The bridge to possible

Cisco SD-Access Transit: Advanced Design Principles

Jerome Dolphin, Technical Marketing Engineer
CCIE#17805 (R&S, SEC), CCDE#2013::3

BRKENS-2816

cisco Live!

#CiscoLive

Cisco Webex App

https://ciscolive.ciscoevents.com/ciscolivebot/#BRKENS-2816

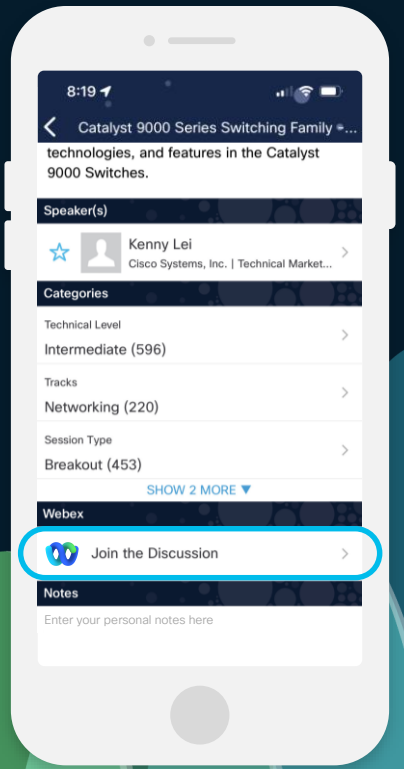
Questions?

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

How

1. Find this session in the Cisco Live Mobile App
2. Click “Join the Discussion”
3. Install the Webex App or go directly to the Webex space
4. Enter messages/questions in the Webex space

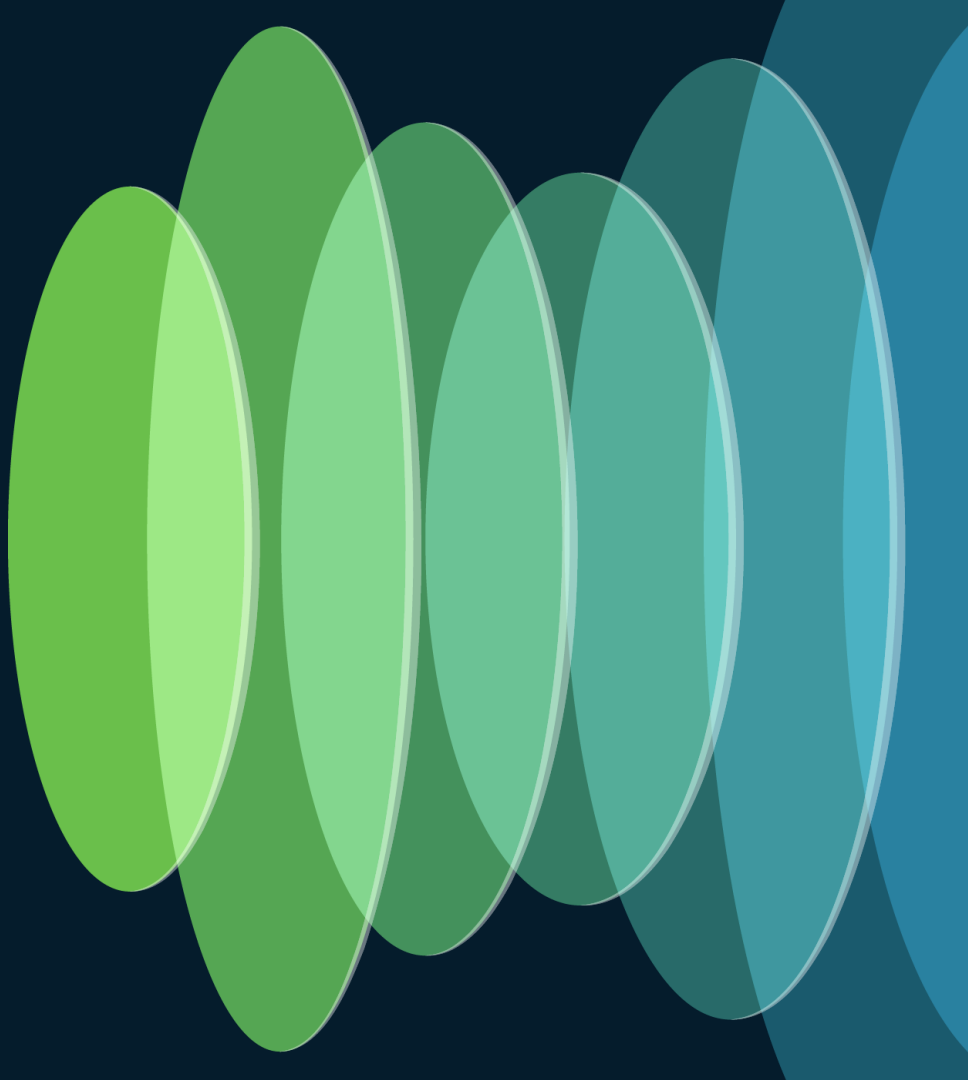
Webex spaces will be moderated by the speaker until June 7, 2024.



Agenda

- Introduction
- Underlay MTU
- Overlay Route Propagation
- Overlay Routing Rules and Features
- Design Myths
- Shared SD-Access Transit
- Demo
- Conclusion

Introduction

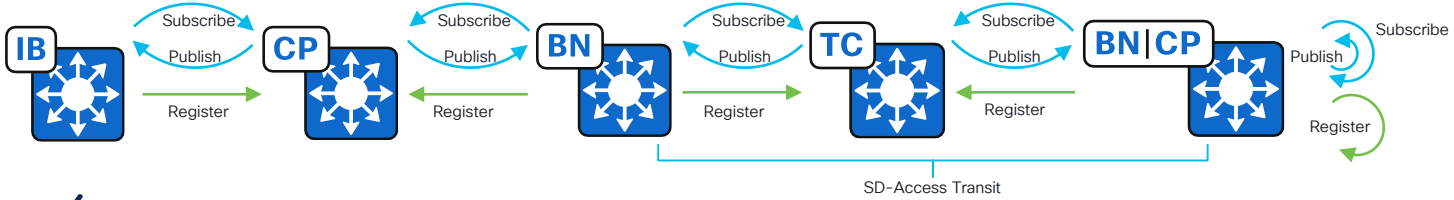


Orientation

- Please note the abstract for this presentation said “Attendees are advised to be familiar with the introductory material discussed in BRKENS-2810 and BRKENS-2811 before attending this session”
- For those wanting to study LISP later:
  - BRKENS-2828 – LISP Architecture Evolution – New Capabilities Enabling SD-Access
- There are additional slides in the PDF accompanying this presentation marked with a “For your Reference” symbol: 

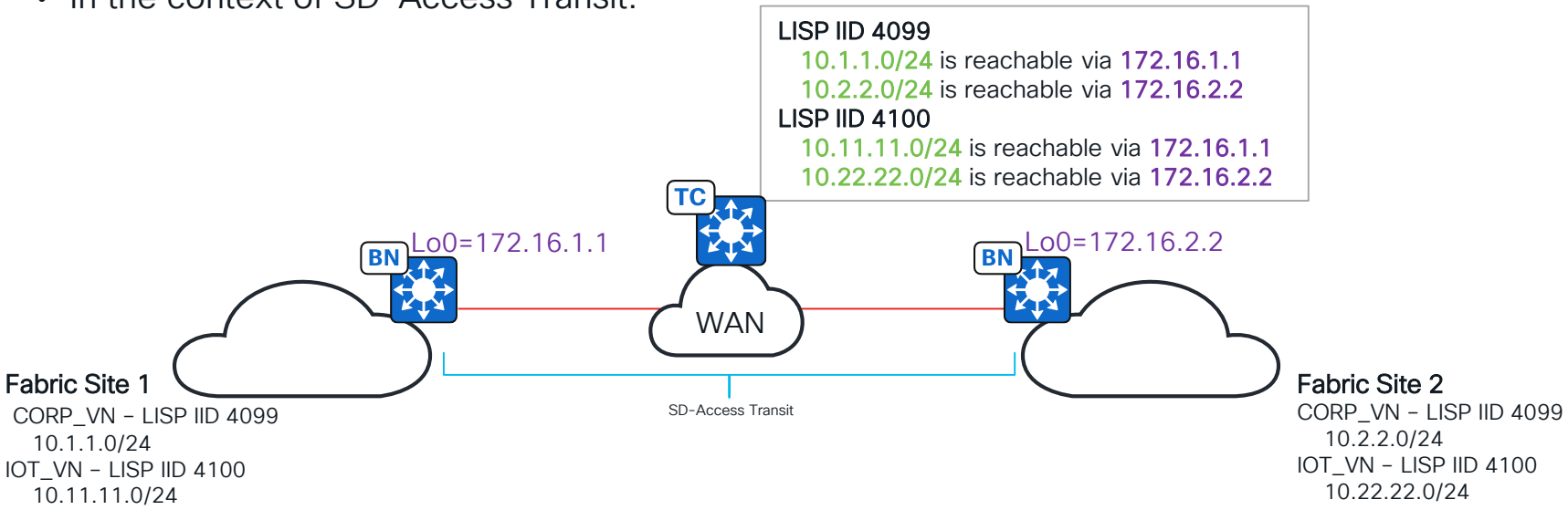
The Context is LISP Publish / Subscribe

- This presentation focuses on the LISP Pub/Sub control plane architecture. Ideas presented may not apply to LISP/BGP.
- Border Nodes:
  - Subscribe to IPv4 and IPv6 reachability information.
  - Register Endpoint IDs.
  - May send map-requests, if there's an Extranet Policy.
- Control Plane Nodes
  - Receive and store Endpoint ID registrations.
  - Publish reachability information to subscribers.
  - May resolve Border Node map-requests, if there's an Extranet Policy.



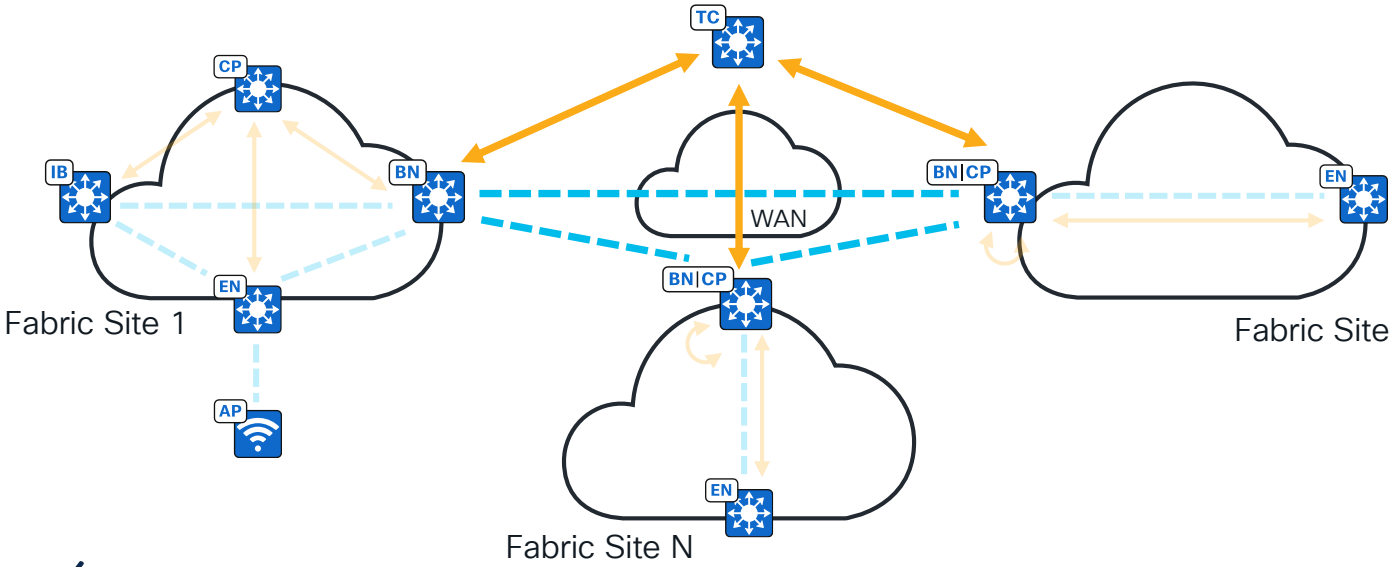
LISP Fundamentals

- LISP is an acronym for Locator ID Separation Protocol.
- An EID (Endpoint ID) is mapped to an RLOC (Routing Locator).
- An EID can be many things: MAC address, IPv4/IPv6 host routes, IPv4/IPv6 summary routes, or network services like default ETR (default route).
- In the context of SD-Access Transit:



What is an SD-Access Transit?

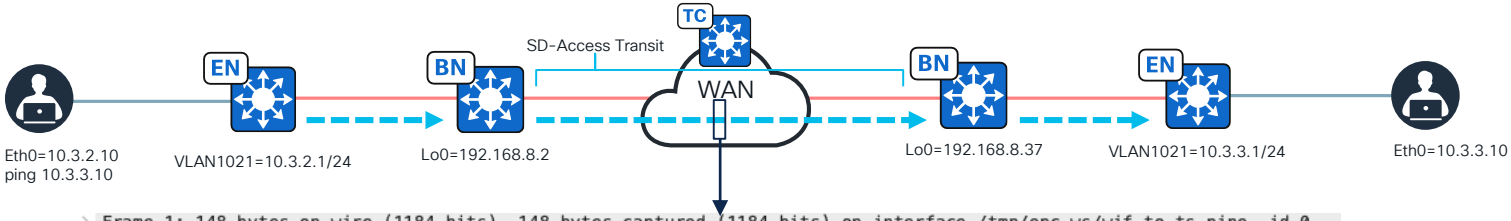
- Catalyst Center automated LISP/VXLAN overlays between Fabric Sites that:
  1. Route between Layer 3 Virtual Networks (L3VN) in different Fabric Sites.
  2. Preserve macro-segmentation context (L3VN).
  3. Preserve micro-segmentation context (SGT).



SD-Access Transit Control Plane Node

- Peers LISP with all External Border Nodes connected to the SD-Access Transit.
- Must be a dedicated device, not co-located with other fabric roles.
- 1-4 Transit Control Plane Nodes (TC) per Pub/Sub SD-Access Transit.
- Tracks Fabric Site summary routes and imported routes, does not track endpoint host routes.
  - Example: 10 Fabric Sites with an aggregate of 100 IPv4 Anycast Gateways and 100,000 IPv4 endpoints = 100 summary routes registered to TC.
- Does not need to be in the traffic forwarding path. TC is a LISP database, not a packet forwarder.
- Recommended to use Catalyst 9000 switch from SD-Access Compatibility Matrix as TC. Switches offer maximum feature flexibility. Routers cannot support Extranet Policy.

SD-Access Transit VXLAN Breakdown




```
Frame 1: 148 bytes on wire (1184 bits), 148 bytes captured (1184 bits) on interface /tmp/epc_ws/wif_to_ts_pipe, id 0
Ethernet II, Src: 00:00:00_00:00:00 (00:00:00:00:00:00), Dst: 00:00:00_00:00:00 (00:00:00:00:00:00)
Internet Protocol Version 4, Src: 192.168.8.2, Dst: 192.168.8.37
   0100 .... = Version: 4
   .... 0101 = Header Length: 20 bytes (5)
   Differentiated Services Field: 0x00 (DSCP: CS0, ECN: Not-ECT)
   Total Length: 134
   Identification: 0x0232 (562)
   010. .... = Flags: 0x2, Don't fragment
      0... .... = Reserved bit: Not set
      .1.. .... = Don't fragment: Set
      ..0. .... = More fragments: Not set
   ...0 0000 0000 0000 = Fragment Offset: 0
   Time to Live: 62
   Protocol: UDP (17)
   Header Checksum: 0xa8bd [validation disabled]
   [Header checksum status: Unverified]
   Source Address: 192.168.8.2
   Destination Address: 192.168.8.37
User Datagram Protocol, Src Port: 65483, Dst Port: 4789
Virtual eXtensible Local Area Network
   Flags: 0x8800, GBP Extension, VXLAN Network ID (VNI)
   Group Policy ID: 19
   VXLAN Network Identifier (VNI): 4101
   Reserved: 0
Ethernet II, Src: 00:00:00_00:39:85 (00:00:00:00:39:85), Dst: ba:25:cd:f4:ad:38 (ba:25:cd:f4:ad:38)
Internet Protocol Version 4, Src: 10.3.2.10, Dst: 10.3.3.10
Internet Control Message Protocol
```

+14B

+20B

+8B

+8B

Don’t Fragment = True

VXLAN Source IP

VXLAN Destination IP (RLOC)

SGT ID

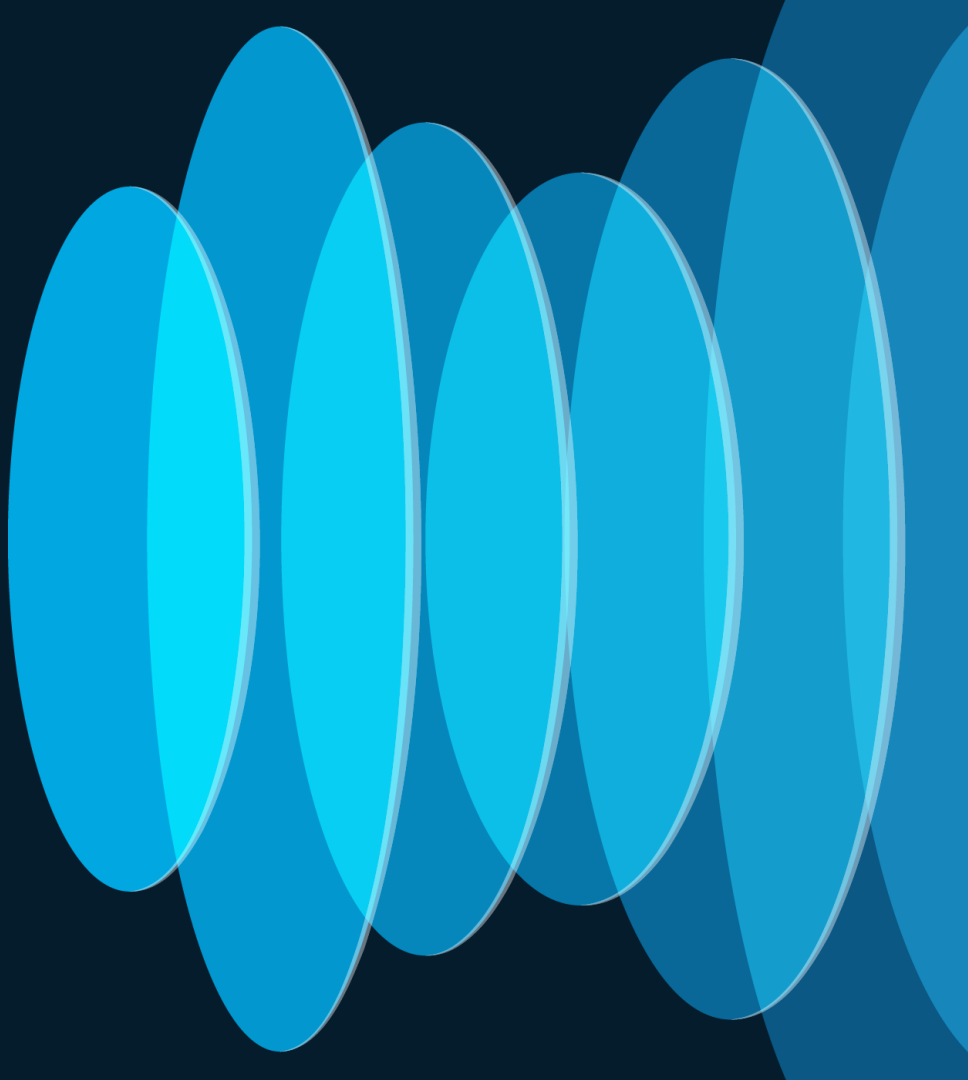
Layer 3 Virtual Network

Original Packet Source IP

Original Packet Destination IP

Original Packet Payload

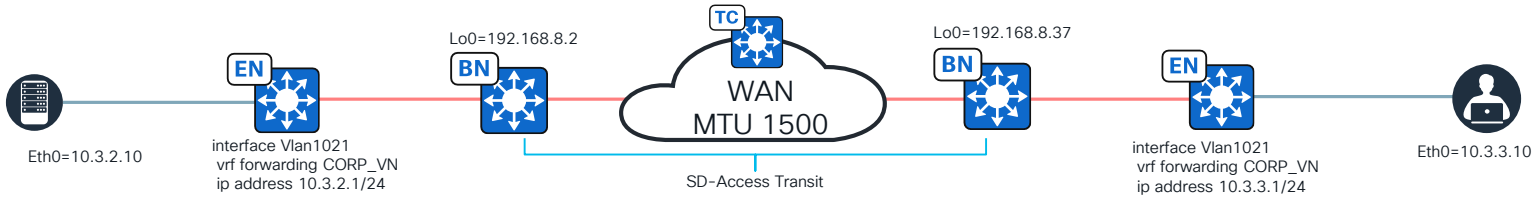
Underlay MTU



Underlay MTU

- All VXLAN has DF bit set as per RFC 7348.
- Underlay (GRT) MTU must accommodate VXLAN encapsulation overhead.
- VXLAN packets that exceed underlay MTU will be silently dropped from the perspective of the originating endpoint.

