    : RT: 1:1 RT: 1:10 RT: 65004:1 RT: 65004:10 Router MAC: 02:00:00:00:07:00
   Ext-Communities    : RT: 1:1 RT: 1:10 RT: 65004:1 RT: 65004:10 Router MAC: 02:00:00:00:07:00
   ...
```

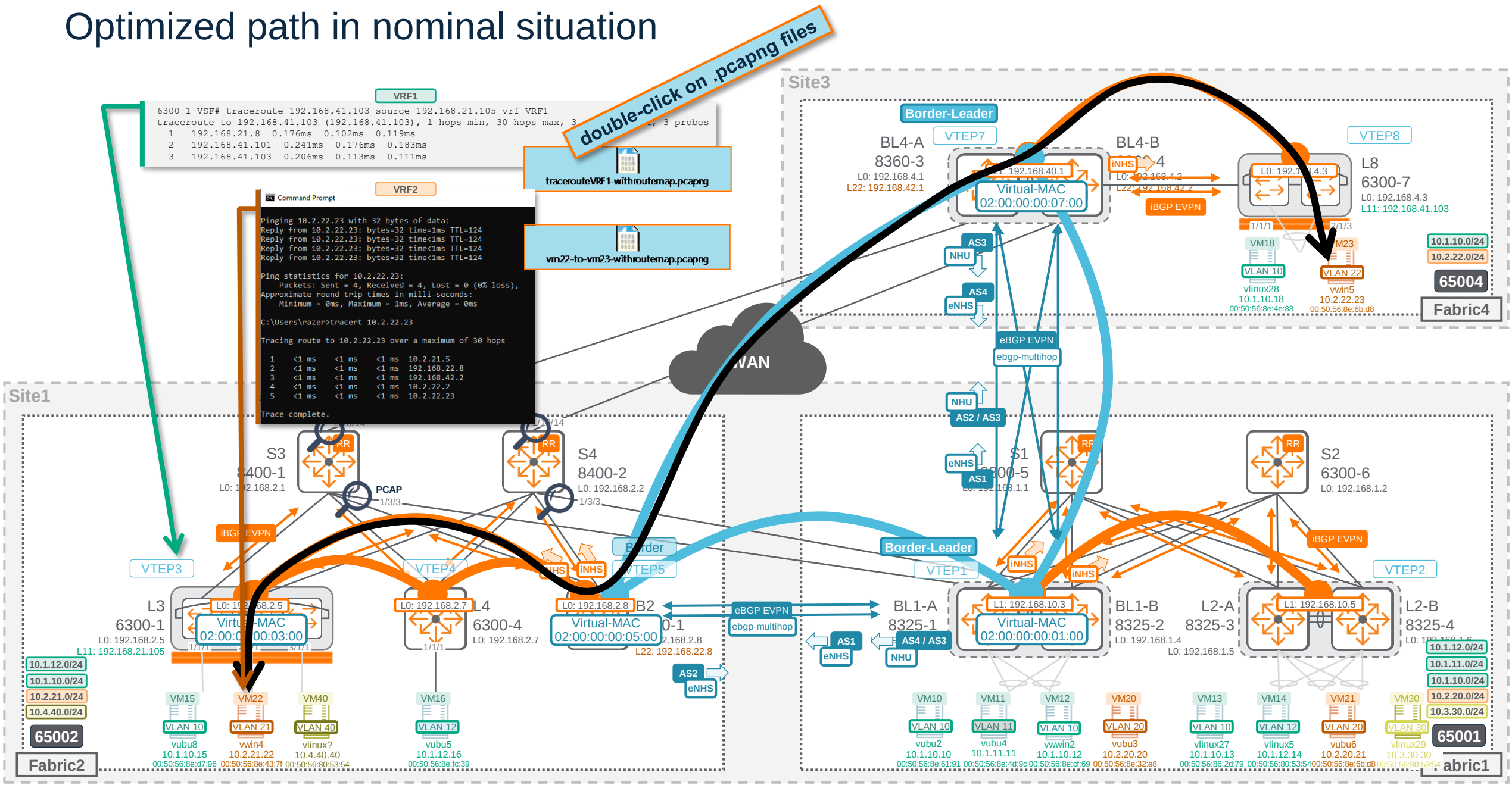


Configurable eBGP to iBGP routes advertisement (iNHS):
From iBGP peers, border-leader-VTEP1 is seen as the next-hop for all external routes
Received iBGP routes:

```
6300-1-VSF# sh bgp l2 e 192.168.4.3:10-[2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18] | inc Ext-Communities
   Ext-Communities    : RT: 1:1 RT: 1:10 RT: 65004:1 RT: 65004:10 Router MAC: 02:00:00:00:05:00
   Ext-Communities    : RT: 1:1 RT: 1:10 RT: 65004:1 RT: 65004:10 Router MAC: 02:00:00:00:05:00
   Ext-Communities    : RT: 1:1 RT: 1:10 RT: 65004:1 RT: 65004:10 Router MAC: 02:00:00:00:05:00
   Ext-Communities    : RT: 1:1 RT: 1:10 RT: 65004:1 RT: 65004:10 Router MAC: 02:00:00:00:05:00
```

Traffic between Fabric2 and Fabric4

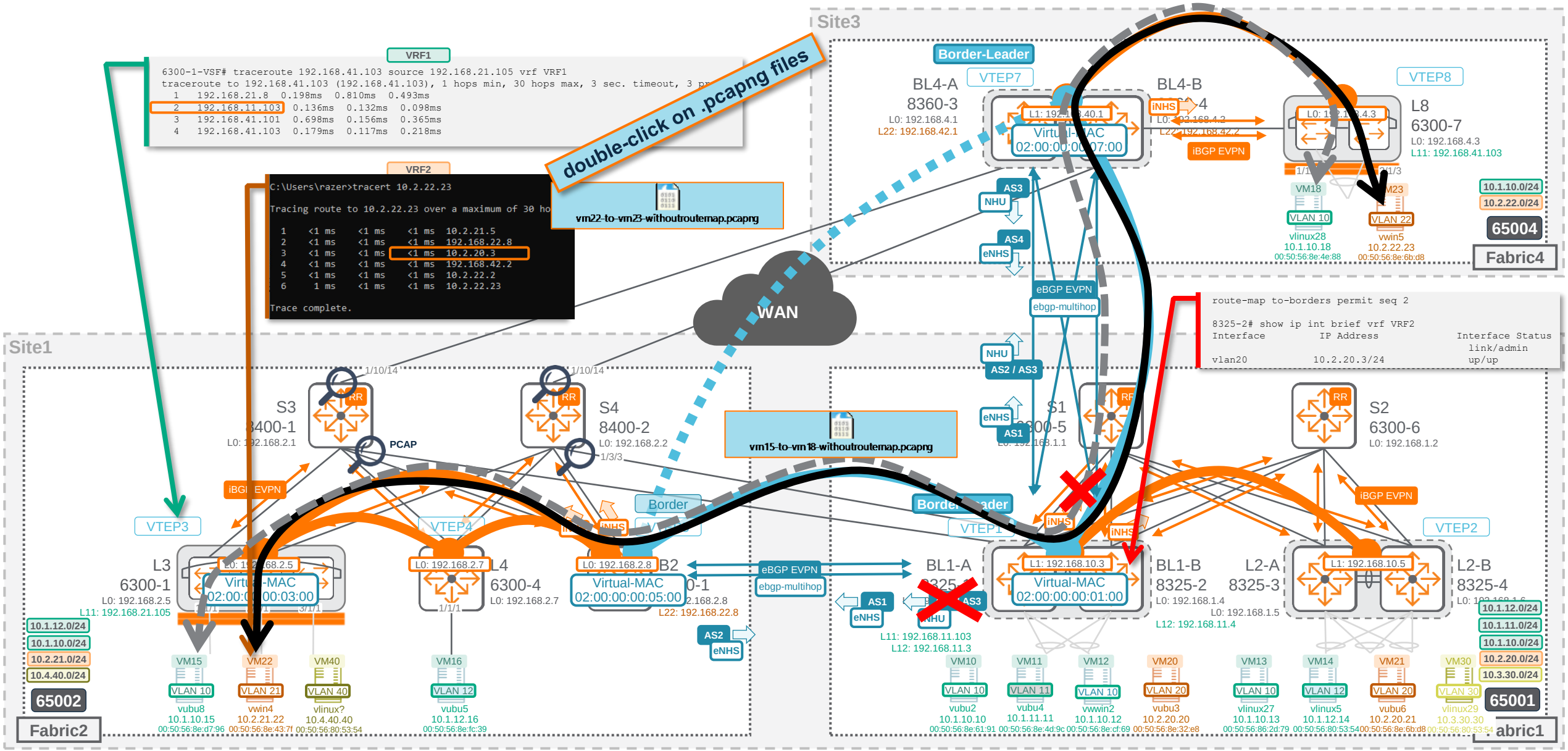
Optimized path in nominal situation



Traffic between Fabric2 and Fabric4

Non-optimized path: route-map removal

Only possible for L3VNI
Split-Horizon rule prevents L2VNI traffic



NEW

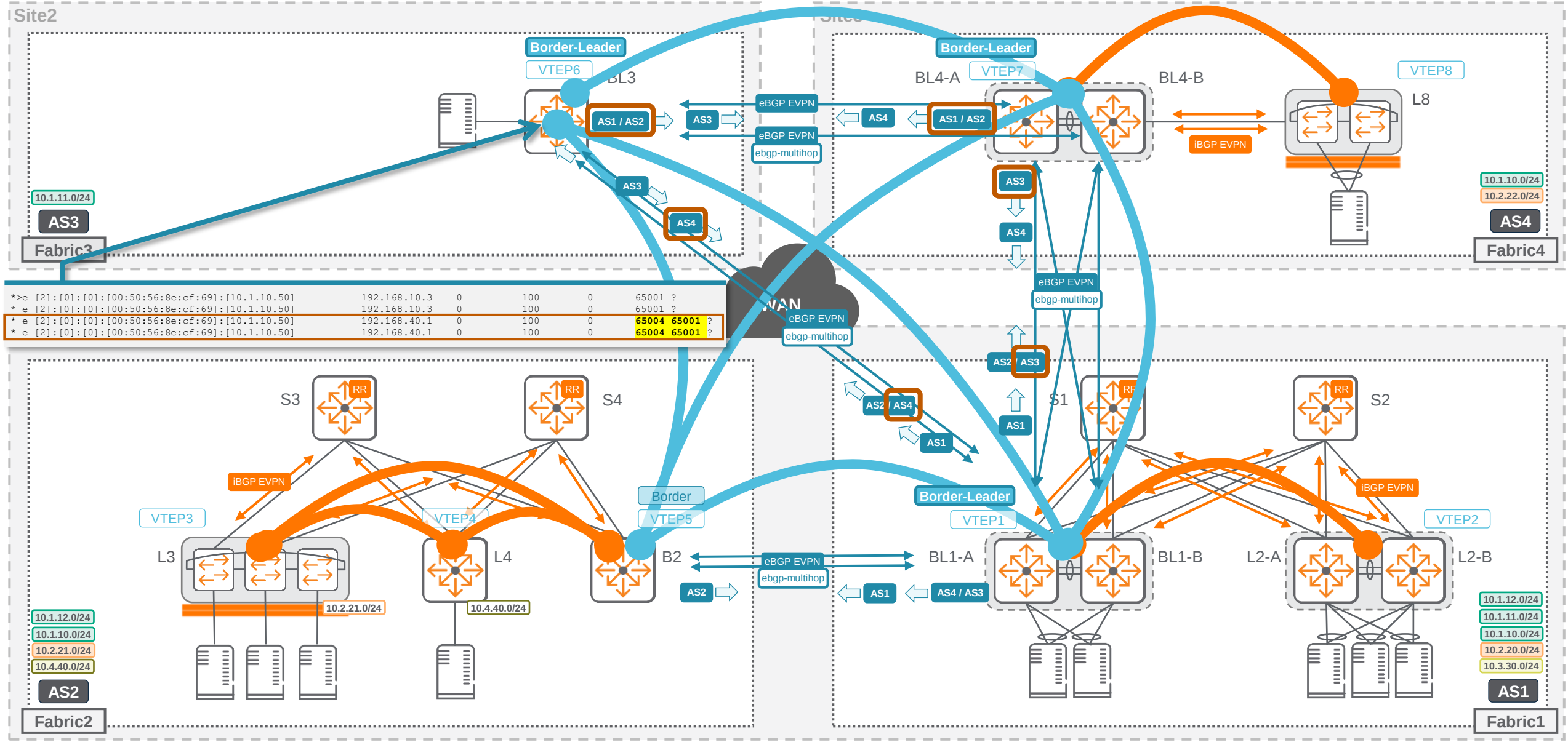
Route-map outcome

- Site Transit Removal

Site-transit functionality for EVPN

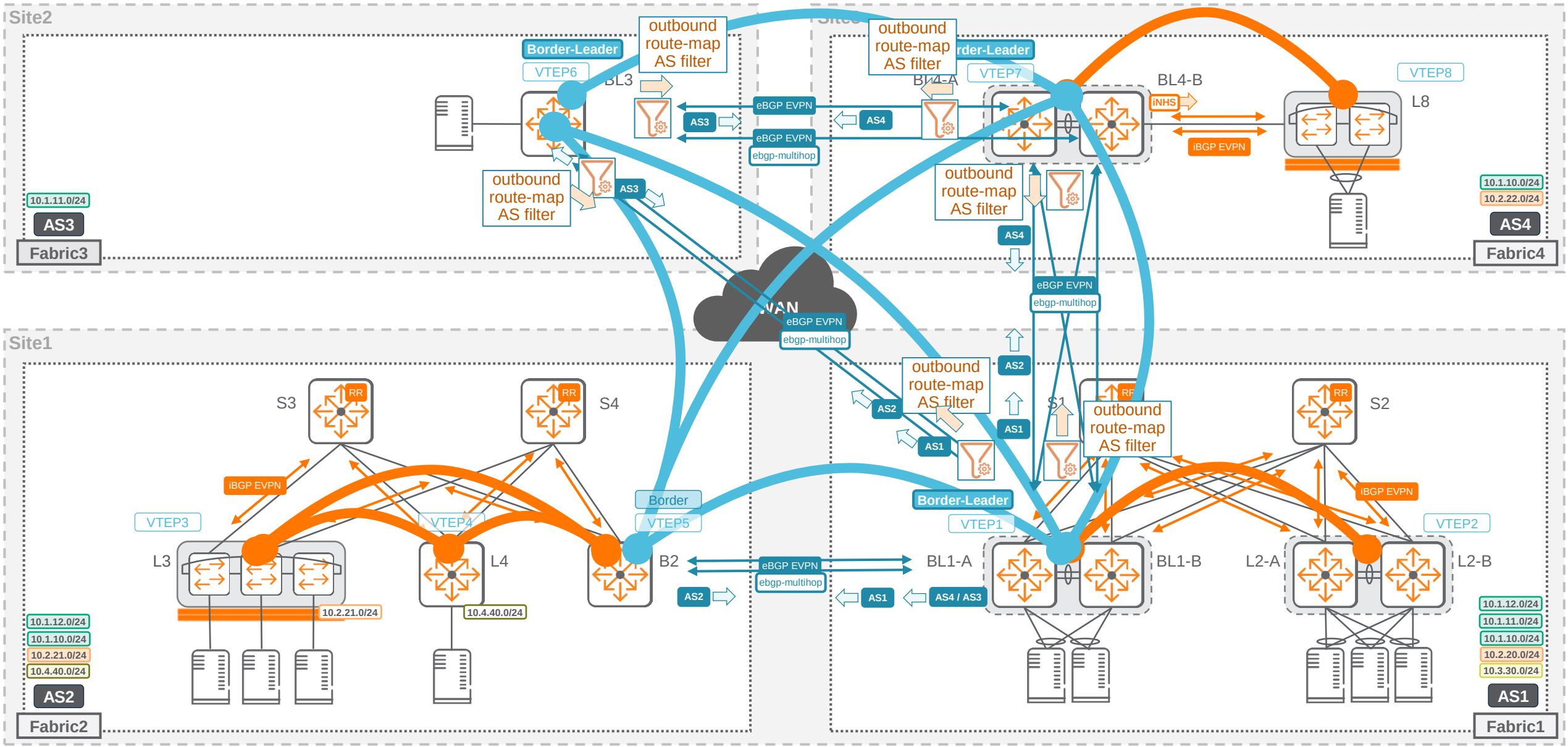
Nominal but unnecessary

Reliability of the EVPN control-plane relies on the underlay IPv4 loopback routing as eBGP EVPN TCP sessions are “multi-hop” between loopbacks



Site-transit functionality removal for EVPN control-plane

Benefit: much less routes on switches and easier troubleshooting



EVPN route-map: combining VTEP next-hop and site-transit removal

Different outbound route-map to borders and to border-leaders

border-leader to borders

Sourcing ASN & Remote VTEP IP

