Multicast with EVPN, Segment Routing, and Traffic Engineering

Multicast deep dive for EVPN and Segment Routing

Mankamana Mishra
Technical Leader

BRKMPL-1123

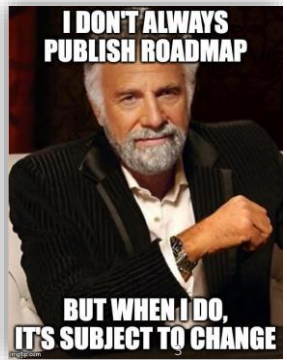
cisco Live!

Agenda

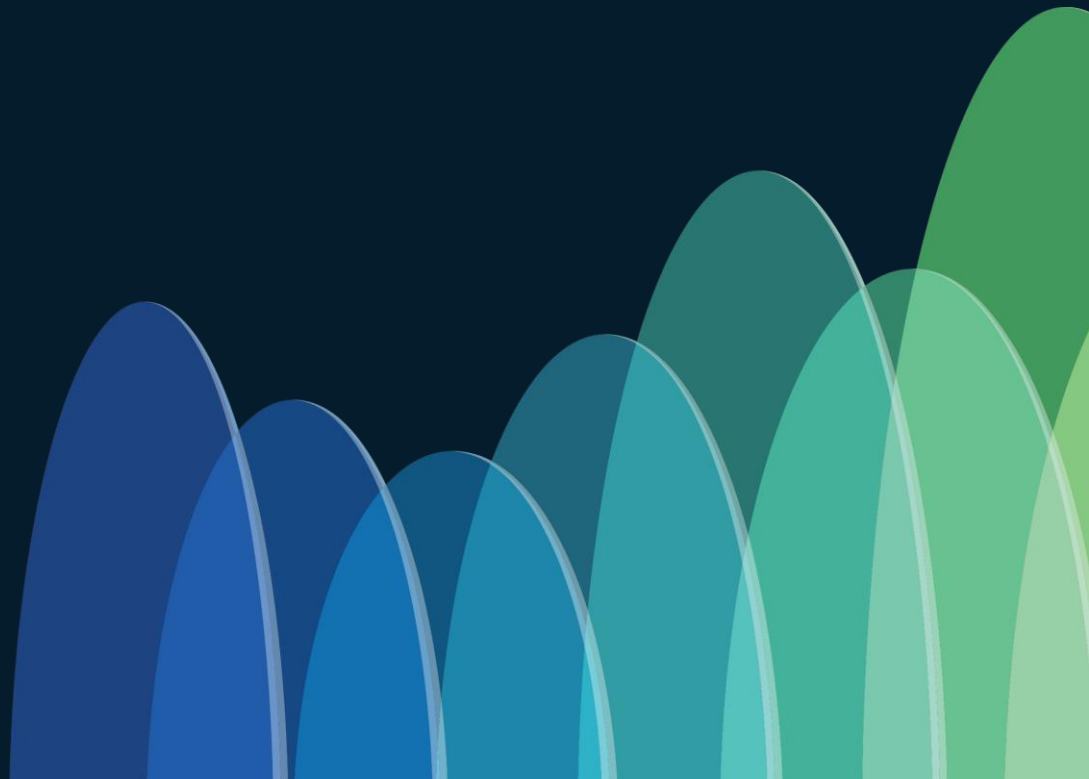
- Introduction
- Multicast Basics
- Multicast service over EVPN network
- Multicast with SRv6
- Conclusion

Disclaimer

- *Many of the products and features described herein remain in varying stages of development and will be offered on a when-and-if-available basis.*
- *This roadmap is subject to change at the sole discretion of Cisco, and Cisco will have no liability for delay in the delivery or failure to deliver any of the products or features set forth in this document.*



What to Expect in next 90 minutes

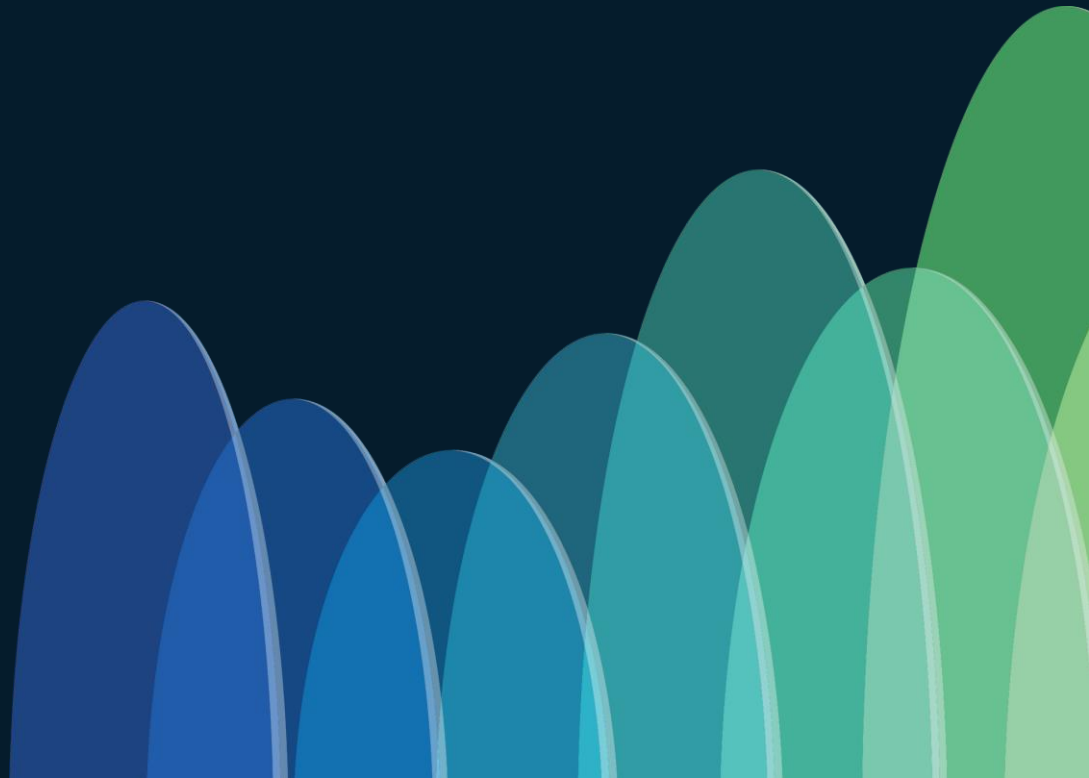


cisco Live!

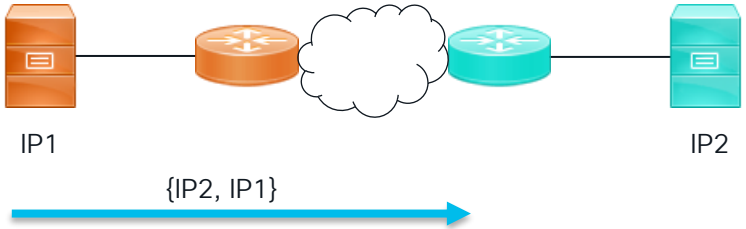
What to Expect from This Session

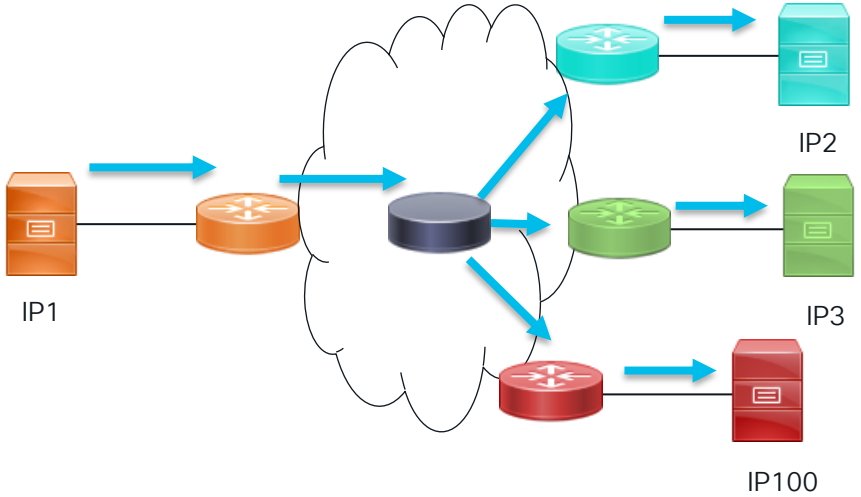
1. Designed for **beginners** to understand EVPN BUM traffic flow.
2. Discussion on the **fundamental building blocks** of BUM traffic in an EVPN network.
3. Exploration of **various optimization techniques** to improve network efficiency.
4. Insights into **applying key EVPN functions** for better network design decisions.
5. **Second part** of the presentation will cover **multicast support** when migrating to **SRv6**.

Introduction



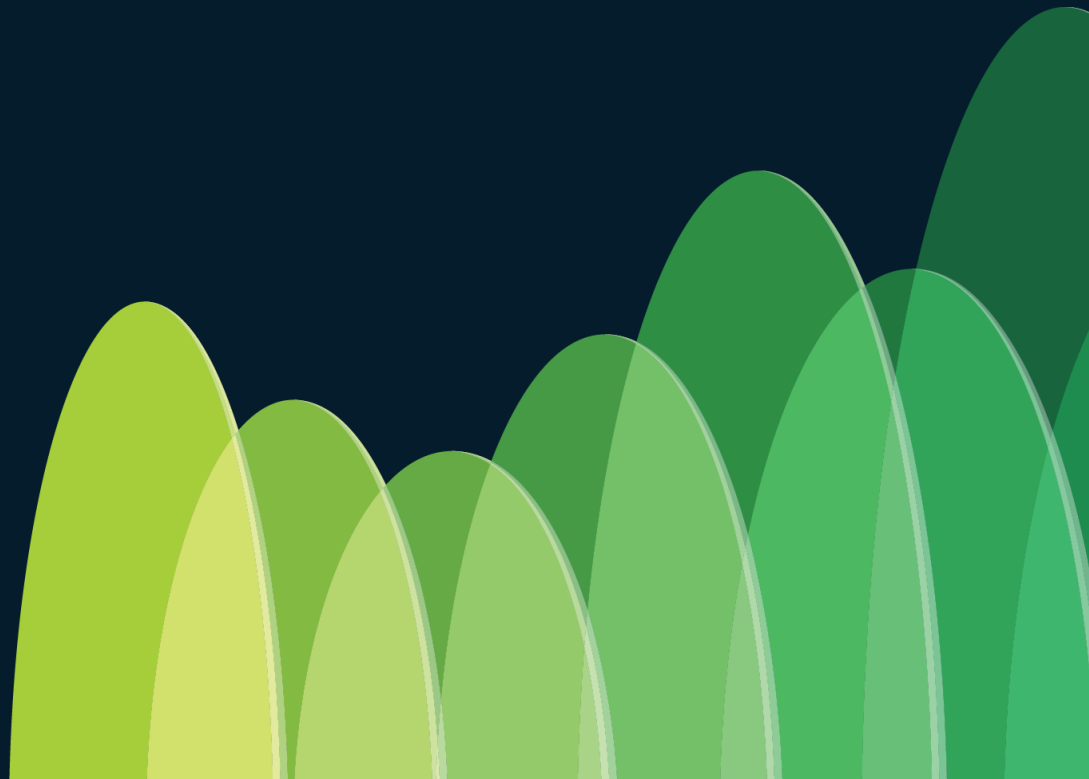
Why Multicast ?





- If total number of end point increase significantly, there is need for replication in network, and it brings requirement for multicast

Multicast Deployments use cases

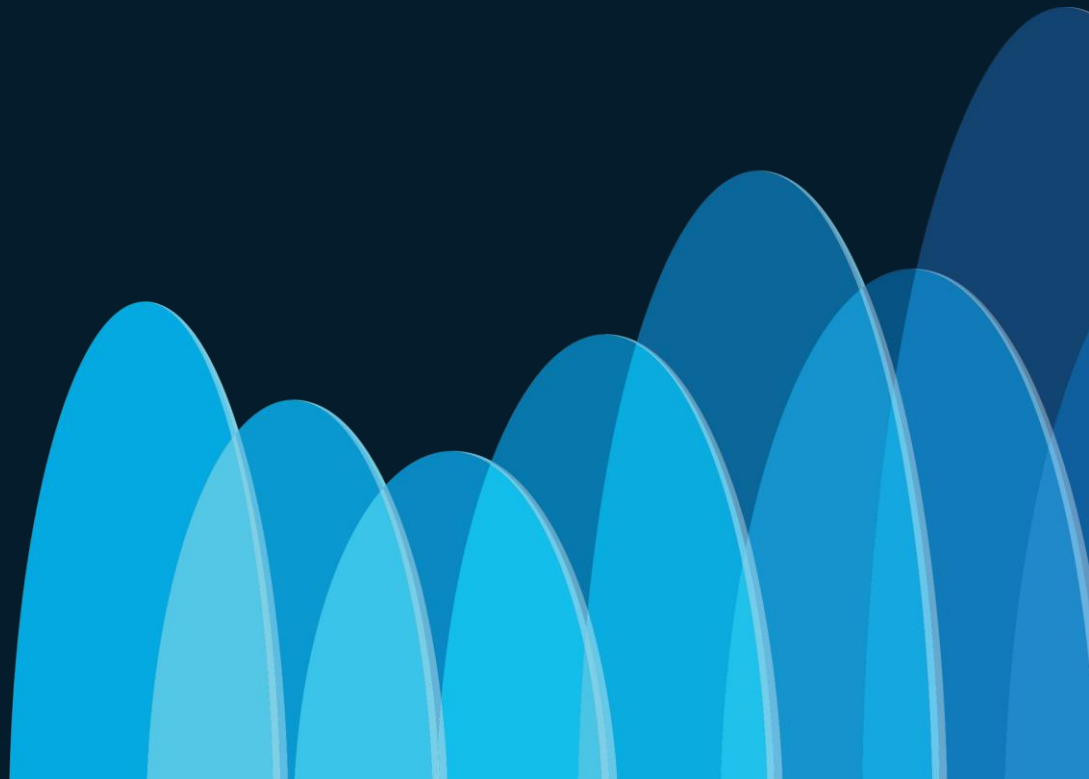


cisco Live!

Major IP Multicast use case

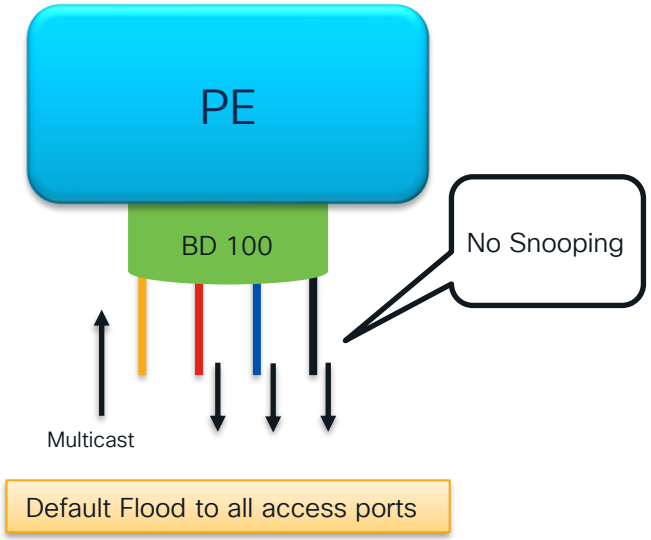
- Media Streaming
  - IPTV, Broadcast channels
- Business Multicast VPN
  - Service provider delivering multicast traffic for enterprise
- Financial data delivery
  - Financial data delivery
  - Strict SLA
- Surveillance
  - Airports, Enterprises, City and more .

IGMP Snooping basics

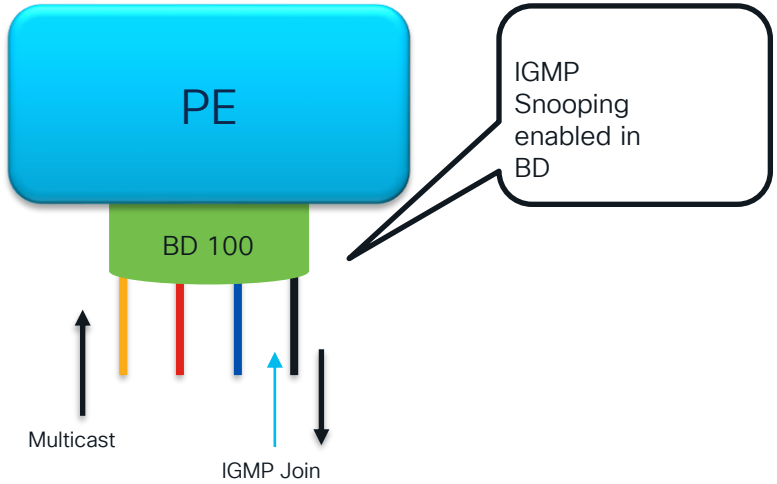


cisco Live!