Consequence of Insufficient Underlay MTU



1. Get somefile.iso from 10.3.2.10

2. Send first fragment of requested file in a 1500B packet to 10.3.3.10

3. Lookup packet destination in CORP_VN
4. Encapsulate in VLXAN, send 1550B packet (DF=1) to 192.168.8.2 in underlay

5. Decapsulate VXLAN
6. Lookup original packet destination in CORP_VN
7. Encapsulate in VXLAN (DF=1), send 1550B packet to 192.168.8.37 in underlay

8. Packet exceeds WAN MTU. DF is bit set, send ICMP T3C4 (Destination Unreachable, Fragmentation Required) to 192.168.8.2

9. 192.168.8.2 receives ICMP T3C4 in underlay. The originating endpoint (10.3.2.10 in CORP_VN) not reachable in underlay. ICMP T3C4 dropped.

Don’t Change the Ingress SVI MTU Configuration

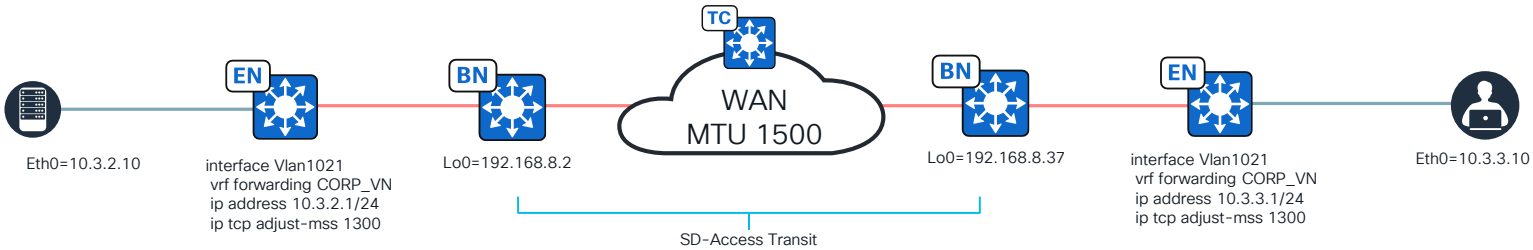
- Do not configure “ip mtu” on overlay ingress SVIs, it will not fragment oversized packets and it will not generate ICMP T3C4 (Destination Unreachable, Fragmentation Required) in response to endpoint Path MTU Discovery attempts.

```
!FABRIC_EDGE_NODE
!
interface Vlan1021
 description Configured from Cisco Catalyst Center
 mac-address 0000.0c9f.f8d5
 vrf forwarding CORP_VN
 ip address 10.3.2.1 255.255.255.0
 ip helper-address 172.31.24.31
 no ip redirects
 ip route-cache same-interface
 no lisp mobility liveness test
 lisp mobility IOT_VN_VLAN-IPV4
 ip mtu 1500       << NOT configured by Catalyst Center. Not helpful to add manually.
```

TCP MSS Adjustment

- Restricts TCP packet size to accommodate restrictive underlay MTUs.
- Automated by Catalyst Center 2.3.7 and later:
  1. On selected Anycast Gateways, for traffic to/from endpoints connected to fabric.
  2. On Border Node Layer 3 Handoff IP handoff interfaces, for traffic in/out of a Layer 3 Handoff.
- MSS Adjust can be templated if granularity is desirable, or version is < 2.3.7.
- If a packet crosses multiple interfaces with MSS Adjust configured, then the lowest value is selected. MSS Adjust does not increase MSS.
- MSS Adjust applies to ingress traffic on a given routed interface.

TCP MSS Adjustment



1. Get somefile.iso from 10.3.2.10
2. TCP MSS adjusted down to 1300

3. Send first fragment of requested file in TCP to 10.3.3.10.
Original Packet size ~1334B
(TCP <=1300B + IP 20B + Ethernet 14B)

3. Lookup packet destination
4. Encapsulate in VLXAN (+50B), send ~1384B packet to 192.168.8.2

5. Decapsulate VXLAN
6. Lookup original packet destination in CORP_VN
7. Encapsulate in VXLAN (DF=1), send ~1384B packet to 192.168.8.37 in underlay

8. Standard SD-Access forwarding

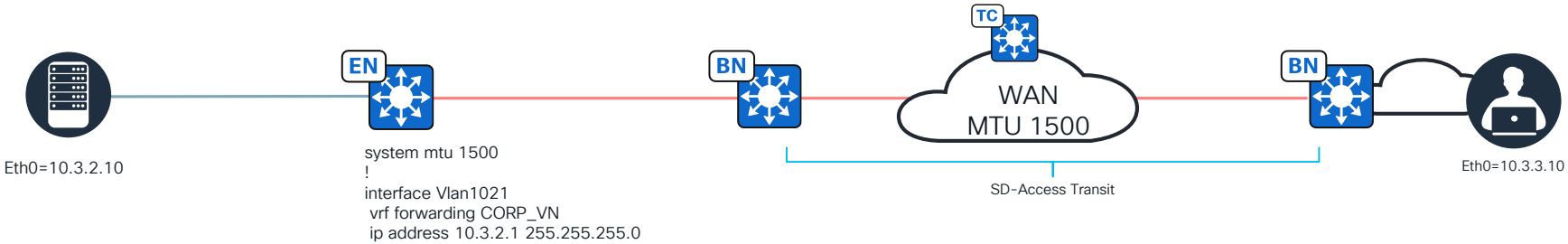
Path MTU Discovery

- PMTUD is implemented by the endpoint to “Discover” the lowest MTU between source and destination.
- Applicable to all upper layer encapsulations (UDP, ICMP, TCP, etc.) riding in IP.
- The packet originator sets IP header DF=1 (Do Not Fragment).
- The ingress SD-Access Fabric Node:
  - Calculates the VXLAN encapsulated packet size.
  - Sends ICMP T3C4 (Destination Unreachable, Fragmentation Required) to packet originator if encapsulated size exceeds Fabric Node egress MTU.
- Fabric Node egress MTU is defined by either a switch system MTU or IP MTU on L3 LISP interface.

```
!FABRIC_EDGE_NODE
!
system mtu 1500
```

```
!FABRIC_EDGE_NODE
!
interface LISP0.4099
 vrf forwarding CORP_VN
 ip mtu 1500
interface LISP0.4101
 vrf forwarding IOT_VN
 ip mtu 1500
```

Path MTU Discovery



1. Send 1500B packet with DF=1 packet to 10.3.3.10

2. Calculate encapsulated VXLAN packet size = 1500 + 50 = 1550
3. Encapsulated VXLAN packet size > system MTU, send ICMP T3C4 from 10.3.2.1 to 10.3.2.10

4. ICMP T3C4 received, reduce packet size and try again
5. Send smaller packet e.g., 1400B with DF=1 to 10.3.3.10

6. Calculate encapsulated VXLAN packet size = 1400 + 50 = 1450, OK to transmit
7. Encapsulate in VXLAN, send for 1450B packet towards destination
8. Standard SD-Access Transit forwarding

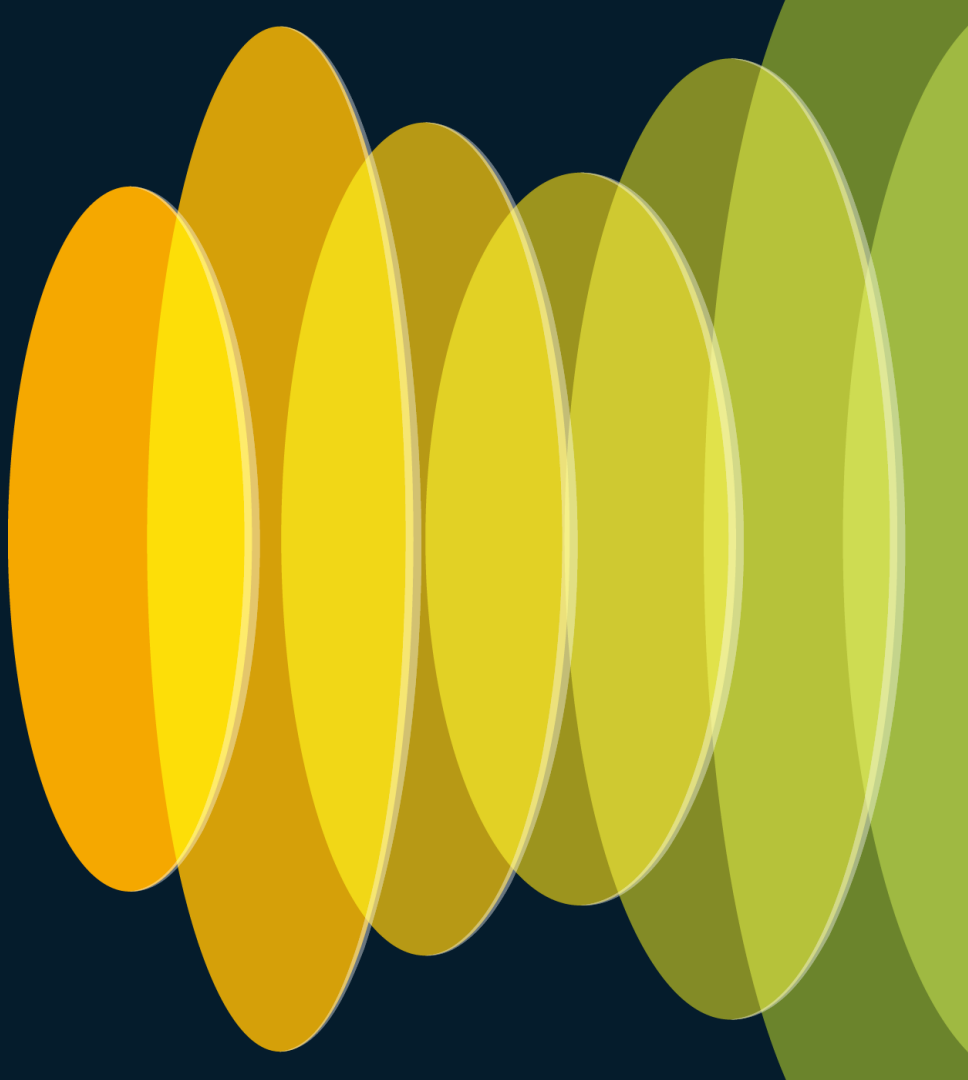
When to use MSS Adjust and PMTUD

- If a network is TCP heavy, then MSS Adjust may be sufficient.
- If network is UDP heavy, then ingress SD-Access switches may need to facilitate PMTUD via system mtu or L3 LISP interface ip mtu.
- If network has a mix or protocols, then both may be necessary.

Great, so the Minimum Underlay MTU is?

- It depends:
  1. On the endpoints. Do they support Path MTU Discovery?
  2. On the applications. Do they generate large packets that exceed the transport MTU?
  3. On the SD-Access fabric configuration. Adjust MSS, System MTU, L3 LISP MTU.
- Theoretically an underlay MTU of 1500B could work with Adjust MSS and PMTUD. There are some basic production deployments with 1500B underlay today.
- Suggested approach:
  - If underlay MTU >=1600B then OK to proceed.
  - If underlay MTU <1600B then pilot, or consult with SD-Access SME. Please share interesting questions or observations with jedolphi@cisco.com or your local Cisco representative.

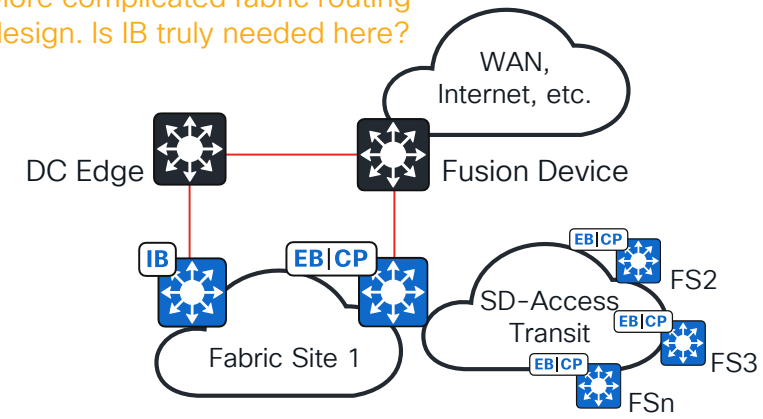
Overlay Route Propagation



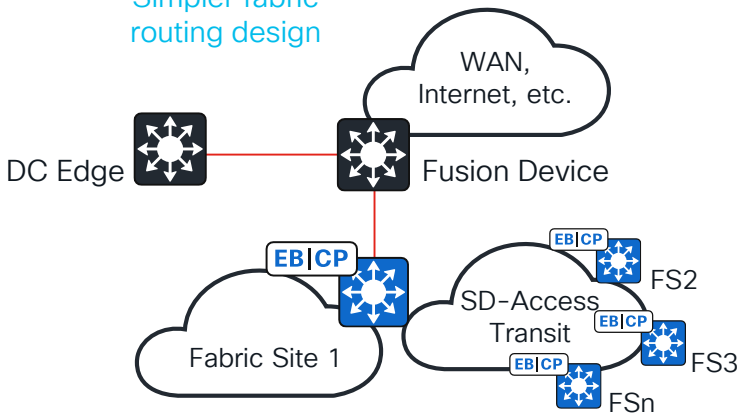
Design for Routing Simplicity

- Make the routing designs as simple as practicable, otherwise you may end up with my hairstyle.

More complicated fabric routing design. Is IB truly needed here?



Simpler fabric routing design



Route Propagation Within the Fabric Domain

Access Networks

Use Case

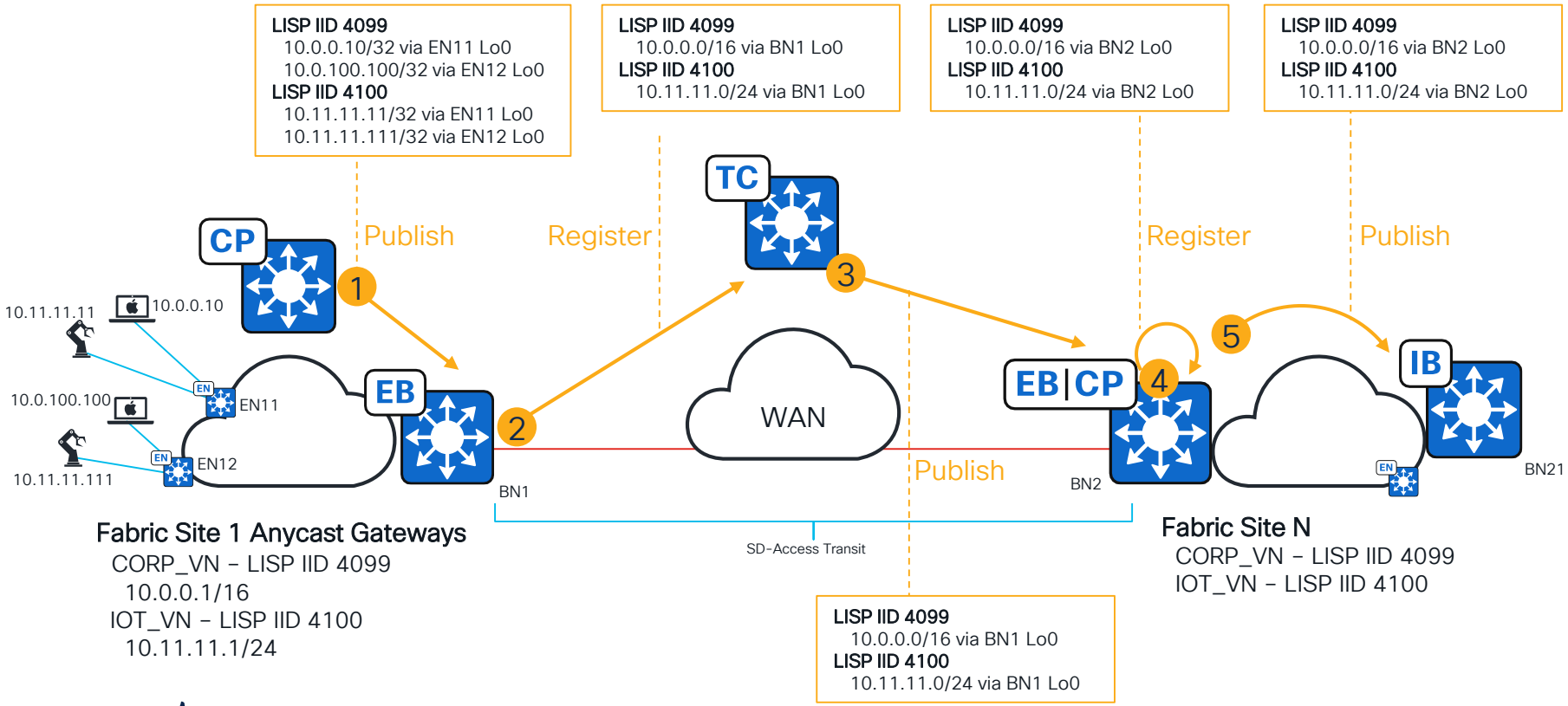
- Site1 endpoints connectivity to SiteN endpoints over SD-Access Transit.

How

- For each L3VN, Site1 Edge Nodes and Fabric WLCs register Endpoint IDs (MAC, IPv4, IPv6) to Site1 CP.
- Site1 CP publishes IPv4 and IPv6 Endpoint IDs to Site1 External Border Nodes and Internal Border Nodes.
- Site1 External Border Nodes register Endpoint ID summary routes to SD-Access Transit CP.
- Transit CP publishes summary routes to SiteN External Border Nodes.
- SiteN External Border Nodes register summary routes to SiteN CP.
- SiteN CP publishes summary routes to SiteN Internal Border Nodes.

Route Propagation Within the Fabric Domain

Access Networks



Fabric Site 1 Anycast Gateways
CORP_VN – LISP IID 4099
10.0.0.1/16
IOT_VN – LISP IID 4100
10.11.11.1/24

Fabric Site N
CORP_VN – LISP IID 4099
IOT_VN – LISP IID 4100

Route Propagation Within the Fabric Domain

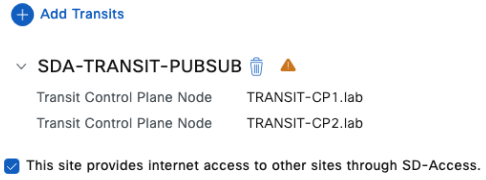
Default Routes

Use Case

- Make Site1 L3VN default route available within Site1 and optionally in SiteN over SD-Access Transit.