```
ip aspath-list fabric2 seq 10 permit _65002$
ip aspath-list fabric3 seq 10 permit _65003$
ip aspath-list fabric4 seq 10 permit _65004$
ip aspath-list fabricid seq 10 permit _<ASN>$
!
route-map to-borders permit seq 10
     match aspath-list fabric2
     set ip next-hop <remote-vtep-ip-fabric2>
route-map to-borders permit seq 20
     match aspath-list fabric3
     set ip next-hop <remote-vtep-ip-fabric3>
route-map to-borders permit seq 30
     match aspath-list fabric4
     set ip next-hop <remote-vtep-ip-fabric4>
route-map to-borders permit seq <n>
     match aspath-list fabricid
     set ip next-hop <remote-vtep-ip-fabricid>
route-map to-borders permit seq 1000
!
router bgp <ASN>
…
address-family l2vpn evpn
   neighbor borders route-map to-borders out
   neighbor borders send-community both
```

border-leader to border-leaders

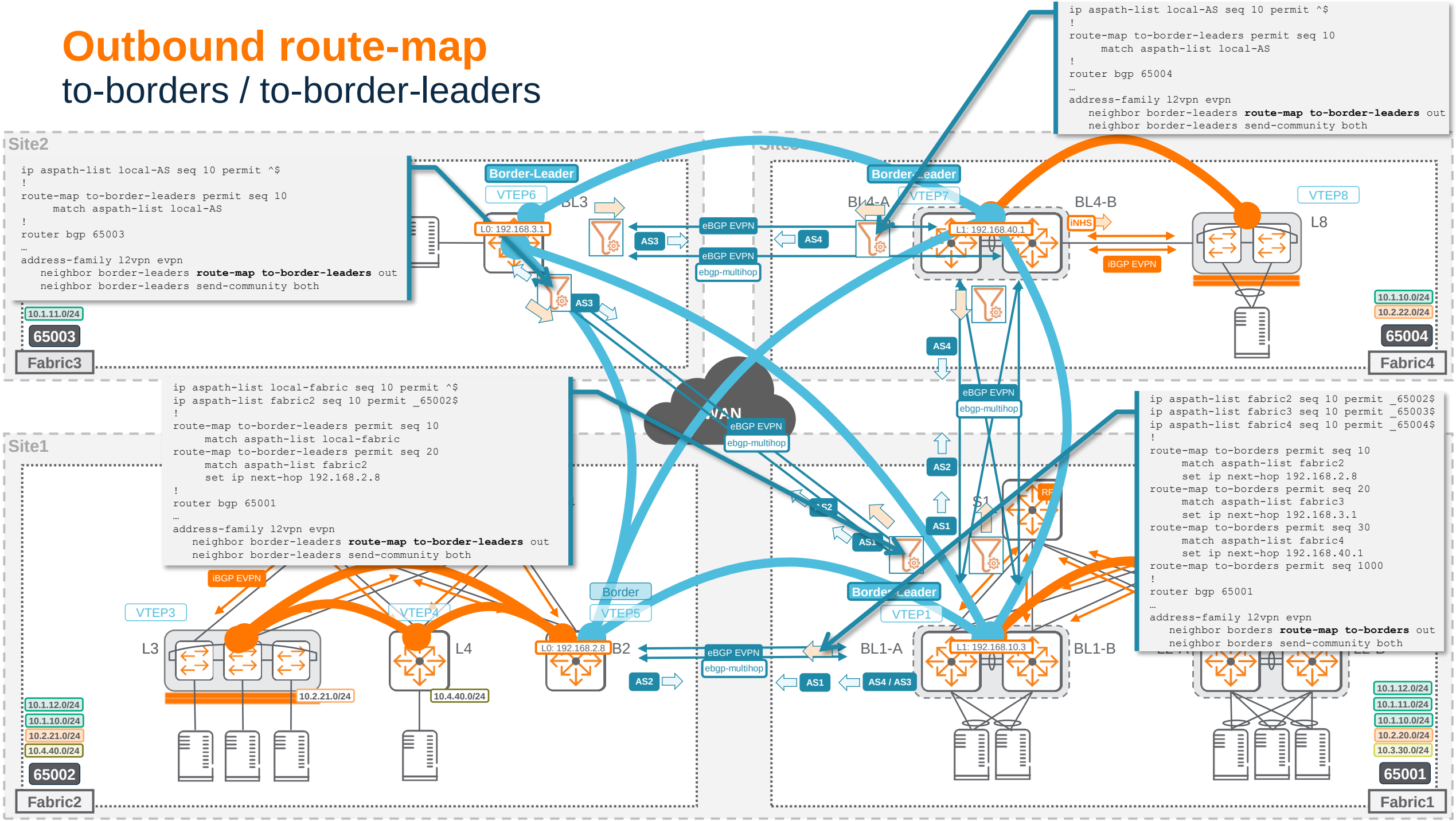
- Single-fabric site:

```
ip aspath-list local-AS seq 10 permit ^$
!
route-map to-border-leaders permit seq 10
     match aspath-list local-AS
!
router bgp <ASN>
…
address-family l2vpn evpn
   neighbor border-leaders route-map to-border-leaders out
   neighbor border-leaders send-community both
```

- Multi-fabric site:

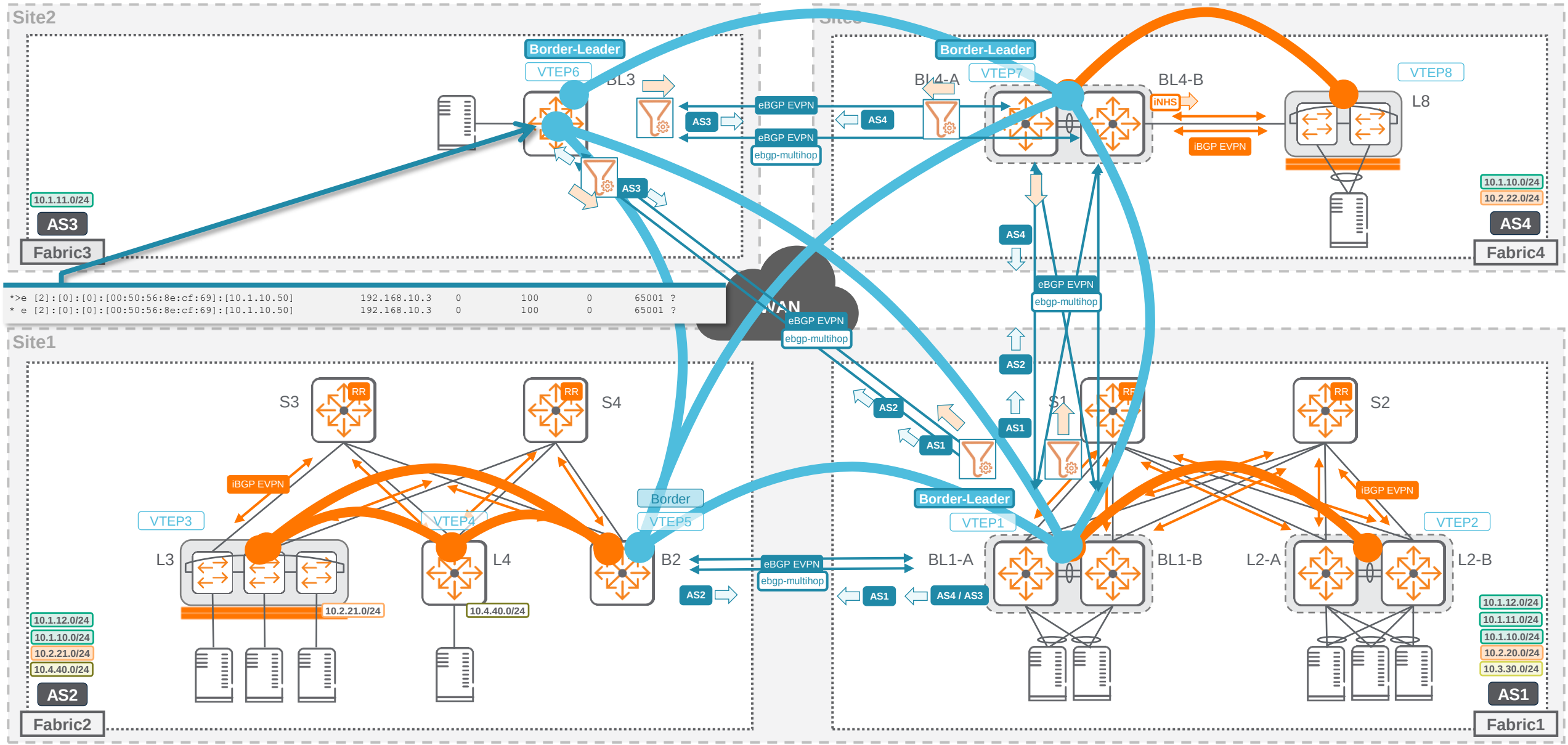
```
ip aspath-list local-fabric seq 10 permit ^$
!
ip aspath-list fabric2 seq 10 permit _65002$
ip aspath-list fabricid seq 10 permit _<ASN>$
!
route-map to-border-leaders permit seq 10
     match aspath-list local-fabric
route-map to-border-leaders permit seq 20
     match aspath-list fabric2
     set ip next-hop 192.168.2.8
route-map to-border-leaders permit seq <n>
     match aspath-list fabricid
     set ip next-hop <remote-vtep-ip-fabricid>
!
router bgp <ASN>
…
address-family l2vpn evpn
   neighbor border-leaders route-map to-border-leaders out
   neighbor border-leaders send-community both
```

Outbound route-map
to-borders / to-border-leaders



Site-transit functionality removal for EVPN control-plane

Example after filtering



NEW

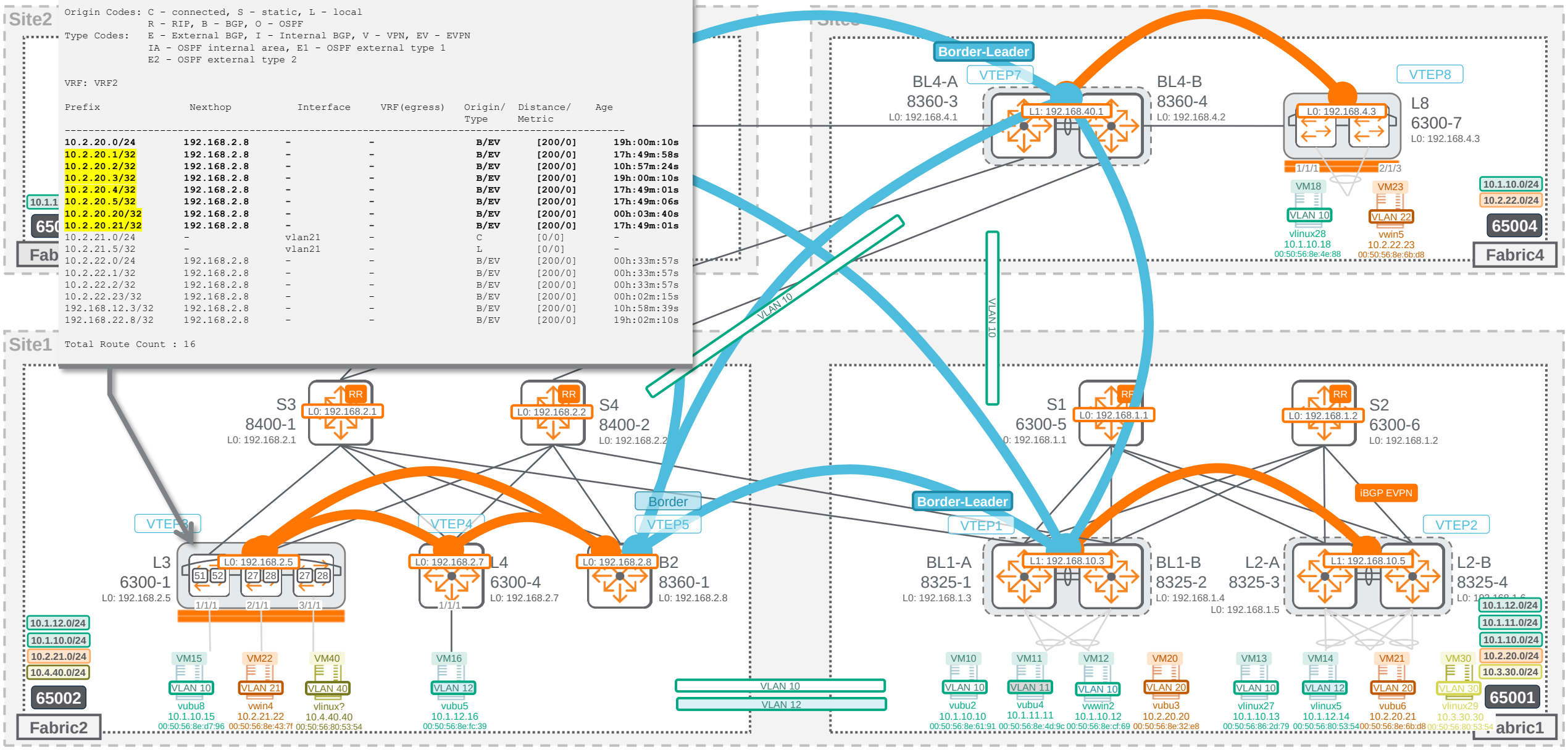
Route-map outcome

- Routes aggregation
- Outbound route-map filtering

Before Route Aggregation

Host routes in 10.2.20.0/24 are not needed as 10.2.20.0/24 is located only in Fabric1

```
6300-1-VSF# show ip route vrf VRF2

Displaying ipv4 routes selected for forwarding

Origin Codes: C - connected, S - static, L - local
              R - RIP, B - BGP, O - OSPF
Type Codes:   E - External BGP, I - Internal BGP, V - VPN, EV - EVPN
              IA - OSPF internal area, E1 - OSPF external type 1
              E2 - OSPF external type 2