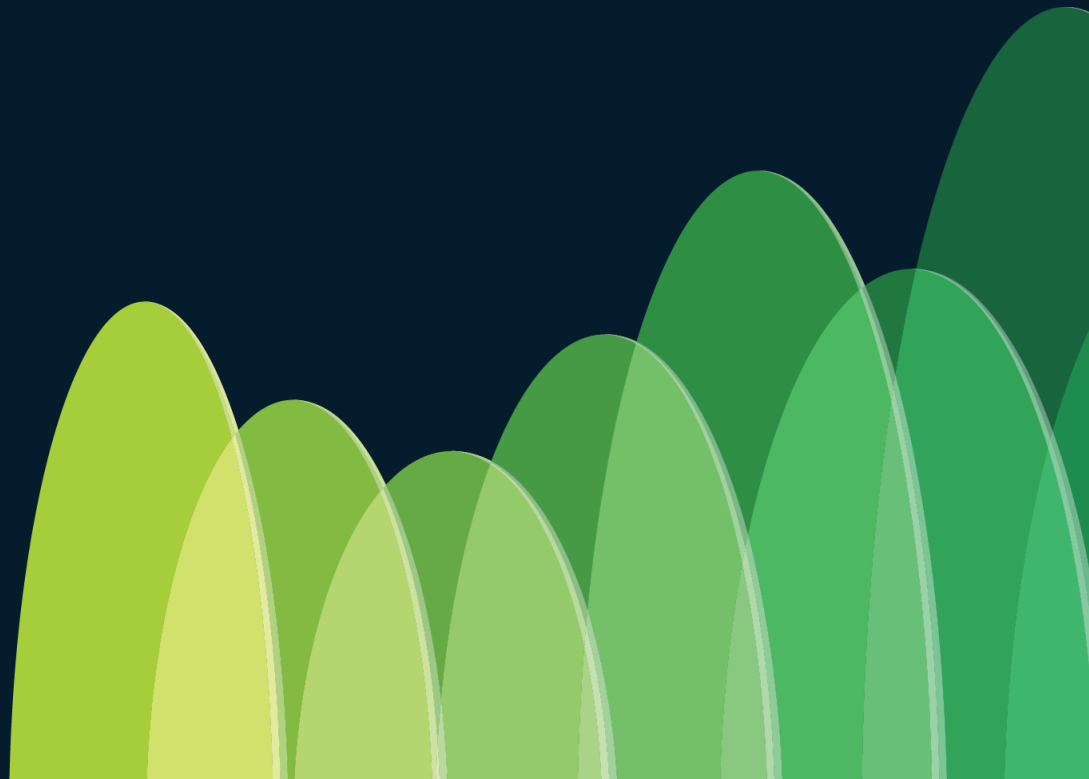
IGMP Snooping Basic



IGMP Snooping Basic



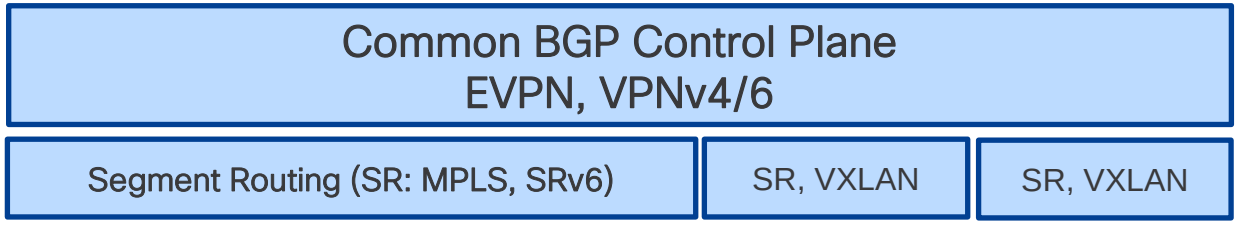
EVPN Basics

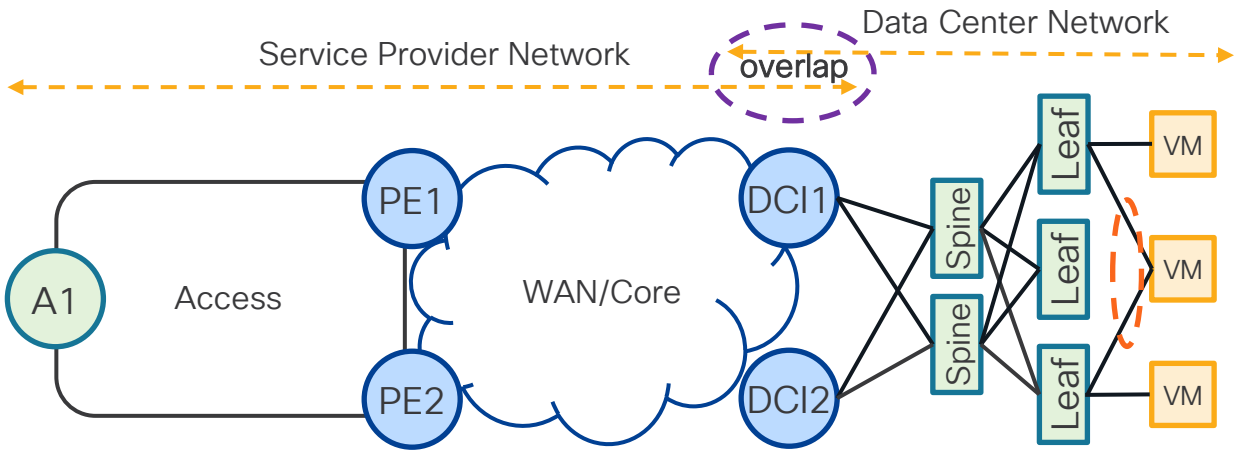


cisco Live!

Unified Control Plane and Data Plane

Next Generation Services Overlay & Data Plane



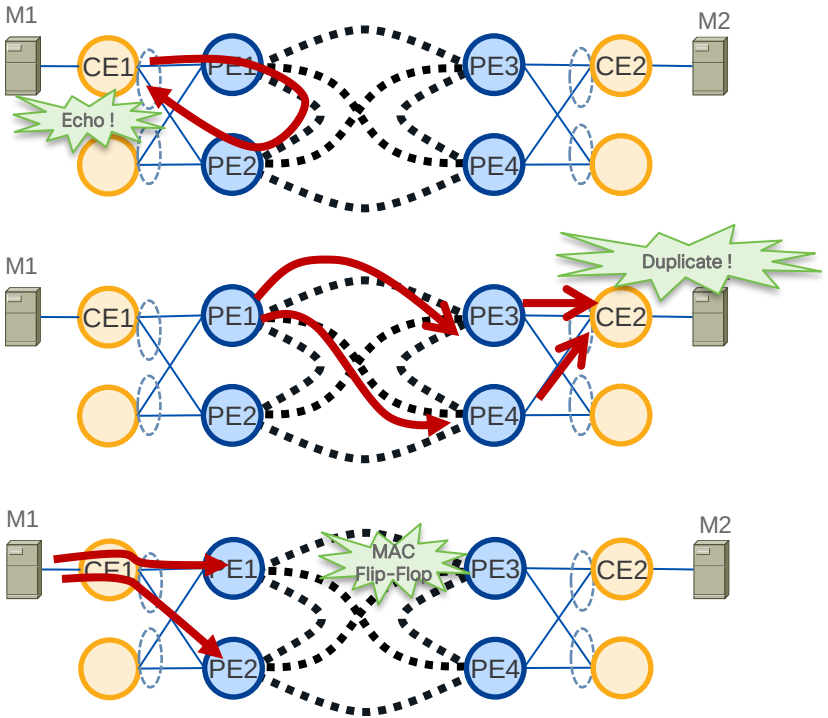


Legacy Solution:

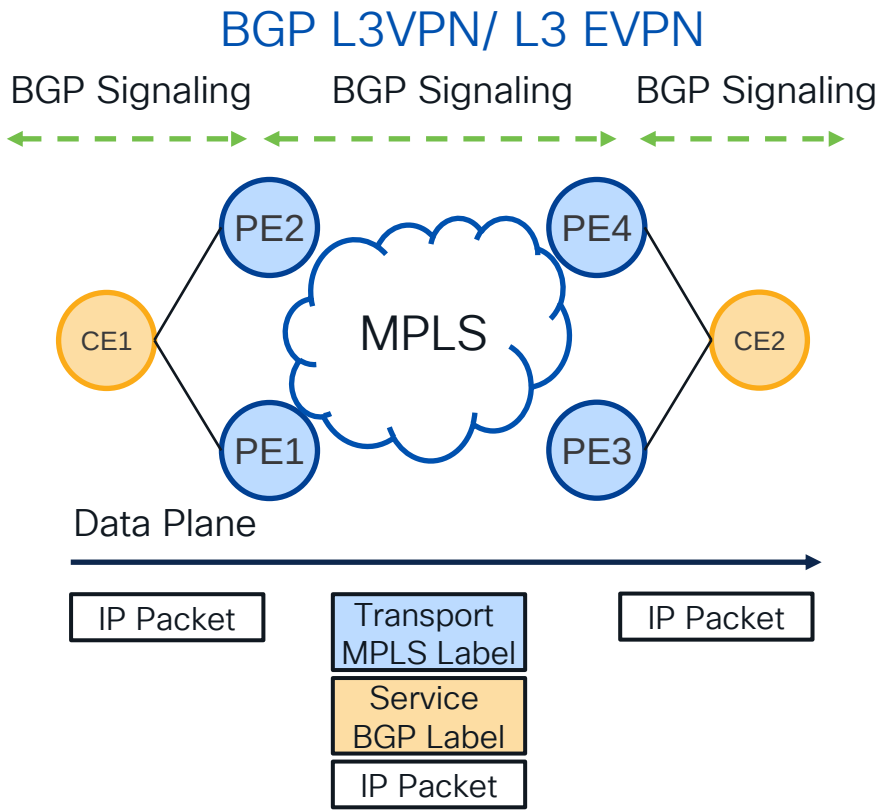
Next-Generation Solutions for L2VPN

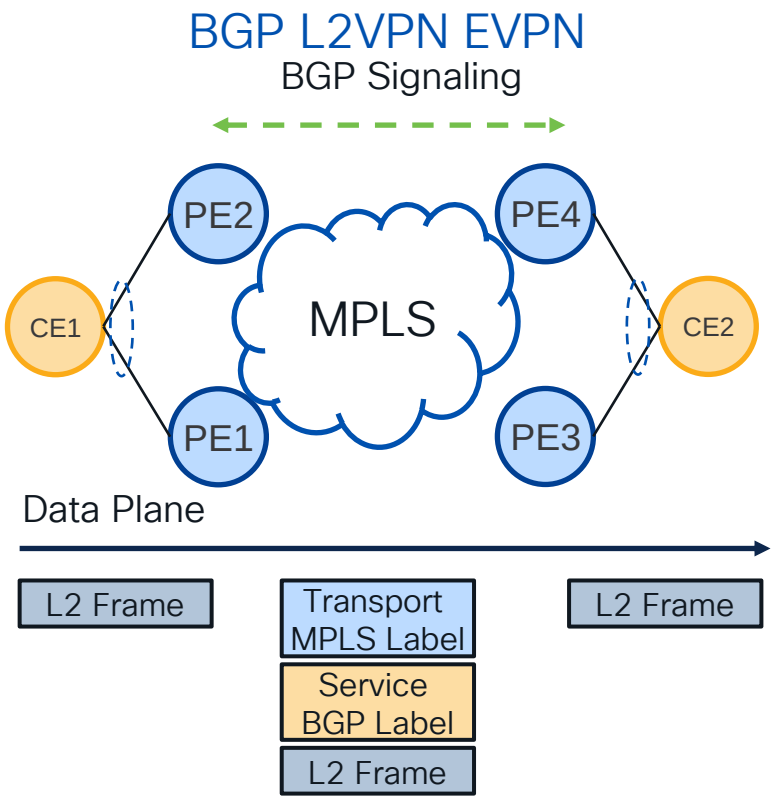
Solving VPLS challenges for per-flow Redundancy

- Existing VPLS solutions do not offer an All-Active per-flow redundancy
- Looping of Traffic Flooded from PE
- Duplicate Frames from Floods from the Core
- MAC Flip-Flopping over Pseudowire
  - E.g. Port-Channel Load-Balancing does not produce a consistent hash-value for a frame with the same source MAC (e.g. non MAC based Hash-Schemes)



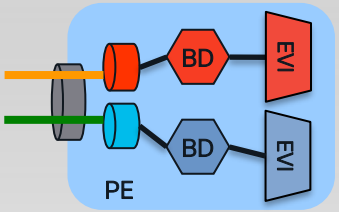
MPLS Transport & BGP Service





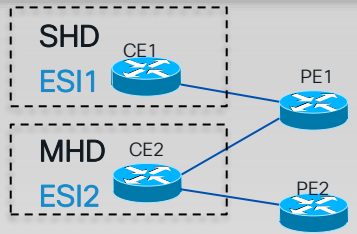
Basic Concepts EVPN

EVPN Instance (EVI)



- EVI identifies a VPN in the network
- Encompass one or more bridge-domains, depending on service interface type
  - Port-based
  - VLAN-based (shown above)
  - VLAN-bundling

Ethernet Segment



- Represents a 'site' connected to one or more PEs
- Uniquely identified by a 10-byte global Ethernet Segment Identifier (ESI)
- Could be a single device or an entire network
  - Single-Homed Device (SHD)
  - Multi-Homed Device (MHD)
  - Single-Homed Network (SHN)
  - Multi-Homed Network (MHN)

BGP Routes

Route Types
[1] Ethernet Auto-Discovery (AD) Route
[2] MAC/IP Advertisement Route
[3] Inclusive Multicast Route
[4] Ethernet Segment Route
[5] IP Prefix Advertisement Route

- New SAFI [70]
- Routes serve control plane purposes, including:
  - MAC address reachability
  - MAC mass withdrawal
  - Split-Horizon label adv.
  - Aliasing
  - Multicast endpoint discovery
  - Redundancy group discovery
  - Designated forwarder election
  - IP address reachability
  - L2/L3 Integration

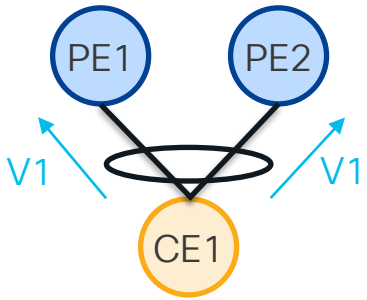
BGP Route Attributes

Extended Communities
ESI MPLS Label
ES-Import
MAC Mobility
Default Gateway
Encapsulation

- New BGP extended communities defined
- Expand information carried in BGP routes, including:
  - MAC address moves
  - Redundancy mode
  - MAC / IP bindings of a GW
  - Split-horizon label encoding
  - Data plane Encapsulation

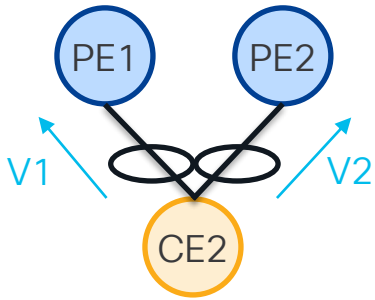
EVPN - Load-Balancing Modes

All-Active (per flow)



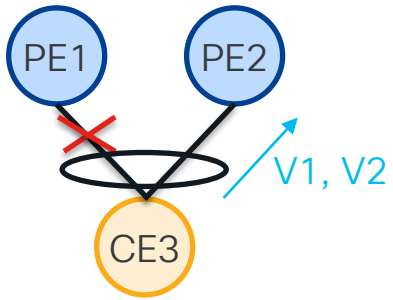
Single LAG at the CE
VLAN goes to both PE
Traffic hashed per flow
Benefits: Bandwidth, Convergence

Single-Active (per VLAN)



Multiple LAGs at the CE
VLAN active on single PE
Traffic hashed per VLAN
Benefits: Billing, Policing

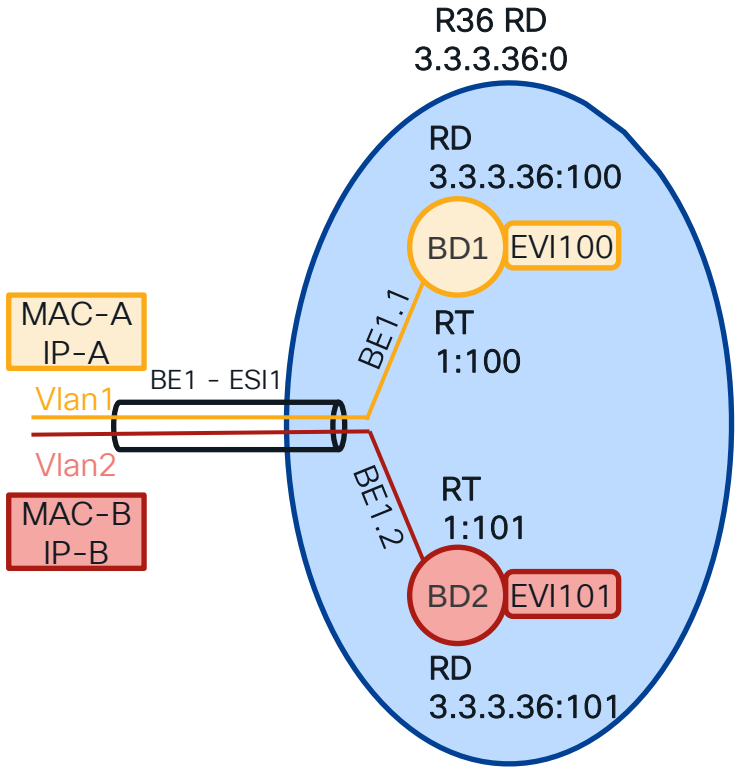
Port-Active (per port)



Single LAGs at the CE
Port active on single PE
Traffic hashed per port
Benefits: Protocol Simplification

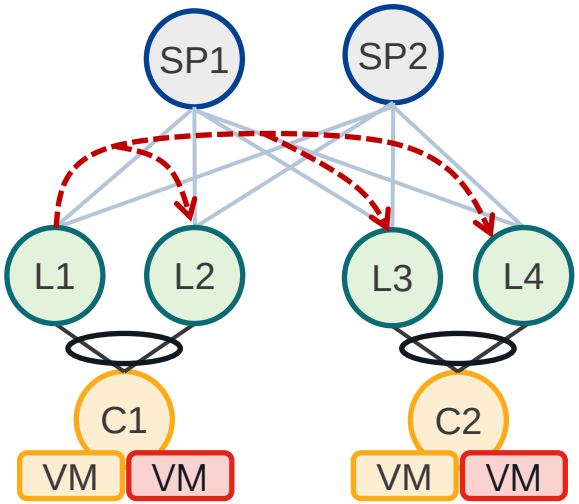
EVPN - Instance

- EVPN Multipoint (ELAN) Instance is identified by **local unique EVI ID**
- Best Practice is to Autogenerate Route-Target (RT) => **EVI ID becomes global unique per EVPN Instance**
  - **IOS-XR Autogenerates RT by default**

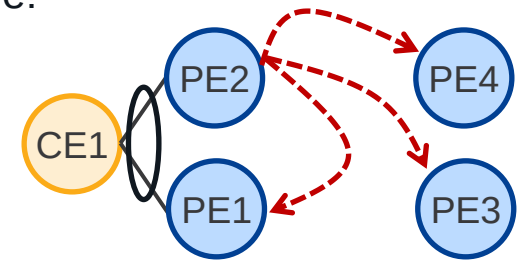


EVPN – Ethernet VPN

- Concepts are same!!! Pick your side!