How

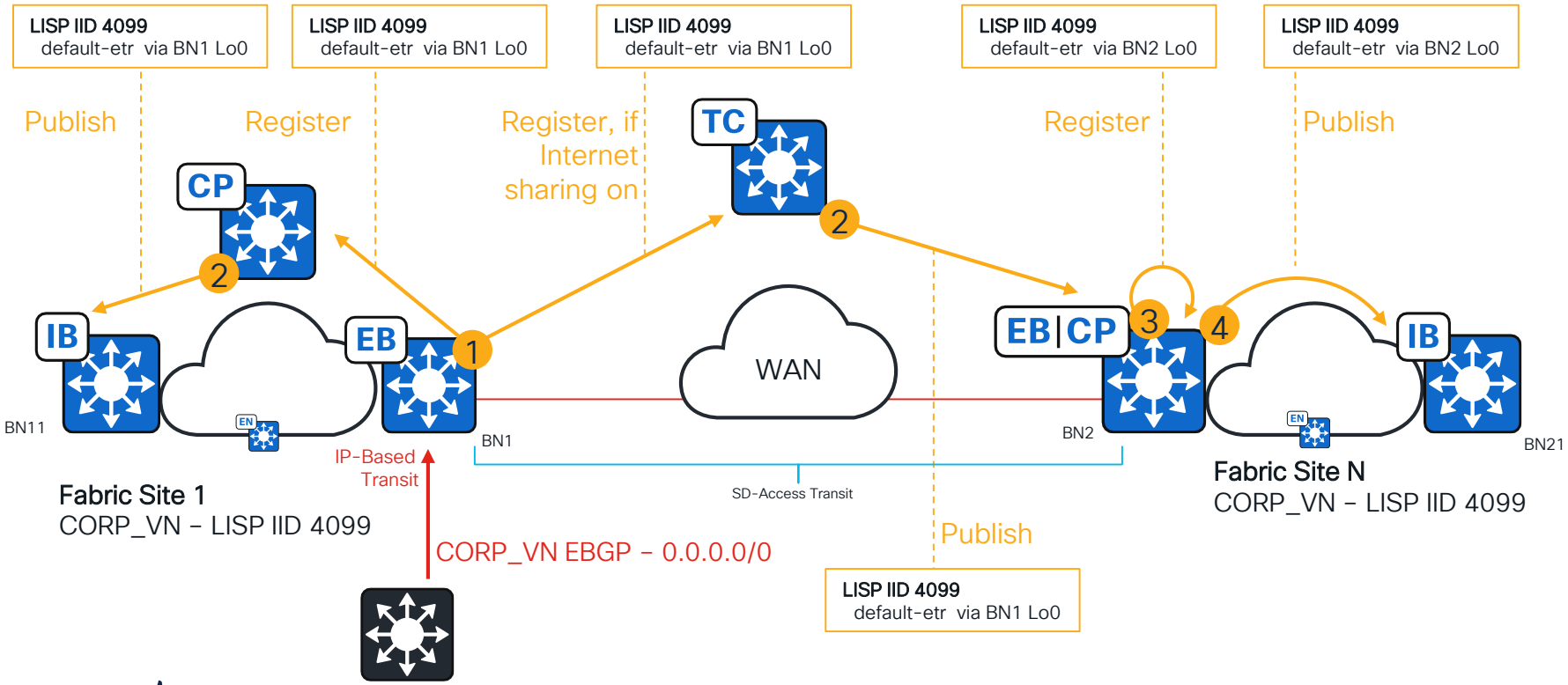
- For each Site1 External Border Node L3VN with 0/0 or ::/0 in RIB, default-etr is registered to the Site1 CP.
- If Internet (default route) sharing is enabled on the Site1 External Border Nodes, then default-etr is registered to Transit CP.



- Transit CP publishes default-etr to SiteN External Border Nodes.
- SiteN External Border Nodes register default-etr to SiteN CP.
- SiteN CP publishes default-etr to SiteN Internal Border Nodes.

Route Propagation Within the Fabric Domain

Default Routes



Route Propagation Within the Fabric Domain

Imported Routes

Use Case

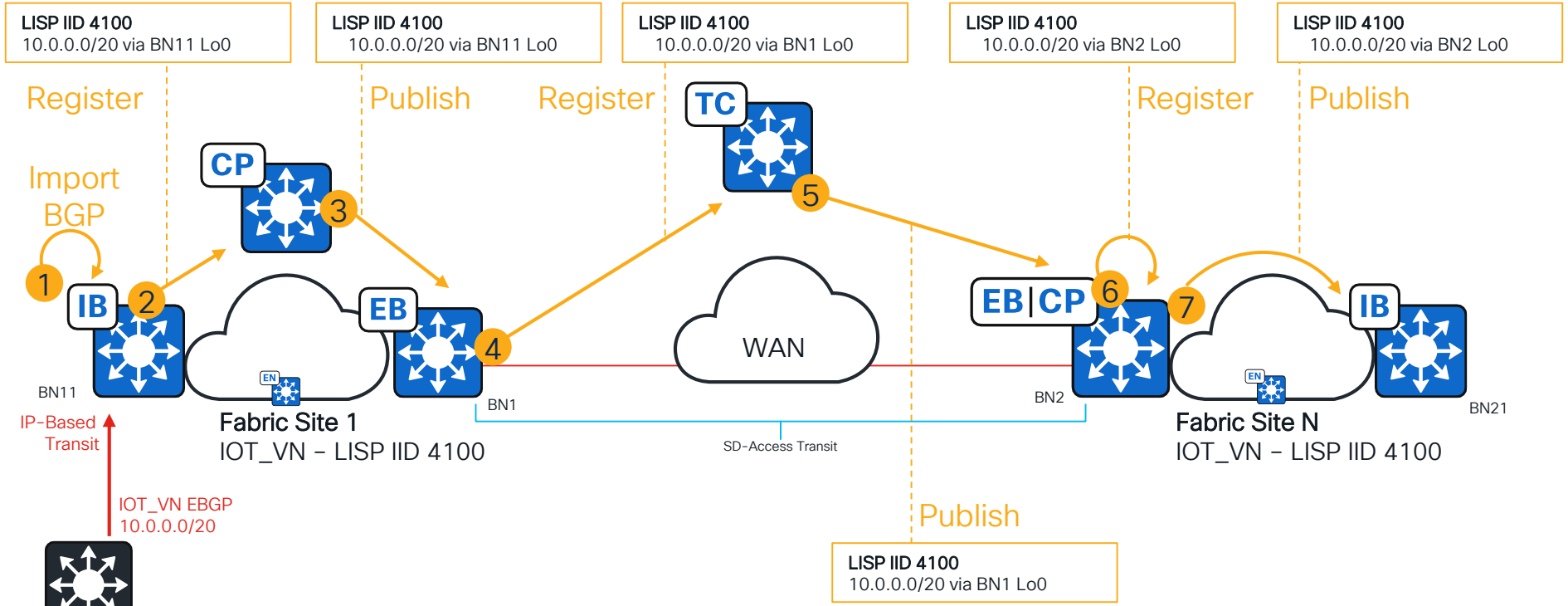
- Import non-default EBGP routes into Site1 for longest-match traffic steering in Site1 and SiteN over SD-Access Transit.

How

- For each L3VN, Site1 Internal Border Nodes import non-default EBGP IPv4 and IPv6 routes into LISP, and register imported routes to Site1 CP.
- Site1 CP publishes imported routes to Site1 External Border Nodes.
- Site1 External Border Nodes register imported routes to Transit CP.
- Transit CP publishes imported routes to SiteN External Border Nodes.
- SiteN External Border Nodes register imported routes to SiteN CP.
- SiteN CP publishes imported routes to SiteN Internal Border Nodes.

Route Propagation Within the Fabric

Imported Routes



Route Propagation to Outside the Fabric Domain

Access Networks

Use Case

- Make endpoints in Site1 reachable to EBGP routing domains outside Site1 and SiteN.

How

- All Border Nodes peering EBGP in Layer 3 Virtual Networks will automatically originate summary routes for SD-Access Anycast Gateways.

```
!BORDER_NODE
router bgp <ASN>
 ...
 address-family ipv4 vrf CORP_VN
  ...
  aggregate-address 10.1.0.0 255.255.255.0 summary-only attribute-map SET_FABRIC_SUBNET_COMMUNITY
  redistribute lisp metric 10 route-map LISP_TO_BGP
  neighbor 172.29.4.49 remote-as 65125
  neighbor 172.29.4.49 update-source Vlan236
  neighbor 172.29.4.49 activate
  neighbor 172.29.4.49 send-community both
  neighbor 172.29.4.49 weight 65535
  neighbor 172.29.4.49 route-map DROP_FABRIC_ROUTES in
 exit-address-family
```

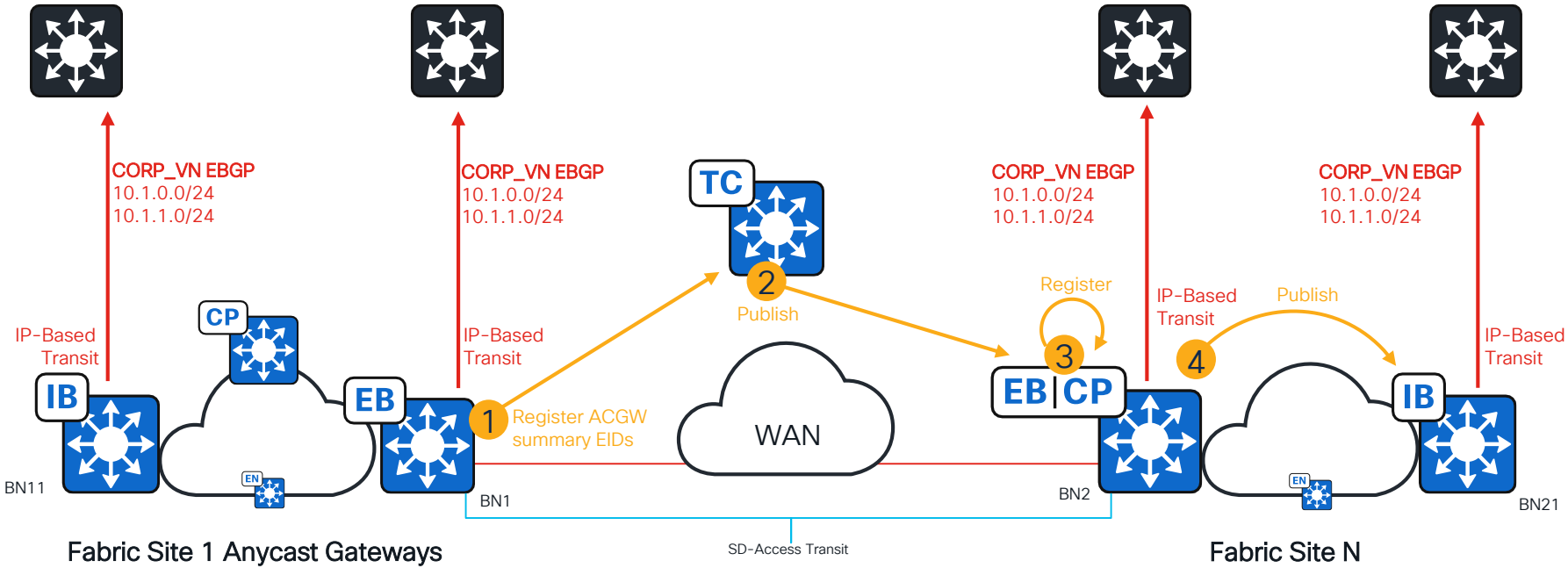
Summary route for local Fabric Site ACGW.

Summary routes for remote ACGWs in Fabric Sites connected to SD-Access Transit.

EBGP peer outside the fabric, aka IP-Based Transit.

Route Propagation to Outside the Fabric Domain

Access Networks



Fabric Site 1 Anycast Gateways
CORP_VN – LISP IID 4099
10.1.0.1/24
10.1.1.1/24

Fabric Site N
CORP_VN – LISP IID 4099

Route Propagation to Outside the Fabric Domain

Default Route

Use Case

- EBGP routing domains outside the fabric need to use an SD-Access Fabric Site as a default route.

How

- Template EBGP default-originate on targeted Border Nodes.
- Design carefully, multiple default routes may lead to unpredictable routing outcomes.

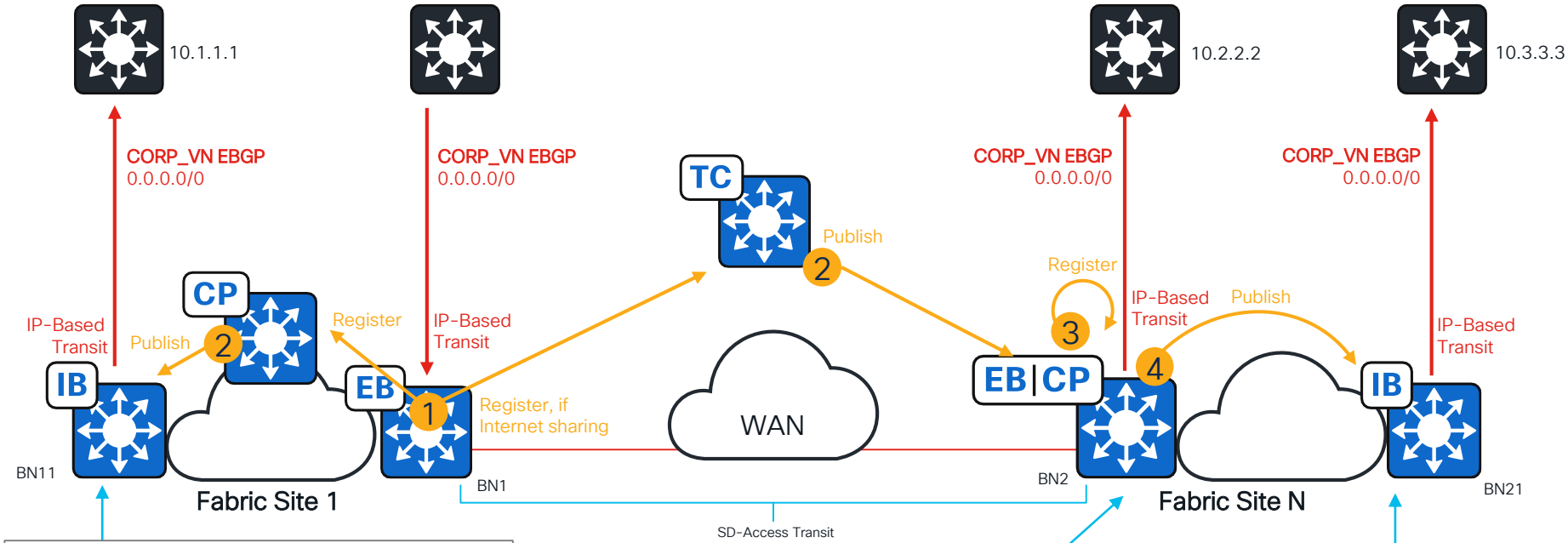
```
!BORDER_NODE
router bgp <ASN>
 ...
 address-family ipv4 vrf CORP_VN
  ...
  aggregate-address 10.1.0.0 255.255.255.0 summary-only attribute-map SET_FABRIC_SUBNET_COMMUNITY
  redistribute lisp metric 10 route-map LISP_TO_BGP
  neighbor 172.29.4.49 remote-as 65125
  neighbor 172.29.4.49 update-source Vlan236
  neighbor 172.29.4.49 activate
  neighbor 172.29.4.49 send-community both
  neighbor 172.29.4.49 weight 65535
  neighbor 172.29.4.49 route-map DROP_FABRIC_ROUTES in
  neighbor 172.29.4.49 default-information originate
 exit-address-family
```

EBGP peer outside the fabric, aka IP-Based Transit

Extra command added by Catalyst Center template

Route Propagation to Outside the Fabric Domain

Default Route




```
router bgp <ASN>
 address-family ipv4 vrf CORP_VN
  neighbor 10.1.1.1 default-information originate
```

```
router bgp <ASN>
 address-family ipv4 vrf CORP_VN
  neighbor 10.2.2.2 default-information originate
```

```
router bgp <ASN>
 address-family ipv4 vrf CORP_VN
  neighbor 10.3.3.3 default-information originate
```

Route Propagation Through the Fabric Domain

Transport between External BGP Autonomous Systems

Use Case

- An EBGP domain connected to SiteN requires connectivity to an EBGP domain connected to Site1 Internal Border Nodes.

How

- SD-Access configuration models do not allow EBGP -> LISP -> EBGP because it eliminates BGP AS Path, which may create BGP loops. It also removes other BGP attributes e.g. communities.
- External BGP route propagation over a Fabric Site, or over SD-Access Transit, is parallel to LISP.
- Within any given Fabric Site a BGP route reflector is automated by Catalyst Center.

Route Propagation Through the Fabric Domain

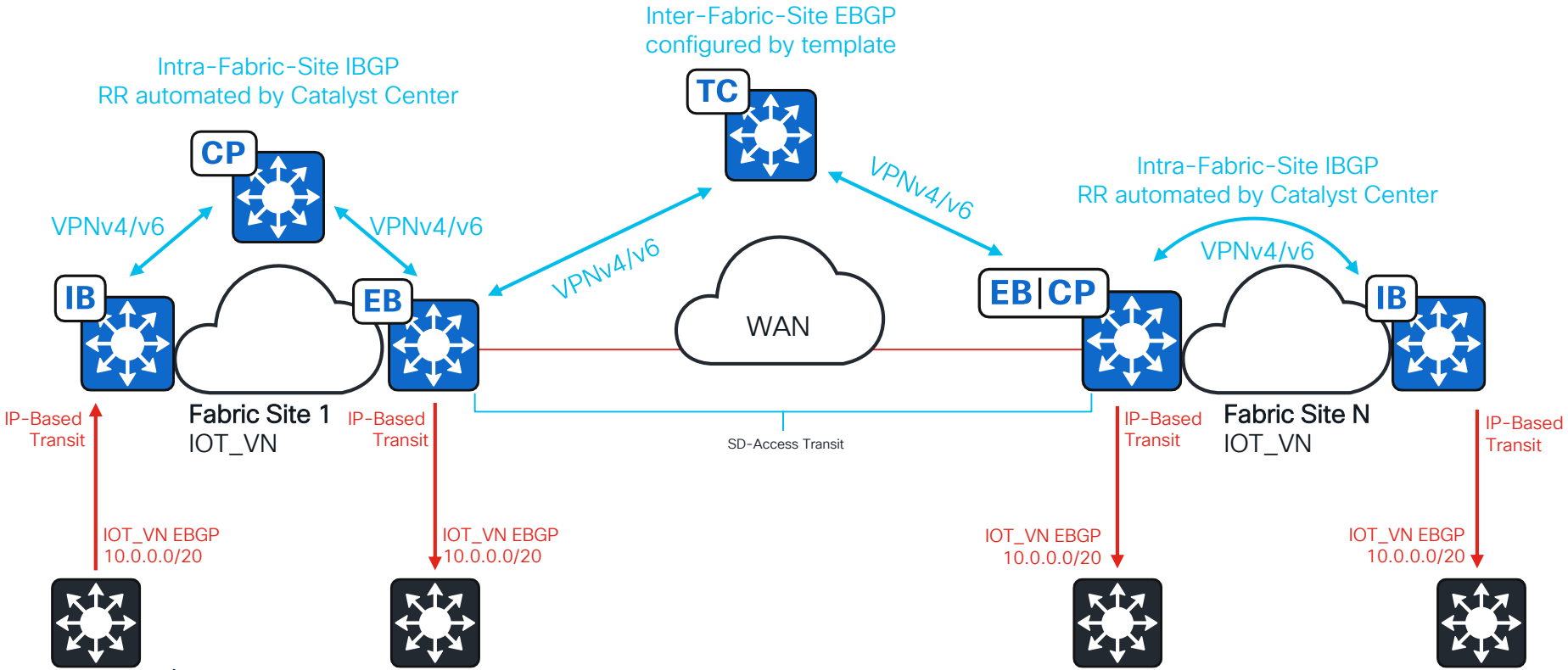
Transport between External BGP Autonomous Systems

How

- Between Fabric Sites connected to SD-Access Transit, if EBGP route propagation is required, then manually template VPNv4/v6 EBGP from External Border Nodes and the SD-Access Transit CPs (TC), or between External Border Nodes and non-Fabric EBGP concentrators.
  - Minimum Catalyst Center version 2.3.7.3.
  - Other VPNv4/v6 EBGP topologies are also possible. Too many permutations to list, please consult an SD-Access SME if variations are necessary.
- Do not modify BGP configuration commands deployed by Catalyst Center, templates should build on the Catalyst Center baseline configurations.
- If unsure, please consult with an SD-Access SME.