VRF: VRF2

Prefix              Nexthop           Interface     VRF(egress)   Origin/  Distance/    Age
                                                                  Type     Metric
-------------------------------------------------------------------------------------------
10.2.20.0/24        192.168.2.8       -             -             B/EV     [200/0]      19h:00m:10s
10.2.20.1/32        192.168.2.8       -             -             B/EV     [200/0]      17h:49m:58s
10.2.20.2/32        192.168.2.8       -             -             B/EV     [200/0]      10h:57m:24s
10.2.20.3/32        192.168.2.8       -             -             B/EV     [200/0]      19h:00m:10s
10.2.20.4/32        192.168.2.8       -             -             B/EV     [200/0]      17h:49m:01s
10.2.20.5/32        192.168.2.8       -             -             B/EV     [200/0]      17h:49m:06s
10.2.20.20/32       192.168.2.8       -             -             B/EV     [200/0]      00h:03m:40s
10.2.20.21/32       192.168.2.8       -             -             B/EV     [200/0]      17h:49m:01s
10.2.21.0/24        -                 vlan21        -             C        [0/0]        -
10.2.21.5/32        -                 vlan21        -             L        [0/0]        -
10.2.22.0/24        192.168.2.8       -             -             B/EV     [200/0]      00h:33m:57s
10.2.22.1/32        192.168.2.8       -             -             B/EV     [200/0]      00h:33m:57s
10.2.22.2/32        192.168.2.8       -             -             B/EV     [200/0]      00h:33m:57s
10.2.22.23/32       192.168.2.8       -             -             B/EV     [200/0]      00h:02m:15s
192.168.12.3/32     192.168.2.8       -             -             B/EV     [200/0]      10h:58m:39s
192.168.22.8/32     192.168.2.8       -             -             B/EV     [200/0]      19h:02m:10s

Total Route Count : 16
```



Aggregation on Border-Leader-VTEP1

EVPN route-map: outbound filtering of unnecessary Host routes in 10.2.20.0/24

```
ip prefix-list vrf2-vlan20-summary seq 10 permit 10.2.20.0/24 ge 25 le 32
!
route-map to-borders deny seq 5
     match ip address prefix-list vrf2-vlan20-summary
!
route-map to-border-leaders deny seq 5
     match ip address prefix-list vrf2-vlan20-summary
!
router bgp 65001
...
   vrf VRF2
        bgp router-id 192.168.1.3
        bgp log-neighbor-changes
        address-family ipv4 unicast
            redistribute local loopback
            redistribute connected route-map connected-bgp-VRF2
            aggregate-address 10.2.20.0/24 summary-only
        exit-address-family
```

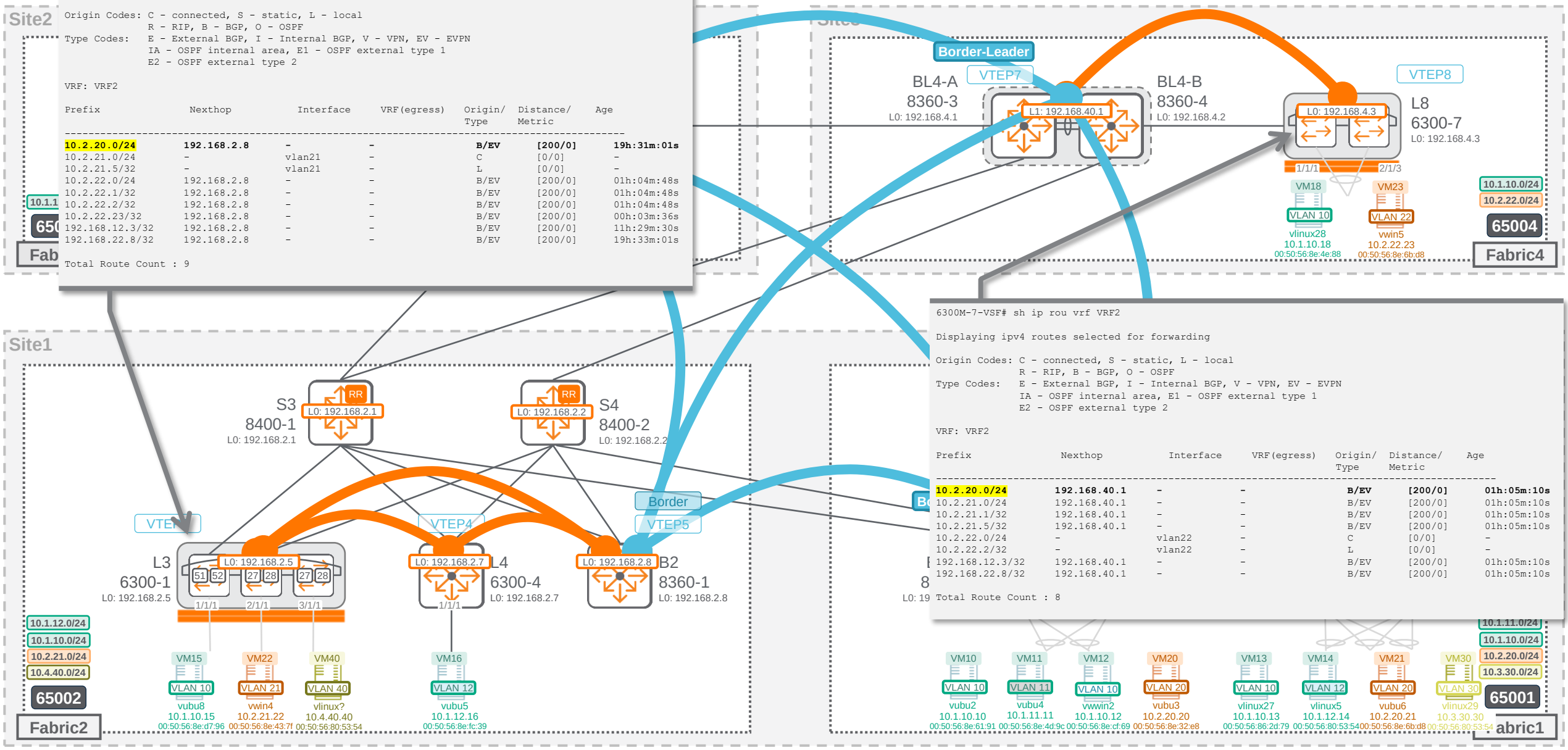
- Objective: in VRF2, other Fabrics do not need to receive host routes or more specific routes than 10.2.20.0/24 as 10.2.20.0/24 is local to Fabric1

- **Summary or summary-only command**: Difference does not matter as the summary-only command will filter routes only inside IPv4 AF. Host routes injected in EVPN AF are not filtered by this summary-only command.

After Route Aggregation

```
6300-1-VSF# sh ip rou vrf VRF2

Displaying ipv4 routes selected for forwarding

Origin Codes: C - connected, S - static, L - local
              R - RIP, B - BGP, O - OSPF
Type Codes:   E - External BGP, I - Internal BGP, V - VPN, EV - EVPN
              IA - OSPF internal area, E1 - OSPF external type 1
              E2 - OSPF external type 2

VRF: VRF2

Prefix              Nexthop          Interface     VRF(egress)   Origin/  Distance/    Age
                                                                 Type     Metric
-------------------------------------------------------------------------------------------
10.2.20.0/24        192.168.2.8      -             -             B/EV     [200/0]      19h:31m:01s
10.2.21.0/24        -                vlan21        -             C        [0/0]        -
10.2.21.5/32        -                vlan21        -             L        [0/0]        -
10.2.22.0/24        192.168.2.8      -             -             B/EV     [200/0]      01h:04m:48s
10.2.22.1/32        192.168.2.8      -             -             B/EV     [200/0]      01h:04m:48s
10.2.22.2/32        192.168.2.8      -             -             B/EV     [200/0]      01h:04m:48s
10.2.22.23/32       192.168.2.8      -             -             B/EV     [200/0]      00h:03m:36s
192.168.12.3/32     192.168.2.8      -             -             B/EV     [200/0]      11h:29m:30s
192.168.22.8/32     192.168.2.8      -             -             B/EV     [200/0]      19h:33m:01s

Total Route Count : 9
```

```
6300M-7-VSF# sh ip rou vrf VRF2

Displaying ipv4 routes selected for forwarding

Origin Codes: C - connected, S - static, L - local
              R - RIP, B - BGP, O - OSPF
Type Codes:   E - External BGP, I - Internal BGP, V - VPN, EV - EVPN
              IA - OSPF internal area, E1 - OSPF external type 1
              E2 - OSPF external type 2

VRF: VRF2

Prefix              Nexthop          Interface     VRF(egress)   Origin/  Distance/    Age
                                                                 Type     Metric
-------------------------------------------------------------------------------------------
10.2.20.0/24        192.168.40.1     -             -             B/EV     [200/0]      01h:05m:10s
10.2.21.0/24        192.168.40.1     -             -             B/EV     [200/0]      01h:05m:10s
10.2.21.1/32        192.168.40.1     -             -             B/EV     [200/0]      01h:05m:10s
10.2.21.5/32        192.168.40.1     -             -             B/EV     [200/0]      01h:05m:10s
10.2.22.0/24        -                vlan22        -             C        [0/0]        -
10.2.22.2/32        -                vlan22        -             L        [0/0]        -
192.168.12.3/32     192.168.40.1     -             -             B/EV     [200/0]      01h:05m:10s
192.168.22.8/32     192.168.40.1     -             -             B/EV     [200/0]      01h:05m:10s

Total Route Count : 8
```



NEW

Route-map outcome

- Inbound route-map VNI filtering

VNI based Route filtering

Inbound L2VNI filter: In Fabric 3, L2VNI routes for VRF1 & VLAN12 are not needed

```
8360-2# sh bgp l2 e route-type 2 | inc 10.1.12.
*>e [2]:[0]:[0]:[00:50:56:80:53:54]:[10.1.12.14]          192.168.10.3    0          100        0       65001 ?
* e [2]:[0]:[0]:[00:50:56:80:53:54]:[10.1.12.14]          192.168.10.3    0          100        0       65001 ?
*>e [2]:[0]:[0]:[12:00:00:00:01:00]:[10.1.12.1]          192.168.10.3    0          100        0       65001 ?
* e [2]:[0]:[0]:[12:00:00:00:01:00]:[10.1.12.1]          192.168.10.3    0          100        0       65001 ?
*>e [2]:[0]:[0]:[54:80:28:fd:f3:00]:[10.1.12.1]          192.168.10.3    0          100        0       65001 ?
* e [2]:[0]:[0]:[54:80:28:fd:f3:00]:[10.1.12.1]          192.168.10.3    0          100        0       65001 ?
*>e [2]:[0]:[0]:[b8:d4:e7:da:28:00]:[10.1.12.1]          192.168.10.3    0          100        0       65001 ?
* e [2]:[0]:[0]:[b8:d4:e7:da:28:00]:[10.1.12.1]          192.168.10.3    0          100        0       65001 ?
*>e [2]:[0]:[0]:[00:50:56:8e:fc:39]:[10.1.12.16]         192.168.2.8     0          100        0       65001 65002 ?
* e [2]:[0]:[0]:[00:50:56:8e:fc:39]:[10.1.12.16]         192.168.2.8     0          100        0       65001 65002 ?
*>e [2]:[0]:[0]:[12:00:00:00:01:00]:[10.1.12.1]          192.168.2.8     0          100        0       65001 65002 ?
* e [2]:[0]:[0]:[12:00:00:00:01:00]:[10.1.12.1]          192.168.2.8     0          100        0       65001 65002 ?
*>e [2]:[0]:[0]:[88:3a:30:ae:73:c0]:[10.1.12.1]          192.168.2.8     0          100        0       65001 65002 ?
* e [2]:[0]:[0]:[88:3a:30:ae:73:c0]:[10.1.12.1]          192.168.2.8     0          100        0       65001 65002 ?
*>e [2]:[0]:[0]:[00:50:56:80:53:54]:[10.1.12.14]          192.168.10.3    0          100        0       65001 ?
* e [2]:[0]:[0]:[00:50:56:80:53:54]:[10.1.12.14]          192.168.10.3    0          100        0       65001 ?
*>e [2]:[0]:[0]:[12:00:00:00:01:00]:[10.1.12.1]          192.168.10.3    0          100        0       65001 ?
* e [2]:[0]:[0]:[12:00:00:00:01:00]:[10.1.12.1]          192.168.10.3    0          100        0       65001 ?
*>e [2]:[0]:[0]:[54:80:28:fd:f3:00]:[10.1.12.1]          192.168.10.3    0          100        0       65001 ?
* e [2]:[0]:[0]:[54:80:28:fd:f3:00]:[10.1.12.1]          192.168.10.3    0          100        0       65001 ?
*>e [2]:[0]:[0]:[b8:d4:e7:da:28:00]:[10.1.12.1]          192.168.10.3    0          100        0       65001 ?
* e [2]:[0]:[0]:[b8:d4:e7:da:28:00]:[10.1.12.1]          192.168.10.3    0          100        0       65001 ?
*>e [2]:[0]:[0]:[00:50:56:8e:fc:39]:[10.1.12.16]         192.168.2.8     0          100        0       65001 65002 ?
* e [2]:[0]:[0]:[00:50:56:8e:fc:39]:[10.1.12.16]         192.168.2.8     0          100        0       65001 65002 ?
*>e [2]:[0]:[0]:[12:00:00:00:01:00]:[10.1.12.1]          192.168.2.8     0          100        0       65001 65002 ?
* e [2]:[0]:[0]:[12:00:00:00:01:00]:[10.1.12.1]          192.168.2.8     0          100        0       65001 65002 ?
*>e [2]:[0]:[0]:[88:3a:30:ae:73:c0]:[10.1.12.1]          192.168.2.8     0          100        0       65001 65002 ?
* e [2]:[0]:[0]:[88:3a:30:ae:73:c0]:[10.1.12.1]          192.168.2.8     0          100        0       65001 65002 ?



8360-2# sh bgp l2 e  route-type 5 | inc 10.1.12.
*>e [5]:[0]:[0]:[24]:[10.1.12.0]                         192.168.10.3    0          100        0       65001 ?
* e [5]:[0]:[0]:[24]:[10.1.12.0]                         192.168.10.3    0          100        0       65001 ?
*>e [5]:[0]:[0]:[24]:[10.1.12.0]                         192.168.10.3    0          100        0       65001 ?
* e [5]:[0]:[0]:[24]:[10.1.12.0]                         192.168.10.3    0          100        0       65001 ?
*>e [5]:[0]:[0]:[24]:[10.1.12.0]                         192.168.2.8     0          200        0       65001 65002 ?
* e [5]:[0]:[0]:[24]:[10.1.12.0]                         192.168.2.8     0          200        0       65001 65002 ?
```

Type-5 routes are enough for 10.1.12.0/24 subnet in Fabric3

```
8360-2# sh bgp l2 evpn route-type 2 | inc 10.1.12. | count
28
8360-2# sh bgp l2 evpn route-type 5 | inc 10.1.12. | count
6
```

```
8360-2(config)#
route-map BL-BL deny seq 10
     match vni 10012
route-map BL-BL permit seq 100
!
router bgp 65003
…
 address-family l2vpn evpn
        neighbor border-leaders route-map BL-BL in
```