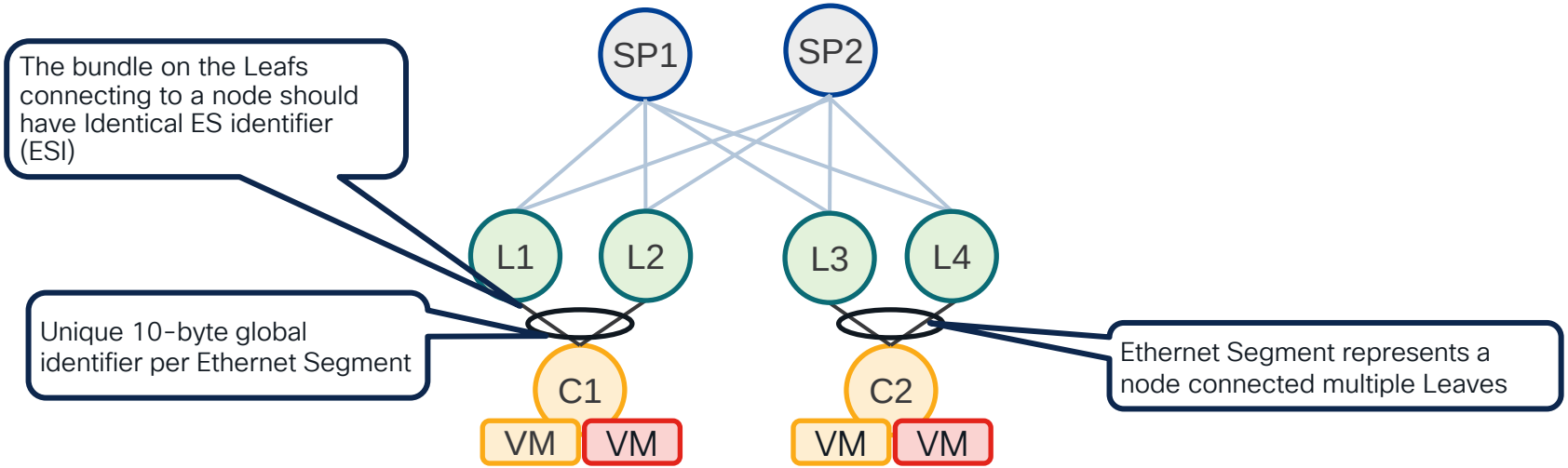


Pick your side!



EVPN – Ethernet-Segment for Multi-Homing

L1 and L2 (L3 and L4) have to know if they multi-home same broadcast domain



EVPN Configuration

CE has to receive same lacp system MAC

```
lacp system mac 3637.3637.3637

interface Bundle-Ether100
 l2transport
 !
!

evpn
 evi 100
  advertise-mac
 !
 interface Bundle-Ether100
 ethernet-segment
  identifier type 0 36.37.00.00.00.00.00.11.00
 !
!
```

RT-2 MAC advertise

```
l2vpn
 bridge group 100
 bridge-domain 100
  interface Bundle-Ether100
  !
  evi 100
  !
 !
 !
!
```

EVPN Configuration – BGP

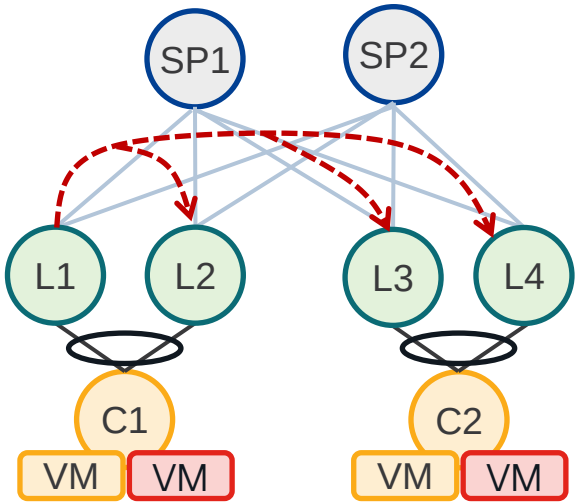
```
router bgp 1
 bgp router-id 3.3.3.36
 address-family l2vpn evpn
 !
 neighbor-group rr
 remote-as 1
 update-source Loopback0
 address-family l2vpn evpn
 !
 neighbor 3.3.3.103
 use neighbor-group rr
 !
 neighbor 3.3.3.104
 use neighbor-group rr
 !
!
```

EVPN Instance View

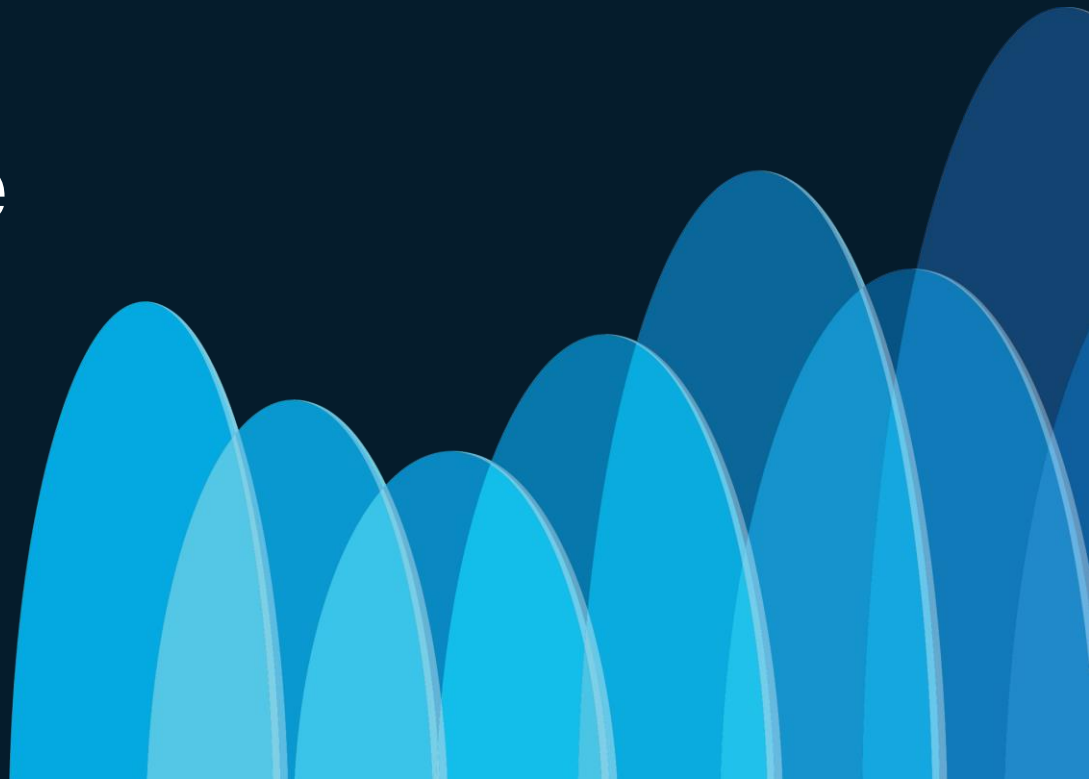
```
L1#show evpn evi vpn-id 100 detail

VPN-ID     Encap  Bridge Domain                 Type
---------- ------ ----------------------------- -------------------
100        MPLS   100                           EVPN
   Stitching: Regular
   Unicast Label  : 68096
   Multicast Label: 64000
   Flow Label: N
   Control-Word: Enabled
   Forward-class: 0
   Advertise MACs: Yes
   Advertise BVI MACs: No
   Aliasing: Enabled
   UUF: Enabled
   Re-origination: Enabled
   Multicast source connected: No

   Statistics:
     Packets            Sent                Received
       Total          : 0                   0
       Unicast        : 0                   0
       BUM            : 0                   0
     Bytes              Sent                Received
       Total          : 0                   0
       Unicast        : 0                   0
       BUM            : 0                   0
   RD Config: none
   RD Auto  : (auto) 3.3.3.36:100
   RT Auto  : 1:100
   Route Targets in Use           Type
   ------------------------------ --------------------
   1:100                          Import
   1:100                          Export
```

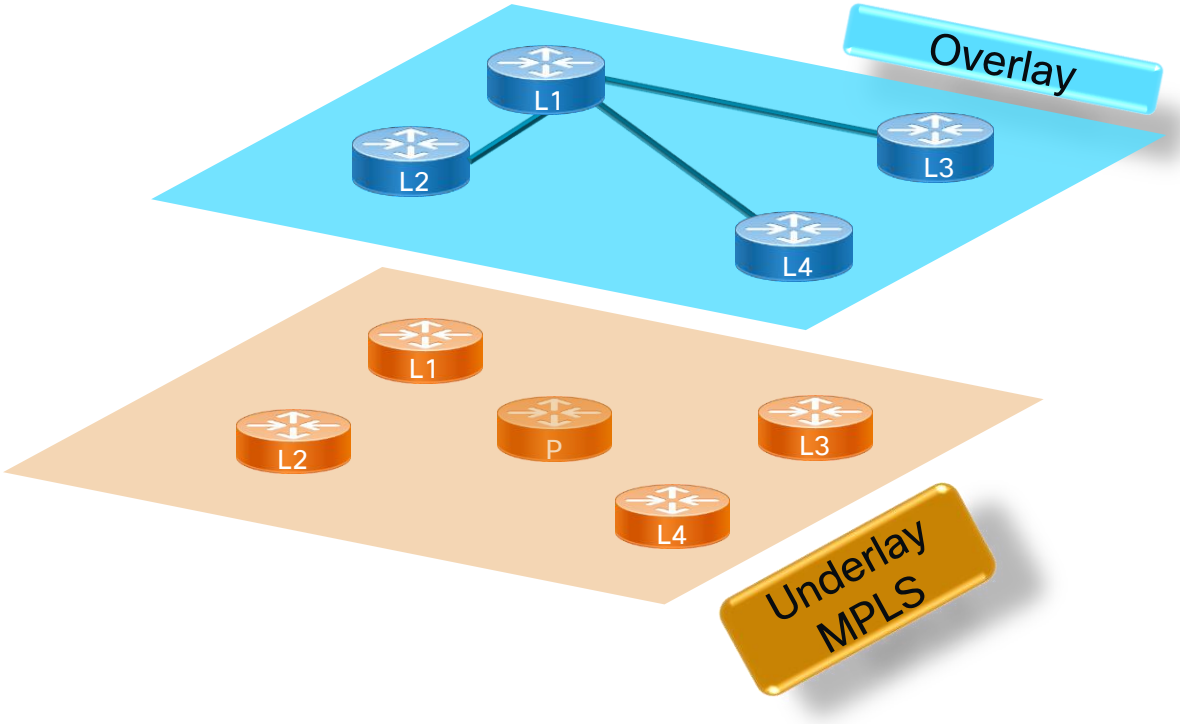


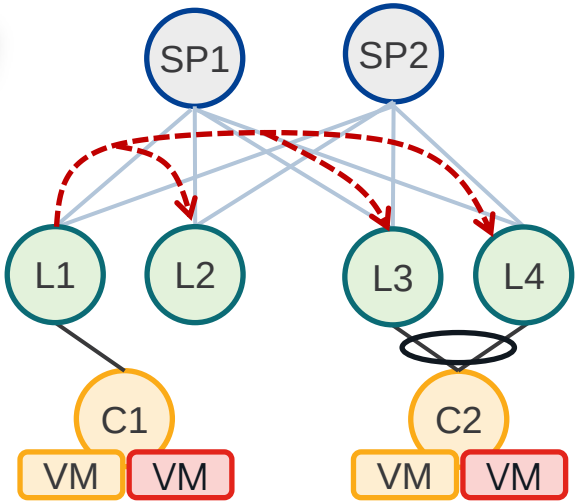
Multicast service over EVPN



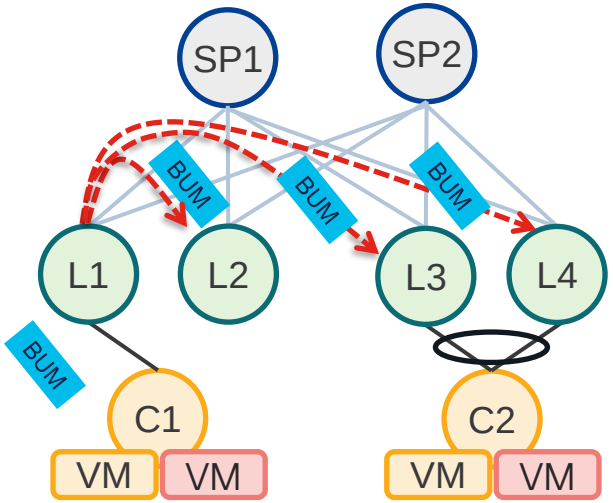
cisco Live!

Operator network view

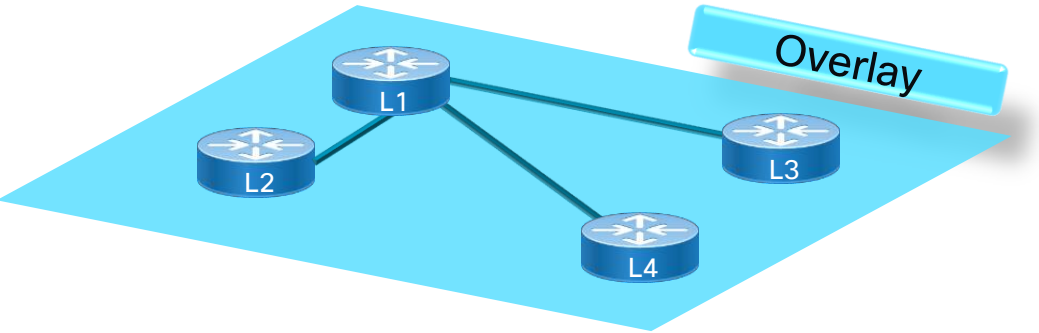




EVPN BUM Ingress Replication



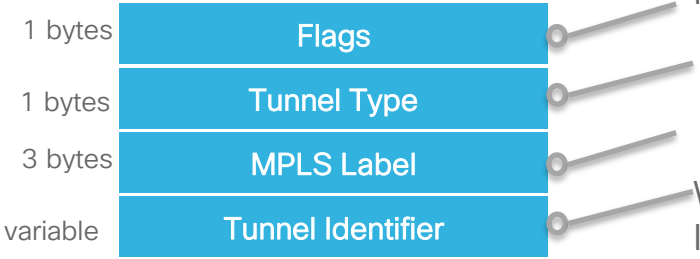
EVPN BGP – Inclusive Multicast Route 0x3



8 bytes	RD	Unique per Advertising PE per EVI
4 bytes	Ethernet Tag ID	Set to VLAN or I-SID for VLAN-Aware Bundling Service interface, otherwise 0
1 bytes	IP Address Length	IP address length
4 or 16 bytes	Originating Router's IP add.	IPv4 or IPv6 address

Route Type specific encoding of E-VPN NLRI

PMSI Tunnel Attribute – RFC6514



Route Type specific encoding of E-VPN NLRI

Flags based on RFC6514

Ingress Replication/mLDP etc.

Multicast MPLS Label

When the Tunnel Type is set to Ingress Replication, the Tunnel Identifier carries the unicast tunnel endpoint IP address of the local PE that is to be this PE's receiving endpoint address for the tunnel.