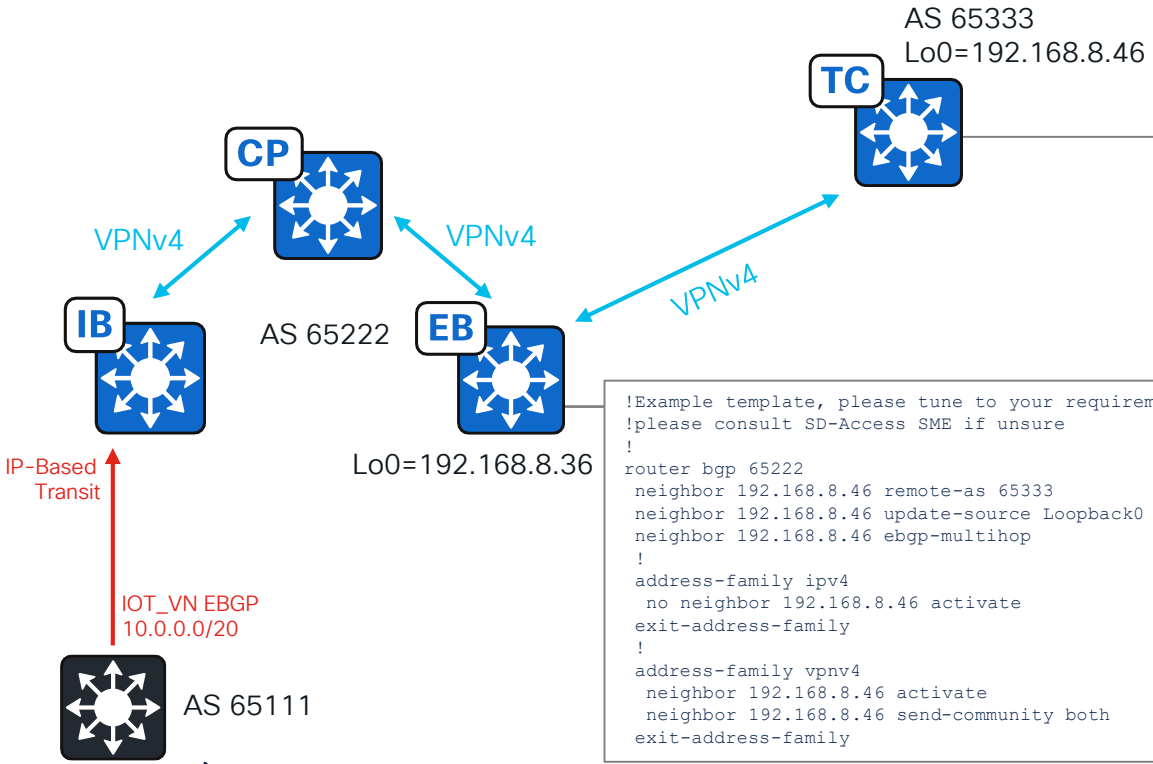
Route Propagation Through the Fabric Domain

Example: Transport between External BGP Autonomous Systems



Route Propagation Through the Fabric Domain

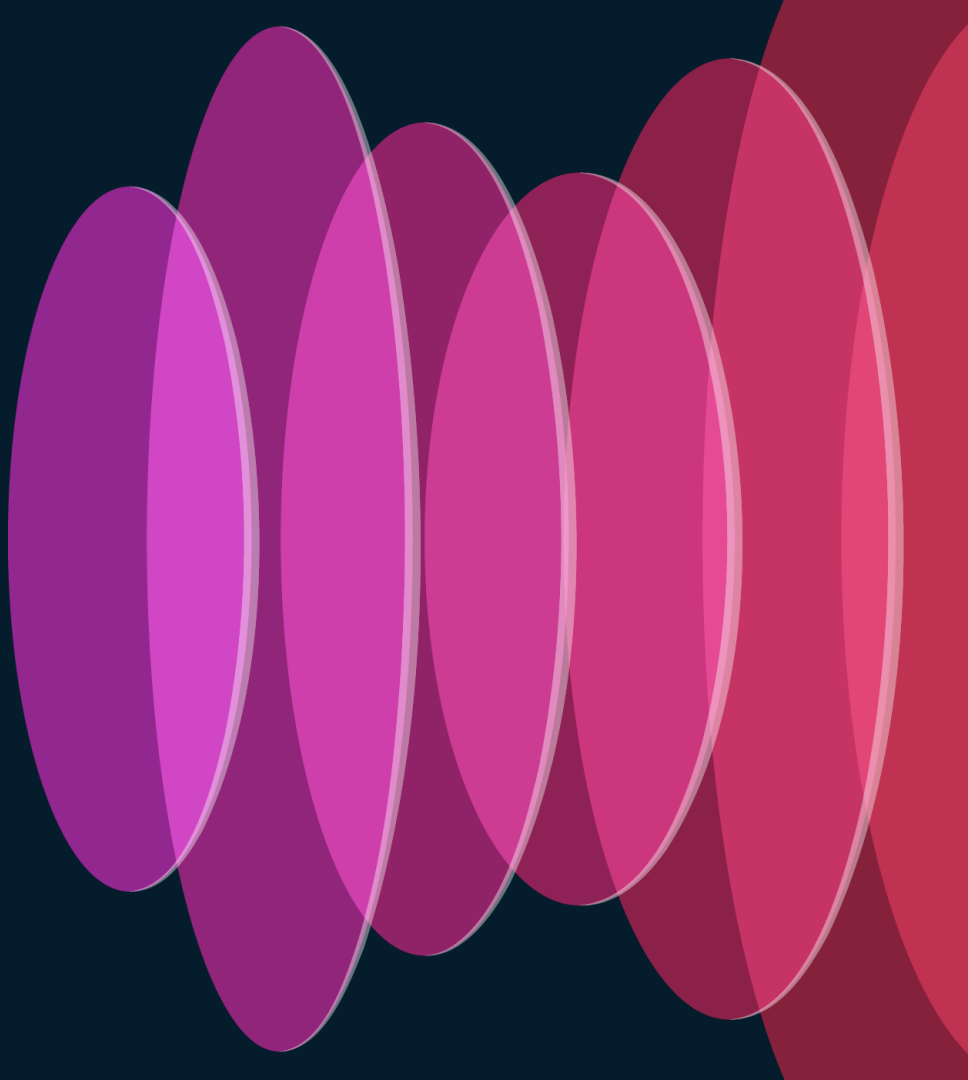
Example: Transport between External BGP Autonomous Systems




```
!Example template, please tune to your requirements,
!please consult SD-Access SME if unsure
!
router bgp 65222
 neighbor 192.168.8.46 remote-as 65333
 neighbor 192.168.8.46 update-source Loopback0
 neighbor 192.168.8.46 ebgp-multihop
 !
 address-family ipv4
  no neighbor 192.168.8.46 activate
 exit-address-family
 !
 address-family vpnv4
  neighbor 192.168.8.46 activate
  neighbor 192.168.8.46 send-community both
 exit-address-family
```

```
!Example template, please tune to your requirements,
!please consult SD-Access SME if unsure
!
ip as-path access-list 1 permit .*_65111
!
route-map BGP_ACCEPT_IN permit 10
 match as-path 1
!
router bgp 65333
 bgp router-id interface Loopback0
 no bgp default route-target filter
 neighbor 192.168.8.36 remote-as 65222
 neighbor 192.168.8.36 update-source Loopback0
 neighbor 192.168.8.36 ebgp-multihop
!
 address-family ipv4
  no neighbor 192.168.8.36 activate
 exit-address-family
 !
 address-family vpnv4
  neighbor 192.168.8.36 activate
  neighbor 192.168.8.36 send-community both
  neighbor 192.168.8.36 route-map BGP_ACCEPT_IN in
 exit-address-family
```

Overlay Routing Rules and Features



Intra-Fabric Site Fast Overlay Convergence

- Overlay convergence times are entirely dependent on the underlay (GRT) IGP performance.
- Underlay should have /32 RLOC and Control Plane Node Lo0 routes.
- On Edge Nodes, anything less specific than /32 RLOC RIB entries will be ignored in SD-Access 2.3.7.3 and later.
- On Border Nodes, default route will be ignored for RLOC reachability. All other routes are valid, which can make summary routes problematic.

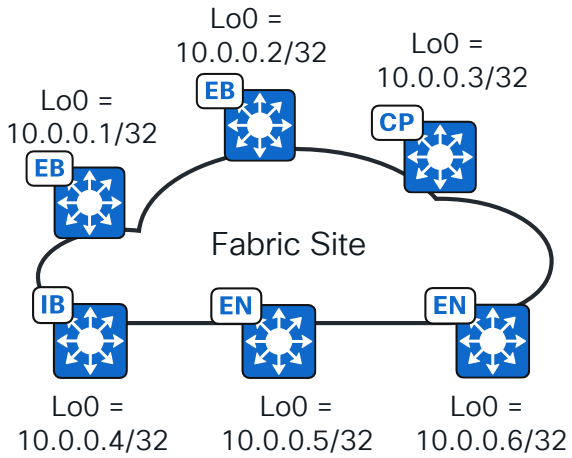
```
!FABRIC_EDGE_NODE
!
router lisp
  …
  ipv4 locator reachability minimum-mask-length 32
```

There must be a /32 routes for RLOCs in underlay

```
!BORDER_NODE
!
router lisp
  …
  ipv4 locator reachability exclude-default
```

There must be a non-default routes for RLOCs in underlay. RLOC summary routes should not exist if fast convergence is required

Intra-Fabric Site Fast Overlay Convergence




```
INTERMEDIATE_NODE#show ip route ospf (or ISIS, or EIGRP, or static)
O     10.0.0.1    [110/11] via 192.0.2.9, 1d00h, HundredGigE1/0/10
O     10.0.0.2    [110/11] via 192.0.2.11, 1d00h, HundredGigE1/0/12
O     10.0.0.3    [110/11] via 192.0.2.13, 1d00h, HundredGigE1/0/14
O     10.0.0.4    [110/11] via 192.0.2.15, 1d00h, HundredGigE1/0/16
O     10.0.0.5    [110/11] via 192.0.2.17, 1d00h, HundredGigE1/0/18
O     10.0.0.6    [110/11] via 192.0.2.19, 1d00h, HundredGigE1/0/20
0     10.0.0.0/24* [110/11] via 192.0.2.9, 1d00h, HundredGigE1/0/10
O*E2 0.0.0.0/0 [110/1] via 192.0.2.9, 1d00h, HundredGigE1/0/10

INTERMEDIATE_NODE#
```

*This will hide a fabric node reload from EB/IB. Overlay traffic may be sent to an RLOC (Lo0) that does not exist, potentially causing unnecessary packet loss.

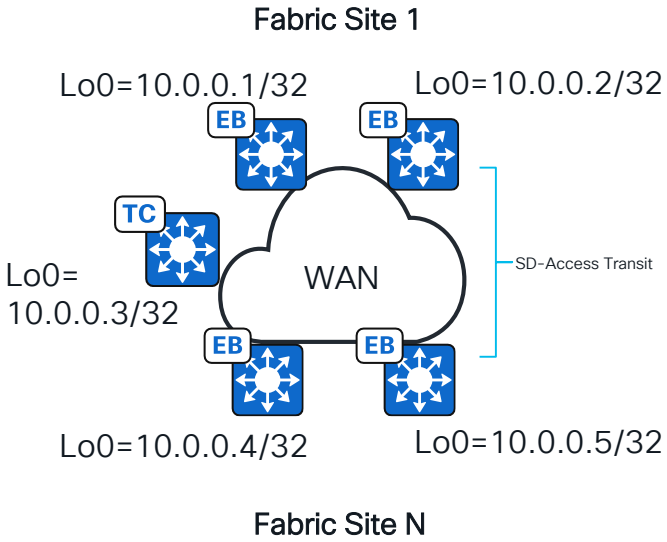
SD-Access Transit Fast Overlay Convergence

- Overlay convergence times are entirely dependent on the underlay (GRT) IGP performance.
- Underlay between Border Nodes connected to SD-Access Transit should have have /32 Border Node RLOC and Transit Control Plane Node Lo0 routes.
- Edge Node /32s are not required between Fabric Sites, assuming no Multisite Remote Border (out of scope for this presentation).
- On Border Nodes, summary routes are considered valid for remote BN RLOC reachability, which can make summary routes problematic.

```
!EXTERNAL_BORDER_NODE_CONNECTED_TO_SD-ACCESS_TRANSIT
!
router lisp
  …
  ipv4 locator reachability exclude-default
```

There must be non-default routes for SD-Access Transit RLOCs in underlay. RLOC summary routes should not exist if fast convergence required

SD-Access Transit Fast Overlay Convergence




```
WAN_ROUTER#show ip route ospf (or ISIS, or EIGRP, or static, or BGP)
O    10.0.0.1    [110/11] via 192.0.2.1, 1d00h, HundredGigE1/0/1
O    10.0.0.2    [110/11] via 192.0.2.3, 1d00h, HundredGigE1/0/3
O    10.0.0.3    [110/11] via 192.0.2.5, 1d00h, HundredGigE1/0/5
O    10.0.0.4    [110/11] via 192.0.2.7, 1d00h, HundredGigE1/0/7
O    10.0.0.5    [110/11] via 192.0.2.9, 1d00h, HundredGigE1/0/9
0    10.0.0.0/24* [110/11] via 192.0.2.1, 1d00h, HundredGigE1/0/1
O*E2 0.0.0.0/0 [110/1] via 192.0.2.1, 1d00h, HundredGigE1/0/1

WAN_ROUTER#
```

*This will hide an External Border Node failure from other External Border Nodes connected to an SD-Access Transit. If a BN fails then overlay traffic may be sent to an RLOC that does not exist, potentially causing unnecessary packet loss.

Multicast Over SD-Access Transit

- Supported with LISP Pub/Sub, not LISP/BGP.
- Headend replication:
  - Overlay IPv4 multicast encapsulated in underlay IPv4 unicast VXLAN.
  - No multicast routing required in the underlay within Fabric Sites and between Fabric Sites.
  - Underlay Typically simpler to configure and maintain.
- Native replication:
  - Overlay IPv4 multicast encapsulated in underlay IPv4 multicast VXLAN.
  - PIM-SSM required in the underlay within Fabric Sites and between Fabric Sites.
  - More scalable if there are many multicast egress nodes (distributed receivers) connected to the SD-Access Fabric.
- All Fabric Sites sending or receiving Any-Source Multicast over SD-Access Transit should use the same overlay RPs. Talk to an SME for design variations.
- All Fabric Sites sending or receiving ASM or SSM over SD-Access Transit must use the same replication method: Headend or Native.
- More information in BRKENS-3850 - Demystifying Multicast Operations in a multi-site SD-Access deployment

LISP Next Hop Selection Logic

General notes:

- This section assumes IOS XE 17.12.1 or later.
- By default, Affinity ID is off.
- By default, all LISP priorities are 10.
- If there is multiple equally preferable routes, then traffic is load balanced across all equally preferable RLOCs.

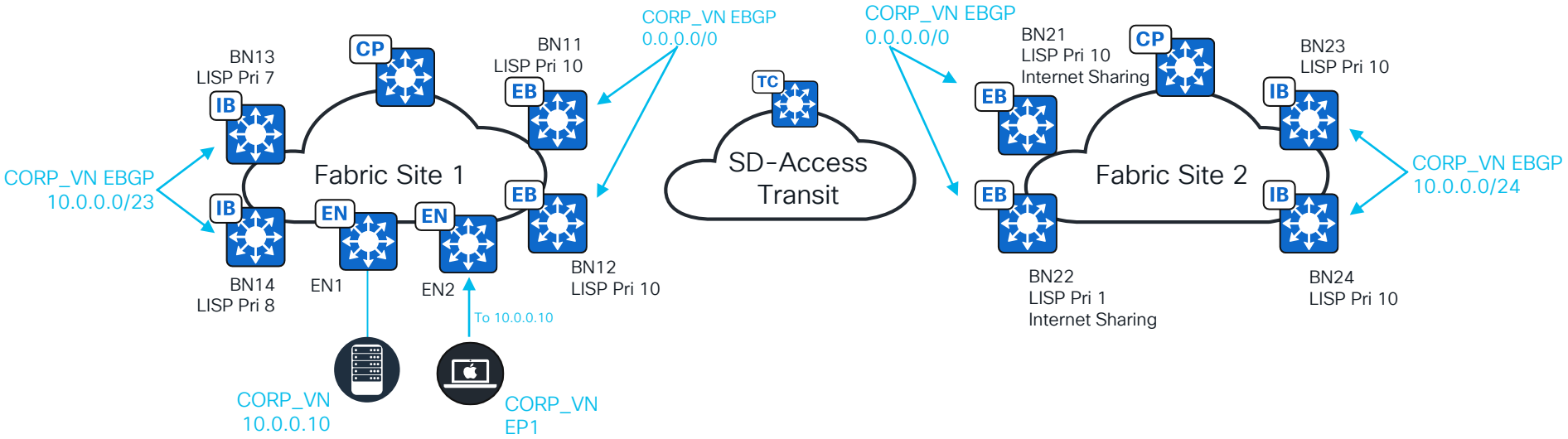
LISP Next Hop Selection Logic

When a routed packet ingresses an SD-Access fabric:

1. Match most specific non-default route. If there are multiple:
   a. Match routes with lowest LISP Priority registered by the local Fabric Site.
   b. Match routes with lowest LISP Priority learned from SD-Access Transit CP.
2. Match the default routes registered by the local Fabric Site with lowest LISP Priority.
3. Match default routes learned from SD-Access Transit CP with the closest Affinity ID.
4. Match default routes learned from SD-Access Transit CP with the lowest LISP Priority.