```
8360-2# sh bgp l2 evpn route-type 2 | inc 10.1.12. | count
0
8360-2# sh bgp l2 evpn route-type 5 | inc 10.1.12. | count
6
```

- Notes:
  - Host routes are EVPN route-type 2
  - IP routing table is populated with RT-5 and RT-2

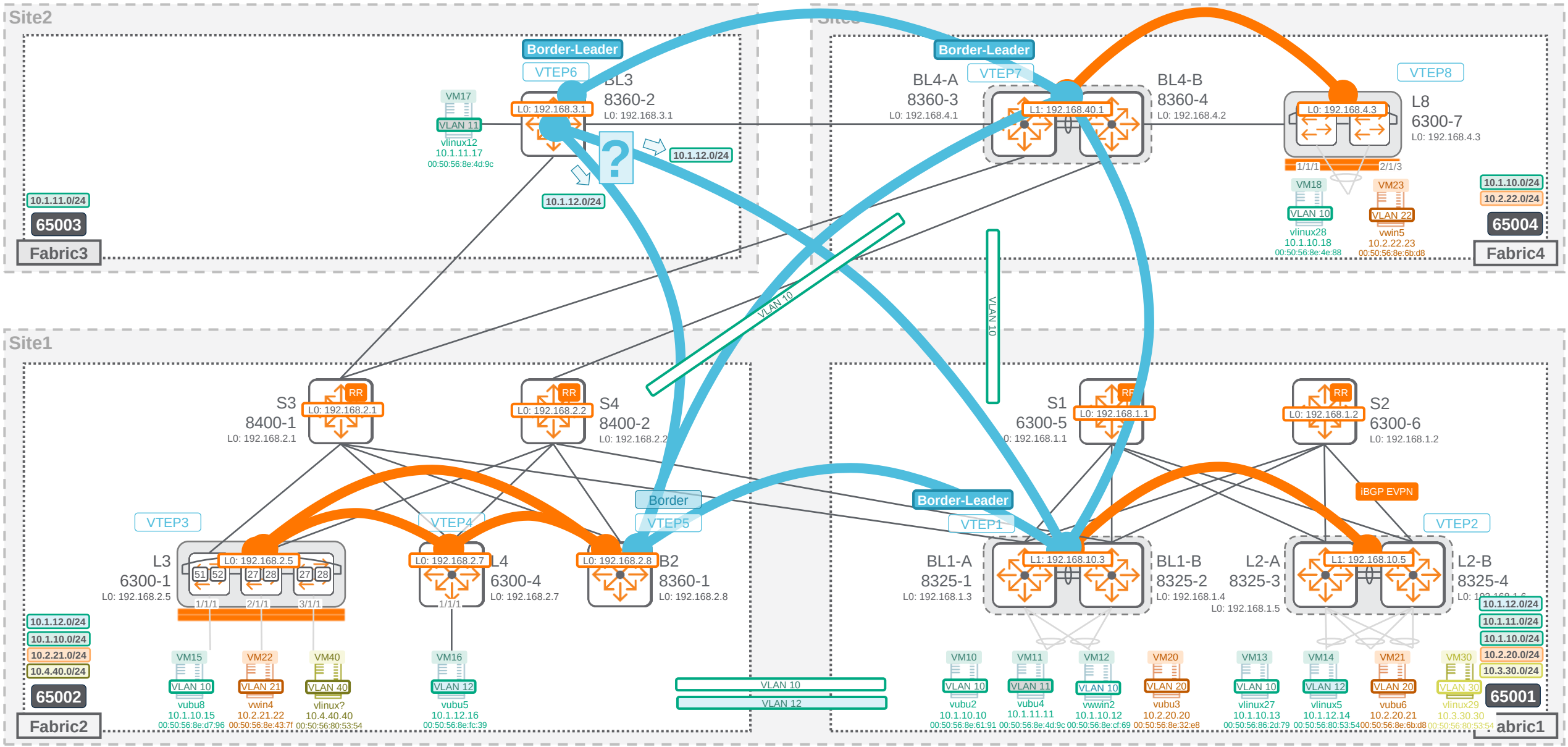
NEW

Route-map outcome

- Influencing path slection with inbound route-map and local-preference

On VTEP6: Influencing outgoing traffic to 10.1.10.0/24

With Local-Preference



Best route selection

Multiple sources/candidates

```
8360-2# sh bgp l2 evpn | inc 10.1.12.0
*>e [5]:[0]:[0]:[24]:[10.1.12.0]                    192.168.10.3    0          100        0       65001 ?
* e [5]:[0]:[0]:[24]:[10.1.12.0]                    192.168.10.3    0          100        0       65001 ?
* e [5]:[0]:[0]:[24]:[10.1.12.0]                    192.168.40.1    0          100        0       65004 65001 ?
* e [5]:[0]:[0]:[24]:[10.1.12.0]                    192.168.40.1    0          100        0       65004 65001 ?
*>e [5]:[0]:[0]:[24]:[10.1.12.0]                    192.168.10.3    0          100        0       65001 ?
* e [5]:[0]:[0]:[24]:[10.1.12.0]                    192.168.10.3    0          100        0       65001 ?
* e [5]:[0]:[0]:[24]:[10.1.12.0]                    192.168.40.1    0          100        0       65004 65001 ?
* e [5]:[0]:[0]:[24]:[10.1.12.0]                    192.168.40.1    0          100        0       65004 65001 ?
*>e [5]:[0]:[0]:[24]:[10.1.12.0]                    192.168.2.8     0          100        0       65001 65002 ?
* e [5]:[0]:[0]:[24]:[10.1.12.0]                    192.168.2.8     0          100        0       65001 65002 ?
* e [5]:[0]:[0]:[24]:[10.1.12.0]                    192.168.40.1    0          100        0       65004 65001 65002 ?
* e [5]:[0]:[0]:[24]:[10.1.12.0]                    192.168.40.1    0          100        0       65004 65001 65002 ?


8360-2# sh ip rou 10.1.12.0/24 vrf VRF1

VRF: VRF1

   Prefix          : 10.1.12.0/24                       VRF(egress)       : -
   Nexthop         : 192.168.10.3                       Interface         : -
   Origin          : bgp                                Type              : bgp_evpn
   Distance        : 200                                Metric            : 0
   Age             : 00h:00m:19s                        Tag               : 0
   Encap Type      : vxlan                              Encap Details     : l3vni 100001
```

Multiple candidate-routes are compared from different NLRIs:

- same local-preference
- Fabric 1 (192.168.10.3) is preferred as being shortest AS-path

```
8360-2# sh bgp l2 e 192.168.2.7:1-[5]:[0]:[0]:[24]:[10.1.12.0]

VRF : default
BGP Local AS 65003       BGP Router-id 192.168.3.1

   Network           : 192.168.2.7:1-[5]:[0]:[0]:[24]:[10.1.12.0]
   Nexthop           : 192.168.2.8
   vni               : 100001                    vni_type          : L3VNI
   Peer              : 192.168.1.3               Origin            : incomplete
   Metric            : 0                         Local Pref        : 100
   Weight            : 0                         Calc. Local Pref  : 100
   Best              : Yes                       Valid             : Yes
   Type              : external                  Stale             : No
   Originator ID     : 0.0.0.0
   Aggregator ID     :
   Aggregator AS     :
   Atomic Aggregate  :

   AS-Path           :  65001 65002

   Cluster List      :
   Communities       :
   Ext-Communities   : RT: 1:1 RT: 65002:1 Router MAC: 02:00:00:00:05:00

   Network           : 192.168.2.7:1-[5]:[0]:[0]:[24]:[10.1.12.0]
   Nexthop           : 192.168.2.8
   vni               : 100001                    vni_type          : L3VNI
   Peer              : 192.168.1.4               Origin            : incomplete
   Metric            : 0                         Local Pref        : 100
   Weight            : 0                         Calc. Local Pref  : 100
   Best              : No                        Valid             : Yes
   Type              : external                  Stale             : No
   Originator ID     : 0.0.0.0
   Aggregator ID     :
   Aggregator AS     :
   Atomic Aggregate  :

   AS-Path           :  65001 65002

   Cluster List      :
   Communities       :
   Ext-Communities   : RT: 1:1 RT: 65002:1 Router MAC: 02:00:00:00:05:00

   Network           : 192.168.2.7:1-[5]:[0]:[0]:[24]:[10.1.12.0]
   Nexthop           : 192.168.40.1
   vni               : 100001                    vni_type          : L3VNI
   Peer              : 192.168.4.1               Origin            : incomplete
   Metric            : 0                         Local Pref        : 100
   Weight            : 0                         Calc. Local Pref  : 100
   Best              : No                        Valid             : Yes
   Type              : external                  Stale             : No
   Originator ID     : 0.0.0.0
   Aggregator ID     :
   Aggregator AS     :
   Atomic Aggregate  :

   AS-Path           :  65004 65001 65002

   Cluster List      :
   Communities       :
   Ext-Communities   : RT: 1:1 RT: 65002:1 Router MAC: 02:00:00:00:07:00

   Network           : 192.168.2.7:1-[5]:[0]:[0]:[24]:[10.1.12.0]
   Nexthop           : 192.168.40.1
   vni               : 100001                    vni_type          : L3VNI
   Peer              : 192.168.4.2               Origin            : incomplete
   Metric            : 0                         Local Pref        : 100
   Weight            : 0                         Calc. Local Pref  : 100
   Best              : No                        Valid             : Yes
   Type              : external                  Stale             : No
   Originator ID     : 0.0.0.0
   Aggregator ID     :
   Aggregator AS     :
   Atomic Aggregate  :

   AS-Path           :  65004 65001 65002

   Cluster List      :
   Communities       :
   Ext-Communities   : RT: 1:1 RT: 65002:1 Router MAC: 02:00:00:00:07:00
```

```
8360-2# sh bgp l2 e 192.168.1.5:1-[5]:[0]:[0]:[24]:[10.1.12.0]

VRF : default
BGP Local AS 65003       BGP Router-id 192.168.3.1

   Network           : 192.168.1.5:1-[5]:[0]:[0]:[24]:[10.1.12.0]
   Nexthop           : 192.168.10.3
   vni               : 100001                    vni_type          : L3VNI
   Peer              : 192.168.1.3               Origin            : incomplete
   Metric            : 0                         Local Pref        : 100
   Weight            : 0                         Calc. Local Pref  : 100
   Best              : No                        Valid             : Yes
   Type              : external                  Stale             : No
   Originator ID     : 0.0.0.0
   Aggregator ID     :
   Aggregator AS     :
   Atomic Aggregate  :

   AS-Path           :  65001

   Cluster List      :
   Communities       :
   Ext-Communities   : RT: 1:1 RT: 65001:1 Router MAC: 02:00:00:00:01:00

   Network           : 192.168.1.5:1-[5]:[0]:[0]:[24]:[10.1.12.0]
   Nexthop           : 192.168.10.3
   vni               : 100001                    vni_type          : L3VNI
   Peer              : 192.168.1.4               Origin            : incomplete
   Metric            : 0                         Local Pref        : 100
   Weight            : 0                         Calc. Local Pref  : 100
   Best              : Yes                       Valid             : Yes
   Type              : external                  Stale             : No
   Originator ID     : 0.0.0.0
   Aggregator ID     :
   Aggregator AS     :
   Atomic Aggregate  :

   AS-Path           :  65001

   Cluster List      :
   Communities       :
   Ext-Communities   : RT: 1:1 RT: 65001:1 Router MAC: 02:00:00:00:01:00

   Network           : 192.168.1.5:1-[5]:[0]:[0]:[24]:[10.1.12.0]
   Nexthop           : 192.168.40.1
   vni               : 100001                    vni_type          : L3VNI
   Peer              : 192.168.4.1               Origin            : incomplete
   Metric            : 0                         Local Pref        : 100
   Weight            : 0                         Calc. Local Pref  : 100
   Best              : No                        Valid             : Yes
   Type              : external                  Stale             : No
   Originator ID     : 0.0.0.0
   Aggregator ID     :
   Aggregator AS     :
   Atomic Aggregate  :

   AS-Path           :  65004 65001

   Cluster List      :
   Communities       :
   Ext-Communities   : RT: 1:1 RT: 65001:1 Router MAC: 02:00:00:00:07:00

   Network           : 192.168.1.5:1-[5]:[0]:[0]:[24]:[10.1.12.0]
   Nexthop           : 192.168.40.1
   vni               : 100001                    vni_type          : L3VNI
   Peer              : 192.168.4.2               Origin            : incomplete
   Metric            : 0                         Local Pref        : 100
   Weight            : 0                         Calc. Local Pref  : 100
   Best              : No                        Valid             : Yes
   Type              : external                  Stale             : No
   Originator ID     : 0.0.0.0
   Aggregator ID     :
   Aggregator AS     :
   Atomic Aggregate  :

   AS-Path           :  65004 65001

   Cluster List      :
   Communities       :
   Ext-Communities   : RT: 1:1 RT: 65001:1 Router MAC: 02:00:00:00:07:00
```

Local-Preference Usage

Fabric2 is now preferred (highest LP)

```
ip aspath-list fabric2 seq 10 permit _65002$
!
route-map BL-BL deny seq 10
     match vni 10012
route-map BL-BL permit seq 20
     match ip address prefix-list subnet12
     match aspath-list fabric2
     set local-preference 200
route-map BL-BL permit seq 100
!
router bgp 65003
...
 address-family l2vpn evpn
        neighbor border-leaders route-map BL-BL in
```

```
8360-2# sh bgp l2 evpn | inc 10.1.12.0
*>e [5]:[0]:[0]:[24]:[10.1.12.0]                    192.168.10.3    0          100        0       65001 ?
* e [5]:[0]:[0]:[24]:[10.1.12.0]                    192.168.10.3    0          100        0       65001 ?
* e [5]:[0]:[0]:[24]:[10.1.12.0]                    192.168.40.1    0          100        0       65004 65001 ?
* e [5]:[0]:[0]:[24]:[10.1.12.0]                    192.168.40.1    0          100        0       65004 65001 ?
*>e [5]:[0]:[0]:[24]:[10.1.12.0]                    192.168.10.3    0          100        0       65001 ?
* e [5]:[0]:[0]:[24]:[10.1.12.0]                    192.168.10.3    0          100        0       65001 ?
* e [5]:[0]:[0]:[24]:[10.1.12.0]                    192.168.40.1    0          100        0       65004 65001 ?
* e [5]:[0]:[0]:[24]:[10.1.12.0]                    192.168.40.1    0          100        0       65004 65001 ?
*>e [5]:[0]:[0]:[24]:[10.1.12.0]                    192.168.2.8     0          200        0       65001 65002 ?
* e [5]:[0]:[0]:[24]:[10.1.12.0]                    192.168.2.8     0          200        0       65001 65002 ?
* e [5]:[0]:[0]:[24]:[10.1.12.0]                    192.168.40.1    0          200        0       65004 65001 65002 ?
* e [5]:[0]:[0]:[24]:[10.1.12.0]                    192.168.40.1    0          200        0       65004 65001 65002 ?


