AOS-CX 10.10 Update
June 2022

Campus Shared border VTEP between EVPN Fabrics

Vincent Giles
Technical Marketing Engineering

aruba
a Hewlett Packard Enterprise company

Agenda

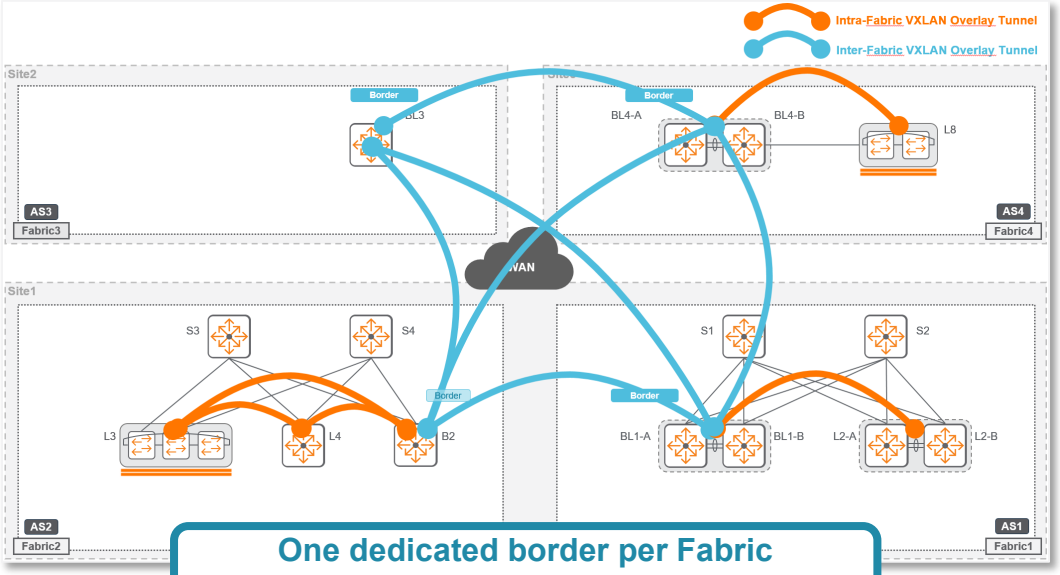
Definitions

Acronyms

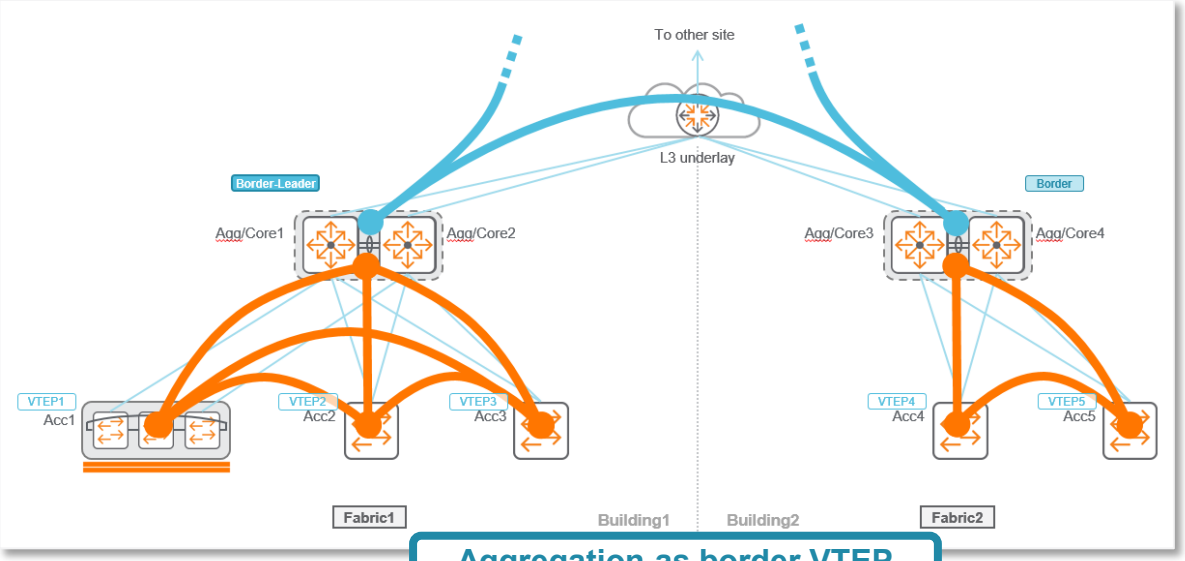
- VXLAN — **V**irtual e**X**tensible **LAN**
- VTEP — **V**XLAN **T**unnel **E**nd **P**oint
- VNI — **V**XLAN **N**etwork **I**dentifier
- L2VNI — **L**ayer2 **V**XLAN **N**etwork **I**dentifier (to extend L2 traffic)
- L3VNI — **L**ayer3 **V**XLAN **N**etwork **I**dentifier (to send routed traffic)
- 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

