AOS-CX 10.10 Update
June 2022

aruba
a Hewlett Packard Enterprise company

Campus Border VTEP as EVPN Route-Reflector

Vincent Giles
Technical Marketing Engineering

Agenda

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

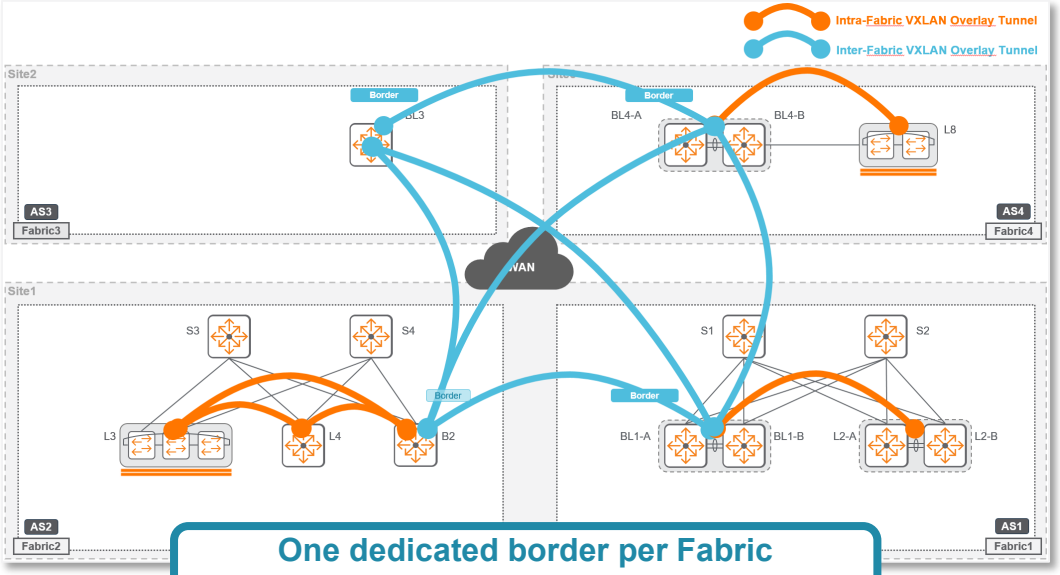
Definitions

Acronyms

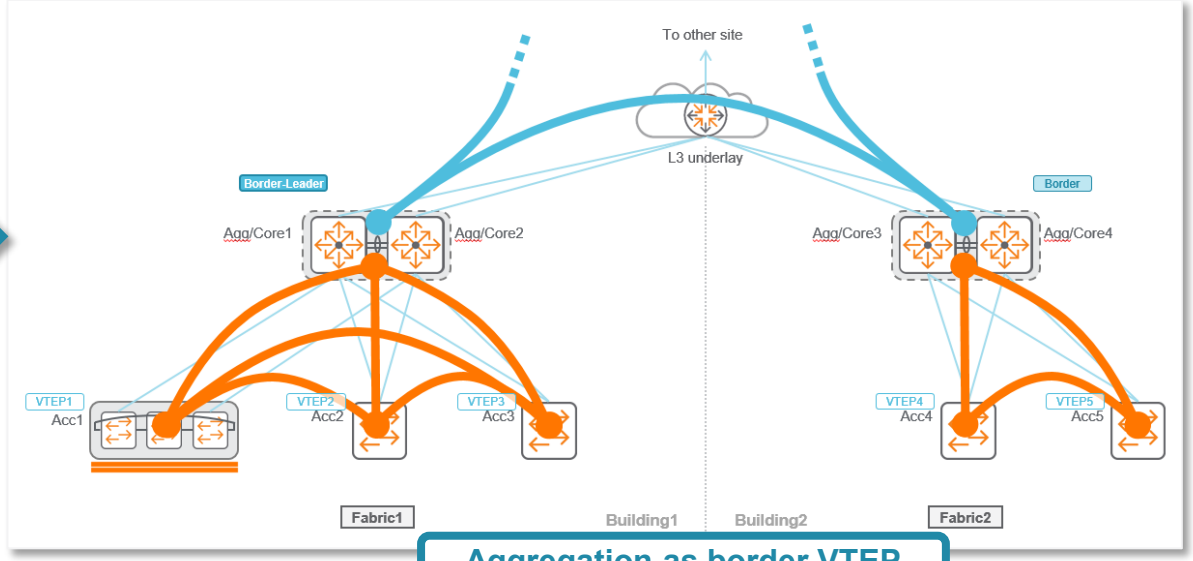
- VXLAN — **V**irtual e**X**tensible **LAN**
- VTEP — **V**XLAN **T**unnel **E**nd **P**oint
- VNI — **V**XLAN **N**etwork **I**dentifier
- L2VNI — **L**ayer2 **V**XLAN **N**etwork **I**dentifier (to extend L2 traffic)
- L3VNI — **L**ayer3 **V**XLAN **N**etwork **I**dentifier (to send routed traffic)
- 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)
- MP-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
- Border-Leader — Border VTEP hosting BGP sessions with other Fabrics
- Fabric — Set of fully-meshed VTEPs for the VXLAN dataplane
- 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
- Routing table — Valid routing entries selected from each active routing protocols based on the administrative distance
- FIB — **F**orwarding **I**nformation **B**ase, active forwarding entries programmed into ASIC based on the routing table
- RIB — **R**outing **I**nformation **B**ase, selected and non-selected candidate routes per routing protocol

EVPN-VXLAN Multi-Fabric

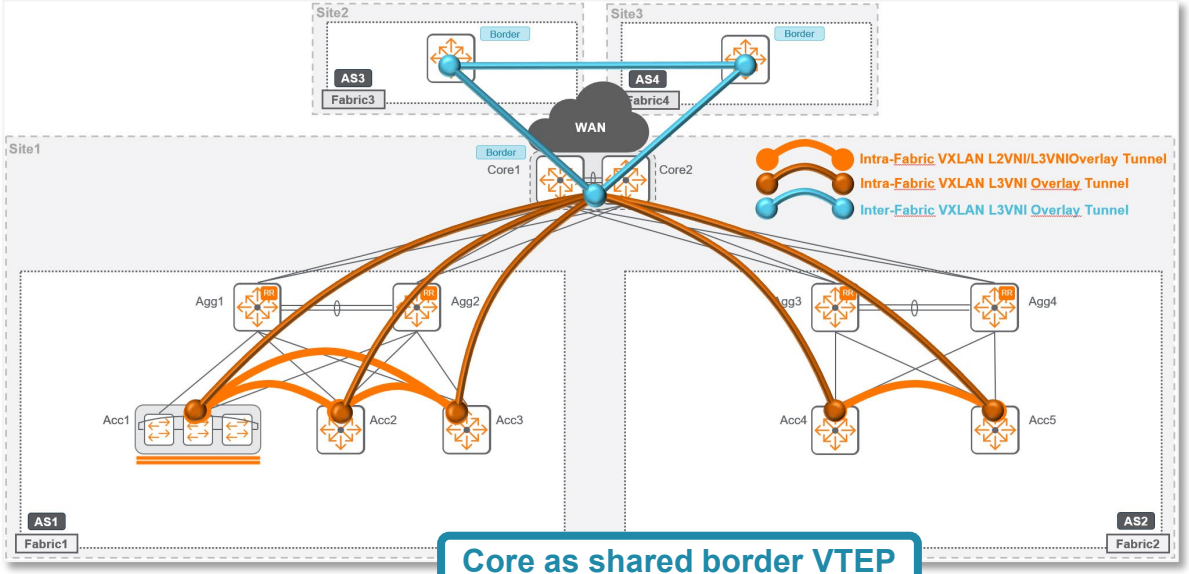
Datacenter / Campus



One dedicated border per Fabric
(DCN, unlikely for Campus access VTEP)



Aggregation as border VTEP

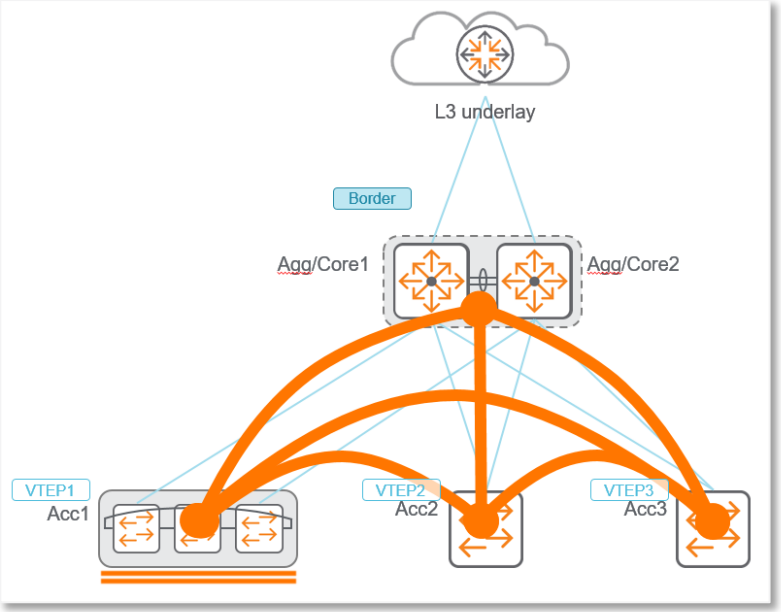


Core as shared border VTEP

Overview

MP-BGP EVPN next-hop-self

AOS-CX behavior in 10.09 and 10.10



- "next-hop-self" for EVPN address-family was introduced in 10.09

```
router bgp 65001
…
address-family l2vpn evpn
   neighbor a.b.c.d next-hop-self
```

New command in EVPN AF in 10.09

- In 10.09, this feature was qualified on border VTEP for the datacenter use-case.
- The current command in EVPN AF resets the BGP next-hop for **BOTH eBGP and iBGP routes**. The general practice for next-hop-self, and the genesis of this feature, is to reset next-hop **ONLY for eBGP routes**, as the standard usage for IPv4/IPv6 AF.
- 2 changes are required for future AOS-CX releases:
  - The next-hop-self command in EVPN AF should, by default, reset next-hop only for eBGP EVPN routes.
  - A command option should be introduced for networks where next-hop must be reset for both eBGP and iBGP EVPN routes (use-case: alternative to route-reflector model).
- For Campus VTEP Route-Reflector, an interim solution is proposed for 10.10, based on route-map.

New route-map set command for EVPN AF

set extcommunity evpn-rmac <border-leader-rmac>

```
switch(config-route-map-test-10)# set extcommunity
  evpn-rmac  Router-MAC extended community
  rt         Route Target extended community

switch(config-route-map-test-10)# set extcommunity evpn-rmac
  MAC-ADDR  Specify the MAC address
```

New command in 10.10

- On Campus border route-reflector, use route-map to mimic next-hop-self behavior for eBGP routes only:

```
ip aspath-list local-fabric seq 10 permit ^$
!
route-map to-RR-client permit seq 10
     match aspath-list local-fabric
!
route-map to-RR-client permit seq 20
     set ip next-hop <border-leader-vtep-ip>
     set extcommunity evpn-rmac <border-leader-rmac>

router bgp <asn>
…
address-family l2vpn evpn
   neighbor access route-reflector-client
   neighbor access route-map to-RR-client out
   neighbor access send-community both
```

No set action for iBGP routes reflected to RR-clients

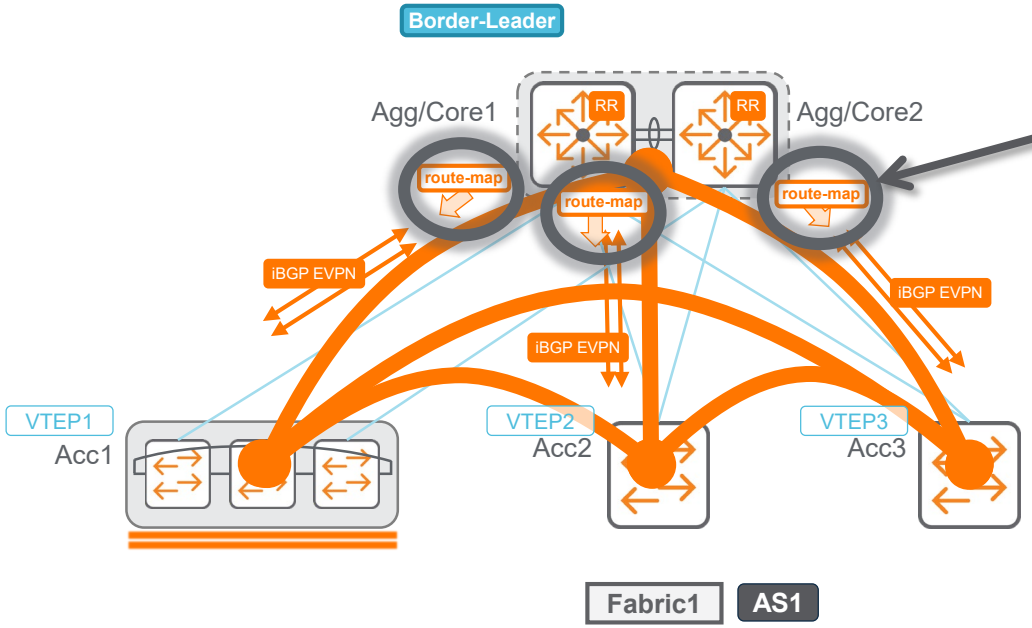
Next-hop IP and router-MAC reset for eBGP routes reflected to RR-clients

10.09 reminder

EVPN Type-3 routes optimization

New enhancement for 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.
- Automatic RT-3 filtering process is triggered through:
  - **next-hop-self** (towards iBGP neighbor with next-hop-self command or towards eBGP neighbor as regular eBGP behavior)
  - through outbound route-map to iBGP speaker (**new in 10.10**).