BUM Tunnel setup post EVPN base config

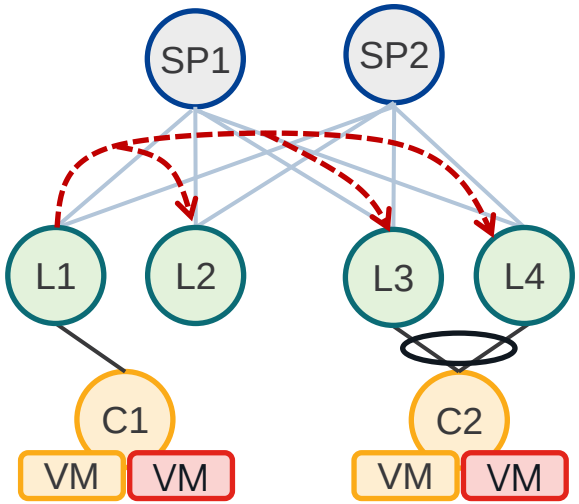
```
RP/0/0/CPU0:L1#show bgp l2vpn evpn rd  10.1.1.1:4 [3][0][32][10.1.1.1]/80
Mon Feb  3 18:54:25.495 PST
BGP routing table entry for [3][0][32][10.1.1.1]/80, Route Distinguisher: 10.1.1.1:4
Versions:
  Process           bRIB/RIB   SendTblVer
  Speaker                19           19
Last Modified: Feb  3 16:25:27.000 for 02:28:58
Paths: (1 available, best #1)
  Advertised to update-groups (with more than one peer):
    0.2
  Path #1: Received by speaker 0
  Advertised to update-groups (with more than one peer):
    0.2
  Local
    0.0.0.0 from 0.0.0.0 (10.1.1.1)
      Origin IGP, localpref 100, valid, redistributed, best, group-best, import-candidate
      Received Path ID 0, Local Path ID 1, version 19
      Extended community: EVPN L2 ATTRS:0x04:0 RT:100:4
      PMSI: flags 0x00, type 6, label 24003, ID 0xc0a80005
```

RT-3

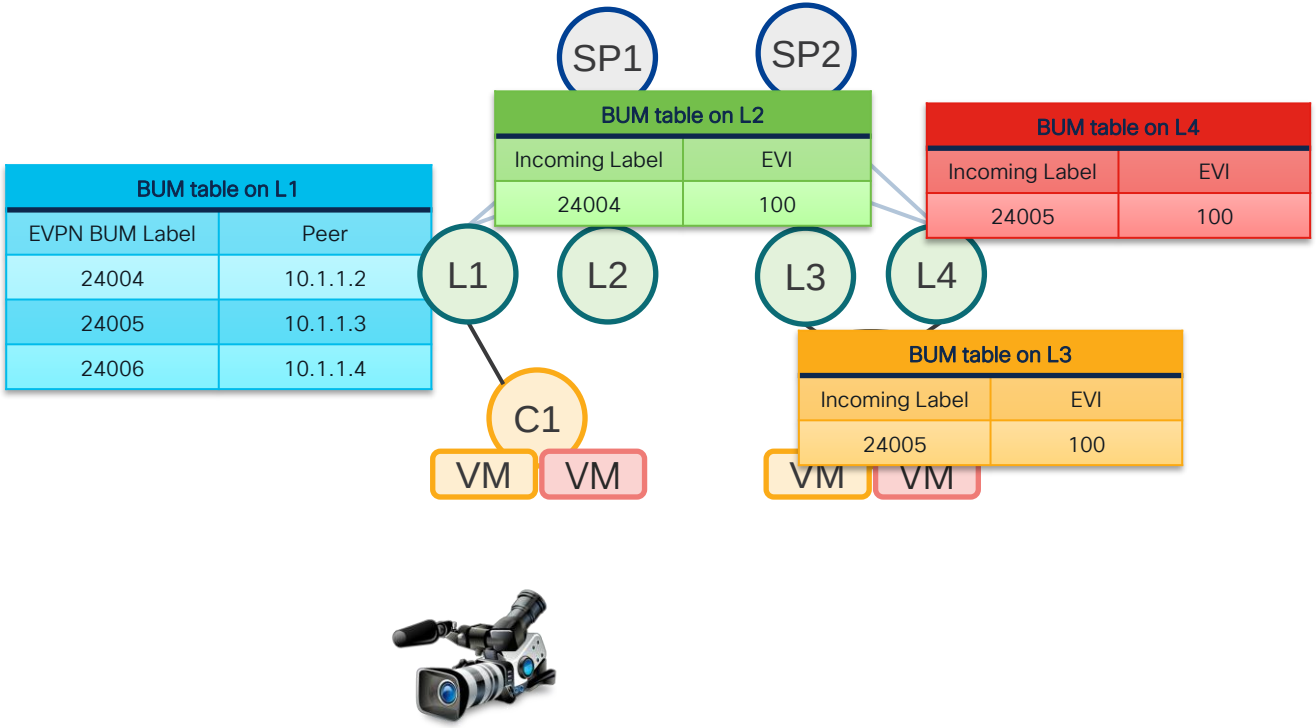
Ingress Replication

Multicast (BUM) Label

EVI 4 Route-Target

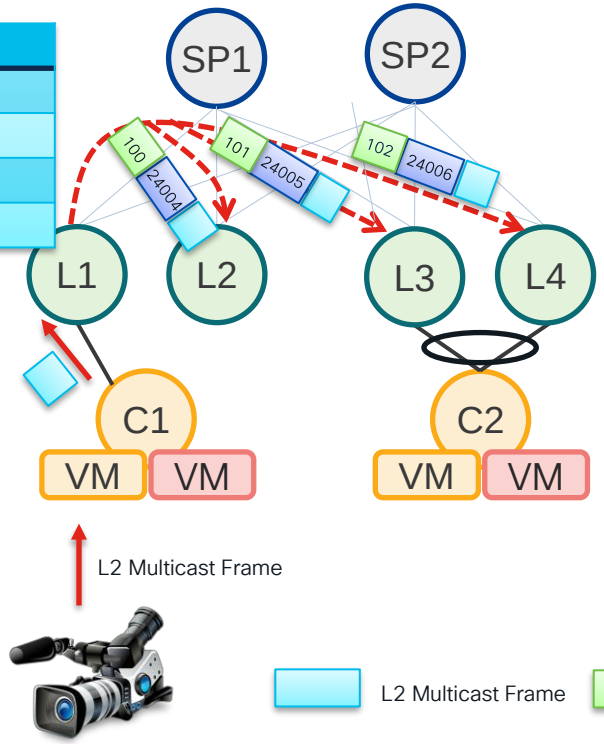


Overlay Control plane Setup



Multicast Data plane

BUM	
EVPN BUM Label	Peer
24004	10.1.1.2
24005	10.1.1.3
24006	10.1.1.4

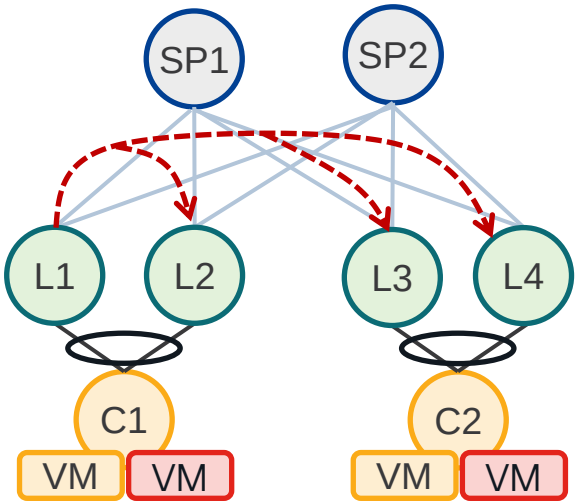


L2 Multicast Frame Transport Label BUM Label

Single Home recap with Ingress replication

1. **EVPN Provisioning** initiates the origination of **EVPN Route Type 3 (RT-3)**.
2. **EVPN Route Type 3 (RT-3)** facilitates the establishment of an **Ingress Replication BUM (Broadcast, Unknown Unicast, Multicast) tunnel**.
3. By default, **BUM traffic** received on the network is **flooded to all peers**.

EVPN L2 All-Active Multihomed Service




```
interface Bundle-Ether100
 l2transport
 !
!

evpn
 evi 100
  advertise-mac
 !
  interface Bundle-Ether100
 ethernet-segment
  identifier type 0 36.37.00.00.00.00.00.11.00
 !
```

```
l2vpn
 bridge group 100
 bridge-domain 100
  interface Bundle-Ether100
  !
  evi 100
  !
 !
 !
!
```

EVPN – Designated Forwarder (DF)

Challenge:

How to prevent duplicate copies of flooded traffic from being delivered to a multi-homed Ethernet Segment?

If (L3 and L4) Multi-Homing access via same Ethernet Segment -> only one of them can forward traffic to access

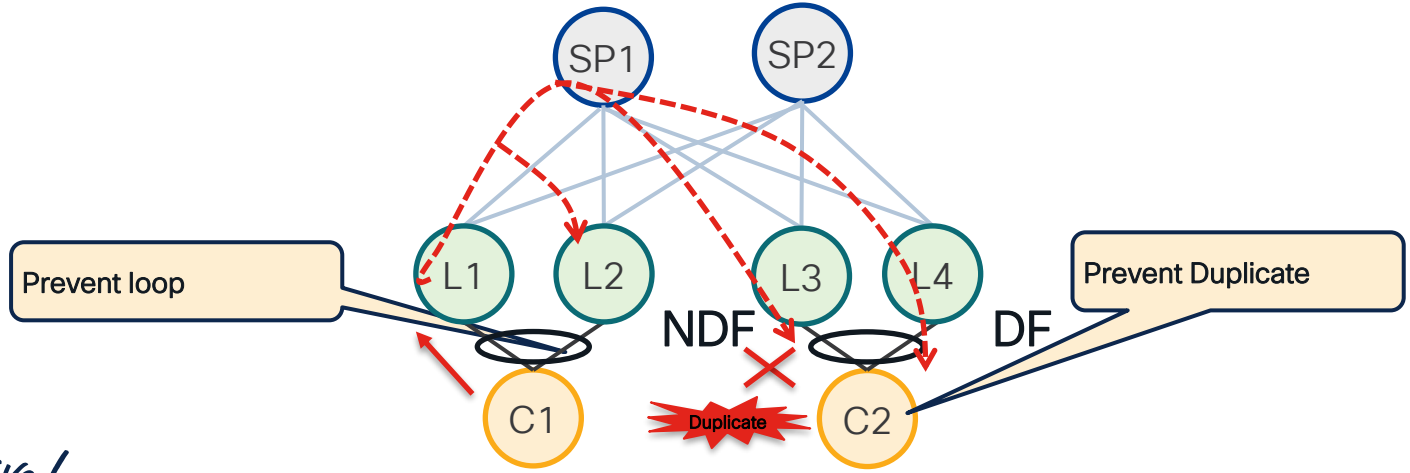
Same for (L1 and L2)

Why extra BUM Label?

What if Unicast Traffic is sent to L3 or L4 (not flooded)? -> DF Election applies only to BUM (from Core to Access)

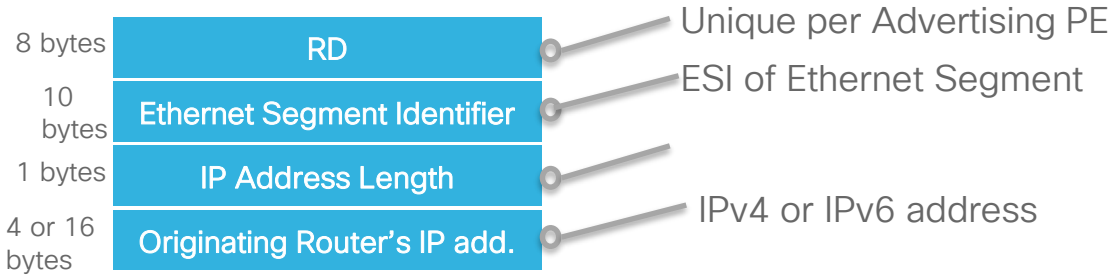
DF, Redirect, Fast Re-Route (FRR), etc.

Service Label informs egress Leaf if traffic is BUM or Unicast



EVPN BGP – Ethernet Segment Route 0x4

- Usage:
  - Auto-discovery of multi-homed Ethernet Segments
  - Designated Forwarder election
- Tagged with ES-Import Extended Community
  - PEs apply route filtering based on ES-Import community. Thus, Ethernet Segment route is imported only by the PEs that are multi-homed to the same Ethernet segment
  - ES-Import extended community is not the same as the Route Target (RT) extended community



Route Type specific encoding of E-VPN NLRI

RT-4 Ethernet Segment Route

```
R36#show bgp l2vpn evpn rd 3.3.3.36:0 [4][0036.3700.0000.0000.1100][32][3.3.3.36]/128

BGP routing table entry for [4][0036.3700.0000.0000.1100][32][3.3.3.36]/128, Route Distinguisher: 3.3.3.36:0
Versions:
 Process         bRIB/RIB SendTblVer
 Speaker             82835     82835
Last Modified: Oct 14 21:32:13.399 for 05:52:37
Paths: (1 available, best #1)
 Advertised to update-groups (with more than one peer):
   0.2
 Path #1: Received by speaker 0
 Advertised to update-groups (with more than one peer):
   0.2
 Local
   0.0.0.0 from 0.0.0.0 (3.3.3.36)
     Origin IGP, localpref 100, valid, redistributed, best, group-best, import-candidate, rib-install
     Received Path ID 0, Local Path ID 1, version 82835
     Extended community: EVPN ES Import:3637.0000.0000 DF Election:00:0:00
```

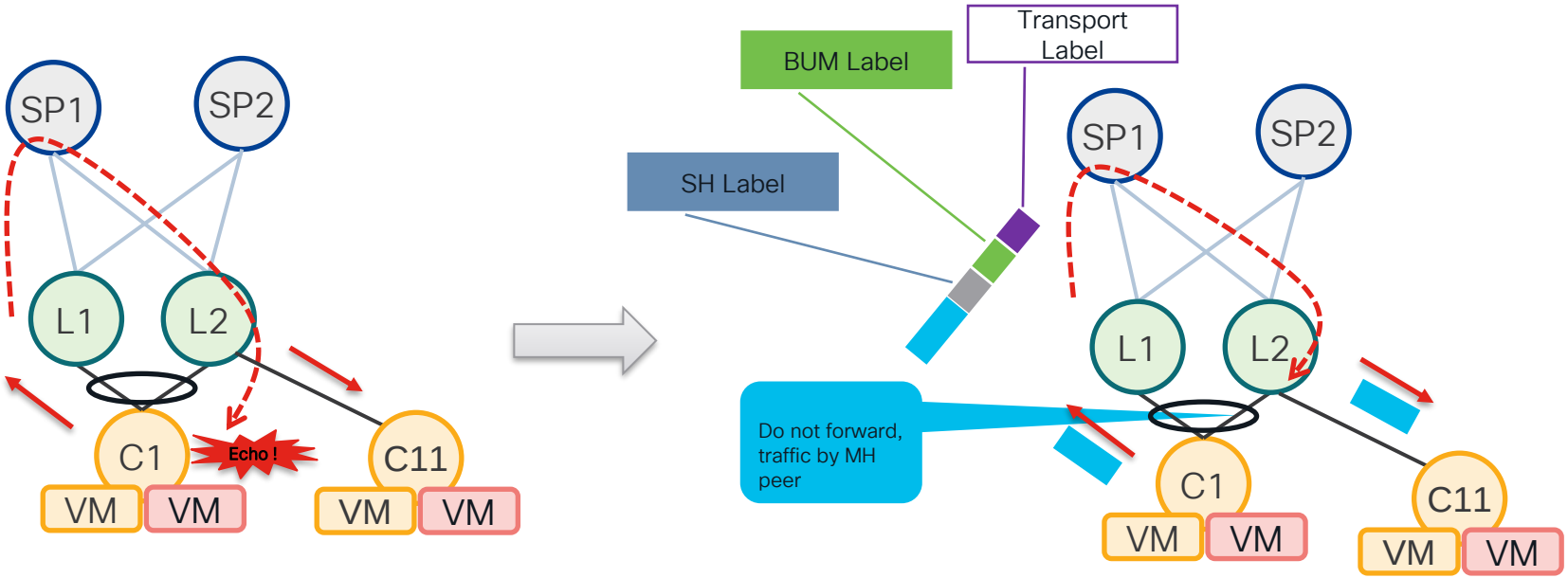
RT-4

Ethernet Segment Identifier (ESI)

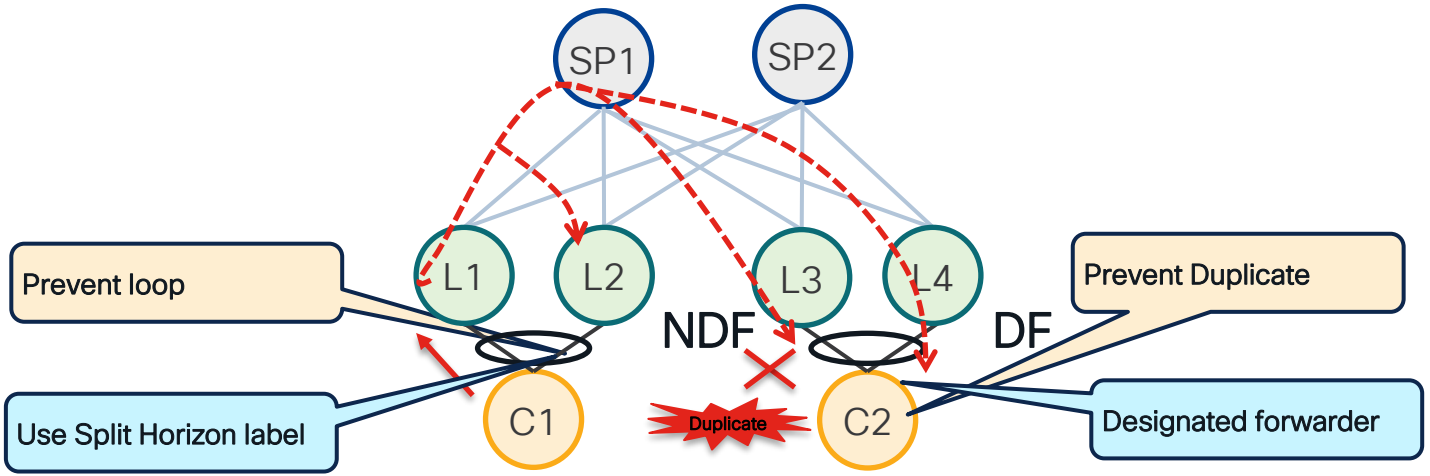
Nodes which share same ESI import this route

EVPN – Split Horizon

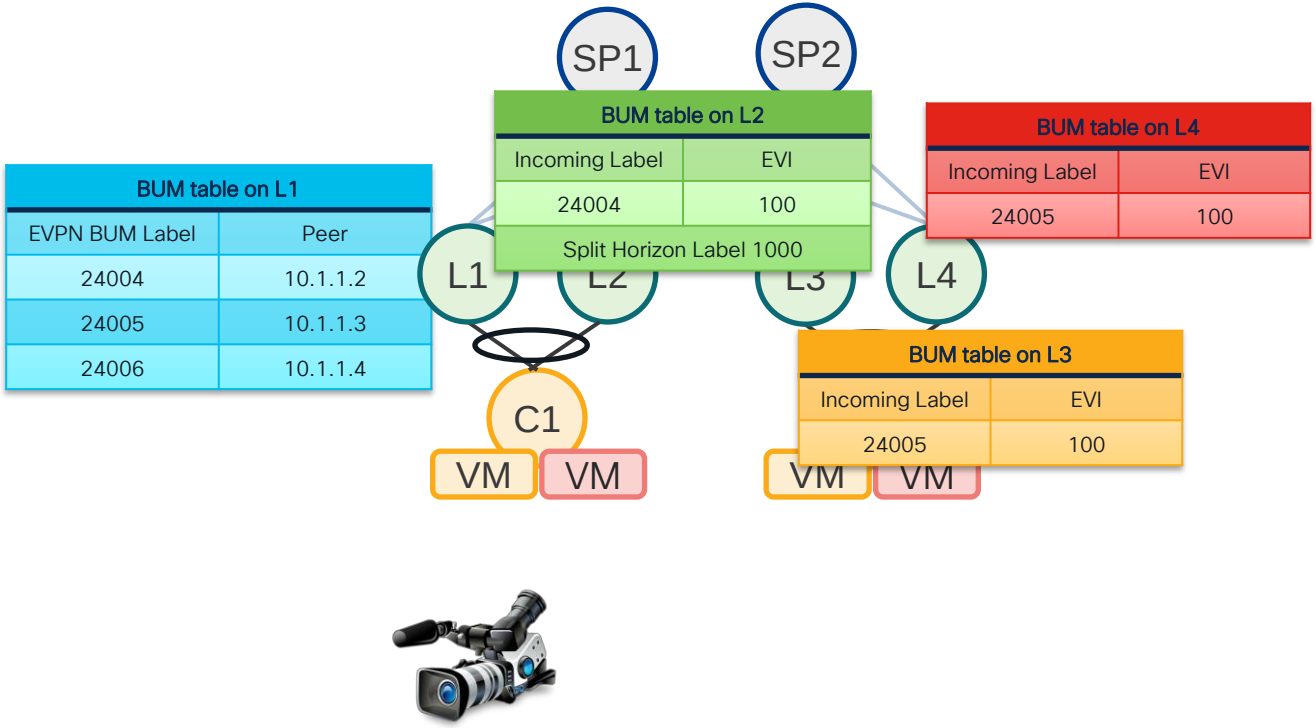
Challenge:
How to prevent flooded traffic from echoing back to a multi-homed Ethernet Segment?