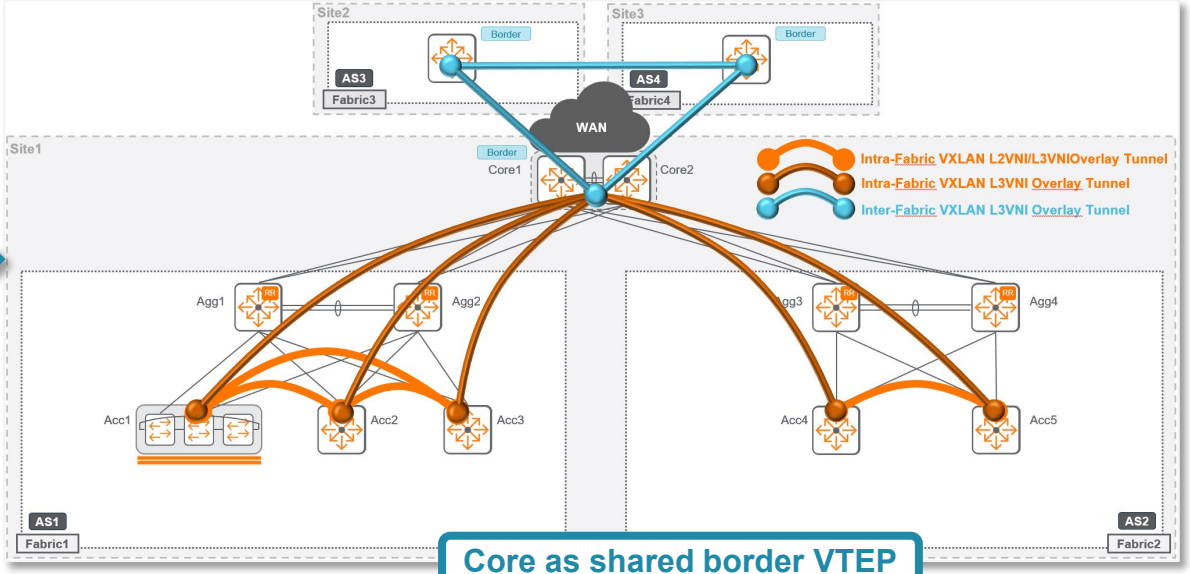
Datacenter / Campus



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



Aggregation as border VTEP



Core as shared border VTEP

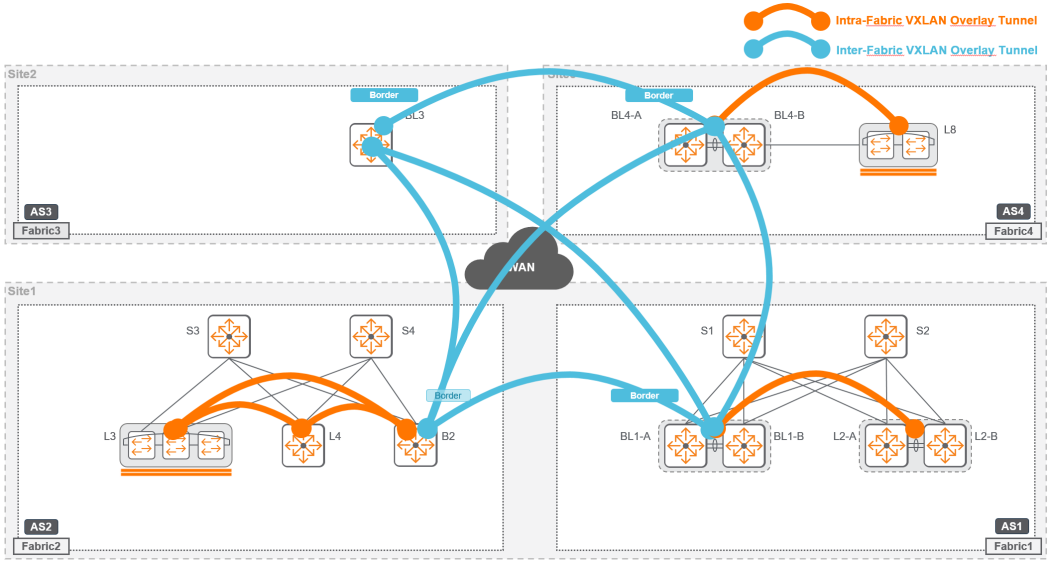
Overview

Campus shared border VTEP

10.09 versus 10.10

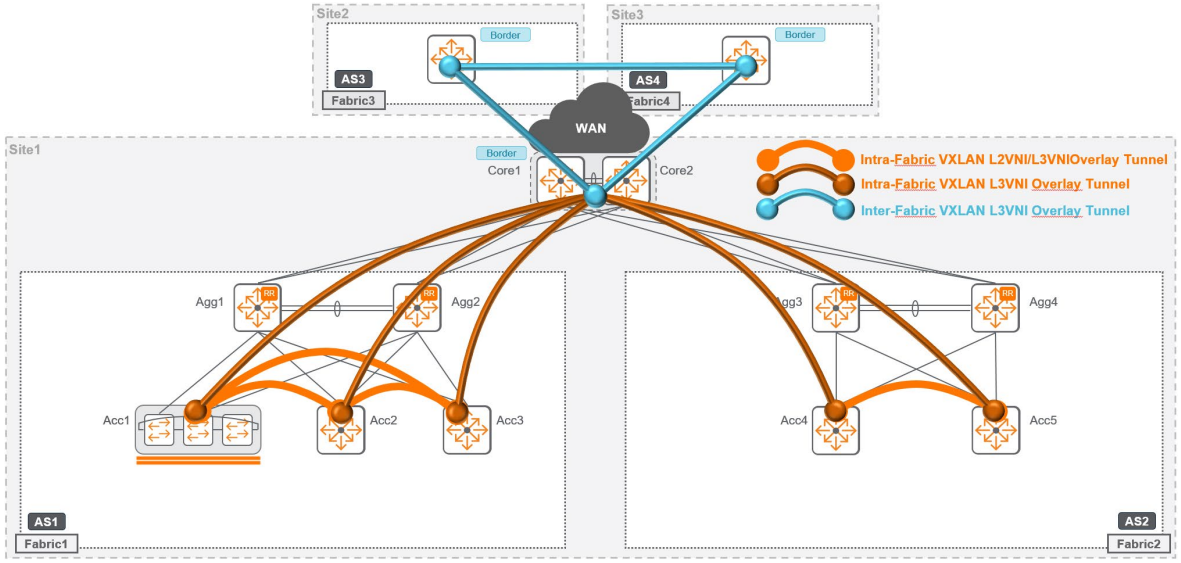
AOS-CX 10.09

- Not supported.
- DC use-case: One border per Fabric (L2VNI/L3VNI)



AOS-CX 10.10

- No new command or feature.
- Campus solution validated and supported (L3VNI only).



Use Cases

Campus use-case

L3VNI only to border

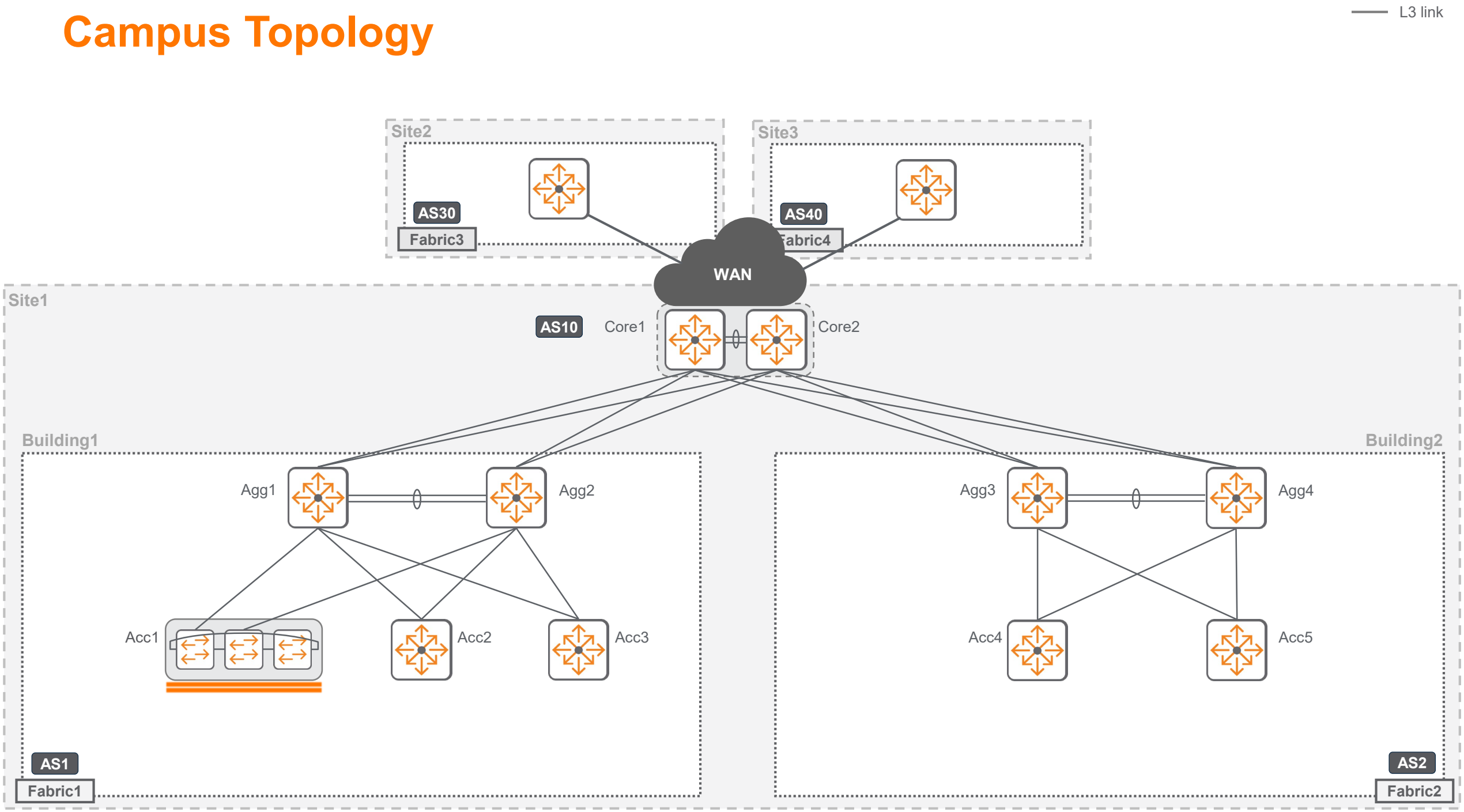
- Unlike datacenter with lot of stretched subnets, the main campus use-case requires less VLAN extension between fabrics.
- AOSCX-10.10 scope is limited to L3VNI, no L2VNI outside the Fabric.
  It is due to the current configuration method of the VXLAN tunnel broadcast domain:

```
dyn-vxlan-tunnel-bridging-mode ibgp-ebgp
```