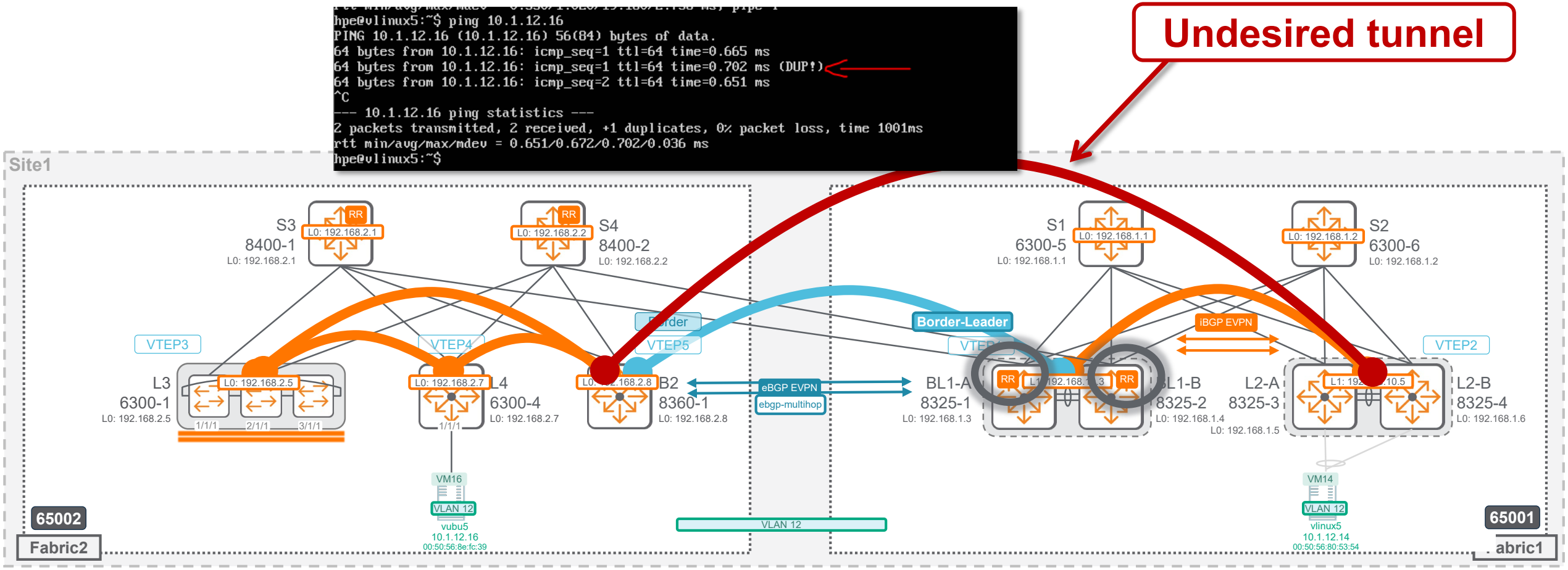

```
ip aspath-list local-fabric seq 10 permit ^$
!
route-map to-RR-client permit seq 10
     match aspath-list local-fabric
route-map to-RR-client permit seq 20
     set ip next-hop <border-leader-vtep-ip>
     set extcommunity evpn-rmac <border-leader-rmac>
!
router bgp <asn>
…
address-family l2vpn evpn
   neighbor access route-reflector-client
   neighbor access route-map to-RR-client out
   neighbor access send-community both
```

Without optimized Type-3 routes

Potential duplicate packets impact

- Does not happen with optimization triggered by outbound route-map to iBGP peer

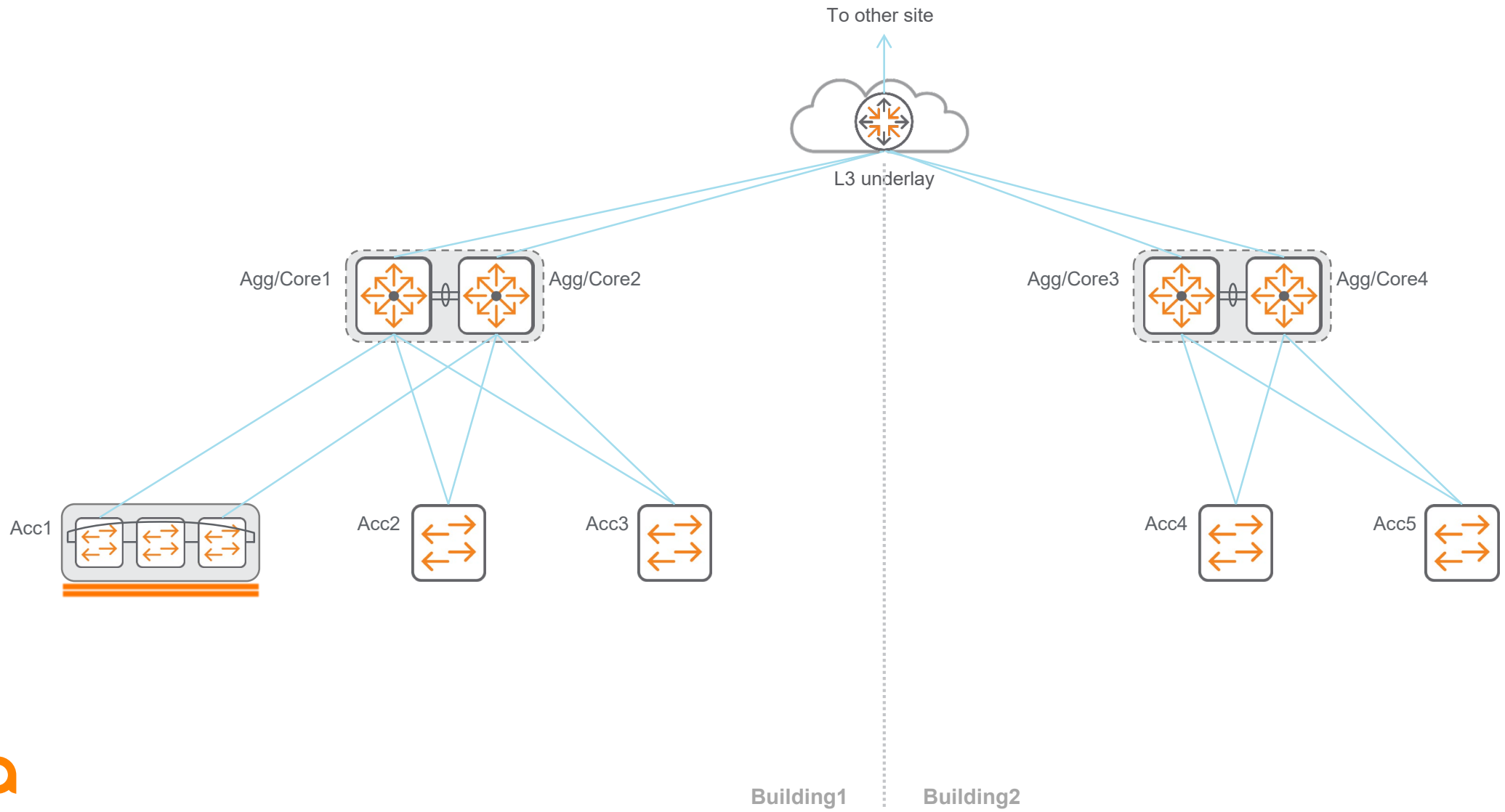
```
hpe@vlinux5:~$ ping 10.1.12.16
PING 10.1.12.16 (10.1.12.16) 56(84) bytes of data.
64 bytes from 10.1.12.16: icmp_seq=1 ttl=64 time=0.665 ms
64 bytes from 10.1.12.16: icmp_seq=1 ttl=64 time=0.702 ms (DUP!)
64 bytes from 10.1.12.16: icmp_seq=2 ttl=64 time=0.651 ms
^C
--- 10.1.12.16 ping statistics ---
2 packets transmitted, 2 received, +1 duplicates, 0% packet loss, time 1001ms
rtt min/avg/max/mdev = 0.651/0.672/0.702/0.036 ms
hpe@vlinux5:~$
```