8360-2# sh ip rou 10.1.12.0/24 vrf VRF1

VRF: VRF1

   Prefix          : 10.1.12.0/24                    VRF(egress)     : -
   Nexthop         : 192.168.2.8                     Interface       : -
   Origin          : bgp                             Type            : bgp_evpn
   Distance        : 200                             Metric          : 0
   Age             : 00h:00m:11s                     Tag             : 0
   Encap Type      : vxlan                           Encap Details   : l3vni 100001
```

```
8360-2# sh bgp l2 e 192.168.2.7:1-[5]:[0]:[0]:[24]:[10.1.12.0]

VRF : default
BGP Local AS 65003       BGP Router-id 192.168.3.1

   Network           : 192.168.2.7:1-[5]:[0]:[0]:[24]:[10.1.12.0]
   Nexthop           : 192.168.2.8
   vni               : 100001                    vni_type          : L3VNI
   Peer              : 192.168.1.3               Origin            : incomplete
   Metric            : 0                         Local Pref        : 200
   Weight            : 0                         Calc. Local Pref  : 200
   Best              : Yes                       Valid             : Yes
   Type              : external                  Stale             : No
   Originator ID     : 0.0.0.0
   Aggregator ID     :
   Aggregator AS     :
   Atomic Aggregate  :

   AS-Path           :  65001 65002

   Cluster List      :
   Communities       :
   Ext-Communities   : RT: 1:1 RT: 65002:1 Router MAC: 02:00:00:00:05:00

   Network           : 192.168.2.7:1-[5]:[0]:[0]:[24]:[10.1.12.0]
   Nexthop           : 192.168.2.8
   vni               : 100001                    vni_type          : L3VNI
   Peer              : 192.168.1.4               Origin            : incomplete
   Metric            : 0                         Local Pref        : 200
   Weight            : 0                         Calc. Local Pref  : 200
   Best              : No                        Valid             : Yes
   Type              : external                  Stale             : No
   Originator ID     : 0.0.0.0
   Aggregator ID     :
   Aggregator AS     :
   Atomic Aggregate  :

   AS-Path           :  65001 65002

   Cluster List      :
   Communities       :
   Ext-Communities   : RT: 1:1 RT: 65002:1 Router MAC: 02:00:00:00:05:00

   Network           : 192.168.2.7:1-[5]:[0]:[0]:[24]:[10.1.12.0]
   Nexthop           : 192.168.40.1
   vni               : 100001                    vni_type          : L3VNI
   Peer              : 192.168.4.1               Origin            : incomplete
   Metric            : 0                         Local Pref        : 200
   Weight            : 0                         Calc. Local Pref  : 200
   Best              : No                        Valid             : Yes
   Type              : external                  Stale             : No
   Originator ID     : 0.0.0.0
   Aggregator ID     :
   Aggregator AS     :
   Atomic Aggregate  :

   AS-Path           :  65004 65001 65002

   Cluster List      :
   Communities       :
   Ext-Communities   : RT: 1:1 RT: 65002:1 Router MAC: 02:00:00:00:07:00

   Network           : 192.168.2.7:1-[5]:[0]:[0]:[24]:[10.1.12.0]
   Nexthop           : 192.168.40.1
   vni               : 100001                    vni_type          : L3VNI
   Peer              : 192.168.4.2               Origin            : incomplete
   Metric            : 0                         Local Pref        : 200
   Weight            : 0                         Calc. Local Pref  : 200
   Best              : No                        Valid             : Yes
   Type              : external                  Stale             : No
   Originator ID     : 0.0.0.0
   Aggregator ID     :
   Aggregator AS     :
   Atomic Aggregate  :

   AS-Path           :  65004 65001 65002

   Cluster List      :
   Communities       :
   Ext-Communities   : RT: 1:1 RT: 65002:1 Router MAC: 02:00:00:00:07:00
```

```
8360-2# sh bgp l2 e 192.168.1.5:1-[5]:[0]:[0]:[24]:[10.1.12.0]

VRF : default
BGP Local AS 65003       BGP Router-id 192.168.3.1

   Network           : 192.168.1.5:1-[5]:[0]:[0]:[24]:[10.1.12.0]
   Nexthop           : 192.168.10.3
   vni               : 100001                    vni_type          : L3VNI
   Peer              : 192.168.1.3               Origin            : incomplete
   Metric            : 0                         Local Pref        : 100
   Weight            : 0                         Calc. Local Pref  : 100
   Best              : Yes                       Valid             : Yes
   Type              : external                  Stale             : No
   Originator ID     : 0.0.0.0
   Aggregator ID     :
   Aggregator AS     :
   Atomic Aggregate  :

   AS-Path           :  65001

   Cluster List      :
   Communities       :
   Ext-Communities   : RT: 1:1 RT: 65001:1 Router MAC: 02:00:00:00:01:00

   Network           : 192.168.1.5:1-[5]:[0]:[0]:[24]:[10.1.12.0]
   Nexthop           : 192.168.10.3
   vni               : 100001                    vni_type          : L3VNI
   Peer              : 192.168.1.4               Origin            : incomplete
   Metric            : 0                         Local Pref        : 100
   Weight            : 0                         Calc. Local Pref  : 100
   Best              : No                        Valid             : Yes
   Type              : external                  Stale             : No
   Originator ID     : 0.0.0.0
   Aggregator ID     :
   Aggregator AS     :
   Atomic Aggregate  :

   AS-Path           :  65001

   Cluster List      :
   Communities       :
   Ext-Communities   : RT: 1:1 RT: 65001:1 Router MAC: 02:00:00:00:01:00

   Network           : 192.168.1.5:1-[5]:[0]:[0]:[24]:[10.1.12.0]
   Nexthop           : 192.168.40.1
   vni               : 100001                    vni_type          : L3VNI
   Peer              : 192.168.4.1               Origin            : incomplete
   Metric            : 0                         Local Pref        : 100
   Weight            : 0                         Calc. Local Pref  : 100
   Best              : No                        Valid             : Yes
   Type              : external                  Stale             : No
   Originator ID     : 0.0.0.0
   Aggregator ID     :
   Aggregator AS     :
   Atomic Aggregate  :

   AS-Path           :  65004 65001

   Cluster List      :
   Communities       :
   Ext-Communities   : RT: 1:1 RT: 65001:1 Router MAC: 02:00:00:00:07:00

   Network           : 192.168.1.5:1-[5]:[0]:[0]:[24]:[10.1.12.0]
   Nexthop           : 192.168.40.1
   vni               : 100001                    vni_type          : L3VNI
   Peer              : 192.168.4.2               Origin            : incomplete
   Metric            : 0                         Local Pref        : 100
   Weight            : 0                         Calc. Local Pref  : 100
   Best              : No                        Valid             : Yes
   Type              : external                  Stale             : No
   Originator ID     : 0.0.0.0
   Aggregator ID     :
   Aggregator AS     :
   Atomic Aggregate  :

   AS-Path           :  65004 65001

   Cluster List      :
   Communities       :
   Ext-Communities   : RT: 1:1 RT: 65001:1 Router MAC: 02:00:00:00:07:00
```

Switch Configuration

For reference

8325-1

```
vrf VRF1
    rd 192.168.1.3:1
    route-target export 1:1 evpn
    route-target export 65001:1 evpn
    route-target import 1:1 evpn
    route-target import 65001:1 evpn
vrf VRF2
    rd 192.168.1.3:2
    route-target export 1:2 evpn
    route-target export 65001:2 evpn
    route-target import 1:2 evpn
    route-target import 65001:2 evpn
!
vlan 10
    vsx-sync
vlan 11
    vsx-sync
vlan 20
    vsx-sync
vlan 1019
    vsx-sync
vlan 1110
    vsx-sync
!
virtual-mac 02:00:00:00:01:00
!
evpn
    arp-suppression
    nd-suppression
    redistribute local-svi
    dyn-vxlan-tunnel-bridging-mode ibgp-ebgp
    vlan 10
        rd auto
        route-target export 1:10
        route-target export 65001:10
        route-target import 1:10
        route-target import 65001:10
        redistribute host-route
    vlan 11
        rd auto
        route-target export 1:11
        route-target export 65001:11
        route-target import 1:11
        route-target import 65001:11
        redistribute host-route
    vlan 12
        rd auto
        route-target export 1:12
        route-target export 65001:12
        route-target import 1:12
        route-target import 65001:12
    vlan 20
        rd auto
        route-target export 65001:20
        route-target import 65001:20
        redistribute host-route
```

```
interface 1/1/23
    no shutdown
    mtu 9198
    ip mtu 9198
    ip address 192.168.19.1/31
    ip ospf 1 area 0.0.0.0
    no ip ospf passive
    ip ospf network point-to-point
    ip ospf authentication message-digest
    ip ospf message-digest-key 1 md5 ciphertext
AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
!
interface 1/1/24
    no shutdown
    mtu 9198
    ip mtu 9198
    ip address 192.168.19.9/31
    ip ospf 1 area 0.0.0.0
    no ip ospf passive
    ip ospf network point-to-point
    ip ospf authentication message-digest
    ip ospf message-digest-key 1 md5 ciphertext
AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
!
interface loopback 0
    ip address 192.168.1.3/32
interface loopback 1
    ip address 192.168.10.3/32
!
interface loopback 11
    vrf attach VRF1
    ip address 192.168.11.103/32
interface loopback 12
    vrf attach VRF1
    ip address 192.168.11.3/32
!
interface loopback 22
    vrf attach VRF2
    ip address 192.168.12.3/32
!
interface vlan 10
    vsx-sync active-gateways
    vrf attach VRF1
    ip mtu 9000
    ip address 10.1.10.1/24
    active-gateway ip mac 12:00:00:00:01:00
    active-gateway ip 10.1.10.1
    ipv6 address link-local fe80:10:1:10::1/64
    ipv6 address fd00:10:1:10::1/64
    active-gateway ipv6 mac 12:00:00:00:01:00
    active-gateway ipv6 fe80:10:1:10::1
    no ipv6 nd suppress-ra
    ipv6 nd router-preference high
    ip helper-address 10.10.129.30 vrf default
```

```
interface vlan 11
    vsx-sync active-gateways
    vrf attach VRF1
    ip mtu 9000
    ip address 10.1.11.1/24
    active-gateway ip mac 12:00:00:00:01:00
    active-gateway ip 10.1.11.1
    ipv6 address link-local fe80:10:1:11::1/64
    ipv6 address fd00:10:1:11::1/64
    active-gateway ipv6 mac 12:00:00:00:01:00
    active-gateway ipv6 fe80:10:1:11::1
    no ipv6 nd suppress-ra
    ipv6 nd router-preference high
!
interface vlan 20
    vsx-sync active-gateways
    vrf attach VRF2
    ip mtu 9000
    ip address 10.2.20.2/24
    active-gateway ip mac 12:00:00:00:01:00
    active-gateway ip 10.2.20.1
!
interface vlan 1019
    ip mtu 9198
    ip address 192.168.19.200/31
    ip ospf 1 area 0.0.0.0
    no ip ospf passive
    ip ospf cost 50
    ip ospf network point-to-point
    ip ospf authentication message-digest
    ip ospf message-digest-key 1 md5 ciphertext
AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
!
interface vlan 1110
    vrf attach VRF1
    description VRF1 transit for loopbacks in VSX
    ip mtu 9198
    ip address 192.168.110.0/31
!
interface vxlan 1
    source ip 192.168.10.3
    vxlan-counters aggregate
    no shutdown
    vni 10010
        vlan 10
    vni 10011
        vlan 11
    vni 10020
        vlan 20
    vni 100001
        vrf VRF1
        routing
    vni 100002
        vrf VRF2
        routing
```

8325-1

```
ip prefix-list endpoint-VRF1 seq 10 permit 10.1.0.0/16 le 32
ip prefix-list endpoint-VRF2 seq 10 permit 10.2.0.0/16 le 32
ipv6 prefix-list v6-endpoint-VRF1 seq 10 permit fd00:10:1:10::/64 le 128
ip prefix-list vrf2-vlan20-summary seq 10 permit 10.2.20.0/24 ge 25 le 32
!
ip aspath-list fabric2 seq 10 permit _65002$
ip aspath-list fabric3 seq 10 permit _65003$
ip aspath-list fabric4 seq 10 permit _65004$
ip aspath-list local-fabric seq 10 permit ^$
!
route-map BGP-OSPF deny seq 10
     match tag 1000
route-map BGP-OSPF permit seq 20
!
route-map OSPF-BGP permit seq 10
     match tag 1000
!
route-map connected-bgp-VRF1 permit seq 10
     match ip address prefix-list endpoint-VRF1
route-map connected-bgp-VRF1 permit seq 20
     match ipv6 address prefix-list v6-endpoint-VRF1
!
route-map connected-bgp-VRF2 permit seq 10
     match ip address prefix-list endpoint-VRF2
!
route-map connected-ospf permit seq 10
     set tag 1000
!
route-map to-border-leaders deny seq 5
     match ip address prefix-list vrf2-vlan20-summary
route-map to-border-leaders permit seq 10
     match aspath-list local-fabric
route-map to-border-leaders permit seq 20
     match aspath-list fabric2
     set ip next-hop 192.168.2.8
!
route-map to-borders deny seq 5
     match ip address prefix-list vrf2-vlan20-summary
route-map to-borders permit seq 10
     match aspath-list fabric2
     set ip next-hop 192.168.2.8
route-map to-borders permit seq 20
     match aspath-list fabric3
     set ip next-hop 192.168.3.1
route-map to-borders permit seq 30
     match aspath-list fabric4
     set ip next-hop 192.168.40.1
route-map to-borders permit seq 1000
```

```
vsx
    system-mac 02:00:00:00:01:00
    inter-switch-link lag 256
    role primary
    keepalive peer 192.168.0.1 source 192.168.0.0 vrf KA
    vsx-sync aaa copp-policy dhcp-relay dns evpn l2-vlan-mac-cfg-mode mclag-interfaces qos-global route-
map sflow-global snmp ssh stp-global time vsx-global
!
ip route 192.168.11.4/32 192.168.110.1 vrf VRF1
!
ip source-interface dhcp_relay interface loopback1
!
router ospf 1
    router-id 192.168.1.3
    max-metric router-lsa include-stub on-startup 240
    passive-interface default
    timers throttle spf start-time 100 hold-time 500 max-wait-time 1000
    redistribute bgp route-map BGP-OSPF
    redistribute local loopback route-map connected-ospf
    area 0.0.0.0
```

8325-1

```
router bgp 65001
    bgp router-id 192.168.1.3
    trap-enable
    bgp log-neighbor-changes
    bgp deterministic-med
    bgp always-compare-med
!
    neighbor border-leaders peer-group
    neighbor border-leaders description eBGP EVPN peering with remote Fabrics
    neighbor border-leaders password ciphertext AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
    neighbor border-leaders fall-over
    neighbor border-leaders update-source loopback 0
!
    neighbor borders peer-group
    neighbor borders description eBGP EVPN peering with local-site Fabrics
    neighbor borders password ciphertext AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
    neighbor borders fall-over
    neighbor borders update-source loopback 0
!
    neighbor spine-RR peer-group
    neighbor spine-RR remote-as 65001
    neighbor spine-RR description Spine and RR peer-group
    neighbor spine-RR password ciphertext AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
    neighbor spine-RR fall-over
    neighbor spine-RR update-source loopback 0
!
    neighbor 192.168.1.1 peer-group spine-RR
    neighbor 192.168.1.2 peer-group spine-RR
!
    neighbor 192.168.2.8 remote-as 65002
    neighbor 192.168.2.8 peer-group borders
    neighbor 192.168.2.8 description Fabric2 Border2
    neighbor 192.168.2.8 ebgp-multihop 10
!
    neighbor 192.168.3.1 remote-as 65003
    neighbor 192.168.3.1 peer-group border-leaders
    neighbor 192.168.3.1 description Fabric3 BL
    neighbor 192.168.3.1 ebgp-multihop 10
!
    neighbor 192.168.4.1 remote-as 65004
    neighbor 192.168.4.1 peer-group border-leaders
    neighbor 192.168.4.1 description Fabric4 BL-A
    neighbor 192.168.4.1 ebgp-multihop 10
!
    neighbor 192.168.4.2 remote-as 65004
    neighbor 192.168.4.2 peer-group border-leaders
    neighbor 192.168.4.2 description Fabric4 BL-B
    neighbor 192.168.4.2 ebgp-multihop 10
!
    neighbor 192.168.29.6 remote-as 65002
    neighbor 192.168.29.6 description Fabric2 S3
    neighbor 192.168.29.6 password ciphertext AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
```

```
    address-family ipv4 unicast
        neighbor 192.168.29.6 activate
        redistribute local loopback
        redistribute ospf 1 route-map OSPF-BGP
    exit-address-family
    address-family l2vpn evpn
        neighbor border-leaders route-map to-border-leaders out
        neighbor border-leaders send-community both
!
        neighbor borders route-map to-borders out
        neighbor borders send-community both
!
        neighbor spine-RR next-hop-self
        neighbor spine-RR send-community both
!
        neighbor 192.168.1.1 activate
        neighbor 192.168.1.2 activate
        neighbor 192.168.2.8 activate
        neighbor 192.168.3.1 activate
        neighbor 192.168.4.1 activate
        neighbor 192.168.4.2 activate
    exit-address-family
!
    vrf VRF1
        bgp router-id 192.168.1.3
        bgp log-neighbor-changes
        address-family ipv4 unicast
            redistribute local loopback
            redistribute connected route-map connected-bgp-VRF1
        exit-address-family
        address-family ipv6 unicast
            redistribute local loopback
            redistribute connected route-map connected-bgp-VRF1
        exit-address-family
!
    vrf VRF2
        bgp router-id 192.168.1.3
        bgp log-neighbor-changes
        address-family ipv4 unicast
            redistribute local loopback
            redistribute connected route-map connected-bgp-VRF2
            aggregate-address 10.2.20.0/24 summary-only
        exit-address-family
```

8360-1