- A new configuration knob, independent of iBGP versus eBGP EVPN-VXLAN tunnel category is investigated for future release.
- If any VLAN must be extended between fabrics, a dedicated border VTEP must be used.

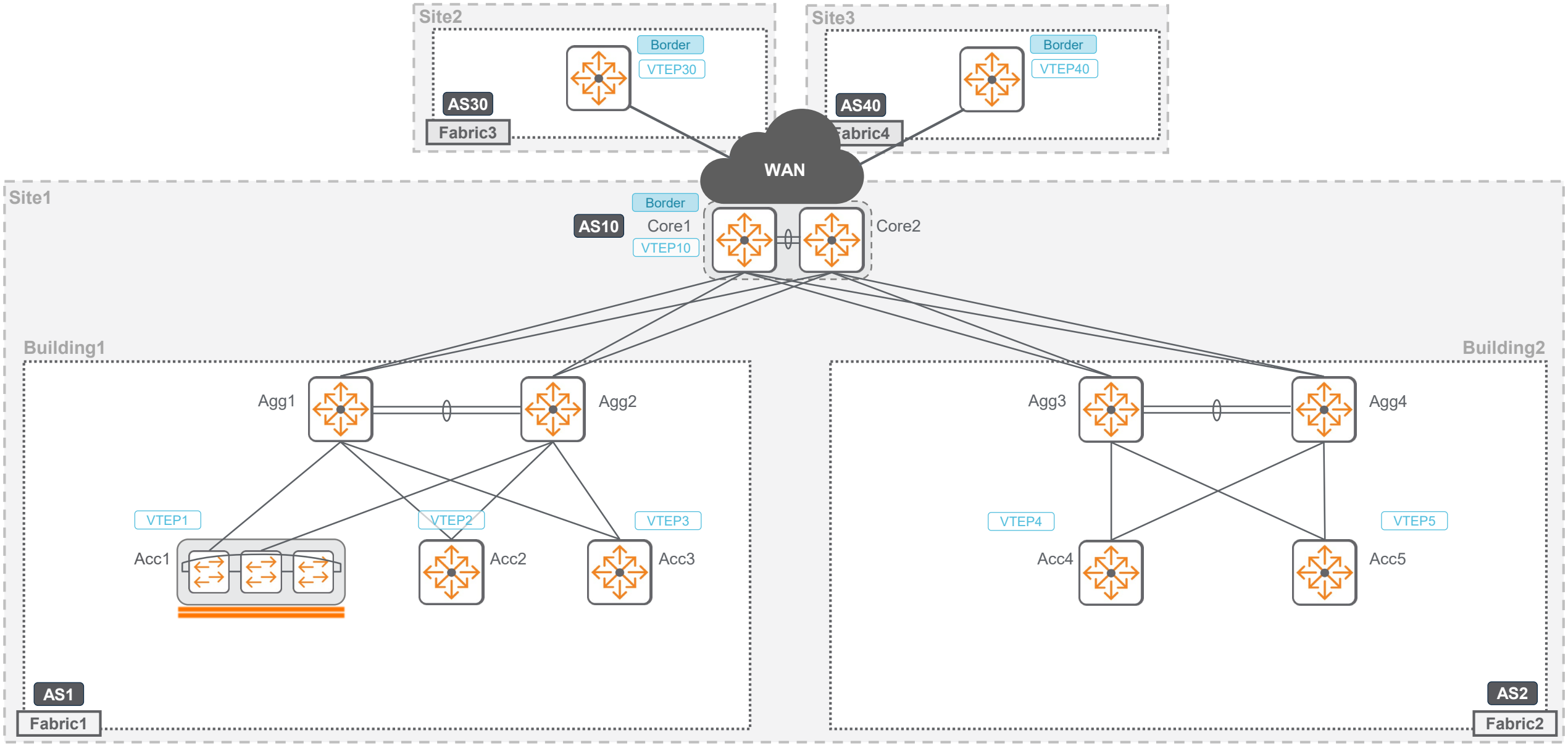
Stretched VLAN and stretched subnet are used interchangeably in this presentation.