Use Cases

Campus use-case

Topology



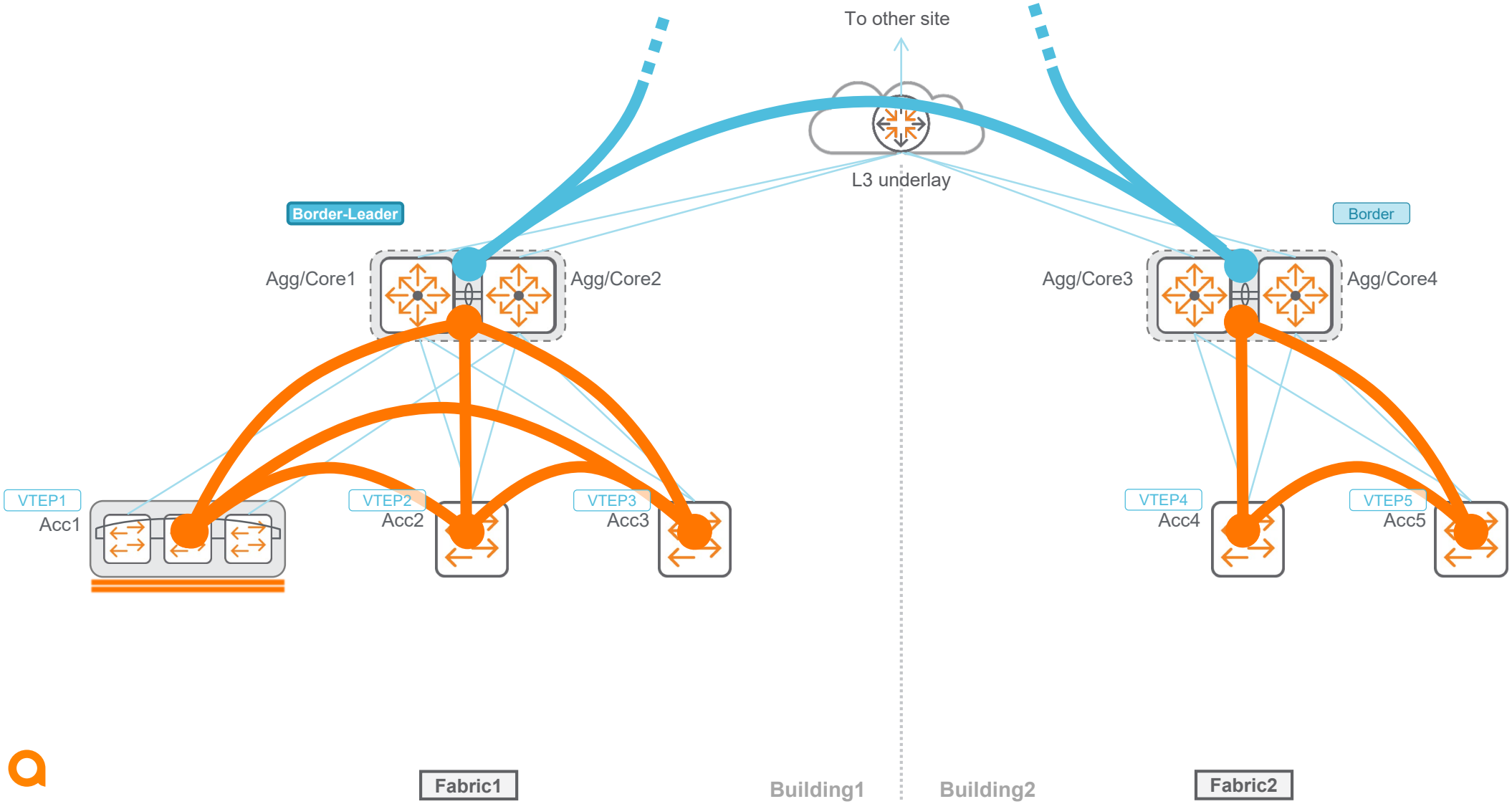
Campus Multi-Fabric

VXLAN tunnels

Intra-Fabric VXLAN Overlay Tunnel

Inter-Fabric VXLAN Overlay Tunnel

10.09 reminder



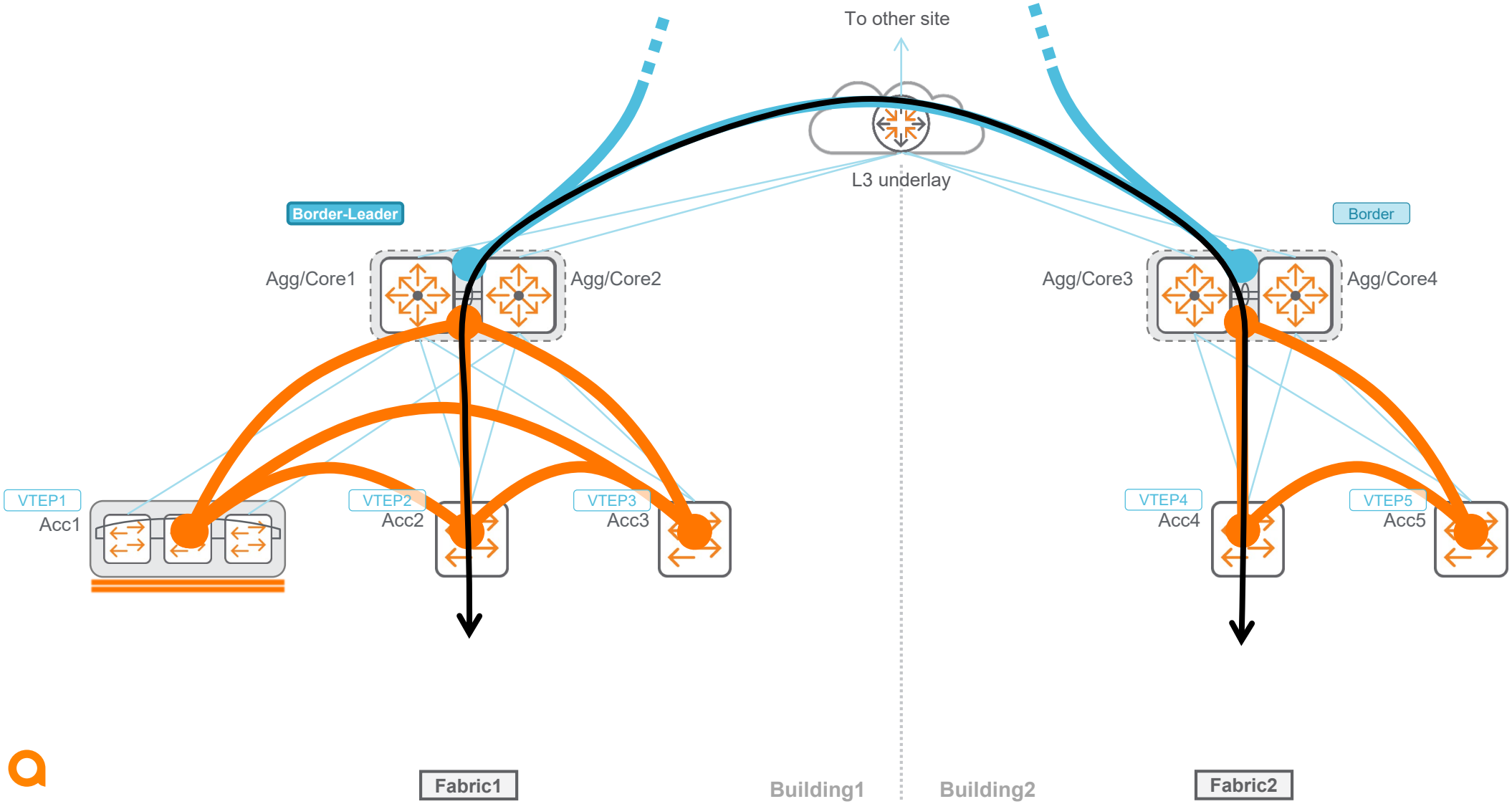
Inter-Fabric L2/L3 overlay traffic

Intra-Fabric VXLAN Overlay Tunnel

Inter-Fabric VXLAN Overlay Tunnel

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

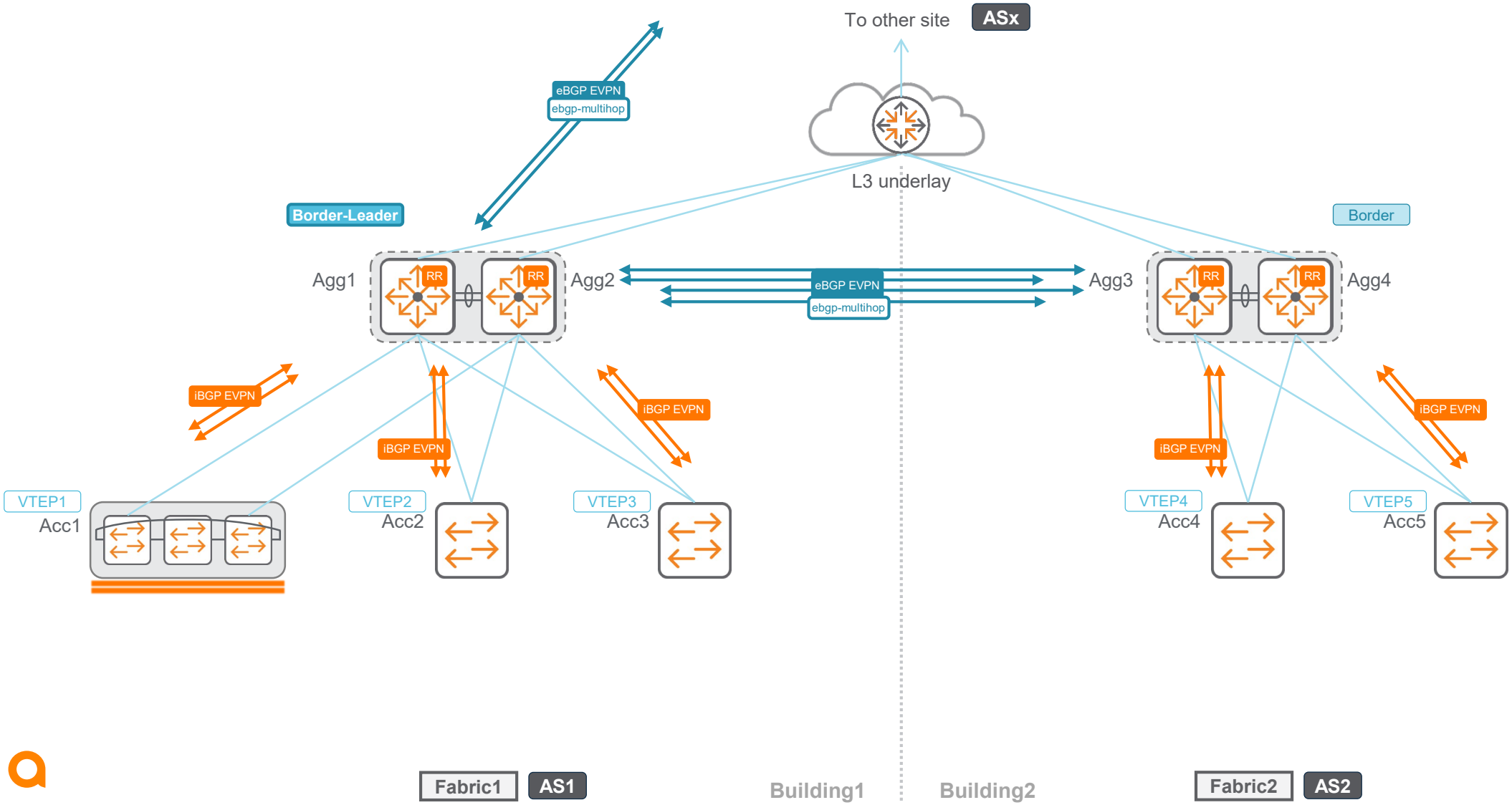
10.09 reminder