Multihome Solution recap

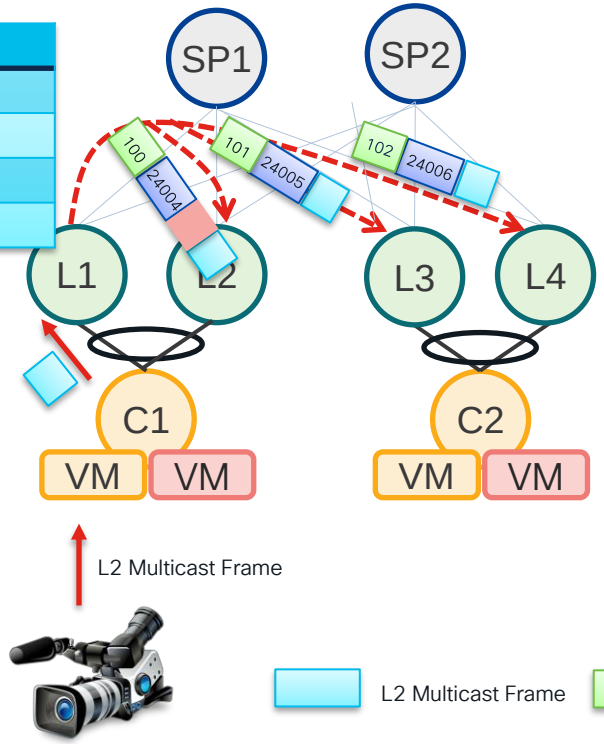


Overlay Control plane Setup



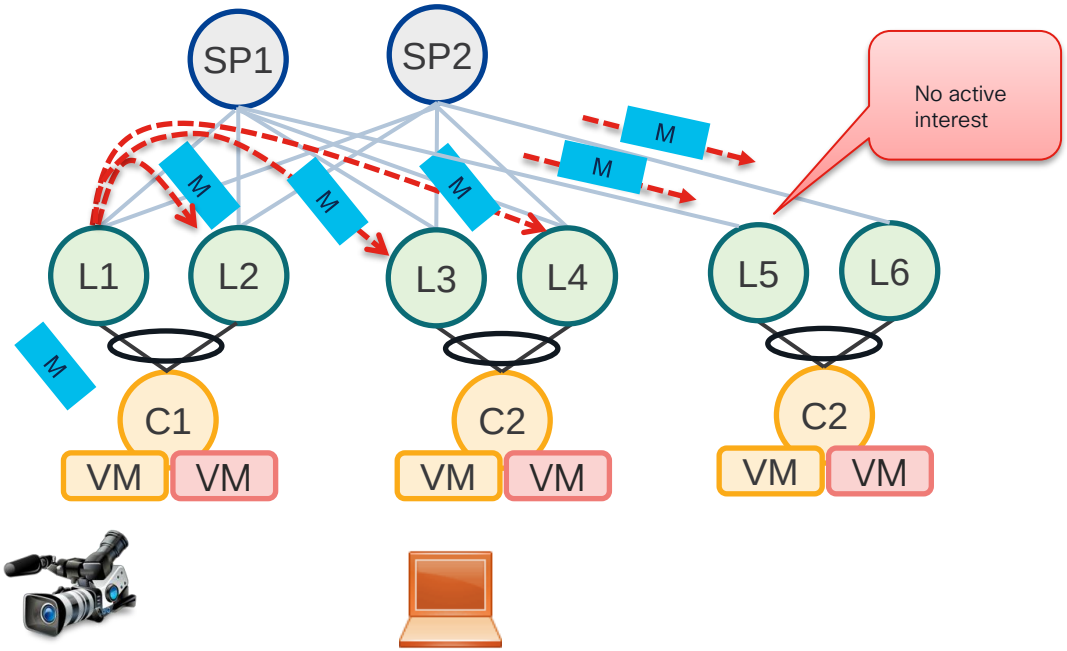
Multicast Data plane

BUM	
EVPN BUM Label	Peer
24004	10.1.1.2
24005	10.1.1.3
24006	10.1.1.4



L2 Multicast Frame Transport Label BUM Label Split Horizon

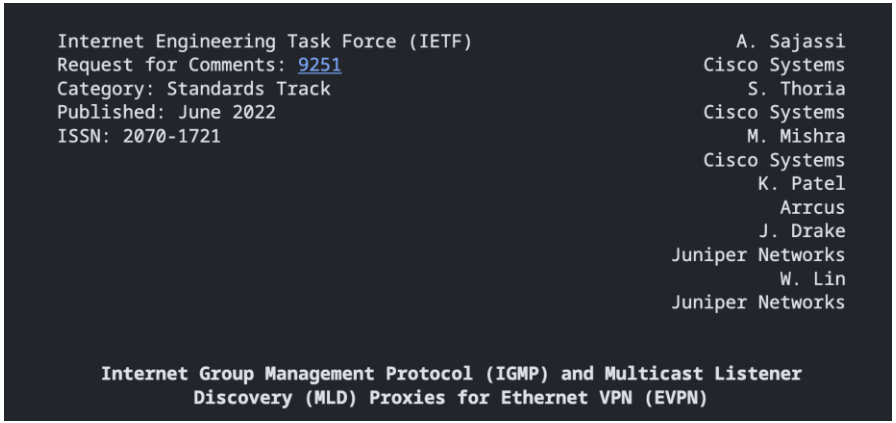
Multicast Traffic Challenges in EVPN with flood model



1. **L5 / L6 devices** receive multicast traffic **even without active receivers**, leading to unnecessary load.
2. In networks with **hundreds of Layer-2 extensions**, the **core can be flooded** with multicast traffic **regardless of interest**, impacting efficiency and scalability.

Optimizing multicast flow in EVPN Fabric

- RFC 9251 Defines a **Selective Multicast** approach to optimize multicast traffic distribution.
- Each **Provider Edge (PE)** can
  - Announce **capabilities** to **pull multicast traffic on demand**.
  - Continue **flooding BUM traffic** as per standard behavior.
  - Ensure **only multicast traffic** is selectively delivered **based on interest**.

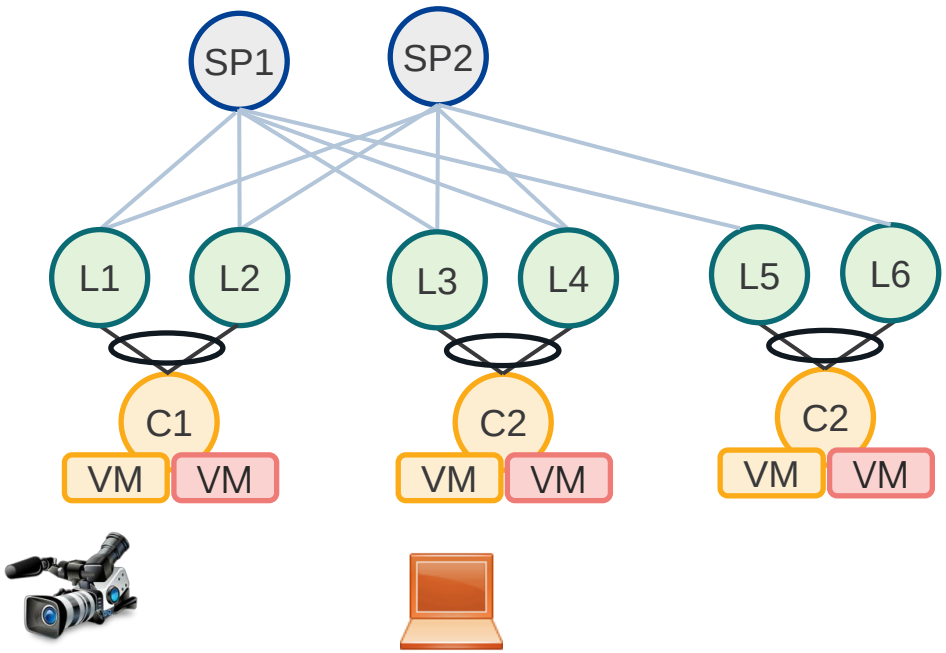


Internet Engineering Task Force (IETF)
Request for Comments: 9251
Category: Standards Track
Published: June 2022
ISSN: 2070-1721

A. Sajassi
Cisco Systems
S. Thoria
Cisco Systems
M. Mishra
Cisco Systems
K. Patel
Arrcus
J. Drake
Juniper Networks
W. Lin
Juniper Networks

Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Proxies for Ethernet VPN (EVPN)

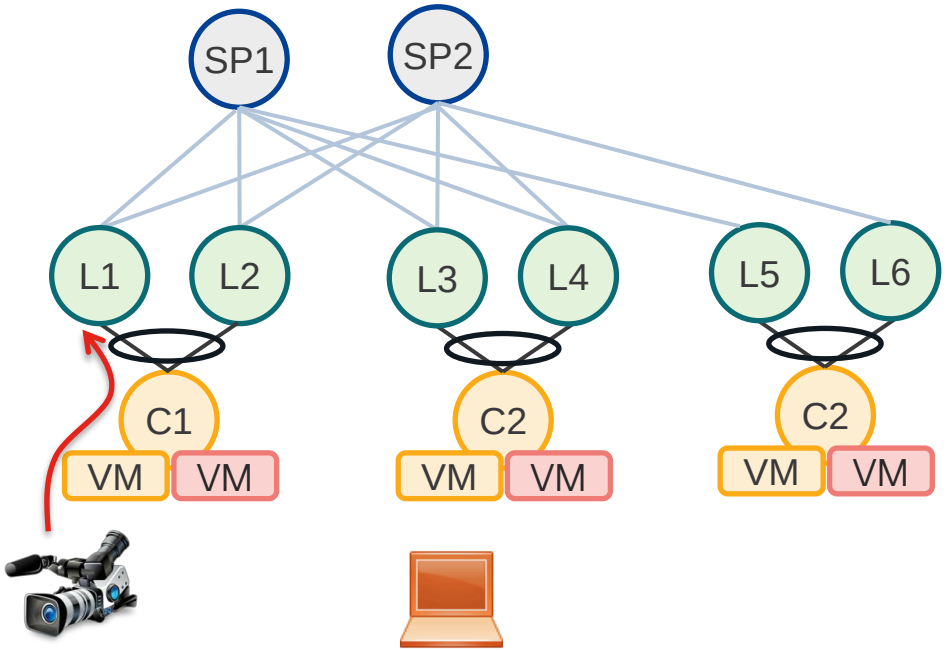
Initial Control plane setup with selective multicast



1. Each **EVPN PE** announces a **flag** with **EVPN Route Type 3 (RT-3)** to indicate **support for selective multicast**.
2. **By default, multicast traffic is blocked** for all PEs, while **only Broadcast and Unknown Unicast (BU) traffic is flooded**.

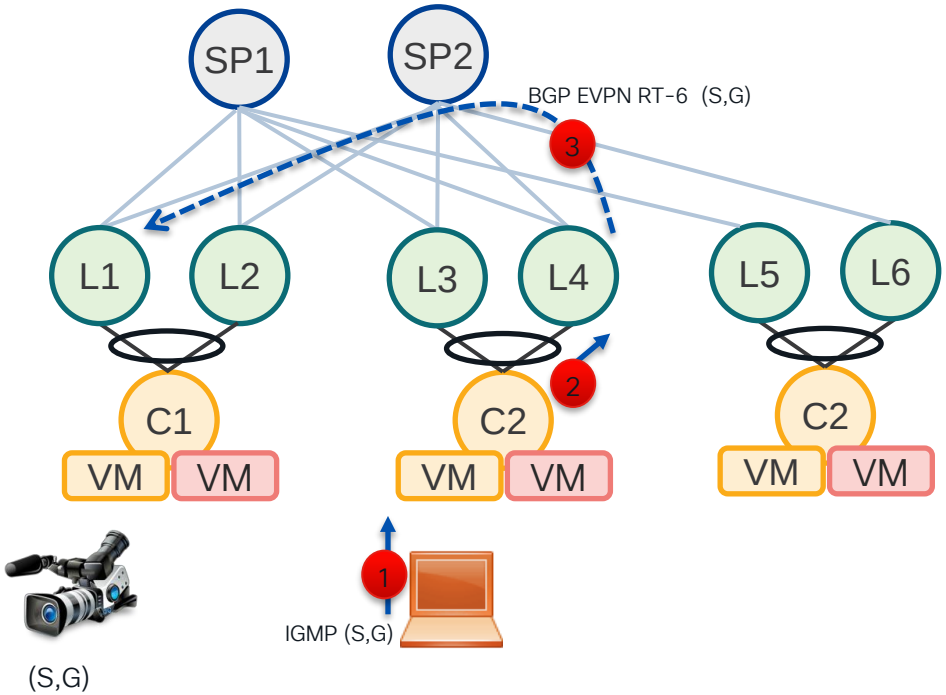
BUM Forwarding table at L1	
BU (Broadcast , Unknown unicast)	L2
	L3
	L4
	L5
	L6
Multicast	Block all

Data Plane



BUM Forwarding table at L1	
BU (Broadcast , Unknown unicast)	L2
	L3
	L4
	L5
	L6
Multicast	Block all

Local Multicast interest

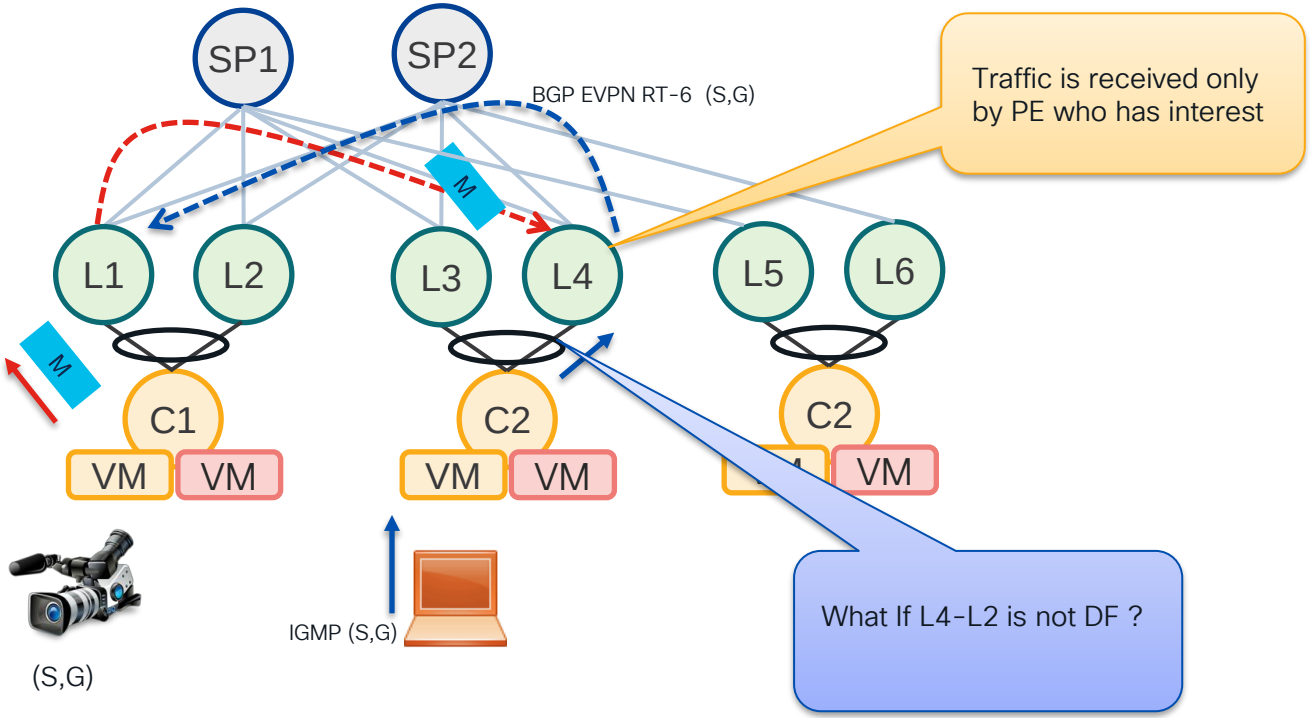


1. Local receiver initiates IGMP join for flow (S,G)
2. C2 has port channel, it hashes join to L4
3. L4 originates selective multicast
4. L1 updates its forwarding table to allow (S,G) to be forwarded to L4

BUM Forwarding table at L1

BU (Broadcast , Unknown unicast)	L2
	L3
	L4
	L5
	L6
Multicast 4	(S,G) L4
	Block All