```
vrf VRF1
    rd 192.168.2.8:1
    route-target export 1:1 evpn
    route-target export 65002:1 evpn
    route-target import 1:1 evpn
    route-target import 65002:1 evpn
vrf VRF2
    rd 192.168.2.8:2
    route-target export 1:2 evpn
    route-target export 65002:2 evpn
    route-target import 1:2 evpn
    route-target import 65002:2 evpn
vrf VRF3
    rd 192.168.2.8:3
    route-target export 65002:3 evpn
    route-target import 65002:3 evpn
vrf VRF4
    rd 192.168.2.8:4
    route-target export 65002:4 evpn
    route-target import 65002:4 evpn
!
vlan 1,10,12
!
virtual-mac 02:00:00:00:05:00
!
evpn
    arp-suppression
    nd-suppression
    dyn-vxlan-tunnel-bridging-mode ibgp-ebgp
    vlan 10
        rd auto
        route-target export 1:10
        route-target import 1:10
        redistribute host-route
    vlan 12
        rd auto
        route-target export 1:12
        route-target import 1:12
```

```
interface 1/1/10
    no shutdown
    mtu 9198
    description 8400-2 1/10/12
    ip mtu 9198
    ip address 192.168.29.13/31
    ip ospf 1 area 0.0.0.0
    no ip ospf passive
    ip ospf network point-to-point
    ip ospf authentication message-digest
    ip ospf message-digest-key 1 md5 ciphertext
AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
!
interface 1/1/13
    no shutdown
    mtu 9198
    description 8400-1 1/10/13
    ip mtu 9198
    ip address 192.168.29.5/31
    ip ospf 1 area 0.0.0.0
    no ip ospf passive
    ip ospf network point-to-point
    ip ospf authentication message-digest
    ip ospf message-digest-key 1 md5 ciphertext
AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
!
interface loopback 0
    ip address 192.168.2.8/32
!
interface loopback 12
    vrf attach VRF1
    description for troubleshooting in VRF1
    ip address 192.168.21.8/32
!
interface loopback 22
    vrf attach VRF2
    description for troubleshooting in VRF2
    ip address 192.168.22.8/32
```

```
interface vxlan 1
    source ip 192.168.2.8
    vxlan-counters aggregate
    no shutdown
    vni 10010
        vlan 10
    vni 10012
        vlan 12
    vni 100001
        vrf VRF1
        routing
    vni 100002
        vrf VRF2
        routing
!
ip prefix-list endpoint-VRF1 seq 10 permit 10.1.0.0/16 le 32
!
route-map connected-bgp-VRF1 permit seq 10
     match ip address prefix-list endpoint-VRF1
!
route-map connected-ospf permit seq 10
     set tag 1000
!
router ospf 1
    router-id 192.168.2.8
    max-metric router-lsa include-stub on-startup
    passive-interface default
    graceful-restart ignore-lost-interface
    redistribute local loopback route-map connected-ospf
    area 0.0.0.0
```

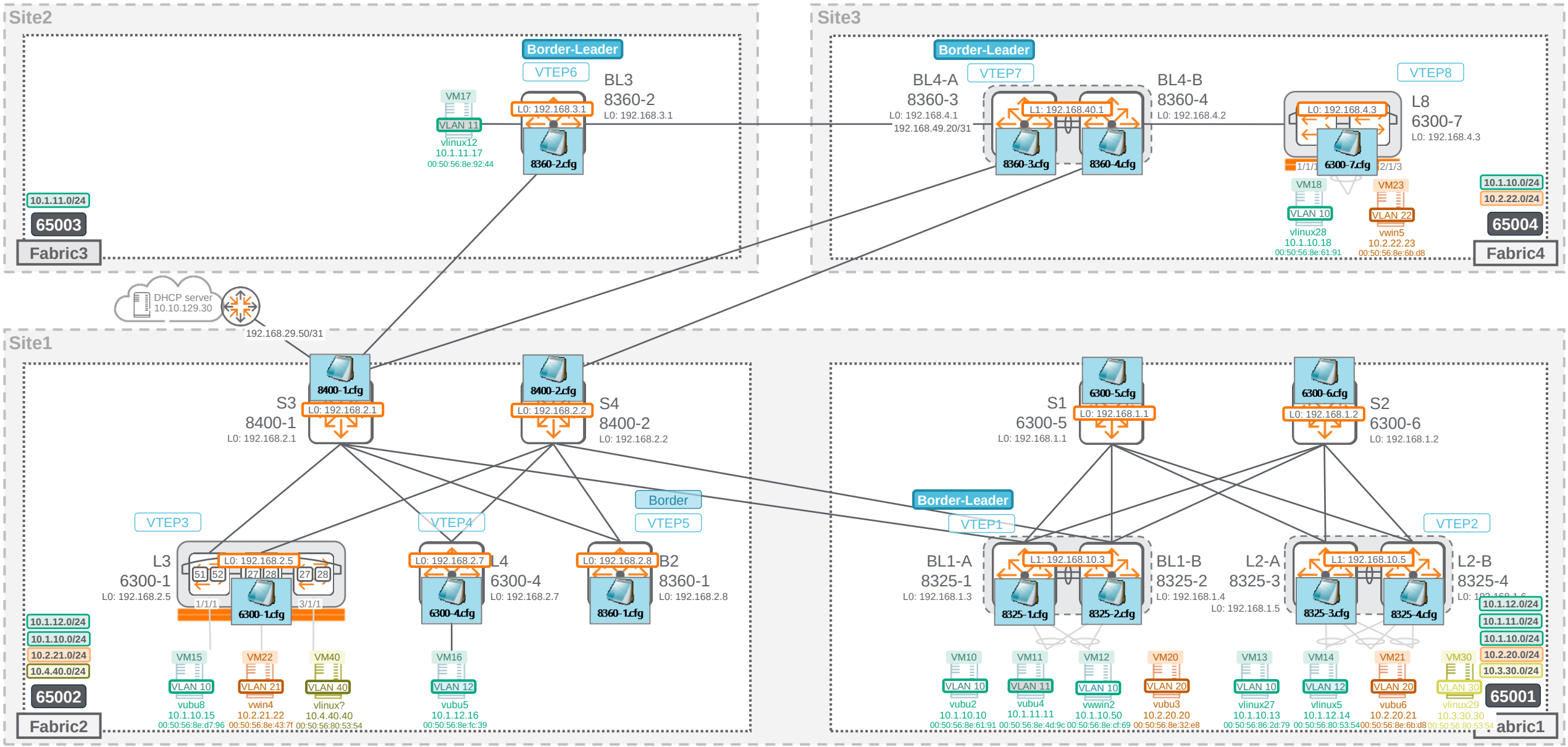
8360-1

```
router bgp 65002
    bgp router-id 192.168.2.8
    trap-enable
    bgp log-neighbor-changes
    bgp deterministic-med
    bgp always-compare-med
    neighbor border-leader peer-group
    neighbor border-leader remote-as 65001
    neighbor border-leader description Fabric1 BorderLeader1
    neighbor border-leader password ciphertext AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
    neighbor border-leader fall-over
    neighbor border-leader ebgp-multihop 10
    neighbor border-leader update-source loopback 0
    neighbor spine-RR peer-group
    neighbor spine-RR remote-as 65002
    neighbor spine-RR description Spine and RR peer-group
    neighbor spine-RR password ciphertext AQBapUbZyuMyDkoDN0zeQbI8qY0p5vpa77xnpPQEngEkpWjWBQAAAIouj7OC
    neighbor spine-RR fall-over
    neighbor spine-RR update-source loopback 0
    neighbor 192.168.1.3 peer-group border-leader
    neighbor 192.168.1.4 peer-group border-leader
    neighbor 192.168.2.1 peer-group spine-RR
    neighbor 192.168.2.2 peer-group spine-RR
    address-family l2vpn evpn
        neighbor border-leader send-community both
        neighbor spine-RR next-hop-self
        neighbor spine-RR send-community both
        neighbor 192.168.1.3 activate
        neighbor 192.168.1.4 activate
        neighbor 192.168.2.1 activate
        neighbor 192.168.2.2 activate
    exit-address-family
!
    vrf VRF1
        bgp router-id 192.168.2.8
        address-family ipv4 unicast
            redistribute local loopback
            redistribute connected route-map connected-bgp-VRF1
        exit-address-family
!
    vrf VRF2
        bgp router-id 192.168.2.8
        address-family ipv4 unicast
            redistribute local loopback
        exit-address-family
!
    vrf VRF3
        bgp router-id 192.168.2.8
        address-family ipv4 unicast
            redistribute local loopback
        exit-address-family
!
    vrf VRF4
        bgp router-id 192.168.2.8
        address-family ipv4 unicast
            redistribute local loopback
        exit-address-family
```

Switch configuration

Double-click on .cfg light-blue box

Some configuration parts do not reflect entirely operational configuration for demo/education purpose addressed during troubleshooting section

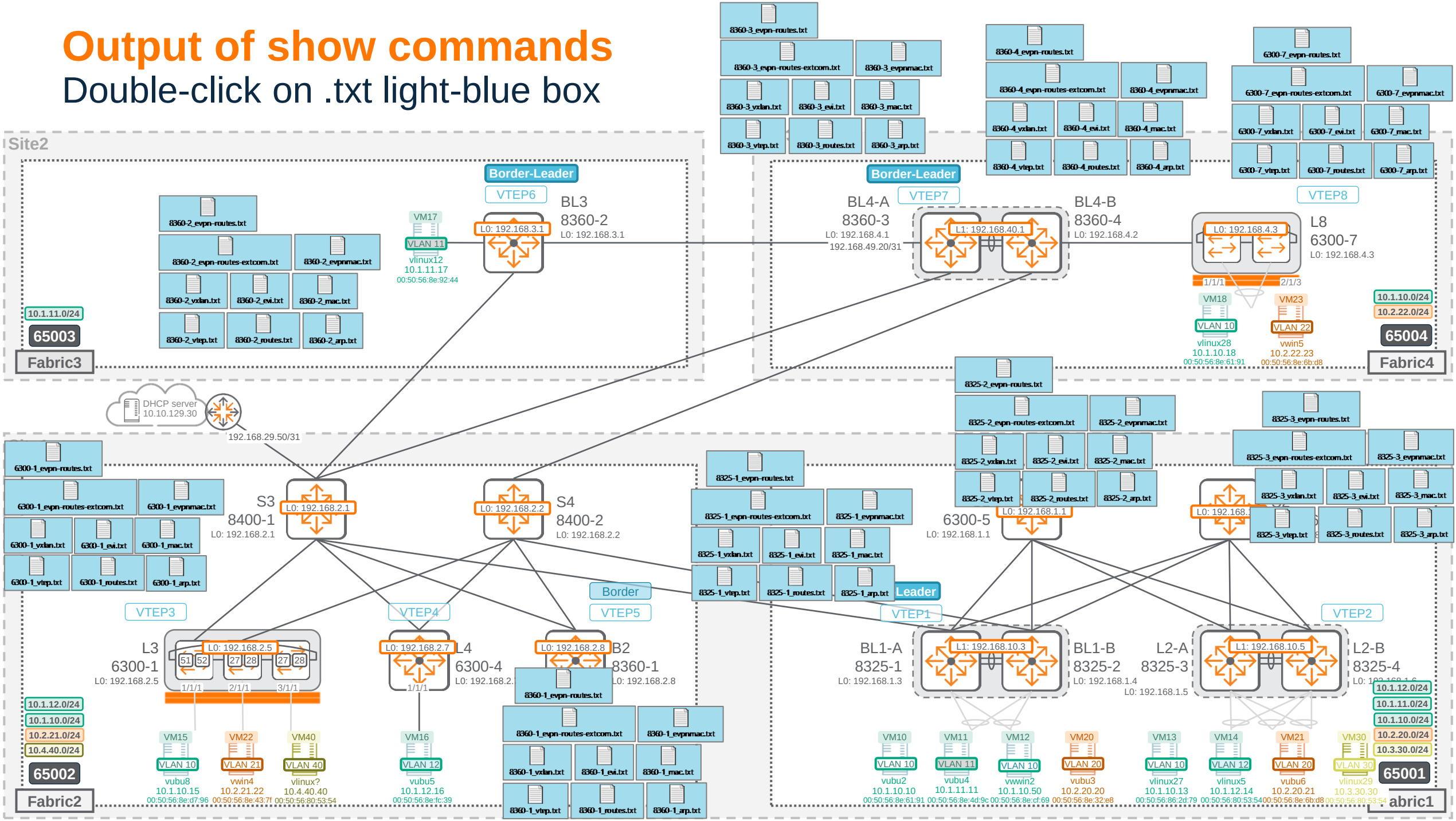


Troubleshooting

Example

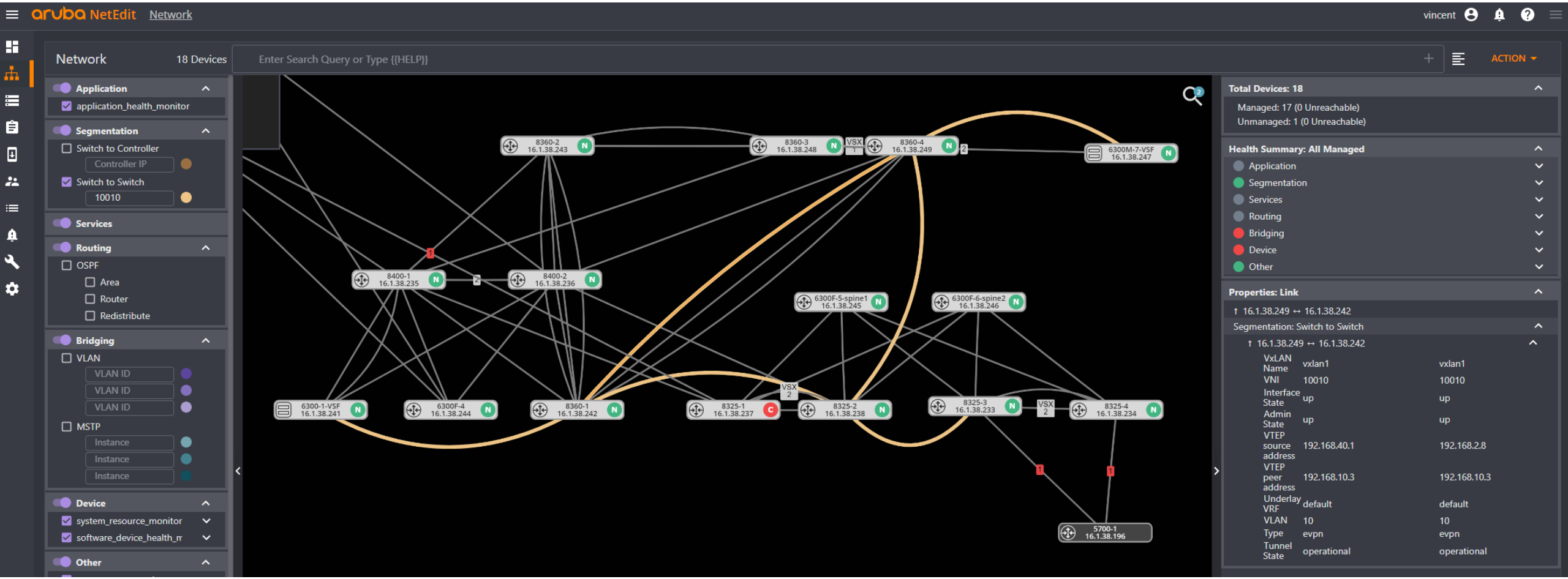
Output of show commands

Double-click on .txt light-blue box



NetEdit - Toplogy View

VNI 10010



Reminders

Cheat Sheet

1. Type-3 routes always show L2VNI routes only as this are IMET routes for BUM traffic (no L3).
2. Type-2 routes are used for L2VNI, and L3VNI in case of host-route redistribution.
   Expect 2 entries per NLRI per EVPN BGP peer: one for L2 traffic (L2VNI), one for routed-traffic (L3VNI).
   It can lead to lot of NLRIs to troubleshoot. Please identify the “Best” route for each VNI type.
3. For L3VNI, next-hop destination MAC (router-MAC) of the remote border-VTEP is found with:

```
8325-2# show evpn vtep-neighbor all-vrfs
VTEP-IP         L3VNI    MAC               VRF                                  State
-------------------------------------------------------------------------------------
192.168.3.1     100001   02:00:00:00:06:00 VRF1                                 Up
192.168.10.5    100001   02:00:00:00:02:00 VRF1                                 Up
192.168.40.1    100002   02:00:00:00:07:00 VRF2                                 Up
192.168.40.1    100001   02:00:00:00:07:00 VRF1                                 Up
192.168.2.8     100002   02:00:00:00:05:00 VRF2                                 Up
192.168.2.8     100001   02:00:00:00:05:00 VRF1                                 Up
192.168.10.5    100002   02:00:00:00:02:00 VRF2                                 Up
```

4. Pay attention to **silent hosts**, especially after clearing MAC/ARP or rebooting switch.
5. Ping to tenant host from VSX VTEP switch may fail if not sourced from loopback in the tenant VRF !
   Ping from host to AG IP is, however, always successful.

How many EVPN NLRI are expected to be received on borders?

Type-3 routes: L2VNI only

- One per remote border-VTEP (logical)
- One per eBGP EVPN peer

→ 2 candidates and 1 route selected (oldest or lowest peer IP)

```
8360-2# sh bgp l2 e route-type 3
Status codes: s suppressed, d damped, h history, * valid, > best, = multipath,
              i internal, e external S Stale, R Removed, a additional-paths
Origin codes: i - IGP, e - EGP, ? - incomplete

EVPN Route-Type 2 prefix: [2]:[ESI]:[EthTag]:[MAC]:[OrigIP]
EVPN Route-Type 3 prefix: [3]:[EthTag]:[OrigIP]
EVPN Route-Type 5 prefix: [5]:[ESI]:[EthTag]:[IPAddrLen]:[IPAddr]
VRF : default
Local Router-ID 192.168.3.1

     Network                                          Nexthop         Metric     LocPrf    Weight   Path
-------------------------------------------------------------------------------------------------------------
Route Distinguisher: 192.168.10.3:10      (L2VNI 10010)
*>e [3]:[0]:[192.168.10.3]                            192.168.10.3    0          100       0       65001 ?
* e [3]:[0]:[192.168.10.3]                            192.168.10.3    0          100       0       65001 ?

Route Distinguisher: 192.168.2.8:10       (L2VNI 10010)
*>e [3]:[0]:[192.168.2.8]                             192.168.2.8     0          100       0       65001 65002 ?
* e [3]:[0]:[192.168.2.8]                             192.168.2.8     0          100       0       65001 65002 ?

Route Distinguisher: 192.168.40.1:10      (L2VNI 10010)
*>e [3]:[0]:[192.168.40.1]                            192.168.40.1    0          100       0       65004 ?
* e [3]:[0]:[192.168.40.1]                            192.168.40.1    0          100       0       65004 ?

Route Distinguisher: 192.168.10.3:11      (L2VNI 10011)
*>e [3]:[0]:[192.168.10.3]                            192.168.10.3    0          100       0       65001 ?
* e [3]:[0]:[192.168.10.3]                            192.168.10.3    0          100       0       65001 ?

Route Distinguisher: 192.168.3.1:11       (L2VNI 10011)
*>  [3]:[0]:[192.168.3.1]                             192.168.3.1     0          100       0       ?

Route Distinguisher: 192.168.10.3:20      (L2VNI 10020)
*>e [3]:[0]:[192.168.10.3]                            192.168.10.3    0          100       0       65001 ?
* e [3]:[0]:[192.168.10.3]                            192.168.10.3    0          100       0       65001 ?
Total number of entries 11
```

```
8360-2# sh bgp l2 e 192.168.2.8:10-[3]:[0]:[192.168.2.8]

VRF : default
BGP Local AS 65003       BGP Router-id 192.168.3.1

   Network           : 192.168.2.8:10-[3]:[0]:[192.168.2.8]
   Nexthop           : 192.168.2.8
   vni               : 10010                    vni_type          : L2VNI
   Peer              : 192.168.1.3              Origin            : incomplete
   Metric            : 0                        Local Pref        : 100
   Weight            : 0                        Calc. Local Pref  : 100
   Best              : Yes                      Valid             : Yes
   Type              : external                 Stale             : No
   Originator ID     : 0.0.0.0
   Aggregator ID     :
   Aggregator AS     :
   Atomic Aggregate  :

   AS-Path           :  65001 65002

   Cluster List      :
   Communities       :
   Ext-Communities   : RT: 1:10

   Network           : 192.168.2.8:10-[3]:[0]:[192.168.2.8]
   Nexthop           : 192.168.2.8
   vni               : 10010                    vni_type          : L2VNI
   Peer              : 192.168.1.4              Origin            : incomplete
   Metric            : 0                        Local Pref        : 100
   Weight            : 0                        Calc. Local Pref  : 100
   Best              : No                       Valid             : Yes
   Type              : external                 Stale             : No
   Originator ID     : 0.0.0.0
   Aggregator ID     :
   Aggregator AS     :
   Atomic Aggregate  :

   AS-Path           :  65001 65002

   Cluster List      :
   Communities       :
   Ext-Communities   : RT: 1:10
```

How many EVPN routes are expected to

Type-5 routes: L3VNI only

- One per subnet (in demo typically /24)
- One per route-distinguisher (common versus distinct choice)
- One per eBGP EVPN peer
- Example: 10.1.10.0/24 subnet sourced from
  2 standalone VTEPs + 2 VSX VTEPs (distinct RD):
  - 2 standalone VTEPs: 2 NLRIs per VTEP due to VSX border-VTEP advertising (primary+secondary peer) => 4 NLRIs
  - 2 VSX VTEPs (distinct RD): 2 RDs per VSX, 2 NLRIs per RD (due to VSX border-VTEP primary + secondary peer) => 8 NLRIs
    It would have been 4 NLRIs with common RD choice for VSX VTEP.
  - **Total per subnet in this example: 12 NLRIs**
- Only one is selected as best route for VRF routing table

```
8360-2# sh ip rou 10.1.10.0/24 vrf VRF1

VRF: VRF1

   Prefix           : 10.1.10.0/24                    VRF(egress)      : -
   Nexthop          : 192.168.10.3                    Interface        : -
   Origin           : bgp                             Type             : bgp_evpn
   Distance         : 200                             Metric           : 0
   Age              : 01h:06m:02s                     Tag              : 0
   Encap Type       : vxlan                           Encap Details    : l3vni 100001
```

```
8360-2# sh bgp l2 e vni 100001 route-type 5
Local Router-ID 192.168.3.1
     Network                                          Nexthop         Metric     LocPrf     Weight   Path
-------------------------------------------------------------------------------------------------------------
Route Distinguisher: 192.168.1.3:1        (L3VNI 100001)
*>e [5]:[0]:[0]:[24]:[10.1.10.0]                      192.168.10.3    0          100        0        65001 ?
*>e [5]:[0]:[0]:[24]:[10.1.11.0]                      192.168.10.3    0          100        0        65001 ?
*>e [5]:[0]:[0]:[31]:[192.168.110.0]                  192.168.10.3    0          100        0        65001 ?
*>e [5]:[0]:[0]:[32]:[192.168.11.103]                 192.168.10.3    0          100        0        65001 ?
*>e [5]:[0]:[0]:[32]:[192.168.11.3]                   192.168.10.3    0          100        0        65001 ?
*>e [5]:[0]:[0]:[64]:[fd00:10:1:10::]                 192.168.10.3    0          100        0        65001 ?
*>e [5]:[0]:[0]:[64]:[fd00:10:1:11::]                 192.168.10.3    0          100        0        65001 ?

Route Distinguisher: 192.168.1.4:1        (L3VNI 100001)
*>e [5]:[0]:[0]:[24]:[10.1.10.0]                      192.168.10.3    0          100        0        65001 ?
*>e [5]:[0]:[0]:[24]:[10.1.11.0]                      192.168.10.3    0          100        0        65001 ?
*>e [5]:[0]:[0]:[31]:[192.168.110.0]                  192.168.10.3    0          100        0        65001 ?
*>e [5]:[0]:[0]:[32]:[192.168.11.103]                 192.168.10.3    0          100        0        65001 ?
*>e [5]:[0]:[0]:[32]:[192.168.11.4]                   192.168.10.3    0          100        0        65001 ?
*>e [5]:[0]:[0]:[64]:[fd00:10:1:10::]                 192.168.10.3    0          100        0        65001 ?
*>e [5]:[0]:[0]:[64]:[fd00:10:1:11::]                 192.168.10.3    0          100        0        65001 ?

Route Distinguisher: 192.168.1.5:1        (L3VNI 100001)
*>e [5]:[0]:[0]:[24]:[10.1.10.0]                      192.168.10.3    0          100        0        65001 ?
* e [5]:[0]:[0]:[24]:[10.1.10.0]                      192.168.10.3    0          100        0        65001 ?
*>e [5]:[0]:[0]:[24]:[10.1.12.0]                      192.168.10.3    0          100        0        65001 ?
* e [5]:[0]:[0]:[24]:[10.1.12.0]                      192.168.10.3    0          100        0        65001 ?
*>e [5]:[0]:[0]:[32]:[192.168.11.105]                 192.168.10.3    0          100        0        65001 ?
* e [5]:[0]:[0]:[32]:[192.168.11.105]                 192.168.10.3    0          100        0        65001 ?
*>e [5]:[0]:[0]:[32]:[192.168.11.5]                   192.168.10.3    0          100        0        65001 ?
* e [5]:[0]:[0]:[32]:[192.168.11.5]                   192.168.10.3    0          100        0        65001 ?
*>e [5]:[0]:[0]:[64]:[fd00:10:1:10::]                 192.168.10.3    0          100        0        65001 ?
* e [5]:[0]:[0]:[64]:[fd00:10:1:10::]                 192.168.10.3    0          100        0        65001 ?
*>e [5]:[0]:[0]:[64]:[fd00:10:1:12::]                 192.168.10.3    0          100        0        65001 ?
* e [5]:[0]:[0]:[64]:[fd00:10:1:12::]                 192.168.10.3    0          100        0        65001 ?

Route Distinguisher: 192.168.1.6:1        (L3VNI 100001)
*>e [5]:[0]:[0]:[24]:[10.1.10.0]                      192.168.10.3    0          100        0        65001 ?
* e [5]:[0]:[0]:[24]:[10.1.10.0]                      192.168.10.3    0          100        0        65001 ?
*>e [5]:[0]:[0]:[24]:[10.1.12.0]                      192.168.10.3    0          100        0        65001 ?
* e [5]:[0]:[0]:[24]:[10.1.12.0]                      192.168.10.3    0          100        0        65001 ?
*>e [5]:[0]:[0]:[32]:[192.168.11.105]                 192.168.10.3    0          100        0        65001 ?
* e [5]:[0]:[0]:[32]:[192.168.11.105]                 192.168.10.3    0          100        0        65001 ?
*>e [5]:[0]:[0]:[32]:[192.168.11.6]                   192.168.10.3    0          100        0        65001 ?
* e [5]:[0]:[0]:[32]:[192.168.11.6]                   192.168.10.3    0          100        0        65001 ?
*>e [5]:[0]:[0]:[64]:[fd00:10:1:12::]                 192.168.10.3    0          100        0        65001 ?
* e [5]:[0]:[0]:[64]:[fd00:10:1:12::]                 192.168.10.3    0          100        0        65001 ?

Route Distinguisher: 192.168.2.5:1        (L3VNI 100001)
*>e [5]:[0]:[0]:[24]:[10.1.10.0]                      192.168.2.8     0          100        0        65001 65002 ?
* e [5]:[0]:[0]:[24]:[10.1.10.0]                      192.168.2.8     0          100        0        65001 65002 ?
*>e [5]:[0]:[0]:[32]:[192.168.21.105]                 192.168.2.8     0          100        0        65001 65002 ?
* e [5]:[0]:[0]:[32]:[192.168.21.105]                 192.168.2.8     0          100        0        65001 65002 ?
*>e [5]:[0]:[0]:[32]:[192.168.21.5]                   192.168.2.8     0          100        0        65001 65002 ?
* e [5]:[0]:[0]:[32]:[192.168.21.5]                   192.168.2.8     0          100        0        65001 65002 ?

Route Distinguisher: 192.168.2.7:1        (L3VNI 100001)
*>e [5]:[0]:[0]:[24]:[10.1.12.0]                      192.168.2.8     0          200        0        65001 65002 ?
* e [5]:[0]:[0]:[24]:[10.1.12.0]                      192.168.2.8     0          200        0        65001 65002 ?
*>e [5]:[0]:[0]:[32]:[192.168.21.7]                   192.168.2.8     0          100        0        65001 65002 ?
* e [5]:[0]:[0]:[32]:[192.168.21.7]                   192.168.2.8     0          100        0        65001 65002 ?

Route Distinguisher: 192.168.2.8:1        (L3VNI 100001)
*>e [5]:[0]:[0]:[32]:[192.168.21.8]                   192.168.2.8     0          100        0        65001 65002 ?
* e [5]:[0]:[0]:[32]:[192.168.21.8]                   192.168.2.8     0          100        0        65001 65002 ?

Route Distinguisher: 192.168.3.1:1        (L3VNI 100001)
*>  [5]:[0]:[0]:[24]:[10.1.11.0]                      192.168.3.1     0          100        0        ?
*>  [5]:[0]:[0]:[32]:[192.168.31.101]                 192.168.3.1     0          100        0        ?
*>  [5]:[0]:[0]:[32]:[192.168.31.11]                  192.168.3.1     0          100        0        ?

Route Distinguisher: 192.168.4.1:1        (L3VNI 100001)
*>e [5]:[0]:[0]:[32]:[192.168.41.101]                 192.168.40.1    0          100        0        65004 ?
*>e [5]:[0]:[0]:[32]:[192.168.41.1]                   192.168.40.1    0          100        0        65004 ?

Route Distinguisher: 192.168.4.2:1        (L3VNI 100001)
*>e [5]:[0]:[0]:[32]:[192.168.41.101]                 192.168.40.1    0          100        0        65004 ?
*>e [5]:[0]:[0]:[32]:[192.168.41.2]                   192.168.40.1    0          100        0        65004 ?