10.09 reminder

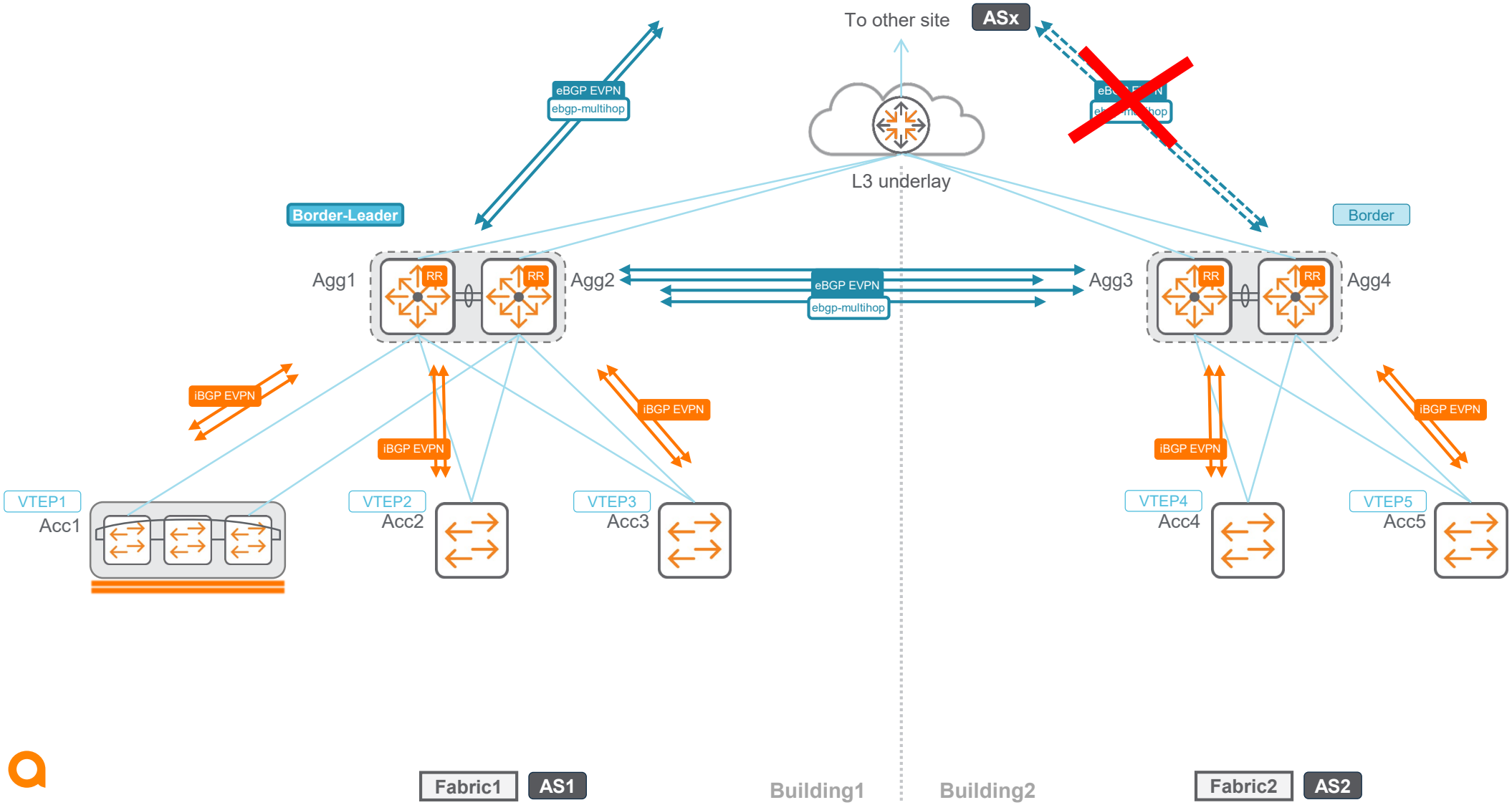
MP-BGP EVPN iBGP / eBGP sessions

Intra-Fabric, Intra-Site B-to-BL, Inter-Site BL-to-BL



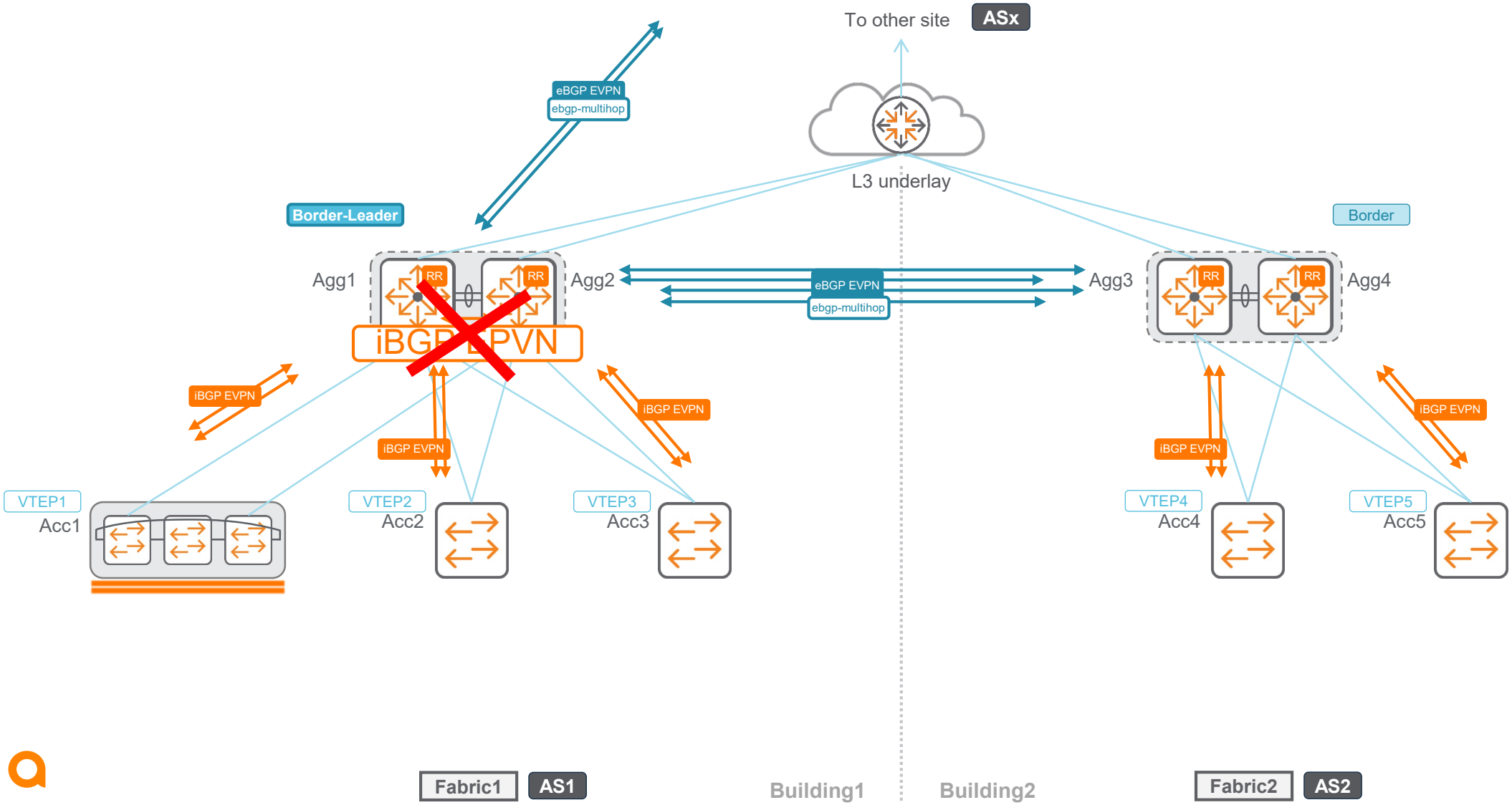
MP-BGP EVPN iBGP / eBGP sessions

Border-Leader concept



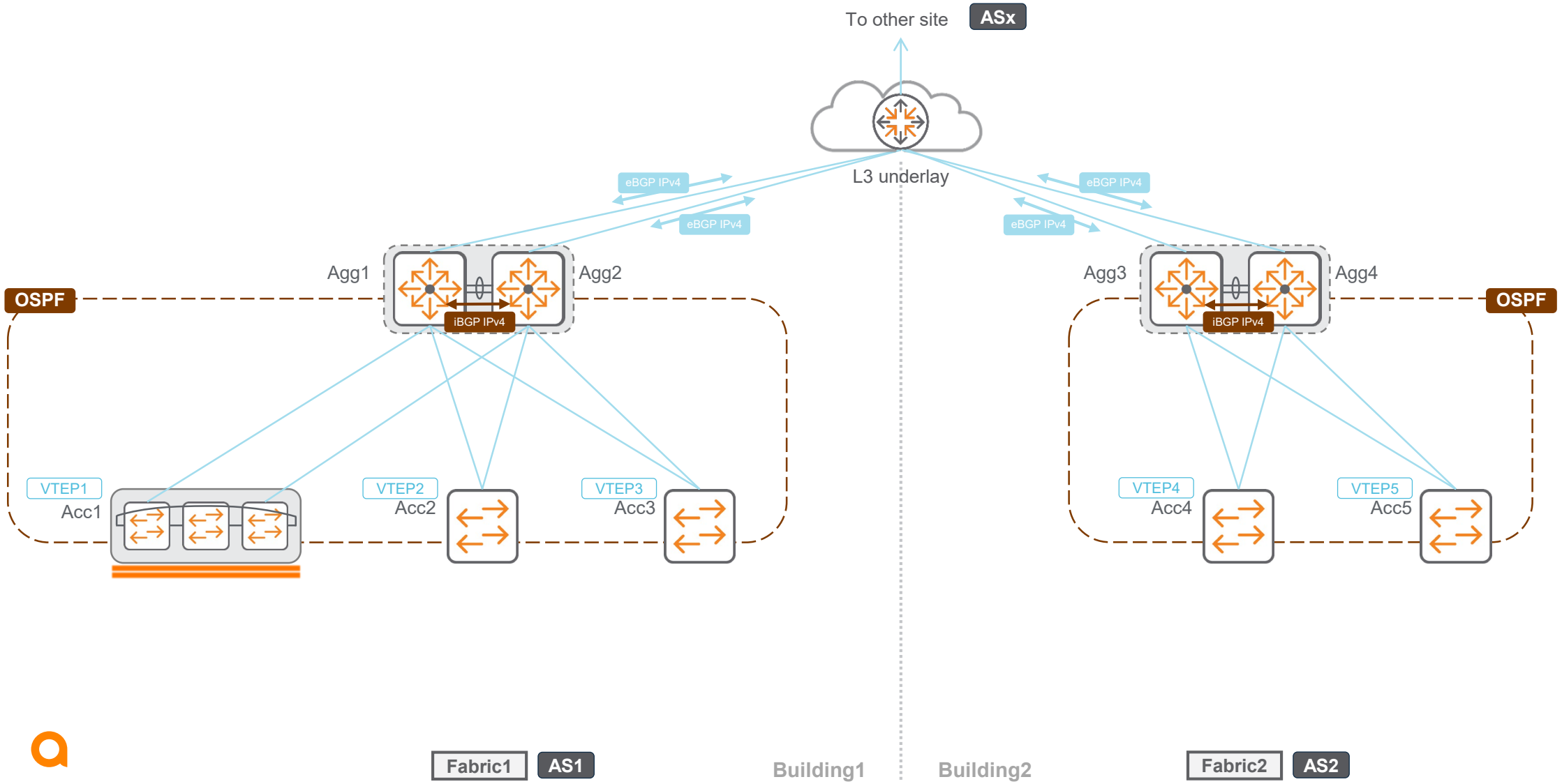
Route-Reflector: no shared BGP cluster-id

No intra-VSX BGP EVPN session between RR



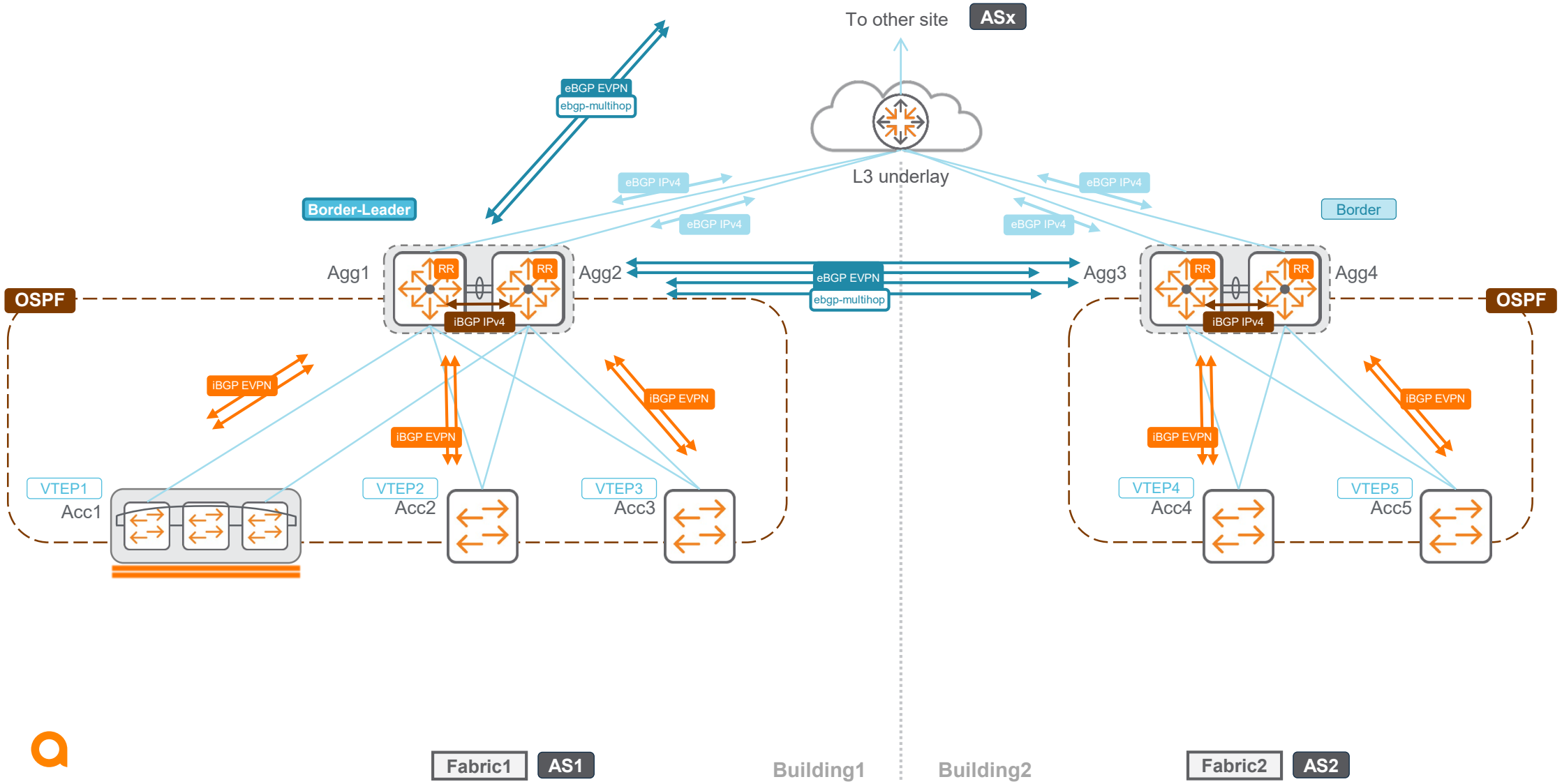
Underlay Control-Plane: eBGP IPv4 + OSPF