Campus Topology



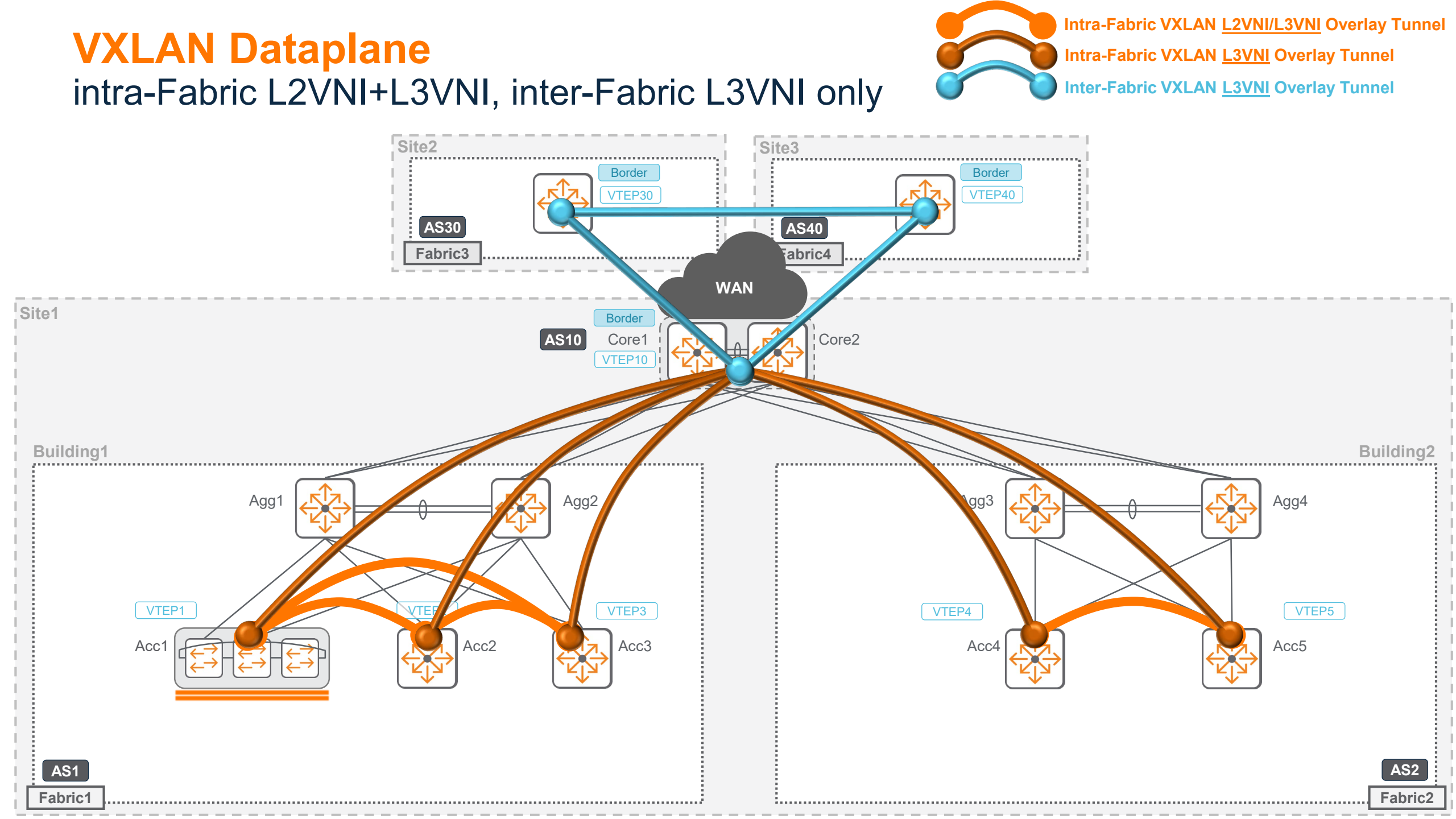
Campus use-case

Topology / VTEPs



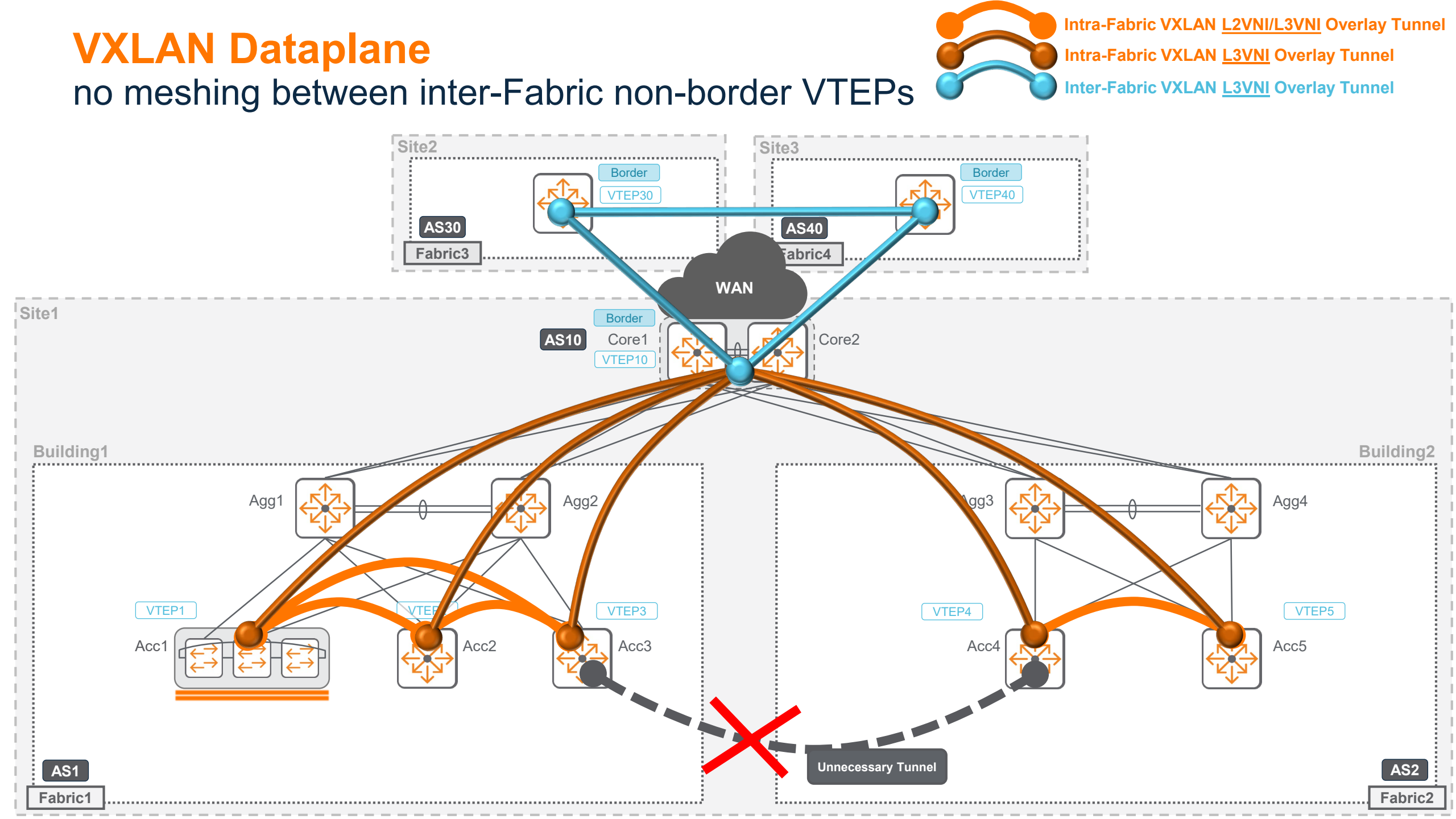
VXLAN Dataplane

intra-Fabric L2VNI+L3VNI, inter-Fabric L3VNI only