Route Distinguisher: 192.168.4.3:1        (L3VNI 100001)
*>e [5]:[0]:[0]:[24]:[10.1.10.0]                      192.168.40.1    0          100        0        65004 ?
* e [5]:[0]:[0]:[24]:[10.1.10.0]                      192.168.40.1    0          100        0        65004 ?
*>e [5]:[0]:[0]:[32]:[192.168.41.103]                 192.168.40.1    0          100        0        65004 ?
* e [5]:[0]:[0]:[32]:[192.168.41.103]                 192.168.40.1    0          100        0        65004 ?
*>e [5]:[0]:[0]:[64]:[fd00:10:1:10::]                 192.168.40.1    0          100        0        65004 ?
* e [5]:[0]:[0]:[64]:[fd00:10:1:10::]                 192.168.40.1    0          100        0        65004 ?
Total number of entries 61
```

How many EVPN NLRI are expected to be received on borders?

Type-2 routes: L2VNI + optionally L3VNI

- One per MAC
- One per MAC+IP (ARP)
- One for L2VNI, one for L3VNI (if redistribute host-route)
- One per eBGP EVPN peer
- Note: RD is Common for type-2 (RD auto)

- Example: 00:50:56:8e:61:91 MAC sourced from one VSX VTEP (RD auto - common):
  - 2 entries for MAC only (L2VNI + L3VNI)
  - 2 entries for MAC+IP (L2VNI + L3VNI)
  - X2 as routes are received from VSX border-VTEP (primary+secondary peer)
  - **Total per MAC (with IP) in this example: 8 NLRIs**
    It would have been 4 NLRIs with no host-route redistribution on the VTEP for this VLAN (L2VNI only)

```
8360-2# sh bgp l2 evpn
Status codes: s suppressed, d damped, h history, * valid, > best, = multipath,
              i internal, e external S Stale, R Removed, a additional-paths
Origin codes: i - IGP, e - EGP, ? - incomplete

EVPN Route-Type 2 prefix: [2]:[ESI]:[EthTag]:[MAC]:[OrigIP]
EVPN Route-Type 3 prefix: [3]:[EthTag]:[OrigIP]
EVPN Route-Type 5 prefix: [5]:[ESI]:[EthTag]:[IPAddrLen]:[IPAddr]
VRF : default
Local Router-ID 192.168.3.1

     Network                                                Nexthop         Metric     LocPrf    Weight   Path
-------------------------------------------------------------------------------------------------------------------
Route Distinguisher: 192.168.10.3:10      (L2VNI 10010)
* e [2]:[0]:[0]:[00:50:56:8e:61:91]:[10.1.10.10]           192.168.10.3    0          100       0       65001 ?
*>e [2]:[0]:[0]:[00:50:56:8e:61:91]:[10.1.10.10]           192.168.10.3    0          100       0       65001 ?
*>e [2]:[0]:[0]:[00:50:56:8e:61:91]:[]                     192.168.10.3    0          100       0       65001 ?
* e [2]:[0]:[0]:[00:50:56:8e:61:91]:[]                     192.168.10.3    0          100       0       65001 ?
...
Route Distinguisher: 192.168.10.3:10      (L3VNI 100001)
* e [2]:[0]:[0]:[00:50:56:8e:61:91]:[10.1.10.10]           192.168.10.3    0          100       0       65001 ?
*>e [2]:[0]:[0]:[00:50:56:8e:61:91]:[10.1.10.10]           192.168.10.3    0          100       0       65001 ?
*>e [2]:[0]:[0]:[00:50:56:8e:61:91]:[]                     192.168.10.3    0          100       0       65001 ?
* e [2]:[0]:[0]:[00:50:56:8e:61:91]:[]                     192.168.10.3    0          100       0       65001 ?
```

Check IMET type-3 routes

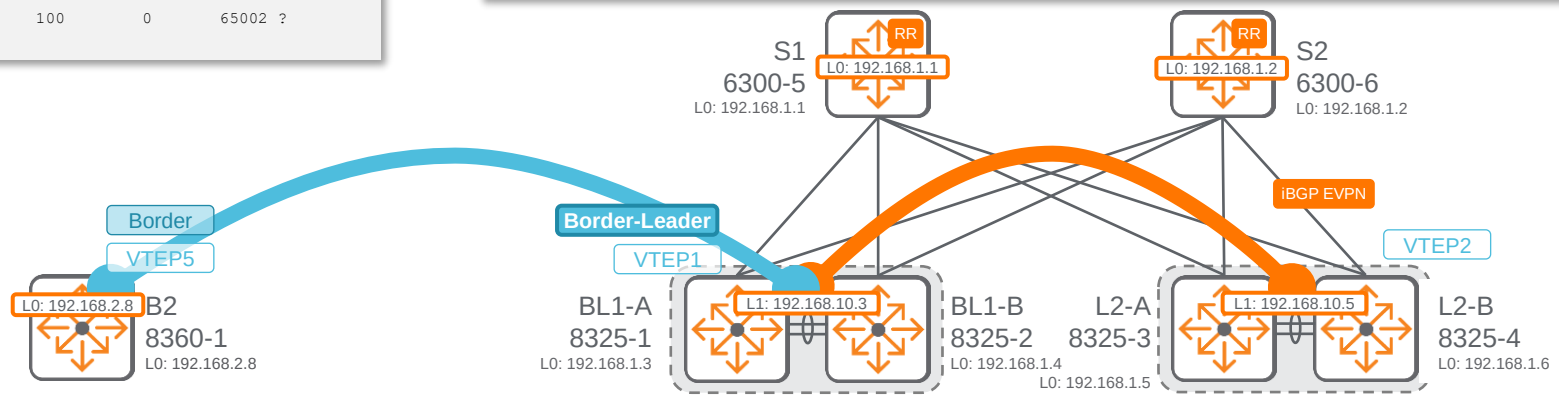
Border-Leader-VTEP1 does not advertise intra-Fabric1 VTEP to other Fabrics and does not advertise external-fabric VTEPs to internal Route-Reflectors

```
8325-1# sh bgp l2vpn evpn nei 192.168.2.8 advertised-routes route-type 3
...
Local Router-ID 192.168.1.3

     Network                                          Nexthop         Metric     LocPrf     Weight   Path
-------------------------------------------------------------------------------------------------------------
Route Distinguisher: 192.168.10.3:10      (L2VNI 10010)
*>e [3]:[0]:[192.168.10.3]                            192.168.10.3    0          0          0        65001 ?

Route Distinguisher: 192.168.40.1:10      (L2VNI 10010)
*>e [3]:[0]:[192.168.40.1]                            192.168.40.1    0          0          0        65001 65004 ?

Route Distinguisher: 192.168.10.3:11      (L2VNI 10011)
*>e [3]:[0]:[192.168.10.3]                            192.168.10.3    0          0          0        65001 ?

Route Distinguisher: 192.168.3.1:11       (L2VNI 10011)
*>e [3]:[0]:[192.168.3.1]                             192.168.3.1     0          0          0        65001 65003 ?

Route Distinguisher: 192.168.10.3:20      (L2VNI 10020)
*>e [3]:[0]:[192.168.10.3]                            192.168.10.3    0          0          0        65001 ?
Total number of entries 5
```

```
8325-1# sh bgp l2 e neighbors 192.168.2.8 routes route-type 3
...
Local Router-ID 192.168.1.3

     Network                                          Nexthop         Metric     LocPrf     Weight   Path
-------------------------------------------------------------------------------------------------------------
Route Distinguisher: 192.168.2.8:10       (L2VNI 10010)
*>e [3]:[0]:[192.168.2.8]                             192.168.2.8     0          100        0        65002 ?

Route Distinguisher: 192.168.2.8:12       (L2VNI 10012)
*>e [3]:[0]:[192.168.2.8]                             192.168.2.8     0          100        0        65002 ?
Total number of entries 2
```

```
8325-1# show bgp l2vpn evpn nei 192.168.1.1 advertised-routes route-type 3
...
Local Router-ID 192.168.1.3

     Network                                          Nexthop         Metric     LocPrf     Weight   Path
-------------------------------------------------------------------------------------------------------------
Route Distinguisher: 192.168.10.3:10      (L2VNI 10010)
*>i [3]:[0]:[192.168.10.3]                            192.168.10.3    0          100        0        ?

Route Distinguisher: 192.168.10.3:11      (L2VNI 10011)
*>i [3]:[0]:[192.168.10.3]                            192.168.10.3    0          100        0        ?

Route Distinguisher: 192.168.10.3:20      (L2VNI 10020)
*>i [3]:[0]:[192.168.10.3]                            192.168.10.3    0          100        0        ?
Total number of entries 3
```

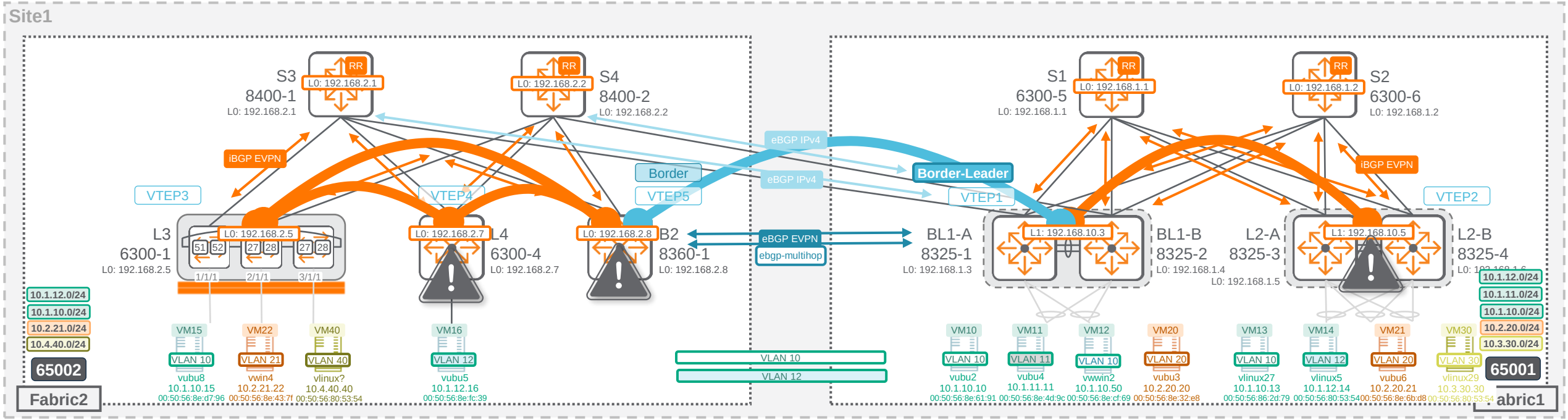
```
8325-1# sh bgp l2vpn evpn nei 192.168.1.1 routes route-type 3
...
Local Router-ID 192.168.1.3

     Network                                          Nexthop         Metric     LocPrf     Weight   Path
-------------------------------------------------------------------------------------------------------------
Route Distinguisher: 192.168.10.5:10      (L2VNI 10010)
*>i [3]:[0]:[192.168.10.5]                            192.168.10.5    0          100        0        ?

Route Distinguisher: 192.168.10.5:12      (L2VNI 10012)
*>i [3]:[0]:[192.168.10.5]                            192.168.10.5    0          100        0        ?

Route Distinguisher: 192.168.10.5:20      (L2VNI 10020)
*>i [3]:[0]:[192.168.10.5]                            192.168.10.5    0          100        0        ?
Total number of entries 3
```



Troubleshooting case: traffic between VM16 and VM14 is broken

Missing L2 configuration

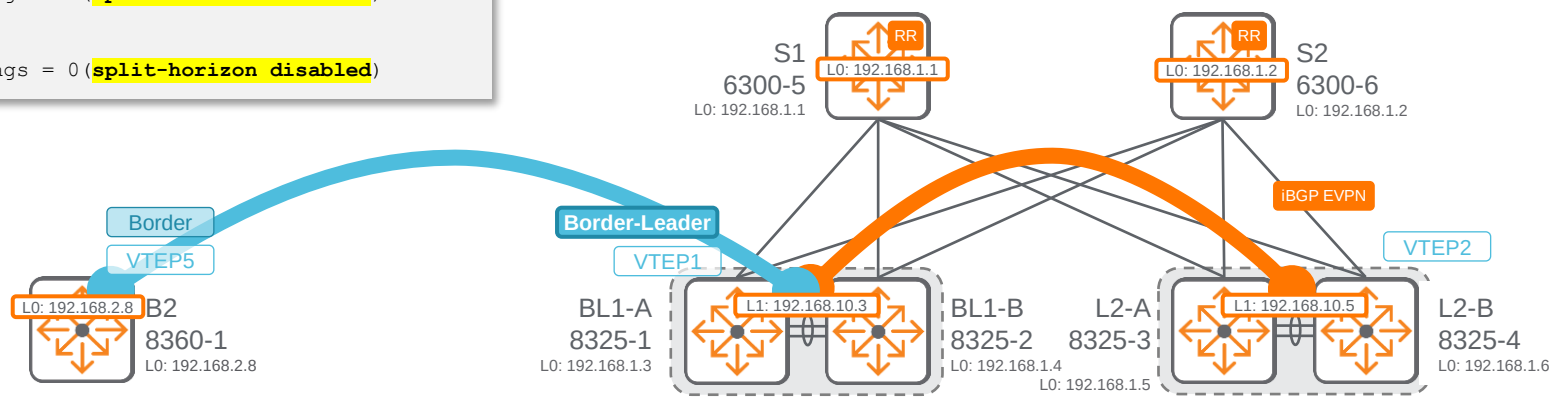
- Check Border EVPN
- Check border VXLAN L2VNI-VLAN mapping
- Check Route-Targets



Check Split-Horizon

ovs-appctl tunnel/vxlanPD/show_split_horizon_asic

```
8325-1:/home/admin# ovs-appctl tunnel/vxlanPD/show_split_horizon_asic

========VxLAN Split Horizon rule programmed in ASIC========
Intra_Fabric_Network to Intra_Fabric_Network:
src_network_group_id = 3, dest_network_group_id = 3, config_flags = 4(split-horizon enabled)

Access to Intra_Fabric_Network:
src_network_group_id = 2, dest_network_group_id = 3, config_flags = 0(split-horizon disabled)

Access to Access:
src_network_group_id = 2, dest_network_group_id = 2, config_flags = 0(split-horizon disabled)

ISL to Intra_Fabric_Network:
src_network_group_id = 4, dest_network_group_id = 3, config_flags = 5(split-horizon enabled)

Inter_Fabric_Network to Inter_Fabric_Network:
src_network_group_id = 5, dest_network_group_id = 5, config_flags = 4(split-horizon enabled)

Access to Inter_Fabric_Network:
src_network_group_id = 2, dest_network_group_id = 5, config_flags = 0(split-horizon disabled)

ISL to Inter_Fabric_Network:
src_network_group_id = 4, dest_network_group_id = 5, config_flags = 5(split-horizon enabled)

Intra_Fabric_Network to Inter_Fabric_Network:
src_network_group_id = 3, dest_network_group_id = 5, config_flags = 0(split-horizon disabled)

Intra_Fabric_Network to Inter_Fabric_Network:
src_network_group_id = 5, dest_network_group_id = 3, config_flags = 0(split-horizon disabled)
```



Resources

BGP decision tree reminder

Step	BGP attribute / criteria	Attribute type	Comment
1	next hop reachability		Next-hop in routing table (ex: in OSPF)
2	highest weight	proprietary	local to router
3	highest local-preference	discretionary	Globally defined within the AS. Default LP=100
4	router originated		route locally originated by itself (like redistribution or route leaking)
5	shortest AS-PATH length	Mandatory	• Can be skipped with bgp bestpath as-path ignore. • Selection can be stopped with bgp bestpath as-path multipath-relax
6	lowest origin type	Mandatory	IGP < EGP < incomplete
7	lowest **M**ulti-**E**xit-**D**iscriminator	Optional non-transitive	MEDs are compared if routes came from the same remote AS or if bgp always-compare-med is enabled
8	eBGP preferred over iBGP or confederation		confederation paths are treated as iBGP
9	lowest IGP cost to the BGP next-hop		closest IGP neighbor
10	Oldest eBGP route	N/A	most stable path
11	lowest router ID OR lowest originator ID (in case of RR)	Mandatory	
12	lowest cluster list length	Optional non-transitive	
13	lowest BGP peer IP address		IPv6 is preferred over IPv4

Feature/Solution References

- User Guides update:
  - VXLAN (10.08: https://www.arubanetworks.com/techdocs/AOS-CX/10.08/PDF/vxlan.pdf)

aruba
a Hewlett Packard Enterprise company

Thank you

vincent.giles@hpe.com
daryl.wan@hpe.com