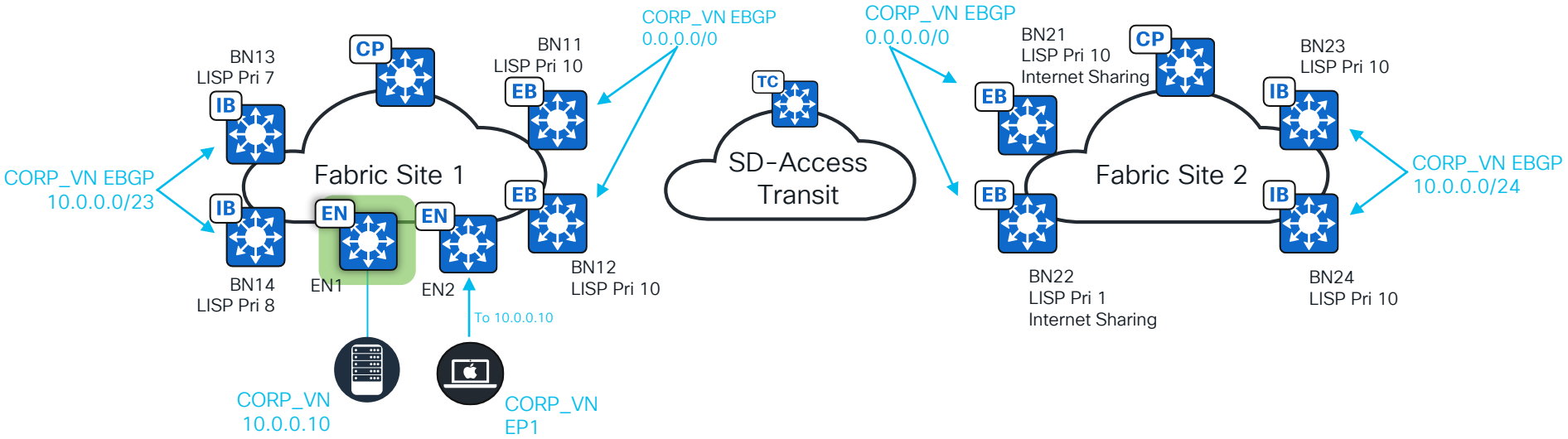
Example: LISP Next Hop Selection

When EP1 sends a packet into CORP_VN destined for 10.0.0.10 EN2 will:



Example: LISP Next Hop Selection

When EP1 sends a packet into CORP_VN destined for 10.0.0.10 EN2 will:
1. Send via EN1 (most specific route)

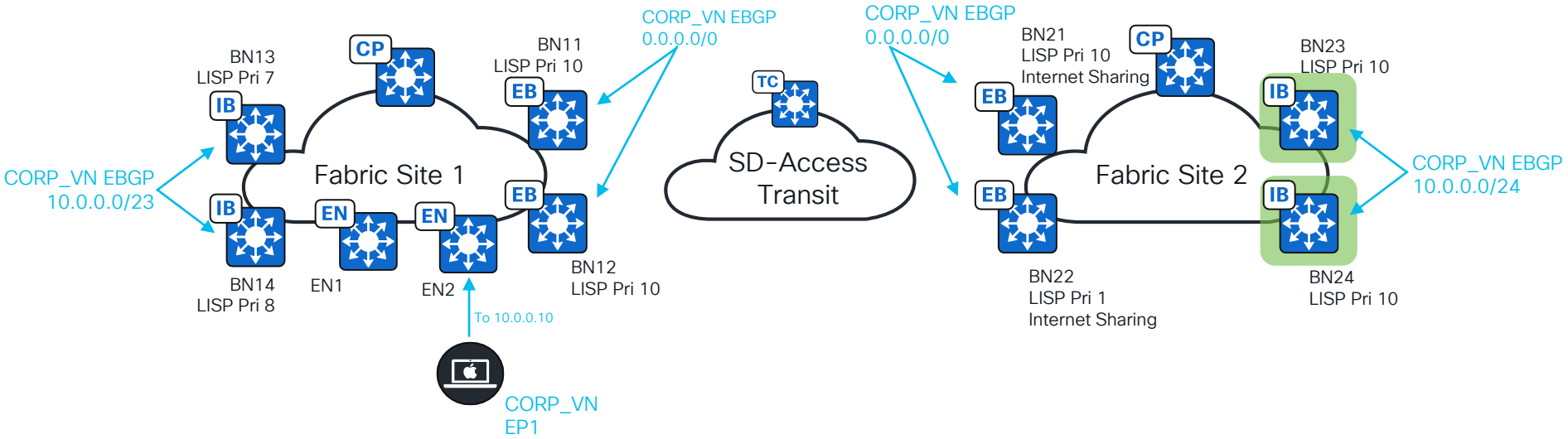


Example: LISP Next Hop Selection

When EP1 sends a packet into CORP_VN destined for 10.0.0.10 EN2 will:

~~1. Send via EN1 (most specific route)~~

2. Send via SD-Access Transit to BN23 and BN24 (most specific routes with equal LISP priority)



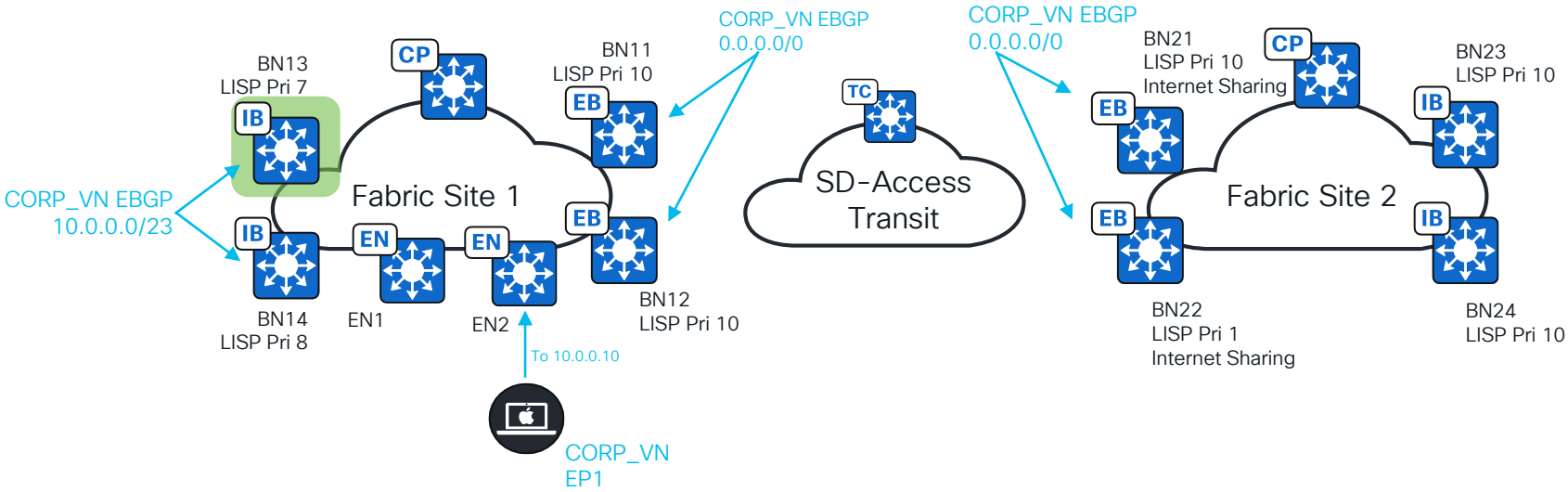
Example: LISP Next Hop Selection

When EP1 sends a packet into CORP_VN destined for 10.0.0.10 EN2 will:

~~1. Send via EN1 (most specific route)~~

~~2. Send via SD-Access Transit to BN23 and BN24 (most specific routes with equal LISP priority)~~

3. Send via BN13 (most specific route with lowest priority)



Example: LISP Next Hop Selection

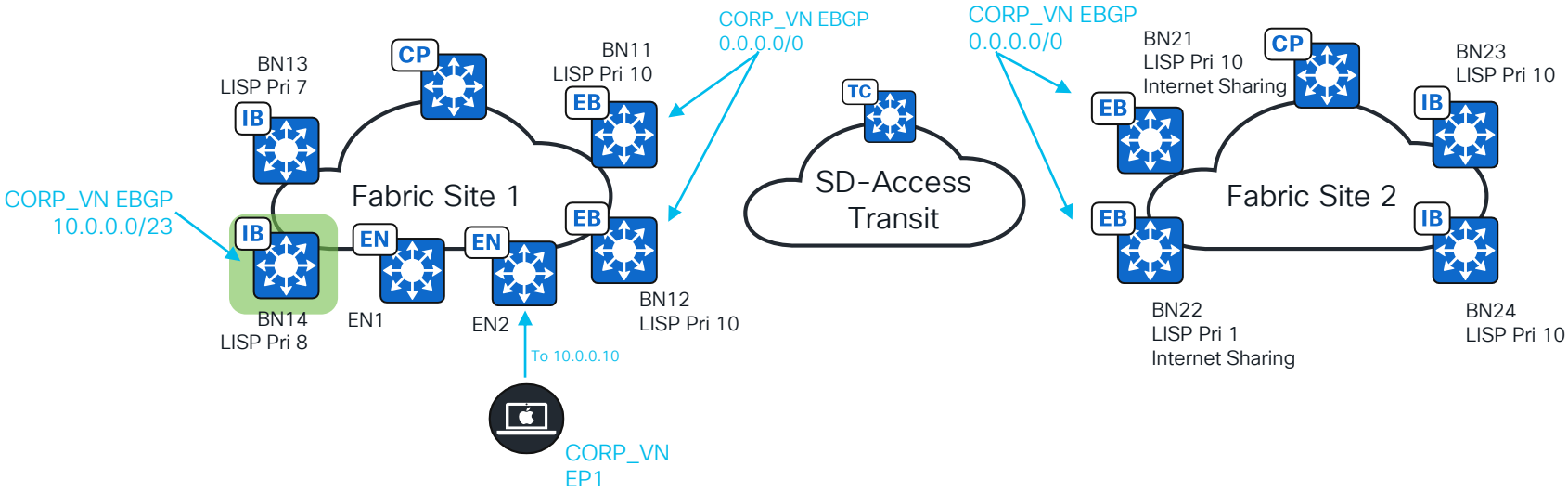
When EP1 sends a packet into CORP_VN destined for 10.0.0.10 EN2 will:

~~1. Send via EN1 (most specific route)~~

~~2. Send via SD-Access Transit to BN23 and BN24 (most specific routes with equal LISP priority)~~

~~3. Send via BN13 (most specific route with lowest priority)~~

4. Send via BN14 (most specific route)



Example: LISP Next Hop Selection

When EP1 sends a packet into CORP_VN destined for 10.0.0.10 EN2 will:

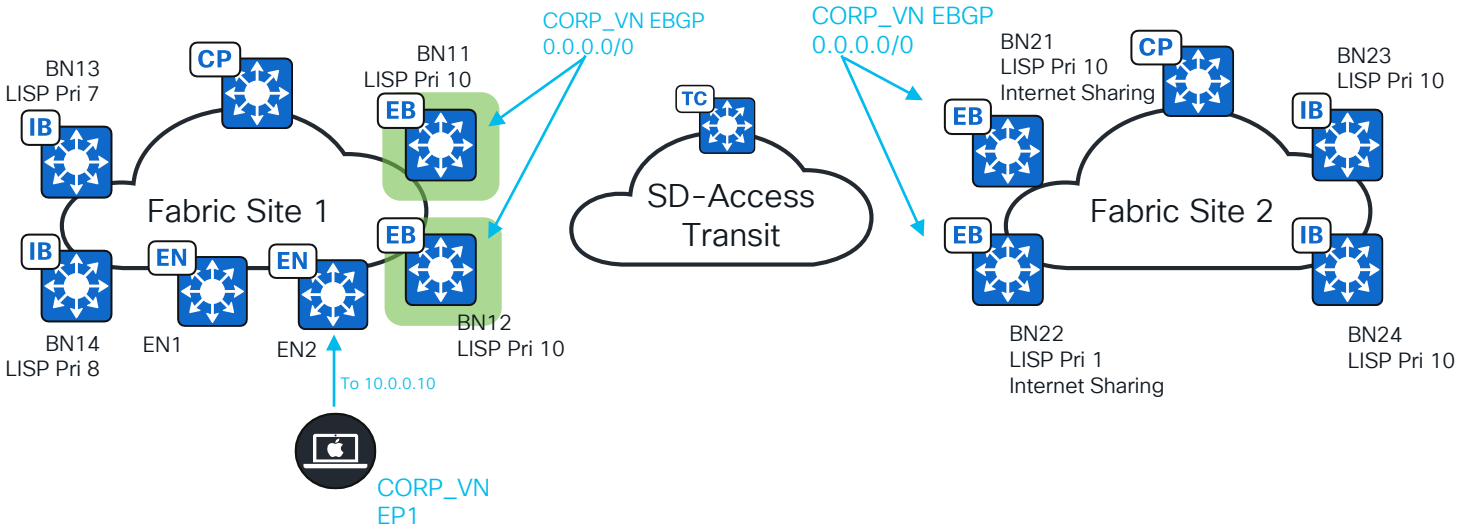
~~1. Send via EN1 (most specific route)~~

~~2. Send via SD-Access Transit to BN23 and BN24 (most specific routes with equal LISP priority)~~

~~3. Send via BN13 (most specific route with lowest priority)~~

~~4. Send via BN14 (most specific route)~~

5. Send via BN11 and BN12 (site-local default routes are preferable to remote default routes. Equal LISP priority)



Example: LISP Next Hop Selection

When EP1 sends a packet into CORP_VN destined for 10.0.0.10 EN2 will:

~~1. Send via EN1 (most specific route)~~

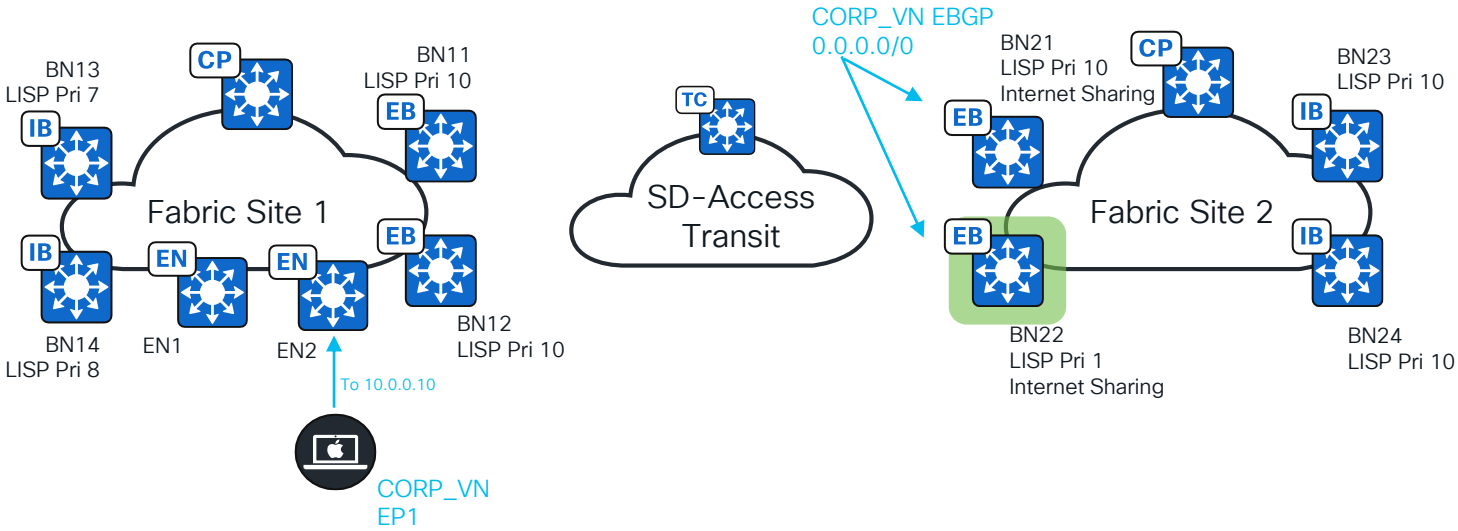
~~2. Send via SD-Access Transit to BN23 and BN24 (most specific routes with equal LISP priority)~~

~~3. Send via BN13 (most specific route with lowest priority)~~

~~4. Send via BN14 (most specific route)~~

~~5. Send via BN11 and BN12 (site-local default routes are preferable to remote default routes. Equal LISP priority)~~

6. Send via SD-Access Transit to BN22 (remote default route with lowest priority)



Example: LISP Next Hop Selection

When EP1 sends a packet into CORP_VN destined for 10.0.0.10 EN2 will:

~~1. Send via EN1 (most specific route)~~

~~2. Send via SD-Access Transit to BN23 and BN24 (most specific routes with equal LISP priority)~~

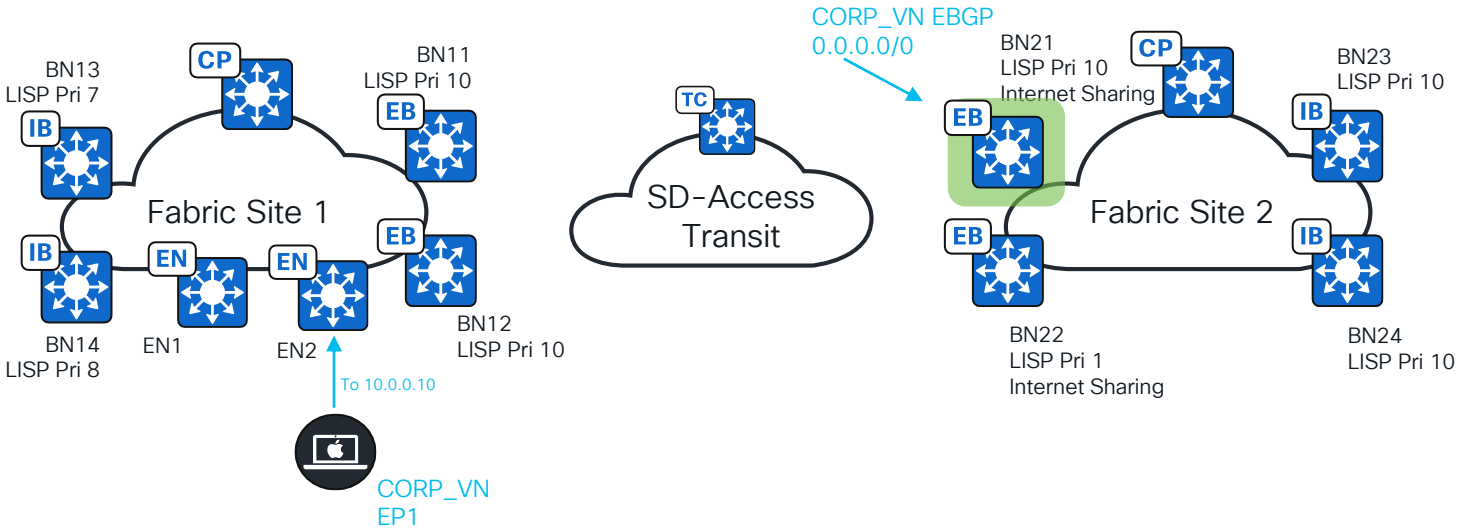
~~3. Send via BN13 (most specific route with lowest priority)~~

~~4. Send via BN14 (most specific route)~~

~~5. Send via BN11 and BN12 (site-local default routes are preferable to remote default routes. Equal LISP priority)~~

~~6. Send via SD-Access Transit to BN22 (remote default route with lowest priority)~~