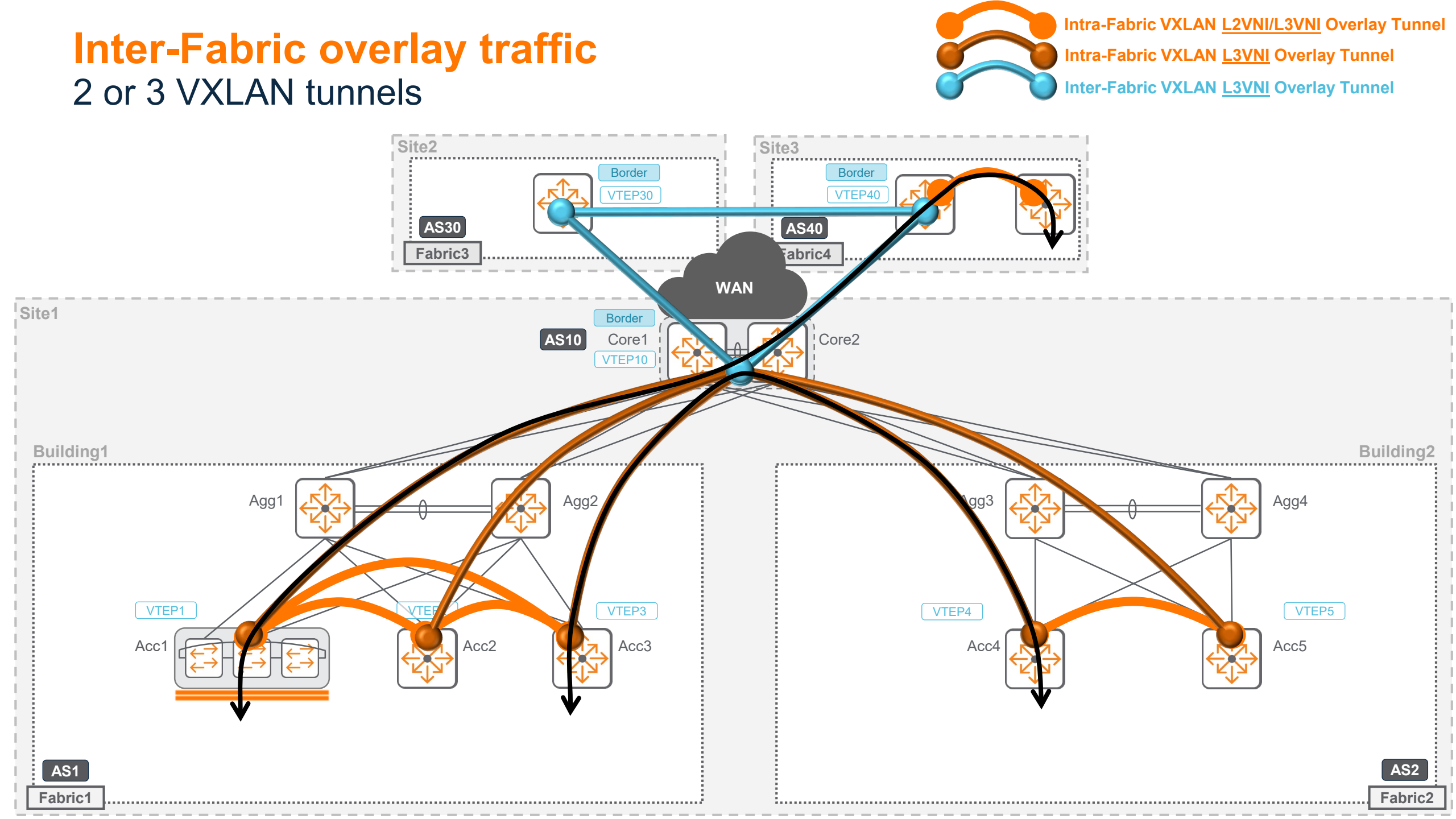
VXLAN Dataplane

no meshing between inter-Fabric non-border VTEPs



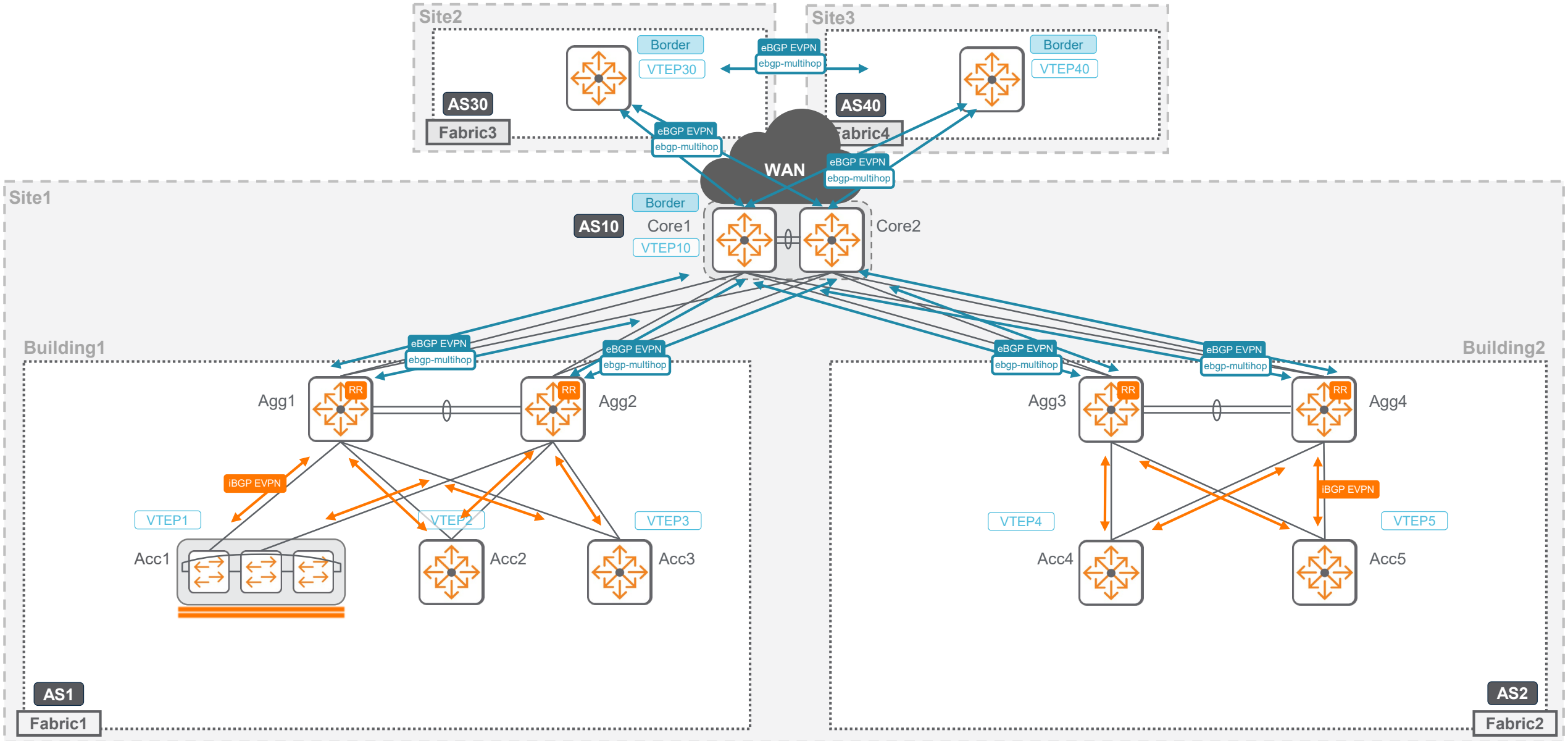
Inter-Fabric overlay traffic

2 or 3 VXLAN tunnels