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



Control-Plane summary

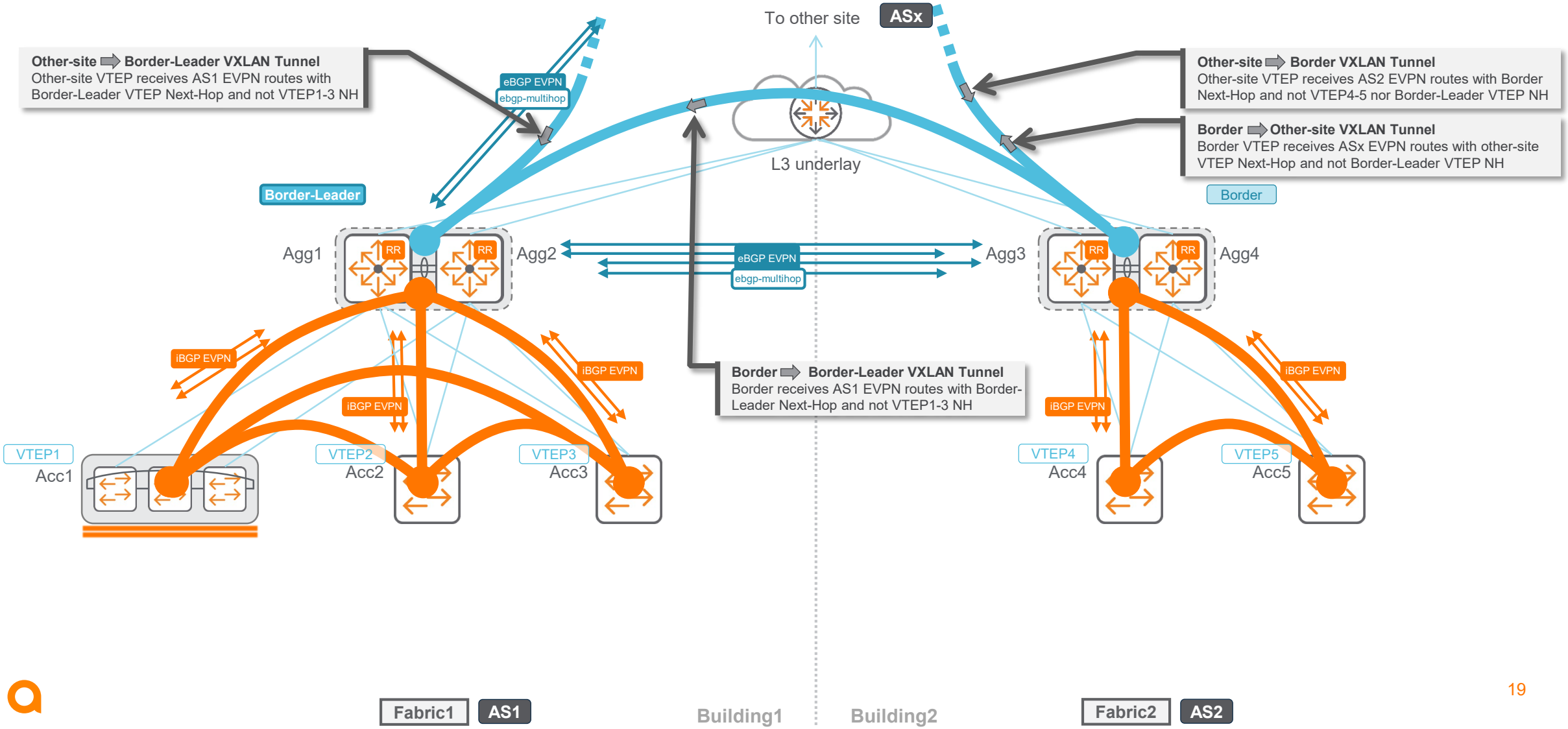
EVPN AF: iBGP and eBGP / IPV4 AF: iBGP and eBGP



10.09 reminder

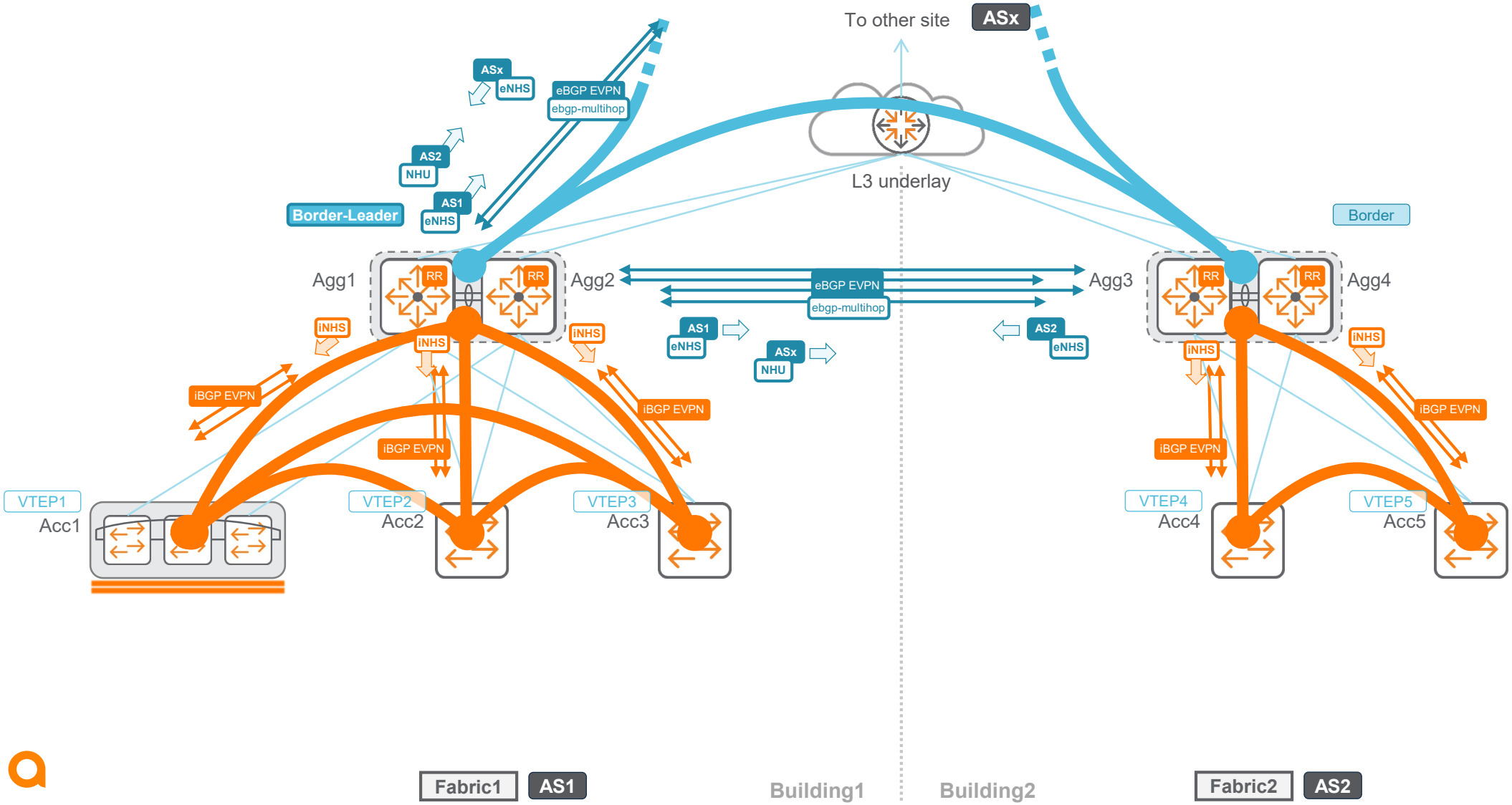
VTEP Next-Hop

Objective



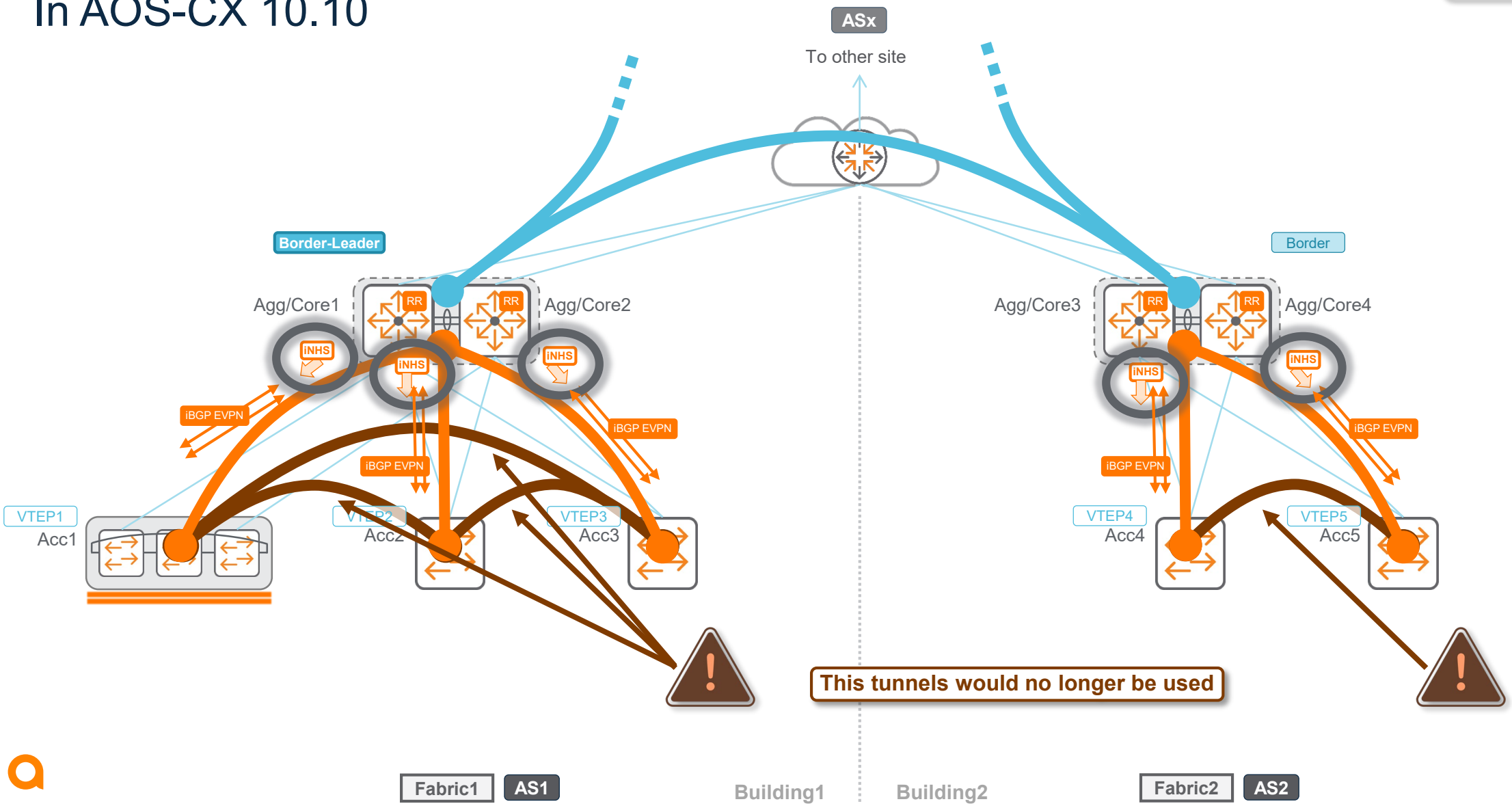
Control-Plane objective

Next-Hop-Self (iNHS/eNHS) and Next-Hop-Unchanged (NHU)



Don’t use next-hop-self command when border = RR !

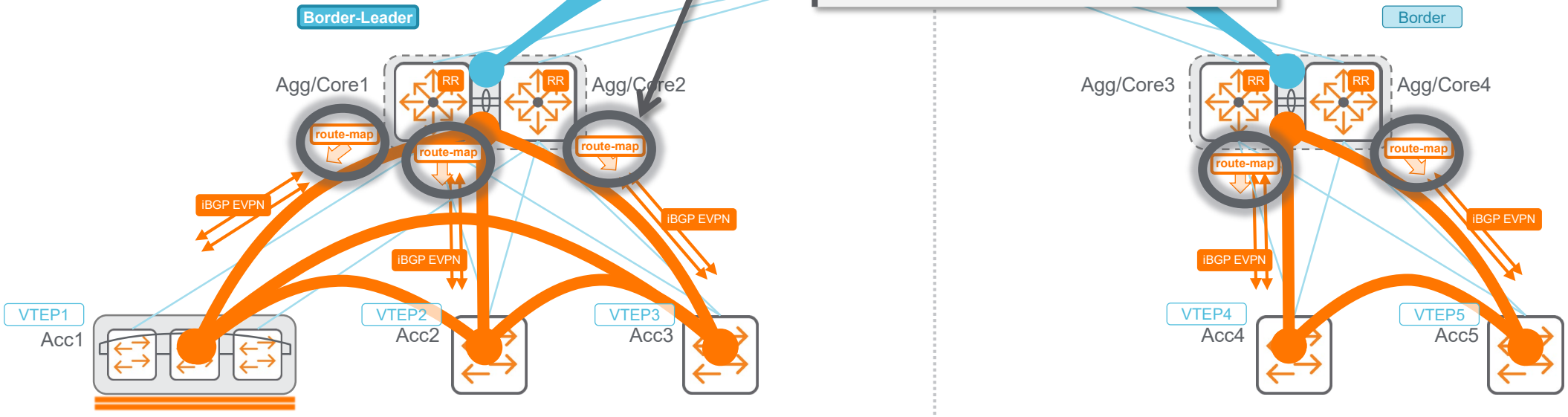
In AOS-CX 10.10



Use route-map command when border = RR !