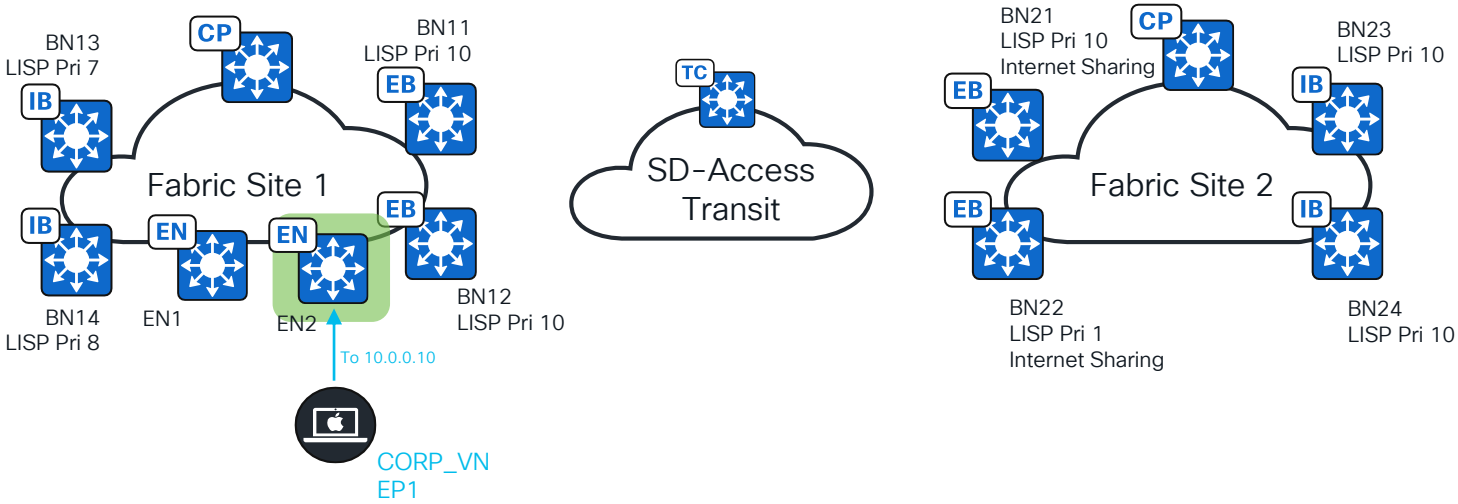
7. Send via SD-Access Transit to BN21(the only route available)



Example: LISP Next Hop Selection

When EP1 sends a packet into CORP_VN destined for 10.0.0.10 EN2 will:

~~1. Send via EN1 (most specific route)~~
~~2. Send via SD-Access Transit to BN23 and BN24 (most specific routes with equal LISP priority)~~
~~3. Send via BN13 (most specific route with lowest priority)~~
~~4. Send via BN14 (most specific route)~~
~~5. Send via BN11 and BN12 (site-local default routes are preferable to remote default routes)~~
~~6. Send via SD-Access Transit to BN22 (remote default route with lowest priority)~~
~~7. Send via SD-Access Transit to BN21(the only route available)~~
8. Drop (no route)



Preserve Priority

Use Case

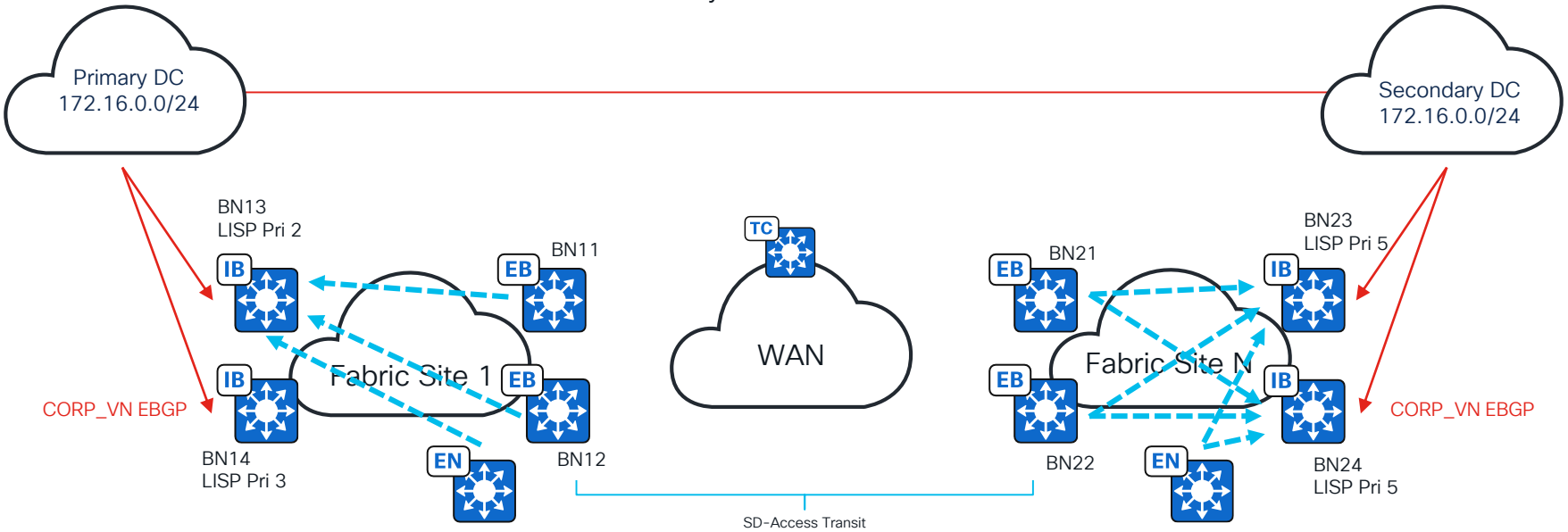
- All traffic for a non-fabric destination connected to multiple Fabric Sites must egress the same Internal Border Nodes e.g. primary DC and secondary DC.

How

- Off by default.
- When on, the check for non-default routes registered by local or remote Fabric Site is ignored, the lowest LISP priority decides best path for duplicate non-default routes.
- Minimum Catalyst Center 2.3.7.3 and minimum IOS XE 17.12.1.
- Not enabled by SD-Access automation yet, requires templates, speak with your Cisco SME to explore this feature.

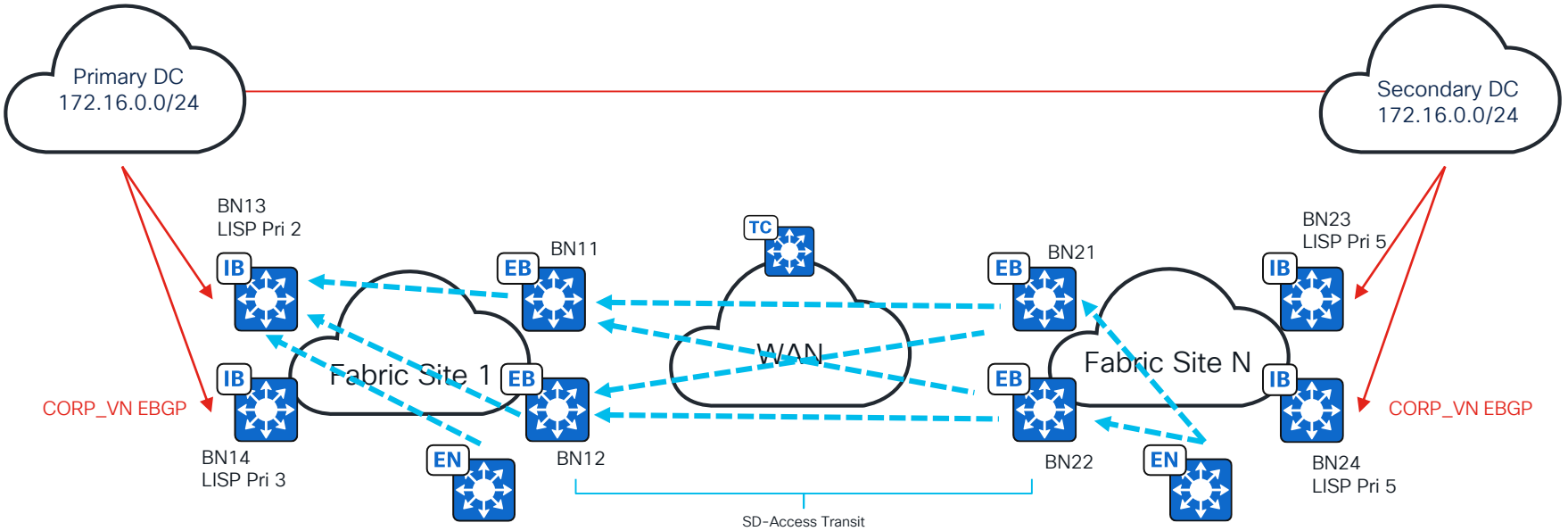
Preserve Priority Off

- Preserve Priority is off by default and LISP next hop selection logic will:
  1. Match most specific non-default route. If there are multiple:
     a. Match routes with lowest LISP Priority registered by the local Fabric Site.
     b. Match routes with lowest LISP Priority learned from SD-Access Transit CP.



Preserve Priority On

- If Preserve Priority is turned on, LISP next hop selection logic will:
  1. Match most specific non-default route. If there are multiple:
     a. Match routes with lowest LISP Priority regardless of local site registration or remote site registration.



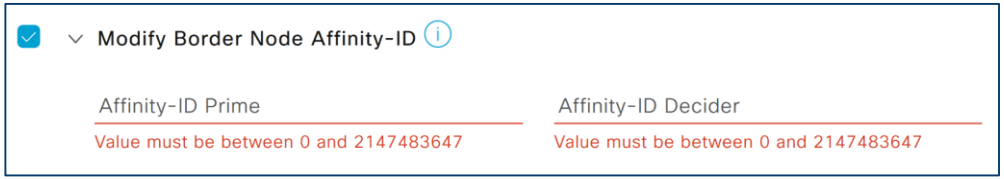
Affinity ID

Use Case

- Each Fabric Site must dynamically select the closest default routes available over SD-Access Transit.

How

- Off by default.
- Applicable to remote default routes reachable over SD-Access Transit.
- Affinity ID comprises two values. In the Catalyst Center UI, there is Affinity ID Prime and Affinity ID Decider:



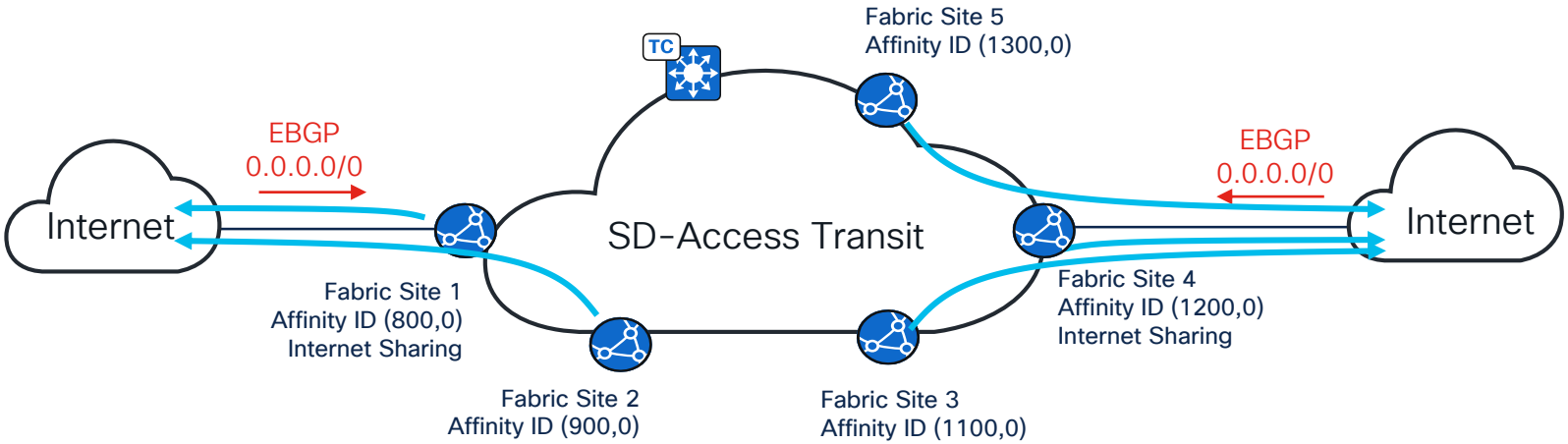
- For each External Border Node, the lowest relative Affinity ID associated with a a default route is selected as most preferable.
- Support for non-default routes is present in LISP but not SD-Access automation yet, please discuss with your Cisco representative.

LISP Affinity ID

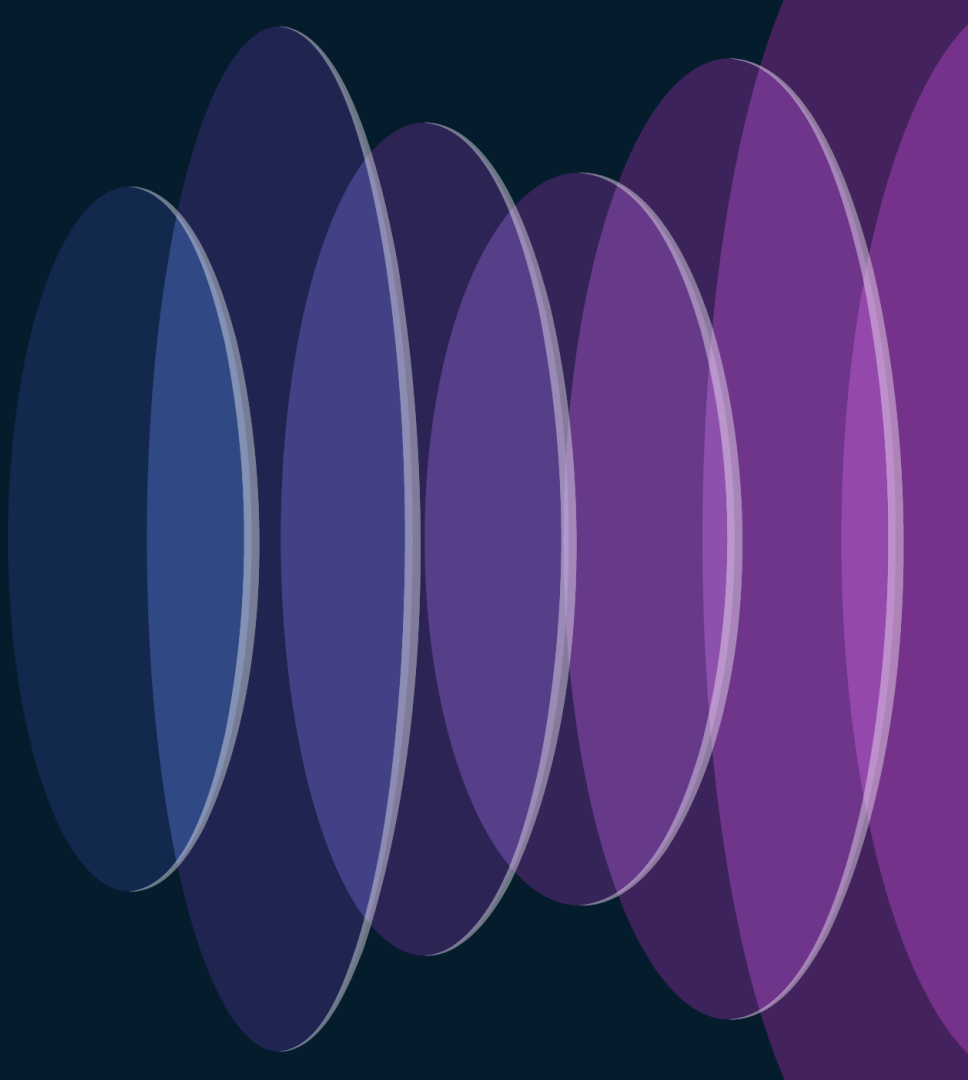
How to Calculate

1. Take the absolute value of the Affinity ID Prime of the local Border Node minus the Affinity ID Prime received from another Border Node.
   a. Repeat this calculation for each Border Node (for each received Affinity ID Prime value).
   b. The Border Node with the lowest relative Prime value based on this calculation is the most preferred Border Node.
2. If the absolute value of the Affinity ID Prime is equal across two or more devices, only then consider the Affinity ID Decider value.
   a. Take the absolute value of the Affinity ID Decider of the local Border Node minus the Affinity ID Decider received from another Border Node.
   b. Repeat this calculation for each Border Node (for each received Affinity ID Decider value).
   c. The Border Node with the lowest relative Decider value based on this calculation is the most preferred Border Node.
3. If the absolute value of Affinity ID Prime and Affinity ID Decider is equal across two or more devices, only then consider Priority value.
   a. The Border Node with the lower Priority value is is the most preferred Border Node.
4. If the Priority value is equal across one or more devices, traffic is load-balanced by CEF.

LISP Affinity ID Example



SD-Access Transit Design Myths



cisco Live!

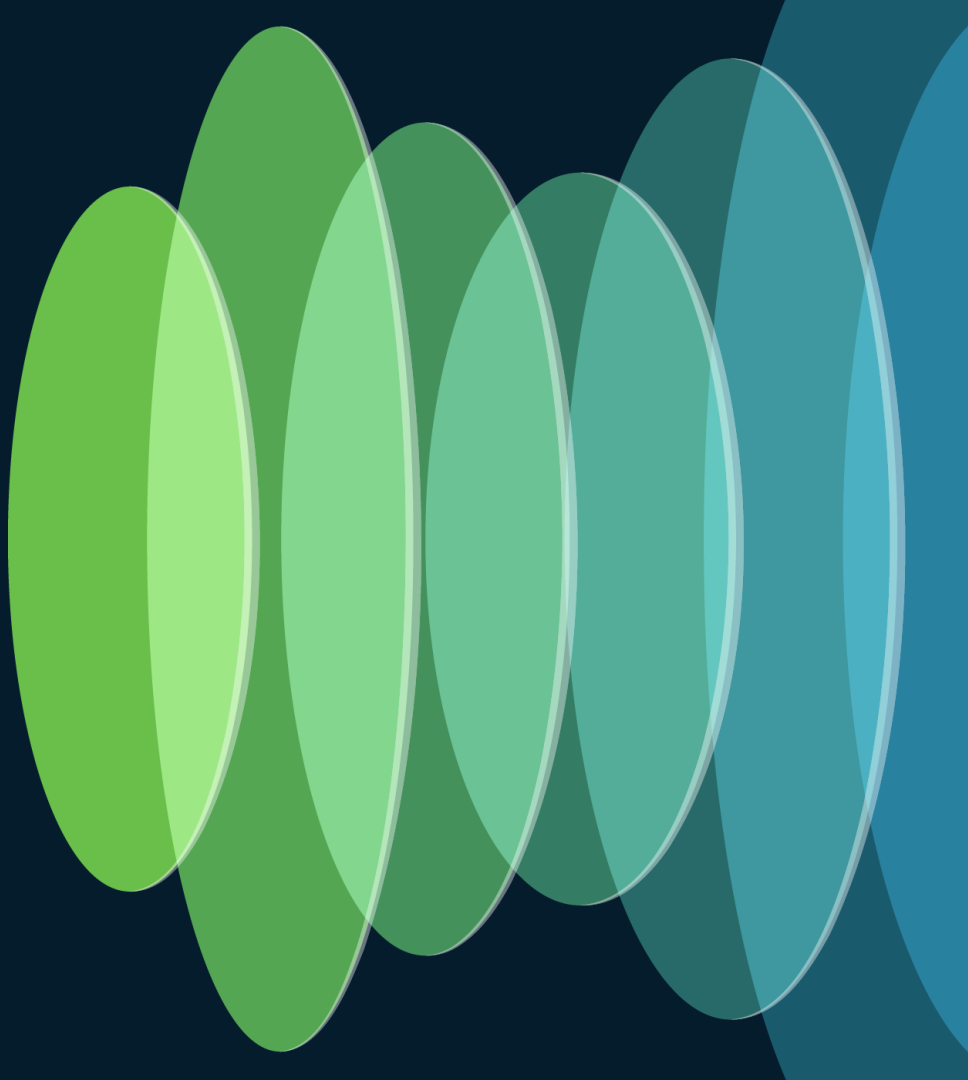
SD-Access Transit Design Myths

- There is no minimum throughput between Fabric Sites, but please be sensible.
- There is no specific type/speed* of WAN link that must be used between Fabric Sites.
- There is no WAN topology restrictions*, all are supported: ring, hub-and-spoke, mesh, partial mesh, daisy chain, etc.
- The Fabric Site and WAN underlay MTU does not have to be a specific value e.g., 9000, 9100, etc. See MTU section of this presentation.
- There is no latency restrictions** between Fabric Sites.
- There is no latency restrictions** between External Border Nodes and SD-Access Transit Control Plane Nodes.