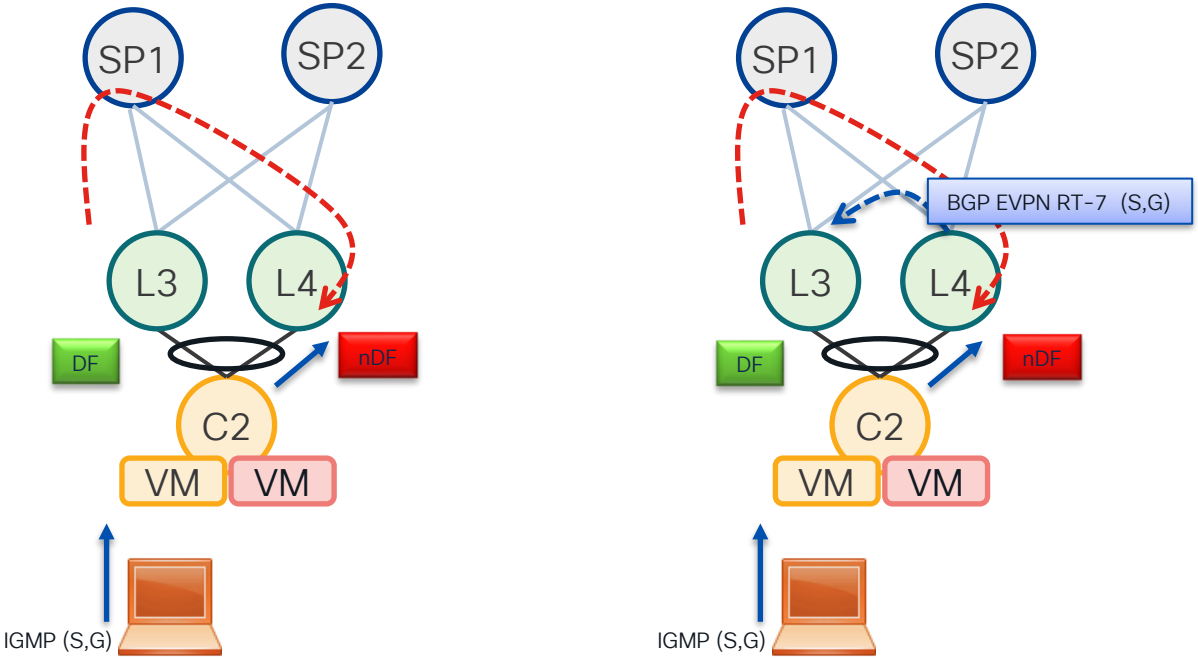
Data plane operation



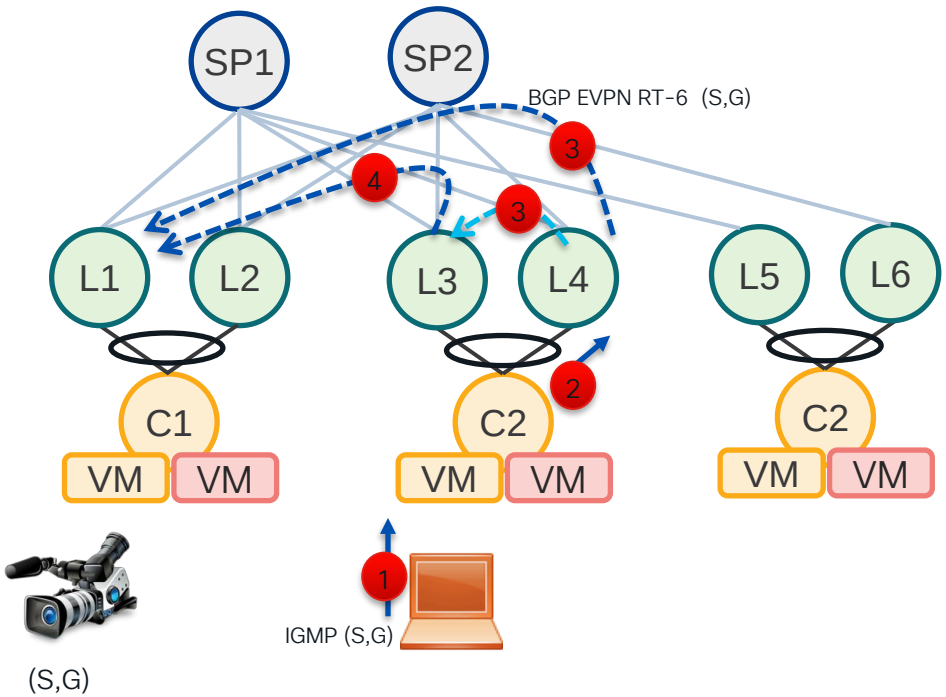
BGP-Based Signaling for IGMP Sync (RFC 9251 - Section 9.2)

- Defines a BGP-based signaling mechanism using EVPN AFI-SAFI to synchronize IGMP state between multihomed peers.
- Ensures that each multihomed peer maintains an identical multicast state.

Problem with All active multihome receiver



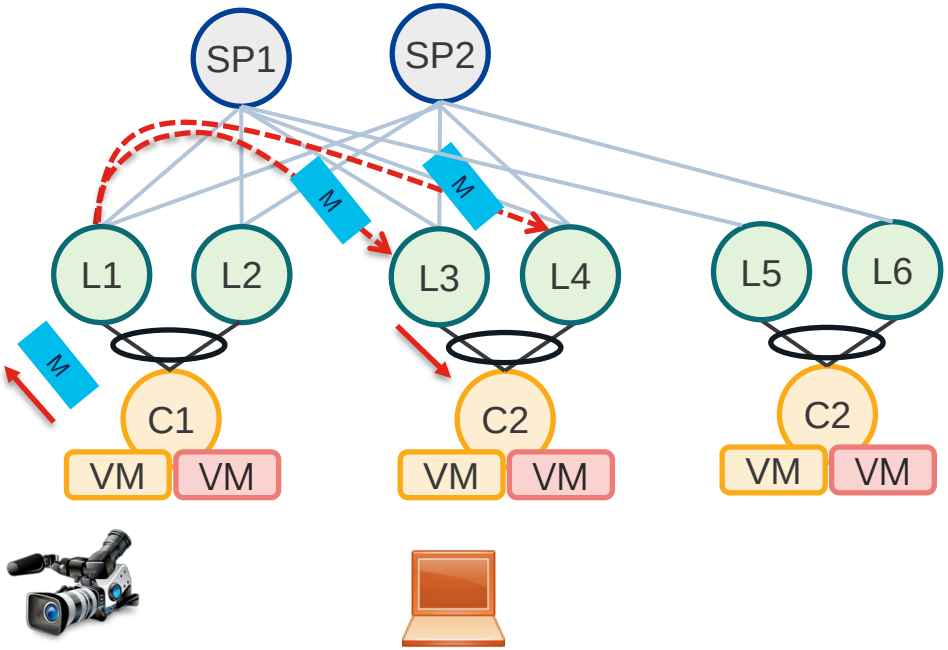
Local Multicast interest , Join Sync and Remote interest



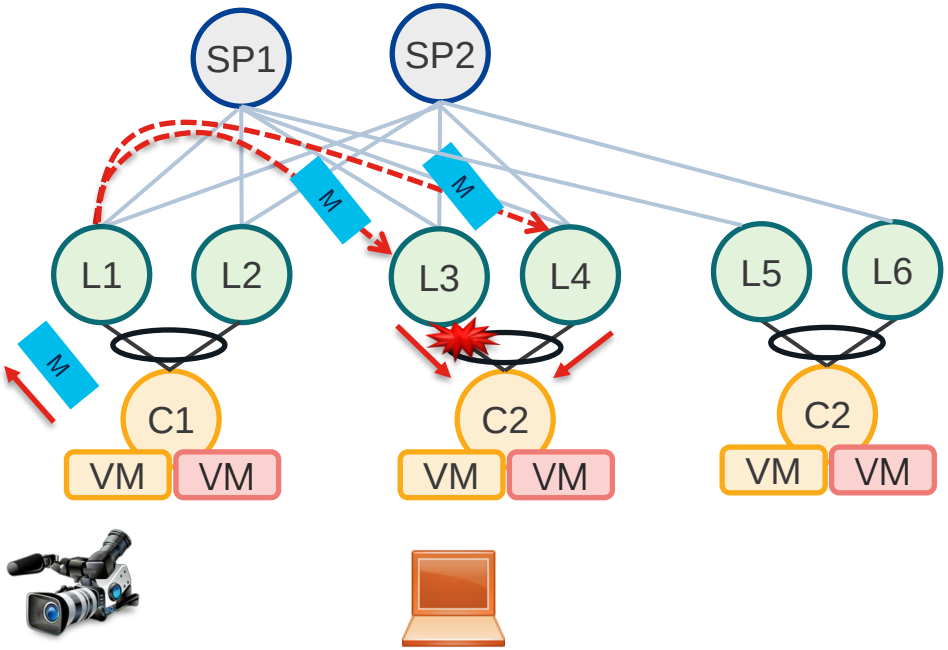
1. Local receiver initiates IGMP join for flow (S,G)
2. C2 has port channel, it hashes join to L4
3. L4 originates selective multicast and join sync to peer
4. L3 also originates selective multicast join
5. L1 updates its forwarding table to allow (S,G) to be forwarded to L4

BUM Forwarding table at L1	
BU (Broadcast , Unknown unicast)	L2
	L3
	L4
	L5
	L6
Multicast 5	(S,G) L4 (S,G) L3
	Block All

Multicast data flow with Selective multicast & all active multihoming

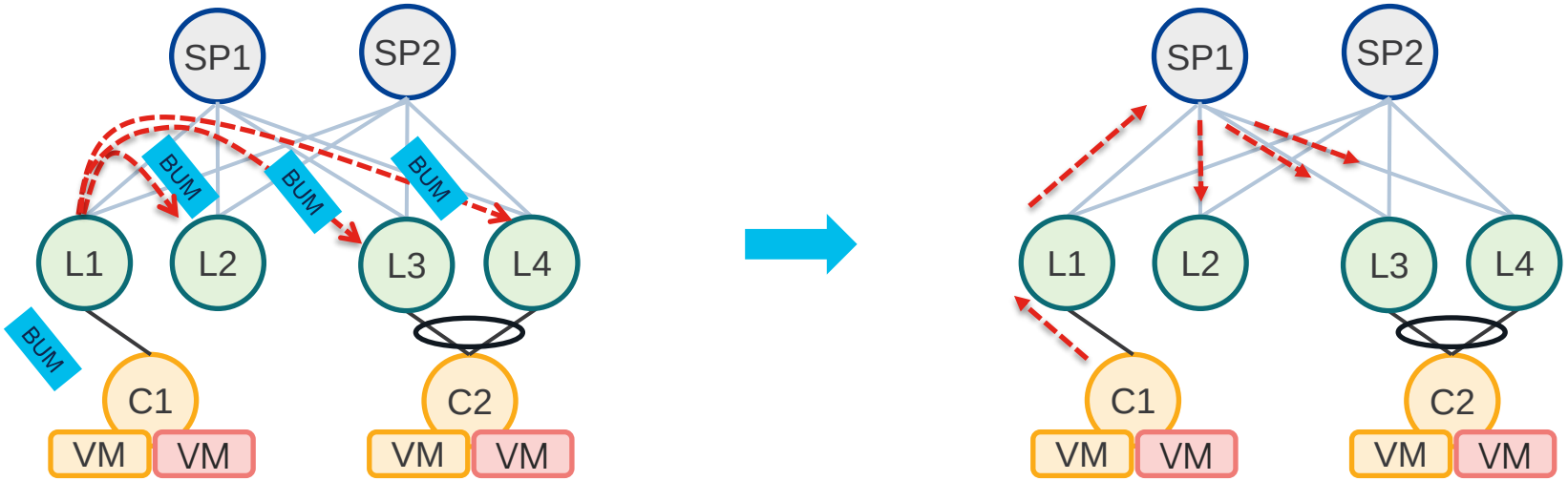


Failure and Recovery



- ES down will lead DF re-election
- As soon as new DF is elected , traffic forwarding will resume

Work In progress ……



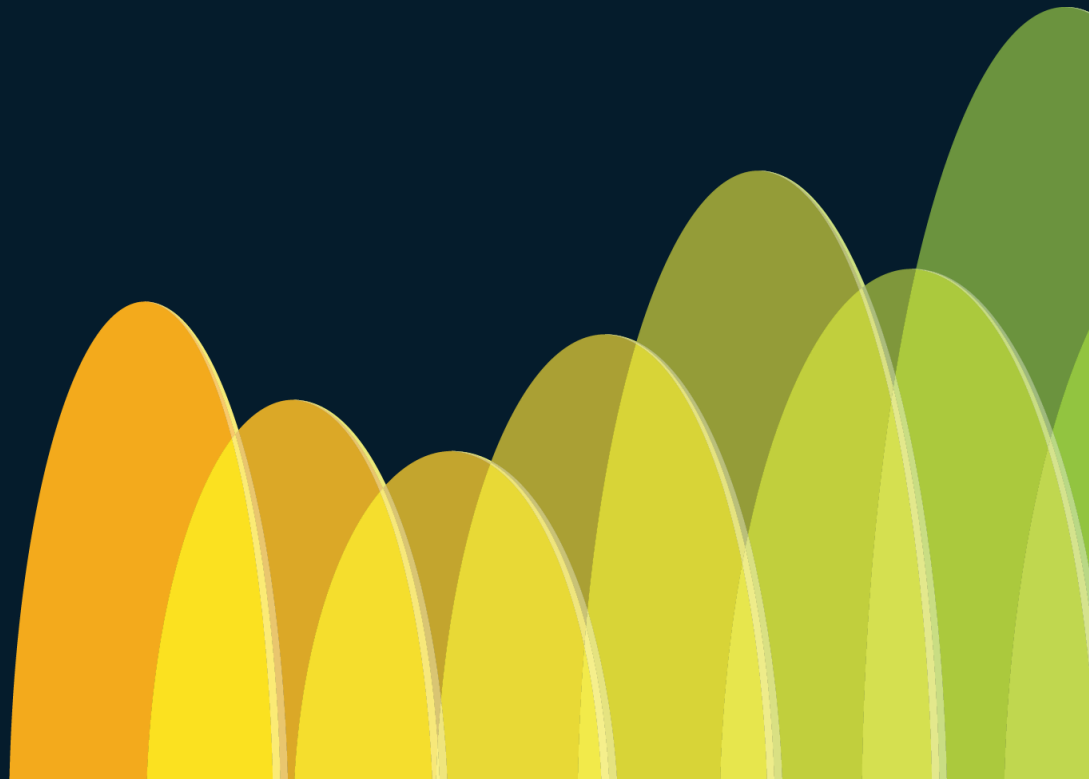
1. **Ingress replication** is **not a preferred choice** when the **Bridge Domain spans hundreds of sites**, as it becomes inefficient.
2. It leads to **excessive packet replication**, creating **hundreds of copies**, which does **not scale well**.
3. **Ongoing developments** aim to enhance **underlay multicast support** for EVPN fabrics, improving scalability and efficiency.

Recap – Key EVPN Multicast Components

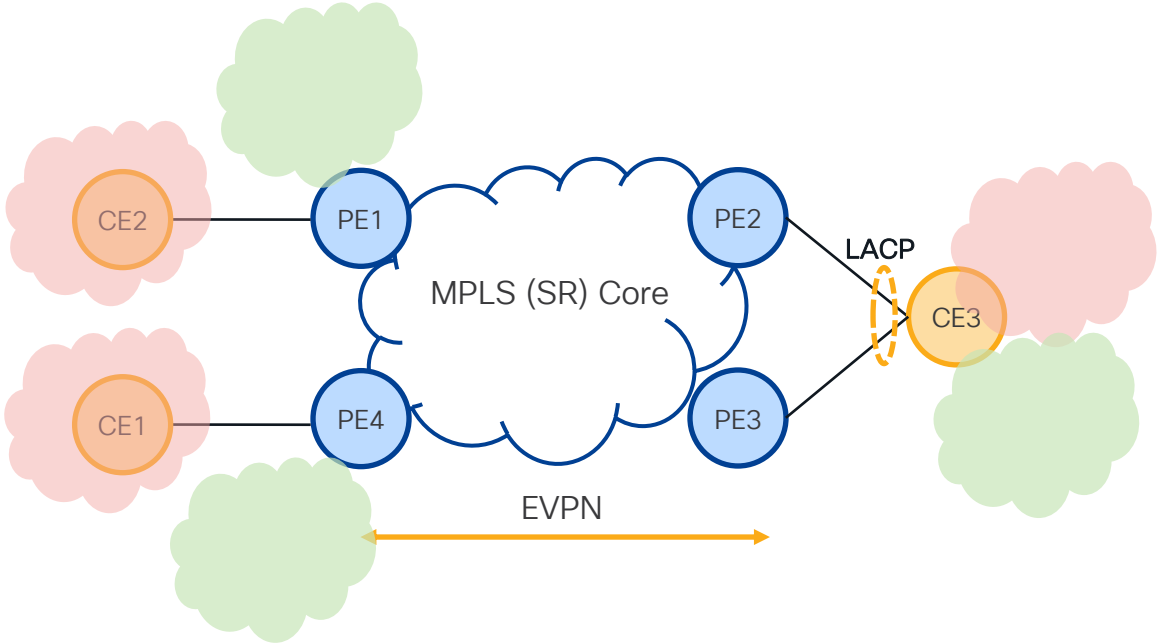
- **Inclusive Multicast Route (I-MET):** Establishes the **BUM tunnel** for Broadcast, Unk
- **Ethernet Segment Route (ES):** Facilitates **Designated Forwarder (DF) election** for all active multihoming
- **Selective Multicast:** Enables **on-demand multicast delivery**, preventing unnecessary traffic flooding.
- **IGMP Route Synchronization:** Ensures **multicast state consistency** across **all-active multihoming** setups.