Outbound route-map to RR-clients

```
ip aspath-list local-fabric seq 10 permit ^$
!
route-map to-RR-client permit seq 10
     match aspath-list local-fabric
route-map to-RR-client permit seq 20
     set ip next-hop <border-leader-vtep-ip>
     set extcommunity evpn-rmac <border-leader-rmac>
!
router bgp <asn>
…
address-family l2vpn evpn
   neighbor access route-reflector-client
   neighbor access route-map to-RR-client out
   neighbor access send-community both
```



Fabric1 AS1

Building1

Building2

Fabric2 AS2

Details / Caveats

Campus Border VTEP RR

Details / Caveats

- Support for both IPv4 and IPv6* unicast overlay.

** No IPV6 support claim for 10.10 public release notes and VXLAN documentation/user-guide. IPv6 unicast overlay official support is planned for 10.11.*

Border VTEP RR - 10.10 Platform Support

set extcommunity evpn-rmac

Platform	4100 6000 6100	6200	6300	6400 (v1/v2)	8320	8325	8360 (v1/v2)	8400	10000	Simulator
set extcommunity evpn-rmac	No	No	**Yes**	**Yes**	No	**Yes**	**Yes**	**Yes**	**Yes**	No

Use-case support of Border VTEP as route-reflector

Platform	4100 6000 6100	6200	6300	6400 (v1/v2)	8320	8325	8360 (v1/v2)	8400	10000	Simulator
set extcommunity evpn-rmac + Type-3 selective filter	No	No	No	**Yes**	No	**Yes**	**Yes**	No	No	No

Border VTEP RR - 10.10 Validated Multi-Dimensional Scale

	Border Leader VTEP		Border VTEP	
	8325	8360	8325	8360
HW profile	Leaf	Agg-Leaf	Leaf	Agg-Leaf
VTEPs per Fabric (standalone or VSX logical VTEP pair)	128	32	128	32
Sites (Number of VSX border-leader VTEPs)	32	32		
Fabrics across sites (Number of VSX border-VTEPs, VXLAN full-mesh)	32 (32x 64VTEPs)	32 (32x 32VTEPs)		
L3 routes across all VRFs and all sites (including host routes)	16K dual-stack	16K dual-stack		
Overlay hosts (MAC / ARP) across sites Notes: - Remote VTEPs share the same UD limit for overlay neighbors - MD test-case: some VNIs are L2 only (MAC only)	14K dual-stack	14K dual-stack		
VLANs local to the Fabric	512	512		
Stretched VLANs among all Fabrics	256	256		
VRFs shared among all Fabrics	32	32		

UD: Uni-Dimensional
MD: Multi-Dimensional

Configuration

Configuration - Campus Border VTEP route-reflector

Outbound route-map applied to route-reflector-clients MP-BGP EVPN peering

- Set local-border IP and loca-border router-MAC

```
ip aspath-list local-fabric seq 10 permit ^$
!
route-map to-RR-client permit seq 10
     match aspath-list local-fabric
!
route-map to-RR-client permit seq 20
     set ip next-hop <border-leader-vtep-ip>
     set extcommunity evpn-rmac <border-leader-rmac>

router bgp <asn>
…
address-family l2vpn evpn
   neighbor access route-reflector-client
   neighbor access route-map to-RR-client out
   neighbor access send-community both
   neighbor access next-hop-self
```

No set action for iBGP routes reflected to RR-clients. This retains iBGP next-hop VTEPs.

Next-hop IP and router-MAC reset for eBGP routes reflected to RR-clients

next-hop-self command must not be used

Configuration - Campus Border VTEP route-reflector

Target for **future** release

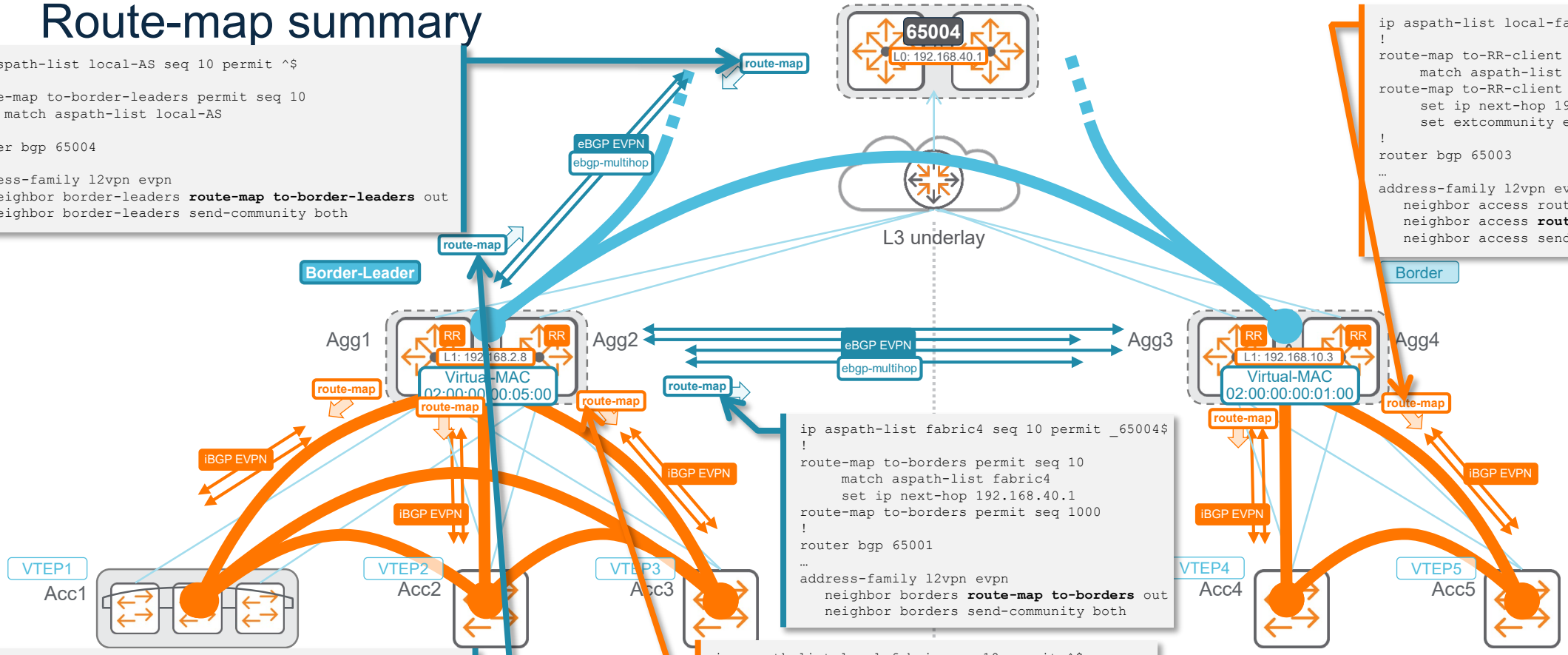
- Plan for homogenous configuration Campus/DC use-cases

```
router bgp <asn>
…
address-family l2vpn evpn
   neighbor access route-reflector-client
   neighbor access send-community both
   neighbor access next-hop-self
```

Campus border VTEP route-reflector for 10.10

Route-map summary

```
ip aspath-list local-AS seq 10 permit ^$
!
route-map to-border-leaders permit seq 10
     match aspath-list local-AS
!
router bgp 65004
…
address-family l2vpn evpn
   neighbor border-leaders route-map to-border-leaders out
   neighbor border-leaders send-community both
```




```
ip aspath-list local-fabric seq 10 permit ^$
!
route-map to-RR-client permit seq 10
     match aspath-list local-fabric
route-map to-RR-client permit seq 20
     set ip next-hop 192.168.10.3
     set extcommunity evpn-rmac 02:00:00:00:01:00
!
router bgp 65003
…
address-family l2vpn evpn
   neighbor access route-reflector-client
   neighbor access route-map to-RR-client out
   neighbor access send-community both
```

```
ip aspath-list fabric4 seq 10 permit _65004$
!
route-map to-borders permit seq 10
     match aspath-list fabric4
     set ip next-hop 192.168.40.1
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
```

```
ip aspath-list local-fabric seq 10 permit ^$
ip aspath-list fabric2 seq 10 permit _65002$
!
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 192.168.10.3
!
router bgp 65001
…
address-family l2vpn evpn
   neighbor border-leaders route-map to-border-leaders out
   neighbor border-leaders send-community both
```

```
ip aspath-list local-fabric seq 10 permit ^$
!
route-map to-RR-client permit seq 10
     match aspath-list local-fabric
route-map to-RR-client permit seq 20
     set ip next-hop 192.168.2.8
     set extcommunity evpn-rmac 02:00:00:00:05:00
!
router bgp 65001
…
address-family l2vpn evpn
   neighbor access route-reflector-client
   neighbor access route-map to-RR-client out
   neighbor access send-community both
```

Best Practices

Campus Border VTEP route-reflector

Best practices

- Border VTEP should be a pair of VSX nodes for better high-availability.
- Next-hop-self command could theoretically be kept in the address-family as outbound route-map, with 'set ip next-hop <addr>' configured, will override any 'next-hop-self' configuration under the BGP address-family for matched routes.

  For clarity, it is recommended to remove next-hop-self command from EVPN address-family.

Troubleshooting

- Refer to 10.09 TOI – Troubleshooting section

Campus border VTEP route-reflector Troubleshooting

1. In BGP EVPN AF configuration, check that route-map is properly applied to RR-clients.

```
switch# show run bgp | begin "address-family l2vpn evpn"
```

2. Check route-map “to-RR-client” and check that both IP and RMAC values are correct.

```
switch# show route-map to-RR-client
```

```
Route-map: to-RR-client
  Seq 10, permit,
    Match :
      aspath-list                  : local-fabric
    Set :
  Seq 20, permit,
    Match :
    Set :
      extcommunity evpn-rmac       : 02:00:00:00:05:00
      ip next-hop address          : 192.168.2.8
```

3. On access VTEP, check next-hop IP of eBGP routes.
   Alternatively, you may filter on eBGP routes from a particular AS-number (here 65004)

```
switch# show bgp l2vpn evpn paths
```

```
switch# show bgp l2vpn evpn paths | inc 65004
```

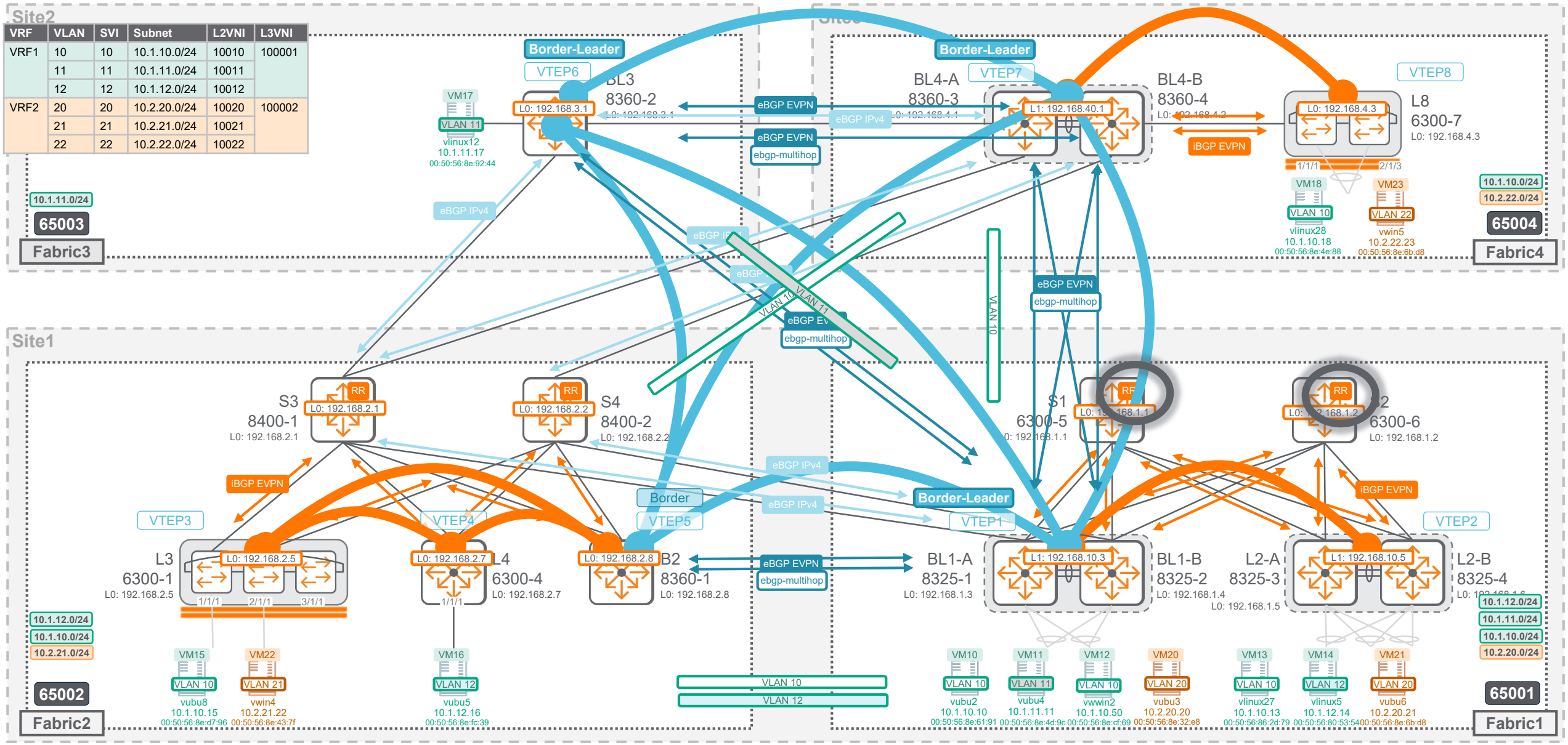
4. On access VTEP, check RMAC of eBGP routes.

```
switch# show bgp l2vpn extcommunity
```