*Noting the WAN-technology-agnostic prerequisites discussed in this presentation: MTU, /32 RLOCs in RIB, multicast routing protocol, etc.

**There is a maximum supported latency within a Fabric Site, and between Catalyst Center and managed devices. Refer to the latest Catalyst Center Data Sheet on cisco.com.

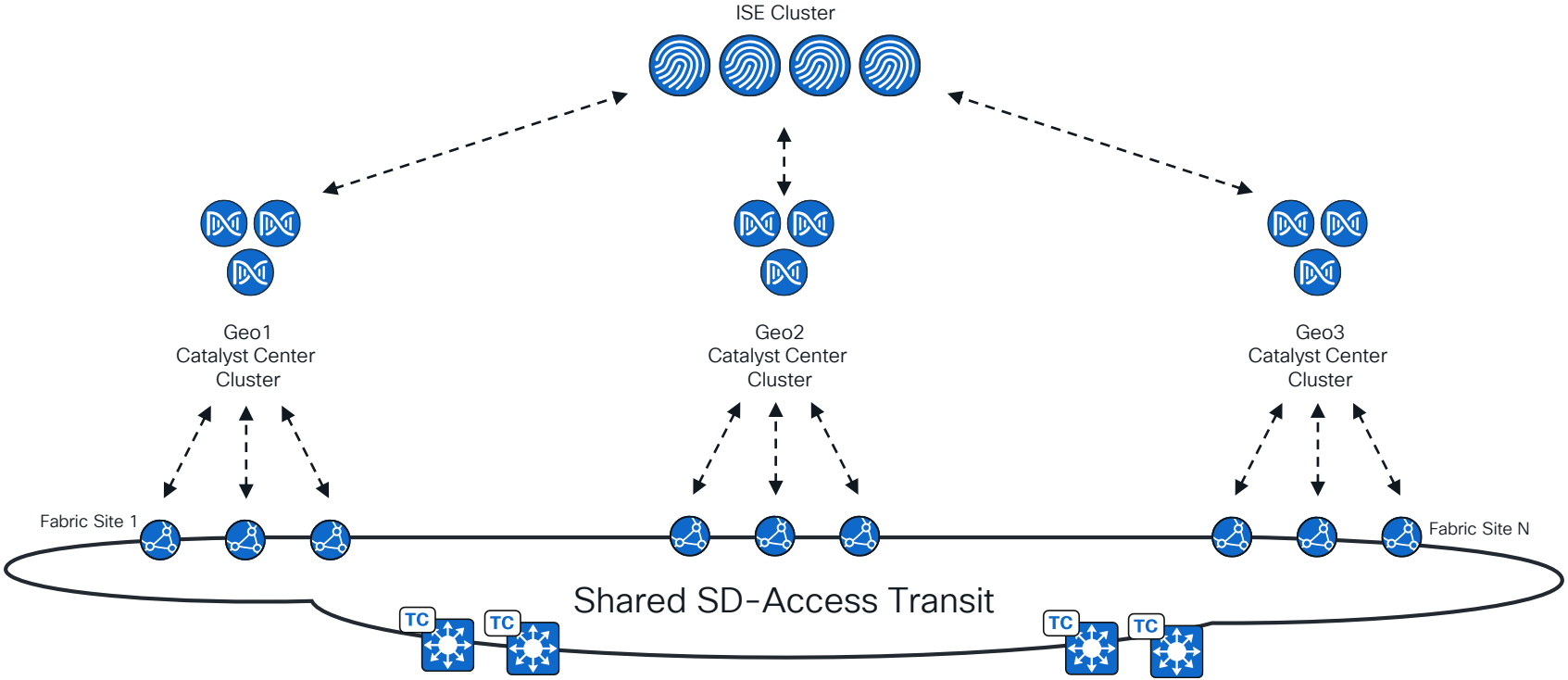
Shared SD-Access Transit



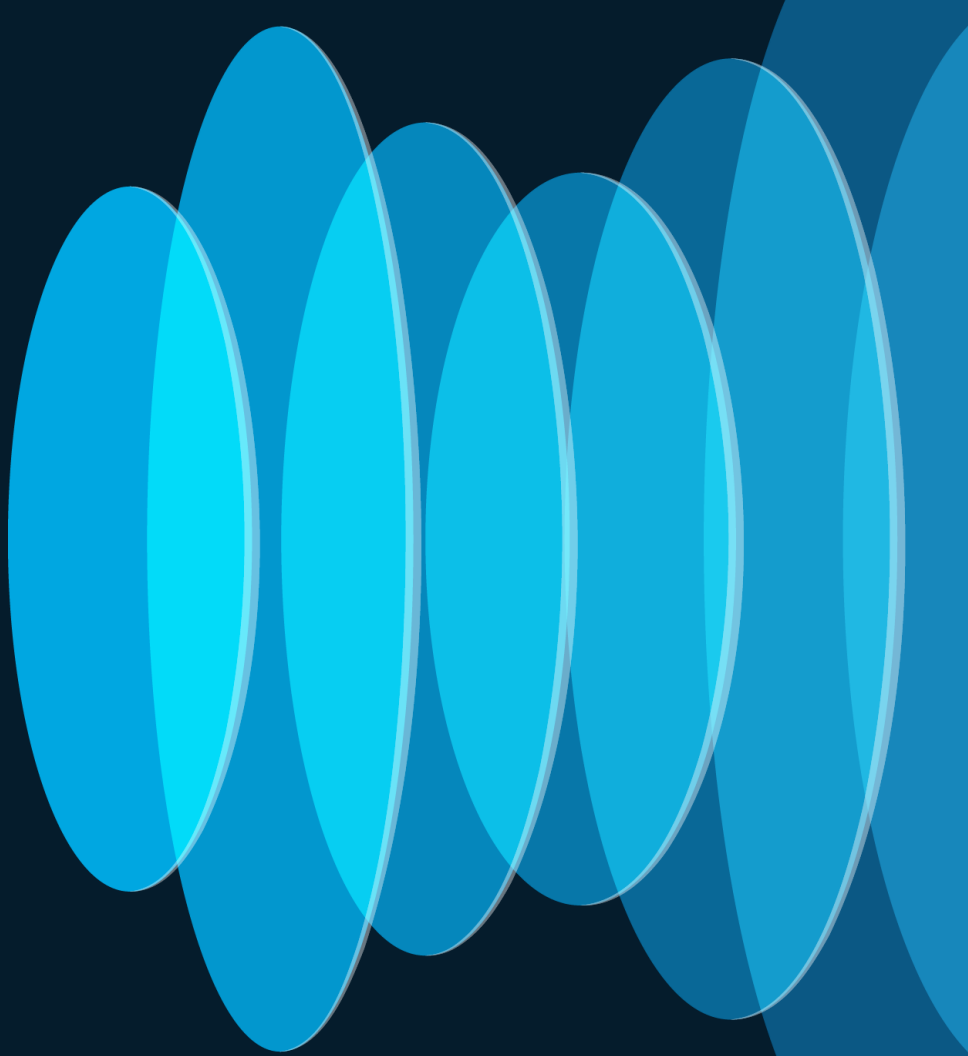
Shared SD-Access Transit

- An SD-Access Transit can be shared amongst multiple Catalyst Center Clusters.
- Requires the Multiple Catalyst Center to ISE software package which is subject to Controlled Availability. Released by exception after a design review, speak to your Cisco CX or Sales representative.
- Typically used when:
  - Different geographies have dedicated Catalyst Center Clusters, OR,
  - Aggregate scale (devices, sites, endpoints, etc.) exceeds the capabilities of a single Catalyst Center Cluster.

Shared SD-Access Transit

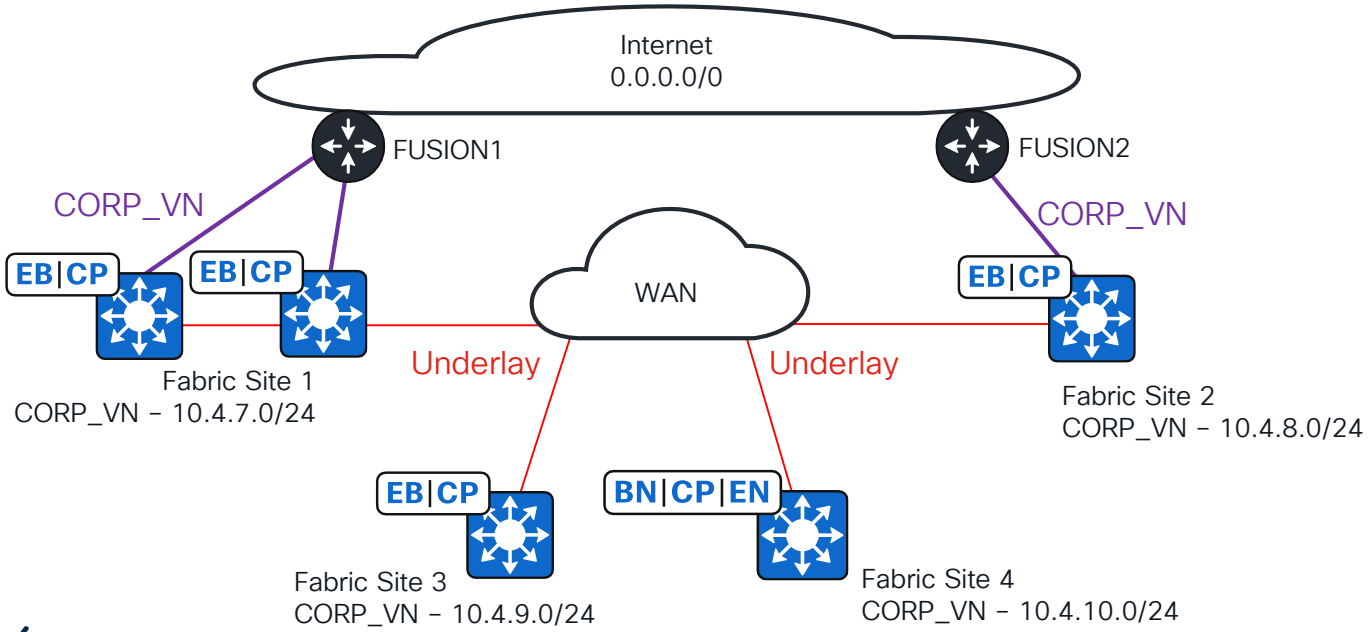


Demo Scenario



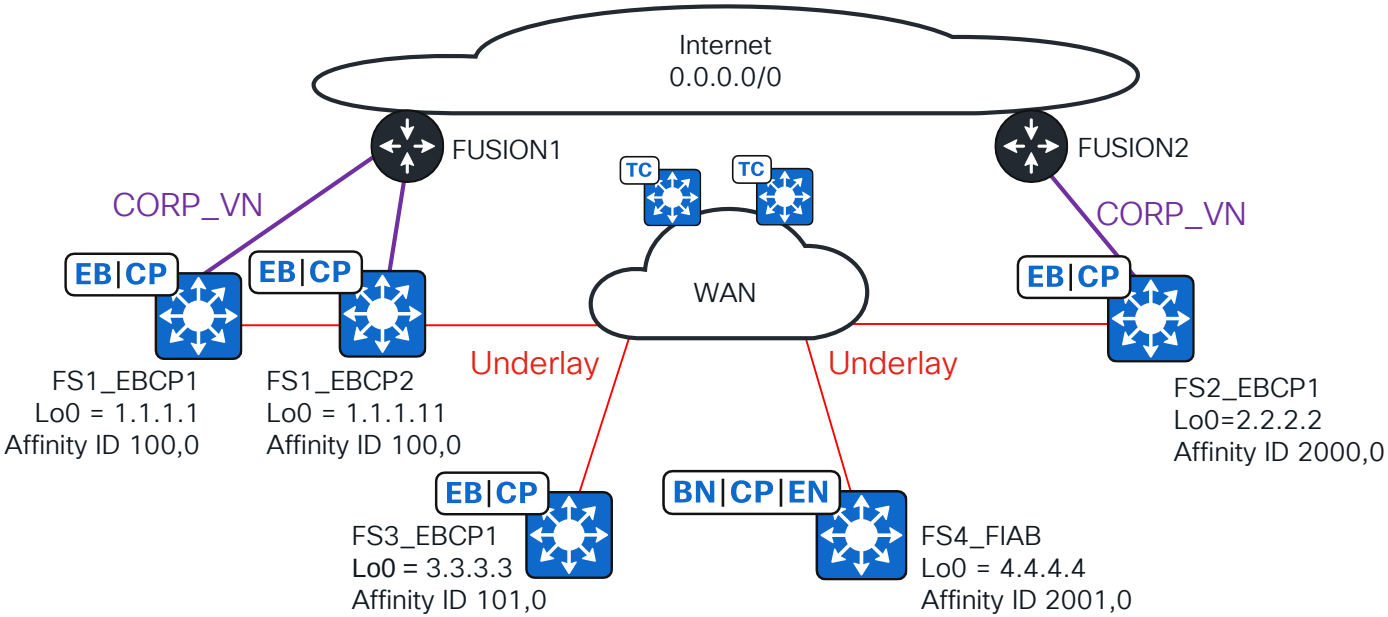
Requirements

- L3VN and SGT preservation between Fabric Sites 1, 2, 3 and 4.
- Fabric Site 1 CORP_VN uses Fabric Site 2 for backup internet access and vice versa.
- Fabric Site 3 CORP_VN uses Fabric Site 1 internet, with backup internet from Fabric Site 2.
- Fabric Site 4 CORP_VN uses Fabric Site 2 internet, with backup default internet from Fabric Site 1.
- Traffic between CORP_VN and internet must be symmetrical.

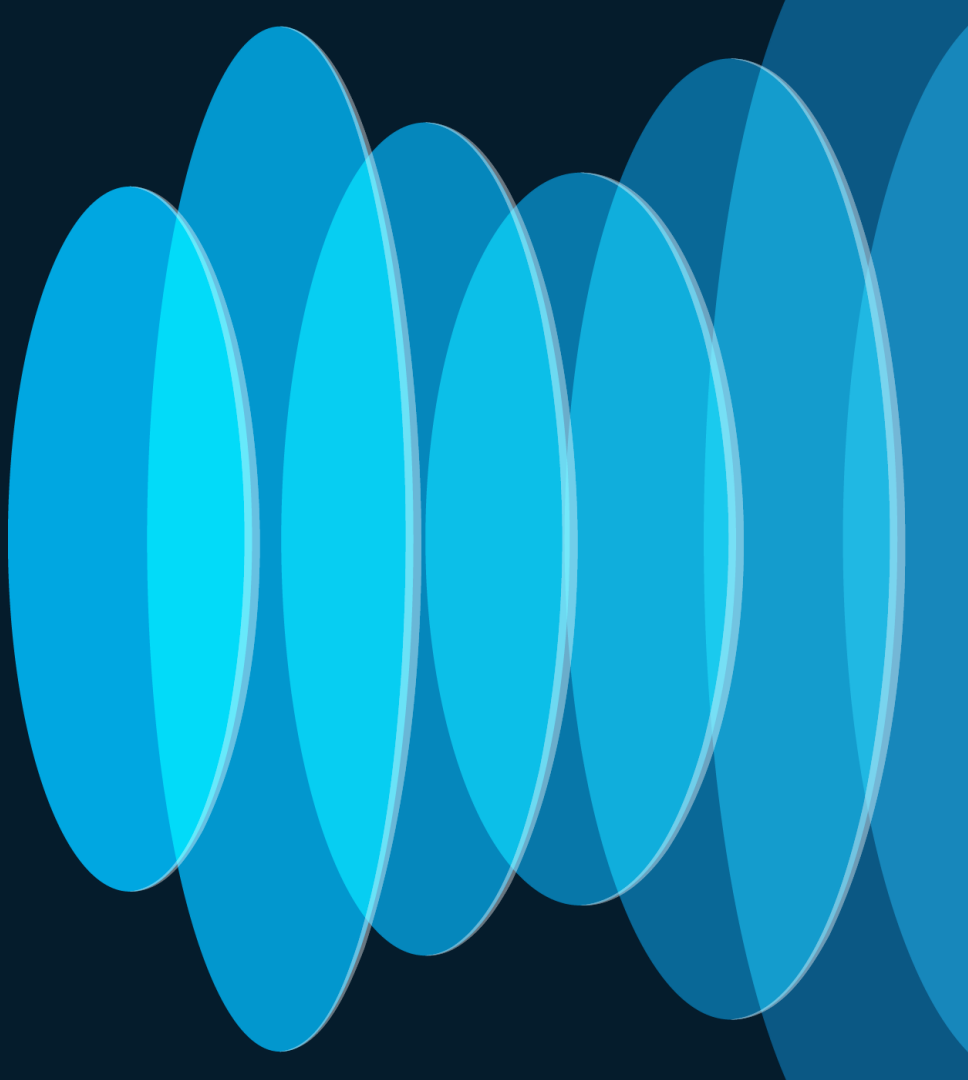


Solution

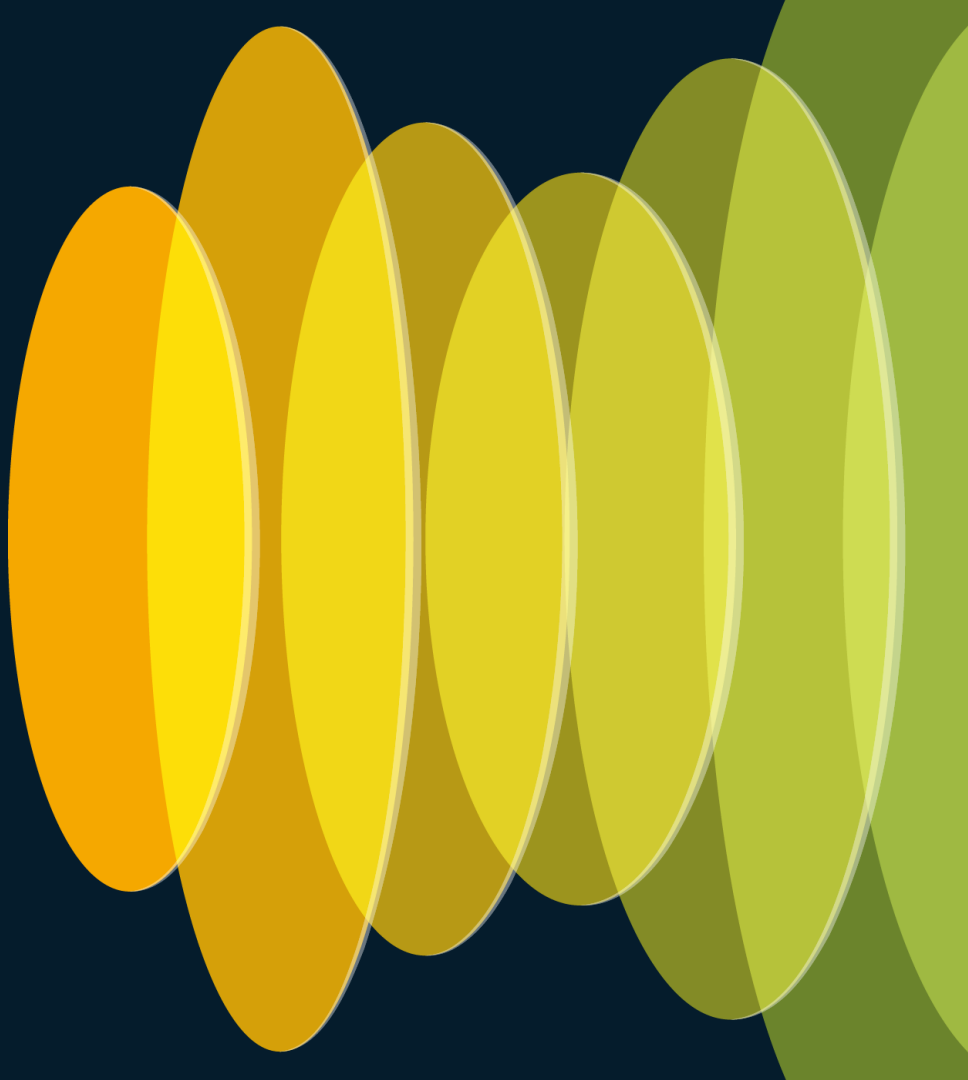
- Interconnect all Fabric Sites with a LISP Pub/Sub SD-Access Transit.
- Assign affinity IDs as shown below.
- On FS1_EBCP1, FS1_EBCP2 and FS2_EBCP1, AS PATH prepend selected routes advertised to FUSION1 and FUSION2.