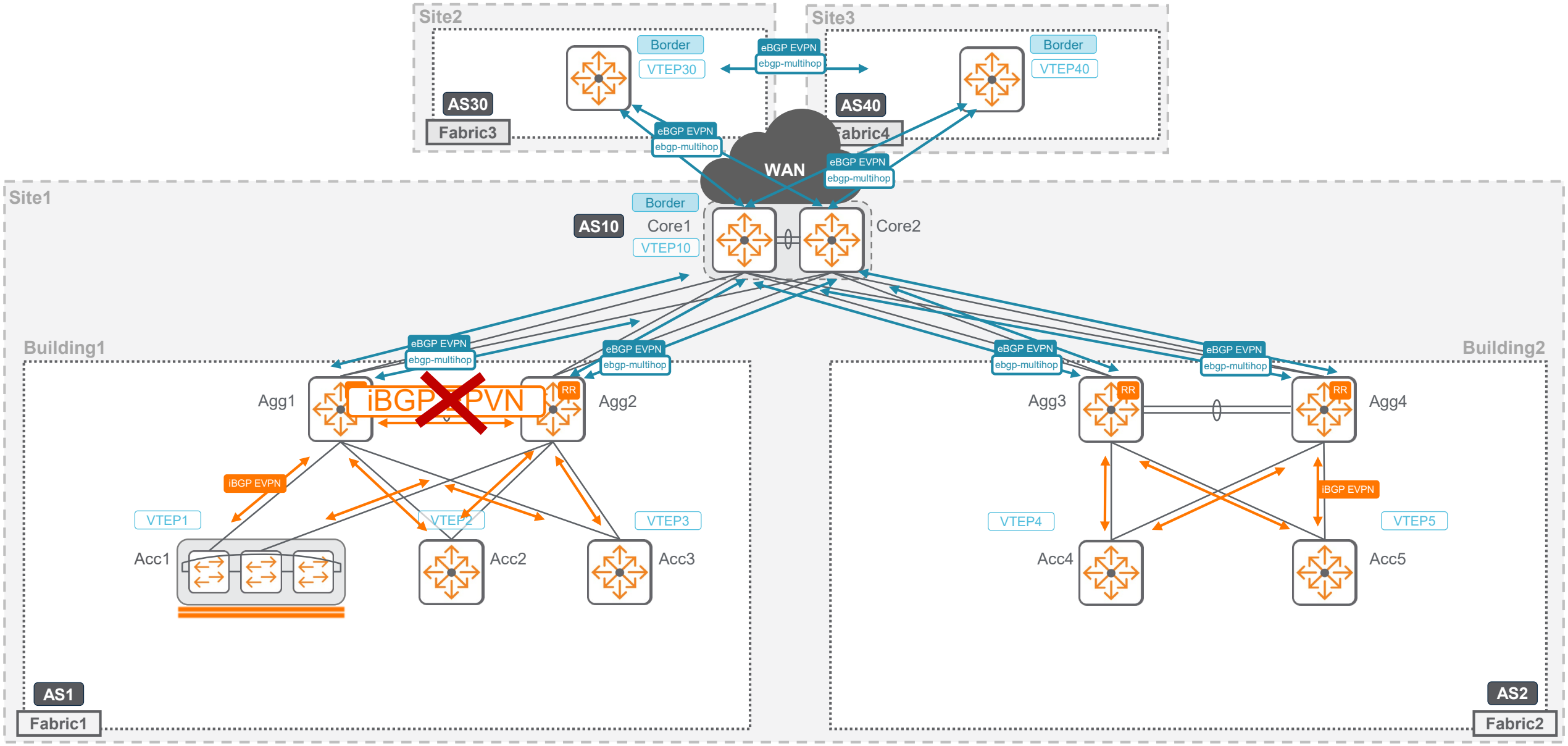
MP-BGP EVPN iBGP / eBGP sessions

intra-Fabric L2VNI+L3VNI, inter-Fabric L3VNI only



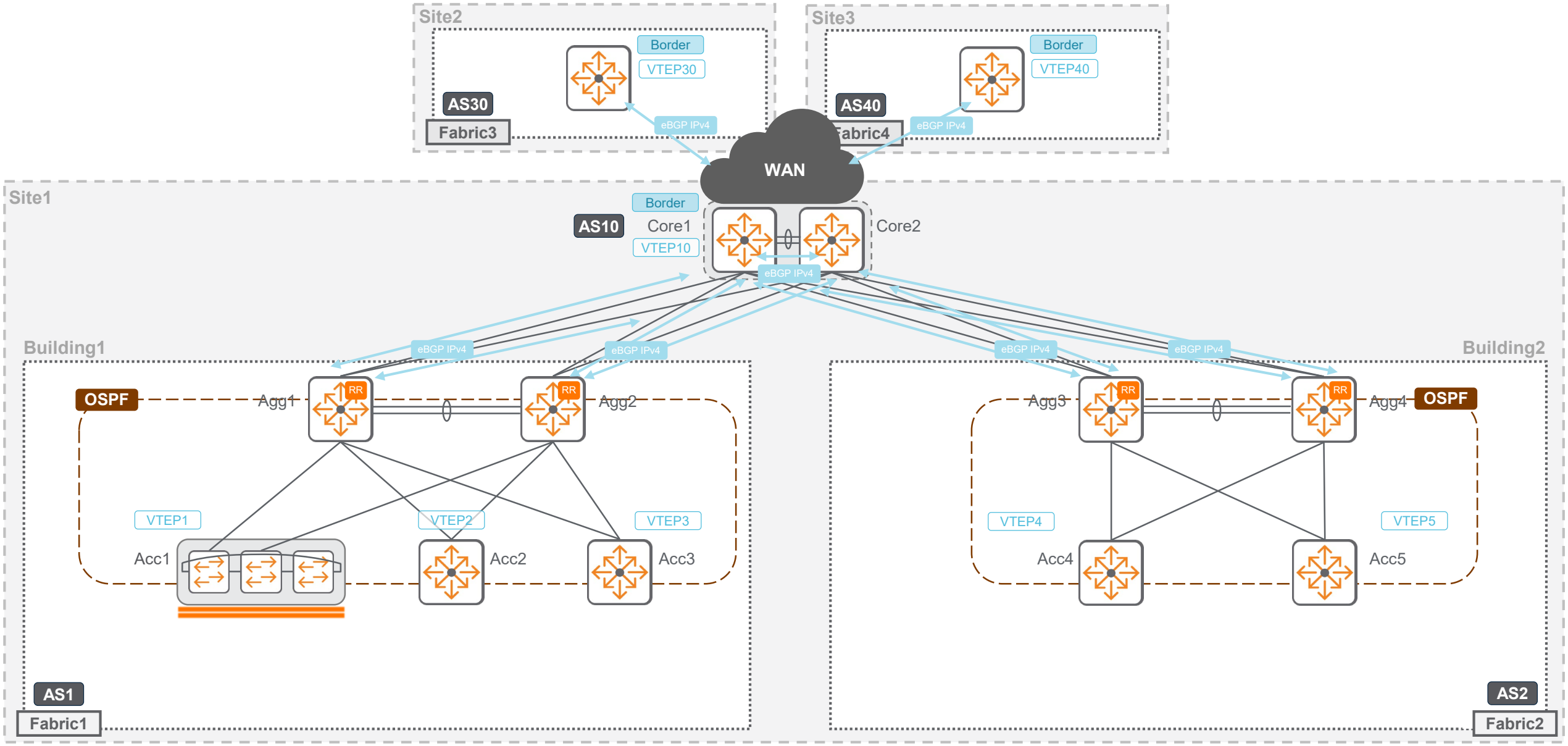
Route-Reflector: no shared BGP cluster-id