How Do we use these concepts to design network

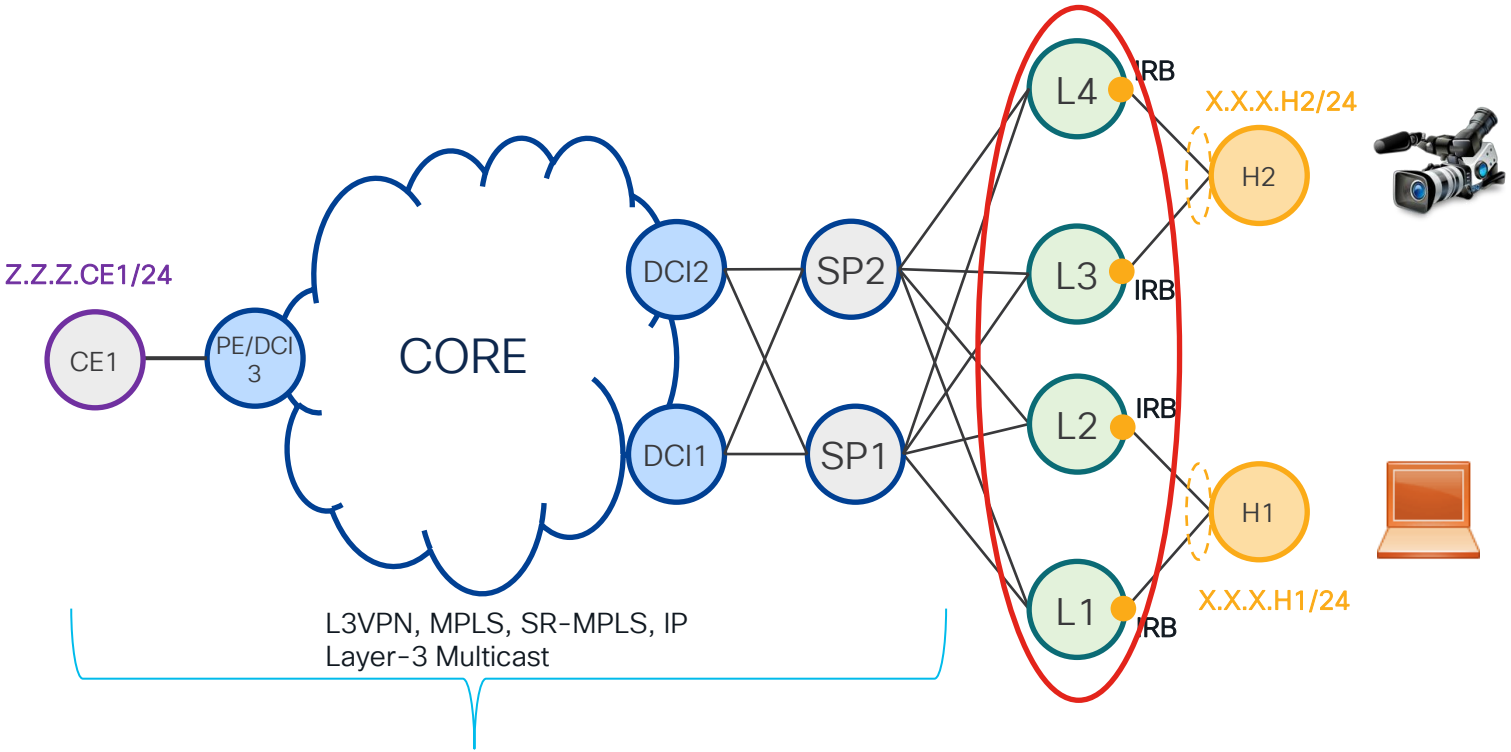
Use cases for EVPN multicast



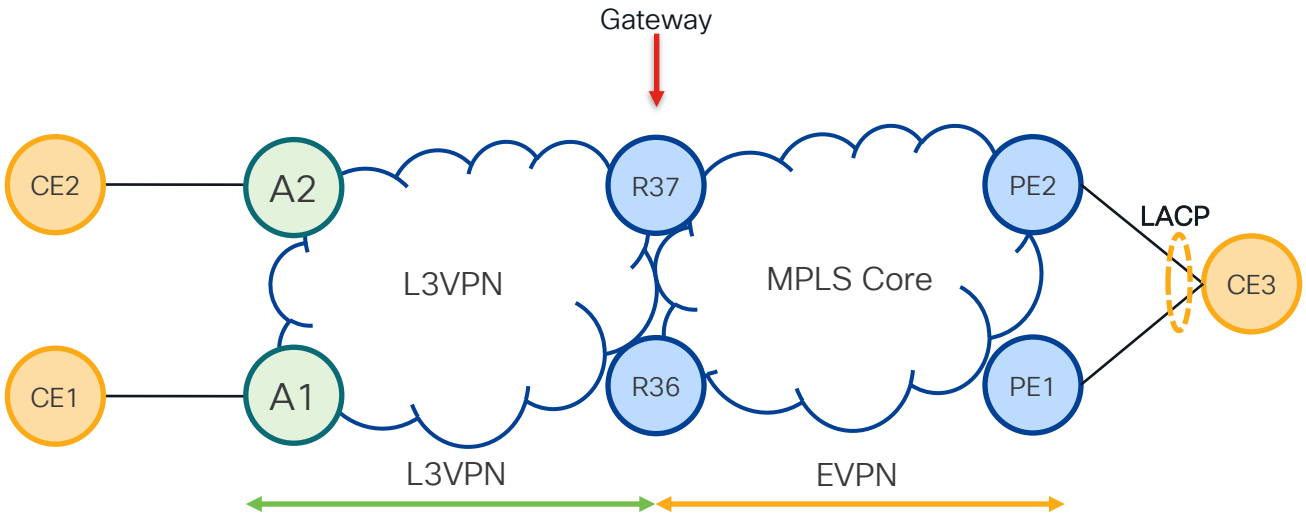
Layer-2 Stretch for wholesale network



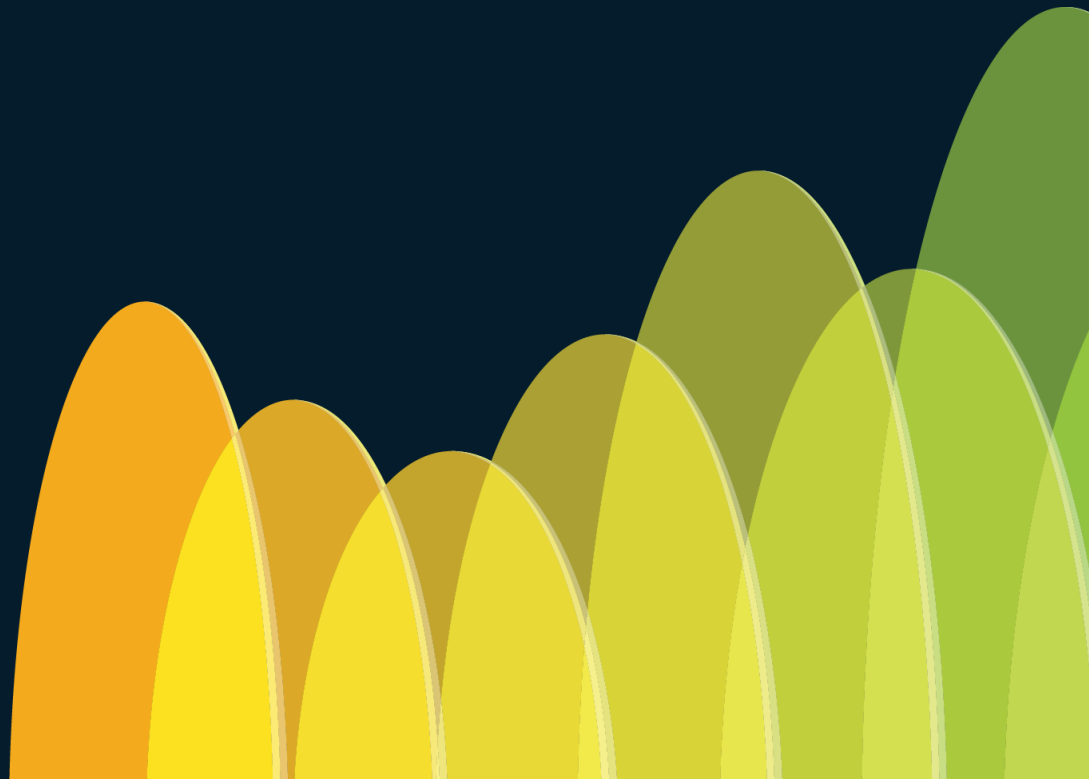
EVPN limited to access multihoming



EVPN Core with L3VPN core



Multicast with SRv6



cisco Live!

Ongoing Developments & Next Steps

- The proposal is being **presented at IETF 122 in March 2025.**
- **Implementation is in progress**—please reach out for the latest updates.
- This section covers **key concepts**, with **more details to follow in the coming months.**

Simplicity Always Prevails

~~LDP~~

~~RSVP-TE~~

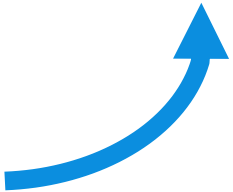
~~BGP 3107~~

~~MPLS~~

~~UDP/VxLAN~~

~~NSH~~

Furthermore, with more scale



and functionality

Simplicity Always Prevails

~~LDP~~

~~RSVP-TE~~

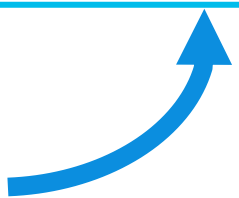
~~BGP 3107~~

~~MPLS~~

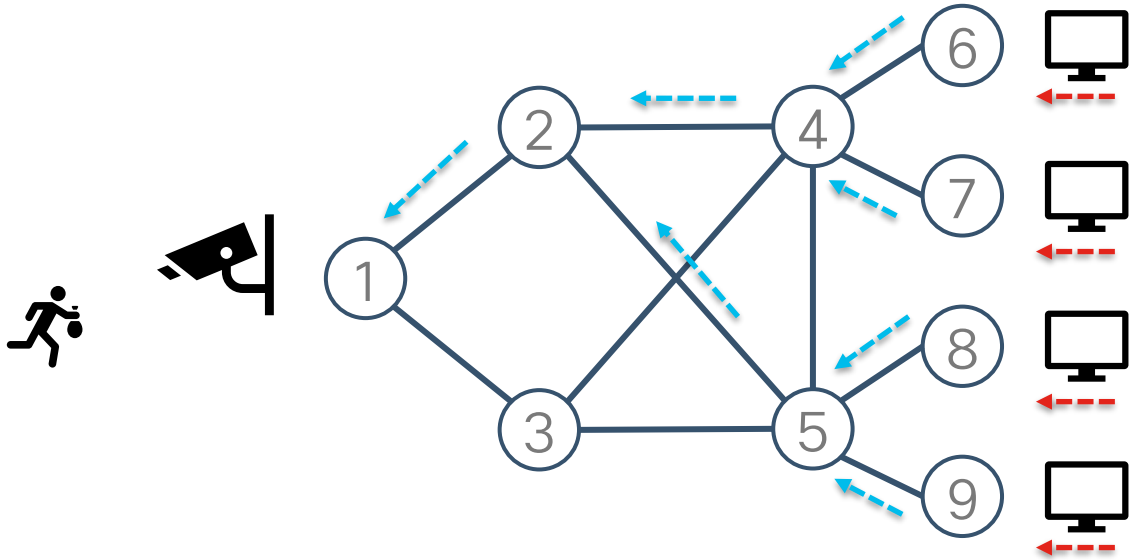
~~UDP/VxLAN~~

~~NSH~~

Continue the simplification for multicast

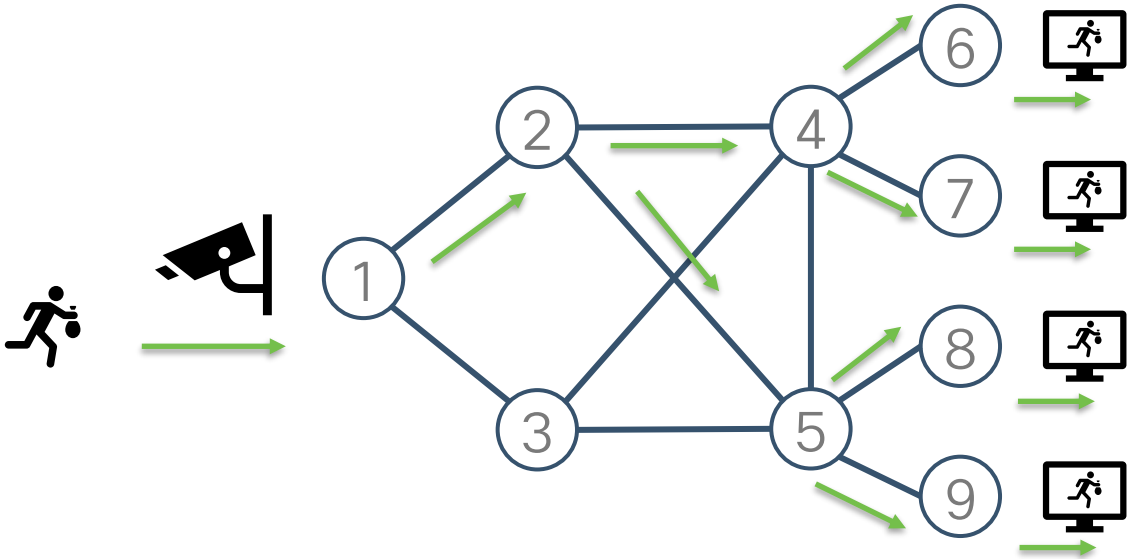


Global table Native PIMv6 Control plane



MLD (S,G) PIM join (S,G)

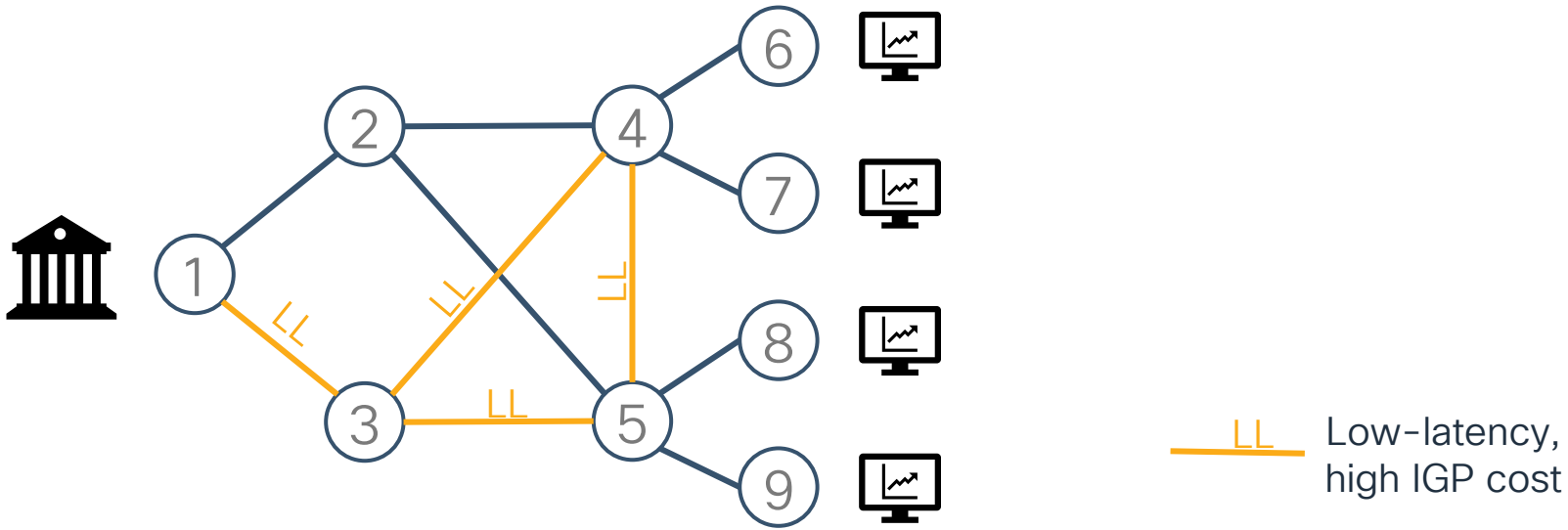
Global table Native PIMv6 Data plane



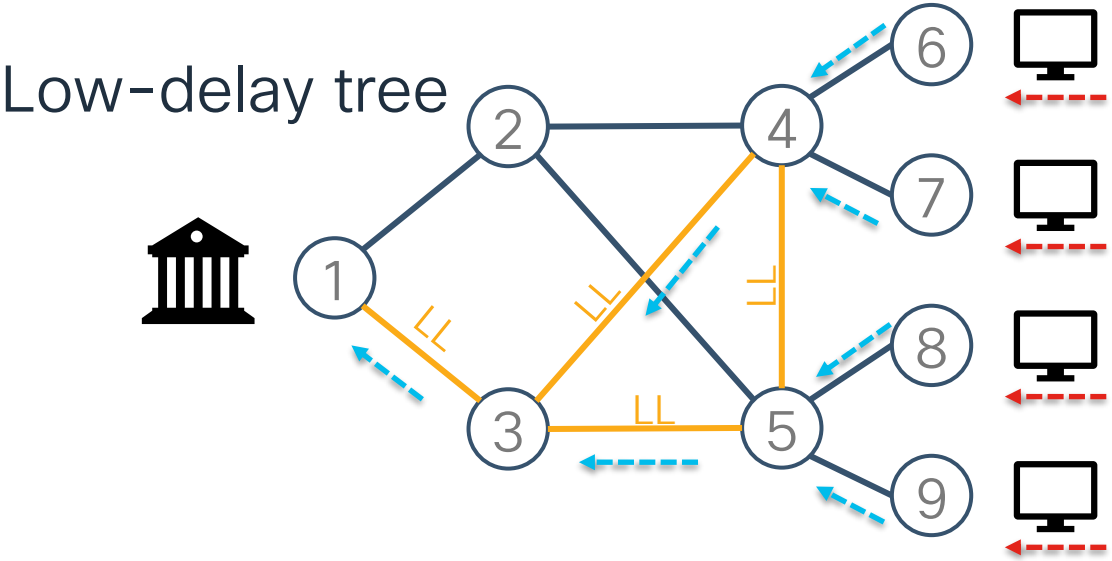
- IGP Low cost tree
- Native IPv6 multicast forwarding
- Decade old tested and proven technology