Demo

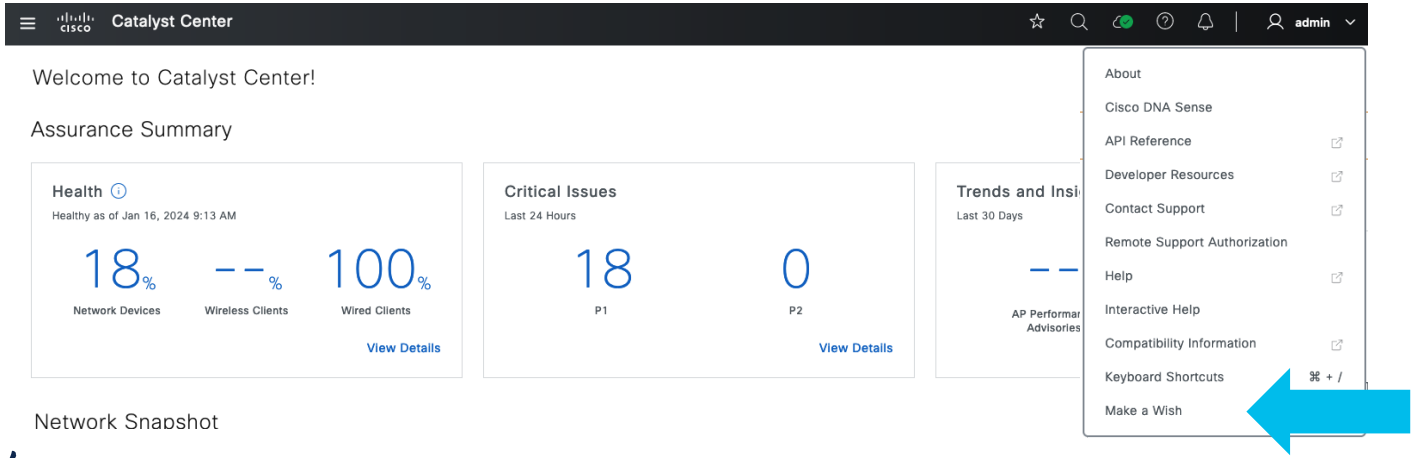


Conclusion

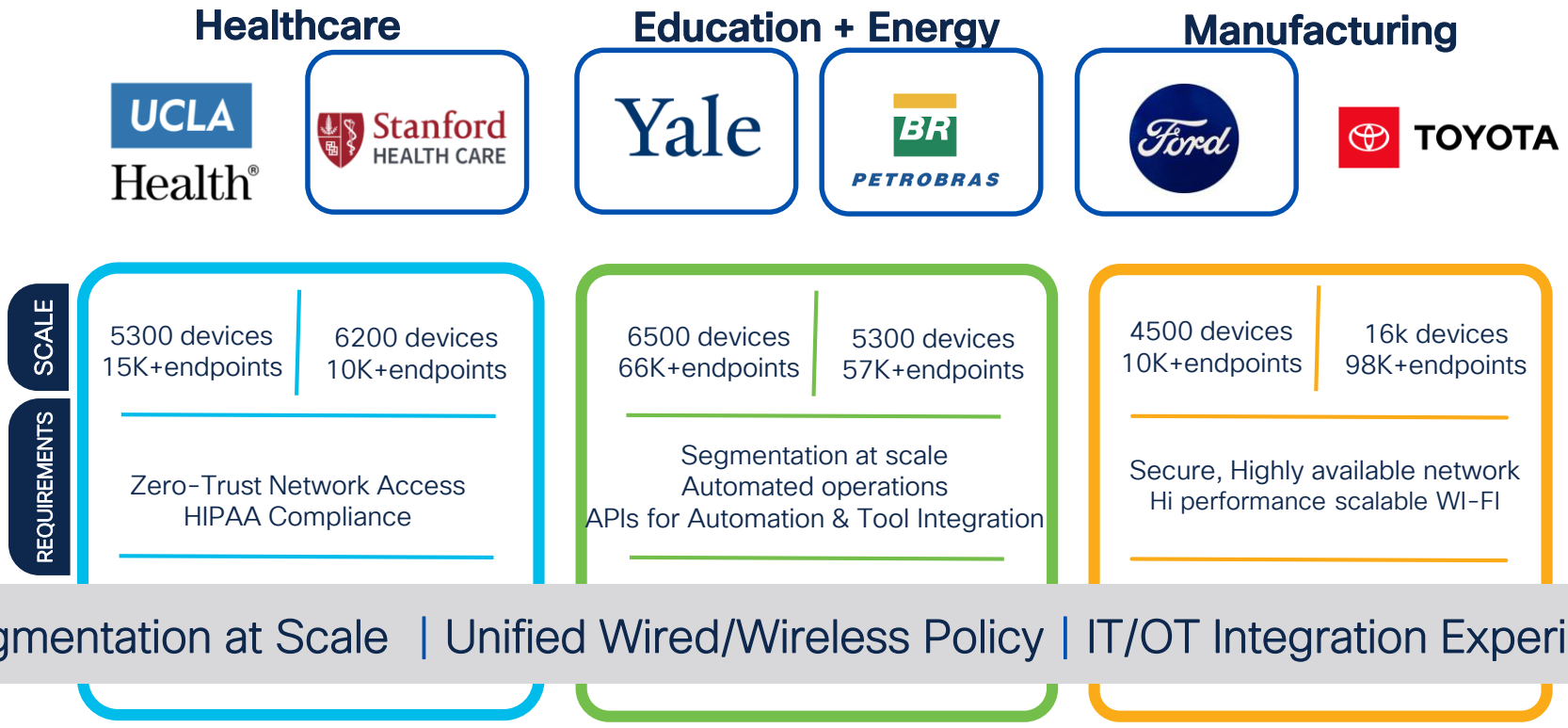


Conclusion

- SD-Access Transit simply automates VN and SGT preservation between Fabric Sites.
- SD-Access Transit offers design flexibility and can solve complex requirements at scale.
- If additional functionality is required, please “Make a Wish” in the Catalyst Center UI. Briefly describe the use case and business value or talk to your Cisco sales/services representative.



SD-Access LISP Customer Success



Segmentation at Scale | Unified Wired/Wireless Policy | IT/OT Integration Experience

BRKENS 1801, BRKENS 1802, CIUG-1003

Speaking at this Cisco Live

Complete Your Session Evaluations

Complete a minimum of 4 session surveys and the Overall Event Survey to be entered in a drawing to **win 1 of 5 full conference passes** to Cisco Live 2025.

Earn 100 points per survey completed and compete on the Cisco Live Challenge leaderboard.

Level up and earn **exclusive prizes!**

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

Continue your education

- Visit the Cisco Showcase for related demos
- Book your one-on-one Meet the Engineer meeting
- Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs
- Visit the On-Demand Library for more sessions at www.CiscoLive.com/on-demand

Contact me at: jedolphi@cisco.com

Cisco Live US SD-Access/ISE Learning Map

Sunday–2nd

- TECENS-2820 9AM
  Cisco Software-Defined Access LISP: Architecture Overview

Monday–3rd

- BRKENS-2810 8:30AM
  Cisco Software-Defined Access LISP Solution Fundamentals
- BRKENS-2800 9:30AM
  Cisco SD-Access Zero-Touch Provisioning Using LAN Automation
- BRKENS-2811 1PM
  Connecting Cisco SD-Access LISP to the World: Use Cases and Segmentation
- LTRENS-2419 1PM
  SD-Access LISP Pub/Sub Wired Lab
- BRKENS-2816 3PM
  Cisco SD-Access Transit: Advanced Design Principles
- BRKSEC-2100 10:30AM
  ISE Your Meraki Network with Group Based Adaptive Policy
- BRKENS-1802 2:30PM
  SD-Access Success Stories: Concept to Reality by Petrobras and Ford Motor
- BRKSEC-2091 3PM
  Cisco ISE Performance, Scalability and Best Practices
- BRKENS-1852 4PM
  TrustSec Refresh Reinforced with Latest Segmentation Innovations

Tuesday–4th

- BRKENS-2502 10:30AM
  Cisco SD-Access LISP VXLAN Fabric Best Practices: Design and Deployment
- BRKENS-1801 4PM
  SD-Access Success Stories: Concept to Reality by Stanford Health and Yale University

Wednesday–5th

- BRKENS-2833 10:30AM
  LISP: Optimized Control Plane for Software-Defined Access
- BRKENS-2819 2:30PM
  Cisco SD-Access and Multi-Domain Segmentation
- CIUG-1003 2:30PM
  Zero Trust with Software-Defined Access Roadmap Update
- BRKENS-2821 4:00PM
  Cisco SD-Access LISP VXLAN Fabric for Manufacturing Verticals

Thursday–6th

- BRKENS-2827 11:00AM
  Cisco SD-Access Migration Tools and Strategies

Catalyst Center

Cisco SD-Access LISP

ISE

Cisco ISE

BU-led sessions

cisco Live!

New Feature
Enhanced

Catalyst Leadership in Enterprise Networks

A Platform based Approach

Catalyst Center and Meraki Dashboard

28M Network Devices Managed

50% Y/Y

19M APs | 6M Switches | 2.5M Routers | 830M Clients

13M
Devices on Catalyst Center

15.3M
Devices on Meraki Dashboard

Catalyst 9000 Family

100,000+ Customers, Millions of Switches

"Catalyst 9K continues to be the fastest ramping product in the company's history"

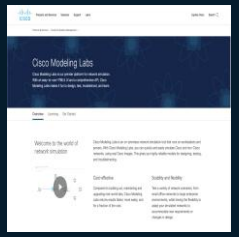
- Chuck Robbins, CEO Cisco Systems

Secure Networking	Digital Experience	Operational Simplicity
Common Policy	Campus Automation	Cloud Managed Catalyst
Secure Equipment Access	AI Endpoint Analytics	Infrastructure as a Code
SD-Access (LISP & EVPN)	Digital Experience ThousandEyes	S3 & CloudWatch Integration
High-speed Encryption	AI Ops & Assurance	Visibility, Control & Rollback

Cisco Validated Profiles (CVP)

Industry Validated Reports

Industry Certifications



Cisco Modeling Labs

cisco Live!

SD-Access LISP **Industry Leading Campus Architecture**

Deployments

4050+

Momentum

40%

YoY growth in customers

Key use case

70%

Wireless

+ 66%

API (YoY)

Usage

24K+

Sites

1.8M+

Devices

Top verticals: Government, Finance, Professional services, and Manufacturing

Adopted by 31% of U.S. **Fortune 100** Companies

EMEA: 52%

Americas 29%

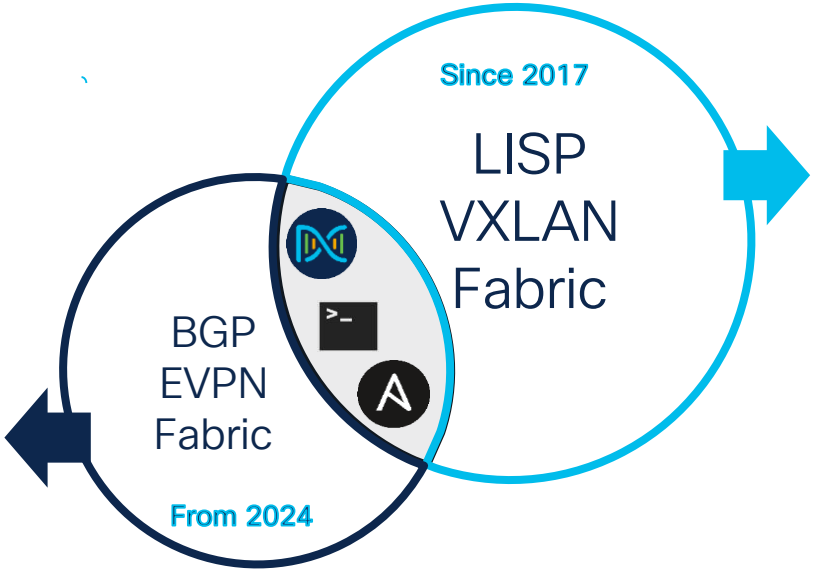
APJC 19%

Effortlessly Deploy Your **Fabric of Choice**

LISP Fabric is the **leading** choice for Enterprise customers!

Ongoing EFT

- Extend fabric across DC & Campus
- Multi-vendor deployment
- Wireless over the top
- Network segmentation



- **Simple**: Standards-based, Campus optimized BU recommended Control Plane
- **Efficient**: Lightweight, unparalleled scale & high performance due to rapid convergence time
- **Extensible**: Highly extensible to drive innovation (PubSub, Multi-site, Extranet)
- **Robust**: Integrated wireless w/ L2 mobility support campus wide

One Infrastructure | Single Data Plane | Consistent Zero-Trust Experience

Global Partner Solution Advisors (GPSA)

NEW - Fully Virtualized, SD-Access Secure Campus Lab

Virtualized SD-Access Lab

- Fully Customizable Topology with virtualized 9kv's and 8kv's
- Access on dCloud or build on your existing Data Center
- Fraction of the cost
- GPSA mentored lab buildout support available!

CTF Mission at Cisco Live

- Experience the SD-Access Virtual Lab at Capture the Flag in The World of Solutions
- Use Cases – Fabric Sites and Virtual Network Provisioning, Fusion Automation, Extranet, Micro Segmentation, and more!

For More Information

- Visit us at the Global Partner Experience booth (4227) across from Capture the Flag!
- Or Reach out to us: gpsa_for_partners@cisco.com
- GPSA is Your source for no-cost, Cisco partner SD-Access Mentorship!

Virtual SD-Access Lab on dCloud

GPSA Sales Connect Page

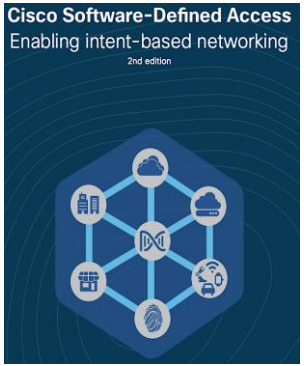
CTF at Cisco Live
Check out Secure Campus Section

cisco Live!

Cisco SD-Access LISP Collaterals

[Cisco Software-Defined Access for Industry Verticals]

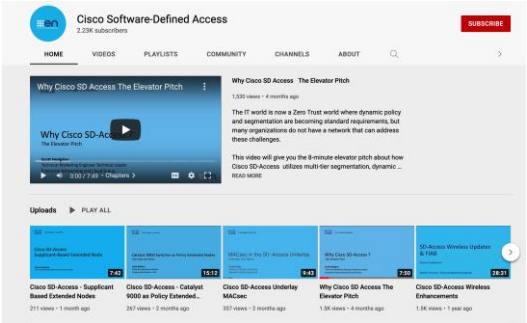
[Cisco Software-Defined Access]
Enabling intent-based networking



[Cisco Solution Validated Profiles (CVPs)]

- [Cisco Large Enterprise and Government Profile]
- [Healthcare Vertical]
- [Financial Vertical]
- [Healthcare Vertical]
- [Manufacturing Vertical]
- [Retail Vertical]
- [University Vertical]

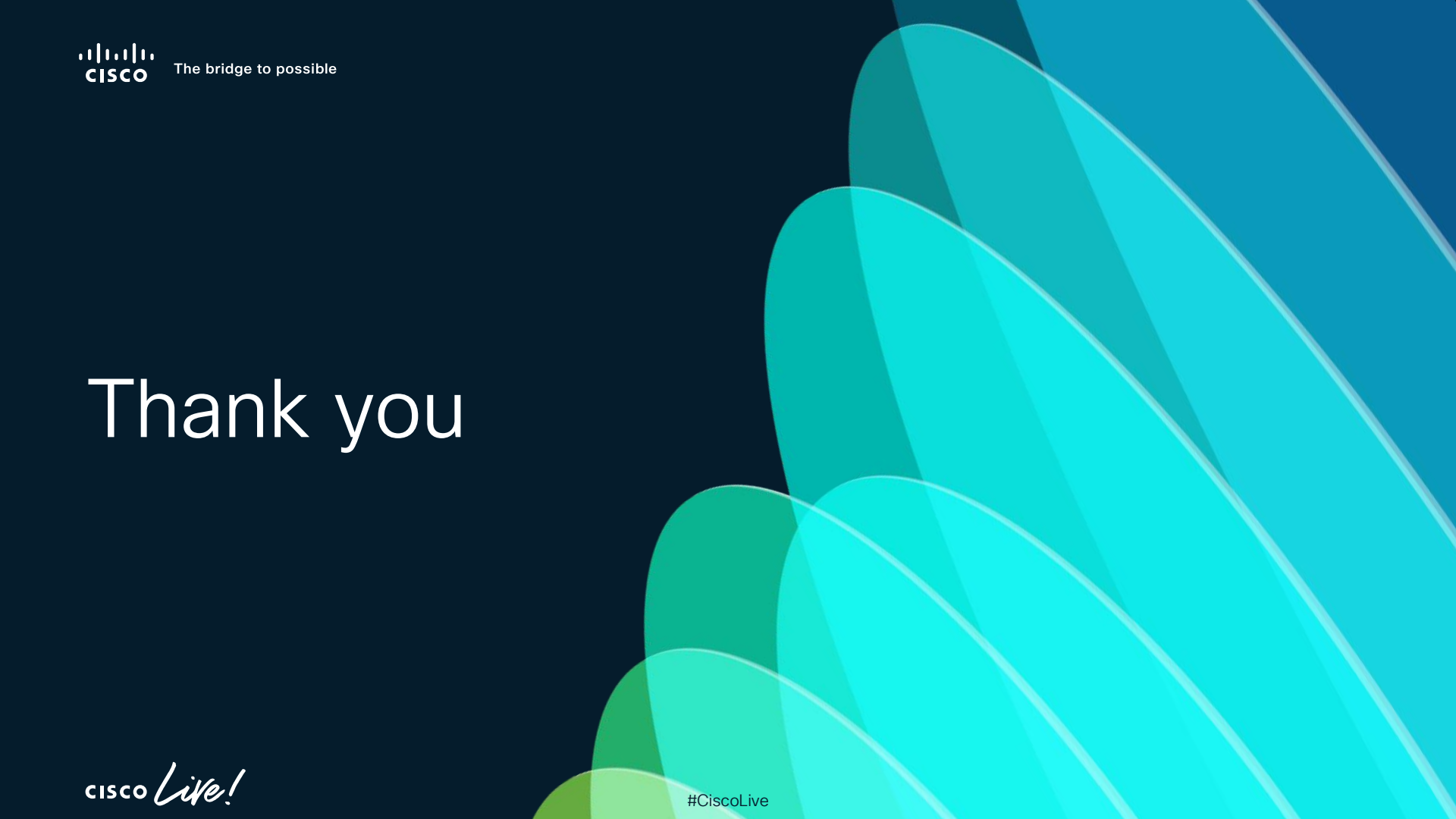
[Cisco SD-Access YouTube Link]



[Cisco SD-Access Design Tool]

[EN&C Validated Designs]

[*The Latest SD-Access Guides*]


The bridge to possible
Thank you
cisco Live!
#CiscoLive