No intra-fabric iBGP EVPN session between RR



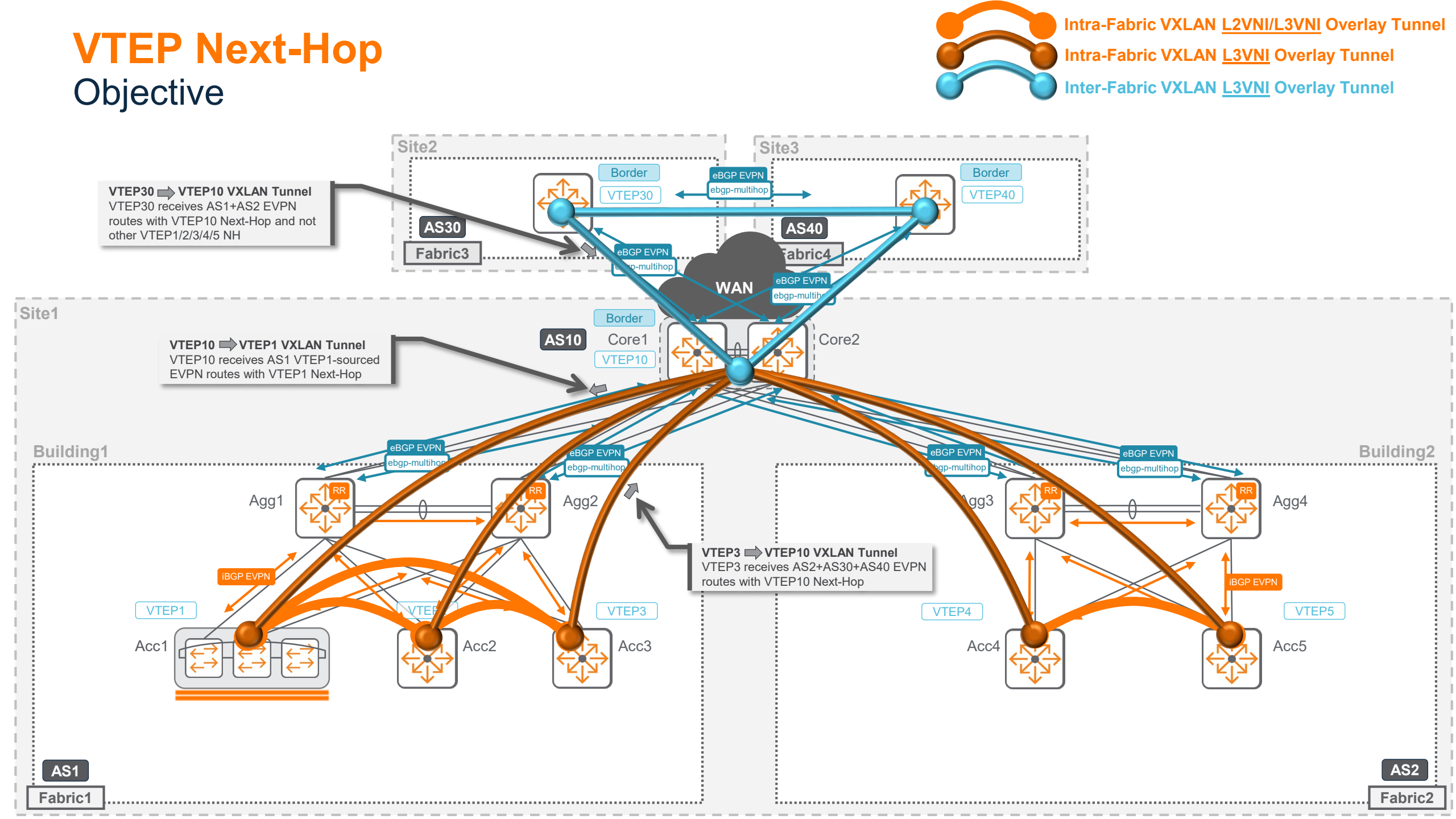
Underlay Control-Plane: eBGP IPv4 + OSPF