SRv6 uSID Flex Algo Tree

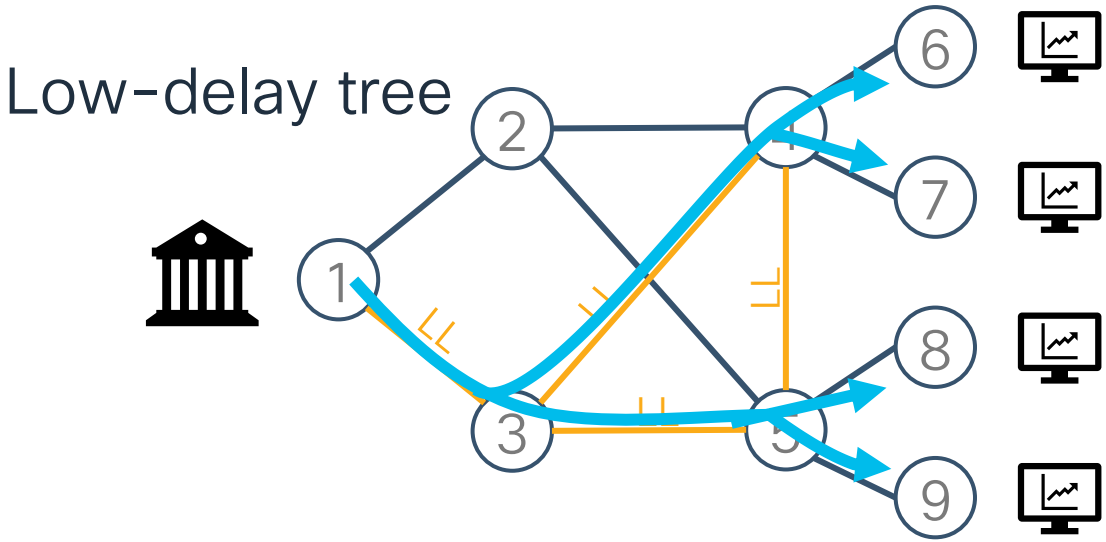
- Mcast tree along Flex-Algo topology
- Advertise Source as a Flex-Algo prefix
- Native IPv6 mcast forwarding
- Leverage all Flex-Algo TE capabilities (low-delay, affinity, …)



SRv6 uSID Flex Algo Tree Control plane

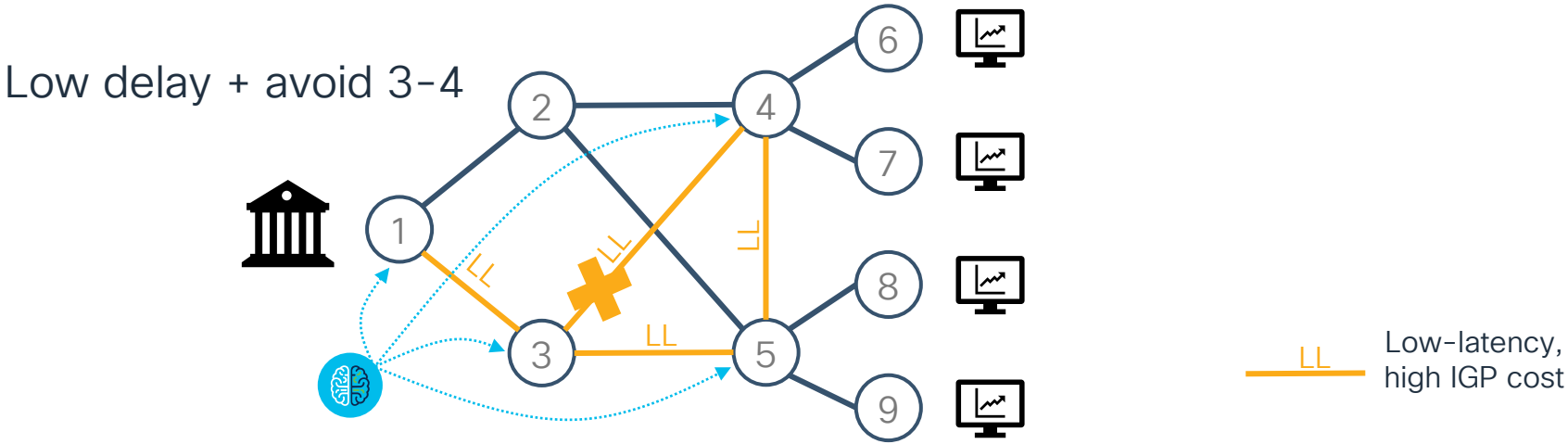


SRv6 uSID Flex Algo Tree Data plane

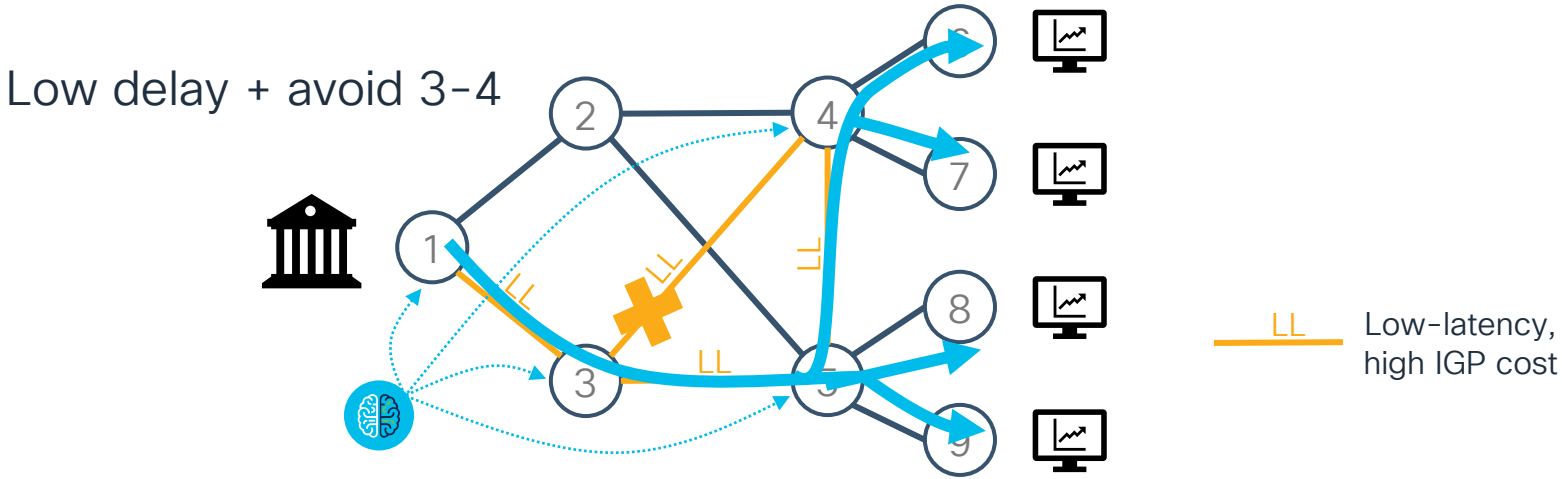


SRv6 uSID TE Tree (Controller-optimized tree)

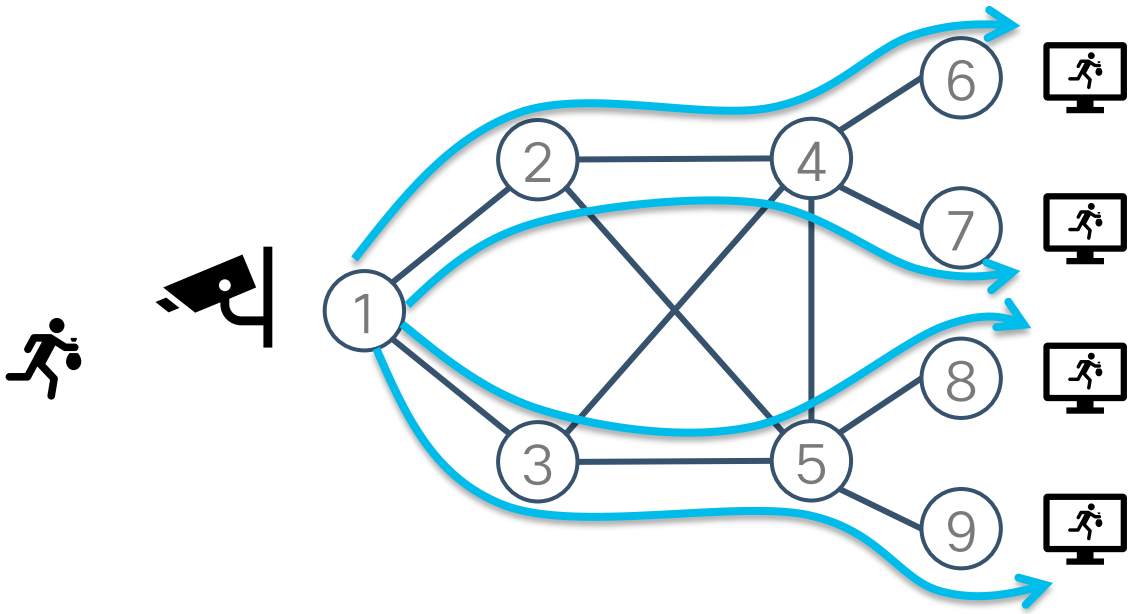
- Solves TE problems that Flex Algo cannot solve
- Controller programs native IPv6 mcast forwarding entries
  - no PIM needed
- Existing dataplane – only SW update, no forklift HW upgrade



Controller-optimized tree Data Plane



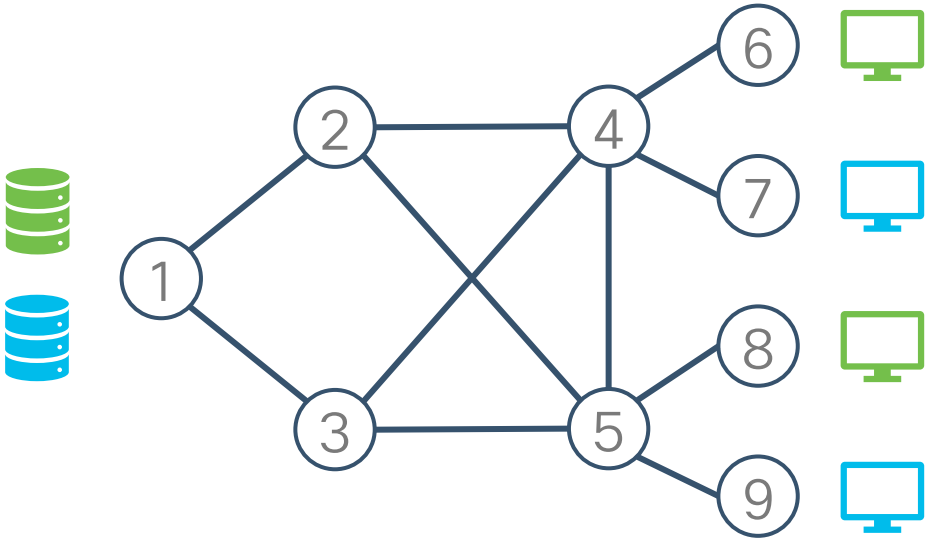
Ingress Replication



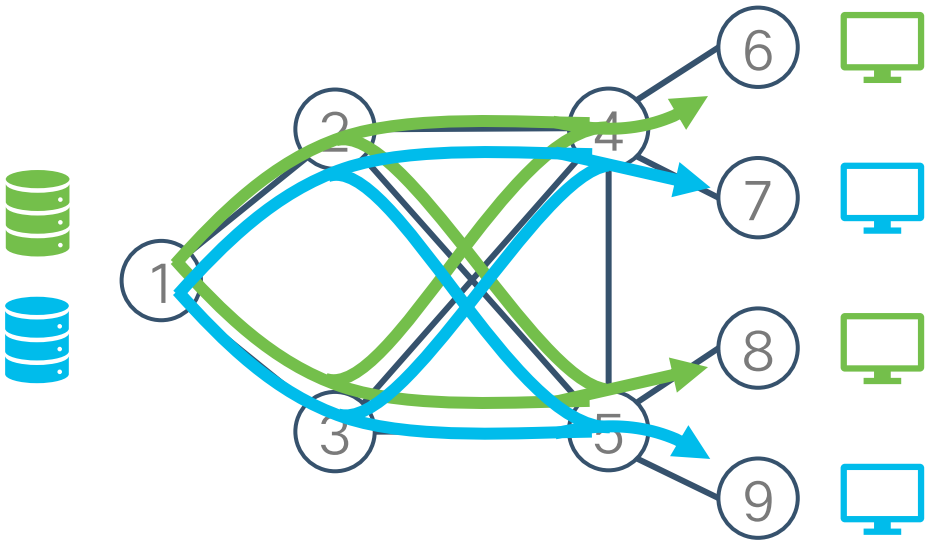
- **Ingress replication** is a viable option for **specific use cases** where the **total replication count is low**.
- The **headend PE** generates a **unicast copy** for each **tail-end PE**, encapsulating the traffic accordingly.
- **Traffic engineering policies**, similar to **unicast traffic**, can be applied to these packets for optimized forwarding.

Multicast service over VPN (mVPN)

- Core IPv6 mcast tree:
  - PIMv6
  - SRv6 uSID FA Tree
  - SRv6 uSID TE Tree
  - Ingress Replication



Multicast use cases – mVPN



Webex App

Questions?

Use the Webex app to chat with the speaker after the session

How

1. Find this session in the Cisco Events 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 February 28, 2025.

Fill Out Your Session Surveys

Participants who fill out a minimum of 4 session surveys and the overall event survey will get a unique Cisco Live t-shirt.

(from 11:30 on Thursday, while supplies last)

All surveys can be taken in the Cisco Events mobile app or by logging in to the Session Catalog and clicking the ‘Participant Dashboard’

Content Catalog

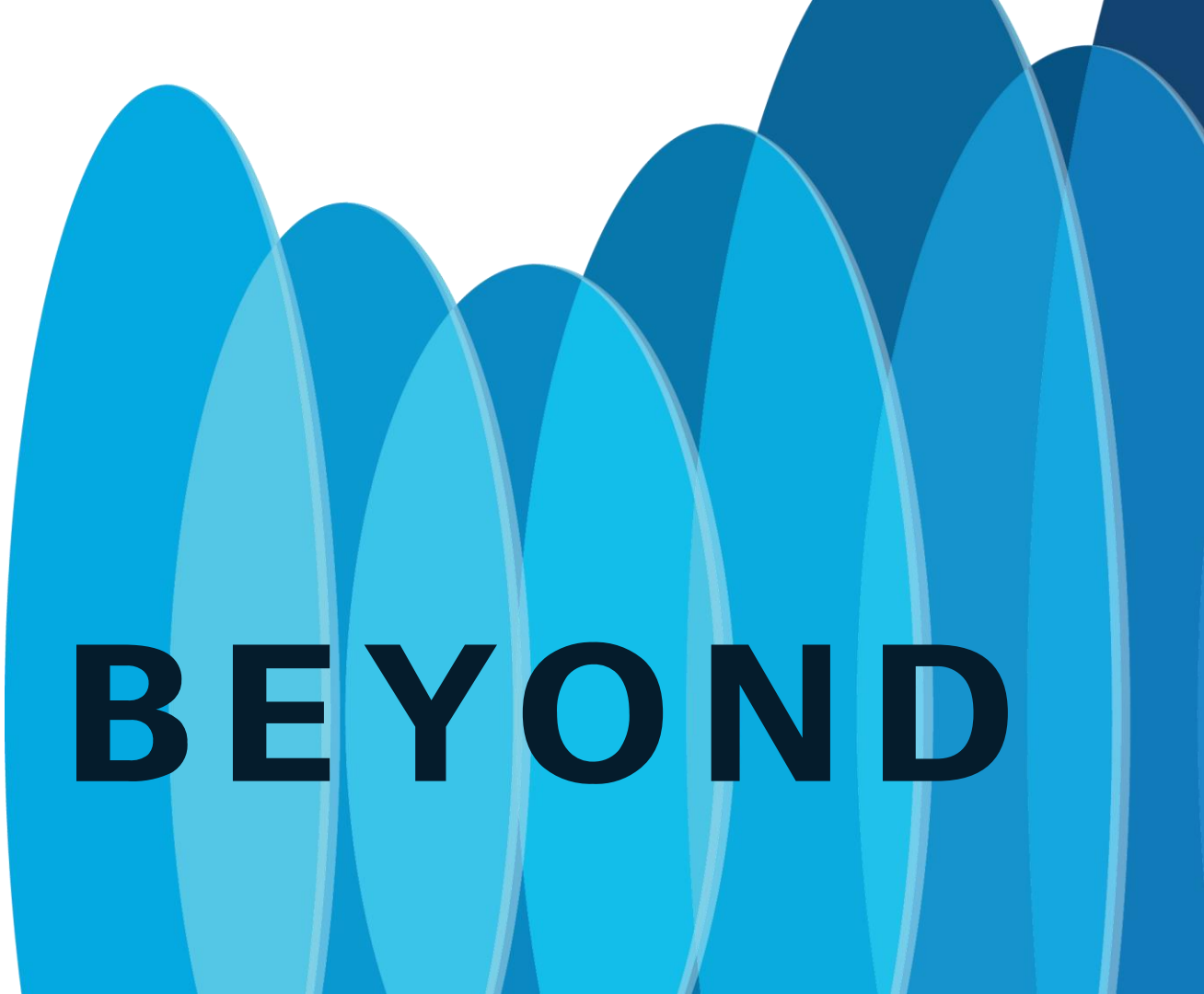
Continue your education

- 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 ciscolive.com/on-demand. Sessions from this event will be available from March 3.

Thank you

cisco Live!

cisco Live!



GO BEYOND