Demo

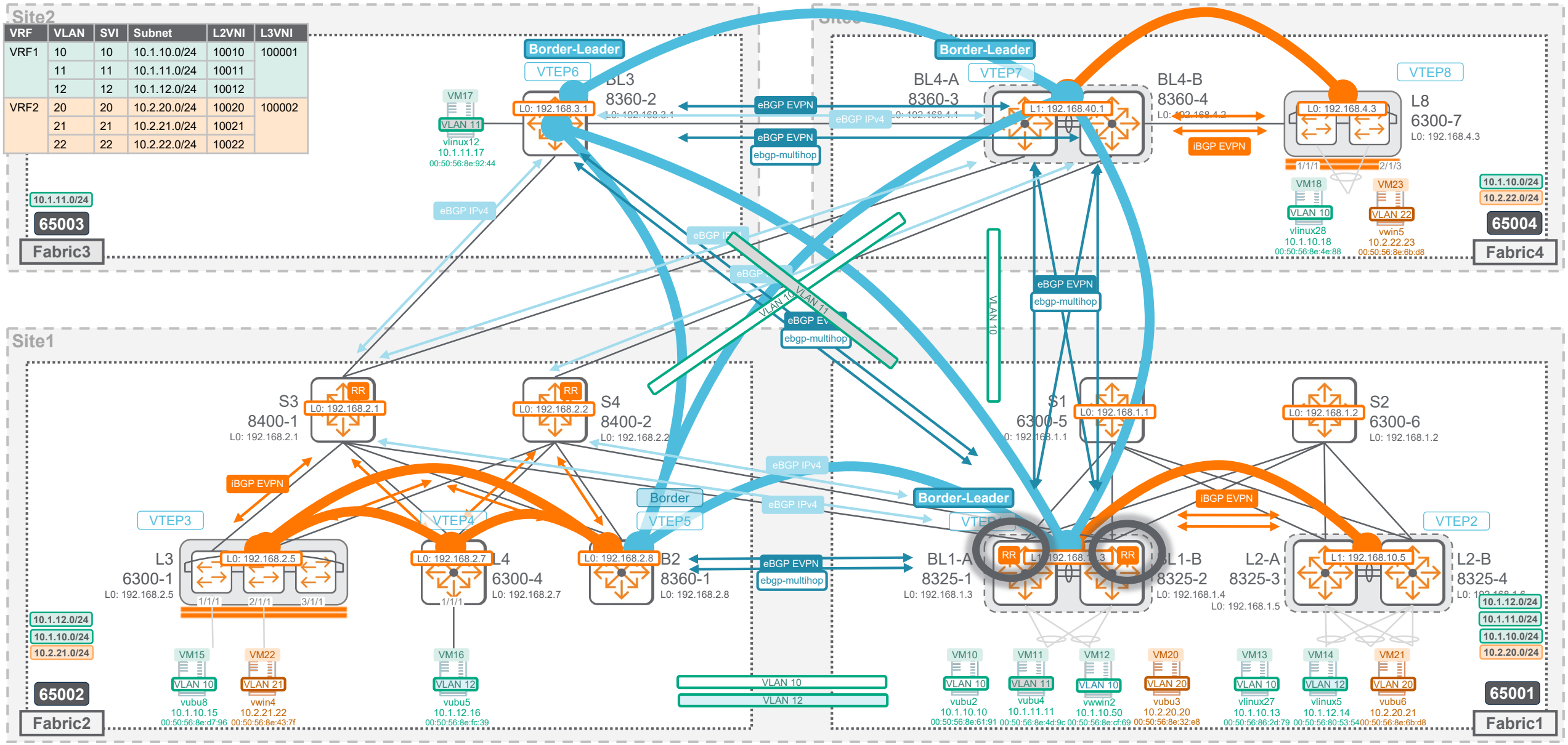
Demo infrastructure

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	



RR function moved to border VTEP in Fabric1

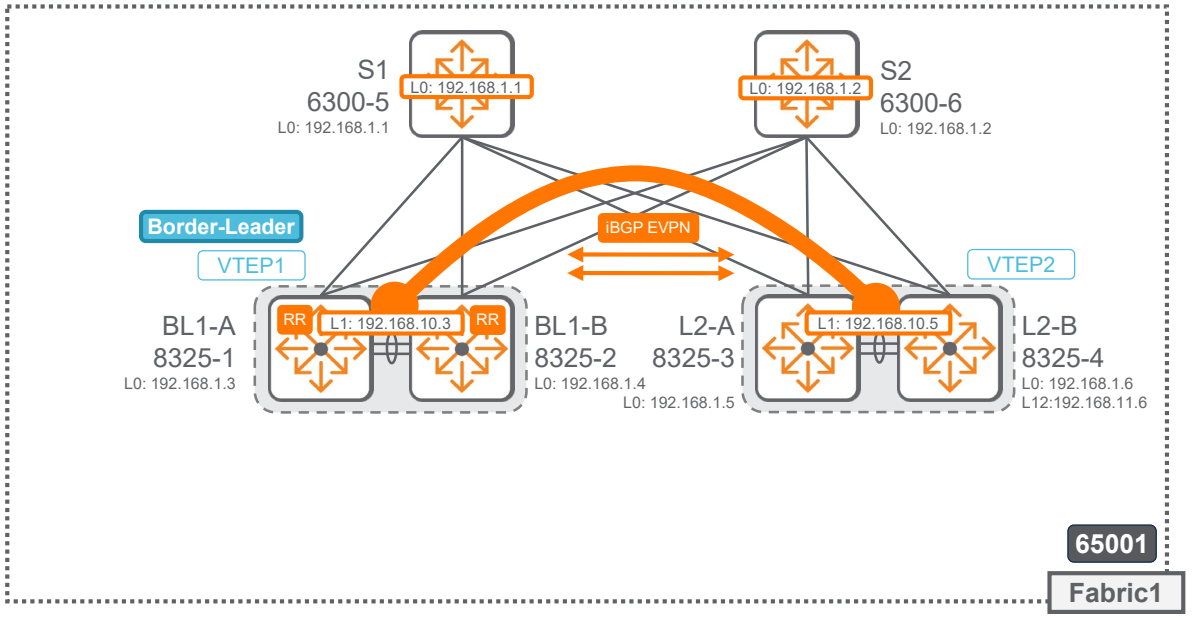
VRF	VLAN	SVI	Subnet	L2VNI	L3VNI
VRF1	10	10	10.1.10.0/24	10010	100001
	11	11	10.1.11.0/24	10011	
	12	12	10.1.12.0/24	10012	
VRF2	20	20	10.2.20.0/24	10020	100002
	21	21	10.2.21.0/24	10021	
	22	22	10.2.22.0/24	10022	



iBGP NH VTEPS are preserved

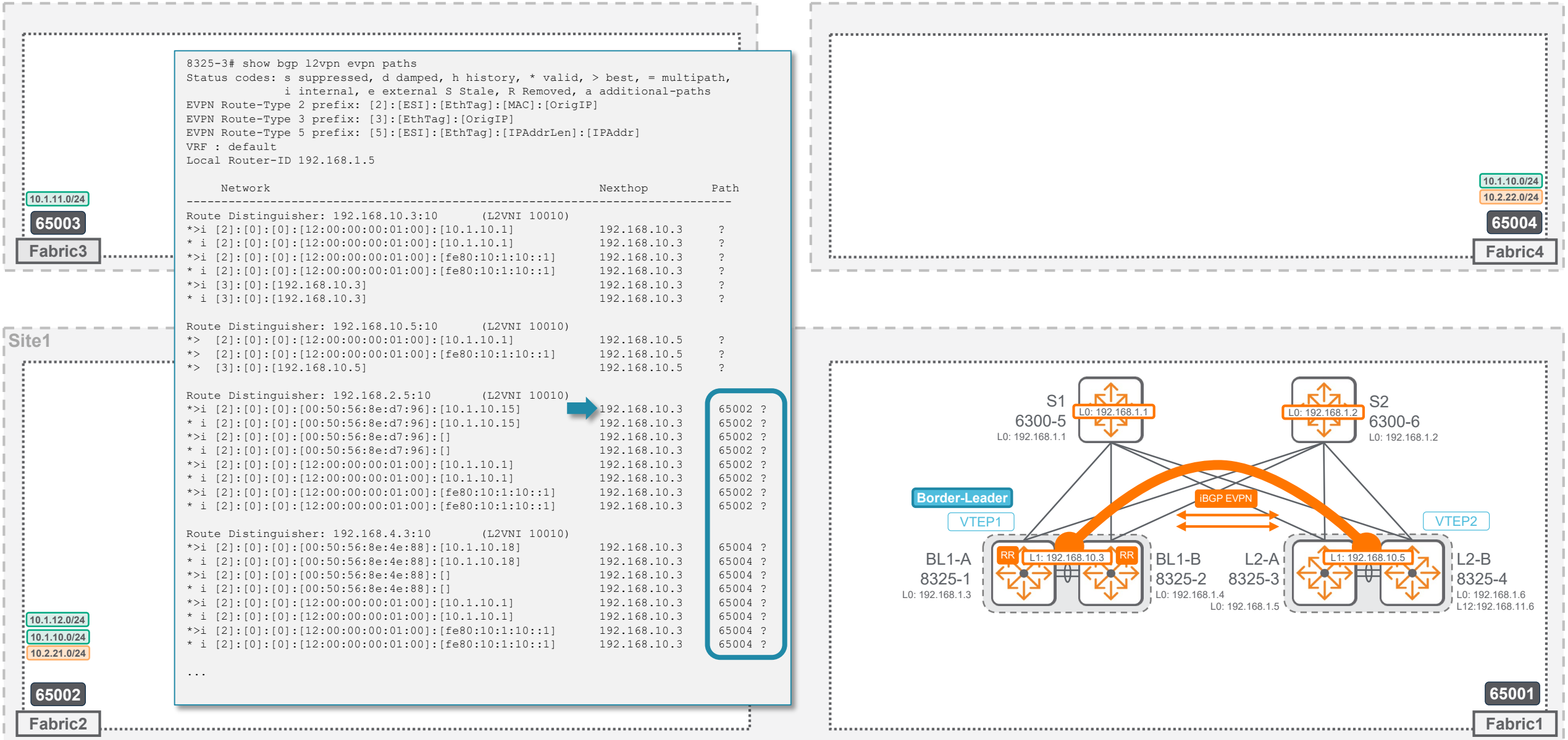
VTEP2 8325-4 L12: 192.168.11.6 in VRF1

```
8325-3# show bgp l2vpn evpn 192.168.10.5:1-[5]:[0]:[0]:[32]:[192.168.11.6]

VRF : default
BGP Local AS 65001       BGP Router-id 192.168.1.5

   Network           : 192.168.10.5:1-[5]:[0]:[0]:[32]:[192.168.11.6]
   Nexthop           : 192.168.10.5
   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              : internal                  Stale               : No
   Originator ID     : 192.168.1.6
   Aggregator ID     :
   Aggregator AS     :
   Atomic Aggregate  :

   AS-Path           :
   Cluster List      : 192.168.1.3
   Communities       :
   Ext-Communities   : RT: 1:1 RT: 65001:1 Router MAC: 02:00:00:00:02:00

   Network           : 192.168.10.5:1-[5]:[0]:[0]:[32]:[192.168.11.6]
   Nexthop           : 192.168.10.5
   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              : internal                  Stale               : No
   Originator ID     : 192.168.1.6
   Aggregator ID     :
   Aggregator AS     :
   Atomic Aggregate  :

   AS-Path           :
   Cluster List      : 192.168.1.4
   Communities       :
   Ext-Communities   : RT: 1:1 RT: 65001:1 Router MAC: 02:00:00:00:02:00
```