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



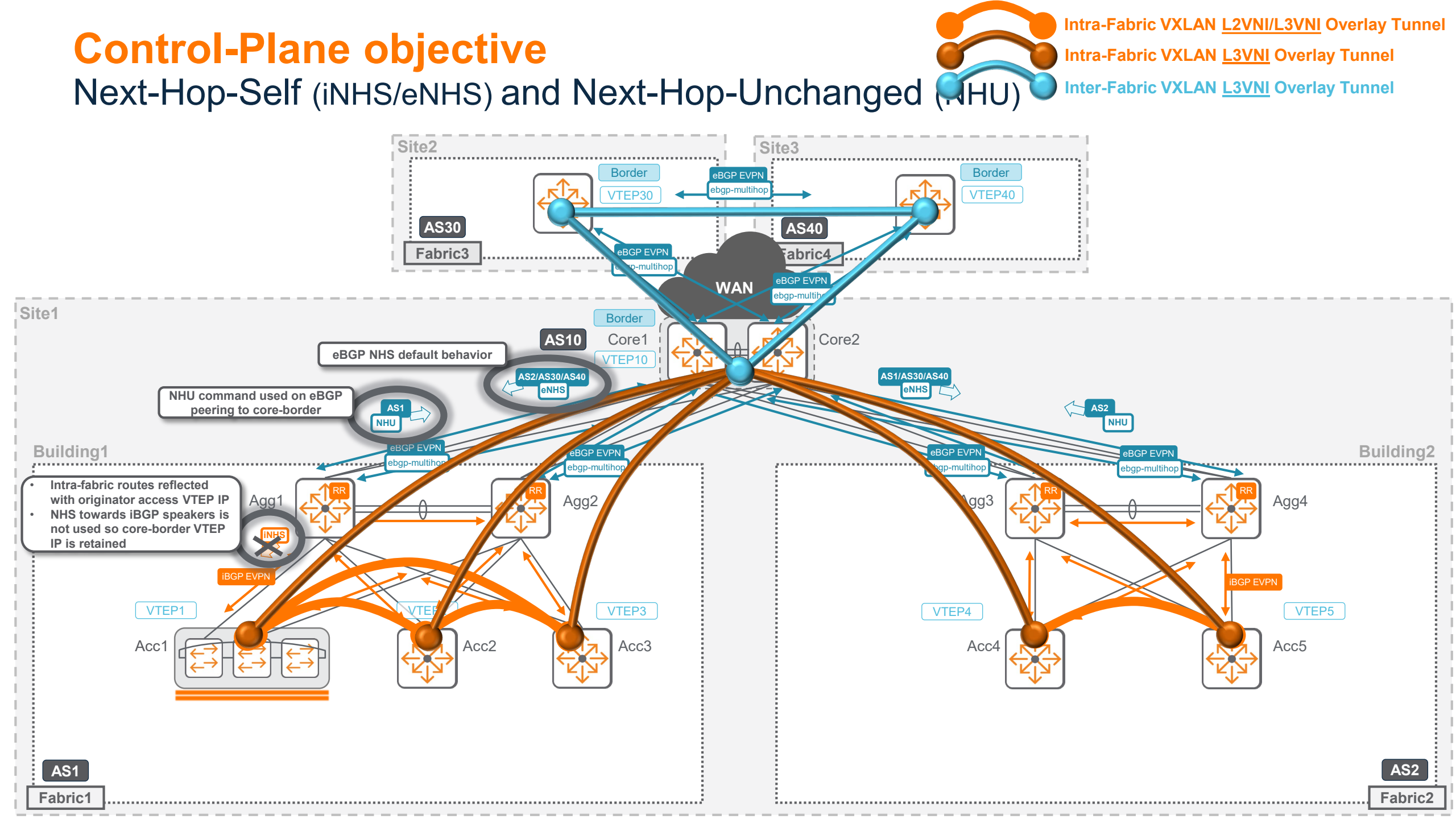
VTEP Next-Hop

Objective



Control-Plane objective

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



Details / Caveats

Campus Shared Border VTEP

Details / Caveats

- No new command. This is a solution validation for AOS-CX 10.10.
- Support for both IPv4 and IPv6* unicast overlay.
- In legacy architecture, Aggregation layer uses VSX for MCLAG and then core layer is generally pure L3 without need of VSX.
  Here, in contrast, Aggregation layer is purely L3 underlay, and the core layer requires VSX for VTEP HA.
  (VSX can be used as well in aggregation layer with SVI underlay instead of ROP.)
- Aggregation devices are interconnected for IGP continuity and offer flexibility for access switch uplinks.
- Route-Reflector function is hosted at the Aggregation layer.
- L2VNI related EVPN routes are exchanged even if the solution is limited to L3VNI. These routes will not get programmed in ASIC on remote VTEPs, but may consume some additional control-plane software resources.
- EVPN Type-3 routes optimization is not needed and not performed.
  If VLAN_ID and L2VNI values are re-used among different fabrics for intra-fabric VLAN extension *(reminder: inter-fabric VLAN extension is not supported)*, ensure that:
  - different EVPN VLAN Route-Target values are used per fabric: route-target value is local to the fabric for local-fabric VLAN stretching, and same route-target value is <u>not re-used</u> in other fabrics.
  - or alternatively, different L2VNI values are used among fabrics, if route-target values can not be different (very unlikely).

** 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.*

10.10 Platform Support

Use-case support of Shared Border VTEP between multiple EVPN domains

Platform	4100 6000 6100	6200	6300	6400 (v1/v2)	8320	8325	8360 (v1/v2)	8400	10000	Simulator
Shared border VTEP between EVPN domains	No	No	No	**Yes**	No	**Yes**	**Yes**	No	No	No

Shared Border VTEP - 10.10 Validated Multi-Dimensional Scale

	Border VTEP		
	8325	8360	6400
HW profile	Leaf	Agg-Leaf	Default
VTEPs per Fabric (standalone or VSX logical VTEP pair)	64	64	64
Number of Fabrics sharing common core VTEP	32	32	32
L3 routes across all VRFs and all sites (including host routes)	23K dual-stack	23K dual-stack	23K dual-stack
Overlay hosts (MAC / ARP) within the Fabric on access VTEP	8K dual-stack	8K dual-stack	8K dual-stack
Overlay hosts (MAC / ARP) across multiple Fabrics	Not supported (L3VNI only)	Not supported (L3VNI only)	Not supported (L3VNI only)
VLANs local to the Fabric	256	256	256
Stretched VLANs among all Fabrics	Not supported (L3VNI only)	Not supported (L3VNI only)	Not supported (L3VNI only)
VRFs shared among all Fabrics	32	32	32

Configuration

Configuration - Campus Shared Border VTEP

On Aggregation Route-Reflector: next-hop-unchanged for EVPN eBGP peering to core

- On Aggregation layer Route-Reflector

```
router bgp <asn>
...
address-family l2vpn evpn
   neighbor core-border send-community both
   neighbor core-border next-hop-unchanged
...
   neighbor leaf route-reflector-client
   neighbor leaf send-community both
   neighbor leaf next-hop-self
```

The next-hop of iBGP access VTEPs must be retained so that, on the shared border VTEP, the VXLAN destination VTEPs are the access VTEPs (and not the RR)

next-hop-self command to iBGP access VTEP RR-clients MUST NOT be configured in order to retain eBGP shared border VTEP as the next-hop (i.e on access VTEP, VXLAN tunnel destination = border, and not the RR)

- For configuration, please refer to 10.09 EVPN-VXLAN multi-fabric section of VXLAN user-guide.

Best Practices

Campus Shared Border VTEP

Best practices

- Shared Border VTEP should be a pair of VSX nodes for better high-availability.
- Aggregation switch should be interconnected for OSPF continuity.
- In case of same VLAN-ID re-use among fabrics, use different EVPN VLAN route-target per fabric.
  (the local scope route-target as defined in 10.09 solution).
  Do not use global scope EVPN VLAN route-target.

Troubleshooting

- Refer to 10.09 TOI – Troubleshooting section

Campus shared border VTEP Troubleshooting

1. On Aggregation RR, check that next-hop-unchanged is properly configured for core-border peering.

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

2. On access VTEP, check that the next-hop VTEP IP of external fabric EVPN routes is retained as the core-border VTEP.
   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
```

3. On access VTEP, check RMAC of external fabric EVPN routes.

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

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