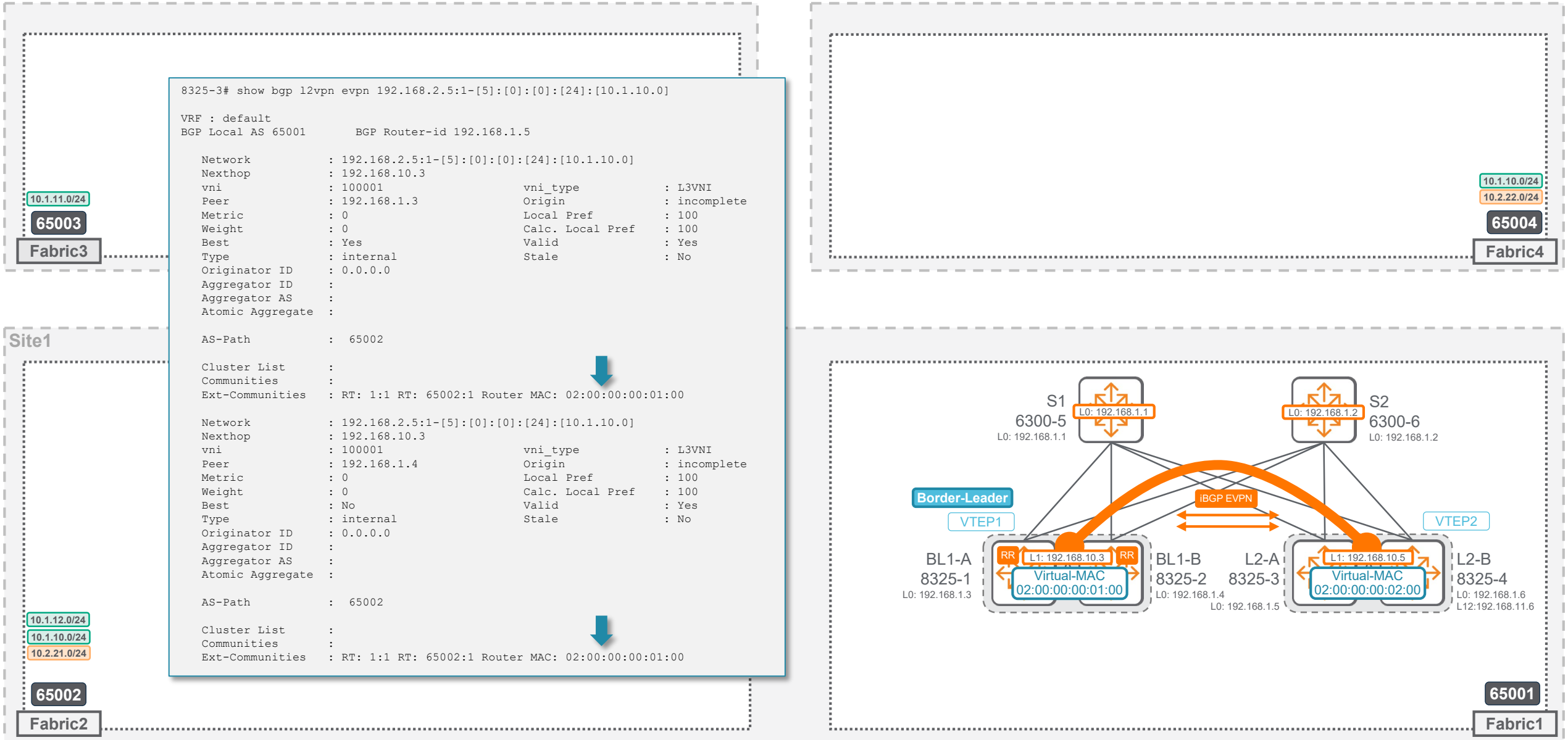
NH VTEP IP for other Fabrics is VTEP1

```
8325-3# show bgp l2vpn evpn paths
Status codes: s suppressed, d damped, h history, * valid, > best, = multipath,
              i internal, e external S Stale, R Removed, a additional-paths
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 192.168.1.5

     Network                                          Nexthop         Path
-------------------------------------------------------------------------------
Route Distinguisher: 192.168.10.3:10      (L2VNI 10010)
*>i [2]:[0]:[0]:[12:00:00:00:01:00]:[10.1.10.1]        192.168.10.3     ?
* i [2]:[0]:[0]:[12:00:00:00:01:00]:[10.1.10.1]        192.168.10.3     ?
*>i [2]:[0]:[0]:[12:00:00:00:01:00]:[fe80:10:1:10::1]  192.168.10.3     ?
* i [2]:[0]:[0]:[12:00:00:00:01:00]:[fe80:10:1:10::1]  192.168.10.3     ?
*>i [3]:[0]:[192.168.10.3]                             192.168.10.3     ?
* i [3]:[0]:[192.168.10.3]                             192.168.10.3     ?

Route Distinguisher: 192.168.10.5:10      (L2VNI 10010)
*>  [2]:[0]:[0]:[12:00:00:00:01:00]:[10.1.10.1]        192.168.10.5     ?
*>  [2]:[0]:[0]:[12:00:00:00:01:00]:[fe80:10:1:10::1]  192.168.10.5     ?
*>  [3]:[0]:[192.168.10.5]                             192.168.10.5     ?

Route Distinguisher: 192.168.2.5:10       (L2VNI 10010)
*>i [2]:[0]:[0]:[00:50:56:8e:d7:96]:[10.1.10.15]       192.168.10.3     65002 ?
* i [2]:[0]:[0]:[00:50:56:8e:d7:96]:[10.1.10.15]       192.168.10.3     65002 ?
*>i [2]:[0]:[0]:[00:50:56:8e:d7:96]:[]                 192.168.10.3     65002 ?
* i [2]:[0]:[0]:[00:50:56:8e:d7:96]:[]                 192.168.10.3     65002 ?
*>i [2]:[0]:[0]:[12:00:00:00:01:00]:[10.1.10.1]        192.168.10.3     65002 ?
* i [2]:[0]:[0]:[12:00:00:00:01:00]:[10.1.10.1]        192.168.10.3     65002 ?
*>i [2]:[0]:[0]:[12:00:00:00:01:00]:[fe80:10:1:10::1]  192.168.10.3     65002 ?
* i [2]:[0]:[0]:[12:00:00:00:01:00]:[fe80:10:1:10::1]  192.168.10.3     65002 ?

Route Distinguisher: 192.168.4.3:10       (L2VNI 10010)
*>i [2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]       192.168.10.3     65004 ?
* i [2]:[0]:[0]:[00:50:56:8e:4e:88]:[10.1.10.18]       192.168.10.3     65004 ?
*>i [2]:[0]:[0]:[00:50:56:8e:4e:88]:[]                 192.168.10.3     65004 ?
* i [2]:[0]:[0]:[00:50:56:8e:4e:88]:[]                 192.168.10.3     65004 ?
*>i [2]:[0]:[0]:[12:00:00:00:01:00]:[10.1.10.1]        192.168.10.3     65004 ?
* i [2]:[0]:[0]:[12:00:00:00:01:00]:[10.1.10.1]        192.168.10.3     65004 ?
*>i [2]:[0]:[0]:[12:00:00:00:01:00]:[fe80:10:1:10::1]  192.168.10.3     65004 ?
* i [2]:[0]:[0]:[12:00:00:00:01:00]:[fe80:10:1:10::1]  192.168.10.3     65004 ?

...
```



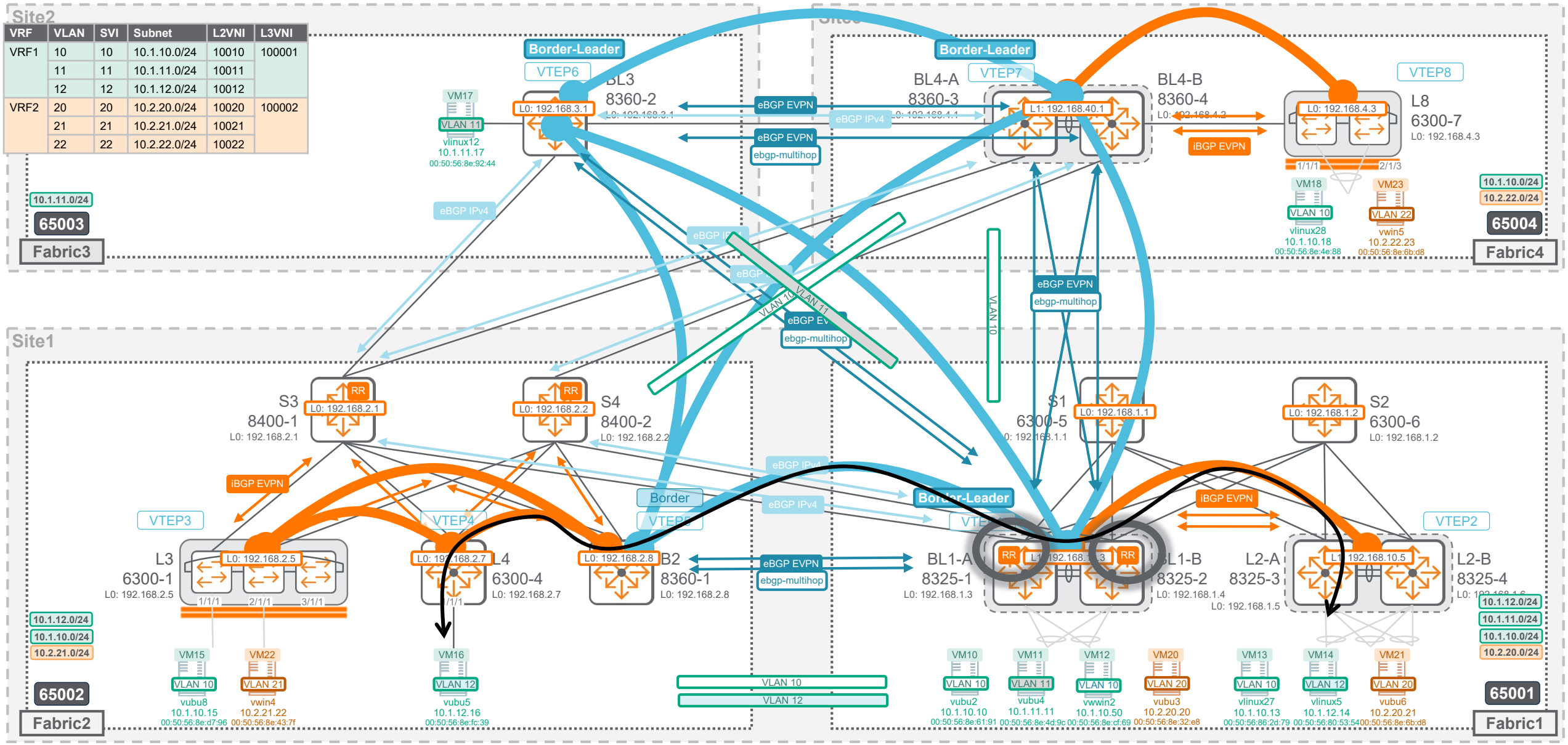
NH Router-MAC for other Fabric is VTEP1

```
8325-3# show bgp l2vpn evpn 192.168.2.5:1-[5]:[0]:[0]:[24]:[10.1.10.0]

VRF : default
BGP Local AS 65001       BGP Router-id 192.168.1.5

   Network           : 192.168.2.5:1-[5]:[0]:[0]:[24]:[10.1.10.0]
   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              : internal                  Stale               : No
   Originator ID     : 0.0.0.0
   Aggregator ID     :
   Aggregator AS     :
   Atomic Aggregate  :

   AS-Path           :  65002

   Cluster List      :
   Communities       :
   Ext-Communities   : RT: 1:1 RT: 65002:1 Router MAC: 02:00:00:00:01:00

   Network           : 192.168.2.5:1-[5]:[0]:[0]:[24]:[10.1.10.0]
   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              : internal                  Stale               : No
   Originator ID     : 0.0.0.0
   Aggregator ID     :
   Aggregator AS     :
   Atomic Aggregate  :

   AS-Path           :  65002

   Cluster List      :
   Communities       :
   Ext-Communities   : RT: 1:1 RT: 65002:1 Router MAC: 02:00:00:00:01:00
```



Ping

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	



Resources

Feature/Solution References

- User Guides update:
  - VXLAN (10.10: https://www.arubanetworks.com/techdocs/AOS-CX/10.10/PDF/vxlan.pdf)
- 10.09 Update: EVPN-VXLAN Multi-Fabric DCI
  - Youtube: https://www.youtube.com/watch?v=vpEaMDKjERM

aruba

a Hewlett Packard Enterprise company

Thank you

vincent.giles@hpe.com